

**Investigation of Mobile Genetic Elements
and Antimicrobial Resistance Genes in
Human Oral Metagenomic DNA**

Thesis submitted by
Supathep Tansirichaiya

For the degree of
DOCTOR OF PHILOSOPHY
University College London

Department of Microbial Diseases

UCL Eastman Dental Institute

256 Gray's Inn Road

London WC1X 8LD

UK

Declaration

I, Supathep Tansirichaiya, hereby certify that the work embodied in this thesis is the results of my own investigations, except where otherwise stated. Investigation into antimicrobial resistance genes in the human oral metagenome was carried out in collaboration with Dr Liam Reynolds and Mr Gianmarco Cristarella (Department of Microbial Diseases, UCL Eastman Dental Institute).

Abstract

Antibiotic resistance is currently one of the major global healthcare problems. Bacteria can become resistant by acquiring resistance genes from other bacteria. This process is usually facilitated by mobile genetic elements (MGEs), a type of DNA that can move from one site to another site within bacterial genome, and often between bacterial cells. The human oral cavity has been shown to harbour various antimicrobial resistance genes (ARGs). The aim of this research is to study the fundamental biology and the association between MGEs and ARGs present in human oral bacteria by both sequence and functional-based metagenomic assays.

Using a PCR-based method, various genes predicted to confer antimicrobial resistance and other adaptive traits were identified on different MGEs (composite transposons, integrons and novel MGEs called translocatable units). This is the first report that showed ARGs in the human oral cavity were associated with these MGEs, especially in integron gene cassettes (GCs).

Some of the integron gene cassettes were predicted to not contain any genes at all. They were predicted to have a regulatory function as a promoter, which could be important for the expression of other genes carried by integrons. Using an enzymatic reporter assay, it was proven that one of the functions of these GCs is as a promoter, which could allow bacteria to survive multiple stresses within the complex environment of the human oral cavity.

Functional screening of a metagenomic library identified a clone that can confer resistance to two commonly used antiseptics agents. This was shown

to be a result of UDP-glucose 4-epimerase enzyme derived from a common oral bacteria *Veillonella parvula*, which altered the cell's surface charge to be more positive, presumably reducing the binding of positively charged antiseptics to the bacteria.

To tackle the antibiotic resistance problems effectively, the understanding of the nature of MGEs is crucial. We have shown the presence of multiple novel MGEs, ARGs and a novel resistance mechanism. Those detected ARGs can be used for the surveillance and increase the understanding of MGEs in other environments.

Impact statement

A fundamental understanding of the evolutionary pressures, which selects for resistance, is a prerequisite to design strategies to stop the spreading of antibiotic resistance genes (ARGs). The work in this study has shown an association of ARGs and various mobile genetic elements (MGEs) (integrons, composite transposons and translocatable units) in the human oral cavity. These associations suggest the possibility for these ARGs to be spread among the oral bacteria and also to a broader bacterial population in other environments.

We have also proven that one of the functions of noncoding integron gene cassettes (GCs), a type of GCs predicted to contain no genes, is to act as a promoter in a GC array. These GCs could allow bacteria to express more than one GCs in response to multiple stresses, which could contribute greatly to our understanding of why these MGEs are so successful in bacteria, even when antibiotics are absent.

A novel resistance mechanism against common antiseptics, quaternary ammonium compounds (QACs), was also shown here for the first time to be conferred by an expression of a heterologous housekeeping gene *galE* in *E. coli*, which altered the cell polarity, subsequently reducing the binding of cationic antiseptics.

As part of this work, numerous methods and reporter systems were also designed and constructed for the detection of MGEs and their associated

ARGs from human oral metagenomic DNA, which could be used to screen for ARGs in a range of metagenomes from diverse samples.

Acknowledgements

I would like to thank my supervisor Dr Adam Roberts and Prof Peter Mullany for their constant encouragement, helpful discussions and guidance throughout my PhD. Their continual support and excellent mentorship were truly inspiring.

I would like to thank Dr Liam Reynolds, Mr Md. Ajijur Rahman, Ms Deena Al Harbi, Mr Gianmarco Cristarella, Ms Agata Antepowicz and Ms Nikoleta Antou for the work they have performed for this project. I would like to thank Professor Dariusz Bartosik and Dr Magdalena Szuplewska (University of Warsaw) for providing *Escherichia coli* strain containing pAK1 entrapment vector, and Dr Morgana Vianna for the 16s rRNA sequencing results.

My thanks also go to all of my colleagues in the Department of Microbial Diseases, especially Dr Haitham Hussain for the help in the lab, Ms Supanan Sucharit, Mr Mehmet Davrandi, Ms Marika Rai and my Starbucks buddies: Ms Asyura Amdan, Ms Norashirene Jamil for their countless help, friendship, discussions and being a stress reliever. My special thanks go to Ms Nichamon Chaianant, the Thai Eastmaners, the Dota team and all of my friends whose friendship and support has helped me survive my PhD!

Lastly, none of this work could have been happen without my family: Papa, Mama, Sister and Brother, so I would like to thank them who always be there and support me (both emotionally and financially) throughout the four years of my PhD and my life in general.

Table of Contents

Title page.....	i
Declaration	ii
Abstract	iii
Impact statement.....	v
Acknowledgements	vii
Table of Contents	viii
List of Tables	xix
List of Figures.....	xxi
List of Equations	xxviii
Abbreviations.....	xxix
Chapter 1 Introduction	1
1.1 Antibiotic Resistance	2
1.1.1 The antibiotic resistance crisis	2
1.1.2 Mechanisms of resistance	5
1.1.2.1 Modification of antibiotic targets	5
1.1.2.2 Inactivation of the antibiotics	7
1.1.2.3 Alteration of biological pathway	7
1.1.2.4 Reduction of drug accumulation	9
1.1.3 Types of resistance.....	10

1.1.3.1	Intrinsic resistance.....	10
1.1.3.2	Acquired resistance.....	10
1.2	Horizontal Gene Transfer.....	11
1.2.1	Transformation.....	12
1.2.2	Transduction.....	16
1.2.3	Conjugation.....	18
1.2.4	Membrane vesicle-mediated gene transfer.....	21
1.3	Mobile Genetic Elements.....	23
1.3.1	Insertion sequences.....	24
1.3.2	Transposons.....	25
1.3.2.1	Composite transposons.....	25
1.3.2.2	Non-composite transposons.....	26
1.3.3	Translocatable Units.....	28
1.3.4	Integrans.....	29
1.3.4.1	Characteristics of integrans.....	29
1.3.4.2	Recombination of gene cassettes.....	32
1.3.4.3	Expression of integron integrase and gene cassettes.....	35
1.4	Metagenomics.....	37
1.4.1	The concept of metagenomics.....	37
1.4.2	Types of metagenomic studies.....	38
1.4.2.1	Sequence-based metagenomics.....	40

1.4.2.2	Function-based metagenomics	42
1.5	The Microbiota of the Human Oral Cavity.....	44
1.5.1	Oral Microbiota.....	44
1.5.2	Antibiotic Resistance in the Oral Cavity	47
1.6	Aims of the Study	49
Chapter 2	Material and Methods	51
2.1	Sources of media, enzymes, and reagents	52
2.2	Bacterial strains, plasmids and growth conditions.	52
2.3	Sampling and processing of saliva samples.....	53
2.3.1	Subjects and sample collection.....	53
2.3.2	Extraction of saliva genomic DNA.....	54
2.4	Molecular biology techniques	55
2.4.1	Genomic DNA extraction	55
2.4.2	Copycontrol induction reaction.....	56
2.4.3	Plasmid extraction.....	56
2.4.4	Oligonucleotide synthesis	57
2.4.5	Standard Polymerase chain reaction (PCR)	57
2.4.6	Agarose gel electrophoresis	58
2.4.7	PCR purification	59
2.4.8	Gel extraction.....	59
2.4.9	Isopropanol precipitation.....	60

2.4.10	Restriction endonuclease reaction.....	61
2.4.11	Dephosphorylation reaction.....	61
2.4.12	DNA ligation reactions.....	62
2.4.13	Desalting of ligation products.....	62
2.4.14	Preparation of chemically <i>E. coli</i> competent cells.....	63
2.4.15	Transformation of <i>E. coli</i>	63
2.4.15.1	Chemical transformation.....	63
2.4.15.2	Electroporation.....	64
2.4.16	Site-directed mutagenesis.....	65
2.5	DNA sequencing reactions.....	66
2.6	DNA sequence analysis.....	66
Chapter 3 Detection of Integron Gene Cassettes in the Human Oral		
Metagenomic DNA.....		
3.1	Introduction.....	68
3.2	Materials and methods.....	69
3.2.1	Extraction of the human oral metagenomic DNA.....	69
3.2.2	Recovery of integron GCs from oral metagenome.....	69
3.2.3	Nomenclature of the recovered integron gene cassettes.....	71
3.2.4	Detection of the circular form of integron gene cassettes.....	72
3.2.5	Functional screening of integron gene cassette PCR products	73
3.3	Results.....	73
3.3.1	Verification of the human oral metagenome.....	73

3.3.2	<i>attC</i> primer design.....	75
3.3.3	Amplification of integron gene cassettes from oral metagenome.	77
3.3.4	Detection of circular forms of gene cassettes	88
3.3.5	Functional screening of gene cassette PCR products	88
3.4	Discussion	90
3.5	Conclusion.....	98
Chapter 4 Determination of the Promoter Activity of Integron Gene Cassettes from the Human Oral Metagenomic DNA		
4.1	Introduction.....	101
4.2	Materials and methods	102
4.2.1	<i>in silico</i> analysis of the human oral cavity gene cassettes.	102
4.2.2	Construction of pUC19-GC- <i>gusA</i> constructs	102
4.2.3	Construction of pCC1BAC-GC- <i>gusA</i> constructs	105
4.2.4	Development of pCC1BAC- <i>lacZ</i> α -GC- <i>gusA</i> constructs	106
4.2.4.1	Optimisation of bi-directional terminators for the pCC1BAC- <i>lacZ</i> α -GC- <i>gusA</i> constructs.....	107
4.2.4.2	Construction of pCC1BAC- <i>lacZ</i> α -GC- <i>gusA</i> constructs....	112
4.2.4.3	The use of NsiI linker to disrupt promoter formed during the construction of pCC1BAC- <i>lacZ</i> α - <i>gusA</i> plasmid	115
4.2.5	Determination of β -glucuronidase enzymatic activity.	118
4.2.6	Determination of β -galactosidase enzymatic activity.....	119

4.2.7	Statistical analysis.....	120
4.2.8	Development of detection system for integron GCs containing promoter sequences.....	120
4.2.8.1	Selection of optimal enzyme substrates for the detection	120
4.2.8.2	Recovery of promoter-containing GCs from the human oral metagenome.....	121
4.3	Results	122
4.3.1	<i>in silico</i> analysis of the promoter sequences on the non-coding gene cassettes	122
4.3.2	Determination of the promoter activity of the noncoding GCs in pUC19-GC- <i>gusA</i> constructs by using β -glucuronidase enzyme assay	125
4.3.3	Determination promoter activity of the TMB4 sample in pUC19-GC- <i>gusA</i> constructs by using β -glucuronidase enzyme assay.....	128
4.3.4	Promoter activity determination of pCC1BAC-GC- <i>gusA</i> constructs by using β -glucuronidase enzyme assay	132
4.3.5	Construction of pCC1BAC- <i>lacZα-gusA</i> plasmid.....	136
4.3.5.1	The modification of luxI bi-directional terminator	136
4.3.5.2	The disruption of promoter formed during the construction of pCC1BAC- <i>lacZα-gusA</i> plasmid.....	139
4.3.6	Promoter activity determination of pCC1BAC- <i>lacZα-GC-gusA</i> constructs by using β -glucuronidase enzyme assay	141

4.3.7	Promoter activity determination of pCC1BAC- <i>lacZ</i> α -GC- <i>gusA</i> constructs by using β -galactosidase enzyme assay	144
4.3.8	Detection of promoter-containing GCs by using Bi-Directional Promoter Detection (BiDiPD) construct.	145
4.3.8.1	Determination of substrates for <i>lacZ</i> α and <i>gusA</i> reporter genes	145
4.3.8.2	Detection of promoter-containing GCs from the oral metagenome using pBiDiPD.....	149
4.4	Discussion	159
4.5	Conclusion.....	176
Chapter 5	Culture-independent Recovery of Composite Transposons and Translocatable Units from the Human Oral Metagenomic DNA.....	177
5.1	Introduction.....	178
5.2	Materials and methods	180
5.2.1	Verification of IS elements in the human oral metagenome ...	180
5.2.2	Screening of composite transposons in the human oral metagenome	181
5.2.3	Detection of Translocatable units in the human oral metagenome	182
5.3	Results	183
5.3.1	Confirmation of IS elements in oral metagenomic DNA	183
5.3.2	Recovery of composite transposons by PCR amplification....	183
5.3.3	Amplification of putative TU structures.....	192

5.4	Discussion	193
5.5	Conclusion.....	200
Chapter 6 The Isolation of DNA from the Human Oral Metagenomic DNA		
	Using Entrapment Vectors.....	201
6.1	Introduction.....	202
6.1.1	Transposable elements (TEs).....	202
6.1.2	Entrapment vectors.....	204
6.2	Materials and methods	207
6.2.1	Construction of BACpAK entrapment vectors	207
6.2.2	Construction of BACpAK-Terminator entrapment vector.	209
6.2.3	The isolation of transposable elements by <i>cl-tetA</i> using entrapment vectors.....	211
6.2.3.1	Direct transformation of the oral metagenomic DNA into <i>E.</i> <i>coli</i> containing <i>tetA</i> -based entrapment vectors	213
6.2.3.2	Direct transformation of the oral metagenomic DNA into <i>E.</i> <i>coli</i> containing promoterless-entrapment vectors.....	213
6.2.3.3	Oral metagenomic library construction on BACpAK entrapment vectors	215
6.2.4	Identification and analysis of the tetracycline resistant colonies.	215
6.2.5	Construction of amplicons to repeat the insertions found during the screening of BACpAK-oral metagenome library	217
6.2.5.1	Circular DNA formation.....	217

6.2.5.2	Preparation of BamHI-digested antibiotic resistance genes ..	
	218
6.3	Results	219
6.3.1	Direct transformation of metagenomic DNA into <i>E. coli</i> containing pAK1 entrapment vectors	219
6.3.2	The detection of TE transposition by constructing a DNA library on BACpAK entrapment vectors.....	222
6.3.2.1	Construction of oral metagenomic DNA on BACpAK entrapment vector.....	222
6.3.2.2	Construction of new BACpAK plasmid	225
6.3.2.3	Detection of TE transposition on the oral metagenomic library constructed on new BACpAK vector (BPG-8).....	226
6.3.2.4	The detection of TE transposition on the <i>Citrobacter freundii</i> genomic library constructed in BACpAK vector	227
6.3.2.5	Detection of the transposition of TEs located in the cloning sites of BACpAK into cl-tetA selection cartridge	229
6.3.3	Direct transformation of metagenomic DNA into <i>E. coli</i> containing BACpAK-terminator entrapment vectors	229
6.3.4	Direct transformation of metagenomic DNA into <i>E. coli</i> containing promoterless-entrapment vectors.....	230
6.3.5	Illegitimate recombination on BACpAK vectors.....	230
6.3.5.1	Sequence analysis on the clones found during the screening BACpAK-oral metagenome library.....	230

6.3.5.2	Investigation of the illegitimate recombination	233
6.4	Discussion	238
6.5	Conclusion.....	249
Chapter 7 Reduced susceptibility to Quaternary Ammonium Compounds		
(QACs) is conferred by a heterologous housekeeping gene.		
7.1	Introduction.....	251
7.2	Materials and methods	252
7.2.1	Construction of the human oral metagenomic library	252
7.2.2	Preparation of antimicrobial stock solutions	253
7.2.3	Minimum Inhibitory Concentration Determination	253
7.2.4	Screening of metagenomic library and resistant clone isolation	253
7.2.5	Plasmid extraction and sequencing of resistance genes	254
7.2.6	Transposon mutagenesis.....	254
7.2.7	Subcloning of udp-4-glucose epimerase and glucose-6- phosphate isomerase genes.	255
7.2.8	Alignment of GalE amino acid sequences and protein structures	258
7.2.9	Lipopolysaccharide (LPS) extraction and purification	258
7.2.10	Cytochrome c binding assay.....	259
7.3	Results	261

7.3.1	Screening of metagenomic library against antimicrobial compounds.....	261
7.3.2	Characterisation of genes conferring CTAB resistance.	262
7.3.3	Identification of putative resistance genes by transposon mutagenesis.....	264
7.3.4	Subcloning of the putative CTAB resistance genes	265
7.3.5	MIC of the resistant clone against CTAB and CPC.....	266
7.3.6	Similarity between A10F2 GalE and <i>E. coli</i> GalE proteins	267
7.3.7	Characterisation of LPS composition of the CTAB resistance clone	270
7.3.8	Bacterial cell's surface charge determination.....	271
7.4	Discussion	273
7.5	Conclusion.....	278
Chapter 8	General Discussion.....	279
References	284
Appendices.....		326
Appendix 1.	Ethics consent form.....	327
Appendix 2.	Composition of media and solutions.....	329
Appendix 3.	Antimicrobial Working/Stock Concentrations.....	331
Appendix 4.	Primers used in this study	332
Appendix 5.	Publications resulting from this study	348

List of Tables

Table 2-1 Lists of bacterial strains and plasmids used in this study	53
Table 3-1 The primer pairs and the sequence analysis results of each library.	77
Table 3-2 Orientation of ORFs in the GCs recovered from metagenomic DNA of saliva samples.	79
Table 3-3 Complementary of the R' and R" cores sites on the recovered gene cassettes.	80
Table 3-4 Characterization of all gene cassettes detected in the saliva metagenomic DNA using <i>attC</i> -based primers.....	83
Table 4-1 The putative promoters for non-protein-coding GCs and GCs with an ORF (SSU17 and MMB3) predicted using BPRM.....	123
Table 4-2 The sequences, values of parameters and termination efficiency of <i>luxI</i> and modified <i>luxI</i> , estimated by an algorithm described by d'Aubenton <i>Carafa et al.</i> (1990).....	138
Table 4-3 Characterisation of the human oral integron GCs containing promoter sequences detected by pBiDiPD.	151
Table 4-4 Complementarity of the core sites R' (1R) and R" (1L) abutting the forward and reverse <i>attC</i> primer sequence on the gene cassettes.....	157
Table 5-1 The list of IS elements that were selected for the presence in the human oral metagenome.....	181
Table 5-2 The details of putative composite transposon structures amplifying from the human oral metagenomic DNA.	188

Table 6-1 The optimisation of the library construction on entrapment vectors.	228
Table 6-2 Details of the human DNA and <i>Veillonella parvula</i> inserts found at the same position between <i>cl</i> and <i>tetA</i> , and their BlastN and BlastX results.	232
Table 7-1 MIC breakpoints for various antimicrobials against <i>E. coli</i> EPI300 containing empty pCC1BAC vector to use for library screening.	262

List of Figures

Figure 1-1 The number of new antibiotics developed and approved since 1980-2015.	4
Figure 1-2 Mechanisms of antibiotic resistance in bacteria.	5
Figure 1-3 The mechanisms of horizontal gene transfer.	12
Figure 1-4 The processes of transformation in Gram-negative and Gram-positive bacteria.....	15
Figure 1-5 Types of transduction.	17
Figure 1-6 Electron microscope image of the of Gram-negative sex pilus in conjugation.	19
Figure 1-7 The conjugative transfer between high-frequency recombination (Hfr) and recipient cells.....	21
Figure 1-8 Membrane vesicle-mediated gene transfer.	22
Figure 1-9 Organisation of an insertion sequence (IS).	24
Figure 1-10 Organisation of composite transposon.	26
Figure 1-11 Organisation of non-composite transposon.....	27
Figure 1-12 Organization of translocatable unit (TU).....	28
Figure 1-13 Organisation of Integron and the associated gene cassettes...	30
Figure 1-14 Structure of gene cassettes and recombination sites.	33
Figure 1-15 Construction and screening of metagenomic libraries.....	39
Figure 3-1 A schematic representation of primer binding sites on integron.	70
Figure 3-2 Schematic presentation of a method to verify the gene cassette.	72

Figure 3-3 Agarose gel electrophoresis of the extracted human oral metagenomic DNA.	74
Figure 3-4 The composition and distribution of the bacterial communities in the extracted oral metagenome, based on 16S rRNA gene sequences.	75
Figure 3-5 Alignment of <i>Treponema denticola attC</i> sequences and the similar sequences retrieved from BlastN by Clustal Omega.	76
Figure 3-6 Agarose gel electrophoresis of <i>Hind</i> III digested products of pUC19 and S34tet1-4 plasmids.....	89
Figure 3-7 Schematic representation of pBR322 plasmid and the <i>Sau</i> 3AI digestion analysis.	90
Figure 3-8 Human oral microbiome hierarchy from the Human Oral Microbiome Database (HOMD).	92
Figure 3-9 The similarity of HS286/HS287 primers to <i>T. denticola attC</i> site....	94
Figure 4-1 Construction of pUC19-GC- <i>gusA</i> constructs.....	103
Figure 4-2 The amplification of the selected GCs from GC-containing pGEM-T easy vectors.	104
Figure 4-3 Construction of pCC1BAC-GC- <i>gusA</i> constructs.....	106
Figure 4-4 Bi-directional <i>rho</i> -independent terminator consisting of a stable hairpin structure, flanked by poly(A) and poly(U) regions.	108
Figure 4-5 A two-dimensional diagram showing the separation between real transcriptional terminators from intracistronic or random structures.	111
Figure 4-6 The correlation between the <i>d</i> score of some <i>rho</i> -independent terminators in <i>E. coli</i> and their efficiency <i>in vitro</i>	112
Figure 4-7 Construction of pCC1BAC- <i>lacZ</i> α -GC- <i>gusA</i> constructs.....	114

Figure 4-8 The disruption of a putative promoter located in <i>Nsi</i> I- <i>Nhe</i> I restriction sites on pCC1BAC- <i>lacZ</i> α - <i>gusA</i> plasmid..	117
Figure 4-9 The promoter activity of the noncoding GCs in pUC19-GC- <i>gusA</i> constructs estimated by β -glucuronidase enzyme assays.	127
Figure 4-10 Nucleotide sequence alignment between the P _C promoter of TMB1 and TMB4 samples.	128
Figure 4-11 The promoter activity of the TMB4 GC in pUC19-GC- <i>gusA</i> constructs estimated by β -glucuronidase enzyme assays.	129
Figure 4-12 Plasmid concentration from 1 ml of <i>E. coli</i> overnight culture (OD ₆₀₀ =3.0) containing pUC19-GC- <i>gusA</i> constructs.	130
Figure 4-13 The normalised concentration of β -glucuronidase enzyme based on the plasmid concentration of pUC19-GC- <i>gusA</i> constructs.	131
Figure 4-14 Plasmid concentration from 10 ml of <i>E. coli</i> overnight culture (OD ₆₀₀ = 1.7) containing pCC1BAC-GC- <i>gusA</i> constructs.	133
Figure 4-15 The promoter activity of the noncoding GCs in pCC1BAC-GC- <i>gusA</i> constructs estimated by β -glucuronidase enzyme assays.	135
Figure 4-16 The predicted hairpin structure of <i>lux</i> and modified <i>lux</i> terminators by Mfold.	137
Figure 4-17 The determination of termination efficiency of <i>lux</i> and modified <i>lux</i> bi-directional terminators.	139
Figure 4-18 Position of putative promoters expressing <i>lacZ</i> α predicted by BPROM.	140
Figure 4-19 The promoter activity from pCC1BAC- <i>lacZ</i> - <i>gusA</i> constructs estimated by β -glucuronidase enzyme assays.	143

Figure 4-20 The promoter activity from pCC1BAC- <i>lacZ</i> α - <i>gusA</i> constructs estimated by β -galactosidase enzyme assays.	145
Figure 4-21 The determination of substrates for the promoter detection activity with <i>lacZ</i> and <i>gusA</i> reporter genes.	147
Figure 4-22 The detection of the integron GCs by using pBiDiPD.	149
Figure 4-23 The transcriptional activity of the TMB4 GC constructs.	165
Figure 4-24 The proposed function of the promoter-containing GC in the first position to control the expression of the integrase in reverse integrons. ...	167
Figure 4-25 The proposed function of promoter-containing GC as a genetic clutch.	169
Figure 4-26 The expression level of gene cassettes with and without a genetic clutch.	170
Figure 4-27 Substrate-induced gene expression (SIGEX) screening procedure.	175
Figure 5-1 The structures of composite transposons and TUs	178
Figure 5-2 The composite transposons and translocatable units identified from the oral bacteria.	179
Figure 5-3 IS1216 and IS257 composite transposon amplicons.	184
Figure 5-4 Schematic representation of CTA1 and CTA2 structure in comparison to Tn6087.	185
Figure 5-5 Schematic representation of plasmid pL5 and CTA3-4 structures.	186
Figure 5-6 Schematic representation of plasmid SAP079A and CTA5 structure.	187
Figure 5-7 TU confirmation PCR of the sample CTA2 and CTA4.	192

Figure 5-8 The possibilities for the amplification of TU amplicons.....	195
Figure 5-9 The formation of composite transposons by IS26 TU.....	197
Figure 5-10 IS6-family composite transposons located on other MGEs. ...	198
Figure 5-11 The excision of <i>rep</i> -containing TU and the formation of plasmid-like structure.	200
Figure 6-1 Excision of the large genomic DNA from <i>Enterococcus faecium</i> D344R due to an interaction between Tn916 and Tn5386.	203
Figure 6-2 The types of entrapment vectors.....	206
Figure 6-3 Construction of BACpAK entrapment vector.	208
Figure 6-4 Construction of BACpAK-Terminator entrapment vector.....	211
Figure 6-5 The approaches for the capturing of TEs by using entrapment vectors.....	212
Figure 6-6 Schematic representation of the target site for TE insertion on pCC1BAC- <i>lacZ</i> α - <i>gusA</i> plasmid.	214
Figure 6-7 Identification and analysis of the tetracycline resistant clones by colony PCR.....	216
Figure 6-8 Sequencing results of the <i>cl-tetA</i> genes of pAK1 tetracycline resistance clones.....	220
Figure 6-9 <i>Hind</i> III digestion of the plasmids extracted from the pAK1 <i>E.coli</i> with the transposition of TEs.....	221
Figure 6-10 The <i>cl-tetA</i> PCR products amplified from the BACpAK-oral metagenome clones found on LB agar containing tetracycline.....	223
Figure 6-11 <i>Bam</i> HI digestion of BACpAK and pUC19 vectors.	225
Figure 6-12 <i>Bam</i> HI digestion on BACpAK-GE plasmids.	226
Figure 6-13 The insertion of Human DNA and <i>Veillonella parvula</i> inserts.	231

Figure 6-14 Predicted model for the insertion of the BPPB1 insert on BACpAK entrapment vector.....	234
Figure 6-15 The transformation of ampicillin resistance gene (<i>AmpR</i>) into <i>E. coli</i> competent cells.	237
Figure 6-16 Homology-facilitated illegitimate recombination (HFIR).....	242
Figure 6-17 Illegitimate recombination of DNA fragment ending with GATC sequences into chromosomal DNA.	245
Figure 6-18 The DNA secondary structures predicted by Mfold on BACpAK entrapment vector.....	247
Figure 6-19 The DNA secondary structures predicted by Mfold on GATC inserts. The red boxes indicate the location of GATC sites.	248
Figure 7-1 Construction of pCC1BAC:: <i>gpi</i> , pCC1BAC:: <i>galE</i> and pCC1BAC:: <i>gpi-galE</i> plasmids.....	257
Figure 7-2 <i>Hind</i> III digestion products of pCC1BAC vector and A10F2 plasmids.	263
Figure 7-3 The putative genes responsible for CTAB resistances identified by transposon mutagenesis.....	264
Figure 7-4 The amplification of putative CTAB resistance genes.	265
Figure 7-5 Minimum inhibitory concentration of CTAB and CPC.....	266
Figure 7-6 The alignment of GalE amino acid sequences and protein structures.....	269
Figure 7-7 Mass spectrometry analysis of the LPS, extracted from <i>E. coli</i> containing pCC1BAC and pCC1BAC:: <i>galE</i>	271
Figure 7-8 The amount of cytochrome <i>c</i> bound to <i>E. coli</i>	272

Figure 7-9 Chemical structures of quaternary ammonium compounds (QACs).	274
Figure 7-10 The UDP-sugar biosynthesis pathway.	276

List of Equations

Equation 2-1: ng of insert = $\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio}$	62
Equation 4-1: $n_T = \sum x_n$	109
Equation 4-2: $Y = \frac{-\Delta G}{L_H}$	110
Equation 4-3: $d = (n_T \times 18.16) + (Y \times 96.59) - 116.87$	110
Equation 4-4: $\beta\text{-glucuronidase Miller units} = \frac{A_{405} \times 1000}{OD_{600} \times \text{time (min)} \times 1.25 \times \text{volume (mL)}}$	119
Equation 7-1: The number of unbound cytochrome c = $\frac{\text{Supernatant } OD_{530}}{\text{Stock } OD_{530}} \times 150 \left(\frac{\mu\text{g}}{\text{mL}}\right)$	260
Equation 7-2: The amount of cytochrome c bound to the cells = $150 \left(\frac{\mu\text{g}}{\text{mL}}\right) - \text{Unbound cytochrome c} \left(\frac{\mu\text{g}}{\text{mL}}\right)$	260

Abbreviations

AMR	Antimicrobial resistance
ARG	Antimicrobial resistance gene
ATP	Adenosine Tri-phosphate
BiDiPD	Bi-Directional Promoter Detection
BLAST	Basic Local Alignment Search Tool
bp	Base pair
°C	Degree Celsius
CIAP	Calf Intestinal Alkaline Phosphatase
cm	Centrimetre
CPC	Cetylpyridinium chloride
CTA	Composite transposon amplicon
CTAB	Cetyltrimethylammonium bromide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polymeric substances
FACS	Fluorescence-activated cell sorting
g	Gravitational force
GC	Gene cassette
GFP	Green fluorescent protein
Hfr	High-frequency recombination
HGT	Horizontal Gene Transfer

hr	Hour
IPTG	Isopropyl- β -D-thiogalactopyranoside
IS	Insertion Sequence
kb	Kilobase
λ	Bacteriophage lambda
LB	Luria-Bertani
LPS	Lipopolysacharide
m/z	Mass-to-charge ratio
μ g	Microgram
μ l	Microlitre
MCS	Multiple cloning site
ml	Millimetre
μ F	Microfarad
μ m	Micrometre
μ M	Micromolar
min	Minute
MMR	Mismatch repair
MUG	4-methylumbelliferyl- β -D-glucuronide
MVs	Membrane vesicles
nm	Nanometre
Ω	Ohm
OD	Optical density
ONPG	o-nitrophenyl- β -D-galactopyranoside

ORF	Open reading frame
PBS	Phosphate-buffered saline
PBP	Penicillin Binding Protein
PCR	Polymerase chain reaction
PNPG	ρ -nitrophenyl- β -D-glucuronide
QAC	Quaternary ammonium compound
RBS	Ribosome binding site
RNA	Ribonucleic acid
rpm	Revolutions per minute
RPP	Ribosomal Protection Proteins
rRNA	Ribosomal Ribonucleic Acid
S/N	Signal-to-noise ratio
sec	Second
SIGEX	Substrate-induced gene expression
SOC	Super Optimal broth with Catabolite repression
UV	Ultraviolet
V	Volt
v/v	Volume percent
w/v	Weight/volume percent
TA	Toxin-Antitoxin
TAE	Tris Acetate EDTA
Tn	Transposon
TOF MS ES+	Positive electrospray ionisation time-of-flight mass spectrometry

TSS	Transformation & Storage Solution
TU	Translocatable unit
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indolyl- β -D-glucuronide

Chapter 1
Introduction

1.1 Antibiotic Resistance

1.1.1 The antibiotic resistance crisis

Antibiotics are natural or synthetic molecules that can kill microbes (bactericidal agents) or inhibit their growth (bacteriostatic agents) with minimal effects on the hosts by targeting specific bacterial targets such as cell wall, cell membrane, ribosome and essential enzymes (Walsh, 2003, Pankey & Sabath, 2004). The discovery and introduction of antibiotics are one of the major achievements in therapeutic medicine as they have contributed to reduce the mortality of infectious diseases (Davies & Davies, 2010). They are also a major part of modern health care as they are regularly used to prevent infections during and after surgery, cancer therapy, organ transplantation and childbirth (Bow, 2013, Bratzler *et al.*, 2013, Mugglestone *et al.*, 2014).

Even though the first antibiotic, penicillin, was discovered by Sir Alexander Fleming in 1929 (Ligon, 2004), the first commercially available antibacterial drug was Prontosil, a sulfonamide drug, discovered in 1932 and introduced in 1935 (Wainwright & Kristiansen, 2011). Since then, various antibiotics were isolated from both fungi and prokaryotes, chemically modified as semi-synthetic derivatives (e.g. amoxicillin from penicillin and minocycline from tetracycline) and also chemically synthesised (fluoroquinolones and oxazolidinones) (Demain, 2009). However, the emergence of resistant pathogens after or even prior to the introduction of new antibiotics have been reported over 70 years, (Lewis, 2013). For example, the first sulfonamide

resistance case was reported in 1942 (7 years after its introduction) (Tillett *et al.*, 1943).

A number of factors have attributed to the acceleration of the antibiotic resistance problem both from the nature of microbes and human-related activities. Microbial populations naturally have the ability adapt to environmental challenges quickly and effectively through mutation, rapid generation time and horizontal gene transfer (HGT), described in section 1.2.

The uses of antibiotics in healthcare and agriculture create a strong selective pressure upon microbial communities to evolve and develop antimicrobial resistance (AMR). These human-related actions accelerating AMR include the overuse of antimicrobials, inappropriate prescription, non-prescription purchase and extensive use in agriculture (Michael *et al.*, 2014, Ventola, 2015). However, the number of new antibiotics had decreased over the past three decades, in which no new class of antibiotics has been found since 1987, leaving fewer options for the treatments against resistant bacteria (Ventola, 2015, Deak *et al.*, 2016) (Figure 1-1).

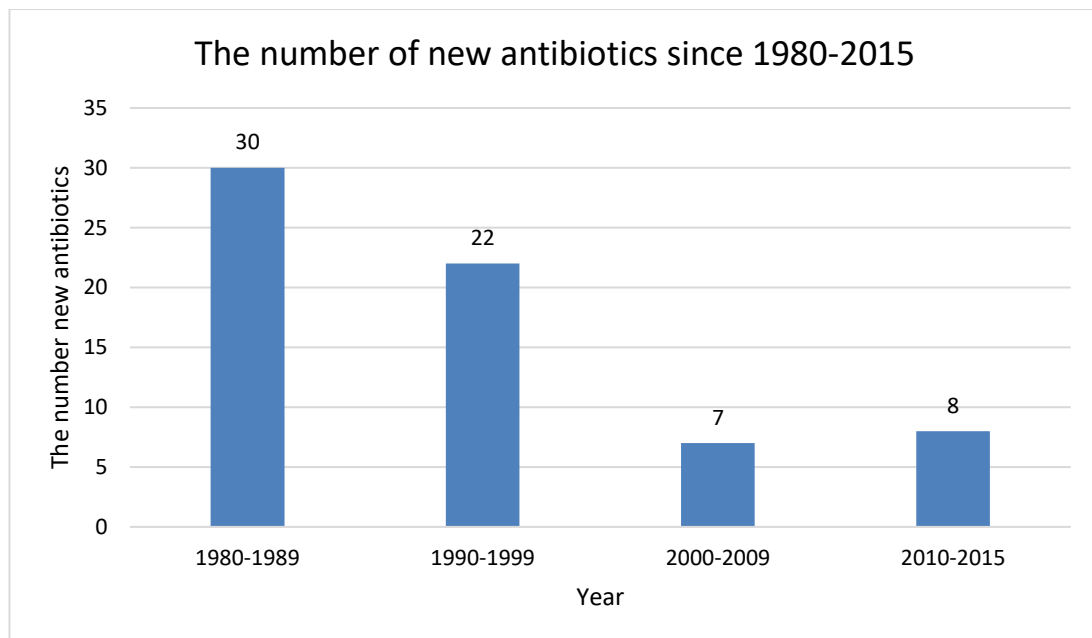


Figure 1-1 The number of new antibiotics developed and approved since 1980-2015. (Adapted from Ventola, 2015)

The World Health Organization (WHO) released the first global report on antibiotic resistance and predicted that we could soon enter the post-antibiotic era, where infections that are usually treatable will become untreatable and able to kill again (World Health Organization, 2014). The estimated deaths from AMR could rise from 700,000 per year in 2016 to 10 million per year in 2050 or one person could die in every three seconds, with the economic cost rising to \$100 trillion per year if no action is taken (O'Neill, 2016). The predictions from these reports could underestimate the real situation as there was a lack of consistency and standard in surveillance among each country. Currently, the inputs from these reports have already generated more awareness about the situation in the news and discussion in global scale organisations including the G7, G20 and United Nations (UN).

1.1.2 Mechanisms of resistance

Bacteria can become resistant to antibiotics through four main mechanisms, which are the modification of antibiotic target sites, inactivation of antibiotics, alteration of metabolic pathways, and reduction of antibiotic accumulation, described below (Figure 1-2).

Figure 1-2 Mechanisms of antibiotic resistance in bacteria. The lists of antibiotic class for each mechanism were shown in the boxes. (Retrieved from Schmieder & Edwards, 2011)

1.1.2.1 Modification of antibiotic targets

The interaction between antibiotics and their targets is usually specific with high affinity, preventing the target to carry out its function. Changing the target

molecules can result in either no binding or lower binding affinity, allowing the drug targets to continue their biological functions within the cells. This can occur through a point mutation in the genes encoding an antibiotic target. For example, the point mutation hotspots for quinolone resistant *E. coli* were in *gyrA* (especially at Ser83 and Asp87), which decreased the binding affinity of norfloxacin to DNA gyrase (Piddock, 1999). Resistance to daptomycin, a cationic antimicrobial lipopeptide, was previously shown to occur through the mutations in genes involved in cell wall and cell membrane homeostasis such as *mprF*, *liaSR* and *pgsA*, resulting in the alteration of the cell surface to be more positively charged (Tran *et al.*, 2015). An increase in net positive charge can cause repulsion and reduction in the binding of daptomycin to the cell membrane, leading to the daptomycin resistant phenotype.

The modification could also occur through the activity of enzymes or proteins in resistant bacteria. Adding a chemical group to the target can prevent antibiotics to bind to their targets. For example, *erm*-methyltransferases catalyse the methylation of an adenine residue on 23S rRNA, which hinders the binding of macrolides, lincosamides, and streptogramin B to 23S rRNA and results in resistance (Skinner *et al.*, 1983, Maravic, 2004). Another example is the target modification through the binding of resistance proteins, such as ribosomal protective proteins (RPPs), to the drug targets. It was shown that when RPPs bind to the tetracycline-blocked ribosome, they trigger conformational change of the ribosome and release tetracycline or prevent the binding of tetracycline, allowing the ribosome to continue protein synthesis (Connell *et al.*, 2003).

1.1.2.2 Inactivation of the antibiotics

Modifying or inactivating antimicrobial agents through an enzyme-catalysed reaction is another approach bacteria use to become resistant. There are two main reactions used to degrade and modify different classes of antibiotics; hydrolysis and transferring of a chemical group. The most widespread and clinically important enzymes are β -lactamases, which deactivate the β -lactam antibiotics through the hydrolysis of a β -lactam ring. Two main mechanisms were identified: the metal-activation of a water molecule for a direct attack on the β -lactam centre (metallo- β -lactamases) and the formation of a covalent acyl-enzyme followed by hydrolysis (serine- β -lactamases) (Wright, 2005).

Transfer of a chemical group catalysed by transferases is another type of reaction that bacteria use to inactivate antibiotics. Attaching these groups, such as acetyl to antibiotics block the binding of the antibiotics to their targets due to steric hindrance. Aminoglycoside antibiotics can be inactivated particularly well by this mechanism because their molecules tend to contain many hydroxyl and amide groups (Ramirez & Tolmasky, 2010). Three main classes of the modifying enzymes are including acetyltransferases, nucleotidyltransferases, and phosphotransferases.

1.1.2.3 Alteration of biological pathway

As antibiotics normally target or block a specific enzyme in the biological pathways, alteration or bypassing the pathway can therefore help bacteria to escape from the effect of those drugs. This can occur through the introduction of new molecules with similar biochemical functions, but are not targeted by

antimicrobial agents. β -lactam antibiotics inhibit cell wall synthesis through the inhibition of penicillin-binding proteins (PBPs), which are important for peptidoglycan biosynthesis. Methicillin resistance in *Staphylococcus aureus*, for example, can occur as a result of the acquisition of *mecA*, a gene encoding an alternative PBP, called PBP2a (Katayama *et al.*, 2000). This gene is often found on a large DNA fragment, called staphylococcal chromosomal cassette *mec* (SCC*mec*). As PBP2a has low affinity for most of β -lactams, it can escape the inhibition and help bacteria to form cell wall and replicate as usual (Lim & Strynadka, 2002).

Another strategy to avoid the action of antibiotics is by overproducing the antibiotic target, increasing the number of drug target molecules escaping from the inhibition by antibiotics, which will be available for the regular activity within bacterial cells. The folate synthesis pathway involves two major enzymes: dihydropteroic acid synthase (DHPS), which forms dihydrofolate from para-aminobenzoic acid, and dihydrofolate reductase (DHFR), which forms tetrahydrofolate from dihydrofolate. DHPS and DHFR can be inhibited by trimethoprim and sulfamethoxazole, respectively. Mutation in the promoter regions of both genes resulted in their overexpression, allowing bacteria to resist both antibiotics (Flensburg & Skold, 1987). Sulphonamide and trimethoprim resistance can be conferred by the acquisition of genes encoding drug-resistant variants of DHPS (*su1* carried by integron and *su2* carried by plasmid pCERC1) and DHFR (*dhrA* carried by Tn4003 and *dhrVII* carried by Tn5086), respectively (Rouch *et al.*, 1989, Sundstrom *et al.*, 1993, Skold, 2001, Anantham & Hall, 2012). Enterococci can also resist trimethoprim and sulfamethoxazole by incorporating exogenous tetrahydrofolic acid and folinic

acid, bypassing the folate synthesis pathway (Zervos & Schaberg, 1985, Hamilton-Miller, 1988).

1.1.2.4 Reduction of drug accumulation

There are two main mechanisms for bacteria to reduce the accumulation of antimicrobial agents in the cell: reducing the permeability and increasing the efflux. As the outer membrane of Gram-negative bacteria can act as a permeability barrier, outer membrane porins are required for hydrophilic antibiotics to cross this barrier. One of the strategies to reduce drug permeability is to downregulate porin expression. For example, some clinically relevant *Enterobacteriaceae* could resist carbapenems without the production of carbapenemases by reducing the expression of porin (Wozniak *et al.*, 2012).

Bacteria can also increase drug efflux to remove drugs from the cell. Overexpression of the efflux pumps can occur by a mutation in the regulators controlling their expression. For instance, the mutation in *acrR*, a local repressor of AcrAB efflux pumps, resulted in overexpression of the pumps and multidrug resistance phenotype in *Salmonella enterica* serovar Typhimurium (Olliver *et al.*, 2004). This can also be induced as a response to environmental signals such as the upregulation of NorA multidrug efflux pump in response to iron limitation in *S. aureus* (Deng *et al.*, 2012). An increase in drug efflux can also occur by the acquisition of genes encoding efflux proteins such as *tet(A)*, *tet(B)* and *tet(C)* tetracycline resistance genes (Roberts, 2005).

1.1.3 Types of resistance

1.1.3.1 Intrinsic resistance

Intrinsic resistance is an innate ability or a type of resistance that is present all members in each bacterial species without any additional genetic alteration (Cox & Wright, 2013). It could involve the absence of the target sites or lower binding affinity to the targets, for example, *Mycoplasma* are resistant to β -lactams as they lack a cell wall (Bebear *et al.*, 2011). Low cell permeability and efflux are also intrinsic resistance mechanisms, for example, the outer membrane of Gram-negative bacteria that can slow the influx of some antibiotics (Zgurskaya *et al.*, 2015).

Pseudomonas aeruginosa is an example of bacteria with multiple intrinsic resistances. As it is a Gram-negative bacterium, it can intrinsically restrict antibiotic passage through its outer membrane. It also has multidrug efflux pumps including MexAB-OprM and MexXY-OprM, reducing the accumulation of antibiotics in cells (Hancock & Speert, 2000, Nikaido, 2001). Furthermore, it produces an AmpC-like chromosomally encoded β -lactamase, which can be induced upon exposure to β -lactams (Tsutsumi *et al.*, 2013).

1.1.3.2 Acquired resistance

Acquired resistance to antibiotics occurs through either genetic mutations within the pre-existing genes in the bacterial genome or via horizontal gene transfer (HGT) (Normark & Normark, 2002). In contrast to intrinsic resistance,

this type of resistance is found only in some strains or subpopulations of strains.

Typically, short generation times and high mutation rate of bacteria, generate genetic variation in bacterial populations. When facing environmental challenges or selective pressures (e.g. antibiotics) bacteria with favourable mutations will be able to survive, replicate and dominate in bacterial population (Woodford & Ellington, 2007). The mutations can occur in the genes encoding drug targets or gene regulatory systems, as described in section 1.1.2 (Martinez & Baquero, 2000).

HGT is another mechanism that bacteria use to acquire antimicrobial resistance genes (ARGs), and also genes encoding proteins for other functions such as biodegradation pathways, metabolism, and pathogenicity (Jain *et al.*, 2002). It is a mechanism that bacteria receive DNA from exogenous sources, particularly those carried on mobile genetic elements (MGEs). The details of HGT are described in next section.

1.2 Horizontal Gene Transfer

HGT involves DNA transfer from donor cells to recipient cells, followed by replication (in the case of plasmids) and integration (in the case of transposons and bacteriophage). MGEs are important mediators of HGT, which include cytoplasmic located replicating plasmids and chromosomally located transposons and genomic islands. Three main mechanisms for HGT are including transformation, transduction and conjugation (Figure 1-3).

Figure 1-3 The mechanisms of horizontal gene transfer. a.) Bacterial transformation is the process that occurs when exogenous DNA is taken up by competent bacteria, which could be integrated into the chromosome or maintained as a plasmid in the recipient cells. b.) Bacterial transduction is the transfer of DNA from one bacterium to another by having bacteriophage as a vector. c.) Bacterial conjugation is the transfer of DNA between bacterial cells by direct cell-to-cell contact. (Retrieved from Furuya & Lowy, 2006)

1.2.1 Transformation

Bacterial transformation was the first identified HGT mechanism, which is a direct take up of exogenous DNA from the environment. It was first observed in *Streptococcus pneumoniae* that can become virulent when exposed to heat-killed virulent cells (Griffith, 1928). This was confirmed later in 1944 which

showed that the nonvirulent strains acquired virulence when inoculated together with the extracted DNA from heat-killed virulent strains (Avery *et al.*, 1944).

Transformation occurs when bacteria take up foreign DNA from the environment and incorporate it into the genome. This DNA could be released from a dead bacterial cell and located in the environment as part of biofilm structures, which were shown to be important for HGT and also the stabilisation of the biofilm matrix (Montanaro *et al.*, 2011). DNA uptake occurs during the competence state of bacteria, which can occur naturally or can be artificially induced. A total of 82 species are known to be naturally transformable, with an approximately equaled number between Gram-negative and Gram-positive bacteria (Johnston *et al.*, 2014). The competence stage of naturally transformable bacteria is usually triggered by specific conditions (such as pH, and nutrient availability) or at a specific phase of bacterial growth (Wilson & Bott, 1968, Solomon & Grossman, 1996). There are also some bacteria, which exhibit a constitutively competence state such as *Helicobacter pylori* and in *Streptococcus pneumoniae* with the mutation in a transmembrane histidine kinase (Lacks & Greenberg, 2001, Baltrus & Guillemin, 2006).

The mechanisms of bacterial transformation are similar between transformable bacteria. The conserved proteins, encoded by gene arrays called *com* regulons, are expressed during the competence state. Double-stranded DNA (dsDNA), as a substrate for transformation, is bound and captured by a transformation pilus (Johnston *et al.*, 2014). The binding sites

on competent cells vary among bacteria (20-50 sites in *Bacillus subtilis*, 33-75 sites in *Streptococcus pneumoniae* and 4-8 sites in *Haemophilus influenzae*) (Singh, 1972, Deich & Smith, 1980, Dubnau, 1999). The captured dsDNA is then internalised as single-stranded DNA (ssDNA) through a transmembrane channel. As Gram-negative bacteria contain two layers of membrane, there are two transmembrane channels for each layer, while Gram-positive bacteria use only one transmembrane channel (Figure 1-4).

Figure 1-4 The processes of transformation in Gram-negative and Gram-positive bacteria. The exogenous double-stranded DNA (dsDNA) is captured by a transformation pilus (Tfp), which consists mainly of ComGC and PilE subunits in Gram-positive and Gram-negative bacteria, respectively. dsDNA is transported into cytosol via transmembrane channels (ComEC for Gram-positive bacteria, and PilQ and ComA for Gram-negative bacteria). The captured dsDNA is internalised as single-stranded DNA (ssDNA) into the cytosol, which is then bound by DNA processing protein A (DprA). The recombinase RecA is then polymerised on ssDNA and promotes homologous recombination. (Retrieved from Johnston *et al.*, 2014)

Once the DNA enters into the cells, it could be degraded to nucleotides that can be used in other metabolic functions such as DNA replication. The internalised ssDNA could be, alternatively, recombined into the bacterial genome through homologous recombination (Johnston *et al.*, 2014). The internalised ssDNA is bound by DNA processing protein A (DprA), which then recruits RecA recombinases to the DNA and promotes recombination (Mortier-Barriere *et al.*, 2007). In the case of plasmids, if it contains an origin of replication that can be recognised by the host, recircularisation and second-strand synthesis will occur and result in plasmid maintenance (Thomas & Nielsen, 2005).

1.2.2 Transduction

Transduction is a HGT mechanism that DNA from a phage-infected bacterium is transferred into a recipient cell via bacteriophage. It was first discovered in 1952, which showed the transfer of chromosomal DNA from one strain of *Salmonella typhimurium* to another by phage P22 (Zinder & Lederberg, 1952). By having phage as a vector, this mechanism does not require a cell-to-cell contact, and the donor DNA can be protected from physical and chemical agents (Calero-Cáceres & Muniesa, 2016).

It can be categorised into two types: generalised and specialised transduction (Ozeki & Ikeda, 1968). Generalised transduction is the process that any bacterial gene could be transferred via a bacteriophage. During the lytic cycle of phage, the bacterial host DNA is usually broken up into fragments, which could be incorporated into the phage head and transferred, instead of the viral

DNA (Figure 1-5A) (Canchaya *et al.*, 2003). Specialised transduction is a process that only the bacterial DNA located close to the integrated phage DNA on the host chromosome is transferred by a bacteriophage (Figure 1-5B). This occurs from the imprecise excision of prophage DNA from the host, resulting in the excised DNA containing phage DNA and adjacent chromosomal DNA from the bacterial host (Canchaya *et al.*, 2003).

Figure 1-5 Types of transduction. A.) Generalised transduction is the transfer of random DNA fragments of host (shown in blue) to neighbouring bacteria B.) Specialised transduction is the transfer of prophage DNA (shown in orange) and the bacterial DNA (located next to the prophage DNA, shown in blue) to a new recipient cell. (Adapted from Salmond & Fineran, 2015)

After packaging of the DNA into transducing particles, they can then infect a new bacterium. Once the DNA is injected into the cell, it could be degraded and used in other metabolic pathways by the host. It could also be integrated

into the host genome by homologous recombination (generalised transduction) and as part of prophage DNA (specialised transduction) (Ozeki & Ikeda, 1968, Ochman *et al.*, 2000).

1.2.3 Conjugation

Unlike transformation and transduction, conjugation is the HGT mechanism through a cell-to-cell contact or a bridge-like connection between donors and recipient cells. It was first reported in 1946 when two auxotroph strains of *Escherichia coli*, which required different nutrients for their growth, were incubated together. The results showed that some of the progenies gained the ability to grow without the nutrient supplementation, suggesting that there was a recombination of genes between both strains (Lederberg & Tatum, 1946, Tatum & Lederberg, 1947). The requirement of cell-to-cell contact was later confirmed by an experiment in a U-shaped tube, which was partitioned into two sections by a frittered glass filter (allow only the media, not bacteria, to move between both partitions) (Davis, 1950). No colony with a recombination was found when inoculated two different auxotroph strains on each arm of the tube, suggesting that a direct physical contact between the two strains is essential for the transfer and recombination event to take place.

Bacterial conjugation mostly depends on the presences of either conjugative plasmids or integrated conjugative elements (ICEs), which contain genes responsible for the transfer, and sometimes also contain other accessory genes for the survival of bacteria in the presence of various environmental stresses.

There are three main steps in conjugation: mating-pair formation, a signaling event to trigger the transfer, and a transfer of DNA (Frost *et al.*, 2005). The physical contact in Gram-negative bacteria occurs by an extracellular filament, called sex pili (Bhatty *et al.*, 2013), shown in Figure 1-6. However, for Gram-positive bacteria, no pili is formed, which could be due to the difference in cell structures, the lack of periplasm and the absence of genes encoding a pili on the MGEs. Instead, their cell attachment is mediated at least in part through adhesins (Goessweiner-Mohr *et al.*, 2014).

Figure 1-6 Electron microscope image of the of Gram-negative sex pilus in conjugation. (Retrieved from Brooks *et al.*, 2012)

Both dsDNA and ssDNA molecules can be transferred from donor to recipient cells by conjugation. The transfer of ssDNA relies on a type IV secretion system (T4SS), which is a membrane-associated transporter complex. The ssDNA molecule usually forms in the donor by a relaxase, which nicks one

strand of DNA at the origin of transfer (*oriT*). The cleaved strand is transferred to the recipient through the T4SS. The DNA is subsequently recircularised and replicated into dsDNA in the cytoplasm of the recipient cell (Zechner *et al.*, 2012). The transfer of dsDNA is found only in Actinobacteria, which relies on a hexameric pore-forming ATPase, called TraB (Vogelmann *et al.*, 2011). The transferred DNA can be either maintained as a plasmid (conjugative plasmid) or integrate into the recipient's genome catalysing by recombinases (conjugative transposon).

The chromosomal DNA from donor can also be transferred to recipient cells by conjugation. It occurs when the donor cell is a high-frequency recombination (Hfr) cell, which a conjugative plasmid integrated into its genome via homologous recombination. The chromosome of Hfr strain is, therefore, considered as a large conjugative plasmid. The conjugative transfer in Hfr cell begins at the *oriT* site, which is then followed by the integrated plasmid and the chromosomal genes closest to the integration site (Figure 1-7) (Smith, 1991). The transfer of a whole genome is unlikely to occur because the physical contact between both cells is temporary, which is too short for an entire genome of Hfr strain to be transferred. The transferred donor DNA can then recombine into a recipient's chromosome.

Figure 1-7 The conjugative transfer between high-frequency recombination (Hfr) and recipient cells. (Adapted from Ma *et al.*, 2014)

1.2.4 Membrane vesicle-mediated gene transfer

In addition to the three main HGT mechanisms described above, membrane vesicle-mediated gene transfer is another HGT mechanism recently described, which relies on membrane vesicles (MVs) released by the donor cells. MVs are lipid-bilayer spheres derived from the cell surface with the size between 10-500 nm, which have been shown to carry and transfer genetic material (both DNA and RNA), polysaccharides and proteins (Figure 1-8) (Domingues & Nielsen, 2017). The release of MVs can be induced by different types of stress conditions, such as starvation, oxygen stress, UV light and antibiotic exposure (Sabra *et al.*, 2003, Fulsundar *et al.*, 2014, Devos *et al.*, 2015).

Figure 1-8 Membrane vesicle-mediated gene transfer. (Retrieved from Domingues & Nielsen, 2017)

The detailed mechanisms for the formation of MVs and their HGT are not yet fully understood. MVs can be released from Gram-negative, Gram-positive bacteria and archaea in which their composition, amounts, sizes of MVs tend to be varied between different species. For Gram-negative bacteria, the most common MVs are formed from the area of the outer membrane, which detaches from the peptidoglycan layer, bulges outward and undergoes fission (Schwechheimer & Kuehn, 2015). The MV formation in Gram-positive bacteria was proposed that their MVs may be forced through the pores or protein channels on the thick cell wall by turgor pressure, and the pore size could be increased by the activity of proteases (Brown *et al.*, 2015).

DNA can be packed into MVs through different ways, such as a cytoplasmic route (donor chromosomal DNA packed into MVs), an extracellular route (extracellular DNA bound to MVs) and an injection of DNA by phages (Figure 1-8) (Domingues & Nielsen, 2017). The released MVs can attach to the surface of recipient cells, followed by the lysis or internalisation of MVs and the introducing donor DNA into the cells (Fulsundar *et al.*, 2014). MVs

therefore act as a vector similar to a bacteriophage in transduction, which can also protect the DNA from degradation caused by nuclease and temperature (Soler *et al.*, 2008, Fulsundar *et al.*, 2014).

1.3 Mobile Genetic Elements

Mobile genetic elements (MGEs) are segments of DNA that can move within bacterial genomes (intracellular transposition) or between bacterial cells (intercellular transposition). They usually contain genes encoding proteins that mediate the movement and also accessory genes such as antibiotic resistance and virulence factors.

The association of resistance genes with MGEs could give evolutionary benefits to all components in the system, including bacterial hosts, resistance genes and MGEs. The bacterial hosts can grow in antibiotic-containing environments in the presence of resistance genes, while the spreading of resistance genes in a bacterial population can be facilitated by MGEs. MGEs can be considered as a selfish or parasitic DNA, as they utilise resources from hosts for their own multiplication, which are often costly. Without providing any benefits to their hosts, MGEs can be lost during the cell division. Therefore, in order for MGEs to be maintained in the bacterial population, they frequently carry genes, which can be benefit to their hosts such as resistance genes, to establish a mutual relationship between hosts and MGEs, ensuring the maintenance and transfer of MGEs.

1.3.1 Insertion sequences

Insertion sequences (ISs) are the simplest type of MGE, and are small with a size of 700-2500 bp (Figure 1-9) (Mahillon & Chandler, 1998, Siguier *et al.*, 2015). They contain only genes encoding proteins for their transposition, comprising nearly the entire length of the element. Most IS elements have inverted repeat (IR) sequences on both ends with the size between 10-40 bp, containing transposase binding sites for the transposase-mediated cleavage during transposition (Mahillon & Chandler, 1998).

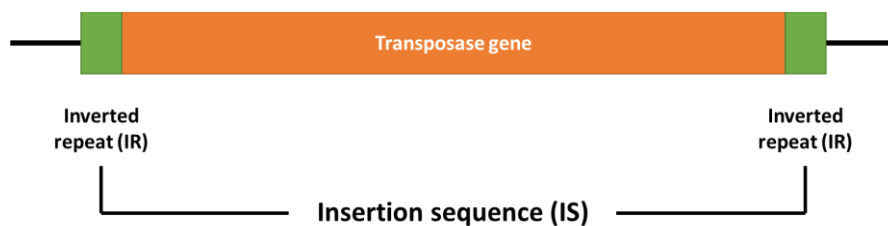


Figure 1-9 Organisation of an insertion sequence (IS). The green and orange boxes represent the inverted repeats and transposase genes, respectively.

The transposition reactions of IS elements normally start with the binding of transposase at the ends of the IS elements, followed by the formation of nucleo-protein assemblies called transpososomes. The transpososomes usually contain two or more transposase monomers, which sometimes also contain other accessory proteins such as DNA-architectural protein IHF (Chalmers *et al.*, 1998). This formation can then lead to conformational changes to mediate DNA cleavage and strand transfer reactions (Mahillon & Chandler, 1998, Siguier *et al.*, 2014).

The transposition of IS elements can cause antibiotic resistance in bacteria by increasing or disrupting the expression of genes downstream of the insertion site. As IS elements often carry either complete promoter sequences (-35 and -10 boxes) or partial promoter sequences (-35 box), their transposition could introduce stronger promoters or form new hybrid promoters, resulting in higher expression of resistance genes or activation of silent genes (Depardieu *et al.*, 2007). For instance, the insertion of IS1-like element upstream from *blaT-6* provided a -35 region, forming a hybrid promoter, which increased the expression of TEM-6 10 times (Goussard *et al.*, 1991). The overexpression of resistance genes could also occur by the insertion of IS elements into repressors of resistance genes. For example, IS186 insertion into the repressor gene *acrR*, controlling the expression of the multidrug efflux pump AcrAB, increased the MICs of fluoroquinolones in *E. coli* (Jellen-Ritter & Kern, 2001).

1.3.2 Transposons

1.3.2.1 Composite transposons

Composite transposons are MGEs which contain two IS elements flanking a segment of DNA (Figure 1-10). Both IS elements could be the same or similar ISs in either direct or inverted orientation relative to each other. The function of DNA segment between ISs can be varied, which usually carries genes for bacterial adaptation and survival genes such as those conferring antibiotic resistance and xenobiotic degradation (Nojiri *et al.*, 2004, Bennett, 2008). For example, Tn10 (flanked by IS10) and Tn5542 (flanked by IS1489) contain

tetracycline resistance gene and benzene catabolism genes, respectively (Foster *et al.*, 1981, Fong *et al.*, 2000).

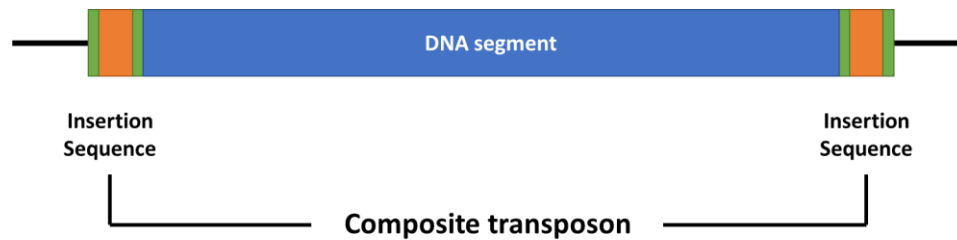


Figure 1-10 Organisation of composite transposon. The green and orange boxes represent the inverted repeats and transposase genes of the IS elements, respectively. The DNA segment of composite transposon is shown as a blue box.

The flanking ISs provide the recombination machinery such as transposase and *cis*-acting sites at the termini of the ISs. With the structure of composite transposons, the transposition activity could occur in two ways: as a whole unit of composite transposon and as an individual IS element excised out from the transposons (Bennett, 2008).

1.3.2.2 Non-composite transposons

Non-composite transposons carry various genes flanked by short inverted repeats (IRs) (Figure 1-11A). Their transposition activity relies on proteins encoded by genes in between the IRs, such as *tnpA* (transposase) and *tnpR* (resolvase for the resolution of cointegrate, which is an intermediate structure forms by a fusion between the donor and target replicons in replicative transposition). Antimicrobial resistance genes can also be found as accessory genes in this region, which can be transferred by non-composite transposons e.g. *bla* in Tn3, mercury resistance *mer* operon in Tn501 and erythromycin

resistance *erm(B)* in Tn551/Tn917 (Figure 1-11B) (Perkins & Youngman, 1984, Brown & Evans, 1991, Nicolas *et al.*, 2015). Whereas most Tns carry one or two genes for transposition, Tn7 carries 5 genes (*tnsABCDE*) and class 2 integron carrying genes conferring trimethoprim and streptomycin resistance (Craig, 1991). Some non-composite transposons, called conjugative transposons, also contain a conjugative module which encodes enzymes to catalyse conjugative transfer of the transposons to recipient cells e.g. Tn916 (Figure 1-11B) (Roberts & Mullany, 2009).

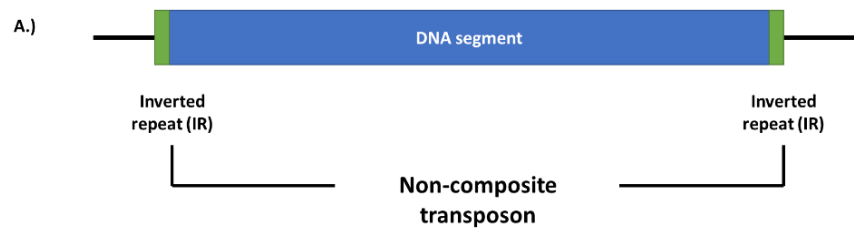


Figure 1-11 Organisation of non-composite transposon. A.) General characteristic of non-composite transposon. The green and blue boxes represent the inverted repeats and DNA segment of non-composite transposon, respectively. B.) Examples of non-composite transposons. (Adapted from Hayes, 2003, Mantengoli & Rossolini, 2005, Poirel *et al.*, 2006, Del Grosso *et al.*, 2009, Roberts & Mullany, 2009)

1.3.3 Translocatable Units

A translocatable unit (TU) is a novel MGE, which was recently identified to be derived from IS26-based composite transposon (Harmer *et al.*, 2014, Harmer & Hall, 2015). TUs are non-replicative circular molecules, which excise from composite transposon and carry one IS element and the DNA segment originally flanked by the ISs, leaving the other IS element at the original genomic location (Figure 1-12). It can also move and integrate with the other IS elements in a different location, forming a composite transposon.

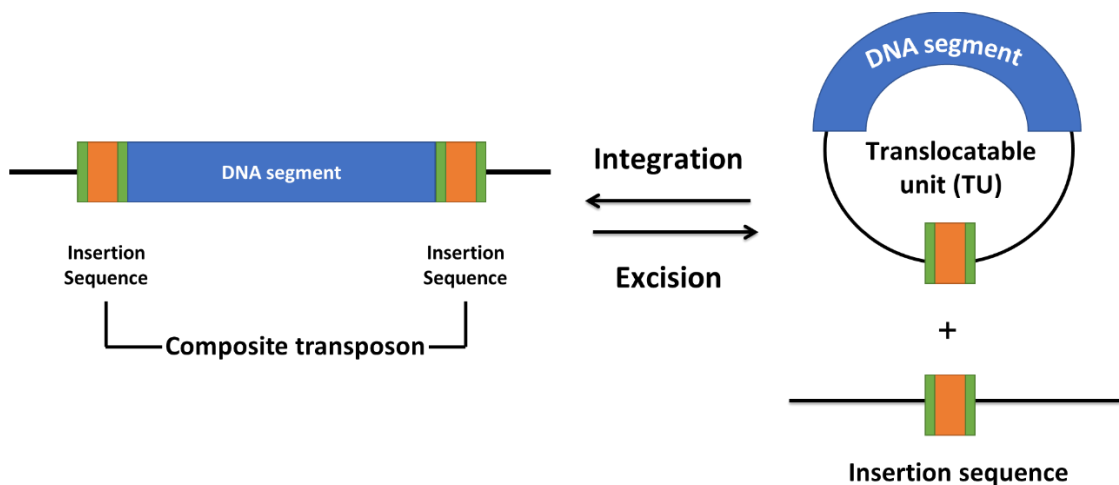


Figure 1-12 Organization of translocatable unit (TU). The green and orange boxes represent the inverted repeats and transposase genes of the insertion sequences, respectively. The DNA segments, carrying by composite transposon and translocatable unit, were shown as blue boxes.

The formation of TUs is, therefore, an alternative route for the transposition of genes from composite transposons. As the DNA segments on TUs are derived from composite transposons, most of the genes found on TUs are also usually conferring antibiotic resistance, such as the kanamycin resistance gene in

IS26-*aphA1a* TU and tetracycline resistance gene in IS1216-*tet(S)* TU (Ciric *et al.*, 2014, Harmer & Hall, 2015).

1.3.4 Integrons

1.3.4.1 Characteristics of integrons

Integrons are genetic elements that are commonly found in bacteria, especially in Gram-negative bacteria. They allow bacteria to capture and express exogenous genes, which are important for the dissemination and expression of genes in the bacterial population. There are two common parts of an integron (Figure 1-13). The first part consists of an integrase gene *intI*, an *attI* recombination site, and a *P_c* promoter. The integron integrase is a site-specific tyrosine recombinase, which can catalyse recombination between incoming circular GCs and the recombination site *attI*, within an integron, resulting in insertion or excision of cassettes (Collis & Hall, 1992).

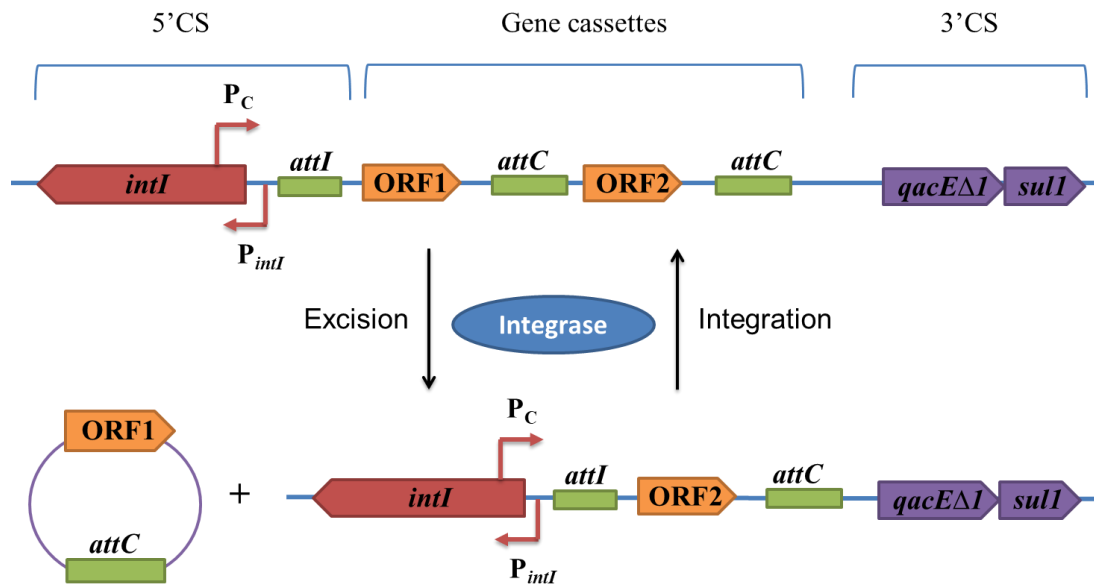


Figure 1-13 Organisation of Integron and the associated gene cassettes. The open arrowed boxes represent ORFs, pointing the direction of transcription. The genes in 5' and 3' conserved segment (CS), the open reading frame (ORF), the recombination site (*attI* and *attC*) are shown in red, purple, orange and green respectively. The red arrows represent a P_C and P_{intI} promoters.

The second part of integron is an array of gene cassettes (GCs). Each GC usually contains a single promoterless open reading frame (ORF) and an *attC* recombination site with the size of 57 to 141 bp (Figure 1-13) (Stokes *et al.*, 1997). GCs can be found in two forms: a free circular non-replicating DNA molecule and a linear form in integrons located on larger DNA molecules such as plasmids or bacterial chromosome. The predicted protein functions of cassette genes are varied, such as antibiotic resistance, virulence, secondary metabolism, and plasmid maintenance, which are likely to be specific for each environment.

After the insertion of GCs into an integron, they can be expressed by the integron-associated promoter P_C , usually located within the integrase (Figure

1-13). This formation of integron gives two main benefits i.e. the new GC can be inserted at the specific site, *attI*, without interrupting the existing cassettes, and genes on the inserted GC can be immediately expressed by P_C (Gillings, 2014).

Integrans can be divided into two groups; mobile integrans (MIs) and chromosomal integrans (CIs). The chromosomal integrans are embedded in bacterial genome and carry a number of GCs, for example, CI in *Vibrio cholerae* carried 176 GCs, which was 3.1% of host genome (Boucher *et al.*, 2006a). They also have homogenous *attC* sequences, reflecting a relationship between the host and the recombination sites. As CIs locate on chromosomal DNA, their mobility mainly relies on the activity of transposases or recombinases, encoded by their hosts or by GCs, such as the transposase gene within *Xanthomonas* cassette arrays (Gillings *et al.*, 2005).

The mobile integrans can carry up to 8 GCs, which usually contain antibiotic resistance genes (Escudero *et al.*, 2015). The sequences of *attC* sites and GCs in MIs are derived from diverse genetic background. Their mobility can occur through the association with plasmids or transposons, allowing them to be spread both intra- and inter-species. Therefore, they are considered as an important factor for bacterial adaptation against antibiotic exposure. It was also the first type of integron identified in the 1980s, which was thought to play a major role in the spreading of ARGs in the 1960s (Stokes & Hall, 1989).

Currently, five classes of mobile integrans, classified according to the sequences of their integrase, have been identified to be associated with

antibiotic resistance, including class 1, 2 and 3 (usually found in clinical environments), class 4 (found on the SXT element of *Vibrio cholerae*) and class 5 (found on the pRSV1 plasmid of *Alivibrio salmonicida*) (Gillings, 2014, Escudero *et al.*, 2015).

Class 1 integrons are the most important class involving in the spreading of ARGs. They were thought to derive from the chromosomal class 1 integron, which can be commonly found in nonpathogenic soil and freshwater *Betaproteobacteria*. This was inferred by the presence of the identical *intI1* sequences from clinical pathogens on the chromosomes of those nonpathogenic environmental *Betaproteobacteria* (Gillings *et al.*, 2008). Introducing antibiotics significantly influenced the evolution of clinical class 1 integrons. The *sul1* gene, conferring sulfonamide resistance, is commonly found to be associated with class 1 integrons, which is likely to correspond to the fact that sulfonamide was the first antibiotic commercialised and therefore used the most.

1.3.4.2 Recombination of gene cassettes

Integrons contain two types of recombination site, *attI* and *attC*, which are important for the insertion, excision and reshuffling of GCs in their systems (Figure 1-14).

Figure 1-14 Structure of gene cassettes and recombination sites. The vertical arrows represent the recombination point. A.) A linear form of single GC contains the conserved recombination site GTTRRY, an open reading frame encoded by the cassette and the *attC* site (B) Detailed structure of a single *attC* site. These elements have integrase binding sites (R'', L'', L' and R'). R'-R'' and L'-L'' are inverted repeats. Therefore they can base pair with each other, forming a secondary structure which can be recognised by integron integrases. An extra base, labelled with an asterisk (*) in L'', ensures correct orientation and insertion of cassettes into the array. Between the inverted repeats is a variable central region that varies in length (16-109 nts). (C) An *attI* site from a class 1 integron. The *attI1* site also has two integrase binding sites (L and R) and direct repeats (DR1 and DR2). The R site contains the conserved recombination point GTTRRRY. (D) The secondary structure is forming by the bottom strand of the *attC_{aadA7}* site. The R and L boxes are indicated with green boxes. (Adapted from Gillings, 2014 and Escudero *et al.*, 2015)

The *attI*, located immediately adjacent to the integrase gene, contains a core site, composing of two integrase binding sites, called L and R boxes. The R

binding site contains the conserved sequence 5'-GTTRRRY-3' (R=A or G, Y=C or T) which has a recombination point located between the G and T residues (Gillings, 2014) (Figure 1-14C). For class 1 integrons, there are two extra integrase binding sites, which are DR1 and DR2 direct repeats, located upstream of the core site (Gravel *et al.*, 1998). Nevertheless, the sequences of L, DR1, and DR2 sites are not conserved and varied between different *attI* sites. Different integron integrases preferentially recognise their related or adjacent *attI* sites. However, they can still recognise and operate on *attI* sites from other systems with lower efficiency (Biskri *et al.*, 2005).

The *attC* sites are located at the end of each integron GCs, containing four integrase binding sites: R'' and L'' at the 5' end, and L' and R' at the 3' end. The sequences of R'' and R' are conserved consisting of 5'-RYYAAC-3' and 5'-GTTRRRY-3', respectively (Gillings, 2014). As the central part of *attC* is highly variable in both size and sequences, it suggests that the recombination activity is unlikely to rely on the sequences of *attC*. However, the recombination relies on the structure of *attC* that is conserved by forming a secondary hairpin structure through the palindromic organisation of *attC* sites, as the sequences of R''- R' and L''- L' are inverted repeats (Figure 1-14B) (Stokes *et al.*, 1997).

Integron integrases are enzymes for the recombination reaction on integrons. The reactions can occur at the double-stranded *attI* sites of integrons and at the single-stranded *attC* sites of GCs. The recombination reaction preferentially occurs on the bottom strand of *attC* with 1000-fold higher than the reaction occurs on the top strand (Bouvier *et al.*, 2005). Integrase can

distinguish the bottom strand of *attC* through an extra nucleotide in the L'' box that has no complementary nucleotide in the L' box, forming as extrahelical or protruding bases in secondary structure (Bouvier *et al.*, 2009).

There are three different recombination reactions catalysed by integron tyrosine recombinases. The recombination between *attI* and *attC* (*attI* x *attC*) is the most common reaction catalysed by integrase, resulting in the insertion of new GC right next to the P_C promoter (Clewell *et al.*, 2014). Even though the reaction can also result in the deletion of GC in the first position from the integron (intramolecular), the chance for *attI* and adjacent *attC* simultaneously forming as ds and ss DNA, respectively, at the same time are very rare (Escudero *et al.*, 2015). Therefore, *attI* x *attC* recombination is more towards integration than excision (intermolecular rather than intramolecular).

The recombination between two *attC* sites (*attC* x *attC*) can also occur with less efficiency. This could occur between *attC* on the same cassette array, leading to the excision of GC into a circular form (Collis & Hall, 1992). The insertion into an array through *attC* x *attC* is also possible, but less preferred compared to the insertion at *attI* (Collis *et al.*, 1993). The last reaction is the recombination between two *attI* sites (*attI* x *attI*), which is the least efficient reaction, resulting in the fusion and rearrangement of the cassette content between different integrons (Hansson *et al.*, 1997, Collis *et al.*, 2001).

1.3.4.3 Expression of integron integrase and gene cassettes.

As the rearrangement of GCs in integron could result in the deleterious effects to the cells, the expression of integrase must be regulated and expressed at

the precise time that the acquisition of new GCs would be advantageous. Integrase expression is controlled through the SOS response, as there is a LexA-binding site located in the *intl* promoter (Guerin *et al.*, 2009). In normal conditions, LexA represses the expression of integrases by its binding to LexA boxes. The SOS response can be induced through the accumulation of ssDNA in the cell, generating during DNA damage, DNA repair, transformation, conjugation and antibiotic exposure (Baharoglu *et al.*, 2010, Baharoglu & Mazel, 2011, Baharoglu *et al.*, 2012). RecA can then recognise these ssDNA and polymerises into RecA nucleofilaments, which can induce the autocleavage of LexA. The expression of integrase is then triggered upon the releasing of LexA from the promoter.

The expression of GCs relies mainly on the external promoters especially P_C . The expression level of GCs depends on the distance of the GCs from P_C as the strength of expression decreases when GCs are located further from the P_C (Collis & Hall, 1995). The strength of P_C promoters also varies between different integrons. Integrons with a weak P_C promoter often have higher expression of integrases, due to the transcription interference with integrase promoter P_{intl} , as they lie facing each other (Figure 1-13) (Jové *et al.*, 2010). The tight relationship between both promoters are important to control the level of GC shuffling and the expression of GCs. Bacteria, exposed to varied stresses such as those in clinical and wastewater environments, were shown to have weak P_C promoter as the cassette shuffling is more important for adaptation than the strong expression of GCs (Vinue *et al.*, 2011, Moura *et al.*, 2012).

Some GCs can also carry promoter sequences such as *cmIA1* chloramphenicol resistance, the *qnrVC1* quinolone resistance, *ere(A)* erythromycin resistance and the toxin-antitoxin (TA) GCs (Stokes & Hall, 1991, Biskri & Mazel, 2003, Szekeres *et al.*, 2007, da Fonseca & Vicente, 2012). The presence of promoter sequences on these GCs ensures the constitutive expression of the genes on GCs, regardless of the P_C promoter and their position on integrons. Having these promoter-containing GCs within the GC array could also have a benefit to integrons as their promoters can initiate transcription of the GCs, located close to them.

1.4 Metagenomics

1.4.1 The concept of metagenomics

Prior to the development of modern molecular techniques, the main approach to investigate bacteria was culture-dependent. The drawback is that not all bacteria can be cultured in the laboratory, which the majority of bacteria in environmental samples remains unculturable by standard culturing techniques (Vartoukian *et al.*, 2010, Stewart, 2012). The recognition of uncultured bacteria was shown by the great plate count anomaly (Staley & Konopka, 1985), which showed that the number of bacteria obtaining from cultivation are orders of magnitude lower than the number of bacteria observed under the microscope, suggesting that only a small proportion of bacteria can be cultured in the laboratory. To overcome this drawback, several culture-independent methods have been developed since the 1980s, such as microarray, fluorescence in

situ hybridization (FISH) and quantitative polymerase chain reaction (qPCR), significantly expanding our knowledge on microbial diversity (Su *et al.*, 2012).

Metagenomics is another culture-independent approach, which is the sequence and function-based analysis of bacterial genetic materials isolated directly from the environment. Therefore, it is a powerful research tool, giving much information for an in-depth understanding of a given microbial ecosystem in many aspects such as the evolutionary history, metabolism and ecological role of microbes (Oh *et al.*, 2011, Abubucker *et al.*, 2012). To date, metagenomic studies have been carried out on a large variety of environments such as ocean, soil, water, human saliva and gut and extreme environments (glacier ice and hot springs) (Frias-Lopez *et al.*, 2008, Schoenfeld *et al.*, 2008, Simon *et al.*, 2009a, Qin *et al.*, 2010, Belda-Ferre *et al.*, 2012, Fierer *et al.*, 2012, Gomez-Alvarez *et al.*, 2012).

1.4.2 Types of metagenomic studies

Metagenomics can be categorised into two types: sequence-based and function-based metagenomics (Figure 1-15).

Figure 1-15 Construction and screening of metagenomic libraries. A collection of DNA is extracted directly from the microbial community. The extracted metagenome can be investigated by either function-based or sequence-based metagenomics. (Retrieved from Sabree *et al.*, 2009)

1.4.2.1 Sequence-based metagenomics

Sequence-based metagenomic studies involve the sequencing and analysis of extracted metagenomic DNA without focusing on the activity or function of DNA. The collected genomic information is then analysed by assembling genetic information and comparing to the known databases to predict on its functions. It could be categorised into two categories: shotgun metagenomics and targeted metagenomics.

The shotgun metagenomics is a technique for microbiome analysis that does not target a specific genomic locus. With the development of next generation sequencing technologies and the significant improvement in bioinformatics, it allows us to perform large-scale sequencing of DNA from multispecies communities. The metagenomic DNA is sheared into small fragments (0.5-5 kb fragments), which are independently sequenced, giving complex and large metagenomic data to be further analysed. The sequences of the DNA fragments, for instance, could be assembled into contigs or complete genomes, or compared to the database to predict for coding regions and assign the functional annotation (Sharpton, 2014). The results from shotgun metagenomics could illustrate the distribution and redundancy of gene functions, the linkage between phylogeny and gene functions, and the genomic organisation in a community (Schmieder & Edwards, 2011). Regarding antibiotic studies, this approach can be used to identify all of the known resistance genes containing in each environmental sample, and possibly identify mutations known to confer resistance.

Targeted metagenomics is another sequence-based approach, focusing only on the gene of interest in metagenomic DNA. PCR primers can be used to amplify the target genes from metagenome, followed by sequencing. For example, taxonomic and phylogenetic analysis studies can be done by performing 16S rRNA gene sequencing with Illumina HiSeq or MiSeq platforms, as 16S rRNA genes contain hypervariable regions which are varied among bacteria and can be used for bacterial identification. The other genes of interest in metagenome can also be targeted by PCR-based approaches. The amplicons can be amplified with specific primers, constructed as a library with cloning vector and sequenced the inserts. To investigate specific sequences or genes in a metagenome, PCR-based approaches was shown to give more sensitive results with a factor of 10 to 100 times higher than the results from shotgun sequencing on the same metagenomics sample (Woodhouse *et al.*, 2013).

Another approach for targeted metagenomics is, for example, to determine the relative abundance of different genes in the different environments by using real-time PCR. Microarrays can also be applied to detect genes in metagenomes such as all known resistance genes.

As the sequence-based approaches rely on the comparison of sequence data to known genes in the databases as references, their major disadvantage is that they cannot identify or predict the function of novel genes that have no similar sequences in the database and also cannot give information on the expression of genes.

1.4.2.2 Function-based metagenomics

The function-based analysis relies on the function of genes containing in metagenomic samples. It involves with the construction of metagenomic libraries by ligating the fragments of metagenomic DNA into appropriate cloning vectors. A suitable vector can be selected based on several factors such as bacterial hosts, the size of inserts and copy number of vectors. The constructed library is then introduced into surrogate hosts and grown on appropriate medium to screen for clones with phenotypes of interest.

Multiple function-based approaches have been developed to identify novel genes with functions of interest, depending on the purpose of the experiment. The first example is the screening for active enzymes by growing the libraries on medium supplemented with selective substances such as chemical dyes and enzyme substrates. The clones with specific metabolic activity can be distinguished and selected as they will have different phenotypes from the others. For example, two novel glycosyl hydrolases were discovered from the fosmid metagenomic libraries, constructed from the bacterial community isolated from a cast of earthworms, by screening for colonies with strong yellow colour on medium containing *p*-nitrophenyl- β -D-glucopyranoside and *p*-nitrophenyl- α -L-arabinopyranoside (Beloqui *et al.*, 2010).

The second type is to screen for the clones that can grow under selective conditions, such as in the presence of antibiotic or nutrient deficiency, which can then be characterised to determine the genes responsible for their survival. For example, a novel ABC transporter gene *tetAB(60)*, conferring

tetracycline and tigecycline resistance, was found from the human oral metagenomic library by screening on agar containing tetracycline (Reynolds *et al.*, 2016). Mutant host strains, which require the targeted genes for growth, can also be used for screening. For example, nine novel genes encoding DNA polymerase I were found by screening metagenomic libraries derived from glacial ice in *E. coli* containing *polA* mutant (Simon *et al.*, 2009b).

The last example is the screening of the metagenomic libraries for genes encoding antimicrobial substances. The libraries can be grown on the bacterial lawn and selected for the clones exhibiting a zone of inhibition. For example, six clones with antibacterial activity were identified by screening soil metagenomic libraries, constructed in *Ralstonia metallidurans*, against *B. subtilis* (Iqbal *et al.*, 2014). These antibacterial compounds were characterised as a lipase, proteases and cell wall lytic enzymes.

The advantage of function-based study is that it has a potential to identify novel gene families and functions, as no known sequence is required in the study. However, there is a limitation regarding the expression of genes in heterologous surrogate hosts. Functional screening of soil metagenomic DNA using six different hosts showed that each strain can express unique sets of clones with minimal overlap (Craig *et al.*, 2010).

1.5 The Microbiota of the Human Oral Cavity

1.5.1 Oral Microbiota

The human oral cavity is the second most complex microbial ecosystems in the human body. It contains many distinct microbial habitats such as the lips, teeth, tongue, cheek, gingival sulcus, soft and hard palates, which are different in physical compositions such as in nutrient availability, pH level and surfaces (Dewhirst *et al.*, 2010). Each of the habitats is, therefore, colonised by distinct microbial communities. Over 700 bacterial species have been identified in the human oral cavity, in which approximately 100 – 200 species can be found per mouth (Wade, 2011). Only two-thirds of these can be cultivated in the laboratory, thus the introduction of culture-independent methods significantly increases our knowledge on the microbial diversity within the human oral cavity.

Bacterial colonisation in oral cavity occurs within seconds of birth and develops throughout childhood which can be influenced by many factors such as mode of delivery and mode of feeding (Lif Holgerson *et al.*, 2011, Al-Shehri *et al.*, 2016). The human oral cavity is a dynamic environment, as it is a major gateway to the human body, which connects to the gastrointestinal tract (food) and the trachea and lungs (air). Also, human-related behaviours (such as brushing, flossing and kissing) and oral treatments (scaling and restoration) can also alter the composition of the microbiota (Corby *et al.*, 2008, Tanner *et al.*, 2011, Kort *et al.*, 2014). Recently, it was shown that shared environment

in the early stage of life was a dominant factor in determining the oral microbiome, rather than genetic factors (Shaw *et al.*, 2017).

Due to the dynamic changes within the oral cavity, it is difficult to define the precise composition of the oral microbiome. However, the human oral microbiome study based on 16S rRNA sequences showed that 96% of detected species were derived from 6 major phyla, including *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, and *Spirochaetes* (Dewhirst *et al.*, 2010). The other minor phyla are including *Chlamydia*, *Chloroflexi*, *Euryarchaeota*, SR1, *Synergistetes*, *Tenericutes*, and TM7. The most abundant bacteria found in the oral cavity are bacteria in the genus *Streptococcus* with at least 18 species identified (Kreth *et al.*, 2009).

The composition of oral bacteria plays an important role in maintaining oral health (Wade, 2013). The changes or shifts in the oral bacterial composition between synergistic and antagonistic species have been shown to be involved in the development of several oral diseases, such as dental caries, oral squamous cell carcinoma, and periodontitis (Sakamoto *et al.*, 2000, Pushalkar *et al.*, 2011, Yang *et al.*, 2012). For example, in periodontitis, the dominant bacteria shifts from Gram-positive (*Streptococcus*) to Gram-negative bacteria (such as *Prevotella intermedia* and *Fusobacterium nucleatum*) (Berezow & Darveau, 2011).

The colonisation by commensals can reduce the opportunity of pathogens to bind and colonise in the oral cavity. Disruption of the commensal microbiota, such as by antimicrobials and in immunocompromised patients, might result

in the infections by opportunistic pathogens such as *Staphylococcus aureus* and *Candida* (Akpan & Morgan, 2002, Loberto *et al.*, 2004). The oral microbiome also has been shown to be a significant risk factor for other systemic diseases such as cardiovascular disease and diabetes mellitus (He *et al.*, 2015).

In the human oral cavity, the surfaces for microbial colonisation could be categorised into two types: shedding (mucosa) and non-shedding or solid surface (teeth). In addition to both types, bacteria attaching to those surfaces continuously shed into saliva as another microenvironment. As saliva samples are easy to collect, it is often used in studies to represent the oral microbiome.

Bacterial colonisation on the tooth can be found as a biofilm, called dental plaque. It can be classified into two types; supragingival and subgingival dental plaques, found above and below the gingival margin, respectively. It is a complex microbial community embedded in a matrix of extracellular polymeric substances (EPS). Different composition of bacteria in dental plaque can affect the balance between health and disease, such as the increasing of acidogenic and aciduric bacteria that can cause dental caries (Marsh, 2006).

The formation of a biofilm initially starts with the binding of initial colonisers, such as *Streptococcus* and *Actinomyces*, which can recognise receptors in the salivary pellicle coated on tooth surface (Diaz *et al.*, 2006). Later colonisers can then attach to early colonisers with specific adhesion-receptor interactions, increasing the diversity in the biofilm. The three-dimensional

structure of mix-culture biofilm is then developed by cell division and the formation of a complex extracellular matrix (Marsh, 2004). The growth of biofilms then become slow or static in their steady state, which some of the bacteria detach from the surface and travel to form new biofilms in other locations.

The formation of multi-species biofilms provides advantages to the oral bacteria. It allows a broader range of bacteria to attach to the surface as not all bacteria can attach to teeth or epithelial tissues (Kolenbrander *et al.*, 2010). Furthermore, oxygen-consuming bacteria within biofilm generate a suitable condition for the growth of obligate anaerobes. The macromolecules within the human oral cavity, especially glycoproteins in saliva and tissue fluid, can be degraded more efficiently through metabolic cooperation, increasing nutrient availability (Carlsson, 1997, Marsh, 2004). The extracellular polymers of biofilms can also act as a barrier to protect bacteria from host defences and increase resistance to antimicrobials, described in section 1.5.2. Bacteria can also adapt to various environmental stresses through a cell-cell communication and gene transfer within the biofilm.

1.5.2 Antibiotic Resistance in the Oral Cavity

Living as a microbial community within the biofilm can increase the resistance of oral bacteria to antimicrobial compounds. With EPS covering the bacteria, less number of antibiotics can diffuse through these layers to bacteria deep within the biofilm. At the same time, if bacteria, secreting antibiotic inactivating compounds such as β -lactamases, are in the community, a local and transient

area with higher resistance will be formed. Moreover, as cells living deep within plaque have less metabolic activities, they will become more resistance to antibiotics targeting metabolic targets such as tetracycline (protein synthesis). Less metabolic activity also increases the resistance against antimicrobials, because less number of compounds will be pumped into the cells. Furthermore, these cells can become insensitive to antibiotics targeting the cell wall such as β -lactams because they do not exhibit cell wall synthesis during their nongrowing state (Lewis, 2001).

Antibiotic-resistant bacteria have been identified in both infant and adult oral samples. Tetracycline-resistant bacteria have been identified from 15 out of 18 children, who are unlikely to have been directly exposed to tetracycline because tetracycline is not prescribed to children under the age of 12 (Lancaster *et al.*, 2005). Ampicillin-, erythromycin-, and penicillin-resistant bacteria were also identified from the oral cavity of healthy children (Ready *et al.*, 2003). It was shown that 2.8% of the oral metagenome were predicted to encode proteins with putative functions related to antibiotic and toxin resistance (Xie *et al.*, 2010). Oral metagenomic studies have recovered novel resistance genes, including tetracycline resistance genes *tet(37)* and *tet(32)*, sulphonamide resistance gene *foIP* and tigecycline resistance gene *tetAB(60)* (Diaz-Torres *et al.*, 2003, Warburton *et al.*, 2009, Card *et al.*, 2014, Reynolds *et al.*, 2016).

The exchange of genetic material, including ARGs, in the human oral microbiota have been described via all HGT mechanisms, described in section 1.2 (Roberts & Kreth, 2014). Many of the resistant genes found in the oral

cavity were shown to be associated with MGEs, which can also facilitate the transfer. One of the common MGEs found in oral bacteria is the Tn916 family of conjugative transposons, which have an exceptionally broad host range and are found in both Gram-negative and Gram-positive bacteria (Roberts & Mullany, 2009). Tn916 normally contains *tet(M)*, conferring tetracycline and minocycline resistance. Different members of the Tn916 family carries various resistance genes, such as kanamycin (Tn1545 and Tn6003), macrolide (Tn1545, Tn6002, Tn6079), cetyltrimethylammonium bromide (CTAB) (Tn6087) and mercury (Tn6009) (Cochetti *et al.*, 2008, Soge *et al.*, 2008, Ciric *et al.*, 2011, de Vries *et al.*, 2011).

1.6 Aims of the Study

A metagenomic approach has a potential to extend our knowledge on antimicrobial resistance, as all of the DNA, from both culturable and unculturable bacteria, contained in the samples can be accessed. The human oral cavity has been shown to contain various ARGs. As only two-thirds of the human oral bacteria can be cultured, the presence of resistance determinants in the oral cavity may be inaccurately estimated. In order to predict the next likely emergence of resistance and tailor antibiotic therapy, a resistance profile and their associated mobile genetic elements for the complete oral microbiota is required.

The objectives of this study are to study the biology of MGEs and their association with ARGs present in human oral bacteria by both sequence-based and function-based metagenomic approaches. Specific aims are to:

1. To develop metagenomic techniques for the detection of MGEs and antimicrobial resistance genes in the human oral metagenome.
2. To investigate the genes associated with MGEs present in the human oral metagenome.
3. To investigate the role and function of the ARGs and other genes associated with MGEs detected in the human oral metagenome.

Chapter 2
Material and Methods

2.1 Sources of media, enzymes, and reagents

Luria-Bertani (LB) broth and agar were obtained from Sigma-Aldrich (Dorset, UK) and Life Technologies (Paisley, UK), respectively. All antibiotics and chemicals were obtained from Sigma-Aldrich (Dorset, UK). The stock and working concentrations of antibiotics were listed in Appendix 3. All restriction enzymes were obtained from New England Bio (Hitchin, UK).

2.2 Bacterial strains, plasmids and growth conditions.

The list of bacterial strains and plasmids were shown in Table 2-1. All bacteria were grown on LB agar or broth at 37°C for 18 hr in aerobic conditions unless otherwise stated. Broth culture was incubated with 200 rpm shaking incubator. Antibiotics and reagents were added in the media as appropriated. The storage of all bacterial isolates was made by adding an equal volume of 20% (v/v) of sterile glycerol to the broth overnight culture, resulting in 10% (v/v) glycerol stocks. One ml aliquots of the glycerol stocks were kept at -80°C.

Table 2-1 Lists of bacterial strains and plasmids used in this study

Bacterial strains/Plasmids	Characteristics/ Resistance Marker	Reference/Source
Bacterial strains		
<i>Escherichia coli</i> α -select	Antibiotic sensitive	Bioline, London, UK
<i>Escherichia coli</i> EPI300	Antibiotic sensitive	Cambio, Cambridge, UK
<i>Escherichia coli</i> TG1	Antibiotic sensitive	Department of Bacterial Genetics, University of Warsaw, Warsaw, Poland
Plasmids		
pUC19	Small insert, High copy number, Ampicillin resistant	NEB, Hitchin, UK
pGEM-T easy	Small insert, High copy number, Ampicillin resistant, TA cloning	Promega, Southampton, UK
pHSG396	Small insert, High copy number, Chloramphenicol resistant	Takara Bio, Saint-Germain-en-Laye, France
pCC1BAC	Large insert, Low copy number but inducible, Chloramphenicol resistant	Cambio, Cambridge, UK
pAK1	Small insert, High copy number, Kanamycin resistant	Department of Bacterial Genetics, University of Warsaw, Warsaw, Poland

2.3 Sampling and processing of saliva samples

2.3.1 Subjects and sample collection

The saliva samples were collected from 11 healthy volunteers, both male and females with age between 21-65, in the Department of Microbial Diseases,

UCL Eastman Dental Institute. All eleven volunteers had not received the antibiotic treatment at least three months before the sample collection date and gave the written consent form shown in Appendix 1. This project received the ethical approval from the University College London (UCL) Ethics Committee (project number 5017/001) for the collection and the processing of the saliva samples. Approximately 2 ml of saliva samples were collected by expectoration into a sterile plastic tube and immediately continued with the extraction of saliva genomic DNA as described in section 2.3.2.

2.3.2 Extraction of saliva genomic DNA

The collected saliva samples were processed in a class I microbiological safety cabinet. All saliva samples were pooled together in a 50 ml sterile plastic tube and aliquoted 1.5 ml in each microcentrifuge tube. All of the subsequent centrifugation steps were performed at 14680 x *g* (13000 rpm, Eppendorf centrifuge 5415 D). The aliquoted samples were centrifuged for 1 min and the supernatant was discarded by pipetting. The oral metagenomic DNA was extracted by using the Genra Puregene Yeast/Bact. Kit (Qiagen, Manchester, UK), following the DNA purification protocol for Gram-Positive bacteria. The cell pellets were resuspended in 300 µl cell suspension solution by pipetting, then 1.5 µl of lytic enzyme solution was added. The mixture was mixed by inverting 25 times and incubated at 37°C for 30 min. The tube was centrifuged for 1 min, and the supernatant was discarded by pipetting. Three hundred microlitres of cell lysis solution was added and mixed. The tube was incubated at 80°C for 5 min, then 1.5 µl of RNase A solution was added and mixed by inverting 25 times. The tube was incubated at 37°C for 60 min and

then incubated the tube on ice for 1 min. One hundred microlitres of protein precipitation solution was added and mixed by vortexing at high speed for 20 sec. After the centrifugation for 3 min, the supernatant was transferred to a 1.5 ml microcentrifuge tube, containing 300 μ l isopropanol, and mixed by gently inverting the tube 50 times. After the centrifugation for 1 min, the supernatant was discarded and the tube was drained by inverting on absorbent paper. Three hundred microliters of 70% ethanol was added and the tube was inverted several times to wash the pellet. The supernatant was discarded after 1 min centrifugation. The tube was dried by inverting the tube on absorbent paper for 5 min. The DNA pellet was dissolved in 400 μ l of DNA hydration solution and incubated at 65°C for 1 hr. The tube was then shaken gently at room temperature for 18 hr. The extracted DNA was kept in -20°C freezer.

2.4 Molecular biology techniques

2.4.1 Genomic DNA extraction

From the overnight bacterial culture, 500 μ l of culture was aliquoted into a 1.5 ml microcentrifuge tube and pelleted by centrifugation at 14680 $\times g$ (13000 rpm, Eppendorf centrifuge 5415 D) for 1 min. The genomic DNA for Gram-positive bacteria was extracted from the pellets by following the protocol for the human oral metagenomic DNA extraction in section 2.3.2. If the bacteria are Gram-negative bacteria, the extraction started from the cell lysis solution step onward. For the DNA hydration, 100 μ l was added, instead of 400 μ l.

2.4.2 Copycontrol induction reaction

Prior to the extraction of the pCC1BAC-based plasmids from *E. coli* EPI300, the Copycontrol induction reaction was prepared to increase the concentration of extracted plasmid as it increases the number of plasmid from 1 copy per cell to 25 copies per cell. The reaction was prepared in a 20 ml sterile plastic tube, containing 1 ml of overnight culture, 10 µl 1000x CopyControl™ Induction Solution (Cambio, Cambridge, UK), 9 ml LB broth supplemented with appropriate antibiotics. The tube was then incubated at 37°C with shaking condition for 4-5 hrs. The plasmids were then extracted from the reaction following the plasmid extraction protocol in section 2.4.3.

2.4.3 Plasmid extraction

Plasmid extraction was performed by using QIAprep Spin Miniprep Kit (Qiagen, Manchester, UK). *E. coli* was subcultured into a 20 ml sterile plastic tube containing 5 ml LB broth supplement with appropriate antibiotics and incubated for 18 hr. The bacterial cells were pelleted by centrifugation at 4500 x *g* (5000 rpm, Eppendorf centrifuge 5804 R) for 15 min at 4°C. The supernatant was then discarded and the pellet was resuspended in 300 µl of Buffer P1 by pipetting up and down. After the resuspension, 300 µl of Buffer P2 was added and mixed by gently inverting the tube 6 times. Three hundred fifty microliters of Buffer N3 was added and mixed gently by inverting the tube 6 times. All of the centrifugations from this step were performed with a table-top centrifuge at 14680 x *g* (13000 rpm, Eppendorf centrifuge 5415 D). The mixture was then transferred to a 1.5 ml microcentrifuge tube and centrifuged

for 12 min. The supernatant was then transferred from a microcentrifuge tube to a QIAprep spin column and centrifuged for 1 min. After discarding the flow-through, 500 µl of Buffer PB was added to the column, and the column was centrifuged for 1 min. The flow-through was discarded and 700 µl of Buffer PE was added to the column. The column was then centrifuged for 1 min, and the flow-through was discarded. Additional centrifugation was then performed for 2 min to remove the remaining Buffer PE from the membrane, the column was then transferred to a new sterile microcentrifuge tube, The DNA was eluted by adding 30 µl of molecular biology grade water (Sigma-Aldrich, Dorset, UK) to the centre of the membrane, left to stand for 3 min and centrifuged for 1 min. The extracted DNA was kept in -20°C freezer.

2.4.4 Oligonucleotide synthesis

All of the oligonucleotides used in this study were synthesised by Sigma-Aldrich (Dorset, UK). The primers were designed by using SnapGene version 3.2.1 and Primer3 web-based software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). The lists and sequences of primers in this study are shown in Appendix 4.

2.4.5 Standard Polymerase chain reaction (PCR)

PCR amplification was performed by using Biometra T3000 Thermocycler (Biometra, Glasgow, UK). The standard PCR reactions were prepared with a total volume of 30 µl, composing of 15 µl 2X Biomix Red (Bioline, London, UK), 2 µl of each primer (10 pmol/µl), 10 µl molecular biology grade water and 1 µl DNA template (50-100 ng). Biomix Red contains the *Taq* DNA polymerase

that can amplify 1kb in 30 sec and generate 'A' overhang for TA cloning. The standard conditions were carried out as followings; (i) initial denaturation at 94°C for 3 min (ii) denaturation at 94°C for 1 min (iii) annealing at 50-65°C (depending on primers) for 30 sec (iv) extension at 72°C for 1-3 min (depending on the expected size of amplicons), repeated step (ii)-(iv) for 35 cycles (v) final elongation at 72°C for 5 min and stored samples at 4°C. The annealing temperature was initially obtained from 5°C lower from the primer with the lowest melting temperature.

2.4.6 Agarose gel electrophoresis

Agarose gel electrophoresis was performed to visualise the DNA products (genomic DNA, the human oral metagenome, plasmids, PCR products, digestion products and ligation products). A standard concentration for gel was 1-1.5% (w/v) agarose (Bioline, London, UK) prepared with 1X TAE (tris-acetate-EDTA) buffer. Gels were stained with either a GelRed (1:10,000 dilution) (Biotium, Cambridge, United Kingdom) or ethidium bromide 0.5 µg/ml (Life Technologies, Paisley, UK). The DNA products (except the PCR products amplified by Biomix Red) were mixed with 5X loading buffer (Bioline, London, UK) and loaded into the gel wells. One microlitre of either HyperLadder 1kb (Bioline, London, UK) or 1 Kb extension ladder (Life Technologies, Paisley, UK) was also added as a size reference. The electrophoresis was run at 50-100 V for 60-90 min. Gels were visualised under UV excitation using an Alpha Imager (Alpha InnoTech, Exeter, UK) and the image was captured by AlphaView software (Alpha InnoTech, Exeter, UK).

2.4.7 PCR purification

PCR purification was performed by using the QIAquick PCR Purification Kit (Qiagen, Manchester, UK). This protocol was conducted to clean the PCR products by removing primers, enzymes, salts and other impurities. All of the centrifugation steps were done at $14680 \times g$ (13000 rpm, Eppendorf centrifuge 5415 D). Five volumes of Buffer PB was added to 1 volume of the PCR products. After mixing the solution, the mixture was transferred to the QIAquick spin column and centrifuged for 1 min. The flow-through was discarded and 700 μl of Buffer PE was then added to the column. After 1 min centrifugation, the flow-through was discarded from the collection tube, and additional centrifugation was done for 1 min. The spin column was then transferred to a new sterile 1.5 ml microcentrifuge tube. The DNA was eluted by adding 30 μl of molecular biology grade water to the centre of the membrane, left to stand for 3 min and centrifuged for 1 min. The purified DNA was kept in -20°C freezer.

2.4.8 Gel extraction

Gel extraction was performed with QIAquick Gel Extraction Kit (Qiagen, Manchester, UK). This protocol was done to purify and retrieve only specific DNA bands of interest on an agarose gel. All of the centrifugation steps were done at $14680 \times g$ (13000 rpm, Eppendorf centrifuge 5415 D). The DNA was subjected to the agarose gel electrophoresis (section 2.4.6). The DNA with the size of interest was then excised from the gel by visualising under UV light and cut with a clean scalpel. The gel slice was transferred to a 1.5 ml

microcentrifuge tube and weighed. Three volumes of Buffer QG was then added to 1 volume of gel and incubated in 50°C heat block. The tube was occasionally vortexed to help dissolve the gel. After the gel was completely dissolved, 1 volume of isopropanol was added and mixed. The mixture was then transferred to the QIAquick spin column and centrifuged for 1 min. After discarding the flow-through, 500 µl of buffer QG was added and centrifuged for 1 min. The flow-through was then discarded and 700 µl of buffer PE was added to the column. After 1 min centrifugation, the flow-through was discarded and then centrifuged for 1 min. The spin column was then transferred to a new sterile 1.5 ml microcentrifuge tube. The DNA was eluted by adding 30 µl of molecular biology grade water to the centre of the membrane, left to stand for 3 min and centrifuged for 1 min. The extracted DNA was kept in -20°C freezer.

2.4.9 Isopropanol precipitation

Isopropanol precipitation was performed to purify large DNA products (>10kb). An equal volume of isopropanol and 1/10 volume of sodium acetate (3 M, pH 5.2) was added to the DNA sample, followed by mixing. The tube was then incubated on ice for 30 min and centrifuged at 14680 x *g* (13000 rpm, Eppendorf centrifuge 5415 D) for 15 min. The supernatant was then discarded and the DNA pellet was rinsed with 1 ml of 70% ethanol. The tube was centrifuged at 14680 x *g* for another 15 min, and discarded the supernatant. The DNA pellet was air-dried for 5 min and resuspended in 30 µl of molecular biology grade water.

2.4.10 Restriction endonuclease reaction

DNA digestion was performed by using restriction enzymes (NEB, Hitchin, UK). The standard digestion reactions were prepared in a 10 μ l total volume, containing 1 μ l restriction enzymes (20 U), 1 μ l 10X digestion buffer, 1-5 μ l DNA samples and topped up with molecular biology grade water. The reactions were incubated at 37 °C for at least 1 hr for complete digestion unless stated otherwise. The reaction was then either purified by QIAquick PCR Purification Kit (Qiagen, Manchester, UK) with the same protocol as in section 2.4.7, or visualised on an agarose gel as described in section 2.4.6.

2.4.11 Dephosphorylation reaction

Dephosphorylation was performed to remove the 5' phosphate groups from the end of single enzyme digested vector prior to ligation. This was done to prevent the self-ligation of the vector during a ligation reaction. After the digestion of vector was completed, 1 μ l of Calf Intestinal Alkaline Phosphatase (CIAP) (1 U/ μ l) was added to the digestion mixture together with appropriate amount of 10X reaction buffer and molecular biology grade water. The reaction was then incubated at 37 °C for 30 min. Another 1 μ l of CIAP (1 U/ μ l) was then added to the reaction mixture and continued the incubation at 37 °C for 30 min. The reaction was then stopped and purified by QIAquick PCR Purification Kit (Qiagen, Manchester, UK), as described in section 2.4.7.

2.4.12 DNA ligation reactions

The standard ligation reaction was prepared with a total volume of 10 μ l, containing 1 μ l T4 DNA ligase (400 U/ μ l) (NEB, Hitchin, UK), 1 μ l 10X ligation buffer, vector, insert DNA and topped up with molecular biology grade water. The ligation mixture was incubated at 16°C for 1 hr. The amount of vector and insert DNA was added according to the molar ratio of 1:3 to 1:10 vector to insert, which was calculated by using the Equation 2-1:

$$\text{Equation 2-1: ng of insert} = \frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio}$$

2.4.13 Desalting of ligation products

The ligation products were desalted to remove salt from the reactions which could cause electric arcing during electroporation. An agarose cone was made by dissolving and heating 0.9 g glucose and 0.5 g agarose in 50 ml water (1.8 and 1% w/v respectively), then aliquoted 600 μ l into each 1.5 ml microcentrifuge tube. A 0.5 ml microcentrifuge tube was placed in the tube and allowed the agarose-glucose solution to be solidified. The 0.5 ml microcentrifuge tube was then removed, forming the agarose cone in the 1.5 ml microcentrifuge tube. The ligation products were transferred into the agarose cone and incubated on ice for 1 hr. The desalted product was then transferred to a new 1.5 ml microcentrifuge tube.

2.4.14 Preparation of chemically *E. coli* competent cells

The *E. coli* competent cells were prepared by following the protocol described previously (Chung *et al.*, 1989). The *E. coli* strain of interest was subcultured into LB containing appropriate antibiotics and incubated for 18 hr. Five millilitres of the overnight culture was then subcultured into a flask containing 50 ml of LB broth supplemented with the same antibiotics. The cells were grown until early exponential phase (OD₆₀₀ 0.3-0.4), then transferred to a 50 ml plastic tube. Cells were pelleted by centrifugation for 10 min at 4°C and 2500 x *g* (3000 rpm, Eppendorf centrifuge 5804 R). The supernatant was then removed by decanting, and the pellets were resuspended with 5 ml of cold TSS buffer (Appendix 2) by gently vortexing the tube. Three hundred microlitres of cells were then aliquoted into pre-chilled 2 ml cryotubes and kept at -80°C.

2.4.15 Transformation of *E. coli*

2.4.15.1 Chemical transformation

The chemical transformation was performed by using α -select silver efficiency *E. coli* (Bioline, London, UK) or the *E. coli* competent cells prepared from section 2.4.14. Fifty microlitres of the competent cells were aliquoted into a 1.5 ml microcentrifuge tube. The ligation product (2-5 μ l) or DNA (10 pg - 100 ng) was then added to the aliquoted competent cell and mixed by gently tapping the tube several times. The reaction was incubated on ice for 30 min and subjected to heat shock treatment in 42°C water bath for 40 sec. The cells were then immediately incubated on ice for 2 min. Nine hundred microlitres of

prewarmed SOC medium (Appendix 2) was added to the tube, then incubated in 37°C shaker for 1 hr. An aliquot of 100 µl of the cells was spread on LB agar containing appropriate antibiotics. The rest of cells were centrifuged at 4000 x g (8000 rpm, Eppendorf centrifuge 5415 D) for 1 min. The supernatant was removed to leave approximately 100 µl in the tube. The cells were resuspended and spread on LB agar containing appropriate antibiotics. All plates were incubated at 37°C for 18 hr.

2.4.15.2 Electroporation

Electroporation was performed by using TransforMax™ EPI300™ Electrocompetent *E. coli* (Cambio, Cambridge, UK). Fifty microlitres of the electrocompetent cells were aliquoted into a pre-chilled 1.5 ml microcentrifuge tube. Two microlitres of desalted ligation products (section 2.4.13) or DNA (10 pg - 100 ng) was added to the competent cells. The mixture was then transferred to a pre-chilled electroporation cuvette 0.1 cm (Bio-Rad, Hertfordshire, UK) without introducing air bubbles in the cuvette. After wiping the condensation on the cuvette, it was transferred and placed in the electroporator (The Gene Pulser II electroporation system, Bio-Rad, Hertfordshire, UK). The condition for electroporation was set to 1.7 kV, 25 µF, 200 Ω. The electric pulse was then applied by pressing both of the pulse buttons at the same time until the beep sound. Nine hundred fifty microlitres of prewarmed SOC medium (Appendix 2) was immediately added to the cuvette and mixed gently by pipetting. The cells were then transferred to a 20 ml sterile plastic tube and incubated in 37°C shaker for 1 hr. An aliquot of 100 µl of the cells was spread on LB agar containing appropriate antibiotics. The

rest of cells were centrifuged at 4000 x g (8000 rpm, Eppendorf centrifuge 5415 D) for 1 min. The supernatant was removed which left approximately 100 µl in the tube. The cells were resuspended and spread on LB agar containing appropriate antibiotics. All plates were incubated at 37°C for 18 hr.

2.4.16 Site-directed mutagenesis

Site-directed mutagenesis was performed by using the Q5[®] Site-Directed Mutagenesis Kit (NEB, Hitchin, UK). It is a method to alter the sequences of plasmid DNA (substitution, insertion, and deletion). The desired mutation can be introduced by PCR with specific primers (Substitution: changing the nucleotides in the centre of the primers, Deletion: designing primers to amplify outward from the deletion region, Insertion: adding the extra nucleotides at the 5' end of each primer).

After designing the primers for mutagenesis, a PCR reaction was prepared with 25 µl total volume composing of 12.5 µl 2X Q5 master mix (NEB, Hitchin, UK), 1.25 µl of each primer (10 pmol/µl), 9 µl molecular biology grade water and 1 µl plasmid template. The PCR cycle was as followings; (i) initial denaturation at 98°C for 30 sec (ii) denaturation at 98°C for 10 sec (iii) annealing at 50-65°C (depending on primers) for 30 sec (iv) extension at 72°C for 1-5 min (depending on the expected size of amplicons), repeated step (ii)-(iv) for 35 cycles (v) final elongation at 72°C for 2 min and stored samples at 4°C.

The PCR products were visualised and verified by agarose gel electrophoresis. If the amplicons were correct in size, KLD (Kinase-Ligase-

DpnI) enzyme reaction is then set up to phosphorylate the 5' end of amplicons, recircularised the amplicons and removed the template DNA. The reaction composed of 1 µl 10X KLD enzyme mix (NEB, Hitchin, UK), 5 µl 2X KLD reaction buffer, 1 µl PCR product and 3 µl molecular biology grade water. The reaction was incubated at room temperature (25°C) for 5 min. The KLD-treated product was then transformed into *E. coli* as described in section 2.4.15.

2.5 DNA sequencing reactions

The sequencing of plasmids and PCR products were performing by Genewiz, formerly Beckman Coulter Genomics (Essex, UK). The concentration of DNA products was determined by NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, Surrey, UK). Ten microlitres of plasmids (100 ng/µl) and 5 µl of PCR products (50 ng/µl) were prepared for sequencing with appropriate primers.

2.6 DNA sequence analysis

The sequencing results were analysed by using BioEdit version 7.2.0 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) and SnapGene version 3.2.1. For the samples sequenced with more than one reaction, the sequences were combined by using the CAP contig function in the BioEdit software (Huang, 1992). The comparison of DNA sequences to the nucleotide and protein database by using the National Centre for Biotechnology Information (NCBI) tools, BlastN and BlastX, respectively (Altschul *et al.*, 1990). The sequence alignment was performed by using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers *et al.*, 2011).

Chapter 3

Detection of Integron Gene Cassettes in the Human

Oral Metagenomic DNA

3.1 Introduction

Integron gene cassettes (GCs) have been identified in various environments such as deep sea hydrothermal vents, soil, river and marine sediment, hot springs, seawater and freshwater biofilms (Stokes *et al.*, 2001, Holmes *et al.*, 2003, Elsaied *et al.*, 2007, Koenig *et al.*, 2008, Gillings *et al.*, 2009, Elsaied *et al.*, 2011). An analysis of GCs in metagenomic DNA has shown that most of the GCs and their encoded proteins are novel and have no known homologs in nucleotide and protein databases (Gillings, 2014). Approximately 20% of cassettes were matched to known homologues, which were predicted to encode proteins with diverse functions and likely to be niche-specific (Boucher *et al.*, 2007).

The targeted metagenomic approach is one of the techniques that has been used to investigate the diversity of integron GCs. Several conserved sequences of integrons were used to design the DNA primers for the amplification of GCs, including the *intl* (5' conserved segment), *qacEΔ1* and *sull* (3' conserved segment) and *attC* (GC), as shown in Figure 3-1 (Holmes *et al.*, 2003).

Even though the human oral cavity contains a complex microbiome and its environment facilitates gene transfer, only two major studies have been reported on integrons in the human oral cavity. The first study described an integron in an oral bacteria *Treponema denticola* ATCC35405 by using whole genome sequencing (Coleman *et al.*, 2004). It is called an unusual integron, which has an integrase gene oriented in the same direction as the GC array,

while the integrase gene of normal integrons are oriented in the opposite direction with the GC array (Figure 3-1). Another study is the *in silico* analysis of the metagenomic data of the Human Microbiome Project to detect integron-containing *Treponema* spp. (Wu *et al.*, 2012). However, the presence of integrons in other oral bacteria remains to be determined.

In this chapter, we investigated the presence of integron GCs in the human oral metagenomic DNA using a PCR-based approach. Different, existing primer sets were used and new primers based on the *T. denticola* oral integron were designed, targeting various conserved regions on different integrons.

3.2 Materials and methods

3.2.1 Extraction of the human oral metagenomic DNA

The human oral metagenomic DNA was extracted from the collected saliva samples as described in section 2.3.2.

3.2.2 Recovery of integron GCs from oral metagenome

The integron GCs were amplified from the human oral metagenome by using two different groups of DNA primers (Figure 3-1). The first group was the primers previously described by other research groups for the GC PCR, including HS286, HS287 (Stokes *et al.*, 2001), HS298 (Nield *et al.*, 2001), 5'CS and 3'CS primers (Martinez-Freijo *et al.*, 1998). The second group contained the newly designed primers based on the *attC* sequence of the *Treponema denticola* integron, the only described integron in oral bacteria to date

(Coleman *et al.*, 2004), which were SUPA3, SUPA4, SUPA5, SUPA6, Flip-SUPA3 and Flip-SUPA4 primers.

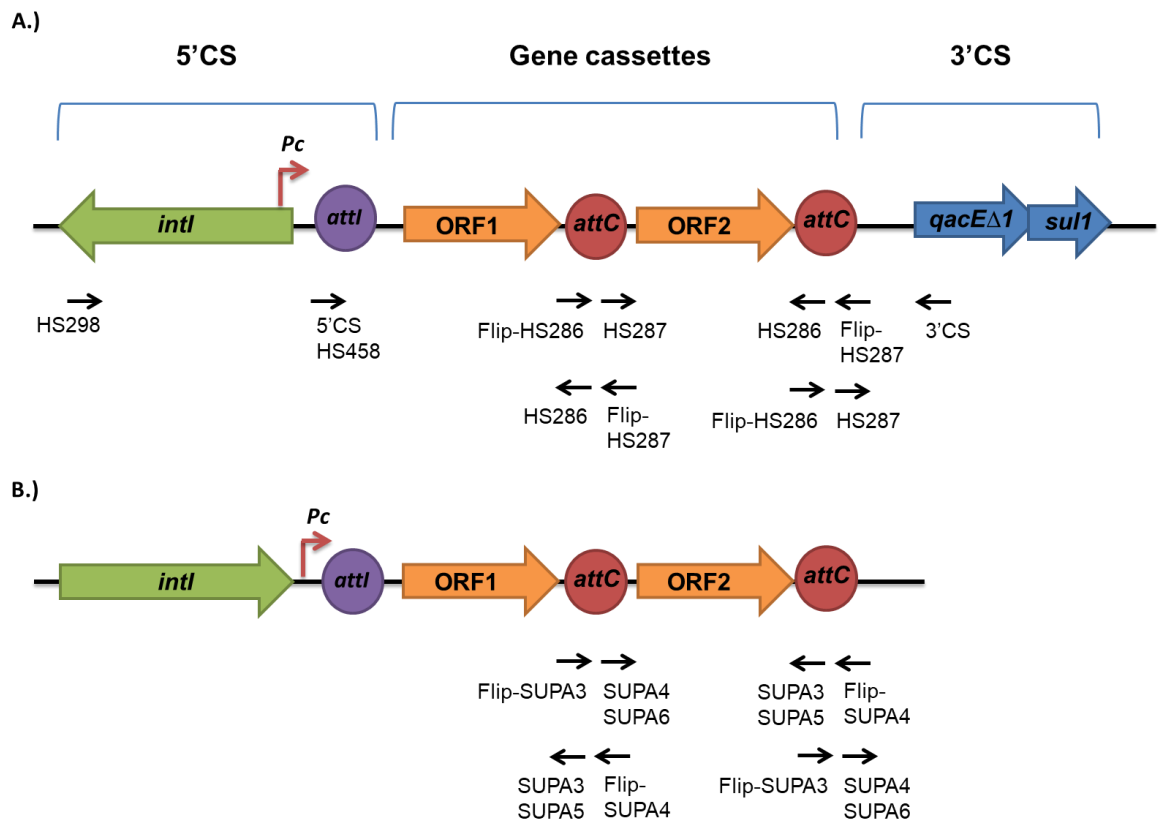


Figure 3-1 A schematic representation of primer binding sites on integron. The black arrows show the primer binding sites of each primer on A.) the class 1 integrons and B.) the unusual integron structure of *T. denticola*. The genes in 5' and 3' conserved segment (CS), the *attI* sites, *attC* sites, and the open reading frame (ORF) are shown in green, blue, purple, red and orange, respectively. The red arrows represent *P_C* promoter. (Adapted from Tansirichaiya *et al.*, 2016b)

Different combinations of primers were used in the PCRs on the oral metagenome. The PCR products were purified (section 2.4.7) and ligated to pGEM-T easy vectors (section 2.4.12). The ligation products were then introduced into *E. coli* α-select silver efficiency competent cells by DNA transformation (section 2.4.15.1). The plasmids of the white colonies,

indicating they contain inserts, were then extracted and sequenced with M13F and M13R primers (Appendix 4). The samples were then analysed to determine whether the inserts were integron GCs based on several criteria, as follows:

- The inserts were not amplified from the human genome.
- Both forward and reverse primer binding sites could be found on the inserts.
- If a potential ORF could be identified, start and stop codons should be located in between the *attC* sites with a potential ribosome binding site (AGGAGG consensus sequence) upstream of the start codon.
- For the amplicons amplified using SUPA3-4 and SUPA5-6 primers, the complementary R' (1R) core [GTTRR(Y)R(Y)Y(R)] and R'' (1L) core [R(Y)Y(R)Y(R)YAAC] were identified. (Degenerate nucleotides: R = A or G; Y = C or T)

3.2.3 Nomenclature of the recovered integron gene cassettes

The recovered GCs were named according to the primers and the source of the oral metagenome. The first two letters represent the forward and reverse primers used in amplification, respectively. The third letter is U to indicate that GCs were amplified from the UK oral metagenomic sample, as there was another parallel project investigating integron GCs from Bangladeshi oral metagenome, carried out by Mr Md. Ajijur Rahman, UCL Eastman Dental Institute (Tansirichaiya *et al.*, 2016b). The last digit indicated a numerical code

for the number of the clone. The sequences of detected GCs were submitted to the GenBank with the accession numbers from KT921474 to KT921495.

3.2.4 Detection of the circular form of integron gene cassettes

Another approach, to verify whether the detected amplicons were GCs, was to detect for a circular form of each GC. As one of the important characteristics of GCs was that they could be excised into a circular form, proving the presence of their circular forms in oral metagenomic DNA, hence, would be strong evidence to confirm that they were GCs. Another set of primers were therefore designed to amplify outwards from each putative GC and used in PCR to verify the presence of the circular forms in the oral metagenomic DNA. If the circular GC exists, the amplicon should contain *attC* and a part of GC sequences, as illustrated in Figure 3-2.

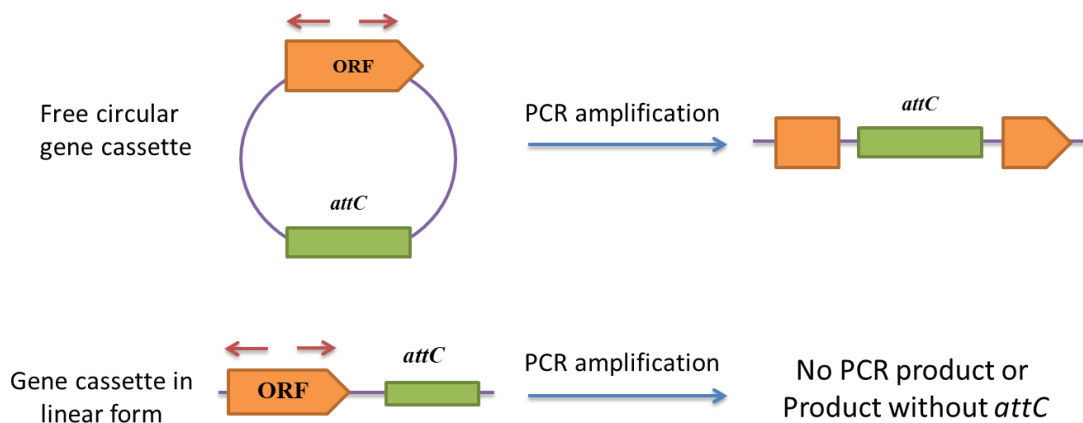


Figure 3-2 Schematic presentation of a method to verify the gene cassette. Primers (red arrows) are designed to bind and amplify outward from the ORF (orange arrow boxes). The green boxes represent *attC* sites.

3.2.5 Functional screening of integron gene cassette PCR products

The integron GCs were subjected to functional screening to detect for the GCs containing ARGs. SUPA3 and SUPA4 primers were modified by adding restriction sites and six extra nucleotides to serve as a GC clamp, in order to directionally clone the amplicons into pUC19. PCR reactions were performed by using five different combinations of PCR primers, based on the compatibility of restriction enzymes on the primers (*Pst*I-Acc65I, *Sal*I-Acc65I, *Xba*I-*Bam*HI, *Xba*I-*Sac*I, *Xba*I-Acc65I). The amplicons were then double digested according to the restriction sites and ligated into pUC19, pre-digested with the same pairs of restriction enzymes. The ligation products were transformed into *E. coli* α -select silver efficiency competent cells and screened on LB agar supplemented with ampicillin (100 μ g/ml) and one of the screening antibiotics [tetracycline (10 μ g/ml), kanamycin (20 μ g/ml), streptomycin (32 μ g/ml) and trimethoprim (20 μ g/ml)]. These antibiotics were selected because streptomycin, kanamycin and trimethoprim resistance genes were the most common resistance genes associated with integrons (Partridge *et al.*, 2009), and several tetracycline resistance genes were previously found in oral metagenomic DNA (Seville *et al.*, 2009).

3.3 Results

3.3.1 Verification of the human oral metagenome

The modified Gram-positive bacteria and yeasts protocol was suitable for the extraction of oral metagenomic DNA from saliva samples. From 1.5 ml of anonymised and pooled saliva sample, the concentrations of extracted DNA

ranged from 80 ng/μl to 175 ng/μl in a volume of 400 μl. Two microliters of three DNA extractions were run on 1% agarose gel together with 1 μl of HyperLadder™ 1kb, which they all showed a high-intensity band above 10 kb (Figure 3-3).

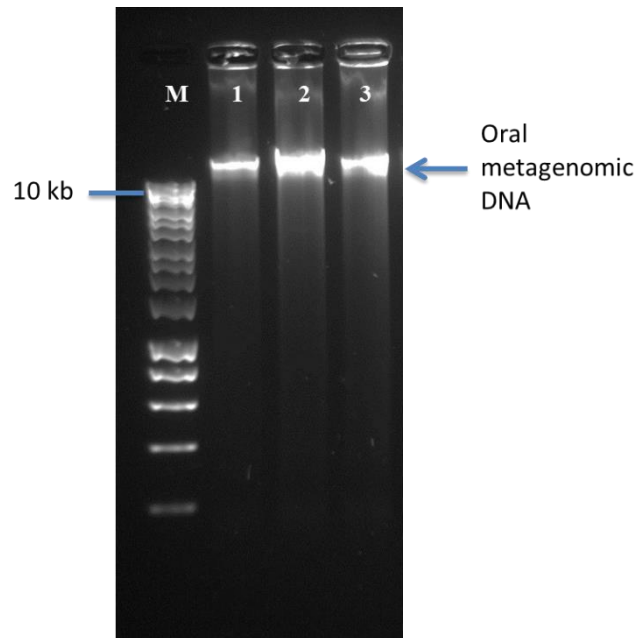


Figure 3-3 Agarose gel electrophoresis of the extracted human oral metagenomic DNA on 1% agarose gel. Lane M, HyperLadder™ 1kb. Oral metagenome showed in lane 1,2 and 3, had a concentration of 80, 175 and 150 ng/μl, respectively.

The microbial composition and diversity of the extracted oral metagenomic DNA were determined by 16S rRNA gene sequencing on the MiSeq Illumina platform. The results of 16S rRNA sequencing were shown in Figure 3-4 in which the bacterial composition was categorised based on their order. The most abundance order was Bacteroidales, following by Clostridiales, Lactobacillales, and Actinomycetales.

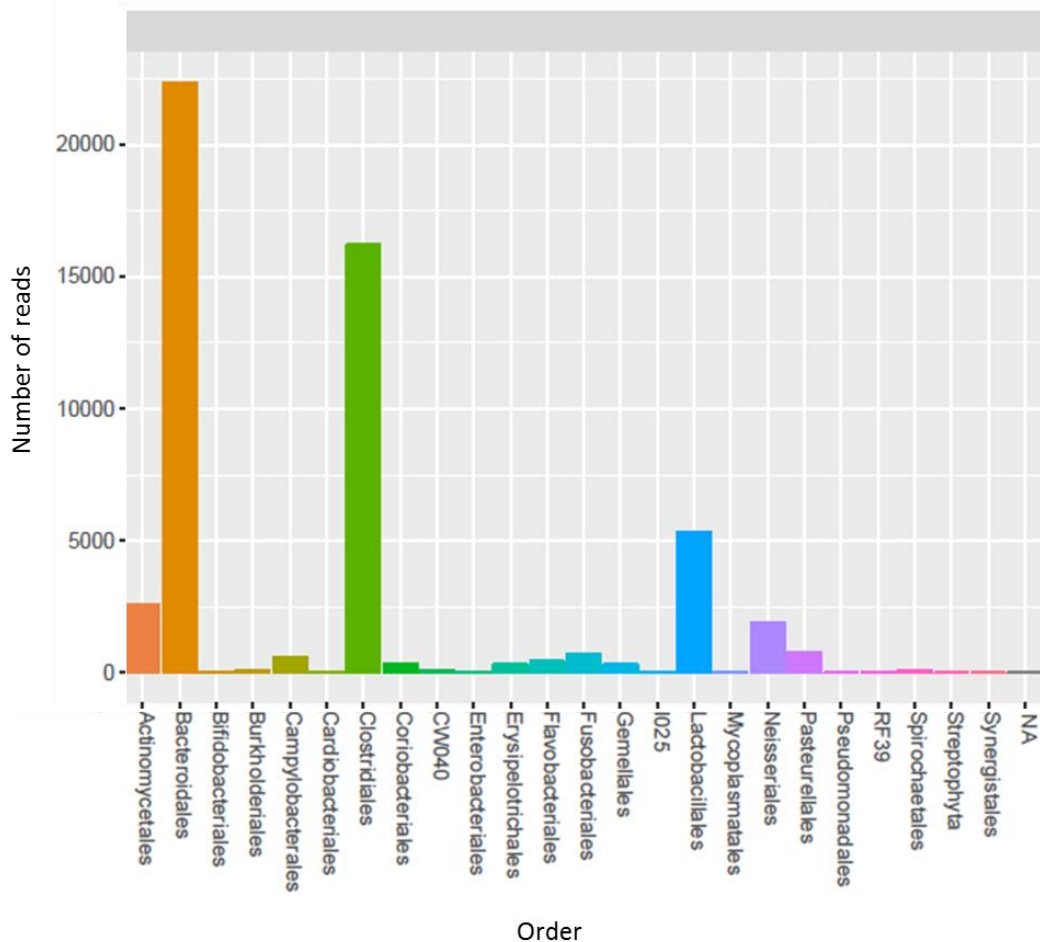


Figure 3-4 The composition and distribution of the bacterial communities in the extracted oral metagenome, based on 16S rRNA gene sequences.

3.3.2 attC primer design

The *Treponema denticola* attC sequence (accession number NC_002967) was subjected to BlastN search to retrieve the similar attC sequences and identify any conserved region for primer design. Forty-seven sequences were retrieved and aligned according to their nucleotide homology by using Clustal Omega (Figure 3-5). The conserved regions were then identified and used to

design the primers reading outward from the *attC* site by using Primer3. Two sets of primers were constructed; SUPA3-SUPA4 (without degenerate nucleotide) and SUPA5-SUPA6 (with degenerate nucleotides).

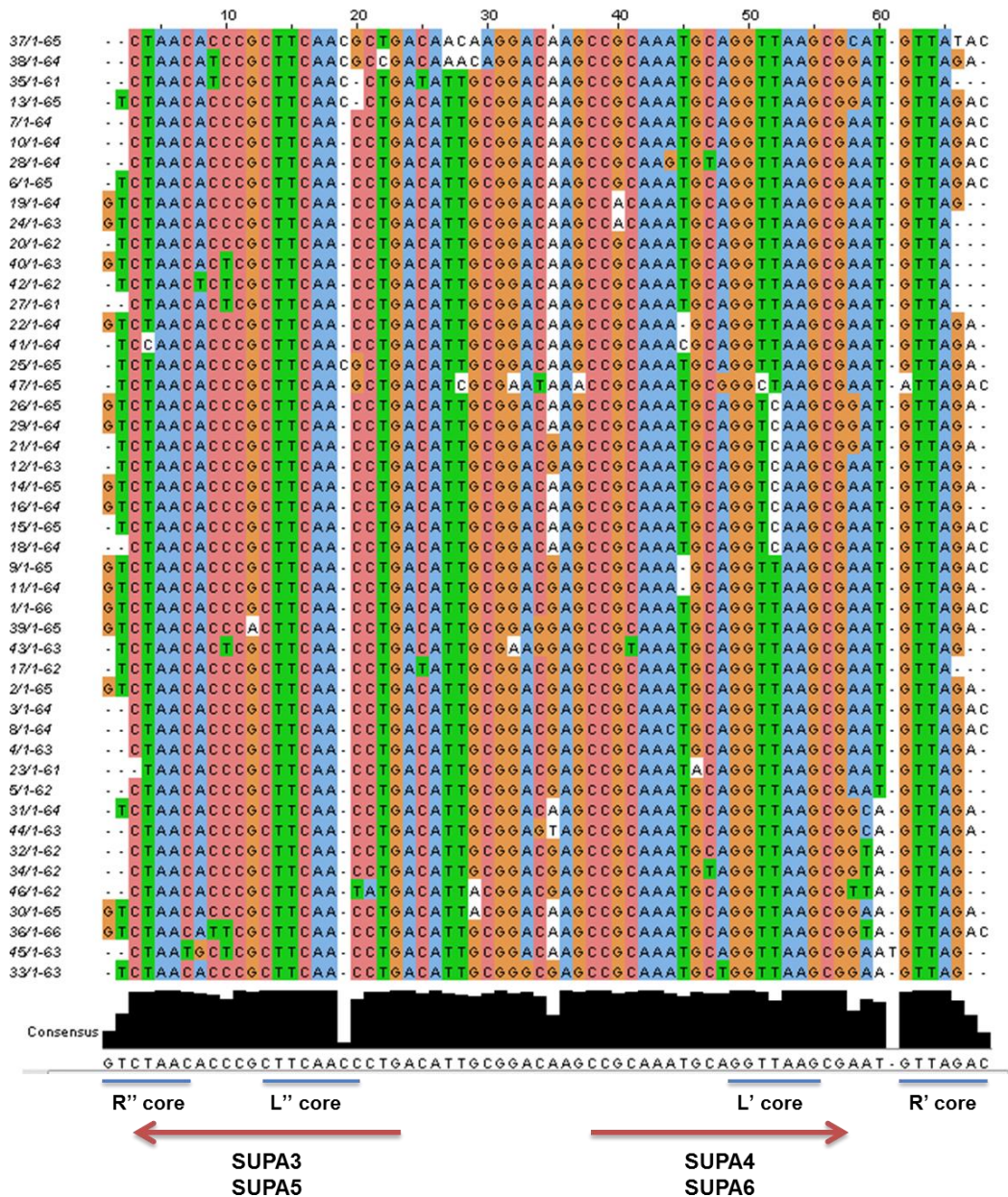


Figure 3-5 Alignment of *Treponema denticola attC* sequences and the similar sequences retrieved from BlastN by Clustal Omega. The red arrows represent the primer binding sites of SUPA3, SUPA4, SUPA5, SUPA6 primers. The blue lines represent L', L'', R' and R'' core sites on *attC*. The black chart represents the conserved percentage of nucleotides from the alignment.

3.3.3 Amplification of integron gene cassettes from oral metagenome

The DNA primers targeting different sites on integrons were selected for the PCR amplification to amplify GCs from oral metagenomic DNA, as shown in Figure 3-1. The libraries of PCR amplicons were constructed as described in section 3.2.2. The combination of primers and the sequence analysis results from each library was summarised in Table 3-1. It was shown that the previously published primers, used to amplify integron GCs from the other environmental samples, could not amplify integron GCs from the human oral metagenome.

Table 3-1 The primer pairs and the sequence analysis results of each library.



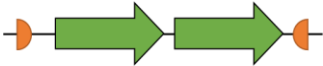
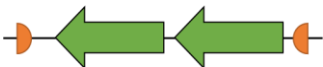


Primer pairs	Expected amplicons	Number of samples sent for sequencing	Sequence analysis		
			Gene cassettes	Human DNA	Mispriming
SUPA3-SUPA4	From one <i>attC</i> to another <i>attC</i>	35	25	0	10
SUPA5-SUPA6	From one <i>attC</i> to another <i>attC</i>	10	8	0	2
HS286-HS287	From one <i>attC</i> to another <i>attC</i>	9	0	6	3
Flip-SUPA3-Flip-SUPA4	From one <i>attC</i> to another <i>attC</i>	6	0	2	4
Flip-HS286-Flip-HS287	From one <i>attC</i> to another <i>attC</i>	3	0	0	3
5'CS-3'CS	From 5' conserved sequence to 3' conserved	20	0	13	7

Primer pairs	Expected amplicons	Number of samples sent for sequencing	Sequence analysis		
			Gene cassettes	Human DNA	Mispriming
5'CS-SUPA3	From 5' conserved sequence to one of <i>attC</i>	10	0	7	3
5'CS-SUPA4	From 5' conserved sequence to one of <i>attC</i>	9	0	0	9
HS298-HS286	From 5' conserved sequence to one of <i>attC</i>	10	0	9	1

* The PCRs performing with newly designed primers were highlighted in yellow.

The newly designed primers based on *T. denticola* integron *attC* were the only primer sets that could amplify the PCR products that followed the criteria for putative integron GCs from the human oral metagenome. Twenty-two different GCs were identified with the size between 425 - 1138 bp. From 22 GCs, 20 of them carried one or more putative open reading frames (ORFs) with the total number of 33 different ORFs. The other two GCs, SSU9 and SSU29, were predicted to contain no ORF and considered as empty or noncoding GCs. The arrangement of the recovered GCs could be categorised into 6 different groups, according to the number, direction, and position of ORFs in the GCs (Table 3-2). The orientation that could be found most was type A, which contained single ORF, transcribed in the forward direction (based on the position of *attC*).

Table 3-2 Orientation of ORFs in the GCs recovered from metagenomic DNA of saliva samples.

Orientation Type	Orientation of ORFs in GCs*	Number of unique gene cassettes
A		9
B		1
C		7
D		1
E		2
F		2
Total GCs		22

*The orange half circles and green arrow boxes are representing *attC* sites and ORFs, respectively.

As SUPA3/SUPA5 (reverse primers) and SUPA4/SUPA6 primers (forward primers) contained the consensus L' (2R) and L'' (2L) core sites of *attC*, respectively, if the putative GCs were not PCR artefacts, the other core sites, the R' (1R) and R'' (1L), should be found in the sequences. Most of the GCs (16 out of 22 GCs) showed 100% complementary between the R' and R'' core sites, while the rest showed 6 out of 7 bp complimentary nucleotides (Table 3-3).

Table 3-3 Complementary of the R' and R'' cores sites on the recovered gene cassettes.

Type (Number of GCs)	Sequence of R' after the forward primer sequence of the <i>attC</i> on GCs	Pattern of R' sequence of the GC^a	Sequence of R'' before the reverse primer of the <i>attC</i> on GCs GC	Pattern of R'' sequence of the <i>attC</i> on GCs GC^a	Clones	Complementarity between R' and R'' core sites of the <i>attC</i> on GCs
A (13)	GTTAGAC	GTTRRRY	GTCTAAC	RYYYAAC	SSU1 SSU9 SSU10 SSU21 SSU29	7/7
	GTTAGAT	GTTRRRY	ATCTAAC	RYYYAAC	SSU12 SSU15 SSU16 SSU17 SSU18	7/7
	GTTAGGC	GTTRRRY	GCCTAAC	RYYYAAC	SSU7	7/7
	GTTAGGT	GTTRRRY	GCCTAAC	RYYYAAC	SSU25	7/7
	GTTGAAC	GTTRRRY	ATCTAAC	RYYYAAC	SSU5	7/7
B (4)	GTTATAC	GTTRYRY	GCCTAAC	RYYYAAC	SSU8	6/7
	GTTATAC	GTTRYRY	GTCTAAC	RYYYAAC	SSU26	6/7
	GTTATGT	GTTRYRY	ACCTAAC	RYYYAAC	SSU6 SSU28	6/7






Type (Number of GCs)	Sequence of R' after the forward primer sequence of the <i>attC</i> on GCs	Pattern of R' sequence of the GC ^a	Sequence of R'' before the reverse primer of the <i>attC</i> on GCs GC	Pattern of R'' sequence of the <i>attC</i> on GCs GC ^a	Clones	Complementarity between R' and R'' core sites of the <i>attC</i> on GCs
C (4)	GTTAAGA	GTTRRRR	TCTTAAC	YYYYAAC	SSU3	7/7
	GTTAGAA	GTTRRRR	GTCTAAC	RYYYYAAC	SSU11 SSU22	6/7
	GTTAGGA	GTTRRRR	TCTTAAC	YYYYAAC	SSU24	7/7
D (1)	ATTAGAC	ATTRRRY	ATCTAAC	RYYYYAAC	SSU27	7/7


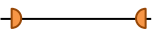




^a Degenerate nucleotides: R = A or G; Y = C or T.






The details and characteristics of the recovered integron GCs are shown in Table 3-4. The sequence analysis with BlastN and BlastX suggested that most of the GCs were from *Treponema* spp. (19 out of 22 GCs, 86.4%). Most of them were homologous with *T. denticola*, while the others were matched with *T. pedis*, *T. putidum*, *T. medium* and *T. vincentii*. Three GCs were predicted to encode non-treponemes proteins from *Rhodonellum psychrophilum*, *Paenibacillus assamensis*, and *Bradyrhizobium* sp. However, the percentages of similarity of these ORFs were lower than 70% at amino acid level.






From 33 ORFs, 29 of the predicted proteins (87.9%) showed a homologue in the protein database, in which half of them were matched with hypothetical proteins and another half were matched with proteins with known function. The functions of detected ORFs were predicted to be stress adaptation (ensuring bacterial survival in adverse environmental stressors such as pH, oxidative stress and nutrient variability), competence, toxin-antitoxin systems and antibiotic resistance. Putative antibiotic resistance encoding ORFs were identified in SSU26 (cof-like hydrolase) and SSU28 (multidrug transporter MatE). Stress adaptation and toxin-antitoxin systems were identified in sample SSU5 (Twitching motility protein PilT) and SSU21/27 (Toxin-antitoxin), respectively. Three putative ORFs were predicted to encode novel proteins with no match in GenBank (e-value <0.001).


Table 3-4 Characterization of all gene cassettes detected in the saliva metagenomic DNA using attC-based primers

Gene cassettes/clone code	Primer pair	Cassette Size (bp)	Orientation*	Distance between attC and ORF (bp)	Accession number	BlastN			BlastX					
						Closest homologue	Percentage identity (%)	Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)	Coverage (%)	The presence of ribosomal binding site	Accession number of the homologous proteins (BlastX)
SSU1/2	SUPA3-SUPA4	799		36	KT921474	No significant similarity found.	-	-	Hypothetical protein [Rhodonellum psychrophilum]	579	66	75	Yes	WP_026333632.1
SSU3/4/30	SUPA3-SUPA4	964		41	KT921475	No significant similarity found.	-	-	Glyoxalase [Treponema pedis]	390	97	100	Yes	WP_009105863.1
									Competence protein TfoX [Treponema pedis]	315	83	100	Yes	WP_024470244.1
SSU5	SUPA3-SUPA4	832		136	KT921476	Treponema sp.	88	27	Hypothetical protein [Treponema putidum]	243	67	100	Yes	WP_044978234.1
									Twitching motility protein PilT [Treponema putidum]	396	71	100	Yes	AIN93467.1
SSU6	SUPA3-SUPA4	1138		34	KT921477	No significant similarity found.	-	-	No significant similarity found.	459	-	-	Yes	-
									Hypothetical protein [Treponema denticola]	339	75	100	Yes	WP_044013590.1
SSU7	SUPA3-SUPA4	491		0	KT921478	Treponema denticola	92	40	No significant similarity found	162	-	-	No	-

Gene cassettes/clone code	Primer pair	Cassette Size (bp)	Orientation*	Distance between attC and ORF (bp)	Accession number	BlastN			BlastX					
						Closest homologue	Percentage identity (%)	Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)	Coverage (%)	The presence of ribosomal binding site	Accession number of the homologous proteins (BlastX)
									No significant similarity found	132	-	-	Yes	-
SSU8	SUPA3-SUPA4	921		41	KT921479	<i>Treponema sp.</i>	98	49	Hypothetical protein [<i>Treponema denticola</i>]	567	98	93.7	Yes	WP_002692239.1
SSU9/13/14/19/20/23	SUPA3-SUPA4	425		-	KT921480	<i>Treponema pedis</i>	78	69	No significant similarity found	-	-	-	-	-
SSU10	SUPA3-SUPA4	612		29	KT921481	No significant similarity found.	-	-	Hypothetical protein [<i>Paenibacillus assamensis</i>]	537	39	91	Yes	WP_028595336
SSU11	SUPA3-SUPA4	648		4	KT921482	No significant similarity found.	-	-	Hypothetical protein [<i>Bradyrhizobium sp.</i> STM 3809]	597	42	89.7	Yes	WP_035659994.1
SSU12	SUPA3-SUPA4	736		77	KT921483	No significant similarity found.	-	-	Hypothetical protein [<i>Treponema vincentii</i>]	582	99	100	Yes	WP_016518887.1
SSU15	SUPA3-SUPA4	989		105	KT921484	<i>Treponema denticola</i>	90	19	Hypothetical protein [<i>Treponema sp.</i>]	396	40	44.3	Yes	WP_044015417.1
									Hypothetical protein [<i>Treponema denticola</i>]	228	39	16.4	No	WP_010689034.1
									Hypothetical protein TPE_0657 [<i>Treponema pedis</i>]	147	96	100	Yes	AGT43153.1

Gene cassettes/clone code	Primer pair	Cassette Size (bp)	Orientation*	Distance between attC and ORF (bp)	Accession number	BlastN			BlastX					
						Closest homologue	Percentage identity (%)	Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)	Coverage (%)	The presence of ribosomal binding site	Accession number of the homologous proteins (BlastX)
SSU16	SUPA3-SUPA4	711		127	KT921485	<i>Treponema pedis</i>	90	21	Hypothetical protein [<i>Treponema sp.</i>]	363	33	41.6	Yes	WP_044015417.1
									Hypothetical protein TPE_0657 [<i>Treponema pedis</i>]	147	96	100	Yes	AGT43153.1
SSU17	SUPA3-SUPA4	787		30	KT921486	<i>Treponema denticola</i>	98	100	Hypothetical protein [<i>Treponema denticola</i>]	711	97	100	Yes	WP_002690335.1
SSU18	SUPA3-SUPA4	853		31	KT921487	No significant similarity found	-	-	Hypothetical protein (Endonuclease) [<i>Treponema putidum</i>]	765	33	99.2	Yes	WP_044978378.1
SSU21	SUPA3-SUPA4	871		273	KT921488	<i>Treponema denticola</i>	97	100	BrnT Toxin [<i>Treponema denticola</i>]	291	96	100	No	WP_010692226.1
									Hypothetical protein (BrnA antitoxin) [<i>Treponema denticola</i>]	258	100	100	Yes	WP_010692225.1
SSU22	SUPA3-SUPA4	1017		72	KT921489	No significant similarity found	-	-	Prevent-host-death family protein (YefM antitoxin) [<i>Treponema medium</i>]	264	94	100	Yes	WP_016523165.1
									Hypothetical protein (Transcriptional regulator) [<i>Treponema medium</i>]	441	97	100	Yes	WP_016523167.1

Gene cassettes/clone code	Primer pair	Cassette Size (bp)	Orientation*	Distance between attC and ORF (bp)	Accession number	BlastN			BlastX					
						Closest homologue	Percentage identity (%)	Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)	Coverage (%)	The presence of ribosomal binding site	Accession number of the homologous proteins (BlastX)
SSU24	SUPA3-SUPA4	962		23	KT921490	No significant similarity found	-	-	Glyoxalase [<i>Treponema pedis</i>]	411	97	100	No	WP_024470245.1
									Competence protein TfoX [<i>Treponema pedis</i>]	222	88	55.2	No	WP_024470244.1
SSU25	SUPA5-SUPA6	789		31	KT921491	<i>Treponema putidum</i>	96	99	Hypothetical protein [<i>Treponema denticola</i>]	705	97	100	Yes	WP_010697531.1
SSU26	SUPA5-SUPA6	848		34	KT921492	<i>Treponema putidum</i>	81	93	Cof-like hydrolase [<i>Treponema denticola</i>]	768	76	100	Yes	WP_010693073.1
SSU27	SUPA5-SUPA6	833		155	KT921493	<i>Treponema denticola</i>	96	100	No significant similarity found	126	-	-	-	-
									Antitoxin HicB [<i>Treponema denticola</i>]	405	99	100	Yes	WP_002669522.1
									Toxin HicA [<i>Treponema denticola</i>]	195	97	100	Yes	WP_002669524.1
SSU28	SUPA5-SUPA6	927		310	KT921494	<i>Treponema putidum</i>	97	91	Multidrug transporter MatE [<i>Treponema putidum</i>]	336	96	100	Yes	WP_044979179.1
									mRNA-degrading endonuclease [<i>Treponema denticola</i>]	231	99	100	Yes	WP_010694033.1

Gene cassettes/clone code	Primer pair	Cassette Size (bp)	Orientation*	Distance between attC and ORF (bp)	Accession number	BlastN			BlastX					
						Closest homologue	Percentage identity (%)	Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)	Coverage (%)	The presence of ribosomal binding site	Accession number of the homologous proteins (BlastX)
SSU29	SUPA5-SUPA6	425		-	KT921495	<i>Treponema pedis</i>	78	69	-	-	-	-	-	-

*The orange half circles and green arrow boxes are representing *attC* sites and ORFs, respectively.

3.3.4 Detection of circular forms of gene cassettes

The presence of the circular form of 22 GCs, identified in previous section, in oral metagenomic DNA was determined. Twenty-two pairs of primers were designed and used in PCR amplification on the original oral metagenomic DNA. However, none of the circular forms were identified by this PCR.

3.3.5 Functional screening of gene cassette PCR products

Four replicates of the functional screening of SUPA3-SUPA4 PCR products were performed. Four colonies, labelled S34tet1-4, were found on LB agar plates supplemented with ampicillin and tetracycline. By the retransformation of extracted plasmid into *E. coli*, it was confirmed that the tetracycline resistance trait was conferred by the genes on the plasmid. Plasmid digestion with *Hind*III suggested that all four tetracycline resistance clones were likely to contain the same insert with a size of 2 kb, comparing to pUC19 (Figure 3-6). The plasmids were sequenced with primers flanking MCS of pUC19 (M13F and M13R) and the primers used to amplify the insert (SUPA3 and SUPA4). However, all sequencing reactions failed.

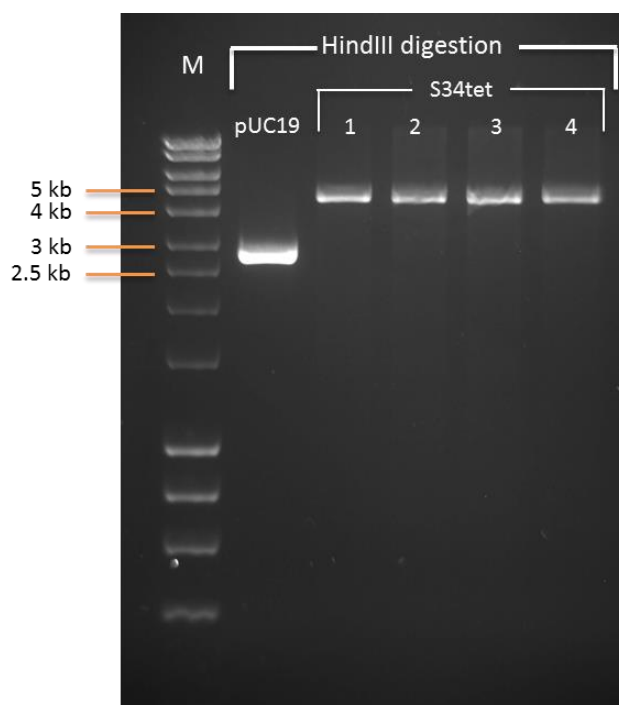


Figure 3-6 Agarose gel electrophoresis of *Hind*III digested products of pUC19 and S34tet1-4 plasmids on 1% agarose gel. Lane M, HyperLadder™ 1kb.

The S34tet plasmid was then subcloned into pUC18 vector by ligating partial *Sau*3AI-digested S34tet plasmid with *Bam*HI-digested pUC18 vector. The clones with tetracycline resistance were selected and subsequently sequenced their plasmids with M13F and M13R primers. The sequencing results suggested that S34tet plasmid was not pUC19 with DNA insert, but it was pBR322 plasmid which contained both ampicillin and tetracycline resistance genes (Figure 3-7A). This was also confirmed by digesting S34tet plasmid with *Sau*3AI, which showed the similar digestion profile as *Sau*3AI-digested pBR322 vector (Figure 3-7B and C). The pBR322 was likely to be a contaminant, however no one in the department used it so its origin is inexplicable.

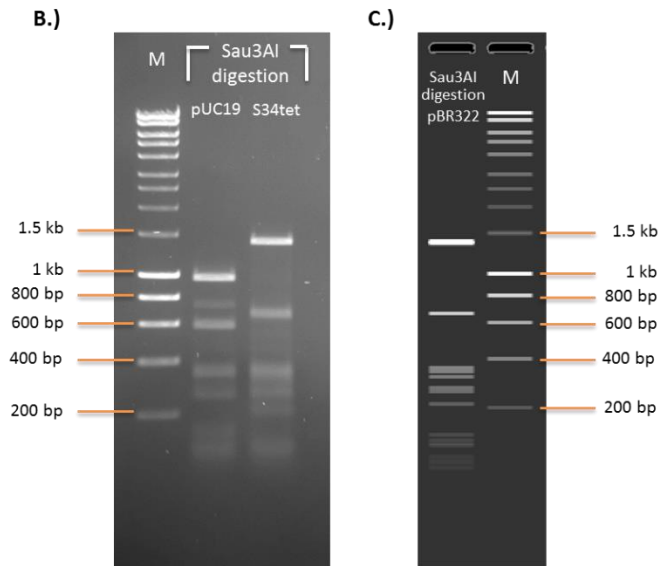


Figure 3-7 Schematic representation of pBR322 plasmid and the *Sau3AI* digestion analysis. A.) Schematic representation of pBR322 plasmid containing ampicillin (Ap^R) and tetracycline (Tc^R) resistance genes. (Retrieved from Balbas *et al.*, 1986) B.) The *Sau3AI* digestion products of pUC19 and Stet34 plasmid were separated on 1% agarose gel. C.) The *Sau3AI* digestion product of pBR322 plasmid on 1% agarose gel was simulated by SnapGene software. Lane M, HyperLadder™ 1kb.

3.4 Discussion

Prior to the detection of integron GCs in oral metagenome, it is important to verify that the extracted metagenome represented the oral microbiome. The 16S rRNA gene sequencing showed that the extracted oral metagenome was composed of DNA from bacteria similar to previous oral metagenomic studies. The major phyla in the human oral saliva are *Actinobacteria*, *Bacteroides*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, *Spirochaetes*, and TM7 (Zaura *et al.*, 2009, The Human Microbiome Project Consortium, 2012), which matched to most of the bacterial orders identified from the 16S rRNA sequencing results

of our oral metagenomic DNA, as shown in Figure 3-8. Therefore, our metagenome is likely to represent the oral microbiome and can be used for the subsequent studies.

Human Oral Microbiome Taxonomic Hierarchy

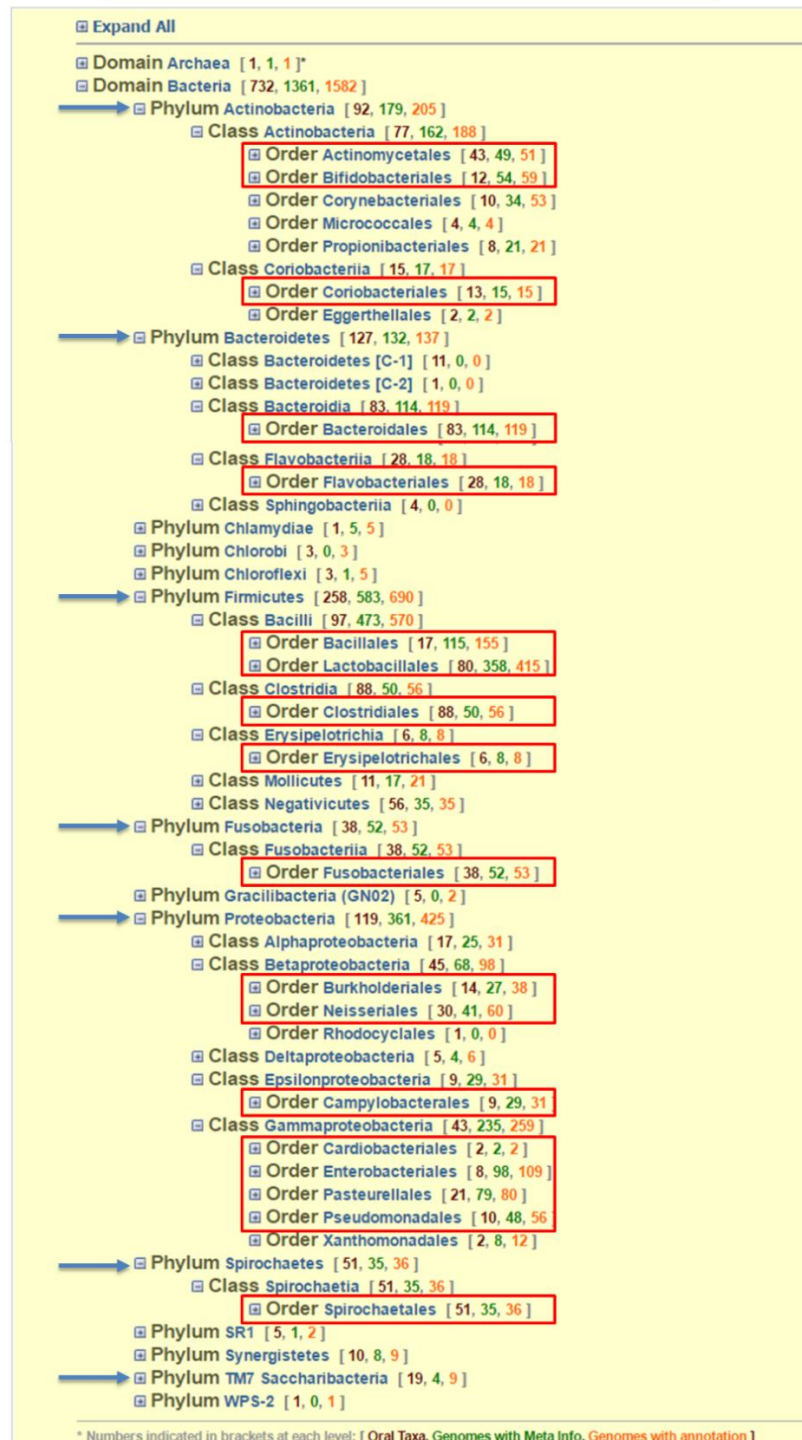


Figure 3-8 Human oral microbiome hierarchy from the Human Oral Microbiome Database (HOMD). The red boxes indicate the bacterial orders identified in the extracted human oral metagenome by the 16S rRNA sequencing. The blue arrows indicate the common phyla of oral bacteria. (Adapted from Chen *et al.*, 2010)

Previously, the PCR-based approach has been used for the recovery of integron GCs from various metagenomes, which are all non-human environments. Most of the metagenomic studies on human microbiota were either sequence-based studies to determine the microbial diversity and all genetic features or functional-based studies to recover genes with the functions of interest. This is the first study that used a PCR-approach to detect integrons in the metagenome extracted from human saliva.

Two different primer sets were tested in this study. The first primer set were developed by other research groups (Martinez-Freijo *et al.*, 1998, Nield *et al.*, 2001, Stokes *et al.*, 2001). They could not amplify GCs from oral metagenomic DNA. This could be due to the variation of the *attC* sequences between environmental samples. For example, the HS286/HS287 degenerate primers were developed based on the *attC* sequences from the class I integrons of soil bacteria (Stokes *et al.*, 2001). The *attC* site is more likely to be conserved through the secondary structure, which can be recognised by the integron integrase. When comparing the complementarity between the HS286/HS287 primers and *T. denticola attC*, there were several mismatch nucleotides, as shown in Figure 3-9, which could explain the random binding of HS286/HS287 primers in oral metagenomic DNA.

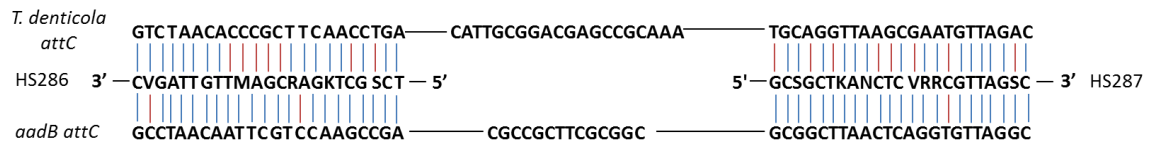


Figure 3-9 The similarity of HS286/HS287 primers to *T. denticola attC* site. The blue lines indicate complementary nucleotides, and the red lines indicate the non-complementary nucleotides. *aadB attC* is an example of *attC* which is complementary with HS286/HS287 primers.

Another set of primers was the novel *attC* primers based on *T. denticola attC* sequences (Coleman *et al.*, 2004), which successfully amplified a diverse array of GCs in oral metagenomic DNA. Most of the GCs were matched with *Treponema* spp. In the previous studies, the integron GCs from *T. denticola* and *T. vincentii* were identified from the whole genome sequencing and *in silico* analysis, respectively (Coleman *et al.*, 2004, Wu *et al.*, 2012). The integron GCs from both species were found in our study. In addition to both species, we also identified the putative GCs related to the other *Treponema* spp., which were *T. pedis*, *T. putidum*, and *T. medium*. Moreover, this PCR approach can also recover novel GCs from other genera, which encoded proteins from *Rhodonellum psychrophilum*, *Paenibacillus assamensis*, and *Bradyrhizobium* sp. However, as the percentage of similarity of these GCs were lower than 70%, there is also a possibility that they were *Treponema*-related GCs which have not been deposited in the database yet.

The functions of the proteins encoded by the detected GCs are mostly related to the adaptation of bacteria to environmental stress. GC SSU3 was predicted to encode a protein similar to the glyoxalase from *Treponema pedis* with 97% amino acid similarity (WP_009105863.1, 100% coverage). It is a

metalloenzyme that can be found in both prokaryotic and eukaryotic organisms, which is important for the detoxification of methylglyoxal. This compound is formed from dihydroxyacetone phosphate (DHAP), which is a by-product of several metabolic pathways, such as glycolysis and phospholipid biosynthesis (Ferguson *et al.*, 1998, Vander Jagt & Hunsaker, 2003). As methylglyoxal is an electrophile, it can react with the nucleophilic centres of macromolecules such as proteins, DNA, RNA, which could lead to protein damage, nucleotide damage and oxidative stress (Lee & Park, 2017). Therefore, methylglyoxal detoxification enzymes, including glyoxalase, are required to prevent the accumulation of methylglyoxal within bacteria.

Another detoxification enzyme was predicted to be encoded by the ORF in GC SSU5. It was a cof-like hydrolase enzyme, which is a member of haloacid dehalogenase superfamily. Cof-like hydrolases are essential for the detoxification of halogenated aliphatic hydrocarbons, such as chlorinated compounds, by hydrolysing the carbon-halogen bonds (Hardman, 1991, Koonin & Tatusov, 1994). For the oral cavity, the halogenated aliphatic hydrocarbons can be introduced and found as part of our diets such as food preservatives (brominated vegetable oils), antibiotics (chloramphenicol), and pesticides (atrazine). Therefore, these detoxification enzymes are required for the oral bacteria to survive and cope with these compounds. By having these genes as a GC within integrons, it allows bacteria to express these enzymes only when they are needed upon the exposure to the stress conditions by the shuffling of the GCs to the first position, next to the P_C promoter.

GC SSU5 contained a gene encoding a twitching motility PilT protein, which is an ATPase, located in the inner membrane motor subcomplex of type IV pili (T4P). It is important for the final stage of bacterial twitching motility, which is a translocation of bacteria over moist surfaces. PilT is required for a pilus retraction by promoting the depolymerisation of pilin subunit from pili (Burrows, 2012). PilT is also involved in bacterial virulence as it is required for the dispersal of bacteria on epithelial cells (Merz & So, 2000). It was previously shown that, in the absence of PilT, bacteria cannot progress from the localised adherence stage on epithelial cells (mediated by pili) into a diffuse adherence stage (nonpilated cells), resulting in bacterial clumps and less virulence (Bieber *et al.*, 1998, Pujol *et al.*, 1999). PilT was also shown to involve bacterial transformation, as the retraction of pili is required for the uptake of DNA (Wolfgang *et al.*, 1998, Graupner *et al.*, 2001). Furthermore, it is necessary for the formation of normal biofilm morphology in *Pseudomonas aeruginosa* (Chiang & Burrows, 2003). Therefore, a *pilT*-containing GC could be needed for bacterial adaptation and the development of biofilm in the oral cavity.

Toxin-antitoxin (TA) containing GCs have previously been detected in other integrons, especially in chromosomal integrons. They are hypothesised to have a function in maintaining and stabilising the chromosomal integrons (Szekeres *et al.*, 2007). In our study, SSU21 and SSU27 samples contained HicA-HicB and BrnT-BrnA TA systems, respectively. The HicA-HicB cassette was previously reported in the fourth position on the integron of *T. denticola* (TDE1838 and TDE1837), which exhibited 96% nucleotide similarity to the SSU27 GC (Coleman *et al.*, 2004). Both TA cassettes are the members of type II TA systems in which the toxin genes encode a small harmful protein

that can act through various mechanisms such as inhibition of DNA replication, degradation of mRNA, inhibition of ribosome and inhibition of cell wall synthesis (Schumacher *et al.*, 2009, Mutschler *et al.*, 2011, Heaton *et al.*, 2012, Aakre *et al.*, 2013). The antitoxins of type II TA systems can inhibit the activity of toxins by protein-protein complex formation (Makarova *et al.*, 2006, Leplae *et al.*, 2011).

Currently, there are three well-established TA systems, type I, II and III, which are classified based on the neutralisation mechanisms of the antitoxins (Page & Peti, 2016). The antitoxins of type I TA systems transcribe into a complementary RNA, which can bind to type I toxin's mRNA and inhibit the translation of toxin, while the type I toxin is a short hydrophobic protein which can insert into and disrupt cell membranes (Brantl, 2012). The antitoxins of type III TA systems are also a small RNA, similar to type I, but the inhibition occurs by the direct binding of RNA antitoxin to a type III toxins (Goeders *et al.*, 2016).

The criteria used to screen the clones were stringent and confirmed that all of 22 clones were amplified from GCs. Even though none of them was successfully verified by the detection of the GC circular form, it did not mean that those clones were not GCs. When GCs excise from integrons, they will form a single-stranded circular molecule that could be degraded by their hosts. Even if the single-stranded circular GCs are converted into double-stranded DNA by DNA synthesis, these molecules tend to be lost after cell division, as they do not contain an origin of replication (Escudero *et al.*, 2015). Another factor is that the excision reaction is catalysed by integron integrases, which

are usually up-regulated by the SOS response (Guerin *et al.*, 2009). The GCs identified in this study, therefore, might not be excised into a circular form during the collection of the saliva samples and cannot be detected in the circular form by PCR.

Even though the PCR-based analysis on the oral metagenome revealed several novel GCs, it can identify only GCs containing known ARGs. Furthermore, PCR amplification can also introduce bias toward specific products. Therefore, the function-based analysis was performed on the SUPA3-SUPA4 PCR products to detect for ARG-containing GCs. However, no positive clone was found. This could be caused by the disruption of genes by restriction enzymes or the constraints from the heterologous protein expression (different in codon usage, tRNA availability and ribosomal binding sites), so the clones with the ARG-containing GCs cannot exhibit resistance phenotypes in the screening. Another possibility is that genes conferring resistance to these drugs might not be located in oral integron, as the previous studies on *Treponema* integrons also did not report ARG-containing GCs. However, the functional screening for ARG-containing GCs still should be done because those previous studies were based on the known sequences in the database. Therefore, there is a chance to recover a novel resistance gene from the oral integrons.

3.5 Conclusion

This is the first study confirming the presence of integron GCs in oral metagenomic DNA using a PCR-based approach, generating new information

regarding the diversity of GCs in the oral cavity. Most of the detected GCs are novel, and the predicted proteins are likely to carry out a multitude of functions contributing to bacterial adaptation against environmental stresses. The oral cavity may contain rich and diverse integron GCs due to its variable physicochemical and stressful environment. These results have been published in PLOS one (Appendix 5) (Tansirichaiya *et al.*, 2016b).

Chapter 4

Determination of the Promoter Activity of Integron Gene Cassettes from the Human Oral Metagenomic DNA

4.1 Introduction

As most of the integron gene cassettes (GCs) are promoterless (Gillings, 2014), their expression mainly relies on the P_C promoter, located downstream from the integron integrase genes. The strength of expression decreases as GCs become more distant from the P_C promoter (Coyne *et al.*, 2010). One of the factors causing this reduction is the *attC* sites, which can form a secondary structure in the mRNA and impede the progression of the ribosome on polycistronic RNAs (Jacquier *et al.*, 2009). Therefore, integron expression driven by only the P_C promoter is unlikely to cover all gene cassettes in a large array, especially in chromosomal integrons (CIs), which could carry up to 200 GCs. However, the studies on *Vibrio* chromosomal integrons showed that most cassettes were transcribed. Thus, internal promoters are likely to be present within the arrays for the expression of these GCs (Michael & Labbate, 2010).

One type of GC, found in 12 out of 63 identified GCs (20%) in our previous study on the detection of integron GCs in the human oral metagenome, is the noncoding GC, which contains no ORF (Tansirichaiya *et al.*, 2016b). Noncoding GCs were also reported by other studies, for example, they were found as a part of *Vibrio* cassette arrays the proportion varied from 6% to 49% (Boucher *et al.*, 2006a). These GCs were predicted to be involved in gene regulation such as promoters or encode regulatory RNAs. For example, the trans-acting small RNA (sRNA)-Xcc1, encoded from the noncoding GC of *Xanthomonas campestris* pv. *campestris* integron, was involved in virulence

regulation (Chen *et al.*, 2011). However, the promoter activity of noncoding GCs has not been proven.

In this chapter, the promoter activity of noncoding GCs was determined by cloning them upstream of the *gusA* reporter gene and measuring β -glucuronidase enzyme activity. Furthermore, a promoter detection system was also developed in this study by utilising a PCR and dual reporter genes, called Bi-Directional Promoter Detection (BiDiPD), which the clones with promoter-containing GCs can be visualised directly on agar plates, allowing the isolation of GC PCR amplicons, and by extension any promoter-containing PCR amplicons, from metagenomic DNA (Figure 4-7).

4.2 Materials and methods

4.2.1 *in silico* analysis of the human oral cavity gene cassettes.

All of the noncoding GCs and some of the GCs containing ORFs identified in Chapter 3 and the previous study (Tansirichaiya *et al.*, 2016b) were analysed for putative promoter sequences by using the web-based software BPPROM in the Softberry package (Solovyev & Salamov, 2011).

4.2.2 Construction of pUC19-GC-*gusA* constructs

The pUC19-GC-*gusA* were constructed based on the pUC19-*Ptet(M)*-*gusA* plasmid described previously (Seier-Petersen *et al.*, 2014) and summarised in Figure 4-1. The *Ptet(M)* promoter was deleted from the plasmid by amplifying the pUC19-*gusA*-only fragment with For916PO and Rev916GO primers and

performing site-directed mutagenesis (section 2.4.16), which resulted in the pUC19-*gusA*-only plasmid.

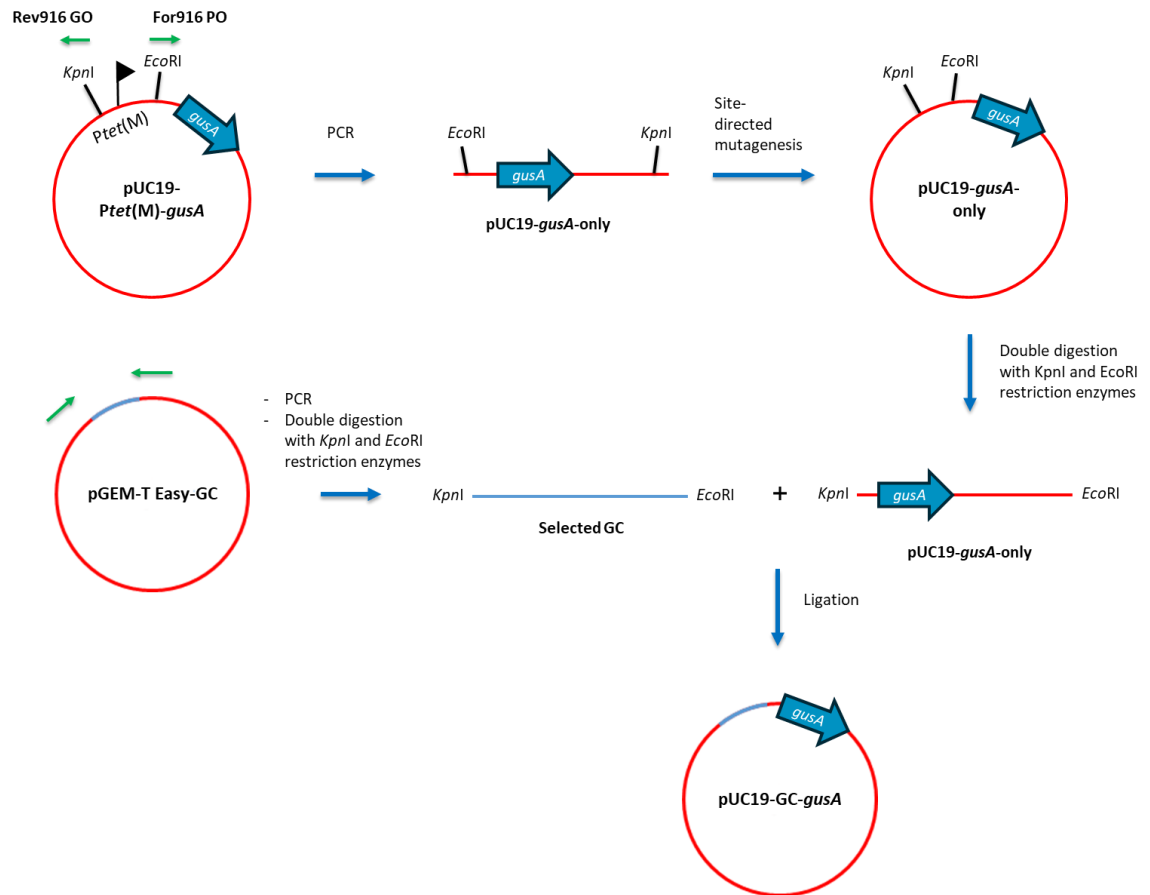


Figure 4-1 Construction of pUC19-GC-*gusA* constructs. The *Ptet(M)* promoter was deleted from pUC19-*Ptet(M)*-*gusA* by performing PCR with Rev916GO and For916GO primers. The pUC19-*gusA*-only amplicon was then recircularised. Each amplified insert, shown as blue lines, was directionally cloned into the *EcoRI* and *KpnI* sites on the pUC19-*gusA*-only plasmids. The green arrows and black lines represent primer binding sites and restriction sites, respectively. The symbol (▶) and blue arrow boxes represent promoter and reporter genes, respectively, pointing in the direction of transcription.

The selected GCs from section 4.2.1 were then amplified from the pGEM-T easy vectors containing the GCs from the previous study (Tansirichaiya *et al.*, 2016b), with the primers listed in Appendix 3 (Figure 4-2). The *KpnI* and *EcoRI*

restriction sites were added to the primers to enable directional cloning. The amplicons and pUC19-*gusA*-only were double digested with *EcoRI* and *KpnI* (section 2.4.10) and ligated together (section 2.4.12), forming pUC19-GC-*gusA* plasmids. The ligation products were then transformed into *E. coli* α -select silver efficiency competent cells by heat shock (section 2.4.15.1).

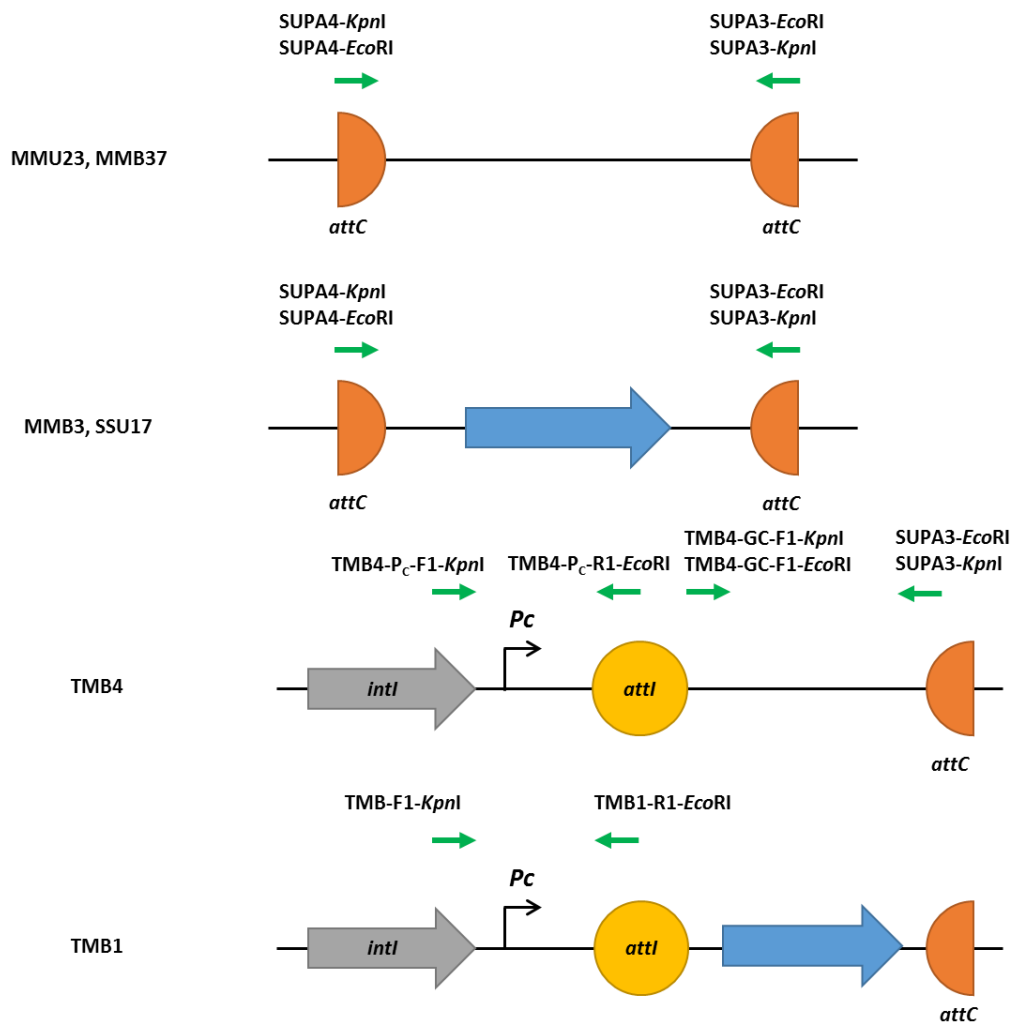


Figure 4-2 The amplification of the selected GCs from GC-containing pGEM-T easy vectors. The green arrows indicate the primer binding sites. The grey and blue arrow boxes represent integrase gene (*intI*) and the ORFs, respectively, pointing in the direction of transcription. The *P_C* promoter is represented by black arrows. The recombination sites, *attI* and *attC*, are represented by yellow circles and orange semi-circles, respectively.

4.2.3 Construction of pCC1BAC-GC-*gusA* constructs

As there was a significant difference in the plasmid copy number in some constructs of the pUC19-GC-*gusA*, new constructs were prepared based on a pCC1BAC vector which can be found as a single copy per cell (Wild & Szybalski, 2004). The construction of pCC1BAC-GC-*gusA* is summarised in Figure 4-3. The GC and *gusA* were amplified from each pUC19-GC-*gusA* construct by using *gusA*-F4-*Hind*III and M13 reverse primers. The amplicons were then digested with *Bam*HI and *Hind*III, and directionally cloned into pCC1BAC vectors, following the protocol in section 2.4.10 and 2.4.12, respectively. The ligation products were introduced into *E. coli* by heat shock transformation (section 2.4.15.1).

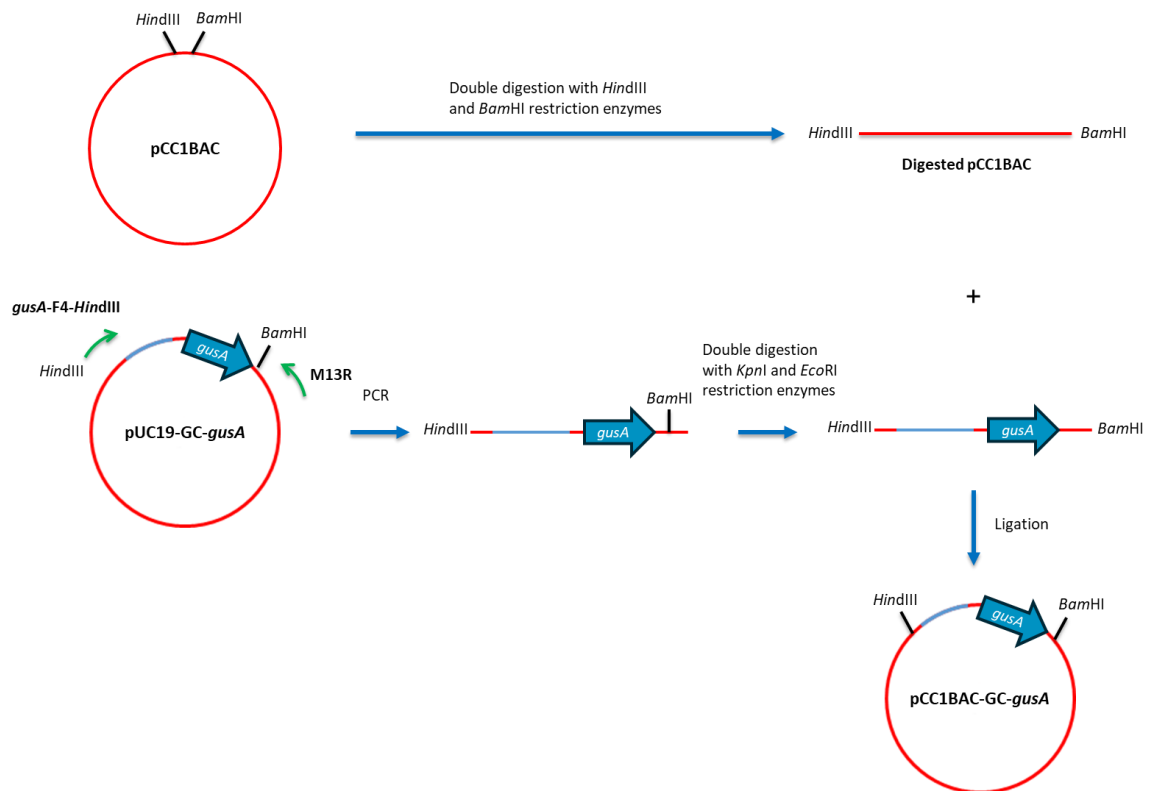


Figure 4-3 Construction of pCC1BAC-GC-gusA constructs. The insert DNAs were amplified from the pUC19-GC-gusA constructs and directionally cloned in between *Hind*III and *Bam*HI restriction sites on pCC1BAC vector. The green arrows and black lines represent primer binding sites and restriction sites, respectively. The blue arrow boxes represent reporter genes, pointing in the direction of transcription.

4.2.4 Development of pCC1BAC-lacZ α -GC-gusA constructs

As the pCC1BAC-GC-gusA constructs were shown to exhibit background promoter activity from the plasmid backbone, new constructs, pCC1BAC-lacZ α -GC-gusA, were therefore designed to contain two reporter genes (*lacZ* and *gusA*), flanked by bi-directional terminators. The uses of two reporter genes to detect gene expressions on both directions have been reported previously such as *uidA-lacZ* reporter system to measure the expression

driven by the P_c and P_{int1} promoters (Guerin *et al.*, 2011), and the dual fluorescence reporter system (a green fluorescent gene from the *Aequorea coerulea* jellyfish and a red fluorescent gene from the *Discosoma striata* reef coral) to study transcriptional output of mammalian bidirectional promoters (Lejard *et al.*, 2014).

4.2.4.1 Optimisation of bi-directional terminators for the pCC1BAC- *lacZ α -GC-gusA* constructs

Bi-directional terminators were added to prevent transcriptional read-through from the promoter in the plasmid backbone and to also prevent promoters from the inserts to interfere with the expression of genes on the plasmid backbone. Bi-directional *rho*-independent terminators are characterised by a GC-rich palindromic region, flanked by a poly(A) region (upstream) and poly(T) region (downstream) (Figure 4-4). When RNA polymerase transcribes the GC palindromic region, it will fold into a stem-loop structure in the nascent transcript, causing a stalling of RNA polymerase (von Hippel & Yager, 1992). A poly(U) region in the nascent RNA will have a very weak base pairing with the template DNA, causing instability of the DNA-RNA hybrid structure (Martin & Tinoco, 1980). Together, it will lead to the release of the mRNA transcript and dissociation of RNA polymerase, resulting in termination of transcription (Martin & Tinoco, 1980, Farnham & Platt, 1981, von Hippel & Yager, 1992).

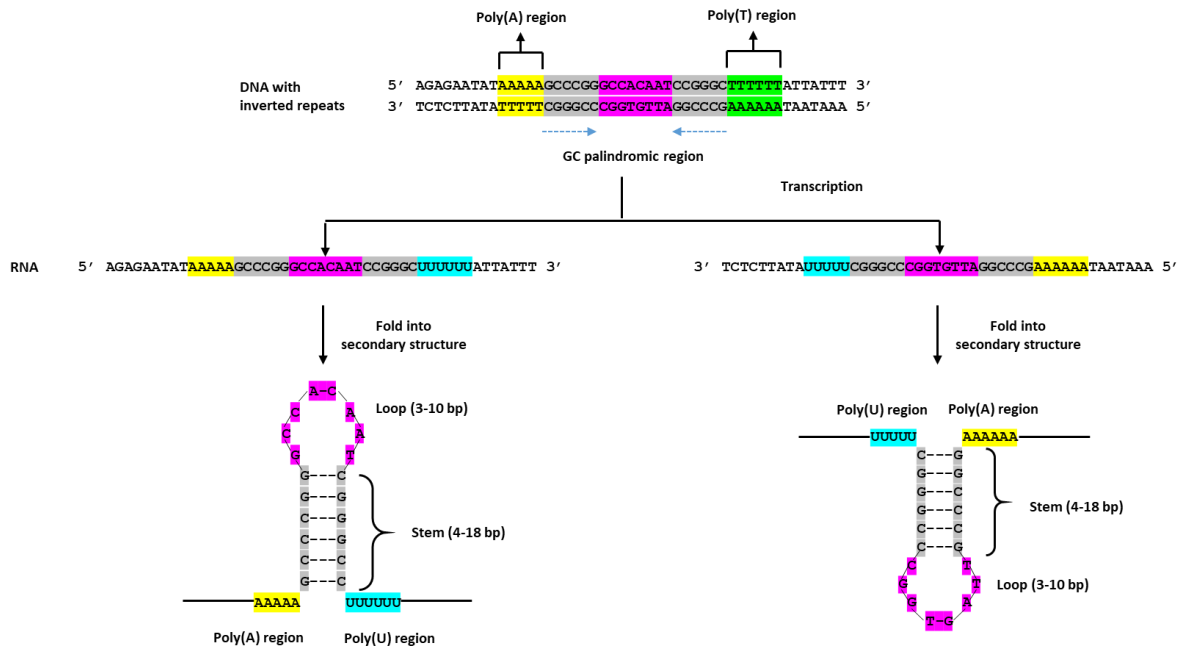


Figure 4-4 Bi-directional *rho*-independent terminator consisting of a stable hairpin structure, flanked by poly(A) and poly(U) regions. The stem of secondary structure is formed by a GC-rich palindromic region (blue arrow). The poly(A), poly(T), poly(U), stem, and loop regions are highlighted in yellow, green, blue, purple and grey, respectively.

luxI bi-directional terminator, found in the *lux* operon of *Vibrio fischeri*, was previously demonstrated to have high efficiency of termination in both directions (Swartzman *et al.*, 1990). To confirm, the termination efficiency was evaluated by using an algorithm described previously by d' Aubenton Carafa *et al.* (1990), based on two parameters; n_T and Y value.

The n_T value is the number of T residues with a weight decreasing in the 5' to 3' direction. To calculate n_T , the x_n value for each nucleotide in the stretch is calculated as followed;

The value for the first T is: $x_1 = 0.9$

$x_n = x_{n-1} \times 0.9$, if the n th nucleotide is a thymine.

$x_n = x_{n-1} \times 0.6$, if the n th nucleotide is other than thymine.

The n_T value is then calculated as the sum of x_n of the T residues only

(Equation 4-1: $n_T = \sum x_n$).

For example, in octamer TTTTATAT, the n_T is calculated as follows:

T: 0.9

T: $0.9 \times 0.9 = 0.81$

T: $0.81 \times 0.9 = 0.729$

T: $0.729 \times 0.9 = 0.656$

A: $0.656 \times 0.6 = 0.394$

T: $0.394 \times 0.9 = 0.355$

A: $0.355 \times 0.6 = 0.213$

T: $0.213 \times 0.9 = 0.192$

Therefore $n_T = 0.9 + 0.81 + 0.729 + 0.656 + 0.355 + 0.192 = 3.642$

The minimum n_T value for a real terminator is 2.895, therefore, any terminator with the n_T value less than 2.895 will be rejected and not considered as a real terminator (d'Aubenton Carafa *et al.*, 1990).

The Y value can be calculated from the following equation:

$$\text{Equation 4-2: } Y = \frac{-\Delta G}{L_H}$$

where ΔG is the Gibbs free energy of an RNA secondary structure and L_H is the number of nucleotides between the 5' end of the stem and the first U in the stretch (LH). The ΔG of hairpin structure was predicted by using Mfold program (Zuker, 2003).

From both parameters, d' Aubenton Carafa *et al.* (1990) plotted a two-dimensional diagram (between Y value on the y-axis and n_T on the x-axis) to differentiate the real terminators from intracistronic or random structures (Figure 4-5). Line D was drawn to obtain the best separation between both types of structures. The *d* score is the distance of between a point (a structure) and the line D (indicated in Figure 4-5). Based on the computational analysis, the following equation was derived;

$$\text{Equation 4-3: } d = (n_T \times 18.16) + (Y \times 96.59) - 116.87$$

where the condition of $d > 0$ is applied to all terminator structures.

Figure 4-5 A two-dimensional diagram showing the separation between real transcriptional terminators from intracistronic or random structures. The real terminators (●) and intracistronic or random structures (○) are plotted, according to their n_T (x-axis) and Y values (y-axis). Line D represents the best separation between both types of structures. The distance between a point (a structure) and the line D was indicated (black arrow). (Retrieved from d'Aubenton Carafa *et al.*, 1990)

The termination efficiency of terminators can then be estimated from a curve, plotted between the d score and the *in vitro* termination efficiency (%) derived from a set of *E. coli rho*-independent terminators (Figure 4-6) (d'Aubenton Carafa *et al.*, 1990).

Figure 4-6 The correlation between the *d* score of some *rho*-independent terminators in *E. coli* and their efficiency *in vitro*. The terminators are indicated based on the preceding gene or operon: *rrnB T1* and bacteriophage T7 *Te* (□); *tonB* (both directions) and *rpIT* (Δ); *ampL* attenuator and *ampL35A* mutant (▲); *trp* attenuator and *trp* mutants (●); *infC*, *pheS* attenuator, *his* attenuator, *trpt* and *trp* mutants, bacteriophage T3 *Te* (○); *thr* attenuator stem mutants (□); *thr* attenuator and mutants (■); *rnpB* and intracistronic signals in *cca* (+). (Retrieved from d'Aubenton Carafa *et al.*, 1990 with permission from Elsevier)

4.2.4.2 Construction of pCC1BAC-*lacZα*-GC-*gusA* constructs

The pCC1BAC-*lacZ*-GC-*gusA* plasmids were constructed as shown in Figure 4-7. As *lacZα* on pCC1BAC vector contained the T7 promoter sequences, it was first deleted by amplifying the pCC1BAC-del*lacZα* fragment with pCC1BAC-del*LacZ*-F1 and pCC1BAC-del*LacZ*-R1 primers. The fragment was then subjected to site-directed mutagenesis, which resulted in the pCC1BAC-del*LacZα* plasmid. The *lacZα* reporter gene was amplified from the

pUC19 vector with *LacZ*-F1 and *LacZ*-R1 primers. For *gusA* reporter gene, it was amplified from pUC19-*P_{tet}(M)*-*gusA* plasmid with *gusA*-F1 and *gusA*-R1 primers. *Nsi*I restriction site was added to *LacZ*-R1 and *gusA*-F1 primers, the *Nsi*I digested *LacZ* α and *gusA* amplicons could be ligated together (Figure 4-7).

As *Avr*II and *Aat*II restriction sites were added to the *LacZ*-F1 and *gusA*-R1 primers, respectively, the *lacZ* α -*gusA* ligation product could be directionally cloned into the pCC1BAC-dell*lacZ* α plasmid. The *lacZ* α -*gusA* ligated product and the pCC1BAC-dell*lacZ* α were digested with *Aat*II and *Avr*II, and then ligated together, resulting in pCC1BAC-*lacZ* α -*gusA* plasmid. The ligated products were transformed into *E. coli* (section 2.4.15.1). The modified *luxI* bi-directional terminators (section 4.2.4.1) was added to *LacZ*-F1 and *gusA*-R1 primers, resulting in two bi-directional terminators flanking the *lacZ* α -*gusA* reporter genes.

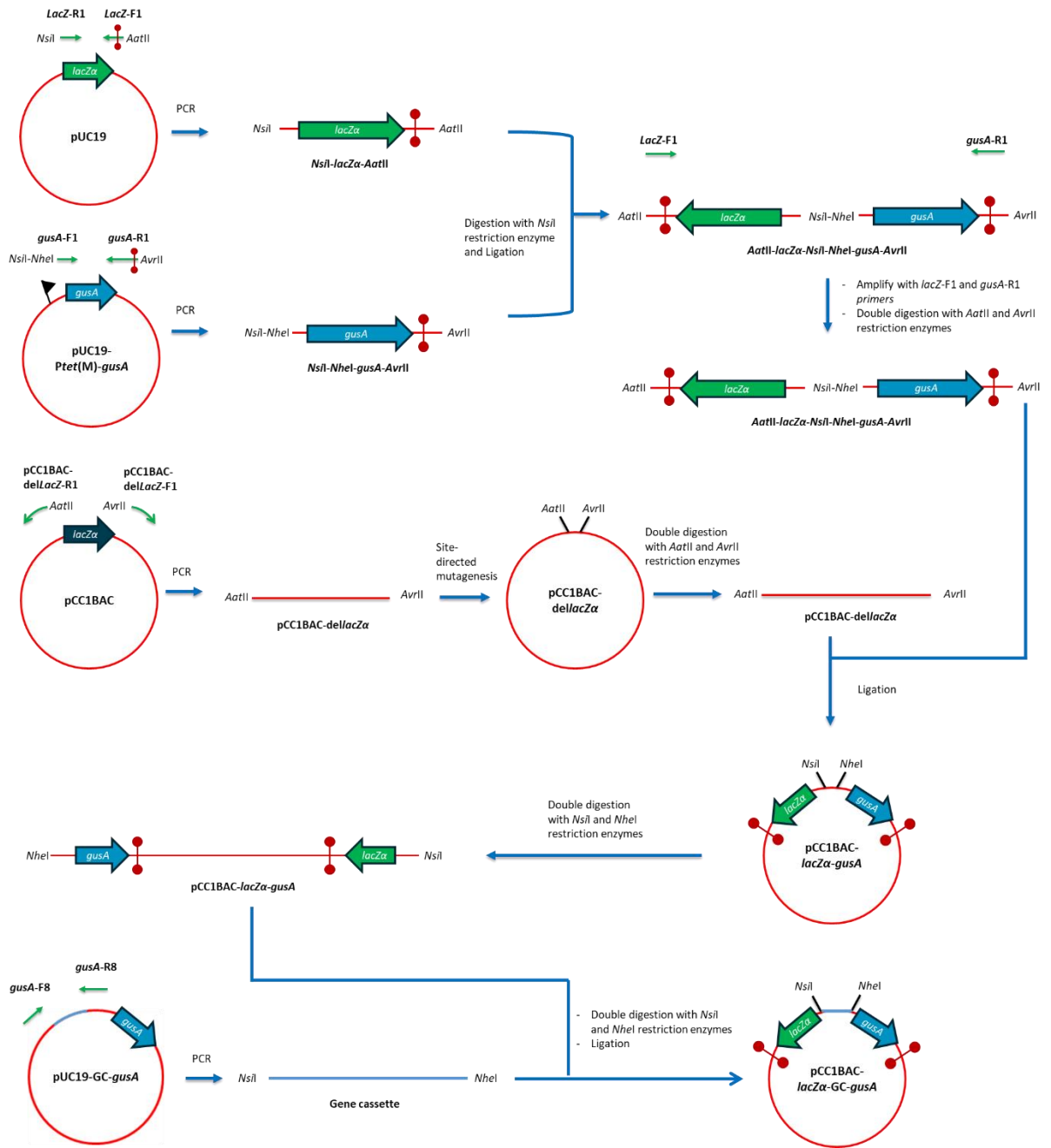


Figure 4-7 Construction of pCC1BAC-lacZα-GC-gusA constructs. The *lacZα* and *gusA* reporter genes were amplified from pUC19 and pUC19-Ptet(M), respectively. Both reporters were ligated together and inserted in between *AatII* and *AvrII* sites of the pCC1BAC-dellacZα plasmid, forming pCC1BAC-lacZα-gusA. The inserts were amplified and cloned into the *NsiI* and *NheI* sites on pCC1BAC-lacZα-gusA plasmid. The green arrows and black lines represent primer binding sites and restriction sites, respectively. The symbol (\blacktriangleright) and (\bullet) represent promoter and bi-directional terminators, respectively. The reporter genes were represented as open arrow boxes, pointing in the direction of transcription.

The selected GCs were amplified from each pUC19-GC-*gusA* construct in section 4.2.2 by using pUC-GC-F1 and pUC-GC-R1 primers (containing *Nsi*I and *Nhe*I restriction sites, respectively). The amplicons were double digested with *Nsi*I and *Nhe*I and directionally cloned into a pre-digested pCC1BAC-*lacZα-gusA* plasmid, then transformed into *E. coli* (section 2.4.15.1).

4.2.4.3 The use of *Nsi*I linker to disrupt promoter formed during the construction of pCC1BAC-*lacZα-gusA* plasmid

During the construction of pCC1BAC-*lacZα-gusA* plasmid in section 4.2.4.2, it was shown that *lacZα* reporter gene was expressed, suggested by the blue colony of *E. coli*:pCC1BAC-*lacZα-gusA* on LB X-gal/IPTG agar plate. An *in silico* analysis by BPROM suggested that a putative promoter was located at *Nsi*I-*Nhe*I restriction sites. A *Nsi*I linker was therefore prepared for the disruption of this putative promoter by increasing the space between the putative -35 and -10 boxes. Two single-stranded oligonucleotides, *Nsi*I-F and *Nsi*I-R (Appendix 4), were designed to have complementary sequences for the annealing and also additional bases of *Nsi*I overhangs for the ligation with *Nsi*I-digested pCC1BAC-*lacZα-gusA* (Figure 4-8A). The complementary region on *Nsi*I linker was designed to contain G and C nucleotides, as AT-rich DNA usually recognises as promoter by RNA polymerase such as -10 hexamers (5'-TATAAT-3') (Zhang *et al.*, 2012).

The *Nsi*I-F and *Nsi*I-R oligonucleotides were phosphorylated by preparing 20 µl phosphorylation reactions: 1 µl *Nsi*I-F/*Nsi*I-R (100 pmol), 1 µl T4 polynucleotide kinase (10 U) (Thermo Scientific, Surrey, UK), 2 µl 10X Buffer

A, 2 μ l ATP (10mM) and 14 μ l molecular grade water. The reactions were incubated at 37°C for 20 min, followed by an inactivation at 75°C for 10 min. Both phosphorylated oligonucleotides were annealed by preparing the 50 μ l mixture, which composed of 10 μ l of each phosphorylated oligonucleotide (50 pmol each), 5 μ l 10X NEBuffer 2.1 (NEB, Hitchin, UK) and 25 μ l molecular grade water. The mixture was incubated at 95°C for 5 min in a heat block, which was then removed from the heat source, allowing the mixture to cool down to room temperature gradually. The pCC1BAC-*lacZ α -gusA* plasmid was digested with *Nsi*I and dephosphorylated, as described in section 2.4.10 and 2.4.11, respectively. The *Nsi*I linker was then ligated to the *Nsi*I-digested pCC1BAC-*lacZ α -gusA* (Figure 4-8B) and introduced into *E. coli* by heat-shock transformation.

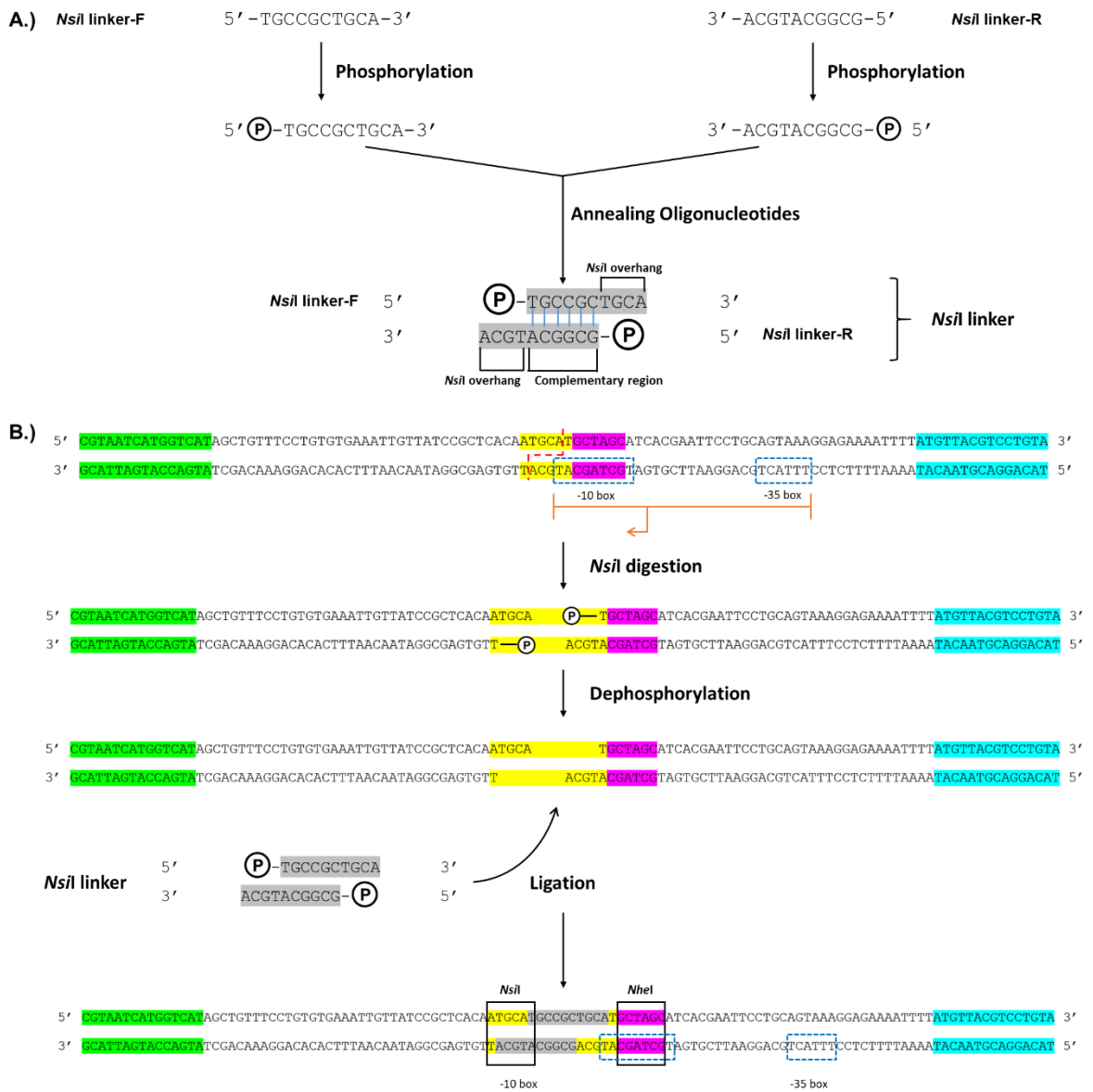


Figure 4-8 The disruption of a putative promoter located in *NsiI-NheI* restriction sites on pCC1BAC-*lacZα-gusA* plasmid. A.) The *NsiI* linker was formed by the annealing of the phosphorylated *NsiI*-F and *NsiI*-R oligonucleotides. B.) The *NsiI* linker was then ligated with the *NsiI*-predigested pCC1BAC-*lacZα-gusA* plasmid. The sequences of *lacZα*, *gusA*, *NsiI* linker are highlighted in green, blue and grey, respectively. The *NsiI* and *NheI* restriction sites are highlighted in yellow and purple, respectively. The predicted promoter on pCC1BAC-*lacZα-gusA* plasmid are indicated with orange arrow and blue boxes.

4.2.5 Determination of β -glucuronidase enzymatic activity.

The β -glucuronidase enzymatic assay was performed to measure the promoter activity based on the expression of *gusA* gene, following the protocol described previously with some modifications (Dupuy & Sonenshein, 1998). The overnight cultures of each *E. coli* strain containing the reporter constructs were prepared in LB broth supplemented with appropriate antibiotic. The optical density of each overnight culture was measured at 600 nm. One millilitre of the overnight culture was centrifuged for 10 min at 3000 x *g*, and discarded the supernatant. The cell pellet was incubated at -70°C for 1 hr, then resuspended in 800 μ l of Z buffer (pH 7, Appendix 2) and 8 μ l of toluene. The mixture was transferred to a 2 ml cryotube, which was filled with glass beads (150–212 μ m in diameter) (Sigma, Dorset, UK). The tubes were vortexed twice for 5 min each with an incubation on ice for 1 min in between, then centrifuged at 3000 x *g* for 3 min at 4°C. An appropriate volume of cell lysate was mixed with Z buffer to make a total volume of 800 μ l, then incubated at 37°C for 5 min. One-hundred sixty microliters of 6 mM *p*-nitrophenyl- β -D-glucuronide (PNPG) was then added to the reaction and incubated at 37°C for 5 min. The reaction was stopped by adding 400 μ l of 1 M Na₂CO₃ and centrifuged at 3000 x *g* for 10 min to remove the cell debris and glass beads. The absorbance of the supernatant was measured with a spectrophotometer at the wavelength of 405 nm. Three biological replicates of β -glucuronidase enzymatic assay were performed. The β -glucuronidase Miller units were calculated by using Equation 4-4.

Equation 4-4: β -glucuronidase Miller units = $\frac{A_{405} \times 1000}{OD_{600} \times \text{time (min)} \times 1.25 \times \text{volume (mL)}}$
(Miller, 1972).

$P_{tet}(M)$ was included as a positive control in the experiment, as its promoter activity had been shown previously with β -glucuronidase reporter assay (Seier-Petersen *et al.*, 2014). As the selected GCs were likely to derive from *Treponema* spp., two other experimentally verified *T. denticola* promoters (P_{Fla} and P_{Tdtro}) were also included to verify that promoters from *Treponema* spp. could be recognised by the *E. coli* host in our study (Limberger *et al.*, 1999, Brett *et al.*, 2008). P_{Fla} and P_{Tdtro} were selected as they rely on different sigma factors to determine the limitations of our enzyme assay in recognising promoters associated with different types of sigma factors. P_{Tdtro} is recognised by sigma factor 70 (σ^{70}) that is responsible for the transcription of most genes during cell growth in both *E. coli* and *Treponema* spp. (Paget & Helmann, 2003, Brett *et al.*, 2008), while the P_{Fla} is associated with sigma factor 28, involving in the expression of flagella-related genes in motile bacteria (Limberger *et al.*, 1999, Koo *et al.*, 2009).

4.2.6 Determination of β -galactosidase enzymatic activity.

The β -galactosidase enzymatic assay was performed to measure the promoter activity based on the expression of a *lacZ α* gene on the pCC1BAC-*lacZ α* -GC-*gusA* constructs. The experiment was performed with the same protocol as in β -glucuronidase enzyme assay (section 4.2.5) with some modifications. The substrate for the assay was 200 μ l of 4 mg/ml o-nitrophenyl- β -D-galactopyranoside (ONPG).

4.2.7 Statistical analysis

The average and standard deviation of β -glucuronidase concentration were calculated from three biological replicates. Statistical analysis was performed by using GraphPad Prism 6 (GraphPad Software Inc., CA, USA). The statistical comparisons between the negative control group and the other constructs were performed by using ordinary one-way ANOVA with either Dunnett's post-hoc test (to compare each construct with a negative control) or Bonferroni's post-hoc test (to compare constructs between themselves). The groups with statistically significant difference from the control had the p -value of less than 0.05.

4.2.8 Development of detection system for integron GCs containing promoter sequences.

4.2.8.1 Selection of optimal enzyme substrates for the detection

The pCC1BAC-*lacZ* α -*gusA* plasmid had the potential to be used for an agar plate based detection assay of integron GCs with promoter sequences on either strand of DNA. This system should allow a direct differentiation between 4 different types of GCs: no promoter activity, with promoter activity on sense strand (*gusA*), promoter activity on antisense strand (*lacZ*) and promoter activity on both strands (both *lacZ* and *gusA*). Therefore, the optimal enzyme substrates for each reporter genes for the differentiation had to be selected.

X-Gal was selected as a substrate for *lacZ* reporter genes (blue-white screening). For *gusA* reporter gene, three β -glucuronidase substrates were

chosen to be tested, which were i.) 300 µg/ml p-nitrophenyl-β-D-glucuronide (PNPG) (detected by a yellow by-product on an agar plate), ii.) 70 µg/ml 4-methylumbelliferyl-β-D-glucuronide (MUG) (detected by visualisation under UV light) and iii.) 60 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc) (detected a blue by-product on an agar plate). The phenotype of an *E. coli* containing *lacZ* and *gusA* grown on the plates containing each substrate was then observed to select for the optimal substrates for the detection.

4.2.8.2 Recovery of promoter-containing GCs from the human oral metagenome

The integron GCs were amplified from the human oral metagenome, extracted in section 3.3.1, by using SUPA4-*NsiI*/SUPA4-*NheI* and MARS5-*NsiI*/MARS2-*NheI* primers, developed in the previous study (Appendix 4) (Tansirichaiya *et al.*, 2016b). The amplified products were double digested with *NsiI* and *NheI* and ligated into pre-digested pCC1BAC-*lacZα-gusA*. The ligated products were transformed into *E. coli* α-select silver efficiency competent cells by heat shock (section 2.4.15.1). Cells were spread on LB agar supplement with 12.5 µg/mL chloramphenicol, 80 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), 50µM IPTG, and 70 µg/mL 4-methylumbelliferyl-β-D-glucuronide (MUG). After incubation at 37°C for 18hr, the colonies with β-galactosidase activity from *lacZ* were detected by blue-white screening on the agar plate, and the β-glucuronidase activity from *gusA* was detected by visualisation under UV light. Colonies exhibiting either activity were selected and subcultured on fresh agar plates. The inserts were amplified by colony PCR using *lacZ-F2* and *gusA-F2* primers (Appendix 4) and sequenced.

The criteria for the sequence analysis of integron GC were the same as described in section 3.2.2. Two additional criteria for the verification of GCs detected with pCC1BAC-*lacZ* α -*gusA* were included. Any clones containing incomplete GCs, caused by digestion at internal *NsiI* and *NheI* restriction sites on the GCs, were excluded from the dataset. Also chimeric inserts, which were the ligation products between digested amplicons, were also excluded. The promoter-containing GCs were named as described in section 3.2.3 with the addition of term “Pro”, indicating the presence of a promoter.

4.3 Results

4.3.1 *in silico* analysis of the promoter sequences on the non-coding gene cassettes

From 63 GCs identified from the human oral metagenome in the previous study, 12 GCs were predicted to contain no ORFs (Tansirichaiya *et al.*, 2016b). By using BPROM promoter prediction software, all noncoding GCs were predicted to contain promoter sequences on both sense and antisense strands (shown in Table 4-1). This suggested that these noncoding GCs might be capable of transcribing the other GCs located upstream and downstream. TMB4 (amplified with primers targeting *IntI* and *attC*) was selected because it was the only noncoding GC, located at the first position in integron GC array, identified in the previous study (Tansirichaiya *et al.*, 2016b). MMU23 and MMB37 were chosen as they had the highest overall score predicted by BPROM. SSU17 and MMB3 were also included in the study as controls, to represent GCs with ORFs.

Table 4-1 The putative promoters for non-protein-coding GCs and GCs with an ORF (SSU17 and MMB3) predicted using BPROM.

Clones	Strand	-10 box	-35 box	Score		
				-10 box	-35 box	Linear discriminant function (LDF)* (Overall score)
TMB4	+	AGGTATAAT	ATAAGA	89	-10	9.78
	-	CATTATTTT	TTGACA	41	66	7.60
SSU9	+	AATTATAAT	TAAAAA	74	0	7.04
	-	TAGTATAAT	TTTATT	80	34	7.11
MMU2	+	AATTATAAT	TTAAAA	74	37	8.36
	-	TAGTATAAT	TTTATT	80	34	8.90
MMU11	+	ATGTAAAAT	TTGCTG	75	47	11.34
	+	AACTATACT	AGGAAA	59	-7	5.99
	-	AAATAAAAAT	TTTTCA	56	34	6.96
	-	CTATAAATT	TTTCAA	44	36	3.24
MMU19	+	AGGTATAAT	TAGAAA	89	23	9.07
	+	TTGAAAAAT	TTGCGG	44	32	3.43
	-	TATTATAAT	TTTCCT	79	37	9.10
MMU23	+	AATTATAAT	TAAAAG	74	-6	9.84
	+	TTTTATTAT	TTGATG	72	52	6.05
	-	TATTATAAT	TTTCCT	79	37	8.66
	-	TAGTATAAT	TTTATT	80	34	8.05
MMB2	+	AATTATAAT	TATAAG	74	-2	8.71
	+	TATTATAAT	TTGATG	79	52	7.88
	-	TATTATAAT	TTTCCT	79	37	9.10
	-	TATTATAAT	TTTATT	79	34	8.84
MMB5	+	AATTATAAT	TTAAAA	74	37	8.36
	-	TAGTATAAT	TTTATT	80	34	7.95

Clones	Strand	-10 box	-35 box	Score		
				-10 box	-35 box	Linear discriminant function (LDF)* (Overall score)
MMB20	+	AATTATAAT	TAAAAG	74	-6	9.09
	-	TATTATAAT	TTTCCT	79	37	9.10
MMB32	+	TATTATAAT	TTGATG	79	52	6.28
	+	AGATATAAA	GTGTAA	39	14	4.84
	-	TATTATAAT	TTGATT	79	53	6.61
	-	TTTTATTTT	TTAAAA	52	37	5.11
MMB36	+	AATTATAAT	TTAAAA	74	37	6.94
	+	TATTATAAT	TTGATG	79	52	6.45
	-	TATTATAAT	TTTATT	79	34	7.44
	-	TTTTAAAAT	TTGACT	79	61	6.13
MMB37	+	AATTATAAT	TAAAAG	74	-6	9.11
	+	TTATATAAT	TTGATG	75	52	8.55
	-	TAGTATTAT	TTTATT	66	34	10.48
	-	TATTATAAT	TTTCCT	79	37	9.10
SSU17	+	CTTTATAAT	ATGAAT	82	25	7.80
	+	TGATAAAAT	GTGAAA	75	27	4.62
	-	TGATATAAT	TTTATT	82	34	9.34
	-	TGATTAGAT	TTTATG	21	33	5.10
MMB3	+	CTGTATATT	TTGATA	63	58	6.74
	+	ATTTATGAT	ATGAAA	65	30	5.18
	-	ATGTATTGT	TTGATG	44	52	6.64
	-	GCATATAAT	TTCTCT	65	28	4.75

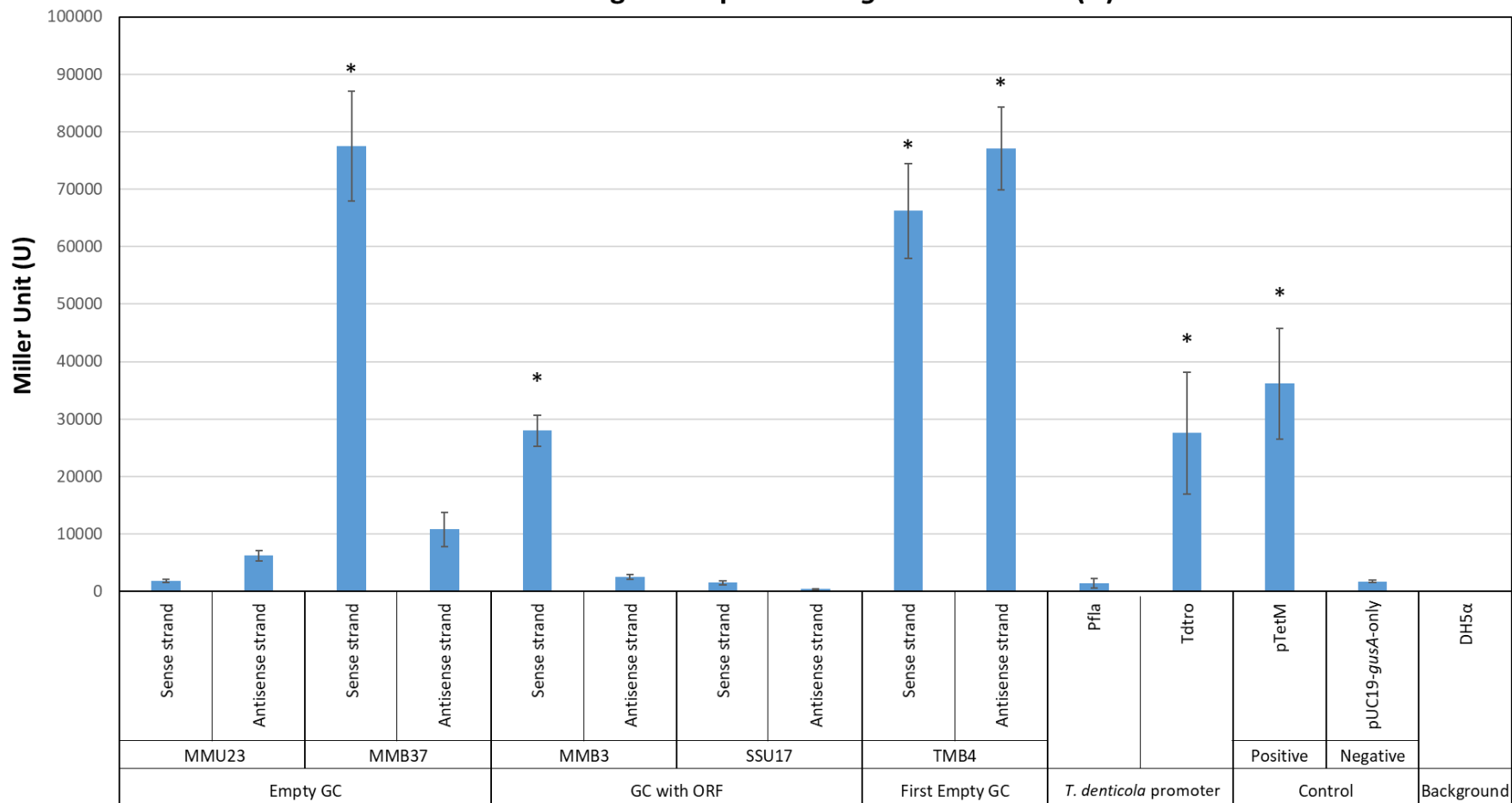
* The LDF takes into account motifs found in promoters: -10 and -35 boxes, a distance between -10 and -35 boxes, and frequencies of certain nucleotides represented in transcription start sites. It can be approximated as $\log(\langle \text{likelihood of a site being promoter} \rangle / \langle \text{likelihood of a site not being promoter} \rangle)$ (Solovyev & Salamov, 2011).

** The selected samples for the enzymatic assay are highlighted in yellow.

4.3.2 Determination of the Promoter activity of the noncoding GCs in pUC19-GC-*gusA* constructs by using β -glucuronidase enzyme assay

The promoter activity of the selected GCs in the pUC19-GC-*gusA* constructs was determined by performing β -glucuronidase enzyme assay and showed in Figure 4-9. The enzymatic assays revealed that MMB3 (ORF-containing GC) had promoter activity on the sense strand, while MMB37 and TMB4 (noncoding GC) had promoter activity on both strands. MMU23 (noncoding GC) showed low activity on the antisense strand, and no activity was found on both strands of SSU17 (ORF-containing GC). The P_{TdTro} from *T. denticola* also showed promoter activity, which verified that promoters from *T. denticola* could be recognised by the *E. coli* host used in our experiments. For the P_{Fla} , no activity was found by the assay, which may occur due to the fact that P_{Fla} was shown to be associated with σ^{28} , which is not a primary sigma factor in *E. coli* (Limberger *et al.*, 1999).

The concentration of β -glucuronidase enzyme in Miller unit from noncoding GCs in pUC19-GC-*gusA* constructs (U)



Constructs

Figure 4-9 The promoter activity of the noncoding GCs in pUC19-GC-*gusA* constructs estimated by β -glucuronidase enzyme assays. Error bars indicate the standard errors of the means from three biological replicates. The asterisks (*) indicate the constructs were statistically significantly different from the negative control group (pUC19-*gusA*-only) with the *p*-value <0.05 by using ordinary one-way ANOVA followed by Dunnett's multiple comparison test.

4.3.3 Determination promoter activity of the TMB4 sample in pUC19-GC-*gusA* constructs by using β -glucuronidase enzyme assay

TMB4 sample was a GC located in the first position of the integron (downstream of the P_C promoter). Therefore, it was interesting to understand the biological function of maintaining a noncoding GC in the first position of a GC array. The total promoter activity was determined by including another two constructs, TMB4- P_C and TMB4 P_C -GC, in the enzyme assay. As the TMB4 P_C promoter was not identical to the P_C of *T. denticola* integron identified previously (Coleman *et al.*, 2004) (Figure 4-10), the P_C of TMB1 sample, which was identical to the P_C of *T. denticola* integron, was also included.

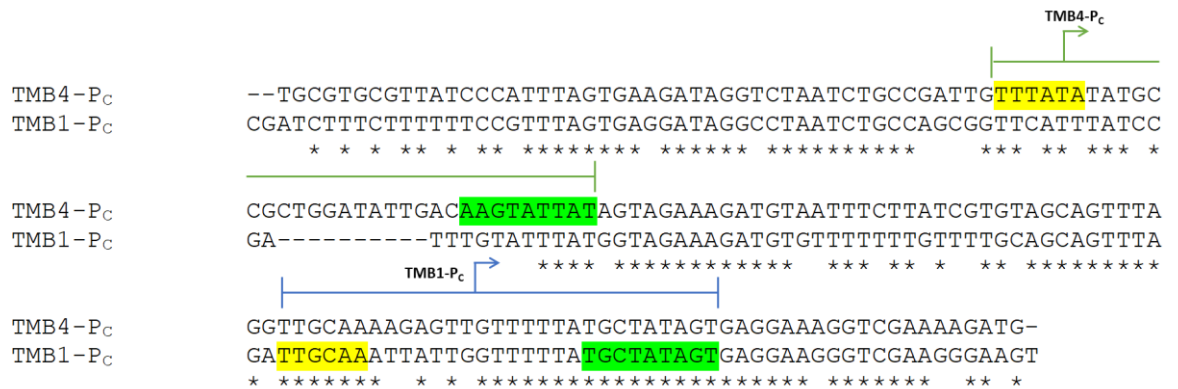


Figure 4-10 Nucleotide sequence alignment between the P_C promoter of TMB1 and TMB4 samples. The asterisks indicate the nucleotides which are identical between the P_C of TMB1 (*T. denticola* integron: Accession number NC_002967) and TMB4 samples. The green and blue arrows indicate TMB1 and TMB4 P_C promoters predicted by BPROM, respectively. The -10 and -35 boxes are highlighted in green and yellow, respectively.

The results from the β -glucuronidase enzyme assay were shown in Figure 4-11. It has been demonstrated that the TMB4 P_C promoter showed low

promoter activity, while the TMB1 P_C promoter had the highest activity among all samples. When TMB4 P_C and the noncoding TMB4 GC located together, the promoter activity was slightly decreased, compared to the activity in the sense strand of TMB4 GC.

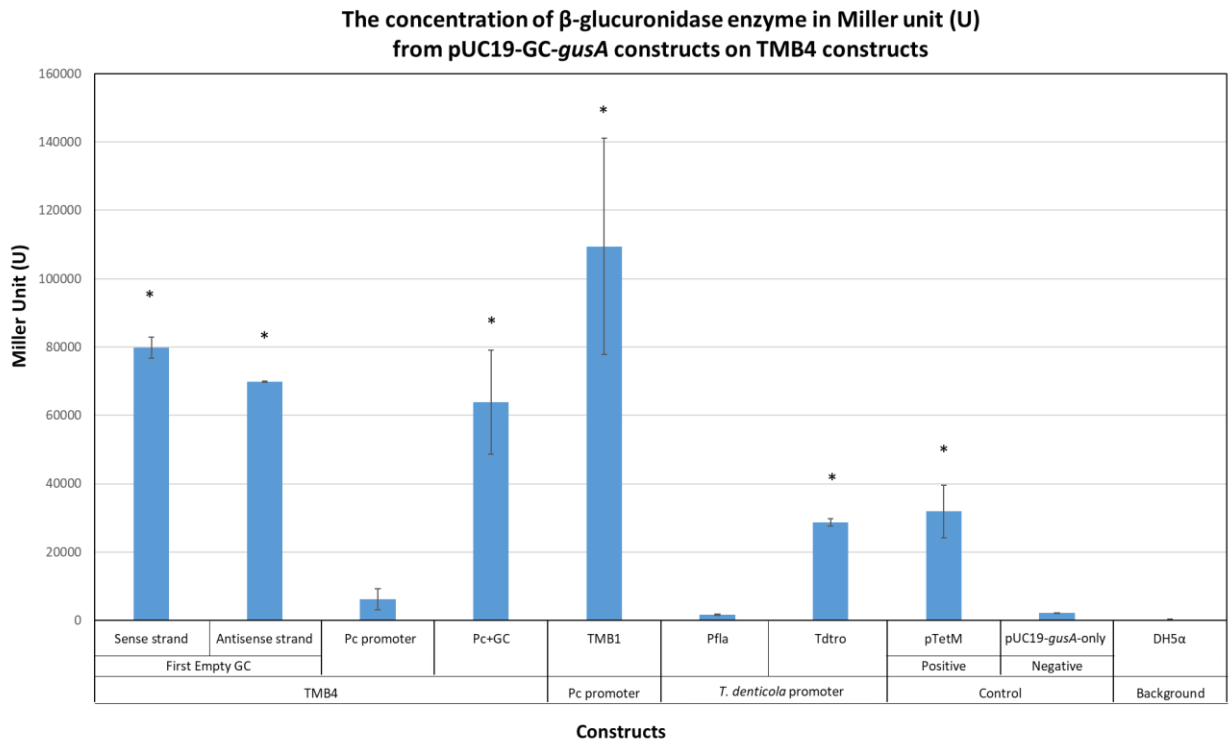


Figure 4-11 The promoter activity of the TMB4 GC in pUC19-GC-*gusA* constructs estimated by β -glucuronidase enzyme assays. Error bars indicate the standard errors of the means from three replicates. The asterisks (*) indicate the constructs were statistically significantly different from the control group (pUC19-*gusA*-only) with the p -value <0.05 by using ordinary one-way ANOVA followed by Dunnett's multiple comparison test.

However, during the plasmid extraction of TMB4-P_C and TMB4 P_C-GC constructs, it was noticed that the plasmid concentrations from the same volume of an overnight culture (1 ml and OD₆₀₀ = 3.0) of both constructs were significantly lower than the other constructs (Figure 4-12). The differences in

the plasmid concentration suggested that there was a difference in the copy number of plasmids between these two constructs and the others.

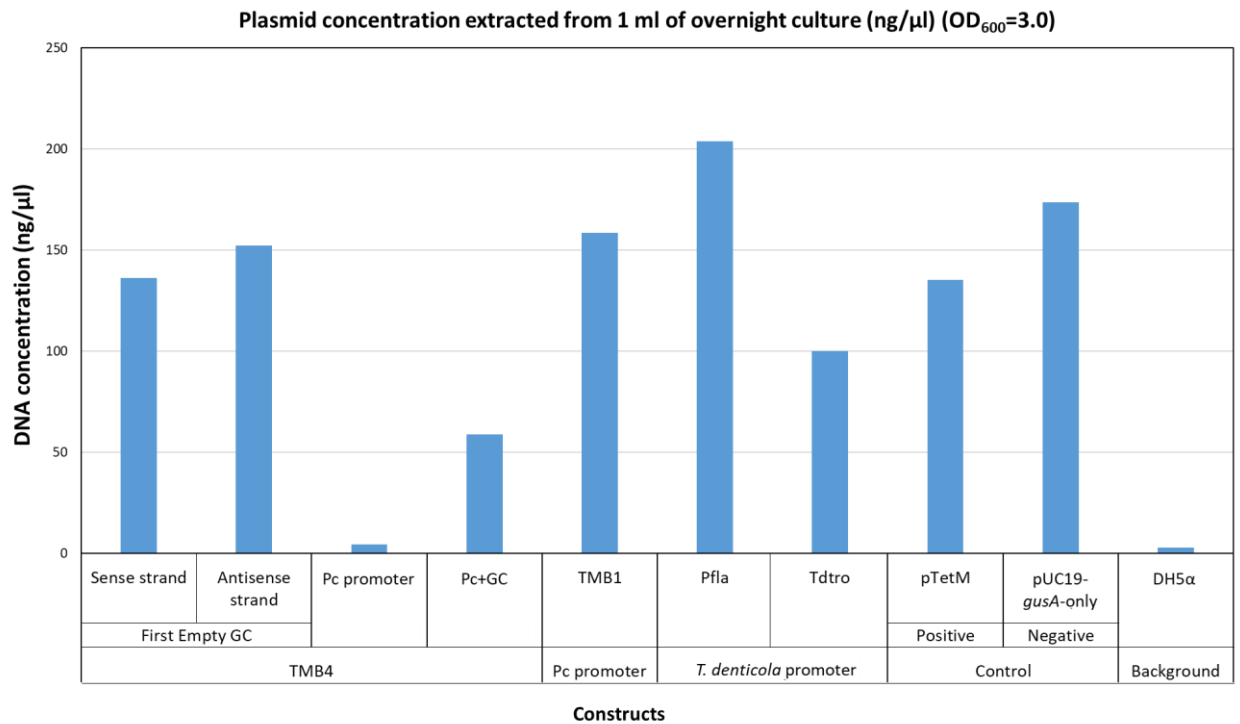


Figure 4-12 Plasmid concentration from 1 ml of *E. coli* overnight culture (OD₆₀₀=3.0) containing pUC19-GC-*gusA* constructs.

The difference in copy number of plasmid per cell could affect the concentration of β -glucuronidase enzyme in the assay, as more β -glucuronidase would be expressed in the strains containing more copies of the plasmid if promoters of equal strength were present. The effect of copy number on the enzyme assay could be estimated by dividing the concentration of β -glucuronidase enzyme with the concentration of plasmid extracted from 1 ml of overnight culture (enzyme Miller unit/ng of plasmid), shown in Figure 4-13. The comparative promoter activity changed when taking the number of plasmids into account, as the TMB4 P_C and TMB4 P_C-GC constructs showed

the highest activity, while they were lower than the activity of TMB4 GC in Figure 4-11. It was, therefore, decided that new constructs should be made to control the copy number of in all strains to be the same.

The normalised concentration of β -glucuronidase enzyme based on the plasmid concentration of pUC19-GC-*gusA* constructs (Miller unit (U)/(ng/ μ l))

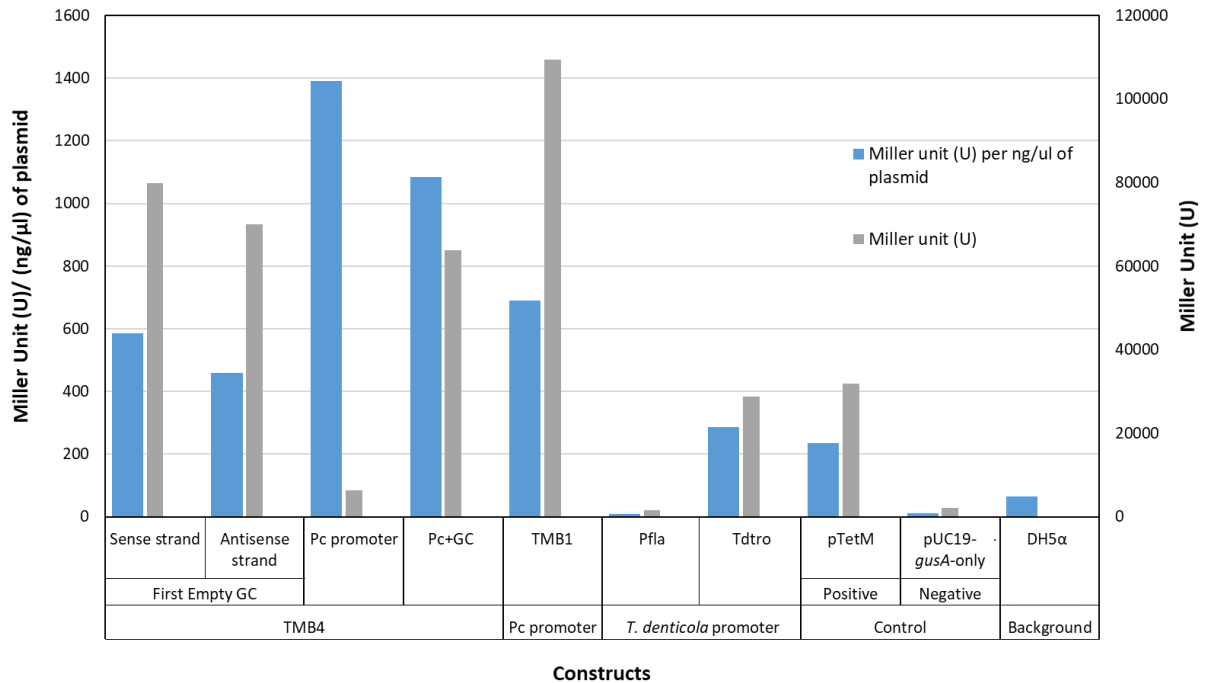


Figure 4-13 The normalised concentration of β -glucuronidase enzyme based on the plasmid concentration of pUC19-GC-*gusA* constructs. The normalised concentration (Miller unit (U)/ (ng/ μ l) of plasmid) is shown in blue, while the concentration without the normalisation (Miller unit (U)) is shown in grey as a reference for the comparison.

4.3.4 Promoter activity determination of pCC1BAC-GC-*gusA* constructs by using β -glucuronidase enzyme assay

A new set of constructs were designed based on pCC1BAC vector, because pCC1BAC will be maintained in *E. coli* cell as one plasmid per cell, and therefore enable us to control the plasmid copy number to be similar between each construct. The relative copy number of each strain was then estimated by performing plasmid miniprep on 10 ml of overnight culture (OD₆₀₀ of 1.7) from every strain (Figure 4-14). The plasmid concentrations were similar between each construct. The TMB4 P_C and TMB4 P_C-GC, which previously showed significantly lower plasmid concentrations in pUC19-GC-*gusA*, were now comparable with the others. It was therefore confirmed that this strategy could solve the copy number issue in the pUC19-GC-*gusA* constructs.

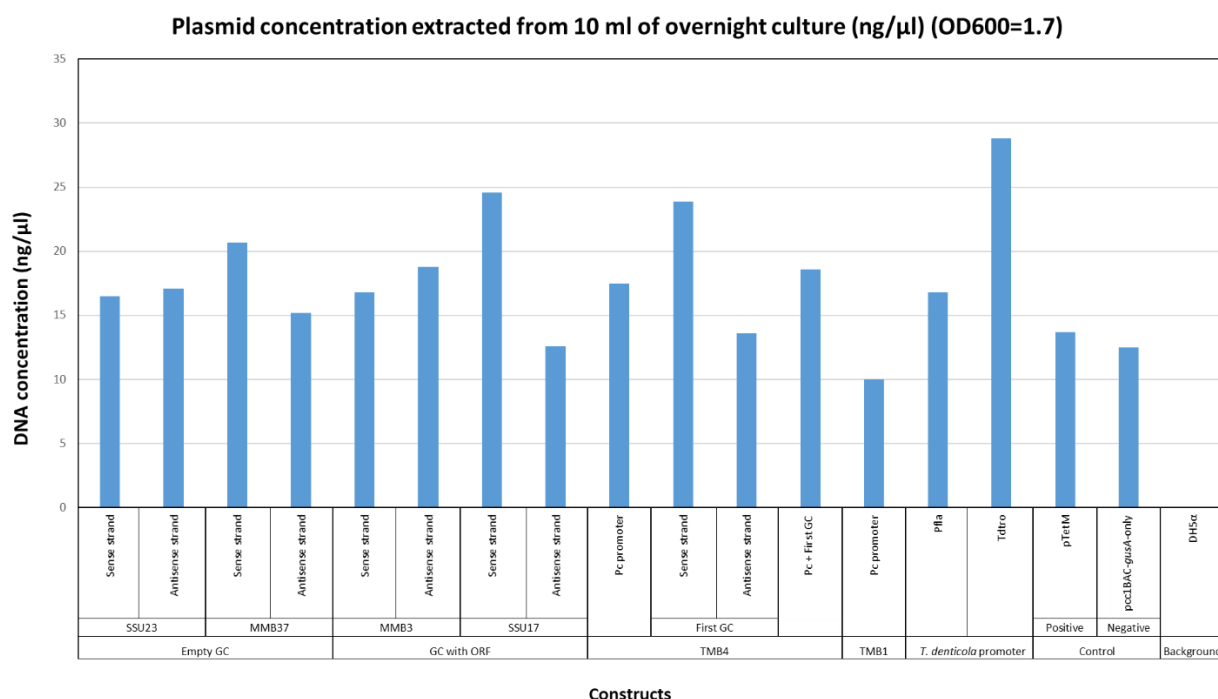


Figure 4-14 Plasmid concentration from 10 ml of *E. coli* overnight culture (OD₆₀₀= 1.7) containing pCC1BAC-GC-*gusA* constructs.

The β-glucuronidase enzyme assay was then performed on *E. coli* containing pCC1BAC-GC-*gusA* constructs, as described in section 4.2.5. The results are shown in Figure 4-15. It was shown that TMB4 P_C and TMB4 P_C-GC exhibited high promoter activity; different from the pUC19-GC-*gusA* constructs (Figure 4-11). However, the amount of enzyme in the negative control (*gusA* only), which was expected to be low, showed promoter activity higher than the positive control (*Ptet*(M)). It was, therefore, suggested that the results from pCC1BAC-GC-*gusA* were not reliable. This could be due to the T7 promoter sequences within *lacZα* of pCC1BAC vector, upstream of the cloning sites.

The concentration of β -glucuronidase enzyme in Miller unit from pCC1BAC-GC-*gusA* constructs (U)

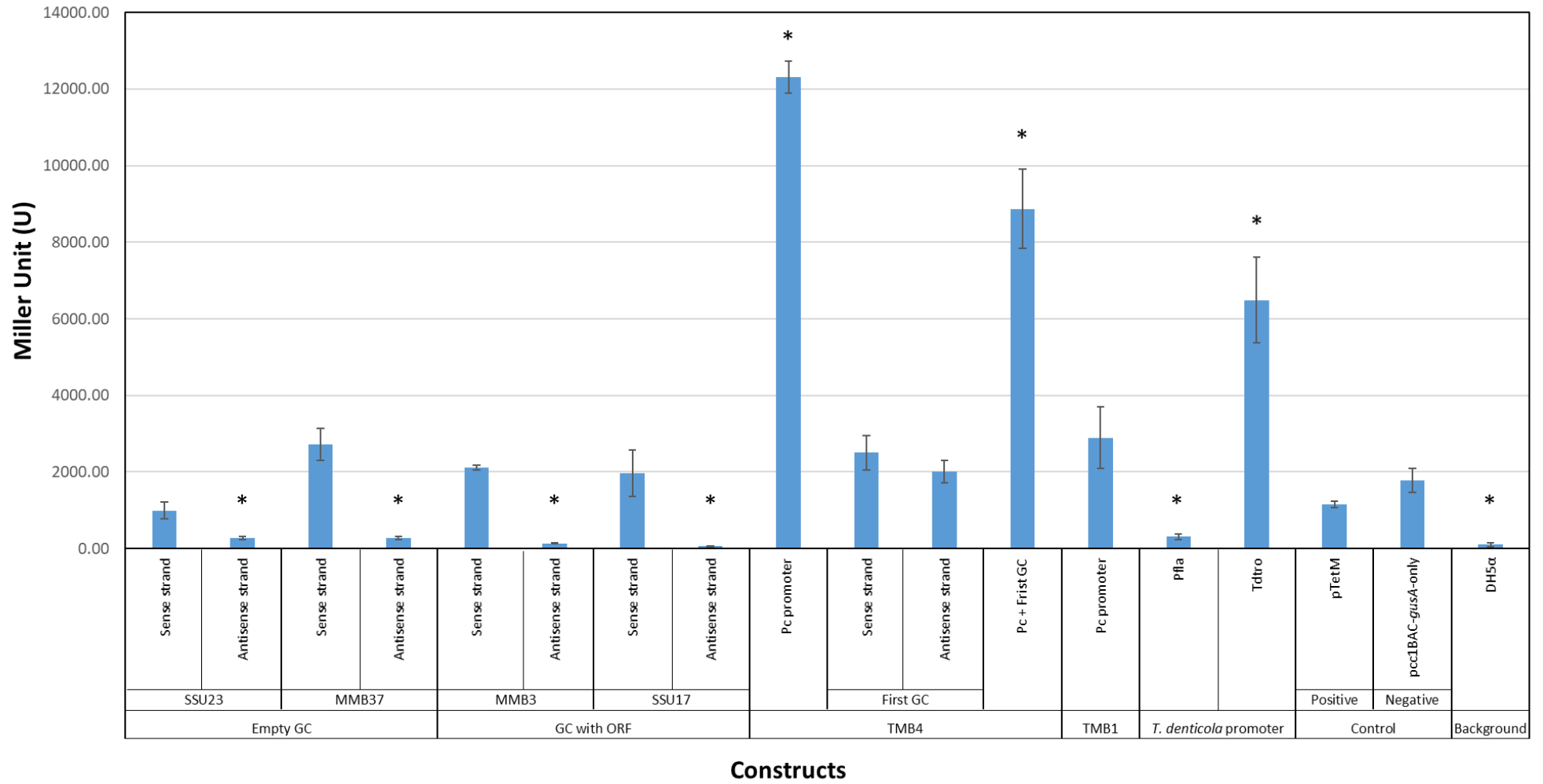


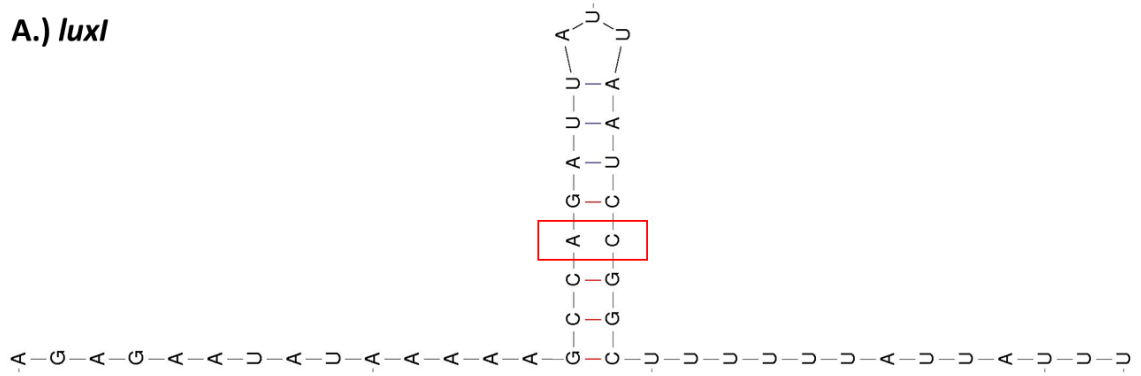
Figure 4-15 The promoter activity of the noncoding GCs in pCC1BAC-GC-*gusA* constructs estimated by β -glucuronidase enzyme assays. Error bars indicate the standard errors of the means from three replicates. The asterisks (*) indicate the constructs were statistically significantly different from the negative control group (pCC1BAC-*gusA*-only) with the p -value <0.05 by using ordinary one-way ANOVA followed by Dunnett's multiple comparison test.

4.3.5 Construction of pCC1BAC-*lacZ* α -*gusA* plasmid

4.3.5.1 The modification of *luxI* bi-directional terminator

The hairpin structure of *luxI* bi-directional terminator was first determined by using Mfold (Figure 4-16A), which showed that one pair of ribonucleotides on the stem structure (A and C ribonucleotides) was not complementary to each other. The termination efficiency on both strands of *luxI* was estimated, as described in section 4.2.4.1, which showed the efficiency on the positive strand and negative strand of 23% and 50%, respectively (Table 4-2 and Figure 4-17). The sequences of *luxI* terminator were, therefore, modified to improve the termination efficiency by changing A to G (Figure 4-16B), resulting in a complementary base pairing and a stronger stem-loop structure. The termination efficiency was then recalculated on the modified *luxI* terminator, which was showed that the termination efficiency was increased from 23% to 75% on a positive strand and from 50% to 85% on a negative strand (Table 4-2 and Figure 4-17). Therefore, the modified *luxI* was used for the construction of pCC1BAC-*lacZ* α -*gusA* plasmid.

A.) *luxI*



B.) Modified *luxI*

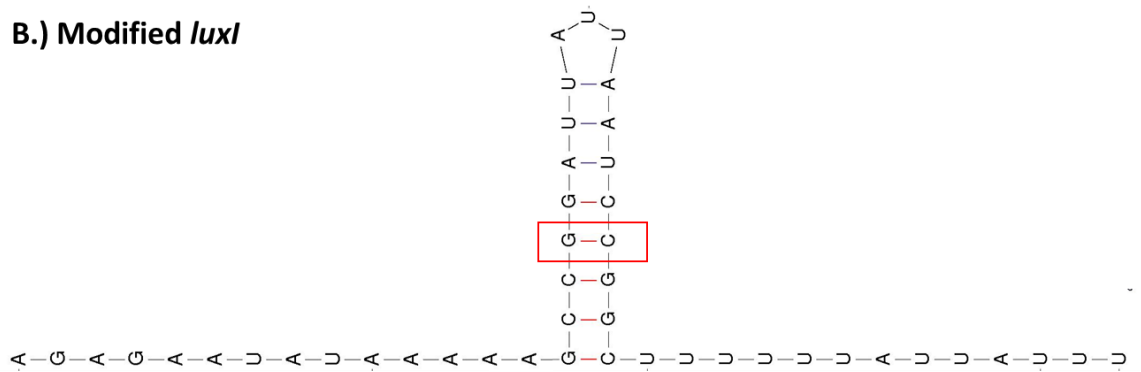


Figure 4-16 The predicted hairpin structure of (A.) *luxI* and (B.) modified *luxI* by **Mfold**. The red boxes indicate the position of the noncomplementary base pair on the stem-loop structure on *luxI* terminator, which was modified by changing A to G to increase the termination efficiency.

Table 4-2 The sequences, values of parameters and termination efficiency of *luxI* and modified *luxI*, estimated by an algorithm described by d'Aubenton Carafa *et al.* (1990).

Terminators	Strand	Sequences (5'->3') ^A	Parameters					Termination efficiency (%)
			n_T	ΔG^B (kcal/mol)	L_H	γ	d	
<i>luxI</i>	Positive strand	GAATATAAAAAAGCCAGATTATTAATCCGGCTTTTTTATTATT	5.027	-5.7	19	0.3	3.4	23
	Negative strand	ATAATAAAAAAGCCGGATTAAATAATCTGGCTTTTTTATATTCTCT	4.46	-9.6	19	0.505	12.9	50
Modified <i>luxI</i>	Positive strand	GAATATAAAAAAGCCGGATTATTAATCCGGCTTTTTTATTATT	5.027	-11.8	19	0.621	34.4	75
	Negative Strand	ATAATAAAAAAGCCGGATTAAATAATCCGGCTTTTTTATATTCTCT	4.46	-11.8	19	0.621	24.1	85

^A The regions forming loop and stem structures of terminators are shown in red and blue colour, respectively.

^B The Gibbs free energy was determined by Mfold (Zuker, 2003).

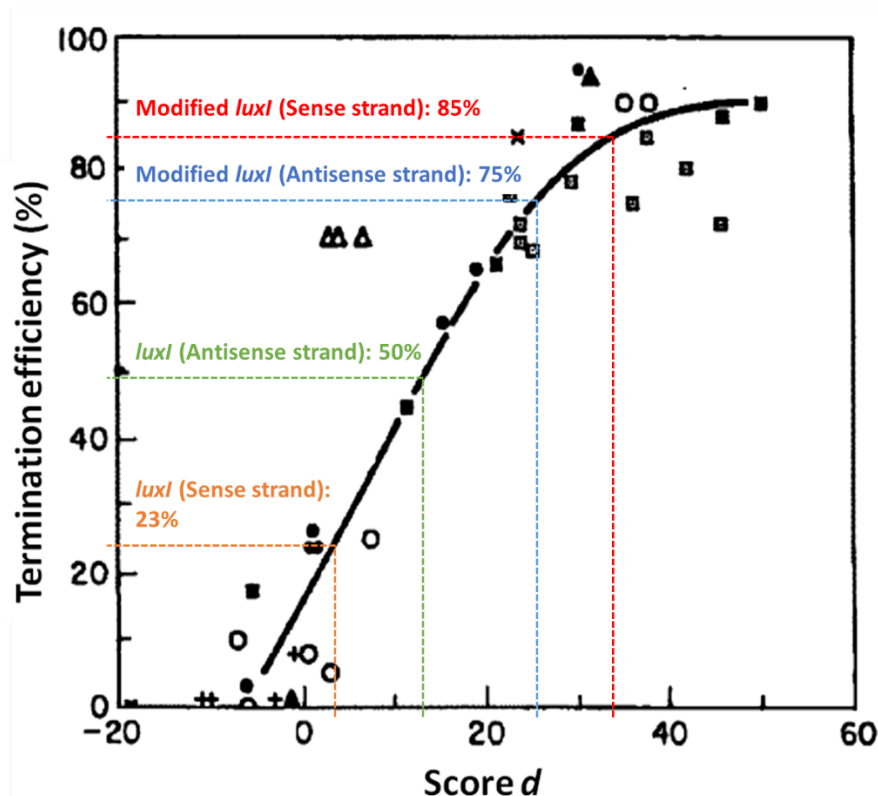


Figure 4-17 The determination of termination efficiency of *luxI* and modified *luxI* bi-directional terminators. The termination efficiency was estimated based on an algorithm described previously by d'Aubenton Carafa *et al.* (1990). The *d* scores of the terminator on positive strand (3.4) and negative strand (12.9) of *luxI* were plotted on the x-axis and correlated with 23% (orange dashed line) and 50% (green dashed line) termination efficiencies, respectively. For the modified *luxI*, the *d* scores of the terminator on positive strand (34.4) and negative strand (24.1) were plotted on the x-axis and correlated with 85% (red dashed line) and 75% (blue dashed line) termination efficiencies, respectively.

4.3.5.2 The disruption of promoter formed during the construction of pCC1BAC-*lacZ* α -*gusA* plasmid

The pCC1BAC-*lacZ* α -*gusA* plasmid was constructed as described in section 4.2.4. The *E. coli*::pCC1BAC-*lacZ* α -*gusA*, however, showed blue colonies on the LB agar containing IPTG/X-gal, suggested that the *lacZ* α was expressed

by some promoters within *lacZα-gusA*. An *in silico* promoter analysis was then performed by using BPROM, which identified two putative promoters: at *NsiI*-*NheI* restriction site in the middle of both reporter genes and at the beginning of *gusA*, as shown in Figure 4-18.

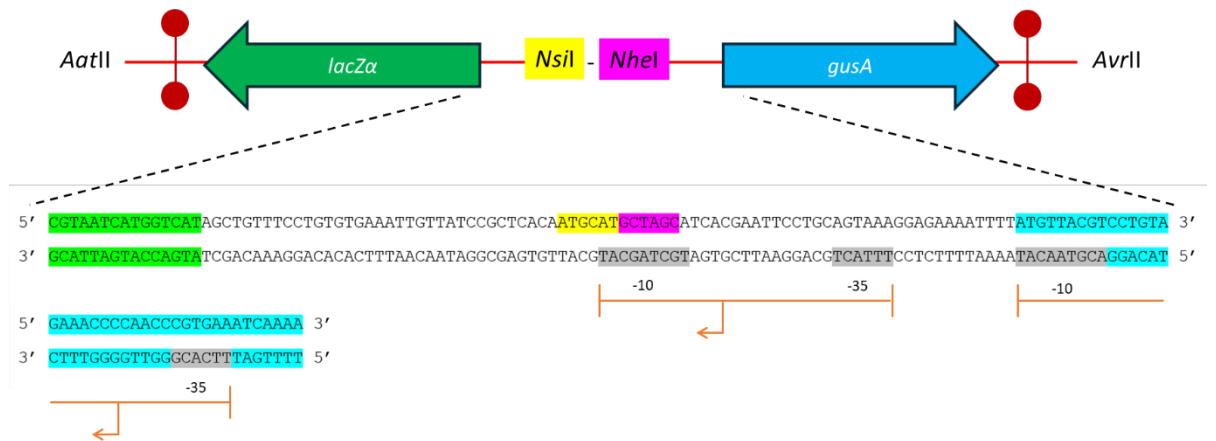



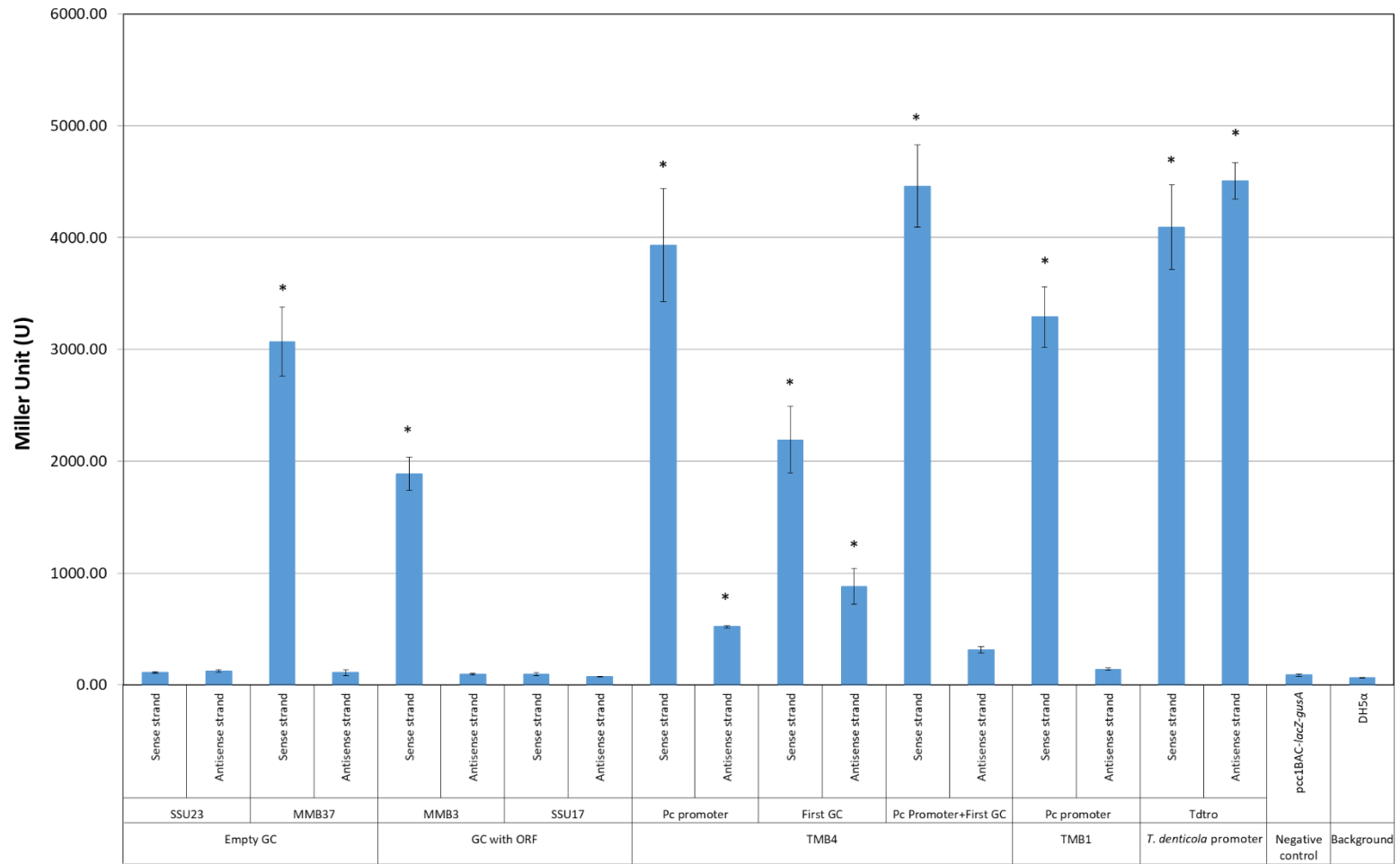
Figure 4-18 Position of putative promoters expressing *lacZα* predicted by BPROM. The open arrow boxes represent reporter genes, pointing the direction of transcription, and the symbols () represent bi-directional terminators. The green, blue, yellow and pink colours represent *lacZα*, *gusA*, *NsiI* and *NheI*, respectively. The predicted putative promoters were indicated with orange arrows, pointing the direction of transcription, and the -10 and -35 boxes were highlighted with grey colour.

The disruption of the putative promoter was first carried out at the restriction sites. The *NsiI* linker was inserted *NsiI* site on pCC1BAC-*lacZα-gusA* plasmid, as described in section 4.2.4.3. The *E. coli* containing pCC1BAC-*lacZα-gusA* with the *NsiI* linker showed white colonies on LB agar containing IPTG/X-gal, suggested that the promoter responsible for the expression of *lacZα* were the one at *NsiI*-*NheI* restriction sites.

4.3.6 Promoter activity determination of pCC1BAC-*lacZ* α -GC-*gusA* constructs by using β -glucuronidase enzyme assay

The β -glucuronidase enzyme assay was performed on *E. coli* containing pCC1BAC-*lacZ* α -GC-*gusA*, and the results were shown in Figure 4-19. The enzymatic results showed that MMB37 and MMB3 had promoter activity on the sense strand, while MMU23 and SSU17 had no promoter activity on either strand, compared to the negative controls. TMB4-P_C, TMB4 GC, TMB4 P_C-GC, and TMB1-P_C constructs, all showed the promoter activities on both sense and antisense strands. As the P_C promoter sequences on TMB1 and TMB4 samples were different in several nucleotides, it was shown that TMB4-P_C had higher promoter activities than the TMB1-P_C in both directions.

The concentration of β -glucuronidase enzyme in Miller unit (U) from pCC1BAC-*lacZ-gusA* constructs



Constructs

Figure 4-19 The promoter activity from pCC1BAC-*lacZ-gusA* constructs estimated by β -glucuronidase enzyme assays. Error bars indicate the standard errors of the means from three replicates. The asterisks (*) indicate the constructs were statistically significantly different from the negative control group (pCC1BAC-*lacZ-gusA*) with the *p*-value <0.05 by using ordinary one-way ANOVA followed by Dunnett's multiple comparison test.

4.3.7 Promoter activity determination of pCC1BAC-*lacZ* α -GC-*gusA* constructs by using β -galactosidase enzyme assay

The promoter activity was also performed by β -galactosidase enzyme assay (section 4.2.6) to determine the expression of *lacZ* α on pCC1BAC-*lacZ* α -GC-*gusA* constructs. However, the reactions were very slow, compared to the β -glucuronidase enzyme assay, which the changes in the colour of solution could be observed immediately after adding substrate. For the β -galactosidase, the solution could only change to faint yellow even after 4 hr 30 min incubation. The Miller units calculated with the formula were less than 1, due to the very long incubation time.

In order for β -galactosidase to function in the *E. coli* containing pCC1BAC-*lacZ* α -GC-*gusA* constructs, another domain of LacZ, LacZ Ω , encoded by the *E. coli* genome needs to be expressed, as our construct encoded only the LacZ α domain of the β -galactosidase enzyme. IPTG should be added in the medium to inactivate *lac* repressors, allowing the expression of LacZ Ω from the *lac* operon (Juers *et al.*, 2012). The β -galactosidase assay was then performed on 4 constructs, including TMB4-GC (sense strand), Tdtro (sense strand), pCC1BAC-*lacZ* α -*gusA* (negative control) and DH5 α (background), with and without the 1mM IPTG. The results showed that adding IPTG increased the promoter activity, however, the rate of reaction was still slow, and the Miller units of all constructs were still below 1, as shown in Figure 4-20.

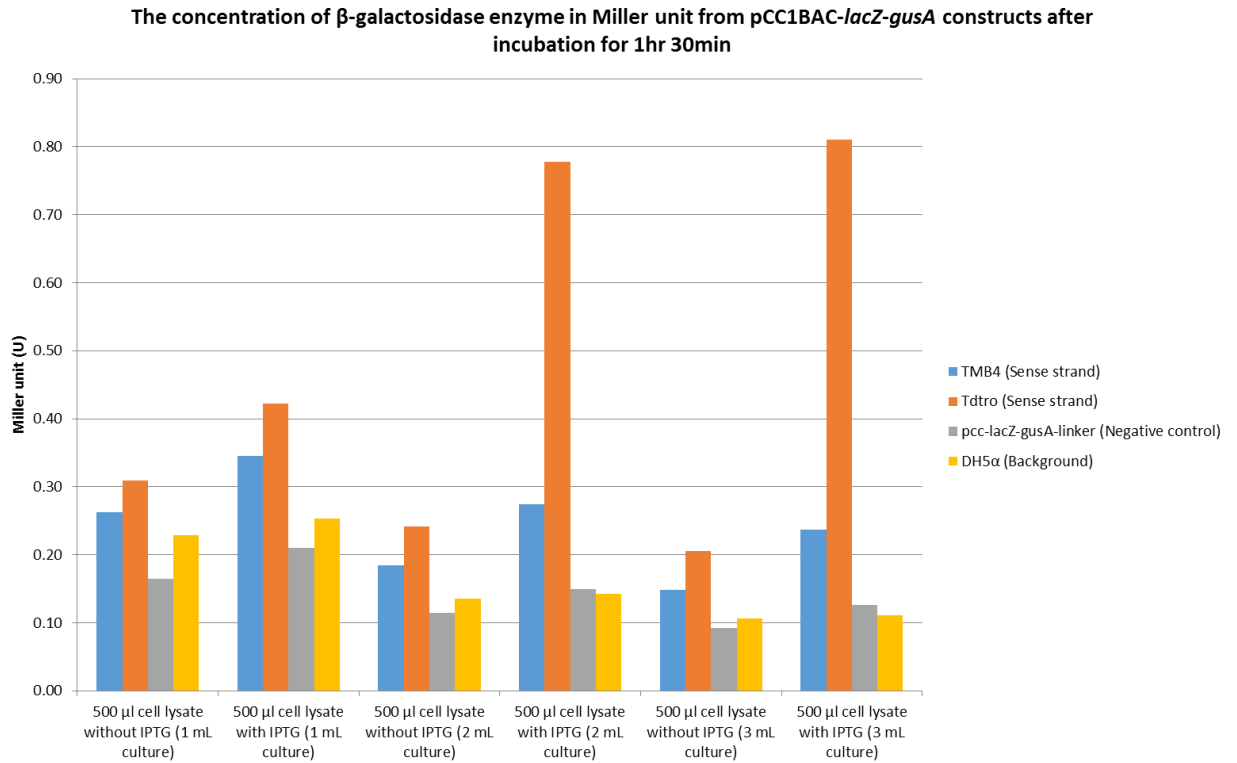


Figure 4-20 The promoter activity from pCC1BAC-*lacZα-gusA* constructs estimated by β -galactosidase enzyme assays.

4.3.8 Detection of promoter-containing GCs by using Bi-Directional Promoter Detection (BiDiPD) construct.

4.3.8.1 Determination of substrates for *lacZα* and *gusA* reporter genes

Different combinations of substrates were added to the LB agars, which were then inoculated with *E. coli* strains (with an active *lacZα*, *gusA* and no reporter gene), to determine the appropriate substrates for the promoter detection, as shown in Figure 4-21.

		<i>gusA</i> substrate			
		Without <i>gusA</i> substrate	+PNPG	+XGluc	+MUG
<i>lacZ</i> substrate	Without X-gal	<p>White (W) = no reporter gene</p> <p>Blue (B) = clone with <i>lacZ</i> gene</p> <p><i>gusA</i> (G) = clone with <i>gusA</i> gene</p> <p>• G • B • W</p>			
	+ Xgal				

Figure 4-21 The determination of substrates for the promoter detection activity with *lacZ* and *gusA* reporter genes. Each plate was divided into 4 sections. Three sections were inoculated with *E. coli* containing no reporter (white: W), *lacZ* (Blue: B) and *gusA* (G). The last section (lower right corner) was inoculated with a single colony of each strain. The plates in the top and bottom row were supplemented with and without X-Gal, respectively. The plates in each column contained different *gusA* substrates, including PNPG, X-Gluc and MUG. The plate containing MUG was visualised under UV light.

The *E. coli::gusA* on the agar plates supplemented with PNPG showed yellow colour around the colonies, while they showed blue colour on the plates containing X-Gluc. The blue colour on the X-Gluc plate was less diffuse from the colonies than the yellow colour on the PNPG plate. By exposing the plates containing MUG under UV light, the *E. coli::gusA* exhibited blue fluorescence surrounding the colonies. It was also shown that β -glucuronidase, encoded by *gusA*, cannot catalyse X-Gal as the *E. coli::gusA* did not show blue colour in the presence of X-Gal (bottom row of Figure 4-21). However, the *E. coli::lacZ α* colonies exhibited a small background activity to all *gusA* substrates, suggesting that β -galactosidase, encoded by *lacZ*, could partially catalyse the substrate for *gusA*.

From the results shown in Figure 4-21, the yellow colour on PNPG plate was too similar to the colour of LB agar, which can be difficult to detect around single colonies. X-Gluc was also not appropriate as it gave a similar blue colour to the byproduct from X-Gal, so it could not be used for the differentiation between clones with promoter activity on sense and/or antisense strands. Therefore, the best substrate for *gusA* to use together with X-Gal was MUG, which allows us to detect and distinguish the clones with promoter activity from different strands of the inserts. The clones with GCs contain a promoter on the sense strand will show blue fluorescence when observed under UV light, reflecting the activity of β -glucuronidase enzymes catalysing MUG. The colour of the clones with promoter activity on the antisense strand will be blue under normal light as a result of β -galactosidase enzymes catalysing X-Gal.

4.3.8.2 Detection of promoter-containing GCs from the oral metagenome using pBiDiPD

The oral integron GCs were amplified and cloned between the reporters, as described in section 4.2.8, to verify the functionality of pBiDiPD as a promoter detection system and also to detect for novel GCs containing promoter sequences in the human oral metagenome. The colonies showing either *lacZ* or *gusA* activities were identified on LB agar containing X-Gal and MUG (Figure 4-22).

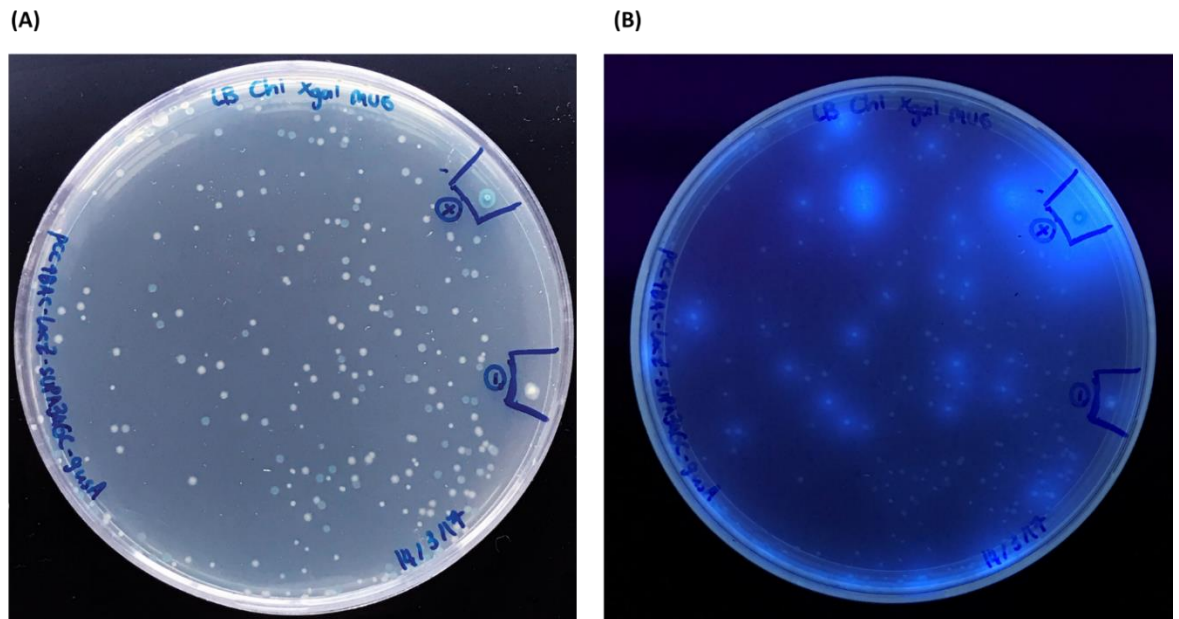









Figure 4-22 The detection of the integron GCs by using pBiDiPD. A.) Blue-white screening to detect for the clones with promoter activity on antisense strand, B.) Exposing the colonies under the UV light to detect clones with promoter activity on the sense strand. The positive (+) and negative (-) colonies were the *E. coli* containing pCC1BAC-*lacZ* α -TMB4-Pc-*gusA* (with experimentally proven promoter activities on either strand of DNA and pCC1BAC-*lacZ* α -*gusA* (no promoter activity), respectively





After screening the clones from 2 libraries (amplified with SUPA3-4 and MARS2-5 primers (Appendix 4, Tansirichaiya *et al.*, 2016b), 23 different GCs with promoter activities were identified (Table 4-3). Fourteen of the identified GCs had a match with the GCs found in the previous study (Tansirichaiya *et al.*, 2016b), while the other 9 GCs had not been reported, including samples SSU-Pro-20, SSU-Pro-27, SSU-Pro-32, SSU-Pro-46, SSU-Pro-65, MMU-Pro-5, MMU-Pro-24, MMU-Pro-48 and MMU-Pro-53. These putative GCs were verified not to be PCR artefacts by detecting the consensus R' (1R) core sites [GTTRR(Y)R(Y)Y(R)] and the complementary R'' (1L) core sites [R(Y)Y(R)Y(R)YAAC] of *attC* located downstream from the *attC* forward primers and upstream from the *attC* reverse primers, respectively (Stokes *et al.*, 1997) (Table 4-4).





The identified GCs could be categorised into two groups. The first group was the non-coding GCs, found in 7 samples, which most were reported in the previous study, except sample MMU-Pro-53. Another group of GCs were predicted to encode toxin-antitoxin systems in 12 out of 23 GCs, including PilT N-terminus (PIN) (toxin)-MazE (antitoxin), plasmid stabilization protein (toxin)-prevent-host-death protein (antitoxin), RelE (toxin)-transcription regulator (antitoxin), PemK (toxin)-MazE (antitoxin), BrnT (toxin)-BrnA (antitoxin), ParE (toxin)-transcription regulator (antitoxin). Most of the samples (14 out of 23 GCs) showed promoter activity on the sense strand only, while 3 GCs showed activity only on the antisense strands, i.e. MMU-Pro-6, MMU-Pro-63 and MMU-Pro-65. The other 10 samples exhibited the promoter activity on both strands.






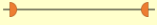
Table 4-3 Characterisation of the human oral integron GCs containing promoter sequences detected by pBiDiPD.



Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
SSU-Pro-7	SUPA3-SUPA4	1001		SSU22	98/95	Prevent-host-death protein (Phd_YefM antitoxin superfamily) [<i>Treponema vincentii</i>]	264	97/100	WP_006188308.1	Y	N
						XRE family transcriptional regulator [<i>Treponema vincentii</i>]	441	98/100	WP_006188306.1		
SSU-Pro-9	SUPA3-SUPA4	834		MMB3	98/99	Hypothetical protein (antitoxin, ribbon-helix-helix domain protein) [<i>Treponema putidum</i>]	246	67/100	WP_044978234.1	Y	N
						Twitching motility protein PilT (PIN toxin domain) [<i>Treponema putidum</i>]	414	71/100	AIN93467.1		
SSU-Pro-13	SUPA3-SUPA4	855		MMB39	98/99	Toxin RelE [<i>Treponema medium</i>]	357	95/100	WP_016522532.1	Y	N
						Transcriptional regulator (Antitoxin, XRE family) [<i>Treponema medium</i>]	330	95/100	WP_016522533.1		

Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
SSU-Pro-16	SUPA3-SUPA4	925		SSU28	98/100	Multidrug transporter MatE (MazE antitoxin superfamily) [<i>Treponema putidum</i>]	231	96/100	WP_044979179.1	Y	N
						mRNA-degrading endonuclease (PemK toxin superfamily) [<i>Treponema denticola</i>]	336	99/100	WP_010694033.1		
SSU-Pro-20	SUPA3-SUPA4	1263		MMU28	77/42	Prevent-host-death protein (Phd_YefM antitoxin superfamily) [<i>Treponema sp. JC4</i>]	249	75/88.3	WP_009103386.1	Y	N
						Plasmid stabilization protein (ParE toxin superfamily) [<i>Treponema sp. JC4</i>]	147	57/43.8	WP_009104800.1		
SSU-Pro-24	SUPA3-SUPA4	425		SSU9	99/100	-	-	-	-	Y	Y
SSU-Pro-27	SUPA3-SUPA4	753		<i>Treponema putidum</i> strain OMZ 758	93/100	Hypothetical protein (Ribonuclease toxin, BrnT superfamily) [<i>Treponema denticola</i>]	273	99/100	WP_002666393.1	Y	N
						Hypothetical protein (BrnA antitoxin superfamily) [<i>Treponema denticola</i>]	288	97/100	WP_002676616.1		

Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
SSU-Pro-32	SUPA3-SUPA4	972		No significant similarity found.	-	Hypothetical protein (ParE toxin superfamily) [<i>Treponema denticola</i>]	354	98/100	WP_002683264.1	Y	N
						Transcriptional regulator (Antitoxin HigA) [<i>Treponema denticola</i>]	273	100/100	WP_002683262.1		
SSU-Pro-34	SUPA3-SUPA4	832		SSU5	99/100	Hypothetical protein (antitoxin, ribbon-helix-helix domain protein) [<i>Treponema putidum</i>]	246	67/100	WP_044978234.1	Y	N
						Twitching motility protein PilT (PIN toxin domain) [<i>Treponema putidum</i>]	414	71/100	AIN93467.1		
SSU-Pro-39	SUPA3-SUPA4	1137		MMU25	99/99	Hypothetical protein [uncultured bacterium]	462	99/100	ANC55535.1	Y	N
						Hypothetical protein [<i>Treponema maltophilum</i>]	213	88/100	WP_016526060.1		
						mRNA interferase MazF2 toxin [<i>Treponema bryantii</i>]	351	48/91	WP_022932935.1		
SSU-Pro-46	SUPA3-SUPA4	971		No significant similarity found	-	Hypothetical protein [<i>Treponema socranskii</i>]	267	80/100	WP_021329686.1	Y	N

Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
						Hypothetical protein [<i>Treponema socranskii</i>]	228	84/100	WP_021329641.1		
						Hypothetical protein [<i>Treponema sp.</i> C6A8]	276	67/100	WP_027729334.1		
SSU-Pro-65	SUPA3-SUPA4	811		<i>Treponema sp.</i> OMZ 838	91/21	Hypothetical protein (MazE antitoxin) [<i>Treponema denticola</i>]	228	93/100	WP_010693782.1	Y	N
						PIN toxin domain-containing protein [<i>Treponema denticola</i>]	402	93/100	WP_010693784.1		
MMU-Pro-4	MARS5-MARS2	520		MMU2	99/100	-	-	-	-	Y	Y
MMU-Pro-5	MARS5-MARS2	983		<i>Treponema putidum</i> strain OMZ 758	94/78	Prevent-host-death protein (Phd_YefM antitoxin superfamily) [<i>Treponema denticola</i>]	240	98/98.8	WP_002669519.1	Y	Y
						RelE/StbE family addiction module toxin [<i>Treponema denticola</i>]	318	94/100	WP_002688980.1		
MMU-Pro-6	MARS5-MARS2	737		MMB36	86/100	-	-	-	-	N	Y

Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
MMU-Pro-18	MARS5-MARS2	634		MMB37	95/100	-	-	-	-	Y	N
MMU-Pro-22	MARS5-MARS2	431		MMU19	91/100	-	-	-	-	Y	Y
MMU-Pro-24	MARS5-MARS2	904		No significant similarity found	-	Universal stress protein UspA [<i>Marinobacter manganoxydans</i>]	348	30/79.5	WP_008177208.1	Y	Y
						Hypothetical protein [<i>Methylobacter tundripaludum</i>]	213	79/100	WP_031438379.1		
						Prevent-host-death protein [<i>Treponema pedis</i>]	84	76/27.8	WP_024469914.1		
MMU-Pro-31	MARS5-MARS2	574		MMB5	88/70	-	-	-	-	Y	N
MMU-Pro-48	MARS5-MARS2	817		<i>Treponema sp.</i> OMZ 838	91/25	Hypothetical protein (MazE antitoxin) [<i>Treponema denticola</i>]	228	93/100	WP_010693782.1	Y	N
						PIN toxin domain-containing protein [<i>Treponema denticola</i>]	402	93/100	WP_010693784.1		
MMU-Pro-53	MARS5-MARS2	430		No significant similarity found	-	-	-	-	-	Y	Y

Gene cassettes	Primer pair	Cassette Size (bp)	Orientation*	BlastN		BlastX				Promoter activity	
				Closest homologue	Percentage identity (%)/ Coverage (%)	Closest homologue	ORF size (bp)	Percentage identity (%)/Coverage (%)	Accession number of the homologous proteins (BlastX)	Sense Strand (<i>gusA</i>)	Antisense strand (<i>lacZ</i>)
MMU-Pro-63	MARS5-MARS2	927		SSU8	99/99	Hypothetical protein [<i>Treponema denticola</i>]	531	98/93.7	WP_002692239.1	N	Y
MMU-Pro-65	MARS5-MARS2	896		MMU27	99/100	Hypothetical protein [uncultured bacterium]	399	99/84.2	ANC55539.1	N	Y
						Hypothetical protein [uncultured bacterium]	357	99/100	ANC55540.1		

**The orange half circles and green arrow boxes are representing *attC* sites and ORFs, respectively.

** The GC samples that have not been reported in Tansirichaiya et al. (2016) are highlighted in yellow.

Table 4-4 Complementarity of the core sites R' (1R) and R'' (1L) abutting the forward and reverse *attC* primer sequence on the gene cassettes.

Type (Number of GCs)	Sequence of R' after the forward primer sequence of the <i>attC</i> on GCs	Pattern of R' sequence of the GC ^a	Sequence of R'' before the reverse primer of the <i>attC</i> on GCs GC	Pattern of R'' sequence of the <i>attC</i> on GCs GC ^a	Clones	Complementarity between R' and R'' core sites of the <i>attC</i> on GCs
A (15)	GTTAAGC	GTTRRRY	ATCTAAC	RYYAAC	SSU-Pro-9	7/7
	GTTAAGC	GTTRRRY	CTTCAAC	YYAAC	MMU-PRO-32	6/7
	GTTAGAC	GTTRRRY	GCCTAAC	RYYAAC	MMU-PRO-6	7/7
	GTTAGAC	GTTRRRY	CTTCAAC	YYAAC	SSU-Pro-24, MMU-PRO-5, MMU-PRO-18, MMU-PRO-22, MMU-PRO-53	6/7
	GTTAGGT	GTTRRRY	ACCTAAC	RYYAAC	SSU-Pro-20, SSU-Pro-32, SSU-Pro-65, MMU-PRO-24	7/7
	GTTAGGT	GTTRRRY	ACCTAAC	RYYAAC	MMU-PRO-48	7/7
	GTTGAAC	GTTRRRY	ATCTAAC	RYYAAC	SSU-Pro-34	7/7

Type (Number of GCs)	Sequence of R' after the forward primer sequence of the <i>attC</i> on GCs	Pattern of R' sequence of the GC ^a	Sequence of R'' before the reverse primer of the <i>attC</i> on GCs GC	Pattern of R'' sequence of the <i>attC</i> on GCs GC ^a	Clones	Complementarity between R' and R'' core sites of the <i>attC</i> on GCs
	GTTGAAC	GTTRRRY	CTTCAAC	YYYYAAC	MMU-PRO-31	6/7
B (3)	GTTAGAA	GTTRRRR	TTCTAAC	YYYYAAC	SSU-Pro-7	7/7
	GTTAGAG	GTTRRRR	CTCTAAC	YYYYAAC	SSU-Pro-46	7/7
	GTTAGGA	GTTRRRR	TCCTAAC	YYYYAAC	MMU-PRO-65	7/7
C (3)	GTTATAC	GTTRYRY	GCCTAAC	RYYYYAAC	MMU-PRO-63	6/7
	GTTATGT	GTTRYRY	ACCTAAC	RYYYYAAC	SSU-Pro-16	6/7
	GTTATGT	GTTRYRY	ACCTAAC	RYYYYAAC	SSU-Pro-39	6/7
D (2)	GTTAGCT	GTTRRYY	AGCTAAC	RRYYAAC	SSU-Pro-13	7/7
	GTTAGCT	GTTRRYY	ATCTAAC	RYYYYAAC	SSU-Pro-27	6/7

^a Degenerate nucleotide: R = A or G; Y = C or T.

4.4 Discussion

As integrons play an important role in the dissemination of ARGs, it is essential to understand the diversity of GCs and their expression. Even though the integron GCs, recovered by the PCR-based approaches from previous environmental metagenomes, mostly contained a single ORF pointing in a forward direction, noncoding GCs were also found as in most of the studies (Stokes *et al.*, 2001, Elsaied *et al.*, 2011, Tansirichaiya *et al.*, 2016b).

In this study, we determined the promoter activity based on the expression of a *gusA* reporter gene by measuring the amount of the β -glucuronidase enzyme from multiple GC containing constructs. As the noncoding GCs were recovered from the oral metagenome, there is no information on the bacterial host and we can only test the promoter activities in a suitable surrogate host. The nucleotide sequence analysis showed that the GCs recovered from the oral metagenome were likely to be derived from *Treponema* spp., so the recognition of *T. denticola* promoter sequences by *E. coli* host had to be confirmed. The *T. denticola* promoter identified from a previous study, called P_{TdTro} (Brett *et al.*, 2008), was therefore included. P_{TdTro} showed high promoter activity from both sense and antisense strands in *E. coli*, demonstrating that *E. coli* could be used as a surrogate host for *Treponema* promoter expression to determine the promoter activity of the selected GCs. It is also a suitable promoter to serve as a positive control. However, as no promoter activity detected from P_{Fla}, it suggested that our enzymatic assay cannot detect promoters associated with σ^{28} from *Treponema* spp. This could be due to either an inability for the *E. coli* host to recognise the *Treponema* σ^{28} promoter

or a low flagella gene expression and nonmotile phenotype of the *E. coli* α -select strain (Wood *et al.*, 2006).

Promoter-containing GCs have been identified previously by using several approaches. For example, the *ere(A)*-containing GC conferred an erythromycin resistance phenotype independent from P_{lac} promoter of the vector (Biskri & Mazel, 2003), and the *qnrVC1*-containing GC could express the green fluorescent protein (GFP) on the pGlow vector (da Fonseca & Vicente, 2012). In our study, we used the *gusA* reporter gene to detect the promoter activity on noncoding GCs, which was first developed in the pUC19 cloning vector. However, there was an issue with two of the pUC19-GC-*gusA* constructs, TMB4 P_C and TMB4 P_C -GC, which had significantly lower plasmid copy number relative to the other strains containing variant constructs. It was hypothesised that both inserts could affect the plasmid stability due to their high promoter activities. For example, they could increase metabolic burden to maintain and replicate the plasmids in the host, or the accumulation of plasmid multimers due to over-replication (Summers & Sherratt, 1984, Xu *et al.*, 2006).

To solve the copy number issue, new constructs were made based on a single-copy pCC1BAC vector by cloning the inserts into multiple cloning sites on the vectors. However, the enzymatic assay on the pCC1BAC-GC-*gusA* constructs showed promoter activity in the negative control (containing only *gusA*). Therefore, the new construct was designed to include two reporter genes, *lacZ α* and *gusA*, flanking the insert. Initially, the purpose of this design was to measure promoter activities on both strands of DNA by performing β -

glucuronidase and β -galactosidase enzymatic assays at the same time on the cell lysate from only one construct. However, the β -galactosidase enzymatic assay showed a very slow kinetic reaction, even with the induction by IPTG. As the β -galactosidase enzyme in this construct was split into LacZ α and LacZ ω , the cell lysate extraction in the assay could interrupt or break the reassembly of both peptides to form a functional enzyme.

The promoter activities of the GCs on pCC1BAC-*lacZ* α -GC-*gusA* constructs were, therefore, measured based on only the *gusA* reporter genes by the β -glucuronidase enzymatic assay. This is the first time that the promoter activity of noncoding GCs was demonstrated *in vitro*, as shown by the activity on both strands of the TMB4 and the sense strand of the MMB37. A previous study on the *Vibrio* integron, containing an 116-cassette array, showed that most of the GCs were transcribed (Michael & Labbate, 2010). Therefore, these non-coding GCs could be responsible for the transcription of the other GCs that cannot be transcribed by P_C promoter of integrons.

For the TMB4 (noncoding GC in the first position), we first hypothesised its function in the first GC position to increase the expression of the GCs located downstream from P_C promoter. This was previously shown that coupling P_C promoter with another promoter, such as a P₂ promoter (located 119 bp downstream), could result in a significantly higher expression of GCs (Lévesque *et al.*, 1994, Papagiannitsis *et al.*, 2009). Therefore, the promoter activity of TMB4 P_C and TMB4 P_C+GC were also measured to test this hypothesis. The results showed that coupling promoter TMB4 P_C with TMB4 GC slightly increased the promoter activity (Figure 4-19). However, the

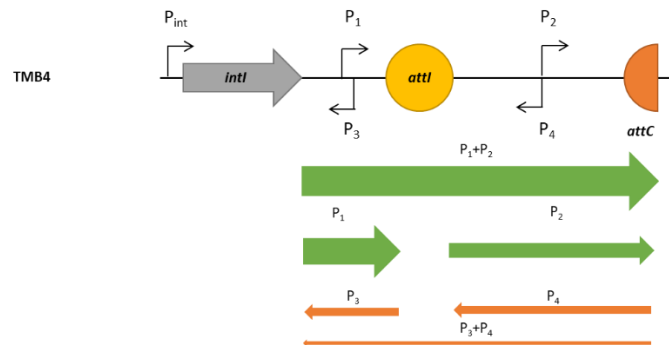
increase in promoter activities was not significant (p -value >0.99 by using ordinary one-way ANOVA followed by Bonferroni's post-hoc).

With more promoters introduced, there could be more competition for enzymes and proteins involved in transcription such as RNA polymerases (RNAP) or sigma factors for each promoter, resulting in lower transcriptional levels (Lamberte *et al.*, 2017). As the transcripts from the promoters on sense (TMB4 P_C promoter) and antisense strands (TMB4 GC) were overlapping, it could result in RNA-RNA hybridisation, reducing the number of available transcripts for the expression of the reporter genes (Figure 4-23A).

Insertion of TMB4 P_C in the first position could also result in transcriptional interference (TI) between four predicted promoters on the TMB4 P_C+GC construct (Figure 4-23C). The analysis by BPRM and enzymatic assay suggested that there were 4 putative promoters on the TMB4 P_C and GC constructs (two promoters on each strand), represented by P₁, P₂, P₃ and P₄ in Figure 4-23 A and B. The promoter P₁ on TMB4 P_C was convergent to P₄ and in-tandem with P₂, while the promoter P₃ was convergent to P₄. TI could occur through several mechanisms (Shearwin *et al.*, 2005). There might be promoter competition by the occupation of RNAP at P₂ and P₄, which can restrict and preclude the occupation of RNAP at P₁ and P₃ (Figure 4-23C-1). The transcription-elongation complex at promoter P₂ and P₄ can be considered as a sitting duck, which could be hit and dislodged by the arrival of the transcriptional bubble initiated from the strong promoter P₁ (Figure 4-23C-2) (Callen *et al.*, 2004). The transcriptional bubble at the strong promoter P₁ can also act as a roadblock, preventing the progress of RNAP

from promoter P_4 (Figure 4-23C-5) (Epshtein *et al.*, 2003). TI can also happen at the elongation stage in which the RNAP progressing from two promoters collide to each other causing the stalled state or fall off of the RNAPs (Figure 4-23C-4) (Prescott & Proudfoot, 2002). Another mechanism, called occlusion, occurs when the occupation of RNAP at promoter P_2 and P_4 is blocked by an elongating RNAP initiating at a promoter P_1 (Figure 4-23C-3) (Adhya & Gottesman, 1982).

(A)



(B)

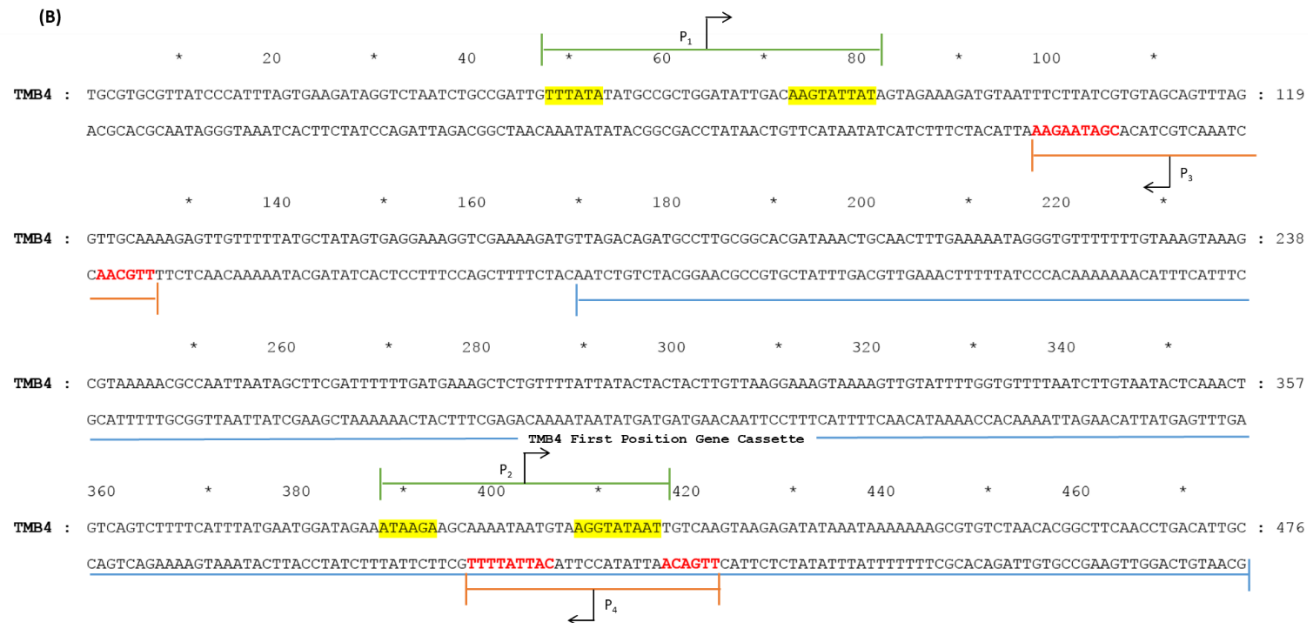


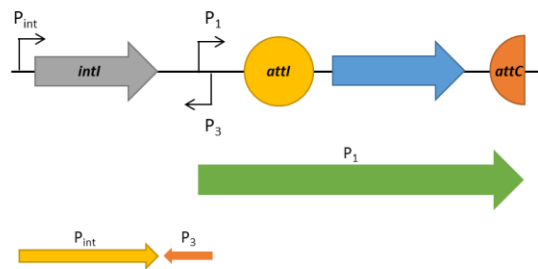
Figure 4-23 The transcriptional activity of the TMB4 GC constructs. (A) Schematic representation of the structure of TMB4 in the integron. The grey open arrowed box represents integrase gene (*intI*), pointing in the direction of transcription. The recombination sites, *attI* and *attC*, were represented by yellow and orange circles, respectively. The black arrows indicate promoters in the constructs. The green and orange arrows indicate the promoter activity in sense and antisense strands, respectively. The thickness of the arrows represents the relative promoter activity as determined by our β -glucuronidase assays in TMB4 P_C (P₁ and P₃), GC (P₂ and P₄) and PC+GC (P₁+P₃ and P₂+P₄). (B) The nucleotide sequence and organisation of the promoters in TMB4 constructs predicted by BRPOM. The promoters in the sense strand and antisense strands are indicated with green and orange lines. The -10 and -35 boxes of the promoters on the sense and antisense strands are highlighted in yellow and shown in bold-red text, respectively. (C) An overview of the mechanisms for transcriptional interference (TI), which are 1.) promoter competition, 2.) sitting duck interference, 3.) occlusion, 4.) collision and 5.) roadblock. P_A and P_S promoters represent a strong (aggressive) promoter and a weak (sensitive) promoter, respectively. (Retrieved from Shearwin *et al.*, 2005)

In usual integrons, TI occurred between the P_C promoter, located in *intl*, and the integron integrase promoter P_{intl} , which are oriented convergent to each other. TI between P_C and P_{intl} have been shown to control the expression of integrase and the subsequent recombination of GCs. The stronger P_C promoter could result in lower expression of integrase, which decreases the recombination of GCs from integrons (Jové *et al.*, 2010, Guerin *et al.*, 2011). This relationship of P_C promoter and P_{intl} might also apply to the reverse integrons found in *Treponema* spp., even though the direction and position of their P_{intl} , P_C and *intl* gene are different from the usual integrons (Figure 3-1). With the differences in P_C sequences between TMB1 and TMB4, it was shown that TMB4 P_C had higher activity on either strand. As the transcripts from antisense P_C promoter overlap with the transcripts from P_{intl} , the stronger activity of the antisense P_C promoter could result in a lower number of integrase as a result of RNA-RNA hybridisation and TI, leading to more stable GC arrays.

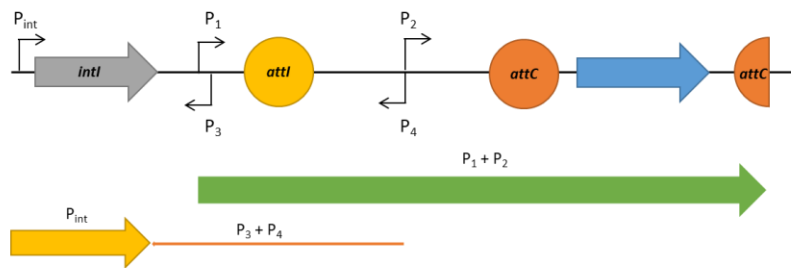
The integration of the promoter-containing GC in the first position on a GC array is hypothesised to be a controlling mechanism for the expression of the integrase in reverse integrons (Figure 4-24). In the absence of promoter-containing GC at the first position, the expression of integrase could be repressed by the activity of antisense TMB4 P_C , maintaining the GCs in their position. By inserting the promoter GC in the first position, the number of transcripts from the antisense TMB4 P_C will be reduced, which will subsequently decrease the effect of RNA-RNA hybridisation or TI. Therefore, more transcripts from P_{intl} will be available for translation and increase the expression of integrase. This could result in an increase of recombination

events and will catalyse the introduction of new GCs into the first position to be expressed by P_C and presumably selected for by external stressors. When the new GC is inserted in the first position, it will push the promoter GC further down the array, which will diminish the interference effect and therefore the integrase expression, preserving the new GCs in the first position.

A.) Without Promoter GC at the first position



B.) With Promoter GC at the first position



C.) After insertion of new GC

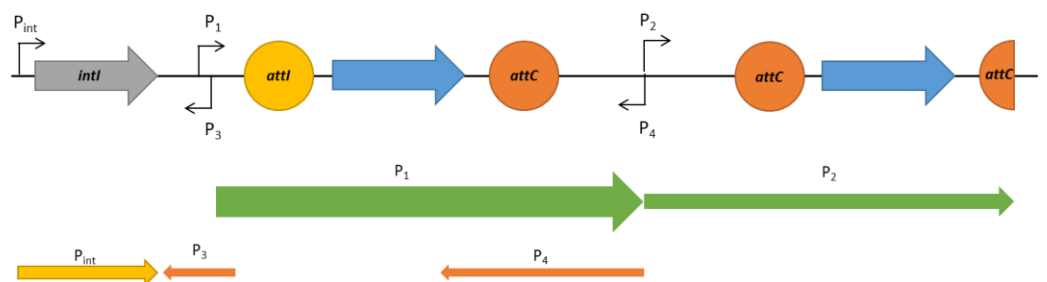


Figure 4-24 The proposed function of the promoter-containing GC in the first position to control the expression of the integrase in reverse integrons. (A) Without the promoter GC, (B) With Promoter GC in the first position, (C) After insertion of new GC. The grey and blue open arrowed boxes represent integrase gene (*intI*) and the open reading frames (ORFs), respectively, pointing in the direction of transcription. The recombination sites, *attI* and *attC*, were represented by yellow and orange circles, respectively. The green and orange arrows indicate the promoter

activity from the P_C promoter and promoter GC in sense and antisense strands, respectively, while the yellow arrows represent the promoter activity from P_{int} . The thickness of the arrows represents the predicted relative promoter activity in each stage.

The expression level of cassette genes located further down in the array normally decreases due to the formation of a stem-loop structure on mRNA at *attC* sites, which impede the progression of the ribosome (Jacquier *et al.*, 2009). It was previously shown that the level of streptomycin resistance was reduced four times, when the *aadA2*-containing GC was located in the second position (Collis & Hall, 1995). However, in our result, it was shown that the insertion of a promoter-containing GC in the first position did not decrease the *gusA* expression significantly (considered as the expression of the gene in the second GC), comparing between TMB4 P_C and TMB4 P_C +GC (Figure 4-19). Therefore, we hypothesised that promoter-containing GCs could act as a genetic clutch, where the expression of the original first GC is being disengaged from the P_C promoter and replaced by the one on the promoter GC (Figure 4-25). This can compensate and prevent a significant change in the expression of the original first GC while a new one is sampled from the pool of GCs within the cell in order to adapt to a stress which occurs at the same time as the stress which selected for the previous first GC. Failure to disengage the original first GC from the P_C whilst inserting a new one would lead to a reduction in expression and would leave the cell potentially at a disadvantage. A genetic clutch could occur with the insertion of any GCs containing promoters in the sense strand to provide an additional promoter for the expression of the downstream GC, so it could be the insertion of either

noncoding GCs like TMB4 GC, or other promoter-containing GCs such as TA-containing GCs.

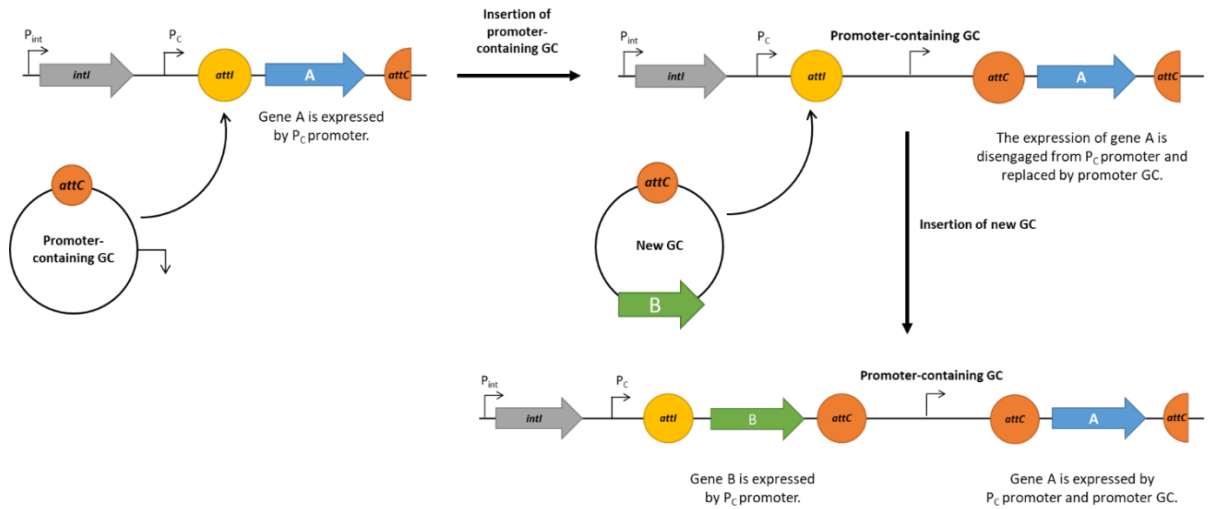


Figure 4-25 The proposed function of promoter-containing GC as a genetic clutch. When a promoter-containing GC inserts into the first position, it can act as a genetic clutch by disengaging the original first GC (A) from P_C promoter and replaced with the one on promoter GC. When a new GC (B) inserts, it can be expressed by P_C promoter, while A is expressed by promoter-containing GC and P_C promoter.

A genetic clutch can be of benefit to bacteria in the situation where they are exposed to multiple environmental stresses at once such as being exposed to two different antibiotics at the same time. The first resistance gene (blue ORF in Figure 4-26A) can be expressed by the P_C promoter. When the promoter-containing GC is inserted (Figure 4-26B), it will increase the integron integrase expression, catalysed the insertion of the new GC. The new GC containing resistance gene (the green ORF in Figure 4-26C) will then be inserted and expressed by P_C promoter. At the same time, the blue GC, now located in the third position, will be expressed by P_C promoter and the promoter GC. Therefore, both resistance GCs will have high expression level, allowing

bacteria to survive in both drugs. Without a promoter-containing GC located upstream from the blue ORF, cells could lose an ability to survive against the first antibiotic as there will be a significant decrease in the expression of the blue ORF after the insertion of the second resistance GC in the first position (Figure 4-26).

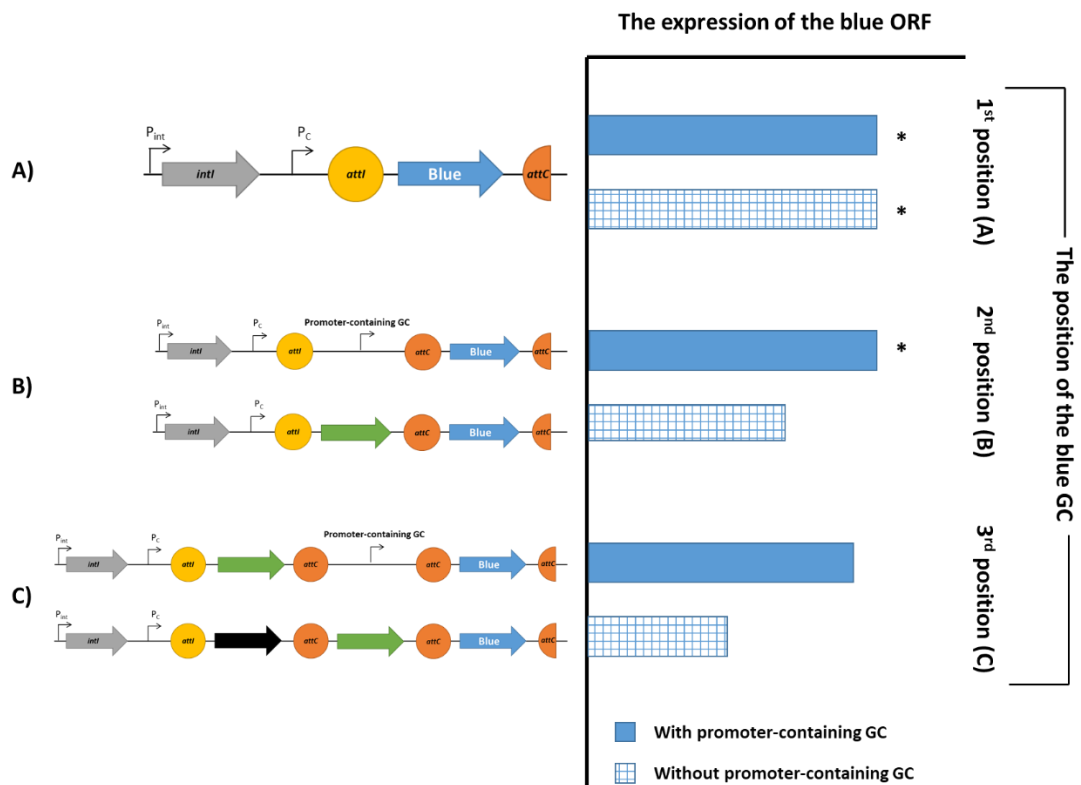


Figure 4-26 The expression level of gene cassettes with and without a genetic clutch. The estimated level of expression of the blue ORF in A.) the first, B.) the second and C.) the third position was shown in the bar chart. The solid bars represent the situation when promoter-containing GC was inserted upstream of the blue GC, while the gridded bars represent the situation when no promoter-containing GC was inserted. The expression of the blue ORF should decrease when more GCs are inserted without the presence of a promoter-containing GC as a genetic clutch. The asterisks indicate the experimentally verified expression level, suggested by the results in Figure 4-19 (TMB4 P_C and TMB4 P_C+GC).

As the noncoding GC MMU23 showed no promoter activity in the enzymatic assay, it suggested that not all noncoding GC have a function as a promoter and might have other functions in integrons. The MMU23 GC could also carry a promoter that can be recognised in its native host but not in *E. coli*, or promoter associated with other sigma factors like P_{Fla} promoter that did not show promoter activity in our enzymatic assay. For the ORF-containing GC MMB3, the promoter activity was found on the sense strand. This GC was predicted to contain toxin-antitoxin (TA) ORFs, the PIN toxin and ribbon-helix-helix antitoxin, which were previously suggested to contain their own promoter sequences. Sample SSU17 and MMU23 are also be considered as a control to justify that not all of GCs amplified from the oral metagenome exhibited promoter activity.

Another possible application of the pCC1BAC-*lacZ* α -*gusA* plasmid, constructed for the β -glucuronidase assay, was the detection of promoter activity in either direction of GCs. The clones with promoters on the sense strand can be detected under UV light and showed blue fluorescence because β -glucuronidase can cleave the substrate, MUG, on the plate, which produces a fluorescent compound called methylumbelliferone. Promoters on the antisense strand can be detected by blue-white screening as β -galactosidase can cleave X-Gal, producing an intensely blue product, 5,5'-dibromo-4,4'-dichloro-indigo which can be viewed by eye under normal light.

To verify the application of pCC1BAC-*lacZ* α -*gusA* plasmids as promoter detection system, integron GCs were amplified from the human oral metagenome by using SUPA3-SUPA4 and MARS2-MARS5 primers, as they

successfully amplified integron GCs from the oral metagenome previously (Tansirichaiya *et al.*, 2016b). After cloning the amplified GCs between both reporter genes, two main groups of GCs were identified with promoter activities, which were noncoding GCs and TA-containing GCs. By detecting 7 clones containing noncoding GCs with promoter activity by this approach, it further supported that one of the functions of noncoding GCs in integrons is to provide promoter activities within an array of GCs.

TA-containing GCs were previously shown to be abundant in CIs. Their functions were suggested to prevent the random deletion of GCs and to stabilise the large arrays of GC in CIs (Rowe-Magnus *et al.*, 2003, Szekeres *et al.*, 2007, Guerout *et al.*, 2013). TA systems normally encode a stable toxin and a labile antitoxin (Van Melderen & Saavedra De Bast, 2009). Consequently, these TA cassettes have to carry their own promoters to ensure their expression and subsequent selection. Most of the GCs amplified with our primers were homologous with *Treponema* spp. The integron from *T. denticola* was previously described as CI, which contained the HicA-HicB TA-containing GC in the fourth position within the GC array (Accession number NC_002967) (Coleman *et al.*, 2004). Therefore, TA-containing GCs should be present in our oral metagenome, which were detected by our pBiDiPD.

Two of the identified GCs, SSU-Pro-9 and MMU-Pro-18 (section 4.3.8.2), were similar to the MMB3 and MMB37 GCs, respectively, in which their promoter activity on the sense strand was shown by the β -glucuronidase enzymatic assay (section 4.3.6). The phenotypes of SSU-Pro-9 and MMU-Pro-18 colonies showed only a blue fluorescent phenotype, reflecting the promoter

activity on the sense strand. This was in correspondence with the enzymatic assay results of MMB3 and MMB37, which also showed the promoter activity only on the sense strand. Thus, it justified that our pCC1BAC-*lacZ* α -*gusA* plasmid can be used as a detection system for the promoter-containing integron GCs, called Bi-Directional Promoter Detection (BiDiPD).

BiDiPD could be improved in several aspects. During the screening, some samples contained incomplete GCs and chimeric inserts, generated by the internal restriction sites within the GCs. Therefore, different combinations of restriction enzymes could be used to overcome this problem and extend the number of samples. The *lacZ* α on pBiDiPD could be changed into the *lacZ* gene, which would allow measurement of the promoter activity on the antisense strand by β -galactosidase enzyme assay. It will be useful as we will be able to determine the promoter activity on either strand of DNA directly from the detected clones without an additional cloning of the inserts. The dual reporter genes (*lacZ* α -*gusA*) could be changed for two different fluorescent proteins, such as GFP and mCherry (red fluorescent protein), so fluorescence-activated cell sorting (FACS) could be used to sort cells into four different groups according to the promoter activity rapidly (no activity, activity on sense strand, activity on antisense strand and activity on both strands).

Our BiDiPD construct not only has the potential to detect promoter sequences in integron GCs, but also can be used with the other type of DNAs. For example, it can be used to screen for inducible promoters, similar to a metagenomic approach called substrate-induced gene expression (SIGEX). SIGEX has been used to identify novel catabolic operons by detecting

promoters induced in response to a particular catabolite, as the expression of catabolic genes is usually induced by relevant substrates or metabolites (Uchiyama *et al.*, 2005). Genes involved in aromatic degradation were previously identified by using SIGEX such as benzoate degradative genes and salicylate oxygenase genes (Uchiyama & Miyazaki, 2010, Meier *et al.*, 2016). The inducible promoters are screened by using FACS to screen a metagenomic library for clones with the expression of a downstream fluorescent reporter gene on an operon-trap vector only in the presence of a particular catabolite (Figure 4-27). This approach allows a rapid and economical screening of promoters from the metagenome; however, it can screen only the promoters in the same direction with the fluorescent protein and also prone to have false positive/negative clones if the gate setting in FACS is not properly optimised.

Figure 4-27 Substrate-induced gene expression (SIGEX) screening procedure.

Clones containing self-ligated vectors and constitutive promoters are first excluded by sorting for non-fluorescent cells in the absence of the substrate. The non-fluorescent cells are then induced with the substrate and clones with the expression of fluorescent proteins are then isolated by FACS, followed by characterisation for inducible promoters and the genes downstream. (Retrieved from Uchiyama & Miyazaki, 2010)

A metagenomic library can be constructed on pBiDiPD, followed by screening for clones exhibiting promoter activity only in the presence of a particular catabolite or antibiotic on agar plates and subsequently characterised the genes downstream from the detected promoter. The advantage of our system is that we can screen for promoters on either strand of DNA. It can also be used to screen for promoters and operons, which are repressed in response to catabolites, by screening clones with promoter activity for a loss of

phenotype (change from blue to white colony or no fluorescence under UV) in the presence of the substrate on agar plates.

4.5 Conclusion

In conclusion, the promoter activities of the noncoding integron GCs were experimentally demonstrated by using a robust β -glucuronidase enzyme assay, confirming that one of the functions of noncoding GCs is to provide promoters for the expression of GCs, in addition to the expression from P_C promoter. The plasmid system, called pBiDiPD, was also developed for the direct visualisation of clones containing gene cassettes with promoter activity on agar plates. This system also can be applied as a detection system for promoter activity for other DNA fragments.

Chapter 5

Culture-independent Recovery of Composite Transposons and Translocatable Units from the Human Oral Metagenomic DNA

5.1 Introduction

Composite transposons are often associated with ARGs and responsible for the spreading of these genes. For example, composite transposon Tn4001 confers kanamycin resistance in different bacteria such as *Enterococcus*, *Staphylococcus* and *Streptococcus* (Lyon *et al.*, 1984, Chow *et al.*, 2007, Leelaporn *et al.*, 2008). Recently, another mode of transposition of ARGs from composite transposons has been reported, which can occur through the excision of ARGs and one of the IS elements from composite transposons, forming as a circular MGEs, called translocatable units (TUs) (Harmer *et al.*, 2014) (Figure 5-1).

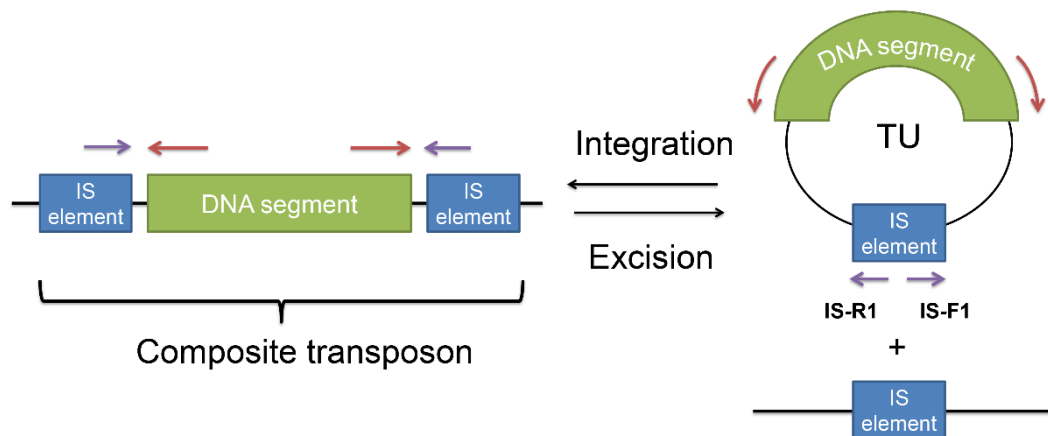


Figure 5-1 The structures of composite transposons and TUs. The composite transposons consist of two IS elements (blue box) flanking DNA segment (green box). The TU circular molecule can be excised from the composite transposon. The purple and red arrows represent the binding site of the DNA primers for the amplification of composite transposon and TUs, respectively. (Retrieved from Tansirichaiya *et al.*, 2016a)

Both elements have been identified from the human oral microbiome. Composite transposons are commonly found in both Gram-positive and Gram-negative bacteria, such as Tn4001 and Tn4003 from *Staphylococcus aureus*, and Tn1547 from *Enterococcus faecalis* (Figure 5-2A, B and C) (Lyon *et al.*, 1984, Rouch *et al.*, 1989, Quintiliani & Courvalin, 1996). Several TUs were reported in oral bacteria, including IS1216-*tet*(S) and IS1216-*qrg* TUs isolated from *Streptococcus infantis* and *S. oralis*, respectively (Figure 5-2D and E) (Ciric *et al.*, 2011, Ciric *et al.*, 2014). However, most of the studies on composite transposons and TUs focused on culturable bacteria, which are often a minority within microbial communities as discussed in Chapter 1.

Figure 5-2 The composite transposons and translocatable units identified from the oral bacteria. A.) Tn4001 B.) Tn4003 C.) Tn1547 D.) IS26-*tet*(S) TU E.) IS26-*qrg* TU (Adapted from Rouch *et al.*, 1989, Quintiliani & Courvalin, 1996, Prudhomme *et al.*, 2002b, Ciric *et al.*, 2011, Ciric *et al.*, 2014)

Therefore, in this study, we detected composite transposons and TUs in metagenomic DNA by using PCR-based techniques. To detect the presence of composite transposons, PCR primers targeting IS elements were designed to amplify DNA segments flanked by IS elements. Then, another set of primers was designed to determine whether the TU forms from each putative composite transposon can be detected in the human oral metagenome.

5.2 Materials and methods

5.2.1 Verification of IS elements in the human oral metagenome

Prior to the screening of composite transposons in the oral metagenome, the presence of insertion sequences within metagenomic DNA was confirmed by performing PCR amplification with various IS element-specific primers, following the protocol in section 2.4.5.

Three different group of IS elements were verified for their presence in the human oral metagenome, including the IS elements found in *Streptococcus* spp. (the most prevalent bacteria in oral cavity (Kreth *et al.*, 2009)), IS6 family (the main IS family associated with the excision of TUs (Harmer *et al.*, 2014)), and IS element commonly associated with transposons and plasmids (Clewell *et al.*, 2014), as listed in Table 5-1.

Table 5-1 The list of IS elements that were selected for the presence in the human oral metagenome.

IS elements found in <i>Streptococcus</i> . (Kreth <i>et al.</i>, 2009)	IS elements in IS6 family (Harmer <i>et al.</i>, 2014)	IS elements commonly found on Tns or plamids (Clewell <i>et al.</i>, 2014)
IS861	IS26	IS3
IS1161	IS240	IS256
IS1167	IS257	IS1485
IS1381		
IS1548		

5.2.2 Screening of composite transposons in the human oral metagenome

The primers amplifying outwards from the detected IS elements were designed for the screening of composite transposons in the oral metagenome. The amplification was performed with two methods: standard PCR (section 2.4.5) and long PCR. The long PCR reaction contained 0.25 μ L of TaKaRa Ex Taq (5 units/ μ L) (Takara Bio, Saint-Germain-en-Laye, France), 5 μ L of 10X Ex Taq buffer, 4 μ L of dNTP mixture (2.5 mM each), 2 μ l of each primer (10 pmol/ μ l), 1 μ l DNA template (50-100 ng), and molecular grade water up to a total volume of 50 μ L. The PCR program for long PCR was similar to the standard PCR except for the extension step which was 10 min, instead of 3 min.

The amplicons were subjected to either PCR purification or gel extraction (Section 2.4.7 and 2.4.8), depending on the profile of amplicons. The amplicons from standard PCR were ligated to linearise pGEM-T easy vector (Promega, Southampton, UK) via TA cloning, and subsequently transformed into *E. coli* α -Select Silver Efficiency competent cells by heat shock, as described in section 2.4.12 and 2.4.15.1, respectively. The white colonies (containing inserts) were subcultured into LB medium supplemented with ampicillin and the plasmids were extracted as described in section 2.4.3. The gel extraction products and plasmids were then sent for sequencing.

The sequences were first analysed with VecScreen analysis tool to remove the vector sequences, and then both primer binding sites on the amplicons were identified. The sequences were then analysed by BlastN, BlastX and ISFinder to search for homologues in the nucleotide, protein and IS element databases, respectively. The sequences of all composite transposons (CTA1 to CTA5) were submitted to the DNA database with the accession numbers from KX305930 to KX305934.

5.2.3 Detection of Translocatable units in the human oral metagenome

As the amplicons identified in section 5.2.2 may have been amplified from a translocatable unit (TU) as well as a linear molecule integrated within a replicon, another set of primers were designed to amplify outward from the DNA segments of each putative composite transposon identified. The TU verification PCR was performed with both Biomix Red (standard PCR, section 2.4.5) and highly processive Q5 polymerase (to confirm that the amplicons

from the standard PCR were not a false positive due to the early fall-off of the standard DNA polymerase). The highly processive PCR reaction was described in section 2.4.16, and the PCR program was similar to the standard protocol, except the extension time that was 5 min instead of 1 min.

5.3 Results

5.3.1 Confirmation of IS elements in oral metagenomic DNA

By performing PCR with the primers targeting all of the selected IS elements, the amplicons of the expected size were sequenced to confirm their presence in the human oral metagenome. Among twelve IS elements, the sequencing results showed that six were present in the extracted oral metagenome, which were including IS26, IS257, IS1216, IS1161, IS1167, and IS1485.

5.3.2 Recovery of composite transposons by PCR amplification

By designing primers to amplify outwards from the detected IS elements, it allows DNA segment carried as either composite transposons or translocatable units to be amplified (Figure 5-1). All six primer pairs were used in the standard and long PCR on the human oral metagenome. The amplicons were either cloned into pGEM-T easy vector or extracted by gel purification and subsequently sequenced. After the screening and sequence analysis, five different putative composite transposons amplicons (CTA1-5) were identified: four with IS1216 primers, and another with IS257 primers, as shown in Figure 5-3 and Table 5-2.

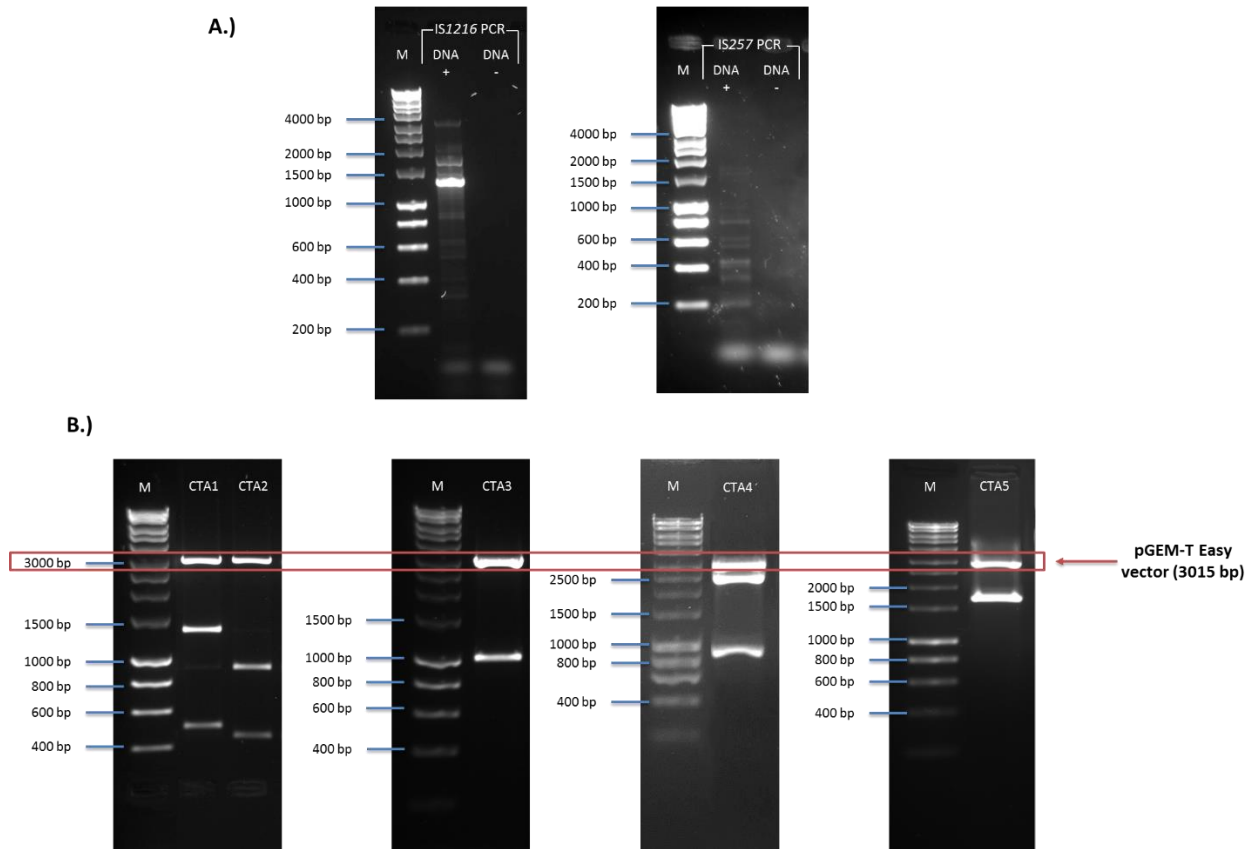


Figure 5-3 IS1216 and IS257 composite transposon amplicons. A.) The amplicons from IS 1216 and IS257 composite transposon PCR B.) The pGEM-T Easy plasmids containing CTA1-5 inserts were digested with *EcoRI* and separated on 1% agarose gel to determine the size of inserts. The red box and arrow indicate the bands of pGEM-T easy vector. Lane M, HyperLadder™ 1kb.

The first amplicon (CTA1) was shown to contain two ORFs, which were predicted to encode a small multidrug resistance protein and a hypothetical protein (similar to NAD⁺ diphosphatase). CTA1 composite transposon was similar to the previously identified IS 1216 composite transposon carried by Tn6087 isolated from *Streptococcus oralis* F.MI.5 with 99% nucleotide identity (Figure 5-4) (Ciric *et al.*, 2011). The second amplicon, CTA2, was similar to the CTA1. The difference between both samples was that the CTA2 insert was 451 nucleotides shorter than the CTA1 (229 bp of the gene encoding a

hypothetical protein and 222 bp of the flanking region between IS 1216 and *hyp*) (Figure 5-4).

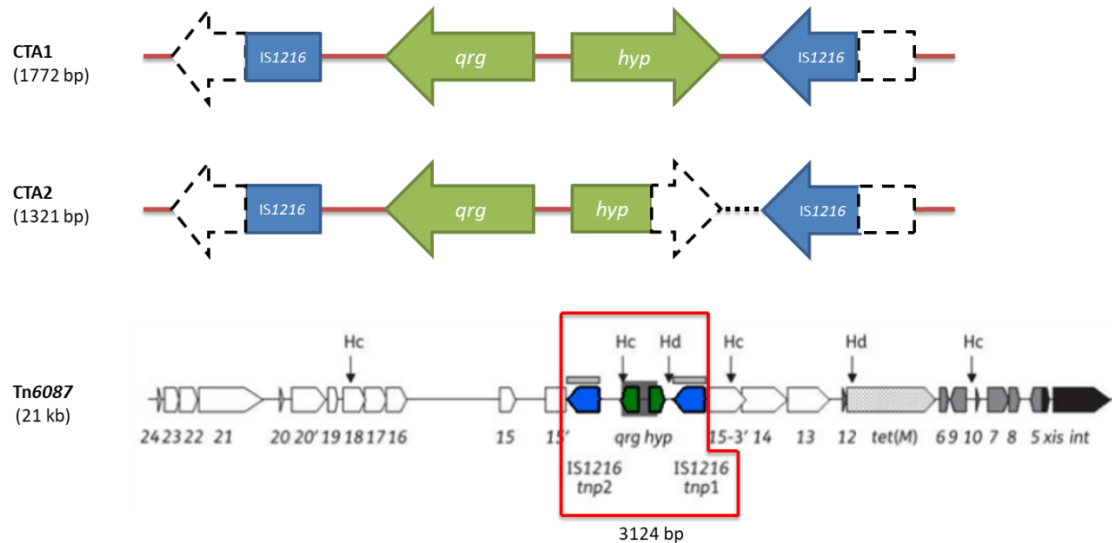


Figure 5-4 Schematic representation of CTA1 and CTA2 structure in comparison to Tn6087. The open arrowed boxes represent ORFs, pointing in the probable direction of transcription. The IS elements and ORF in the DNA segment are shown in blue and green, respectively. The dashed arrow box and dashed line represent truncated gene. The red box indicates the IS 1216 composite transposon on Tn6087. (Tn6087 figure was adapted from Ciric *et al.*, 2011)

The next amplicon (CTA3) was shown to partially match with a part of two plasmids: plasmid pIL5 and plasmid pBL1 from *Lactococcus lactis* subsp. *lactis* (Sanchez *et al.*, 2000, Gorecki *et al.*, 2011). The main part of the amplicon (84% of the amplicon) matched to *orf 14*, *orf15* and a part of *orf 16* from plasmid pIL5 (98% nucleotide identity), encoding a transposase, a universal stress-like protein, and a Mn²⁺/Fe²⁺ transporter-like protein, respectively (Figure 5-5). Another part of the amplicon was matched to the ISS1-like element, located on a pBL1 plasmid, with 100% nucleotide identity. By analysing the sequences with ISFinder, it was shown that both *orf14* and

ISS1-like element had 100% nucleotide similarity to IS1216. The CTA4 structure was similar to CTA3, but contained an additional 2329 bp. The extra nucleotides include the rest of *orf16* missing from CTA3 and a transposase gene *orf17* from plasmid pIL5 (Figure 5-5).

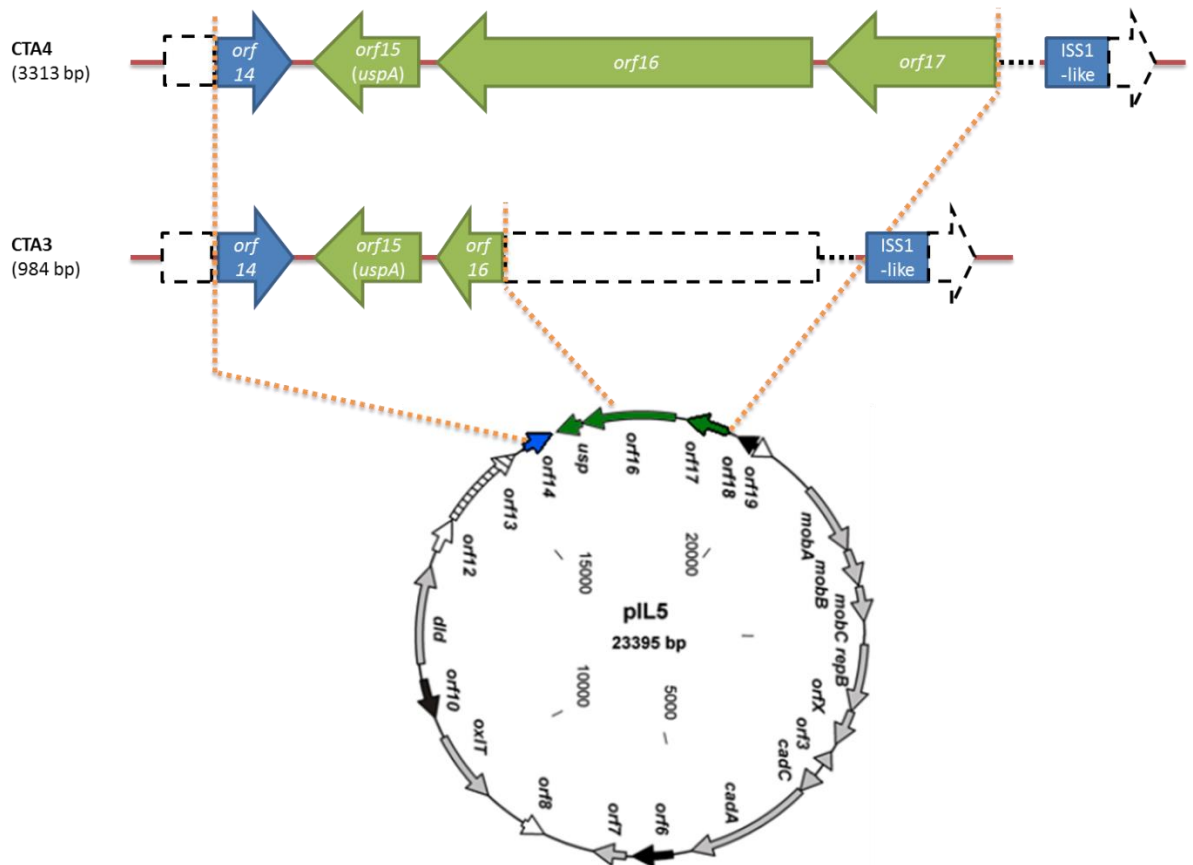


Figure 5-5 Schematic representation of plasmid pIL5 and CTA3-4 structures.

The open arrowed boxes represent ORFs, pointing in the probable direction of transcription. The IS elements and ORF in the DNA segment are shown in blue and green, respectively. The dashed arrow box represents truncated gene. The orange dotted lines indicate the regions on pIL5 that matched to the CTA3 and 4 structures. (The pIL5 figure was adapted from Gorecki *et al.*, 2011)

The last structure, CTA5, was the only amplicon amplified with IS257-based primers. It contained two ORFs predicted to encode a kanamycin resistance

Table 5-2 The details of putative composite transposon structures amplifying from the human oral metagenomic DNA.

Sample (Accession number)	BlastN/ BlastX	Query	Gene/Protein	Position	Identities	Coverage (%)	Accession number of BLASTn/BLASTx matched	Predicted Structure (Size)*
CTA1 (KX305930)	BLASTn (<i>S. oralis</i> Tn6087)	1-1772	IS1216 tnp1- <i>qrg</i> - <i>hyp</i> IS1216 tnp2	9771-11542	1755/1772 (99%)	1772/1772 (100%)	HQ663849.2	<p>(1772 bp)</p>
	BLASTx	287-686	Hypothetical protein (NAD ⁺ diphosphatase) [<i>Streptococcus</i>]	1-132	131/132 (99%)	132/132 (100%)	WP_009732039.1	
		858-1184	QacE family quaternary ammonium compound efflux SMR transporter [<i>Streptococcus</i>]	1-108	108/108 (100%)	108/108 (100%)	WP_009732038.1	
CTA2 (KX305931)	BLASTn (<i>S. oralis</i> Tn6087)	118-1179	<i>qrg</i> - <i>hyp</i>	9965-11026	1049/1062 (99%)	1321/1321 (100%)	HQ663849.2	<p>(1321 bp)</p>
		1-120	IS1216 tnp1	11423-11542	120/120 (100%)			
		1172-1321	IS1216 tnp2	9771-9921	148/151 (98%)			

Sample (Accession number)	BlastN/ BlastX	Query	Gene/Protein	Position	Identities	Coverage (%)	Accession number of BLASTn/BLASTx matched	Predicted Structure (Size)*
	BLASTx	102-287	Hypothetical protein (NAD ⁺ diphosphatase) [<i>Streptococcus</i>]	1-57	55/57 (96%)	57/132 (43.2%)	WP_009732039.1	
		460-786	QacE family quaternary ammonium compound efflux SMR transporter [<i>Streptococcus</i>]	1-108	108/108 (100%)	108/108 (100%)	WP_009732038.1	
CTA3 (KX305932)	BLASTn (<i>Lactococcus lactis</i> plasmid pIL5)	1-842	<i>orf14 (tnp)-orf15 (UspA)-orf16 (Mn²⁺/Fe²⁺ transporter)</i>	15920-16767	830/848 (98%)	848/984 (86.2%)	HM021330.1	<p>Query 1 ————— 842 839 ————— 1073</p> <p>(984 bp)</p>
	BLASTn (<i>Lactococcus lactis</i> plasmid BL1)	839-984	ISS1-like element	1598-1743	146/146 (100%)	146/984 (14.8%)	AF242367.1	
	BLASTx	167-604	Universal stress protein UspA [<i>Lactococcus lactis</i>]	1-145	144/145 (99%)	145/145 (100%)	WP_047687110.1	

Sample (Accession number)	BlastN/ BlastX	Query	Gene/Protein	Position	Identities	Coverage (%)	Accession number of BLASTn/BLASTx matched	Predicted Structure (Size)*
		591-860	Manganese transporter [<i>Lactococcus lactis</i>]	443-525	74/83 (89%)	83/525 (15.8%)	WP_011669120.1	
CTA4 (KX305933)	BLASTn (<i>Lactococcus lactis</i> plasmid pIL5)	1-3170	<i>orf14</i> (<i>tnp</i>)- <i>orf15</i> (<i>UspA</i>)- <i>orf16</i> (<i>Mn2+</i> / <i>Fe2+</i> transporter)- <i>orf17</i> (<i>tnp</i>)	15920-19096	3121/3178 (98%)	3178/3313 (96%)	HM021330.1	<p style="text-align: center;">(3313 bp)</p>
	BLASTn (<i>Lactococcus lactis</i> plasmid pBL1)	3167-3311	ISS1-like element	1598-1742	145/145 (100%)	145/3313 (4.4%)	AF242367.1	
	BLASTx	167-604	Universal stress protein <i>UspA</i> [<i>Lactococcus lactis</i>]	1-145	142/145 (98%)	145/145 (100%)	WP_047687110.1	
		591-2216	Manganese transporter [<i>Lactococcus lactis</i>]	1-535	521/535 (99%)	525/525 (100%)	WP_031297139.1	
		2382-3188	Transposase [<i>Lactococcus lactis</i>]	4-272	259/269 (96%)	269/273 (98.5%)	WP_063283651.1	

Sample (Accession number)	BlastN/ BlastX	Query	Gene/Protein	Position	Identities	Coverage (%)	Accession number of BLASTn/BLASTx matched	Predicted Structure (Size)*
CTA5 (KX305934)	BLASTn [<i>Staphylococcus aureus</i> plasmid SAP079A]	1-110	<i>tnpB</i>	26358-26468	110/111 (99%)	1658/1658 (100%)	GQ900432.1	<p style="text-align: center;">(1658 bp)</p>
		108-1658	<i>tnpB</i> -kanamycin resistance <i>knt</i> -Truncated <i>rep</i> -transposase	26366-27913	1545/1551 (99%)		GQ900432.1	
	BLASTx	483-1247	Kanamycin nucleotidyltransferase [<i>Staphylococcus aureus</i>]	1-254	255/257 (99%)	257/257 (100%)	EZV91554.1	
		1413-1559	Replication protein, partial [<i>Staphylococcus aureus</i>]	1-55	55/55 (100%)	55/55 (100%)	EVZ23238.1	

* The open arrowed boxes represent ORFs, pointing in the probable direction of transcription. The IS elements and the other ORFs in the composite transposons are shown in blue and green, respectively. The dash boxes, arrow boxes, and dotted lines represent the regions that are not present compared to the sequences in the database. (Retrieved from Tansirichaiya et al., 2016a)

5.3.3 Amplification of putative TU structures

As the amplicons identified from composite transposon PCR in section 5.3.2 may have been amplified from a TU template, another set of primers were designed to amplify in the outward direction from the DNA segments toward the IS elements (Figure 5-1). If the TU form can be found in the human oral metagenome, the amplicon will contain the flanking region and one IS element. The PCR results showed that the amplicons with the expected size containing IS1216 were identified from the primers targeting the TU form of CTA2 and CTA4 structures (Figure 5-7), confirming that TU form of both structures could be found in the human oral metagenome.

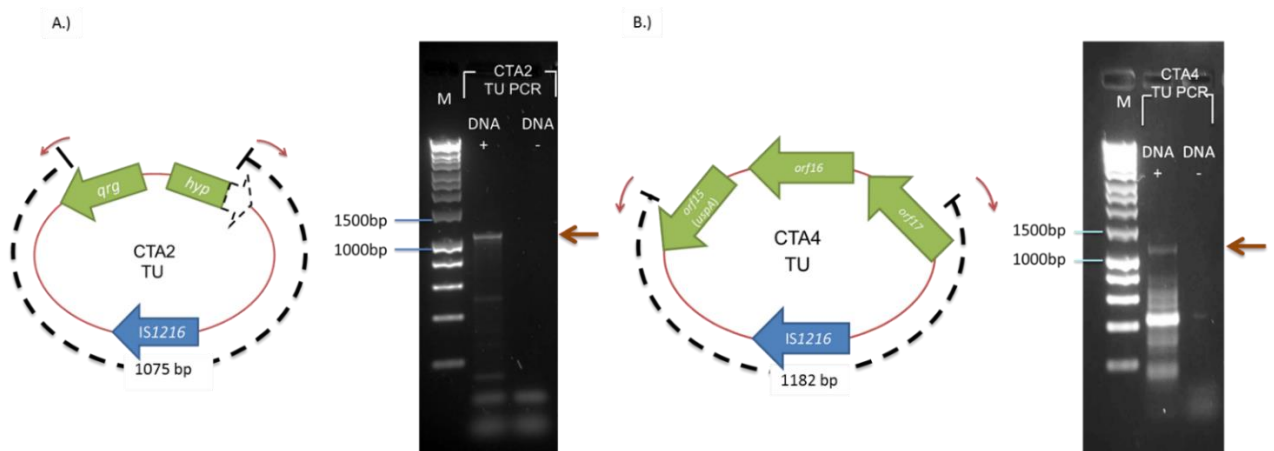


Figure 5-7 TU confirmation PCR of the sample (A) CTA2 and (B) CTA4. The confirmation of TU was done by performing PCR in the outward direction from the DNA segments toward the IS elements. If the TU form is present, the amplicon will contain IS element. Lane M, HyperLadder 1 kb (Bioline, UK). The open arrowed boxes represent ORFs, pointing in the probable direction of transcription. The IS elements and ORFs in the DNA segment are shown in blue and green, respectively. The dashed arrow box represents truncated gene. The red arrows represent the binding site of the TU

verification primers. The brown arrows indicate the expected PCR products of the TU verification PCR. The black dashed lines indicated the expected size of the amplicons. (Retrieved from Tansirichaiya *et al.*, 2016a)

The TU verification PCR was first carried out with standard PCR using Biomix Red, which can amplify up to 5 kb DNA. However, there was a chance that the amplicons might occur by the early fall-off of the DNA polymerase, leaving partial amplicons that could act as primers in a subsequent round of PCR. The verification PCR was, therefore, repeated with highly processive Q5 high-fidelity DNA polymerase (amplifying up to 20 kb). The amplicons containing IS 1216 still could be identified from the highly processive PCR. The lack of positive results for the TU PCR of CTA1, CTA3 and CTA5 also further supported that the PCR results were unlikely to be artefacts.

5.4 Discussion

By designing the primers amplifying outward from IS elements, five different putative composite transposons were identified from a sample of human oral metagenomic DNA. This strategy is similar to the approach used in the detection of oral integron GCs in Chapter 3, which were amplified by using primers targeting *attC* sites. As the primers were designed to target the IS elements, it also has the potential to recover novel genes carried by composite transposons.

All of the putative composite transposons were predicted to encode proteins involved in environmental adaptation, including those conferring antimicrobial resistance. CTA1 and CTA2 contained *qrg*, encoding a small multidrug resistance

efflux protein which confers antiseptic resistance to cetyltrimethylammonium bromide (CTAB) and acriflavine (Ciric *et al.*, 2011). Kanamycin resistance gene *knt* was found on CTA5 sample, encoding kanamycin nucleotidyltransferase protein. This protein can inactivate kanamycin by transferring a nucleotidyl group to the 4'-hydroxyl group of the aminoglycoside antibiotic (Goffic *et al.*, 1976, Pedersen *et al.*, 1995).

CTA3 and CTA4 contained the *uspA* gene, encoding a universal stress protein. The precise biological function of UspA remains unknown. It has been reported that the amount of UspA increased upon exposure to various stress conditions including nutrient starvation, oxidant exposure, heat and antibiotic exposure (polymixins and cycloserine) (Nyström & Neidhardt, 1994, Kvint *et al.*, 2003, Nachin *et al.*, 2005). Cells with mutated UspA were shown to die prematurely during growth-arrested state (Nyström & Neidhardt, 1994). UspA is a serine and threonine phosphoprotein, which is phosphorylated in response to stasis such as during glucose deprivation, so it was hypothesised that UspA regulates the activity of certain proteins through kinase activities (Freestone *et al.*, 1997). An increased in UspA expression was also shown to result in changes in the pattern of protein synthesis (Nyström & Neidhardt, 1996). The study in *E. coli* showed that UspA reprogrammed cells towards defence and escape by decreased the cell growth rate, increasing the cell's capacity to withstand stresses and modulating activities related to motility and adhesion (Nyström & Neidhardt, 1996, Nachin *et al.*, 2005).

To determine the presence of TU forms of each detected composite transposons, PCR with another set of primers based on each amplicon was performed. It was confirmed that CTA2 and CTA4 were likely to be present as small circular molecules as the amplicons containing a single copy of the expected IS element were found from both samples. This is the first time that TUs were detected in metagenomic DNA and also the first time that the stress adaptation gene *uspA* was found on a TU. Even though the TU amplicon could also have arisen by the repetition of the entire composite transposon in the host genome; it is unlikely because of the inherent instability of large repeated units of mobile DNA (Figure 5-8) (Bzymek & Lovett, 2001).

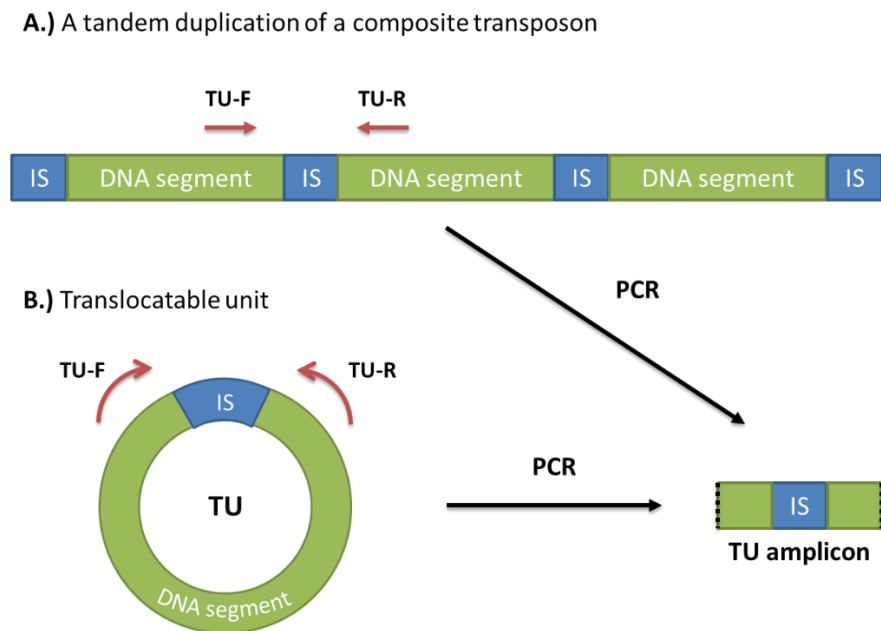


Figure 5-8 The possibilities for the amplification of TU amplicons. A.) A tandem duplication of a composite transposon and B.) Translocatable unit. The red arrows indicate forward and reverse primers. The IS elements and DNA segments are shown in blue and green, respectively.

The detection of the TUs in a metagenome could be more promising with our approach, compared to the detection by metagenomic sequencing. As these molecules are likely to be rare and comprise a small fraction in metagenomic samples, which could be missed from the sequencing. They could be missed during an assembly, as it is difficult to distinguish between composite transposons and TUs which have identical sequences. Therefore, the reads of TU could be considered and assembled with those more abundant reads of the composite transposon, instead of recognising as a circular structure. Several approaches have been developed to study circular MGEs in metagenomic sequencing. For example, metagenomic samples can be treated with exonuclease to remove chromosomal DNA and followed by multiple displacement amplification to increase the number of DNA sequences (Brown Kav *et al.*, 2013, Jørgensen *et al.*, 2014). However, this method could introduce biases toward specific circular structures during the amplification step (Jørgensen *et al.*, 2014).

The studies on TUs are still in an early stage, and a small number of TUs have been identified. The integration and excision mechanisms of TUs have been studied based on IS26 composite transposons. IS26 is a member of the IS6 family, which is the same family as IS 1216 and IS257, identified in our samples. An intact transposase gene was required for the integration and excision of the TUs (Harmer *et al.*, 2014, Harmer & Hall, 2015). The integration of IS26 TU with a pre-existing IS26 element preferred to occur via a conservative reaction catalysed by the Tnp26 transposase, rather than a replicative transposition to a new site (Figure 5-9) (Harmer *et al.*, 2014). The RecA-dependent homologous

recombination could be used for the integration of a TU as well, but it was 100-fold less efficient than the reaction catalysed by Tnp26 (Harmer & Hall, 2016).

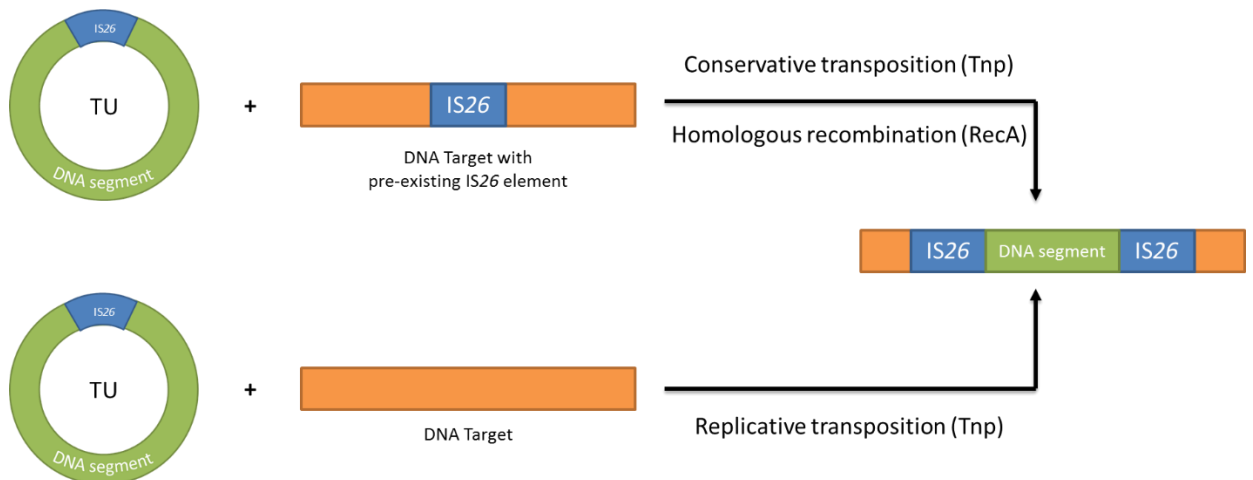


Figure 5-9 The formation of composite transposons by IS26 TU. IS26-composite transposon can be formed by conservative transposition or homologous recombination with a pre-existing IS26 element, and replicative transposition to a new site. Transposition reaction is catalysed by transposase (Tnp), while homologous recombination is catalysed by RecA. The IS26 elements, DNA segments and DNA targets are represented with green, blue and orange boxes, respectively. (Adapted from Harmer & Hall, 2016)

Recently, there is another mobile genetic element similar to TUs, called an unconventional circularizable structure (UCS) (Palmieri *et al.*, 2013). UCSs are non-replicative circular molecules, which excise from a structure that contains two direct repeats (DRs) flanking a DNA segment. After excision, it generates a circular structure containing one of the DRs and the DNA segment. TUs and UCSs can both transpose in RecA-deficient bacteria, suggesting that homologous recombination is not major mechanisms for their movement (Azpiroz

et al., 2011, Harmer *et al.*, 2014). The major difference between both MGEs is that UCS does not carry a recombinase gene to catalyse their integration (Palmieri *et al.*, 2013). With their similarity in structures, the PCR strategy used in this study could have a potential to be applied to UCSs by designing the primers based on the DRs of the UCSs.

As the IS6-family composite transposons can be transposed either as a whole unit of composite transposons or as TUs, their associated resistance genes could have more chance to be spread in bacterial population, comparing to the composite transposons containing other IS families (Harmer & Hall, 2016). Moreover, their transfer can be facilitated by other MGEs because composite transposons are often located on other MGEs, such as *tet(S)*-containing IS1216 composite transposon located on plasmid pSI01 and *tetC*-containing IS26 composite transposon located on transposon Tn1404 (Figure 5-10) (Schnabel & Jones, 1999, Ciric *et al.*, 2014).

Figure 5-10 IS6-family composite transposons located on other MGEs. a.) *tet(S)*-containing IS1216 composite transposon located on plasmid pSI01 and b.) *tetC*-containing IS26 composite transposon located on transposon Tn1404. (Retrieved from Schnabel & Jones, 1999, Ciric *et al.*, 2014)

In a situation, similar to the circular integron gene cassettes, where the excised TUs do not integrate into a replicon, they will likely be lost from bacterial population during cell division. However, as we found a truncated *rep* gene on a putative IS257 composite transposon (CTA5), there is the possibility that TUs can facilitate the movement of *rep* genes between DNA molecules further adding to the complexity of MGE biology. Indeed, there are some composite transposons that could give rise to *rep*-containing TUs such as $\Delta repA-repC$ which is flanked by IS26 on Tn6029 (Figure 5-11A) (Reid *et al.*, 2015). If a *rep*-containing TU integrates with an origin of replication-containing TU, it may result in a plasmid-like structure, which could be replicated and maintained in the bacteria (Figure 5-11B).

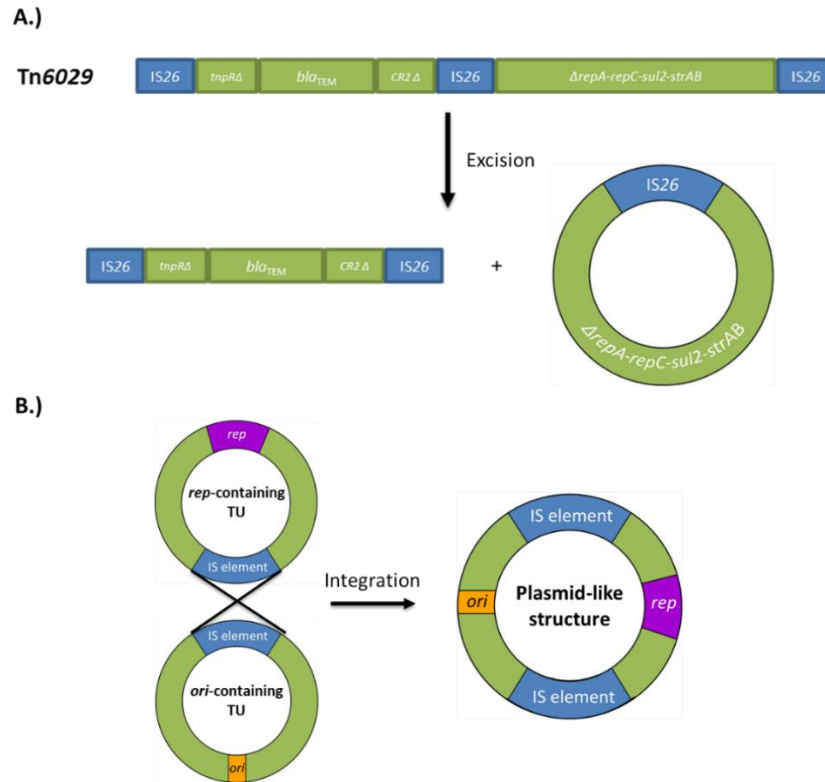


Figure 5-11 The excision of *rep*-containing TU and the formation of plasmid-like structure. A.) The excision of *rep*-containing TU from Tn6029 B.) The formation of plasmid-like structure from the integration of *rep*-containing TU and origin of replication (*ori*)-containing TU. The IS elements, *rep*, *ori* and other genes are shown in blue, purple, orange and green, respectively.

5.5 Conclusion

This is the first study showing the detection of composite transposons and TUs from the oral metagenome using PCR amplification with DNA primers based on the IS elements. This method also has the potential to amplify novel genes carried by those composite transposons. These results have been published in FEMS Microbiology Letters (Appendix 5) (Tansirichaiya *et al.*, 2016a).

Chapter 6

The Isolation of DNA from the Human Oral Metagenomic DNA Using Entrapment Vectors

6.1 Introduction

6.1.1 Transposable elements (TEs)

Transposable elements (TEs) are DNA structures which can move from one location in a genome to another. Insertion sequences and transposons are important TEs in the bacterial population, as their movement increase variability in bacterial genomes leading to evolutionary and adaptive abilities of their hosts. Both IS elements and Tns are also frequently located on other MGEs such as plasmids, which can facilitate their inter-species transfer. The analysis of functional TEs could identify novel elements and their accessory genes, and also increase our understanding of the genome dynamics of bacteria. For example, an ampicillin-resistant *Enterococcus faecium* D344R was converted to ampicillin-susceptible strain D344S due to an interaction between Tn916 (18 kb) and Tn5386 (29 kb), which shared 15 similar ORFs (Figure 6-1) (Rice *et al.*, 2005). The strand exchange between both elements resulted in a 178-kb deletion from the bacterial genome, including *pbp5* gene (conferring ampicillin resistance).

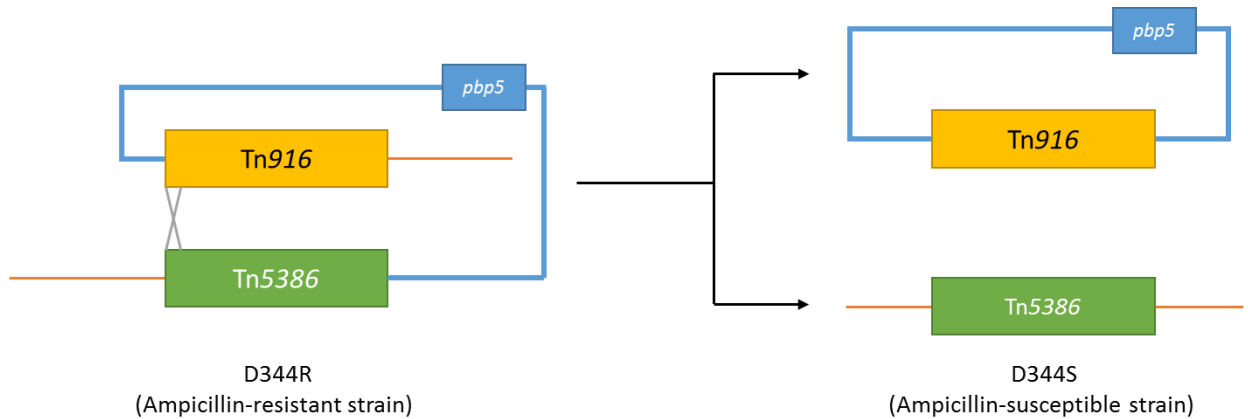


Figure 6-1 Excision of the large genomic DNA from *Enterococcus faecium* D344R due to an interaction between Tn916 and Tn5386. The strand exchange (indicated with grey cross) between Tn916 (yellow boxes) and Tn5386 (green boxes) results in the generation of two DNA products: A.) circular DNA composed of genomic DNA (represented in blue) and Tn916 and B.) a regenerated chromosomal region that contains Tn5386 linked to the region that flanked Tn916 in D344R. (Adapted from Rice *et al.*, 2005)

TEs could be identified through the phenotypic changes of the hosts, conferred by the accessory genes on TEs or by the mutations caused by TE transposition. They could also be identified by chance such as TEs located near genes of interest. Sequence analysis of bacterial genomes is another approach to identify TE, by searching and comparing the sequences to known transposase genes or terminal inverted repeats (IRs) from the nucleotide and protein databases. However, the sequence-based analysis cannot demonstrate the transposition activities of putative TEs.

6.1.2 Entrapment vectors

Using an entrapment vector is another approach to detect and isolate TEs with an active transposition activity. They can isolate TEs by having a gene system that will result in different phenotypes for selection when the TEs transpose into it. They have been shown to be effective in the isolation of TEs within single bacteria species (Solyga & Bartosik, 2004). They can also determine the frequency of transposition of TEs in bacteria. For example, Tn5393 was detected in *Paracoccus pantotrophus* LMD 82.5 by using pMEC1 entrapment vector (Bartosik *et al.*, 2003). By calculating the ratio between the number of the positive clones obtained on selective medium and the total number of bacteria, it also showed that Tn5393 had a very high transposition frequency (10^{-3}) in this bacterial strain.

The appropriate types of capturing systems and selective marker genes mainly depend on the bacterial species of interest. There are three main types of entrapment vectors. The first type contains conditionally lethal genes. Their TE detection is based on disruption of lethal genes by the insertion of a TE, which enabled growth of the mutant clones. For example, pUCD800 contained a *sacB* gene, encoding levansucrase (Figure 6-2A) (Gay *et al.*, 1985). This enzyme catalyses the breakdown of sucrose and results in the formation of levan, which is toxic to Gram-negative bacteria (Reyrat *et al.*, 1998). Therefore, only the clones with *sacB* mutation will be able to grow in the presence of sucrose, which can be subsequently characterised for TE insertion.

The next type contains antibiotic selection cartridges, composing of a silent antibiotic resistance gene, whose expression is under the control of P_R promoter, and the *cl* gene, which is constitutively expressed by the P_{RM} promoter and encoding for a λ repressor (as shown in Figure 6-2D) (Woody *et al.*, 1993). The expression of antibiotic resistance gene is completely blocked as the λ repressor inhibits a transcription from the P_R promoter. The entrapment vector pGBG1, for example, contained a silent tetracycline resistance gene *tetA* and the *cl* gene (Figure 6-2B) (Schneider *et al.*, 2000). The clones with a disruption of *cl* (i.e. by insertion of TEs) can be positively selected in the presence of an antibiotic, as their resistance gene is expressed as a result of the derepression of the P_R promoter (Figure 6-2D).

Finally, entrapment vectors can carry promoter-less selective genes. As many TEs contain outward directed promoters, the transposition of such TEs upstream of promoter-less genes will activate their expression and result in a selectable phenotype. For example, a pAW1326 vector carried two promoter-less resistance genes, conferring resistance to kanamycin and chloramphenicol (Figure 6-2C) (Szeverenyi *et al.*, 1996).

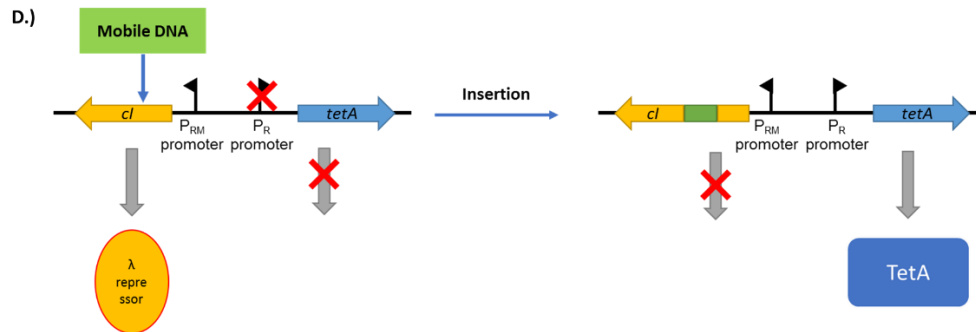


Figure 6-2 The types of entrapment vectors. There are three main types of entrapment vectors, which are A.) Entrapment vectors containing conditionally lethal genes (pUCD800) B.) Entrapment vectors containing antibiotic selection cartridges (pGBG1) and C.) Entrapment vectors carrying promoterless genes (pAW1326). For the *cl-tetA* selection cartridge (D.), the insertion of mobile DNA in *cl* gene disrupts the repression of *tetA* expression, allowing *tetA* to be expressed and conferring tetracycline phenotype. The red boxes and the symbols (\blacktriangleright) represent the selection cartridges and promoters. The red crosses represent a repression on promoters and genes. (Adapted from Gay *et al.*, 1985, Szeverenyi *et al.*, 1996, Schneider *et al.*, 2000)

Entrapment vectors have been used to isolate TEs in single bacterial species at a time, which was also limited to only culturable bacteria. Therefore, in this study, we aimed to directly isolate TEs within the human oral metagenomic DNA using entrapment vectors. Two main approaches were used: i.) direct transformation of the oral metagenome into *E. coli* containing entrapment vectors (to detect the transposition between different replicons) and ii.) construction of the oral

metagenomic libraries by cloning DNA into entrapment vectors (to detect the transposition on same replicon).

6.2 Materials and methods

6.2.1 Construction of BACpAK entrapment vectors

The *cl-tetA* selection cartridge was amplified from pAK1 entrapment vector, received from Prof. Dariusz Bartosik and Dr Magdalena Szuplewska (University of Warsaw, Poland), by using *cl-tetA*(F)-XhoI and *cl-tetA*(R)-XhoI primers (Appendix 4). The *cl-tetA* amplicons and pCC1BAC vector were then digested with XhoI restriction enzyme and were ligated together, forming BACpAK plasmid (Figure 6-3). The ligation products were then desalted and introduced into TransforMax™ EPI300™ Electrocompetent *E. coli* by electroporation (as described in section 2.4.13 and 2.4.15.2).

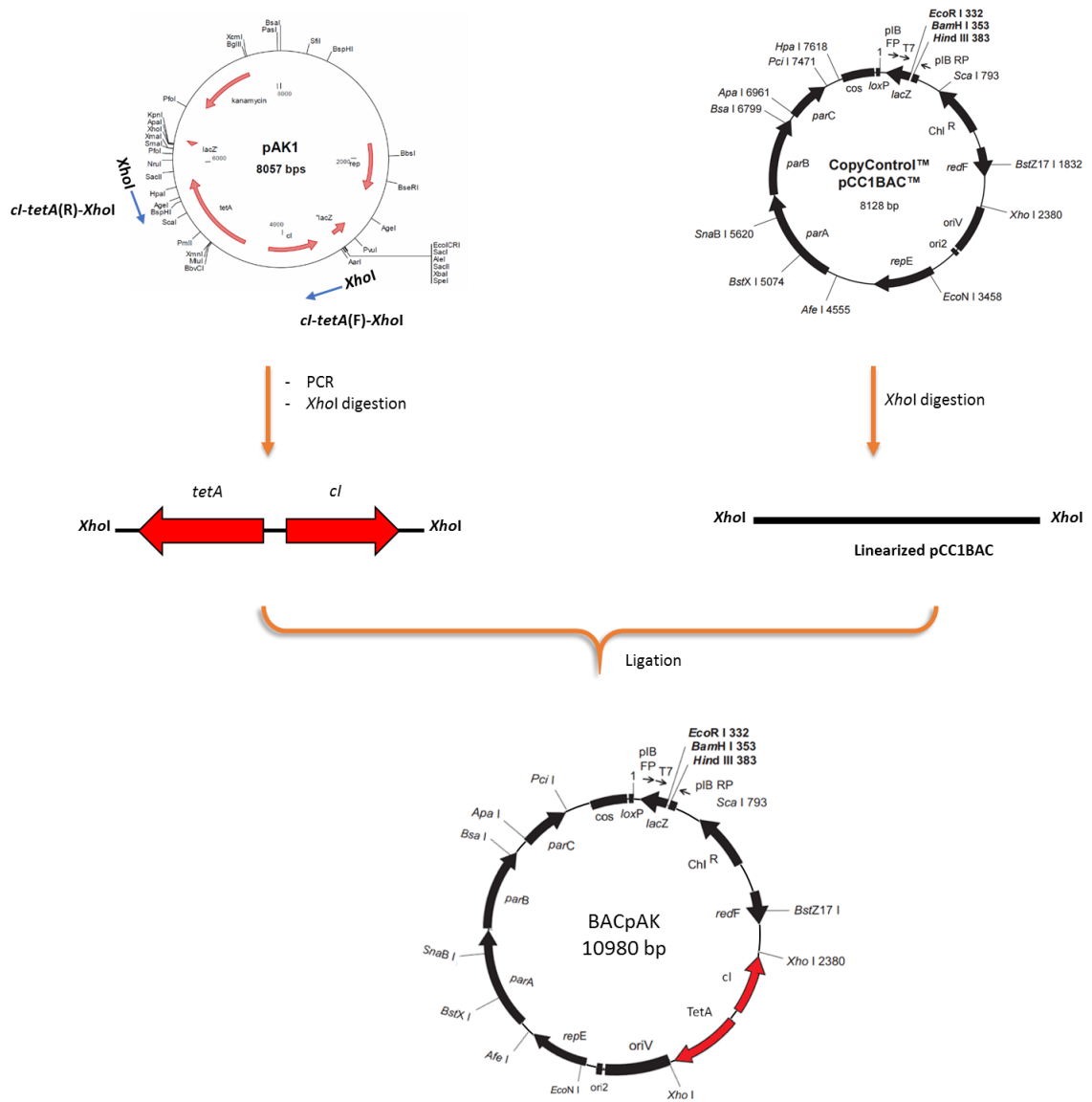


Figure 6-3 Construction of BACpAK entrainment vector. The *cl-tetA* genes (indicated with red open arrow boxes) were amplified from pAK1 entrainment vectors with *cl-tetA(F)-XhoI* and *cl-tetA(R)-XhoI* primers (represented by blue arrows). The *cl-tetA* amplicon was then cloned into pCC1BAC vector at *XhoI* site, forming BACpAK entrainment vector.

6.2.2 Construction of BACpAK-Terminator entrapment vector.

BACpAK-Terminator entrapment vector was constructed by replacing the *cl* gene with *luxI* terminator, described in section 4.2.4.1 (Figure 6-4) (Swartzman *et al.*, 1990). The *cl* gene was removed from BACpAK vector by performing PCR with BACpAK-del-*cl*-F1 and BACpAK-del-*cl*-R1 primers. The amplicons were subjected to site-directed mutagenesis, resulting in BACpAK-del-*cl*. The terminator was added in the forward primers together with *Aat*II restriction site, while *Nhe*I restriction site was added to the reverse primer.

The Tdtro promoter, a promoter found in *Treponema denticola* (Brett *et al.*, 2008), was amplified from pCC1BAC-*lacZ* α -Tdtro-*gusA* plasmid (constructed in Chapter 4) with *gusA*-F8 and Tdtro-R1 primers, containing *Aat*II and *Nhe*I restriction sites, respectively. This promoter was added in front of the terminator by double-digesting the Tdtro amplicons and BACpAK-del-*cl* with *Aat*II and *Nhe*I. Both digestion products were then ligated together, forming BACpAK-Terminator. The ligation product was introduced into TransforMax™ EPI300™ Electrocompetent *E. coli* by electroporation. The Tdtro promoter was included to ensure the expression of *tetA* when there is an insertion of TEs at the terminator.

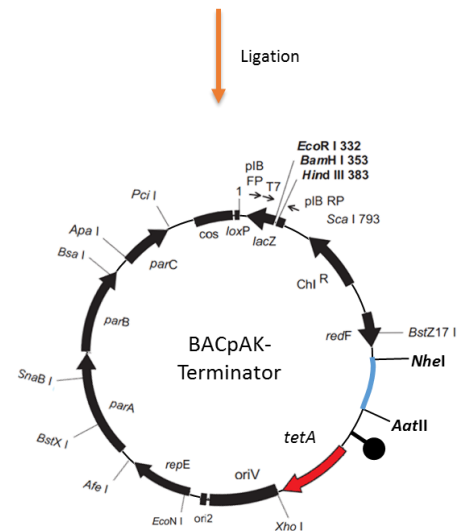
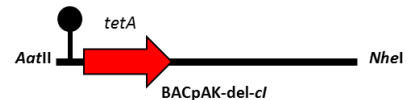
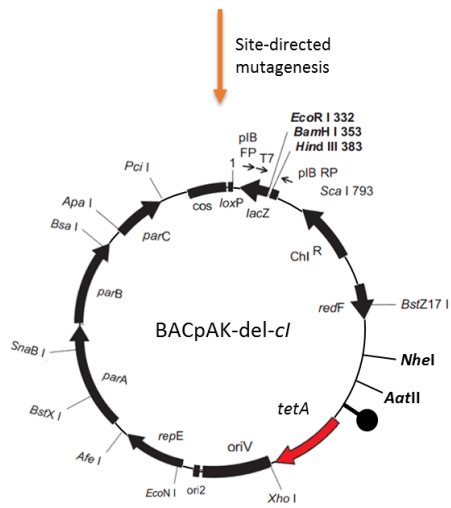
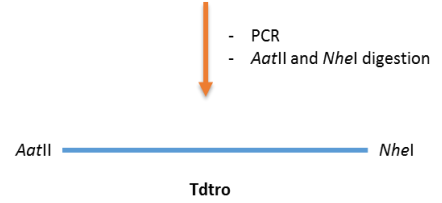
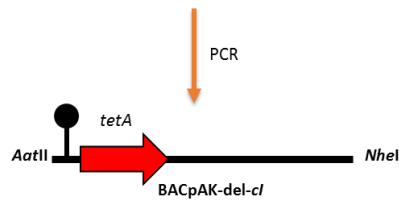
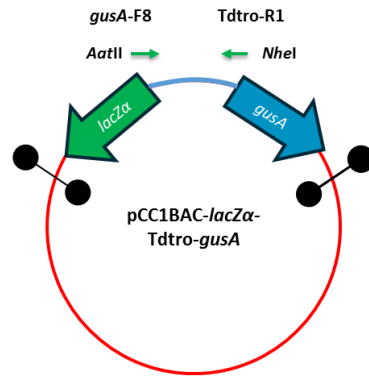
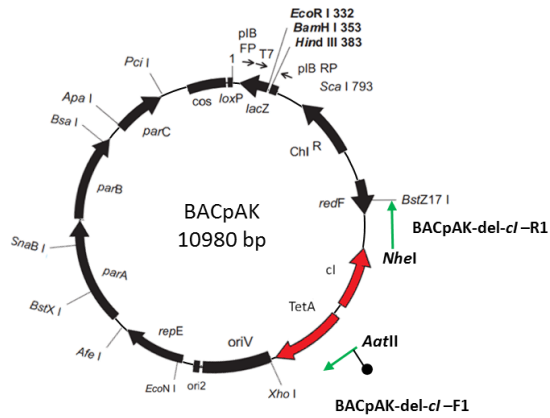


Figure 6-4 Construction of BACpAK-Terminator entrapment vector. The *cl* gene was deleted from BACpAK entrapment vector by site-directed mutagenesis, which resulted in BACpAK-del-*cl*. The promoter P_{Tdtro} was amplified from pCC1BAC-*lacZ* α -Tdtro-*gusA* plasmid, constructed in chapter 4, by using *gusA*-F8 and Tdtro-R1 primers. The promoter was then cloned in between *NheI* and *AatII* sites on BACpAK-del-*cl* plasmid. The symbols (I) and (b) represent unidirectional and bi-directional terminators, respectively. The green arrows and blue lines represent primers and P_{Tdtro}, respectively. The *cl-tetA*, *lacZ* α and *gusA* genes are represented by red, gene and blue open arrow boxes, respectively.

6.2.3 The isolation of transposable elements by *cl-tetA* using entrapment vectors

Two main strategies were performed for the isolation of TEs from the human oral metagenomic DNA (Figure 6-5). The first approach was to directly introduce the human oral metagenome into *E. coli* competent cells containing the entrapment vectors. The second strategy was to clone the human oral metagenomic library into the entrapment vectors and screened for transposition. The transformants from both approaches were incubated at 37°C after either heat shock or electroporation with various duration (1, 3, 7 and 24 hr) before spreading on LB agar supplemented with selective antibiotics, to allow the transposition of TEs.

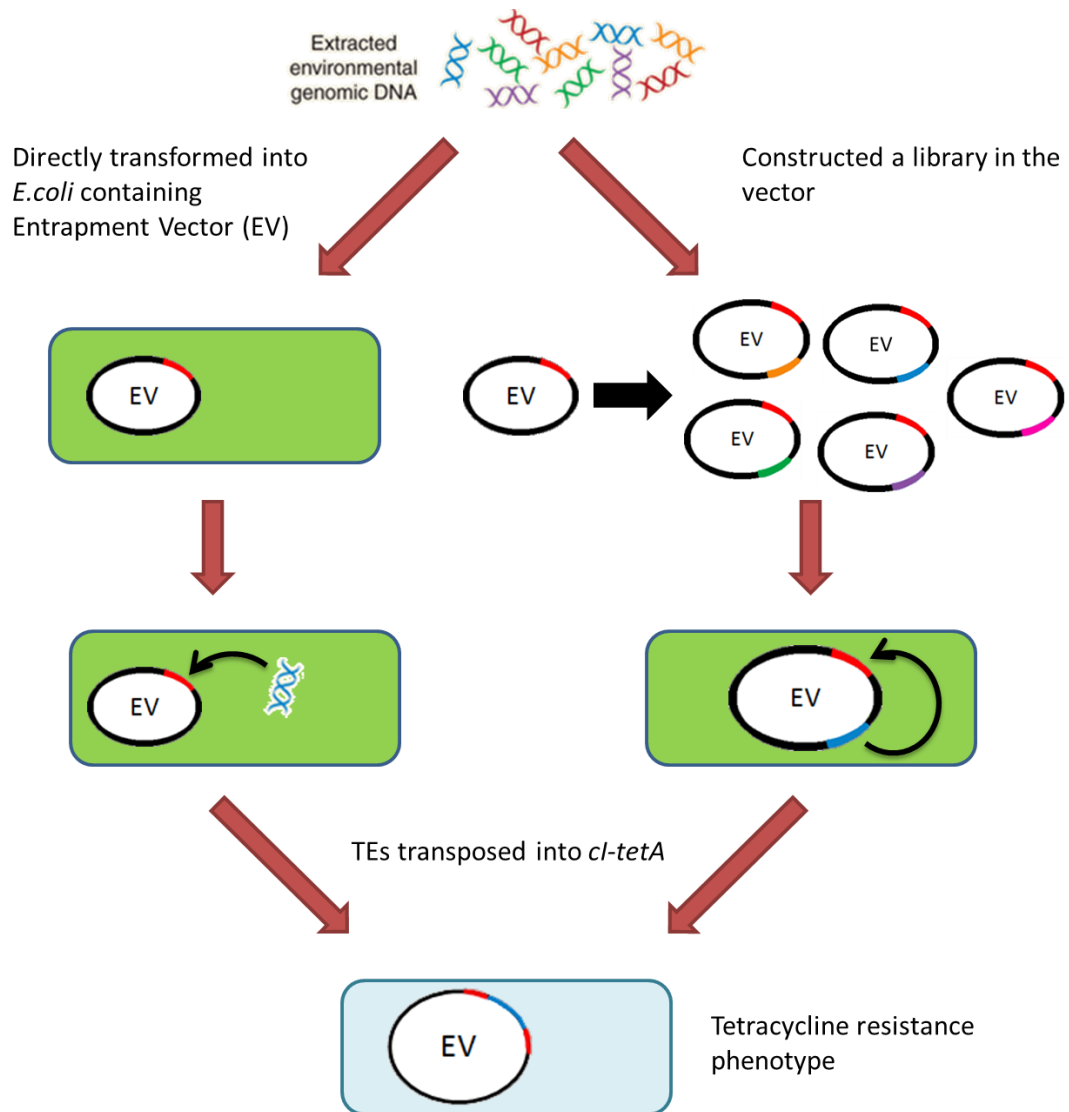


Figure 6-5 The approaches for the capturing of TEs by using entrapment vectors.

The metagenomic DNA was either directly introduced into *E. coli* containing entrapment vector by transformation (left-hand side) or constructed into entrapment vector and transformed into *E. coli* (right-hand side). If TEs located on the metagenome can transpose into a *cl-tetA* selection cartridge (red line), *E. coli* will have tetracycline resistance phenotype (blue rounded rectangular). The green rounded rectangular represents an *E. coli* with tetracycline sensitive phenotype.

6.2.3.1 Direct transformation of the oral metagenomic DNA into *E. coli* containing *tetA*-based entrapment vectors

The *E. coli* competent cells containing entrapment vector (pAK1 and BACpAK-terminator) were prepared as described in section 2.4.14. The oral metagenomic DNA was introduced into the competent cells by heat-shock DNA transformation, as described in 2.4.15.1. The pAK1 transformants were spread on LB agar containing kanamycin (50 µg/ml) and tetracycline (20 µg/ml). The BACpAK-terminator transformants were spread on LB agar containing chloramphenicol (12.5 µg/ml) and tetracycline (2 µg/ml). All culture was incubated at 37°C for 18 hr, then checked for the growth on selection plates.

6.2.3.2 Direct transformation of the oral metagenomic DNA into *E. coli* containing promoterless-entrapment vectors

The pCC1BAC-*lacZ* α -*gusA* plasmid developed in Chapter 4 was used as a promoterless-entrapment vector. As the space for TE insertion between *lacZ* α and *gusA* reporter genes was small (only 90 bp) (Figure 6-6), it would be better to expand the space by adding an insert DNA, which will increase the chance for TE to transpose between the reporter genes. The pCC1BAC-*lacZ*-*gusA*-W1, isolated during the screening of promoter-containing GCs in section 4.2.8.2, was selected because it contained 358 bp insert DNA with no promoter activity in either direction (shown by the phenotypes on agar plate) and no terminator sequence (analysed by ARNold software (Naville *et al.*, 2011)) (Figure 6-6).

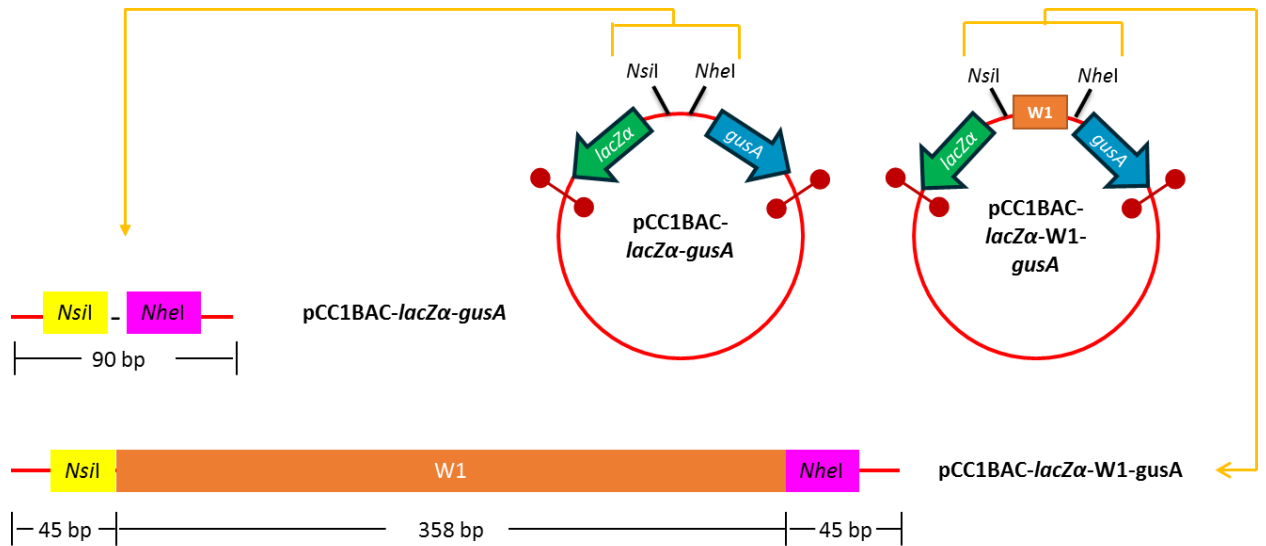



Figure 6-6 Schematic representation of the target site for TE insertion on pCC1BAC-*lacZα-gusA* plasmid. The green and blue open arrow boxes represent *lacZα* and *gusA* reporter genes, respectively. The symbols () represent bi-directional terminators. The yellow, pink and orange boxes represent *NsiI* restriction site, *NheI* restriction site and W1 insert, respectively.

The *E. coli* competent cells containing pCC1BAC-*lacZ-gusA-W1* were then prepared and used in the heat-shock transformation. The transformants were diluted to 10^{-4} and grown on LB agar supplemented with 12.5 $\mu\text{g}/\text{mL}$ chloramphenicol, 80 $\mu\text{g}/\text{mL}$ 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal), 50 μM IPTG, and 70 $\mu\text{g}/\text{mL}$ 4-methylumbelliferyl- β -D-glucuronide (MUG). All cultures were incubated at 37°C for 18 hr, then checked the phenotypes of colonies on selective plates.

6.2.3.3 Oral metagenomic library construction on BACpAK entrapment vectors

The oral metagenome was prepared, following the protocol described previously (Seville *et al.*, 2009). The oral metagenome was digested with either *Sau3AI* or *BamHI*, as both enzymes leave a 5'-GATC overhang which can be ligated to the *BamHI*-digested BACpAK entrapment vector. Two digestion reactions were prepared each containing 22 µl oral metagenome (100 ng/µl), 1 µl diluted *Sau3AI/BamHI* (10 units), 4 µl 10X restriction enzyme buffer and 13 µl molecular grade water. One of the reactions was incubated at 37°C for 1 min, while the other was incubated for 2 min. Both reactions were combined and purified by isopropanol precipitation. The air-dried DNA pellet was resuspended in 30 µl of molecular grade water. BACpAK entrapment vector was digested with *BamHI* and dephosphorylated. The linearised BACpAK was then ligated to the digested oral metagenomic DNA. The ligation was then desalted and electroporated into TransforMax™ EPI300™ Electrocompetent *E. coli*, as described in section 2.4.13 and 2.4.15.2. The size of the library was estimated by extracting and digesting the plasmids from 10 white colonies. The transformants were grown on LB agar containing chloramphenicol (12.5 µg/ml) and tetracycline (2 µg/ml). All culture was incubated at 37°C for 18 hr and checked for the growth on selection plates.

6.2.4 Identification and analysis of the tetracycline resistant colonies.

The tetracycline resistant clones were initially analysed by colony PCR across the selection cartridge. For the *cl-tetA*-based entrapment vectors, *cl-tetA-F1* and

ERIS primers were used in the amplification (Figure 6-7). The positive clones can be classified based on the size of PCR products, including 1.) no change in size: clones with a point mutation(s) or small insertion/deletion of DNA (less than 100 bp), 2.) smaller size of PCR product: clones with deletion and 3.) larger size of PCR product: potential transposition of TEs. The plasmids of the clones with putative insertion were extracted and sequenced to identify and characterise inserted DNA.

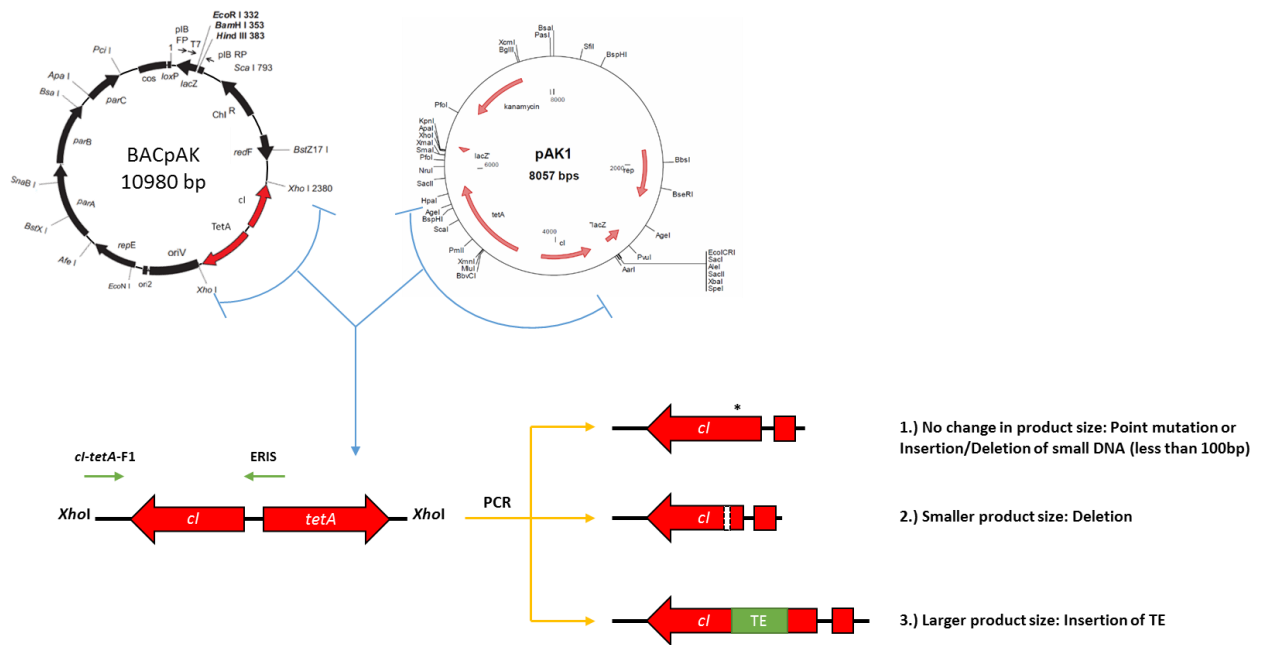


Figure 6-7 Identification and analysis of the tetracycline resistant clones by colony PCR. The colony PCR was performed with *cl-tetA-F1* and ERIS primers (shown as green arrows). The clones were categorised based on the size of PCR products: no change, smaller and larger PCR products. The asterisk indicates point mutations or small insertion or deletion of small DNA. The red, white and green boxes represent a *cl-tetA* gene, gene deletion and TE insertion, respectively.

6.2.5 Construction of amplicons to repeat the insertions found during the screening of BACpAK-oral metagenome library

During the screening in section 6.3.2.1, several clones with an insertion, which was not likely to be a result of classical transposition, were found. Therefore, different amplicons were constructed to test the hypotheses.

6.2.5.1 Circular DNA formation

One of the hypotheses was that it could be a recombination between BACpAK vector and a circular insert DNA. Therefore, circular DNAs were prepared to test the hypothesis. The PCR amplicons, flanked by restriction sites, were digested and purified with QIAquick PCR Purification Kit. After measuring the concentration of the digested products, they were diluted with molecular grade water to get the final concentration of DNA below 1 ng/ μ l. The self-ligation reaction was then set up with a total volume of 120 μ l, containing 12 μ l of 10X ligation buffer, 100 μ l of diluted digested DNA, 1 μ l of T4 DNA ligase (50 U/ μ l) and 7 μ l of molecular grade water. The reaction was incubated at 16°C for 18 hr and heat inactivated at 65°C for 10 min.

The presence of circular DNA was checked by treating the products with Plasmid-Safe™ ATP-Dependent DNase (Cambio, Cambridge, UK), following the instructions from the manufacturer. The reactions contained 2 μ l of 25 mM ATP, 1 μ l of 10x reaction buffer, 1 μ l of Plasmid-safe DNase (10 units), 2 μ l of self-ligation product and 4 μ l of molecular grade water. After incubation at 37°C for 1

hr, the reaction was inactivated by incubation at 70°C for 30 min and visualised on 1% agarose gel.

6.2.5.2 Preparation of *Bam*HI-digested antibiotic resistance genes

Two *Bam*HI-digested resistance genes (ampicillin and tetracycline resistance genes) were prepared. The ampicillin resistance gene and its promoter were amplified from the pGEM-T easy vector with *AmpR*-F1 and *AmpR*-R1 primers (containing *Bam*HI restriction sites) (Appendix 4) and cloned into a pHSG396 vector. The pHSG396::*AmpR* was introduced into *E. coli* and screened on LB agar containing chloramphenicol (pHSG396 selective marker) and ampicillin. As the clones containing pHSG396::*AmpR* can resist to ampicillin, it confirmed that *AmpR* were expressed. The *AmpR* amplicons were amplified from pHSG396::*AmpR* plasmid with *AmpR*-F1 and *AmpR*-R1 primers and digested with *Bam*HI to generate the GATC overhang.

Tetracycline resistance gene *tet*(M) and its promoter located on Tn916 was amplified from the genomic DNA of *Bacillus subtilis* BS34A (extracted by Ms Deena Al Harbi, UCL Eastman Dental Institute) by using *tet*(M)-F1 and *tet*(M)-R1 primers (Appendix 4). The *tet*(M) amplicons were then digested with *Bam*HI. The expression of *tet*(M) amplicons was confirmed by cloning *tet*(M) amplicon into pUC19 and introduced into *E. coli*, which showed that *E. coli*::[pUC19::*tet*(M)] can grow on tetracycline-containing agar plates.

6.3 Results

6.3.1 Direct transformation of metagenomic DNA into *E. coli* containing pAK1 entrapment vectors

Introducing the human oral metagenomic DNA into *E. coli* containing pAK1 entrapment vector showed several colonies on the tetracycline-containing agar plates. The insertion of TEs on tetracycline resistance clones was determined by colony PCR across the *cl-tetA* genes. Most of the clones showed no change in the size of amplicons, suggesting no TE transposition. Some of the plasmids from these clones were sequenced to determine the mutations, as shown in Figure 6-8. Point mutations were found in the *cl* gene, P_{RM} promoter and P_R promoter. Small deletions and insertions were also found within *cl* gene. Therefore, the sequencing of *cl* gene of tetracycline resistance clones were subsequently performed only when there was a change in size of colony PCR product.

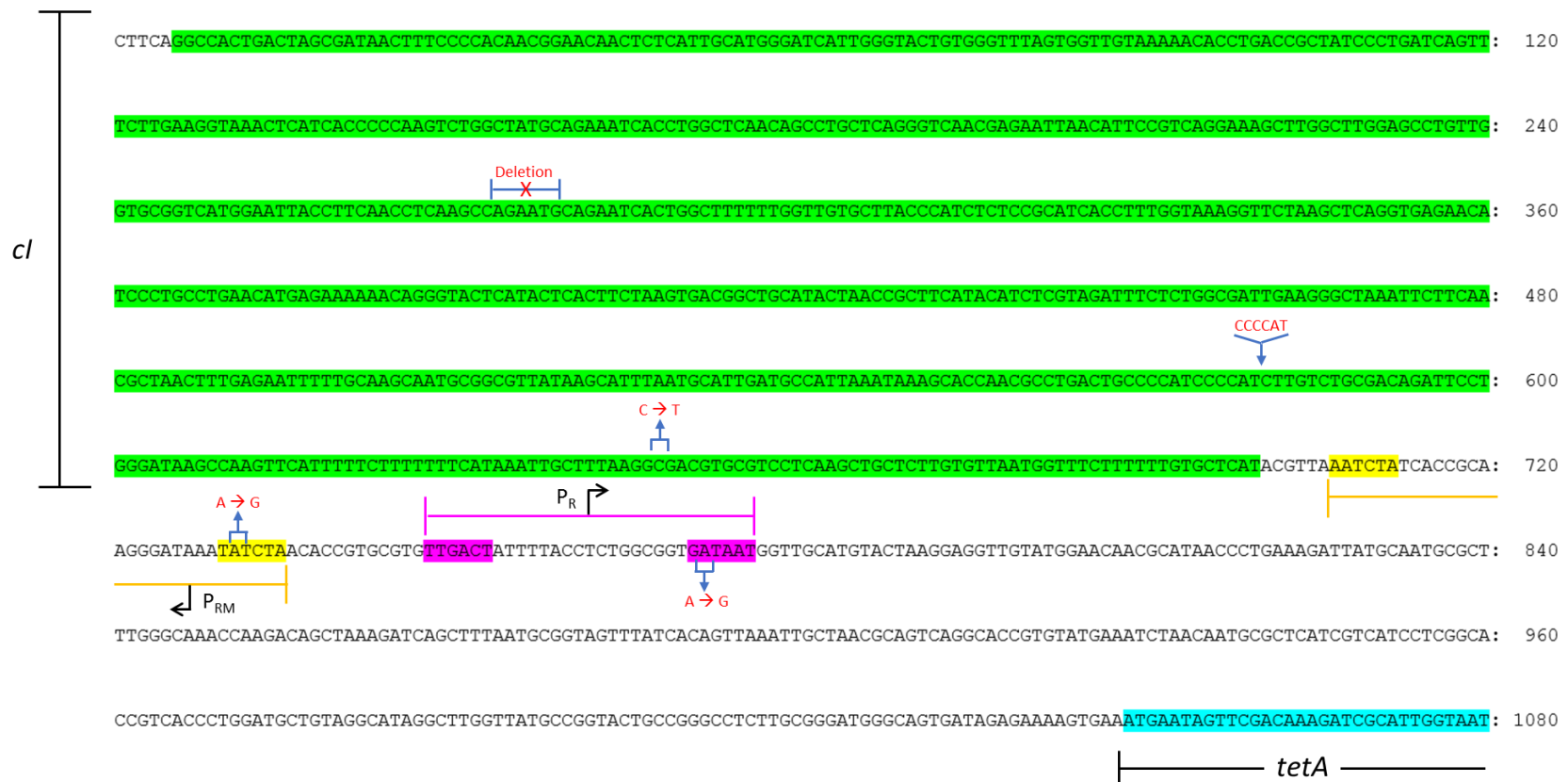


Figure 6-8 Sequencing results of the *cI-tetA* genes of pAK1 tetracycline resistance clones. The sequences of *cI*, *tetA*, P_{RM} and P_R were highlighted in green, blue, yellow and purple, respectively. The black arrows represent promoter, pointing the direction of transcription. The positions of point mutations, small insertion and small deletion are indicated with blue arrows and the nucleotide changes are shown in red.

Three clones (k3-5, k6-1, and t6-6) were shown to have an increase in the size of the *cl-tetA* region by *Hind*III digestion of the plasmids (Figure 6-9). Sequencing result showed the insertion of transposon Tn 1000 (k6-1) and insertion sequence IS5 (k3-5 and t6-6) with 100% and 99% similarity to *E. coli* genome. The presence of both TEs within the *E. coli* host and the human oral metagenome were determined by performing PCR with the primers specific to Tn 1000 and IS5. The results showed that both TEs were present in the *E. coli* host genome and not the metagenome. Therefore, our entrapment vector can successfully detect the transposition of TEs from the *E. coli* chromosome, while the transposition of TEs from the oral metagenome have not been detected yet.

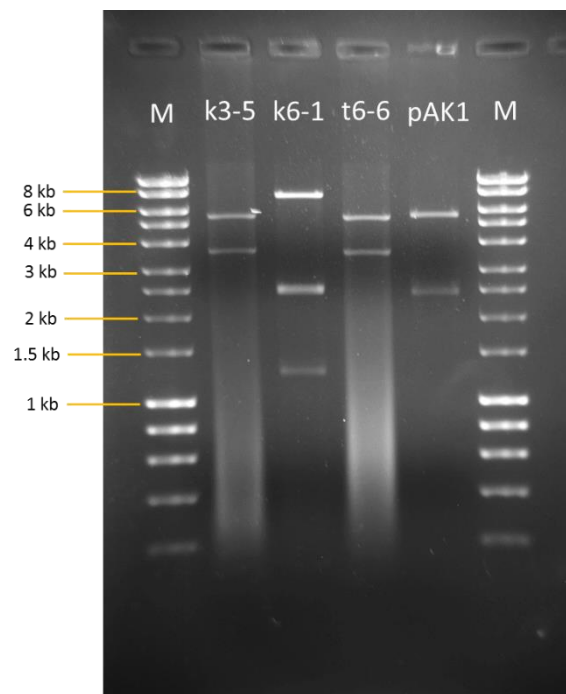


Figure 6-9 *Hind*III digestion of the plasmids extracted from the pAK1 *E.coli* with the transposition of TEs on 1% agarose gel. Lane M, Hyperladder HyperLadder™ 1kb.

The negative control for DNA transformation (transformation with distilled water) also showed several colonies on LB agar (kanamycin and tetracycline supplemented). As there was no change in the size of colony PCR amplicons of these tetracycline resistance plasmids, they were not investigated further because their resistance trait was likely to confer by point mutations within *cl* gene.

6.3.2 The detection of TE transposition by constructing a DNA library on BACpAK entrapment vectors

6.3.2.1 Construction of oral metagenomic DNA on BACpAK entrapment vector.

BACpAK entrapment vector was constructed as described in section 6.2.1. The plasmids from BACpAK-6 and BACpAK-16 clones was extracted as it contained the correct BACpAK construct according to the results from colony PCR and plasmid digestion. They were therefore ligated with the *Sau3AI*-digested oral metagenome. After the electroporation of BACpAK-6/16-*Sau3AI* oral metagenome, the control plate (LB agar supplemented with IPTG/Xgal and chloramphenicol) showed a lawn of blue colonies (uncountable) and 12 white colonies, suggesting that there was a problem with the ligation as the blue colony represents a clone with no insert. On the screening plates containing tetracycline, six and two colonies were found from the cells electroporated with BACpAK::oral metagenomic DNA and BACpAK (negative control), respectively.

All clones on the screening plates were then characterised by *cl-tetA* PCR and sequencing. Two clones (BPNB1 and BPNW1) from the negative control had no insertion within the selection cartridge, which was showed that their resistance occurred due to point mutations. The other six clones (BPPB1, BPPB2, BPPW1, BPPW2, BPPW3 and BPPW4) showed an increase in the size of the PCR products (Figure 6-10). Sequencing showed an insertion of human DNA in four clones and *Veilonella parvula* DNA in one clone. However, sequencing reactions on the BPPW2 plasmid failed with both *cl-tetA*-F1 and ERIS primers, which might be due the loss of primer binding sites, suggested by the nonspecific bands of the *cl-tetA* PCR in lane 6 of Figure 6-10. The details of these clones were discussed in section 6.3.5.

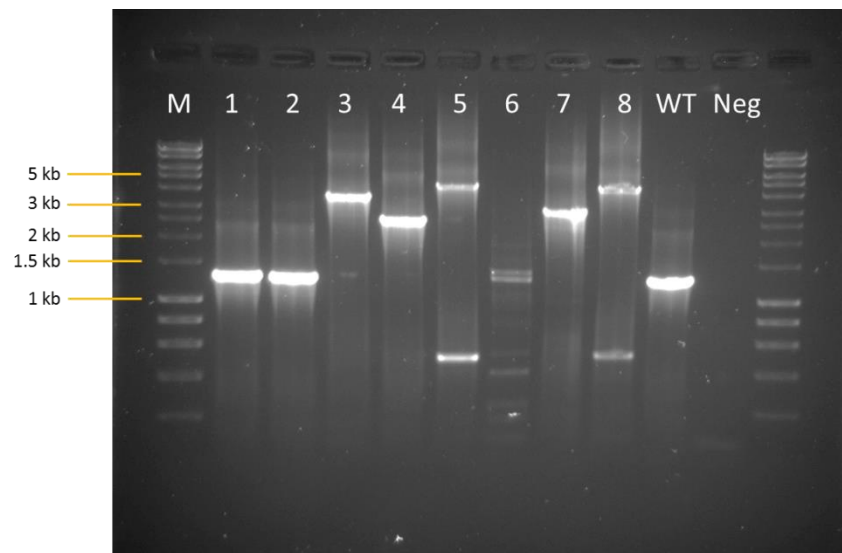


Figure 6-10 The *cl-tetA* PCR products amplified from the BACpAK-oral metagenome clones found on LB agar containing tetracycline (screening plates). Lane M, Hyperladder HyperLadder™ 1kb. Lane 1-8, BPNB1, BPNW1, BPPB1, BPPB2, BPPW1, BPPW2, BPPW3, and BPPW4, respectively. Lane WT and Neg were empty BACpAK and negative control, respectively.

As the results of the library construction showed a problem in ligation. The experiment was repeated by ligating the *Bam*HI-digested BACpAK with the fully *Bam*HI digested oral metagenome. However, the ligation problem still remained as a lawn of blue colonies (uncountable) was found on the control plate. Two additional colonies (BPPB3 and BPPB4) were identified from the tetracycline-containing plates in which the characterisation of these clones suggested that there was an insertion of the same *Veilonella parvula* DNA found in the clone BPPW3 (Table 6-2) and discussed in section 6.3.5.

The ligation problem of BACpAK-oral metagenome was then suspected to occur due to the BACpAK vector because the digestion of BACpAK6 and 16 with *Bam*HI showed multiple bands, instead of a single linearised vector band at 10.8 kb (Figure 6-11B and C). The digestion activity of *Bam*HI was normal, as the *Bam*HI digestion on pUC19 vector showed complete digestion with a single band at 2.8 kb (Figure 6-11D).

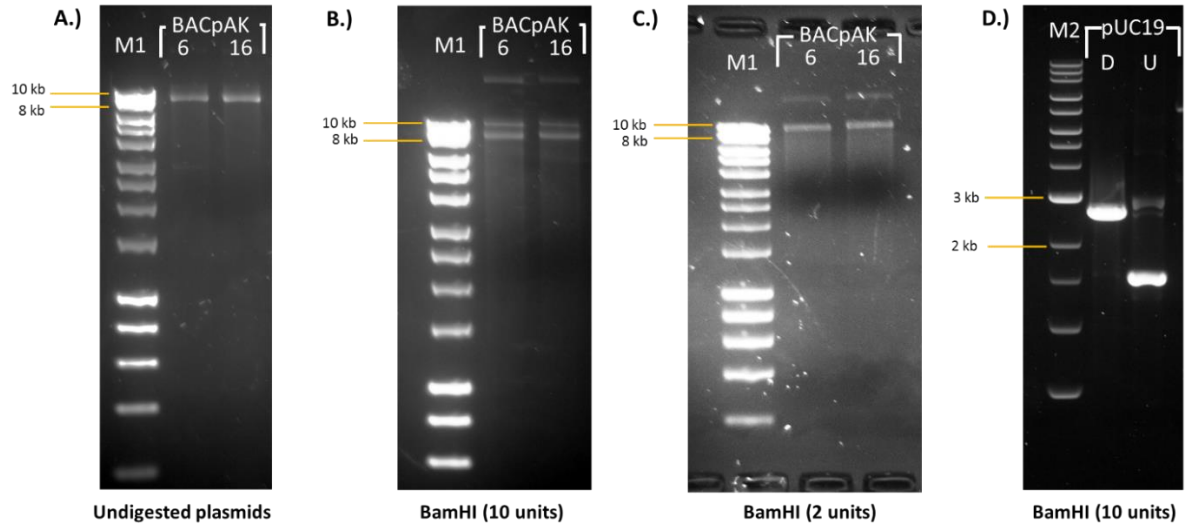


Figure 6-11 *Bam*HI digestion of BACpAK and pUC19 vectors. A.) The undigested BACpAK, the BACpAK digested with B.) 10 units *Bam*HI and C.) 2 units *Bam*HI, and D.) pUC19 were separated on 1% agarose gel. Lane M1, M2, D and U loaded with HyperLadder™ 1kb, Quick-Load® 1 kb Extend DNA Ladder, digested pUC19 and undigested pUC19, respectively. (B and C) It was shown that BACpAK 6 and 16 failed to be completely linearised with *Bam*HI, as shown by multiple bands on the gels. (D) The *Bam*HI enzyme activity was normal as shown by a single linearised pUC19.

The linearised BACpAK-6 and -16, used in the library construction, was then introduced into *E. coli* by transformation. However, the transformation results showed a lawn of blue colonies, suggesting that it was not a ligation problem. It was instead a BACpAK digestion problem, which resulted in a carry-over of the undigested vector into the ligation and transformation reactions.

6.3.2.2 Construction of new BACpAK plasmid

The new BACpAK vector was constructed by digesting BACpAK-6 plasmid with *Bam*HI, then isolated only the 10-kb linearised BACpAK by gel extraction. The

gel extracted product was then re-ligated and transformed into *E. coli*. Eight colonies were found on the LB agar containing chloramphenicol, which one of them showed single band after the digestion with *Bam*HI, called BACpAK-GE-8 (BPG-8) (Figure 6-12).

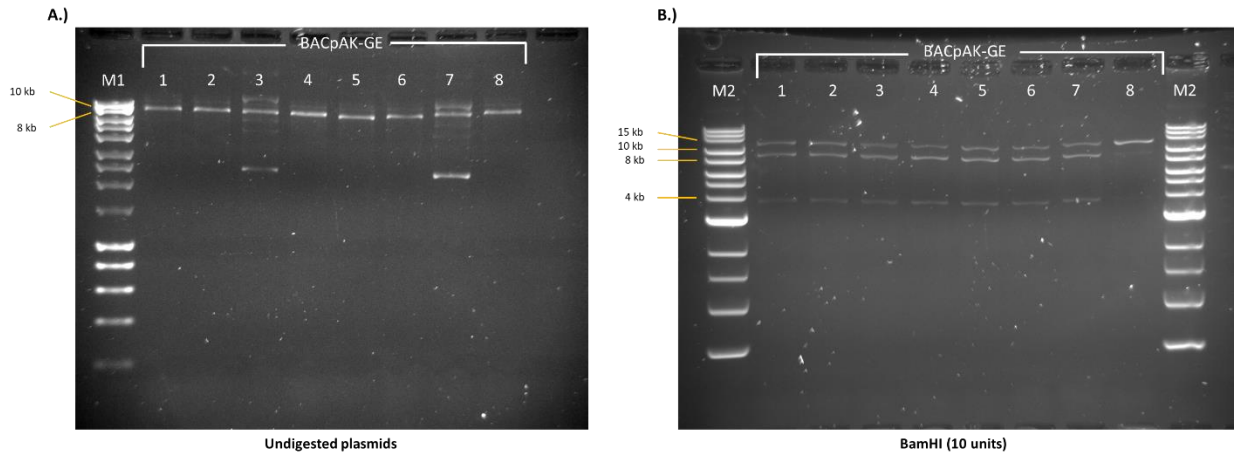


Figure 6-12 *Bam*HI digestion on BACpAK-GE plasmids. The undigested and *Bam*HI digested BACpAK-GE 1-8 were separated on 1% agarose gel. Lane M1 and M2 loaded with HyperLadder™ 1kb and Quick-Load® 1 kb Extend DNA Ladder, respectively.

6.3.2.3 Detection of TE transposition on the oral metagenomic library constructed on new BACpAK vector (BPG-8)

The construction of the oral metagenome on BPG-8 vector was first tested by ligating the vector with the *Bam*HI-digested oral metagenome. By transforming 1, 2 and 5 μ l of ligation products into *E. coli*, 13, 40 and 11 white colonies were found from 100 μ l of cells, respectively. Several blue colonies were found on the plates, not as a lawn of blue colonies (uncountable) found on libraries constructed

with BACpAK-6 and -16 in the previous section (6.3.2.1). Therefore, BPG-8 can be used for the construction of the oral metagenomic library, while the electroporation of 2 µl of ligation products showed the highest number of white colonies. No colony was found on the tetracycline-containing agar plates. Based on the digestion of 10 plasmids from white colonies, the library size was estimated to be approximately 0.418 Mb with the average insert size of 1.045 kb.

Different conditions of the digestion of the oral metagenome were optimised in order to get the larger metagenomic library and higher average insert size, as shown in Table 6-1. During the optimisation of the oral library construction, the cells were also screened on LB agar supplemented with tetracycline, however, all colonies found on the screening plates were a result of either point mutations (no change in *cl-tetA* colony PCR product) or the insertion of TEs from the host (IS1, IS4, and IS10).

6.3.2.4 The detection of TE transposition on the *Citrobacter freundii* genomic library constructed in BACpAK vector

BACpAK library construction was carried out with *Citrobacter freundii* genomic DNA (isolated from Swab and Send project by Dr Liam Reynolds). Several TEs were previously identified from *C. freundii* such as ISKpn19 (Wu *et al.*, 2016), Tn2 (Xiong *et al.*, 2016) and Tn6256 (Antonelli *et al.*, 2015). The screenings of TEs from the bacterial genomic DNA and the oral metagenome are both *ex vivo* screening, so if the BACpAK entrapment vector detects the transposition of TEs from *C. freundii* genomic DNA, it can be inferred that our BACpAK should have

an ability to capture TEs from the oral metagenome as well. However, no colony with the insertion of TEs from *C. freundii* was found on the tetracycline screening plates (Table 6-1).

Table 6-1 The optimisation of the library construction on entrapment vectors.

Source of DNA	Digestion condition	Total volume of ligation (µl)	Number of transformants (in 1000 µl of SOC medium)	Average insert size (kb)	Total insert DNA from 2 µl ligation product (Mb)	Total insert DNA (Mb)
Oral metagenomic DNA	2 U/µl <i>Bam</i> HI for 1hr	30	400	1.045	0.418	6.27
	0.1 U/µl <i>Sau</i> 3AI for 1 and 2 min	30	3540	0.95	3.363	50.45
	0.005 U/µl <i>Sau</i> 3AI for 1 min	30	1350	2.6	3.510	52.65
	0.0005 U/µl <i>Sau</i> 3AI for 1 min	30	380	4.45	1.691	25.37
<i>Citrobacter freundii</i> genomic DNA	0.005 U/µl <i>Sau</i> 3AI for 1 min	20	8390	1.05	8.810	88.1
	0.0005 U/µl <i>Sau</i> 3AI for 1 min	20	1100	5.89	6.479	64.79

6.3.2.5 Detection of the transposition of TEs located in the cloning sites of BACpAK into *cl-tetA* selection cartridge

To verify that TEs located in the cloning site of BACpAK can be transposed into *cl-tetA* selection cartridge (transposition within the same replicon), the BACpAK-IS5 plasmid was constructed by ligating the IS5, isolated in section 6.3.1, into the cloning site of BACpAK. The plasmid was then introduced into *E. coli* EPI300 by electroporation and grown on LB supplemented with tetracycline and chloramphenicol. The transposition of IS5 was determined on 80 clones found on the screening plates by *cl-tetA* colony PCR. Most of the clones showed no change in the size of amplicons. Only 9 clones showed an increased in amplicons' size, which was identified as a transposition of IS1, IS4, and IS10 from an *E. coli* host, while no IS5 was detected.

6.3.3 Direct transformation of metagenomic DNA into *E. coli* containing BACpAK-terminator entrapment vectors

The *E. coli* competent cells containing BACpAK-terminator vector was prepared as described in section 6.2.2. The heat-shock transformations were then set up to introduce different donor DNA into the competent cells separately, which were the oral metagenomic DNA, *Sau3AI*-digested oral metagenomic DNA, and *Sau3AI*-digested *C. freundii* genomic DNA. The transformants were inoculated on LB agar supplemented with chloramphenicol and tetracycline, but no colony was found on any screening plates.

6.3.4 Direct transformation of metagenomic DNA into *E. coli* containing promoterless-entrapment vectors

The oral metagenomic DNA, *Sau3AI*-digested oral metagenomic DNA, and *Sau3AI*-digested *C. freundii* genomic DNA were introduced into *E. coli* competent cells containing pCC1BAC-*lacZ-gusA-W1* by heat-shock transformation. The diluted transformants were grown on LB agar supplemented with chloramphenicol, IPTG/Xgal, and MUG. However, all colonies found on the plates showed no promoter activity on both directions as there was no blue colony and no colony exhibiting blue fluorescent under UV.

6.3.5 Illegitimate recombination on BACpAK vectors

6.3.5.1 Sequence analysis on the clones found during the screening

BACpAK-oral metagenome library

The plasmids from the six tetracycline-resistant clones found in section 6.2.3.2, which had an insertion of human DNA and *V. parvula* DNA, were extracted and sequenced. The sequencing results showed that all inserts were flanked by GATC sequences and were inserted into the same position between *cl* and *tetA* on BACpAK (Figure 6-13). The sequence analysis was shown in Table 6-2. For the human DNA samples, no intron was found as identified by using Ensembl (Yates *et al.*, 2016).

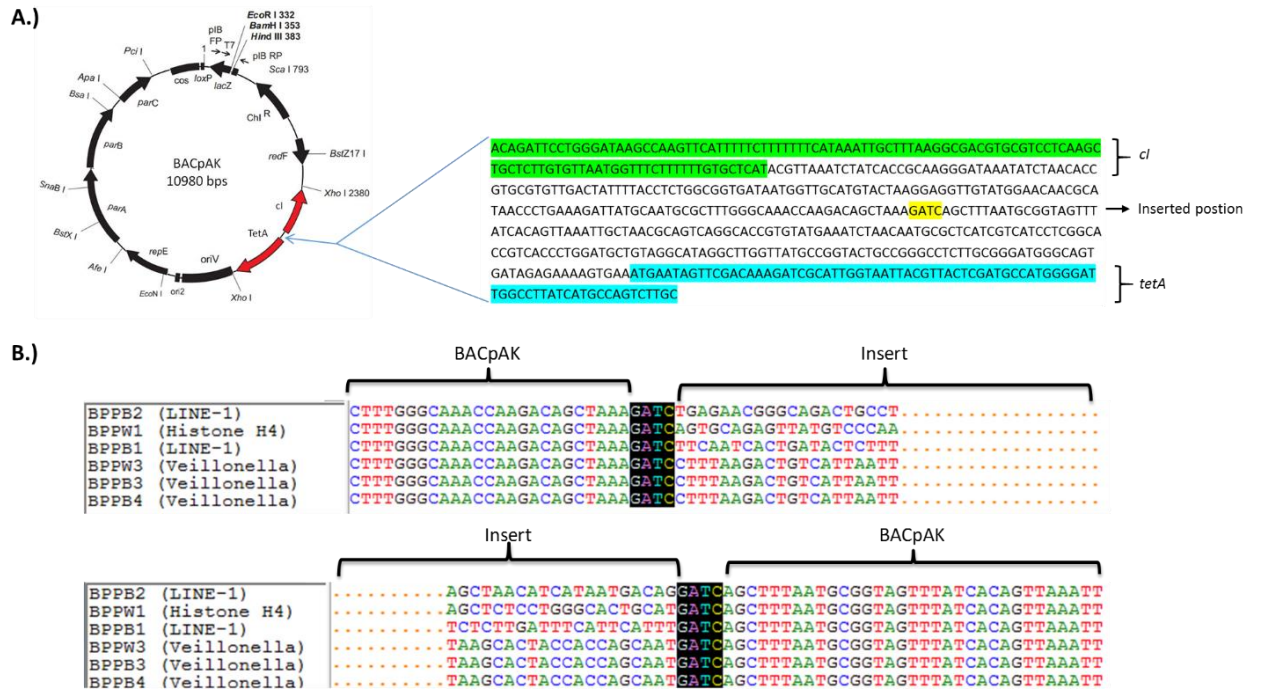

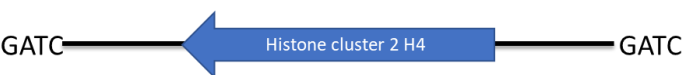




Figure 6-13 The insertion of Human DNA and *Veillonella parvula* inserts. A.) The location of the insertion on BACpAK entrapment vector. The green, yellow and blue highlights indicate *cl*, inserted position and *tetA*, respectively. B.) The alignment of the left and right junction of the inserts.

Table 6-2 Details of the human DNA and *Veillonella parvula* inserts found at the same position between *cl* and *tetA*, and their BlastN and BlastX results.

Sample (size)	BlastN	Accession number (similarity/coverage)	BlastX	Predicted structure*
BPPB2 (1336 bp)	<i>Homo sapiens</i> chromosome 5 clone CTB-99A3	AC008728.7 (100%/100%)	LINE-1 retrotransposon element orf1 [<i>Homo sapiens</i>]**	
BPPW1, BPPW4 (3282 bp)	<i>Homo sapiens</i> BAC clone CH17-164C24 from chromosome 1	AC239868.2 (99%/57%)	Histone cluster 2 H4 family member b [<i>Homo sapiens</i>]**	
BPPB1 (2431 bp)	<i>Homo sapiens</i> chromosome 5 clone RP11-115A1	AC114295.2 (100%/99%)	ORF1, ORF2 [<i>Homo sapiens</i>] Domain: transposable element**	
BPPW3, BPPB3, BPPB4 (1810 bp)	<i>Veillonella parvula</i> DSM 2008	CP019721.1 (83%/80%)	Serine hydroxymethyltransferase [<i>Veillonella atypica</i>]	

* The predicted ORFs are represent in blue arrows. The dash boxes represent the regions that are not present in the inserts compared to the sequences in the database.

** No intron was found in human DNA samples as verified by using Ensembl (Yates *et al.*, 2016)

6.3.5.2 Investigation of the illegitimate recombination

6.3.5.2.1 The transformation of GATC-circular DNAs with BACpAK vector

An attempt to repeat these insertions were then performed. The first hypothesis was that the inserts were circularised during the ligation reaction and electroporated into *E. coli* together with the BACpAK vector, as the sequence analysis of BPPB1 insert by BlastN and BlastX showed the separation of *orf1* on both ends of the insert (Figure 6-14A). The GATC nucleotides on both molecules were then recombined, resulting in the insertion at the same location on BACpAK (Figure 6-14B). As the BACpAK library constructions suggested that the BACpAK was not linearised and showed a lawn of blue colonies, this GATC recombination was likely to happen between a circular GATC insert DNA and a circular BACpAK vector.

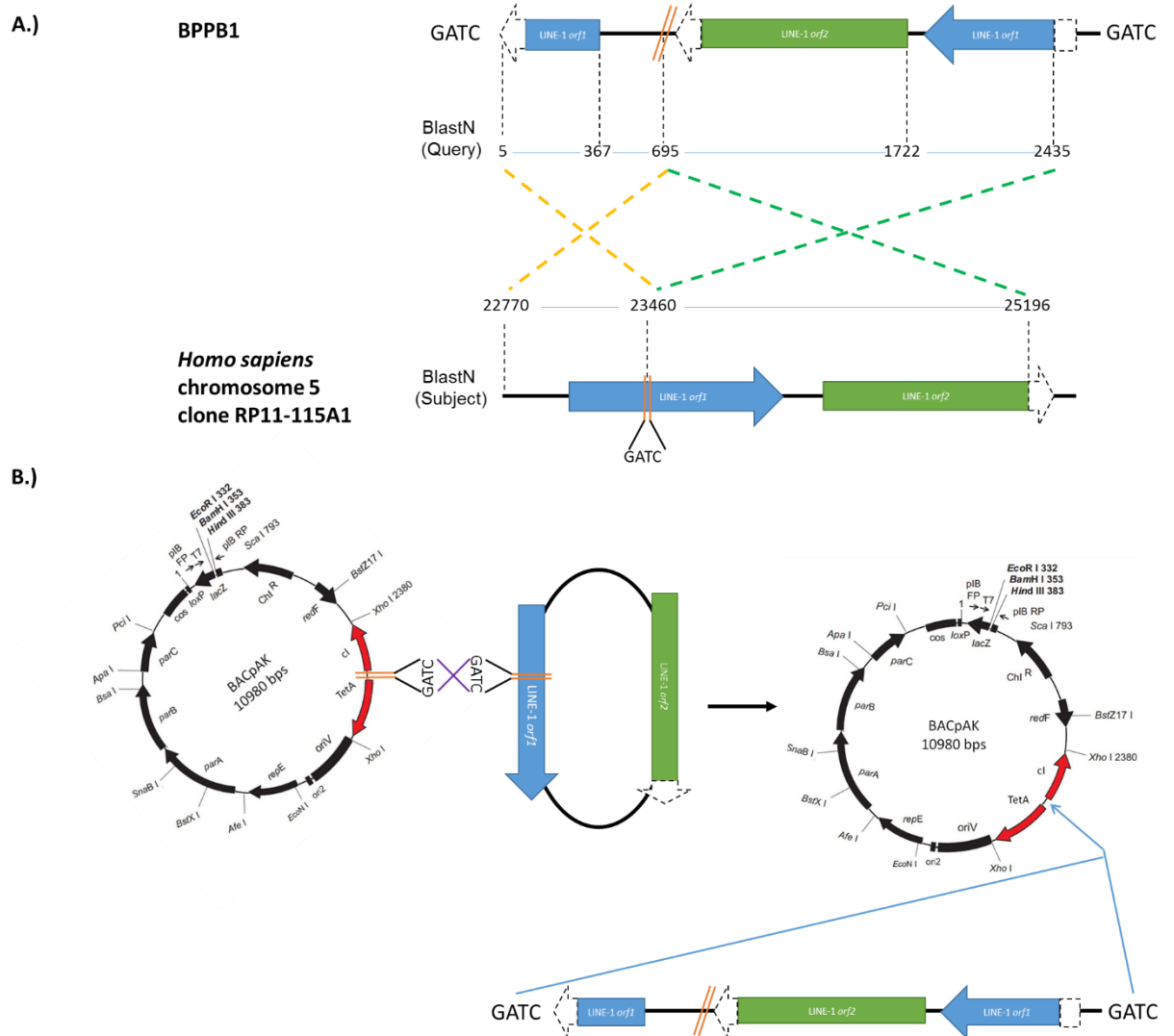


Figure 6-14 Predicted model for the insertion of the BPPB1 insert on BACpAK entrapment vector. A.) The schematic representation and the similarity between the BPPB1 insert and *Homo sapiens* chromosome 5 (Accession number AC114295.2). B.) The proposed hypothesis on the insertion of BPPB1 at the GATC sequences between *cl-tetA* on BACpAK entrapment vector. The GATC sequences on BACpAK (between *cl* and *tetA* genes) may recombine with the GATC sequences on LINE-1 *orf1* circular molecule, resulting in the insertion of BPPB1 in BACpAK with the separation of *orf1* on each side.

The first experiment to repeat this transposition was to amplify BPPB2 and BPPW3 inserts (human and *V. parvula* DNA) with primers containing *Bam*HI restriction sites. After *Bam*HI digestion, the circular molecules were

constructed as described in section 6.2.5. The circular BPPB2 and BPPW3 molecules were then electroporated together with the BACpAK vector. Several colonies were found on tetracycline-containing LB agar, but all of them either contain point mutations or insertion of IS elements from the *E. coli* host. The circular BPPB2 and BPPW3 were also transformed into *E. coli* cells containing BACpAK. However, none of the colonies found on the tetracycline-screening plates contained an insertion in the BACpAK vectors.

6.3.5.2.2 The transformation of *Bam*HI-digested DNA fragments with BACpAK vector

The next hypothesis was based on the previous report, which showed an illegitimate recombination of *Bam*HI-digested DNA fragments (containing GATC-overhangs on both ends) into GATC-sites on the yeast genome (Robert & Thomas, 1991). This was shown by the transformation of *Bam*HI-digested *URA3* DNA into yeast strain RSY12 (lacking *URA3* gene, which is necessary for the synthesis of pyrimidine ribonucleotides), which showed that the *Bam*HI-digested *URA3* fragment inserted at different GATC sites on the yeast chromosome.

The insertions on BACpAK vector could, therefore, occur due to the same mechanism in which the GATC-overhang is necessary for the integration. The *Bam*HI-digested BPPB2 and BPPW3 amplicons were, therefore, prepared by amplifying the inserts with primers containing *Bam*HI site, listed in Appendix 4, and digested with *Bam*HI. The *Bam*HI-digested BPPB2 and BPPW3 fragments were introduced into *E. coli* by electroporation together with BACpAK vector and heat-shock transformation into BACpAK-containing

competent cells. The clone with the insertion of either BPPB2 or BPPW3 was not found on screening plates. *Sau3AI*-digested oral metagenomic DNA was also introduced into *E. coli* with both transformation approaches, but no positive clone was detected.

6.3.5.2.3 The transformation of *Bam*HI-digested antibiotic resistance genes with BACpAK vector

The last approach was to use the antibiotic resistance genes as a selective marker, which would allow detection of clones with the insertions of GATC-DNA fragments not only in the GATC sequences between *cl-tetA* on BACpAK vector but also other locations on BACpAK and *E. coli* genome.

The *Bam*HI-digested *AmpR* amplicons, prepared as described in section 6.2.5.2, were electroporated together with BACpAK into *E. coli*. The colonies found on the screening plates (LB plates supplemented with ampicillin) were shown to be resistant due to the presence of pHSG396::*AmpR* plasmids, not due to recombination, as they also had chloramphenicol resistant phenotype (Figure 6-15). To reduce the contamination of pHSG396::*AmpR* plasmids in the transformation, the *AmpR* fragments were amplified by using the 10^{-7} dilution of pHSG396::*AmpR* plasmid as a template for the PCR. Several colonies were found on ampicillin-containing LB plates. However, their resistance phenotype was still conferred by the presence of pHSG396::*AmpR* plasmid.

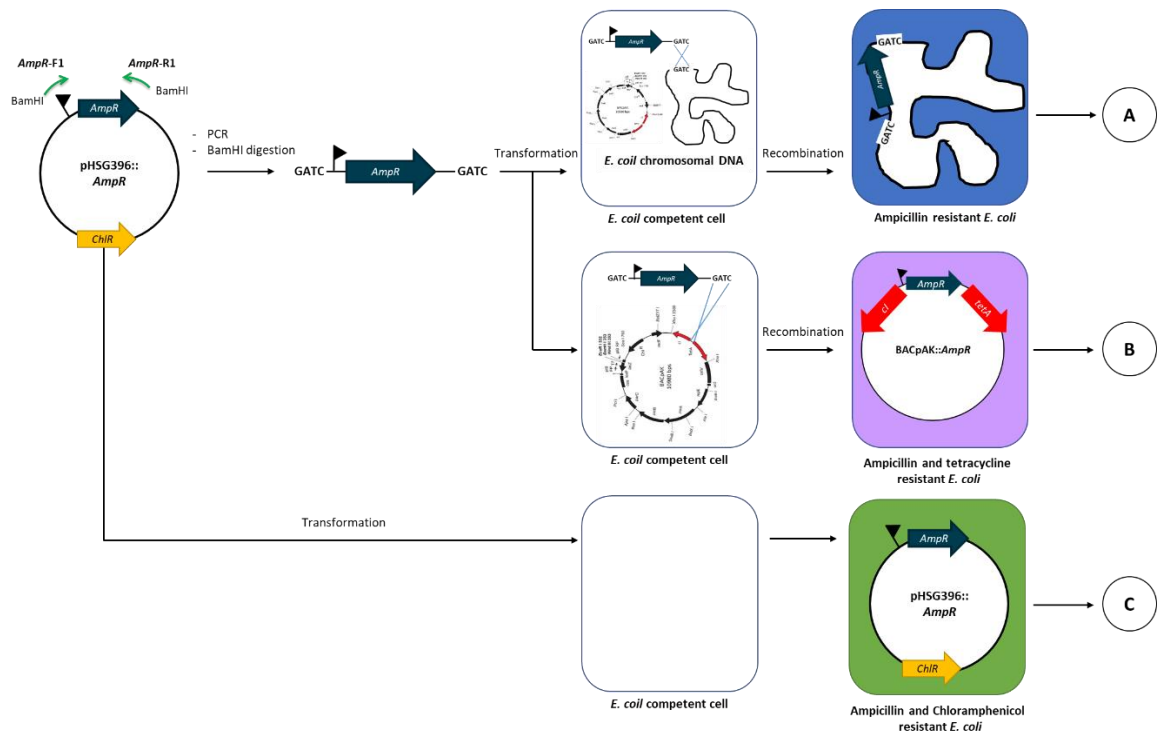


Figure 6-15 The transformation of ampicillin resistance gene (*AmpR*) into *E. coli* competent cells. The *Bam*HI-digested *AmpR* were introduced into *E. coli* together with BACpAK vector, so that it will allow us to detect for the recombination of the GATC-containing DNA with either (A) *E. coli* chromosomal DNA (Ampicillin resistance, shown in blue) or (B) BACpAK vector (Ampicillin and tetracycline resistance, shown in purple). However, all ampicillin resistant clones were contained (C) pHSG396::*AmpR* plasmid (Ampicillin and chloramphenicol resistance, shown in green). The symbols (▶) represent promoters.

To avoid the contamination of plasmid, *Bam*HI-digested *tet*(M) was prepared, as described in 6.2.5.2, from the chromosomal DNA of *B. subtilis*, as they should not be maintained in the competent cells same as pHSG396::*AmpR*. An electroporation of *Bam*HI-digested *tet*(M) and BACpAK did not show colony with the transposition of *tet*(M). The circular form of *Bam*HI-*tet*(M) was also prepared and electroporated into *E. coli* together with BACpAK vector. However, none of the colonies on the tetracycline-screening plates had the insertion of *tet*(M).

6.4 Discussion

The isolation of TEs with entrapment vectors have been shown in both Gram-negative and Gram-positive bacteria, such as *Pseudomonas cepacia*, *Agrobacterium tumefaciens*, *Brevibacterium lactofermentum*, *Rhodococcus fascians* and *Corynebacterium glutamicum* (Scordilis *et al.*, 1987, De Meirsmen *et al.*, 1989, Jager *et al.*, 1995). The advantage of this approach is that no knowledge about TEs is required in the detection, allowing the isolation of novel TEs. This method has been used *in vivo* to capture active TEs in one bacterial species at a time. It would be useful if we can apply this method to metagenomes so that TEs from multiple species can be investigated and isolated in a single experiment.

In our study, none of the clones showed the transposition of TEs from the human oral metagenome. Most of the colonies, containing the *cl-tetA*-based entrapment vectors, showed tetracycline resistance due to point mutations either within the *cl* gene or its promoter. Clones with point mutations were also found as part of the previous entrapment vector studies, especially in those containing lethal genes and antibiotic selection cartridges. For example, pMEC1 (*cl-tetA*) and pMAT1 (*sacB*) entrapment vectors, which previously identified novel TEs such as IS*Ppa1*, Tn3434 and ISAs1, showed that 21% and 2% of the mutants contained point mutations, respectively (Bartosik *et al.*, 2003, Szuplewska & Bartosik, 2009).

The BACpAK-terminator and pCC1BAC-*lacZ-gusA* were developed in this study to reduce the chance of false positive clones by point mutations, as their

detection was not depended on *cI* gene, which was shown to be disrupted by point mutation. The BACpAK-terminator plasmid was constructed by replacing the *cI* gene with terminator sequences, which is more tolerance and less likely to be disrupted by point mutations. The stem-loop structure and poly(U) tail of terminators should still be able to form in the transcripts, blocking the expression of *tetA* gene, even when one nucleotide on terminator sequence is changed. The positive clones for the BACpAK-terminator entrapment vector will, therefore, occur through either the insertion of TEs or the deletion of terminator sequences, but no colony was found in our screening. One of the advantages for the promoter-less entrapment vectors, like pCC1BAC-*lacZ-gusA*, is their lower chance of false positive clones, because the expression of a selective marker is less likely to be conferred by point mutation and deletion (Solyga & Bartosik, 2004). It is more likely to be expressed by the promoters from the TEs. However, this method can capture only the TEs with an outward directed promoter. The results from both vectors showed no false positive clones. However, no transposition activity was detected.

As our objective is to isolate the TEs from the metagenome, the insertion of TEs from the *E. coli* host, including IS1, IS4, IS5, IS10 and Tn1000, was considered as false positive. These transpositions, however, demonstrated our entrapment vectors can detect TEs from the *E. coli* host. One way to reduce this type of false positive clones is to change the surrogate host to insertion sequence-free bacteria, such as *E. coli* MDS42, which has been modified by deleting all IS elements and cryptic prophages (Csörgő *et al.*, 2012). MDS42 also has a low mutation rate, since the stress-inducible error-

prone DNA polymerases genes were deleted from its genome, which can reduce the false positive caused by mutations (Csörgő *et al.*, 2012).

To successfully capture the TEs from the oral metagenome, several factors have to be considered. As functional-based screening depends on the expression of the metagenome in a surrogate host, the genes involving in the transposition of the TEs in the oral metagenome, therefore, have to be able to be expressed in a heterologous *E. coli* host. The transcriptional machinery, such as sigma factors, of the host bacteria is an important factor because it has to recognise heterologous promoters on TEs to initiate transcription (National Research Council, 2007). The 16s rRNA analysis, described in section 3.3.1, showed that the majority of the DNA in our oral metagenome was derived from bacteria in order *Bacteroidales*. The expression of genes from *Bacteroides* in *E. coli* was shown to be impeded at the level of promoter recognition (Mastropaolo *et al.*, 2009). There are differences in both ribosomal binding site sequences (AGGAGGU in *E. coli* and AGAAAGGAG in *B. fragilis*) and promoter consensus sequence (-10 and -35 in *E. coli* but -7 and -33 in *Bacteroides*) (Tribble *et al.*, 1999, Bayley *et al.*, 2000). Therefore, different hosts might be used as a surrogate. Recently, an engineered strain of *E. coli* expressing heterologous sigma factors have been shown to increase the expression of heterologous genes in functional metagenomic studies, which could be applied with our experiment to increase the transposition of TEs (Gaida *et al.*, 2015).

The target DNA for the transposition of TEs is also important as the TEs have to insert into the capturing system on entrapment vectors for detection. It is,

therefore, a random chance for TEs to insert into the correct location for detection on entrapment vectors. Some TEs were described for their sequence-specific target choices, for example, IS1 prefers an AT-rich region, TnGBS2 prefers to insert 15-17 bp upstream of promoters and some members of the IS1111 family prefer to insert into integron *attC* recombination sites (Shiga *et al.*, 1999, Brochet *et al.*, 2009, Post & Hall, 2009). Some TEs, thus, will not be detected by this approach if that specific-target DNA are not located within the transposition marker genes on entrapment vectors.

For the library construction strategy, the digestion of the human oral metagenome can affect the isolation of TEs. For example, if the TEs contain restriction sites, they may be digested and disrupted during the library construction. Therefore, it is crucial to optimise the library to have large inserts, reducing the chance for the restriction enzymes to disrupt TEs. In our experiment, the average insert size might not be large enough to isolate large transposons, but they were large enough for IS elements. The total DNA screened of each library was also too small in which TEs might be missed from the screening. Different restriction enzymes and the digestion of the oral metagenome with a more diluted enzyme or less incubation time could be performed to avoid the disruption of TEs and increase the size of the library.

During the optimisation of oral metagenomic library on BACpAK vector, the insertions of *V. parvula* and human DNA and at the same position on BACpAK were detected, which represent interspecies and interkingdom integration of DNA, respectively. These have been shown to occur by homology-facilitated illegitimate recombination (HFIR). Interspecies integration was shown in

several bacteria such as *Acinetobacter sp.*, *Streptococcus pneumoniae* and *E. coli* (de Vries & Wackernagel, 2002, Prudhomme *et al.*, 2002a, Amarir-Bouhram *et al.*, 2011). Interkingdom integration was found by DNA transformation of tobacco plastid DNA with naturally transformable *Acinetobacter sp* (De Vries *et al.*, 2004). The important characteristic of HFIR is a single homologous region on the donor and recipient DNAs, which act as a recombinational anchor, facilitating an illegitimate recombination on the heterologous parts of the molecules (Figure 6-16). Illegitimate recombination can occur at a region with 3-6 bp identical (microhomologies). The sequence analysis between BACpAK and all inserts showed that there was only a microhomology region (at a GATC region) but not a homologous region between both molecules, so our findings were unlikely to be a result of HFIR.

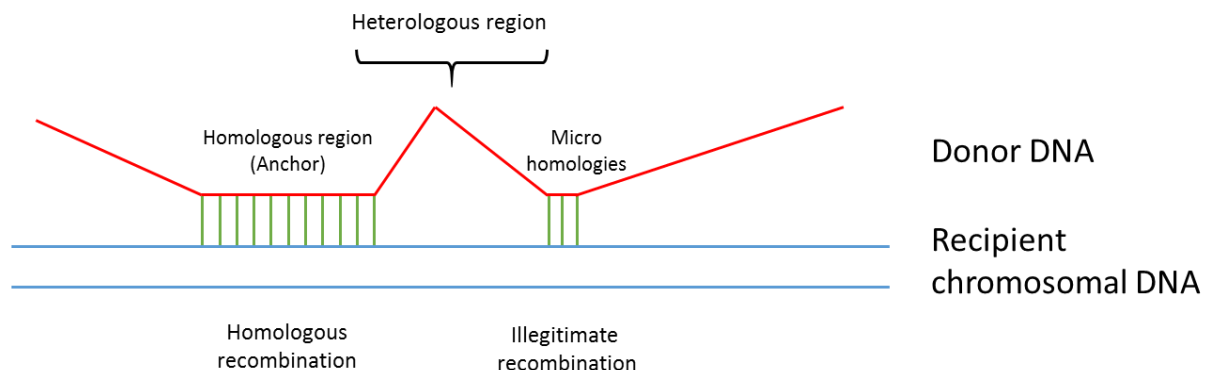


Figure 6-16 Homology-facilitated illegitimate recombination (HFIR). The green lines represent homologous region between donor and recipient chromosomal DNA (red and blue lines, respectively). (Adapted from Amarir-Bouhram *et al.*, 2011)

Another possibility is that these insertions could be artefacts, which occurred as a product from digestion and ligation. If an active *Sau3AI* enzyme from digestion contaminated into the ligation reaction between BACpAK and

Sau3AI-digested oral metagenome, the enzyme could linearise BACpAK at the GATC between *cl-tetA*, which can be ligated with the *Sau3AI*-digested oral metagenome by T4 DNA ligase. However, our results were unlikely to be this case as the BPPB3 and BPPB4 clones were identified from the library constructed with *BamHI*-digested oral metagenome. The insertion site on BACpAK was 5'-AGATCA-3', which cannot be recognised by *BamHI* (5'-GGATCC-3').

Even though two of the samples, BPPB1 and BPPB2, were matched with a long interspersed nuclear element-1 (LINE-1), transposition was unlikely to occur due to the activity of LINE-1. LINE-1 can be found in the human genome with the estimated number of 500,000 copies per genome with a size of 6-7 kb (Rodić & Burns, 2013). It consists of two ORFs: ORF1 (encoding an RNA-binding and chaperone protein) and ORF2 (encoding a protein-complex with endonuclease and reverse transcriptase activities). However, the sequence analysis of both samples showed that they contained ORF1 and partial ORF2. Therefore, the LINE-1 on both samples were unlikely to be functional or responsible for their integration into BACpAK vectors.

The integration of *BamHI* DNA fragments, containing GATC sequence at the ends, into genomic DNA of *Saccharomyces cerevisiae* was previously reported (Robert & Thomas, 1991). It was suggested to be a result of illegitimate recombination as it is a recombination event involved with little or no sequence homology. It was proposed that the 5' single-stranded GATC ends of the fragments invade and pair with the GATC sites in the chromosomal DNA (Figure 6-17). Both strands of chromosomal DNA were nicked, following

by the ligation of 3' ends of chromosomal DNA to 5' ends of the fragment. The GATC sequences, generated by either *Bam*HI or *Sau*3AI, on all inserts could invade the same GATC spot on BACpAK, resulting in integration at the same position.

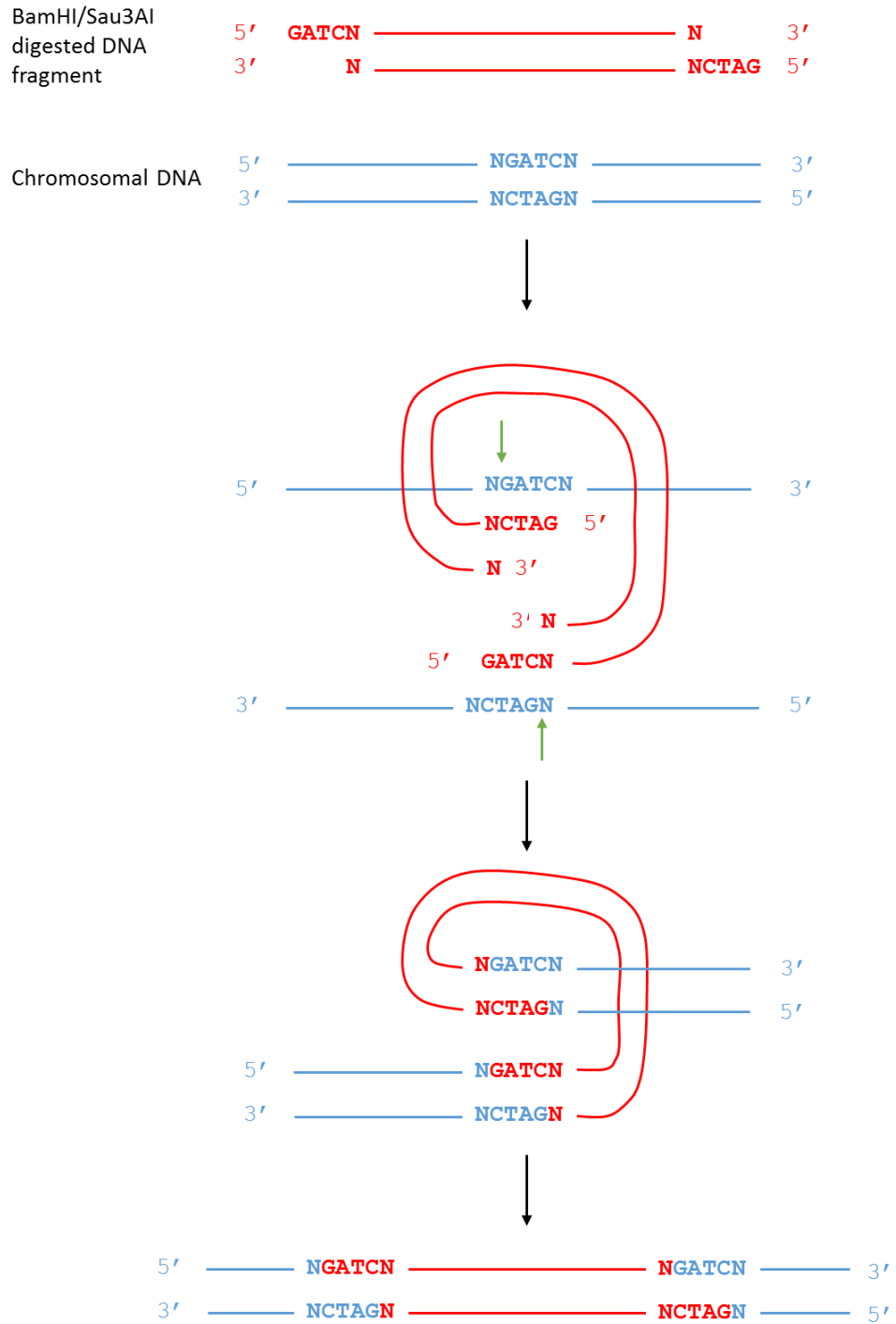


Figure 6-17 Illegitimate recombination of DNA fragment ending with GATC sequences into chromosomal DNA. The blue and red color represent chromosomal DNA and the GATC-containing DNA fragments. The green arrows indicate the position of nicking on the chromosomal DNA. (Adapted from Robert & Thomas, 1991)

One of the common features between these inserts was the GATC sequences on both ends of inserts and the integration site on BACpAK. Also, the GATC sequence is essential for a DNA mismatch repair (MMR) in *E. coli*. As there is an absence of methylation on a newly synthesised (error-containing) DNA strand, it can be used as a signal for the differentiation of the error-containing strand, called as methyl-directed MMR (Fukui, 2010). In *E. coli* MMR, MutH endonuclease nicks the unmethylated strand at the nearest hemimethylated GATC site to the mismatch base as the entry point for the excision. Therefore, the GATC insertion site in between *cl-tetA* on BACpAK could be nicked by MMR, allowing the insertion of the GATC-ended fragments.

The nucleotide sequences of the *cl-tetA* selection cartridge showed that there were multiple GATC sites, not only the one that was used in integration (Figure 6-18). The location of the GATC site could be one of the factors for the preference in integration. If integration occurs at the GATC site within *tetA*, it will disrupt *tetA* and the clone will not have a tetracycline resistant phenotype. However, there were another two GATC sites within *cl* genes that could be used and allow *tetA* to be expressed. By analysing the DNA secondary structure formation by Mfold, it was shown that the GATC site in between *cl-tetA* (used in the insertion) was located on the stem-loop secondary structure, which could be necessary for the integration (Figure 6-18).

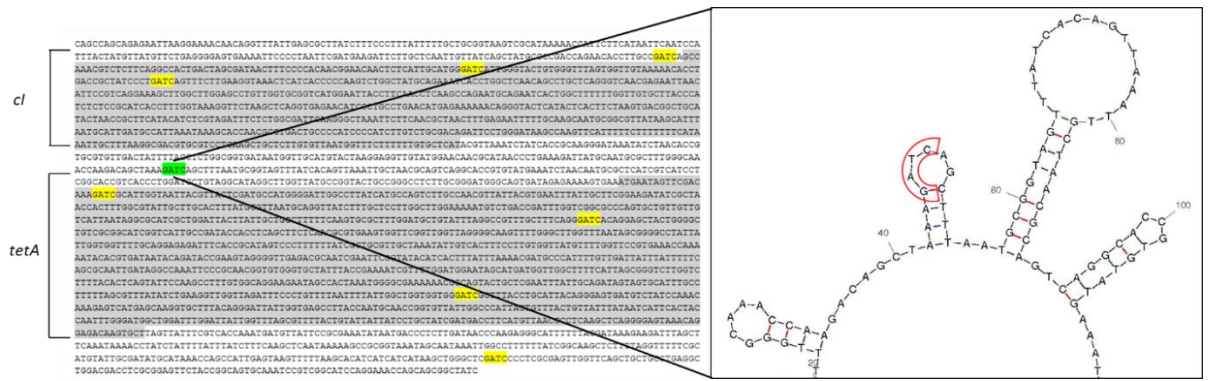


Figure 6-18 The DNA secondary structures predicted by Mfold on BACpAK entrapment vector. The yellow highlights indicate the location of GATC sites, while the green highlight indicated the GATC site used in the insertion. The red box indicates the location of GATC site.

The sequence analysis on the BPPB1 insert suggested that the insertion could occur through the recombination between two circular structures: BACpAK and GATC-containing circularised structures from the human oral metagenomic DNA. The circular structures may form during the ligation reaction in which the *Sau3AI/BamHI*-digested oral metagenomic DNA self-ligated through the GATC-overhang on both ends. Recently, there was a report on an extrachromosomal circular structure, called microDNA, found in mammalian cells with a size between 60 to 2000 bp (Shibata *et al.*, 2012). MicroDNA were usually flanked by 2 to 15 bp direct repeats, which were hypothesised to arise from a replication slippage and the MMR pathway. Some microDNAs were mapped to originate from repetitive elements, including LINE-1 retrotransposons (Dillon *et al.*, 2015). As BPPB1 and BPPB2 samples contained part of LINE-1 elements, their circular molecules could be a microDNA contained in the oral metagenome.

With the hypothesis that integration could occur between two circular structures, the DNA secondary structures around the GATC sites on each insert were then also predicted by Mfold (Figure 6-19). It was shown that the GATC site from all inserts located on the stem-loop secondary structures. Therefore, recombination could occur between two GATC sites, located on DNA secondary structures, leading to the fusion of both structures.

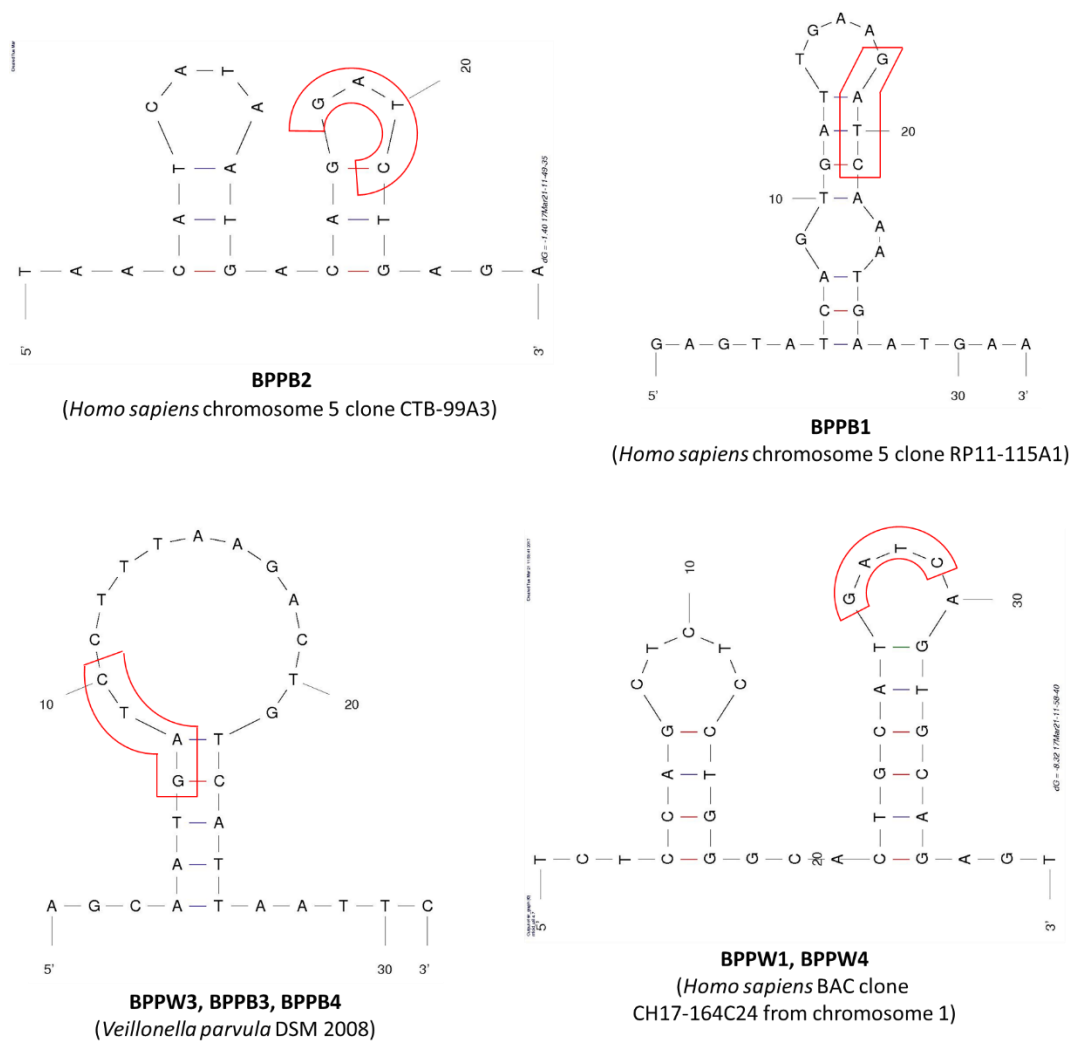


Figure 6-19 The DNA secondary structures predicted by Mfold on GATC inserts. The red boxes indicate the location of GATC sites.

6.5 Conclusion

In this study, the entrapment vectors were developed for capturing TEs from the oral metagenome. Our entrapment vectors could capture the *E. coli* host TEs. The next step is to optimise the conditions for the capturing of TEs from oral metagenomic DNA. During the experiment, we observed the insertions of human and bacterial DNA which were not likely to be caused by a classical transposition or recombination.

Chapter 7

Reduced susceptibility to Quaternary Ammonium Compounds (QACs) is conferred by a heterologous housekeeping gene.

7.1 Introduction

Antimicrobial substances have been extensively used in various industries and are also present in personal hygiene products such as soap, detergents, mouthwash and toothpaste (Russell, 2003). The uses of antimicrobials, especially inappropriate uses (over-applied or using too low concentration), contribute as a selective pressure for bacteria to develop resistance to antimicrobials (Gullberg *et al.*, 2011).

The human oral cavity is an environment that is constantly exposed to antimicrobials in food (such as sodium nitrite in cured meat products, sodium propionate in bakery products and sodium benzoate in acidic foods) and oral care products (such as chlorhexidine) (Marsh, 2010, Carocho *et al.*, 2014). In order to detect novel resistance genes from the entire microbiota, a functional metagenomic screening can be performed. This can be done by constructing a metagenomic library in a surrogate bacterial host and subsequently screen for clones with a gain of resistant phenotype, as described in section 1.4.2.2.

In this study, we performed functional screening on human oral metagenomic libraries against six antimicrobials which oral bacteria are likely to come into regular contact with. Copper sulphate and silver nitrate were chosen because copper and silver metals are present in dental amalgam fillings. Cetylpyridinium chloride (CPC), cetyltrimethylammonium bromide (CTAB) and chlorhexidine were included as they are antiseptics widely used in oral hygiene products. Sodium benzoate, a preservative used in food production,

was also selected. We have identified one clone with a reduced susceptibility to two common quaternary ammonium compounds (QACs): CTAB and CPC.

7.2 Materials and methods

7.2.1 Construction of the human oral metagenomic library

The human oral metagenomic library was constructed by Dr Liam Reynolds, UCL Eastman Dental Institute. The human oral metagenomic DNA was partially digested with *HindIII* as described previously (Seville *et al.*, 2009). The digestion reactions were set up with 40 μ l total volume, containing 1 μ l *HindIII* (4 unit/ μ l), 4 μ l 10X digestion buffer, 22 μ l oral metagenome and 13 μ l molecular grade water. Two reactions were prepared and incubated at 37°C for 1 min and 2 min, respectively. Samples were then combined and purified by isopropanol precipitation (section 2.4.9). The DNA pellet was air-dried and resuspended in 30 μ l of molecular grade water.

The digested oral metagenome was ligated to the pre-digested *HindIII* pCC1BAC vector and electroporated into *E. coli* EPI300 following the protocol in section 2.4.12 and 2.4.15.2. White colonies were then subcultured into an individual well of a 96-well plate containing LB with chloramphenicol. After overnight incubation, 20% glycerol stocks were made in 96-well plates and kept at -80°C.

7.2.2 Preparation of antimicrobial stock solutions

The stock concentrations of each antimicrobial were: copper sulphate (64 mg/ml), silver nitrate (20 mg/ml), CTAB (10 mg/ml), CPC (10 mg/ml), chlorhexidine (10 mg/ml) and sodium benzoate (144 mg/ml). The stock solutions of antimicrobial were prepared by dissolving antimicrobial compounds in either sterilised water (copper sulphate, silver nitrate, chlorhexidine and sodium benzoate) or 70% ethanol (CTAB and CPC). The water-dissolved stock solutions were sterilised by filtration through syringe filters (0.22- μ m pore size).

7.2.3 Minimum Inhibitory Concentration Determination

The minimum inhibitory concentration (MIC) of antimicrobials was determined by following the broth microdilution method described previously (Wiegand *et al.*, 2008). The overnight culture was prepared in LB broth supplemented with chloramphenicol and diluted to OD₆₀₀ of 0.1. In a 96-well plate, 90 μ l of LB broth, containing 12.5 μ g/ml chloramphenicol and different concentration of antimicrobials (diluted from the stock concentrations in section 7.2.2), and 10 μ l of diluted *E. coli* were added to the wells to make a total 100 μ l volume then incubated at 37°C with shaking for 16 hrs.

7.2.4 Screening of metagenomic library and resistant clone isolation

The human oral metagenomic library was screened against antimicrobial substances according to the MIC determined in section 7.3.1 (Table 7-1). An ethanol sterilized 96-pin replicator was used to inoculate the *E. coli* library into

a 96-well plate containing 100 ml of LB broth supplemented with chloramphenicol and antimicrobials. The inoculated 96-well plates were then incubated in a 37°C shaker for 18 hr and determined for resistant clones.

7.2.5 Plasmid extraction and sequencing of resistance genes

The resistant clones were subcultured into 5ml of LB broth containing chloramphenicol and the antimicrobial that they showed resistance to, followed by an incubation in 37°C shaker for 18 hours. The CopyControl induction reaction was then prepared by transferring 1 mL of overnight culture into 9 mL of LB broth supplemented with chloramphenicol (12.5 µg/ml) and 1000X CopyControl™ Induction Solution (10 µl in 10 ml) (Cambio, Cambridge, UK), then incubated in 37°C shaker for 5 hours. Plasmids were extracted from the CopyControl induction reactions following the protocol in section 2.4.3. The insert DNA on the plasmid was sequenced with pCC-F and pCC-R primers (Appendix 4). Additional primers were also designed and used to extend the sequencing.

7.2.6 Transposon mutagenesis

To identify the genes conferring resistance in the large insert, transposon mutagenesis was performed with Template Generation System II kit (Thermo scientific, Surrey, UK), following the manufacturer's instructions. The reaction contained 60 fmoles of extracted plasmid, 1 µl MuA Transposase (0.22 µg/µl), 4 µl 5X reaction buffer, 1 µl Entraceposon (Kan^R-3) (20 ng/µl) and topped up to 20 µl with molecular grade water. The reaction was mixed by pipetting and incubated at 30°C for 1 hr. The reaction was stopped by incubating at 75°C

for 10 min. The mutagenesis reaction was introduced into *E. coli* EPI300 by electroporation, as described in section 2.4.15.2. The transformation reaction contained 5 μ l of the 10-fold diluted mutagenesis reaction and 50 μ l of electrocompetent cells. Cells were spread on LB agar supplement with chloramphenicol. Each colony was then subcultured on the LB agar supplemented with chloramphenicol and CTAB to screen for clones with the loss of phenotype.

7.2.7 Subcloning of udp-4-glucose epimerase and glucose-6-phosphate isomerase genes.

The primers were designed and used to amplify the putative CTAB resistance genes, udp-4-glucose epimerase (*galE*) and glucose-6-phosphate isomerase (*gpi*) genes, in the A10F2 CTAB resistant clone. *Hind*III and *Eco*RI restriction sites were added in the forward and reverse primers, respectively, for directional cloning (Figure 7-1). The amplicons were purified (section 2.4.7) and double digested with *Eco*RI and *Hind*III (section 2.4.10). The digested amplicons were ligated to pre-digested pCC1BAC vector (section 2.4.12) and electroporated into *E. coli* EPI300 electrocompetent cells (section 2.4.15.2). The cells were spread on LB chloramphenicol agar with and without CTAB.

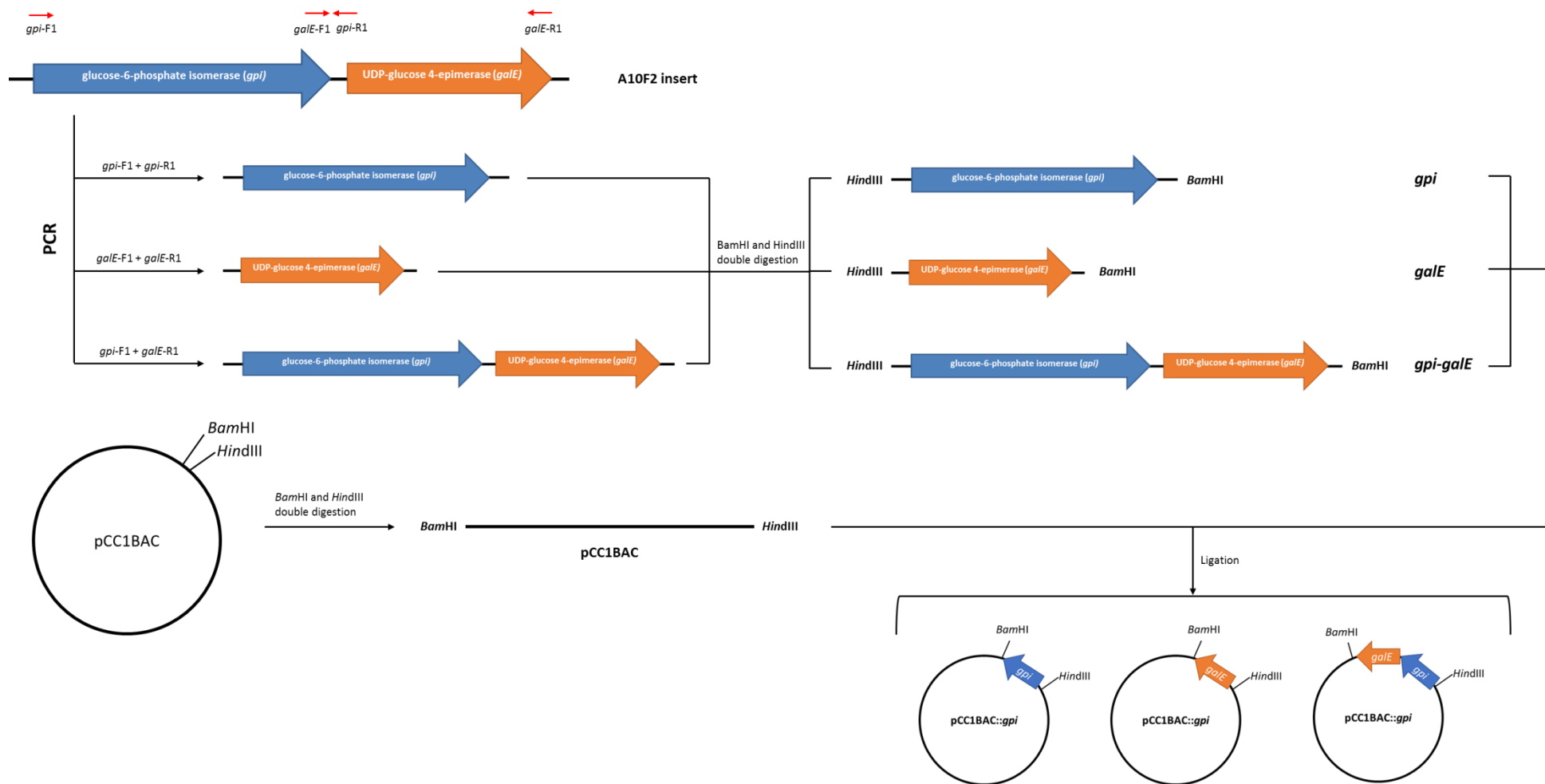


Figure 7-1 Construction of pCC1BAC::*gpi*, pCC1BAC::*galE* and pCC1BAC::*gpi-galE* plasmids. The *gpi*, *galE*, and *gpi-galE* genes were amplified from the A10F2 insert and directionally cloned in between *Hind*III and *Bam*HI restriction sites on pCC1BAC vector. The blue and orange arrow boxes, pointing in the direction of transcription, represent *gpi* and *galE* genes, respectively.

7.2.8 Alignment of GalE amino acid sequences and protein structures

The amino acid sequence of A10F2 GalE was aligned with the sequences from *Veillonella parvula* (Accession number WP_060918982.1) and *E. coli* (Accession number EFJ53418.1) by using Clustal Omega (Sievers *et al.*, 2011). The substrate binding sites and nicotinamide adenine dinucleotide (NAD) binding sites were identified according to the annotation in a protein database.

The protein structure of A10F2 GalE was generated by using SWISS-MODEL (Biasini *et al.*, 2014), which searched and built the target structure based on related evolutionary structures in the protein database. The A10F2 GalE was superimposed with the *E. coli* GalE structure (PDB ID: 1XEL) by using Pymol software (Delano, 2002).

7.2.9 Lipopolysaccharide (LPS) extraction and purification

The lipopolysaccharide (LPS) was extracted as described by the previous study with some modifications (Rezania *et al.*, 2011). A 1.2 ml culture with the OD₆₀₀ of 1.00 was prepared in a microcentrifuge tube from an overnight culture of *E. coli*. Cells were pelleted by centrifugation at 10000 x *g* for 5 min, which was then washed twice with phosphate-buffered saline (PBS) (pH = 7.2, 0.15 M) containing 0.15 mM CaCl₂ and 0.5 mM MgCl₂ (Appendix 2). The cells were resuspended in 5 ml of PBS and sonicated for 2 min on ice. Proteinase K (100 µg/mL) (Thermo Scientific, Surrey, UK) was added to the cell mixture and incubated the tube at 65°C for 1 hr to eliminate contaminating proteins. The nucleic acids were then removed from the samples by adding RNase (40

$\mu\text{g/mL}$) (Qiagen, Manchester, UK) and DNase ($20 \mu\text{g/mL}$) (Thermo Scientific, Surrey, UK) in the presence of $1 \mu\text{L/mL}$ 20% MgSO_4 and $4 \mu\text{L/mL}$ chloroform. The samples were incubated at 37°C for 18 hr.

An equal amount (5 ml) of pre-warmed 90% phenol ($65\text{--}70^\circ\text{C}$) was added to the mixtures, followed by incubation at 65°C for 15 min and vortexing the tube every 5 min during the incubation. The tubes were then incubated on ice and centrifuged at $4500 \times g$ (5000 rpm, Eppendorf centrifuge 5804 R) for 15 min. The supernatant was transferred to a 50 mL conical centrifuge tube. Sodium acetate (0.5 M final concentration), $300 \mu\text{L}$ distilled water and 10 volumes of 95% ethanol were added to the tube, followed by incubating at -20°C for 18 hr. The tube was centrifuged at $2000 \times g$ 4°C for 10 min, and the supernatant was discarded. The pellets were resuspended in $800 \mu\text{L}$ distilled water and transferred to dialysis column (Sigma-Aldrich, Dorset, UK). The dialysis was then performed at 4°C for 18 hr against distilled water to remove the residual phenol from the sample.

To characterise the extracted LPS samples, the samples were sent for Mass Spectrometry at UCL School of Pharmacy, London, UK to be analysed with positive electrospray ionisation time-of-flight mass spectrometry (TOF MS ES+). The results were shown as a plot between the mass-to-charge ratio (m/z) and the relative abundance (%).

7.2.10 Cytochrome *c* binding assay

Cytochrome *c* binding assay was performed to determine the bacterial cell surface charge, following the protocol described in Peschel *et al.*, 1999 with

some modifications. The bacterial strains were subcultured into LB broth with chloramphenicol and incubated for 18 hr. The overnight culture was then transferred into fresh LB broth supplemented with chloramphenicol and incubated for 3 hrs (mid-log phase). The bacterial cells were then collected by centrifugation at 4500 x g (5000 rpm, Eppendorf centrifuge 5804 R) for 10 min. The cells were washed twice with 20 mM MOPS buffer (Sigma-Aldrich, Dorset, UK) at pH 7 (Appendix 2). Cells were then serial diluted into each sterile plastic tube to the final OD₆₀₀ from 1 to 7 and resuspended in 20 mM MOPS buffer supplemented with 150 µg/mL cytochrome c (Sigma, Dorset, UK). The tubes were incubated at room temperature (25°C) with shaking condition at 200 rpm for 10 min. The cells were pelleted by centrifugation at 4500 x g for 10 min, and the absorbance of the supernatant was measured at 530nm.

The amount of cytochrome c bound to the cells was calculated by comparing to the absorbance of the supernatant to the absorbance of the 150 µg/mL cytochrome c stock solution. The amount of cytochrome c bound to the cells were then calculated by using Equation 7-1 and Equation 7-2.

Equation 7-1: The number of unbound cytochrome c = $\frac{\text{Supernatant OD}_{530}}{\text{Stock OD}_{530}} \times 150 \left(\frac{\mu\text{g}}{\text{mL}} \right)$

Equation 7-2: The amount of cytochrome c bound to the cells = $150 \left(\frac{\mu\text{g}}{\text{mL}} \right) - \text{Unbound cytochrome c} \left(\frac{\mu\text{g}}{\text{mL}} \right)$.

Three biological replicates of the assays were performed. Statistical analysis was performed by using Graphpad Prism 6 (GraphPad Software Inc., CA, USA). T-test was carried out to determine whether the amounts of cytochrome *c* bound between an experimental group and a control group were statistically significant different from each other with the *p*-value < 0.05 (*) or *p*-value < 0.005 (**).

7.3 Results

7.3.1 Screening of metagenomic library against antimicrobial compounds

Prior to the screening, the baseline MIC for each antimicrobial against *E. coli* EPI300 containing empty pCC1BAC was determined with broth microdilution method by Mr Gianmarco Cristarella, an MSc student at UCL Eastman Dental Institute, as described in 7.2.2. The concentration ranges and MIC values for each antimicrobial compound are shown in Table 7-1, which were used for the screening of the human metagenomic library. Out of 12,227 clones screened against all antimicrobials (by Mr Gianmarco Cristarella), one clone, labelled A10F2, was confirmed for its resistance to CTAB in the subsequent screening.

Table 7-1 MIC breakpoints for various antimicrobials against *E. coli* EPI300 containing empty pCC1BAC vector to use for library screening.

Antimicrobial Compound	MIC breakpoints (Range)
Copper Sulphate	4 mg/mL (6.4 – 0.8 mg/mL)
Silver Nitrate	18 µg/mL (40 – 4 µg/mL)
CTAB	8 µg/mL (32 – 2 µg/mL)
CPC	6 µg/mL (10 – 0.5 µg/mL)
Chlorhexidine	1.8 µg/mL (10 – 0.125 µg/mL)
Sodium Benzoate	18 mg/mL (40 – 1 mg/mL)

7.3.2 Characterisation of genes conferring CTAB resistance.

The plasmid from the A10F2 clone was digested with *Hind*III, revealing at least 6 DNA fragments on the gel with the estimated size of 17.1 kb (Figure 7-2). The insert DNA on the A10F2 plasmid was sequenced from both ends of the insert.

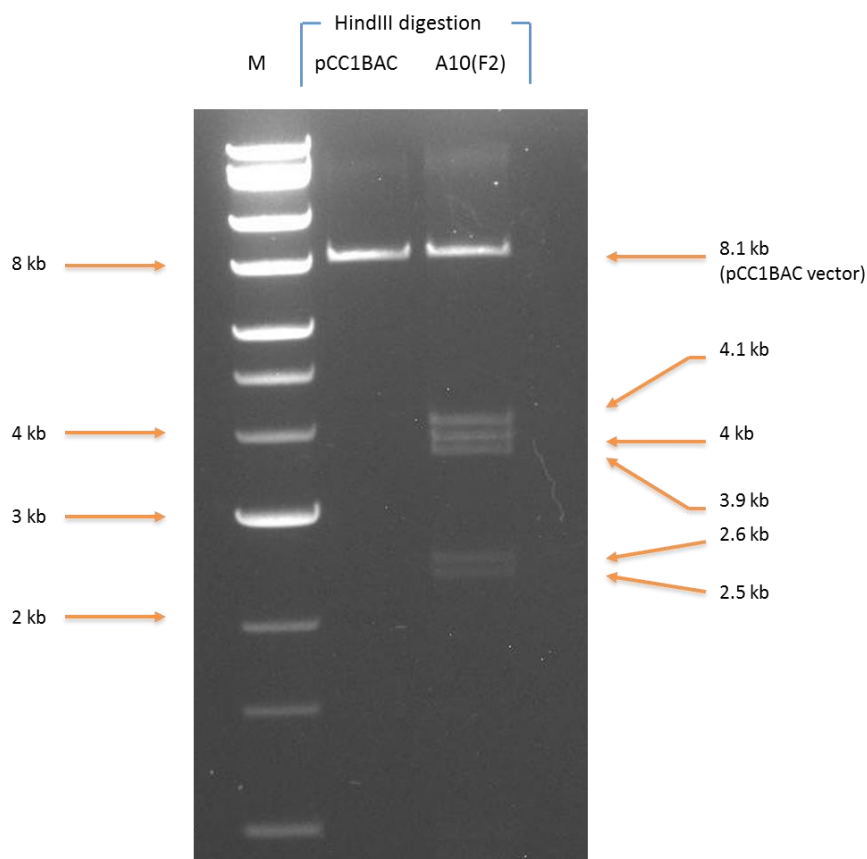


Figure 7-2 *Hind*III digestion products of pCC1BAC vector and A10F2 plasmids were separated on 1% agarose to determine the size of the A10F2 insert. Lane M, Quick-Load® 1 kb Extend DNA Ladder. The brown arrows indicate the size of DNA fragments. (Retrieved from Tansirichaiya *et al.*, 2017)

The sequencing results showed that the insert was likely to be chimeric DNA derived from two different bacterial hosts. The region from one end was most closely related to *Veillonella parvula* (Accession number CP001820.1, 85% nucleotide identity) and the other end was mostly related to *Prevotella melaninogenica* (Accession number CP002122.1, 80% nucleotide identity).

7.3.3 Identification of putative resistance genes by transposon mutagenesis

Transposon mutagenesis was performed to determine the gene(s) conferring CTAB resistance as described in section 7.2.6, which was carried out by Mr Gianmarco Cristarella, UCL Eastman Dental Institute, under my supervision. Ten colonies with the loss of the CTAB resistant phenotype were identified. The plasmids were extracted and sequenced by using primers located at the end of the transposons, revealing that the transposons were inserted in a UDP-glucose 4-epimerase (*galE*) gene in nine clones, and a glucose-6-phosphate isomerase (*gpi*) gene in one clone (Figure 7-3). Both genes were located next to each other, and have 87% nucleotide similarity to DNA from *Veillonella parvula* (accession numbers; CP019721.1 and CP001820.1). The sequence of *galE* and *gpi* was deposited in the nucleotide database with the accession number KY769203.

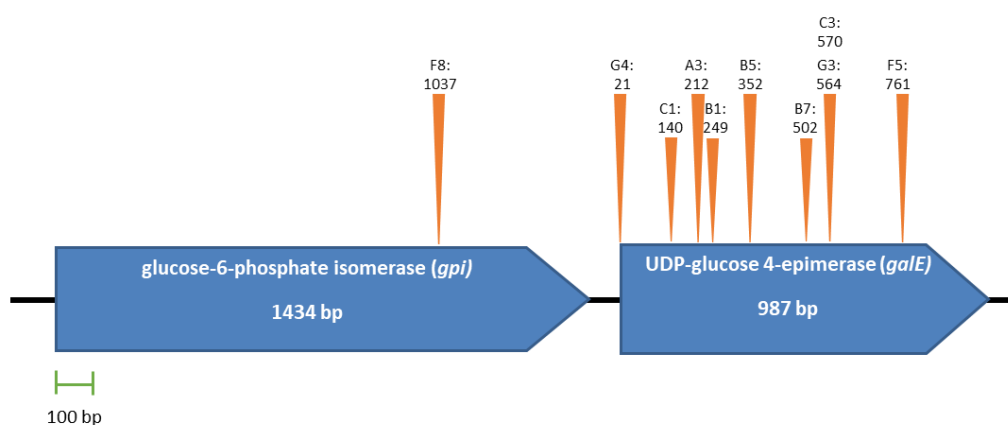


Figure 7-3 The putative genes responsible for CTAB resistances identified by transposon mutagenesis. The putative genes are responsible for the reduced susceptibility to CTAB identified by transposon mutagenesis. The name of mutant clones and the position of the transposons are indicated with orange triangles. The

blue open arrowed boxes represent ORFs, pointing in the probable direction of transcription. (Retrieved from Tansirichaiya *et al.*, 2017)

7.3.4 Subcloning of the putative CTAB resistance genes

The putative CTAB resistance genes were amplified from the A10F2 plasmid and directionally cloned into pCC1BAC vector as described in section 7.2.7 (Figure 7-1 and Figure 7-4). The *galE* and *gpi-galE* transformants showed the CTAB resistant phenotype, while the *gpi* transformant did not. Therefore, it suggested that *galE* was responsible for the CTAB resistance in the A10F2 clone.

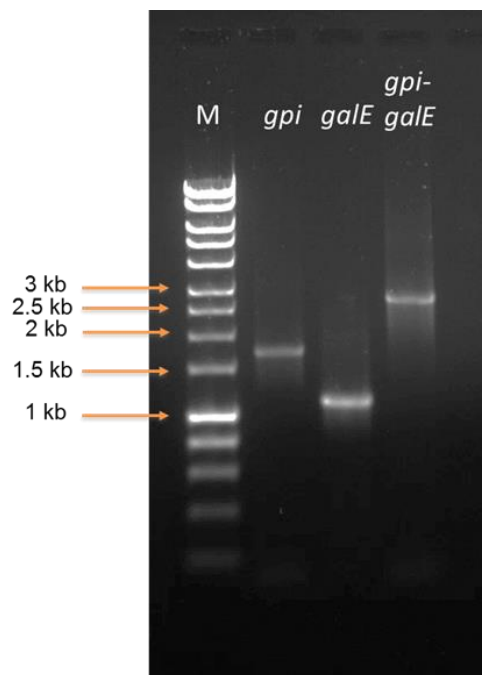


Figure 7-4 The amplification of putative CTAB resistance genes. *gpi*, *galE*, and *gpi-galE* DNA fragments were amplified by PCR and visualised on 1% agarose gel. Lane M, HyperLadder™ 1kb (Bioline, United Kingdom). The size of products were 1727, 1078 and 2609 bp, respectively. (Adapted from Tansirichaiya *et al.*, 2017)

7.3.5 MIC of the resistant clone against CTAB and CPC

The MIC values of CTAB against *E. coli* EPI300 containing pCC1BAC::A10F2, pCC1BAC::galE, and pCC1BAC::galE-gpi increased two-fold (from 8 µg/ml to 16 µg/ml), compared to the *E. coli* EPI300 with empty pCC1BAC (Figure 7-5). As the mode of action of CTAB is similar to CPC (McDonnell & Russell, 1999), the MIC values of CPC were also determined, which showed that the MIC increased from 6 µg/ml to 8 µg/ml, compared to the *E. coli* EPI300 containing the empty pCC1BAC (Figure 7-5). It is noted that the MIC values for CTAB and CPC were the same for all of those nine replicates.

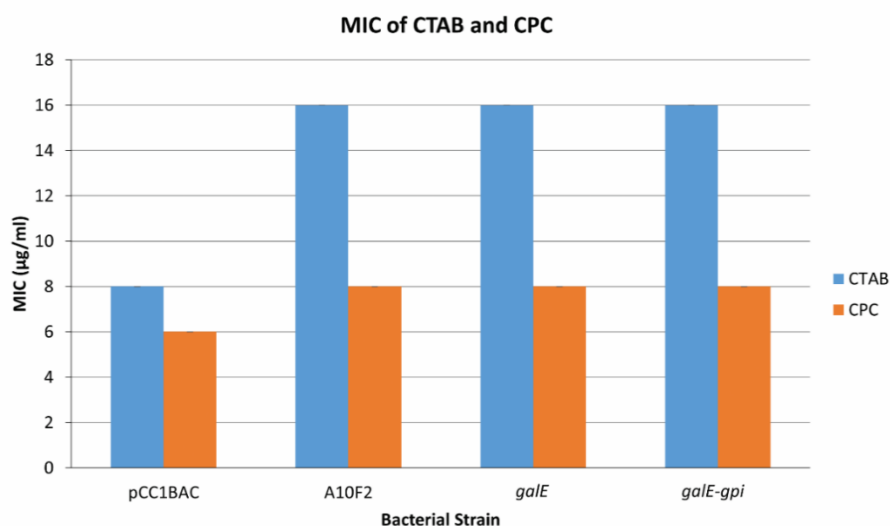


Figure 7-5 Minimum inhibitory concentration of CTAB (blue) and CPC (orange) against *E. coli*::pCC1BAC, *E. coli*::[pCC1BAC::A10F2], *E. coli*::[pCC1BAC::galE], and *E. coli*::[pCC1BAC::galE-gpi] determined by broth microdilution from three independent experiments. Error bars indicate the standard deviation of nine replicates. (Retrieved from Tansirichaiya *et al.*, 2017)

7.3.6 Similarity between A10F2 GalE and *E. coli* GalE proteins

The amino acid sequences of GalE from A10F2 showed homology to GalE from *V. parvula* (Accession number WP_060918982.1) and *E. coli* (Accession number EFJ53418.1) with 97% and 56% similarity, respectively. Amino acid sequence alignment showed that the NAD binding sites and substrate binding sites of all proteins, which are essential for epimerase activity, were located at a similar position (Figure 7-6A).

The protein structure of A10F2 GalE was simulated by SWISS-MODEL based on the best match GalE, which was *Bacillus Anthracis* GalE (PDB ID: 2C20, 54.29% protein sequence similarity). The A10F2 GalE was aligned with the *E. coli* GalE protein structure (Protein database ID: 1XEL) (Thoden *et al.*, 1996) (Figure 7-6B). It was shown that core domain structures (including the NAD binding sites and substrate binding sites) between both proteins are almost identical with some variation in the length and orientation of the secondary structural elements.

A.)

```

A10 (F2) : ---MNILVTGGAGYIGSHTVRLQCAGYTPPIIVDNLSRGHVESIP-----EGVIFYNMDIADPKLV-DIMKDHNIIGVMHFAAHSQVGESMVPATY : 88
V. parvula : ---MNILVTGGAGYIGSHTVRLQCAGYTPPIIVDNLSRGHVESIP-----EGVKFYFNMDIADPKLV-GIMKEHNIIGVMHFAAHSQVGESMVPATY : 88
E. coli : MERINRVLVTGGAGYIGSHTCVQLIINCHDVLIIDNLCNSKRSLVLEVIERLGGKHPTFVEGDIKNEALMTBIIHDEPIIDTVIHFAGLKAVGESVQKLELY : 100

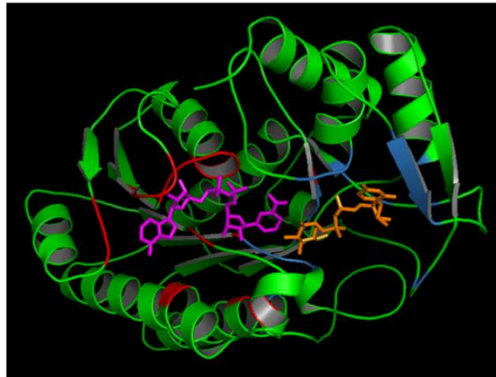
A10 (F2) : YENNVVGSYHLIESARTAGVKHFVFSSTAAVYGEPEVVPVIREI-ALIHPTNVYGRTKLMIEEMLSDYSSIIYGSTYVAL-RYFNAAAGADPSGTIGEDHHPE : 186
V. parvula : YENNVVGSYHLIESGRTAGIKHFVFSSTAAVYGEPEVVPVIREI-ALIQPTNVYGRTKLMIEEMLSDYSSIIYGSTYVAL-RYFNAAAGADPSGMIGEDHHPE : 186
E. coli : YENNVVGTLRLLISAMRPAANVRNFIFFSSAAVYGDQPKPIHYVSEFPTGTETQSHYGKSKLMVEQIILDLQKAQPDWSIALLRYFNFPVGAHPSGDMGED--PC : 198

A10 (F2) : ---THLIPVLDAAARGKREHIVVFGTDYDTADGTQVRRYIHVNDLATAHVLAMDYLRKGGESQVFNLGSGNGFSVKEIIEETAKEVTGIDIPVQYGDERRAG : 283
V. parvula : ---THLIPVLDAAARGKREHIVVFGTDYDTADGTQVRRYIHVNDLATAHVLAMDYLRKGGESQVFNLGSGNGFSVKEIIEETAKEVTGIDIPVQYGDERRAG : 283
E. coli : GIENNLMEYIAQVAVGRRDSIAIFGNLDYFEDGTGVRRYIHVNDLADGHVVAMEKLANKPGVHIYNLCAQVGSVLLVVFVNAFSKACGKKEVNYHFAPRREG : 298

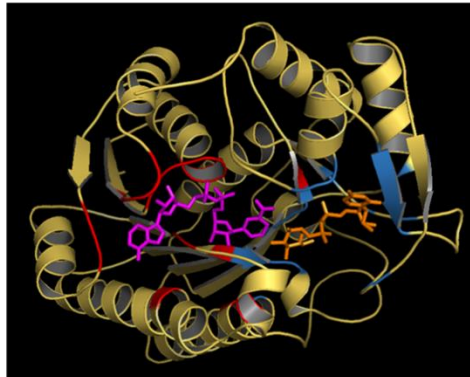
A10 (F2) : DPGTLIASSEKIKNLLGWDPKESNVADVIKDAKWHHTSHPEGFNSK----- : 329
V. parvula : DPGTLIASSEKIKNLLGWDPKESNVADVIKDAKWHHTSHPEGFNSK----- : 329
E. coli : LLPAYWADASKADRELNW--RVRTLDEVAQDTWHWQSRHPQGYPD----- : 342

```

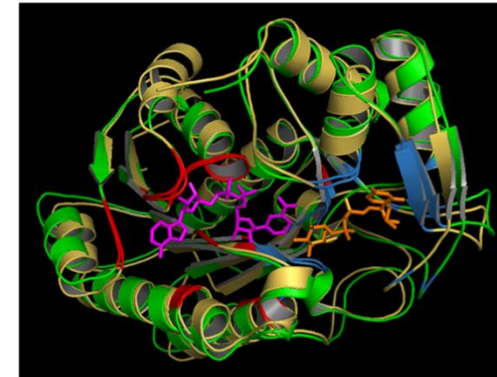
B.)



A10F2 GalE



E. coli GalE



Protein structure alignment of
A10F2 and *E. coli* GalE

Figure 7-6 The alignment of GalE amino acid sequences and protein structures. A.) The amino acid sequences of GalE from A10F2, *V. parvula* and *E. coli* were aligned by using Clustal Omega. The red and blue boxes indicate the NAD binding sites and substrate binding sites, respectively. The black and grey shadings indicate the base pairs with 100% and 50% conservation level. (Retrieved from Tansirichaiya *et al.*, 2017) B.) The protein structure of A10F2 GalE was simulated by SWISS-MODEL (Biasini *et al.*, 2014) and the *E. coli* GalE structure was retrieved from protein database (ID: 1XEL) (Thoden *et al.*, 1996). Both structures were aligned by using Pymol software Version 1.8 (Delano, 2002). The A10F2 GalE, *E. coli* GalE, NAD binding sites and substrate binding sites are shown in green, yellow, red and blue, respectively. The NAD and UDP-4-glucose molecules are shown in purple and orange, respectively.

7.3.7 Characterisation of LPS composition of the CTAB resistance clone

One of the functions of GalE is the interconversion of UDP-glucose and UDP-galactose in the final step of the Leloir pathway in galactose metabolism (Maxwell, 1957). UDP-glucose and UDP-galactose are substrates in the LPS biosynthesis pathway, which is a major component of the outer membrane of Gram-negative bacteria, and also responsible for the overall negative charge on the cell surface. Receiving an insert containing *galE* from *V. parvula* might affect the structure of LPS in *E. coli*, resulting in reduced susceptibility to QACs.

The LPS samples extracted from *E. coli* EPI300 containing an empty pCC1BAC and pCC1BAC::*galE* were, therefore, extracted and analysed with TOF MS ES+, and the results are shown in Figure 7-7. By comparing the MS peaks from pCC1BAC and *galE*, two common peaks can be found at 550.6 and 1051.7. There was also a unique peak that could be found only in the LPS samples extracted from the *E. coli* containing *galE* gene at the m/z value of 1091.7048. This result initially suggested that introducing heterologous *galE* gene into *E. coli* could modify the LPS composition on the cell surface. Tandem mass spectrometry was also performed to characterise the structure of the molecule at 1091.7048. However, the characterisation was failed, possibly due to the degradation of the samples.

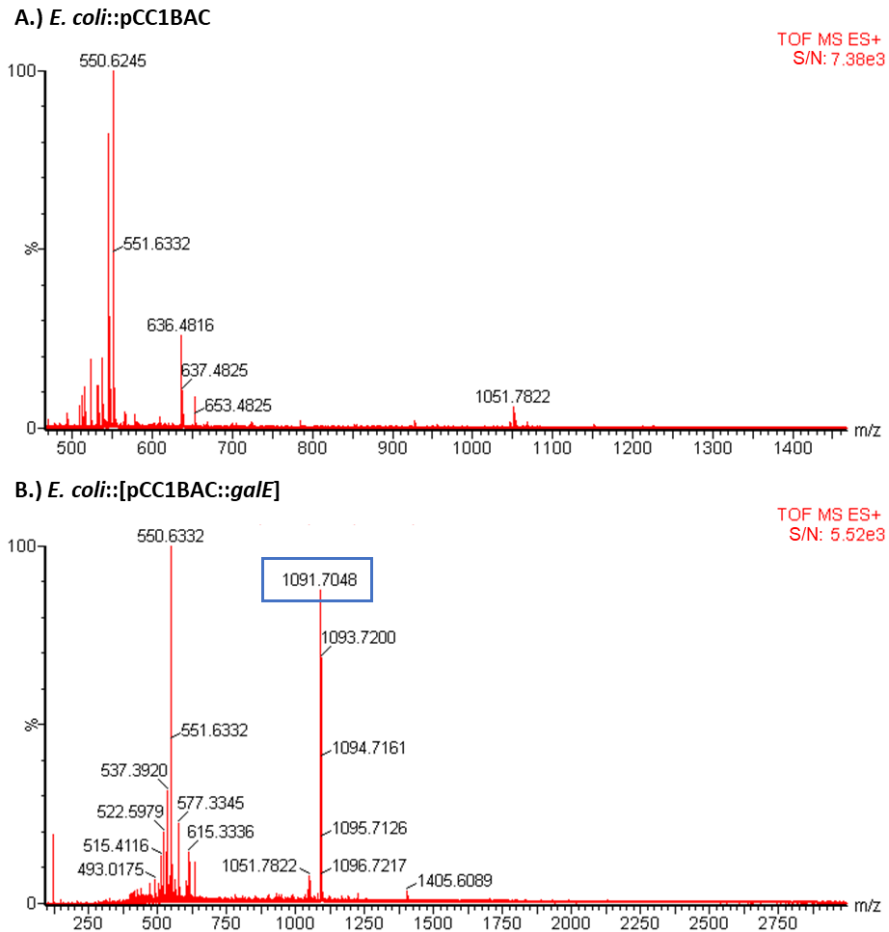


Figure 7-7 Mass spectrometry analysis of the LPS, extracted from *E. coli* containing pCC1BAC and pCC1BAC::*galE*. The samples were analysed with positive electrospray ionisation time-of-flight mass spectrometry (TOF MS ES+). The results were shown as a plot between the mass-to-charge ratio (m/z) and the relative abundance (%). A signal-to-noise ratio (S/N) of each sample indicates the concentration of LPS. The blue boxes indicate the peaks that could be found only in the LPS samples extracted from *galE*-containing *E. coli* cells.

7.3.8 Bacterial cell's surface charge determination

As the LPS characterisation results suggested that there was a change in the LPS composition of *E. coli*::[pCC1BAC::*galE*], which could affect the overall cell surface charge, the changes in cell's surface charge were therefore

determined by performing cytochrome *c* binding assay as described in section 7.2.8 on *E. coli* containing empty pCC1BAC and pCC1BAC::*galE*.

The number of cytochrome *c* bound to the *E. coli* cells containing pCC1BAC::*galE* was lower than the *E. coli* containing empty pCC1BAC at every concentration of the cells (Figure 7-8). As cytochrome *c* is a cationic peptide, which has a high affinity for strongly negatively charged cells, the decrease in binding of cytochrome *c* to *E. coli* with pCC1BAC::*galE* suggested that introducing *galE* from *V. parvula* changed the cell surface of *E. coli* to be less negative. Therefore, CTAB and CPC, which are positively charge compound, would be able to bind less to the cells with *galE*, allowing them to grow in a higher concentration of both QACs.

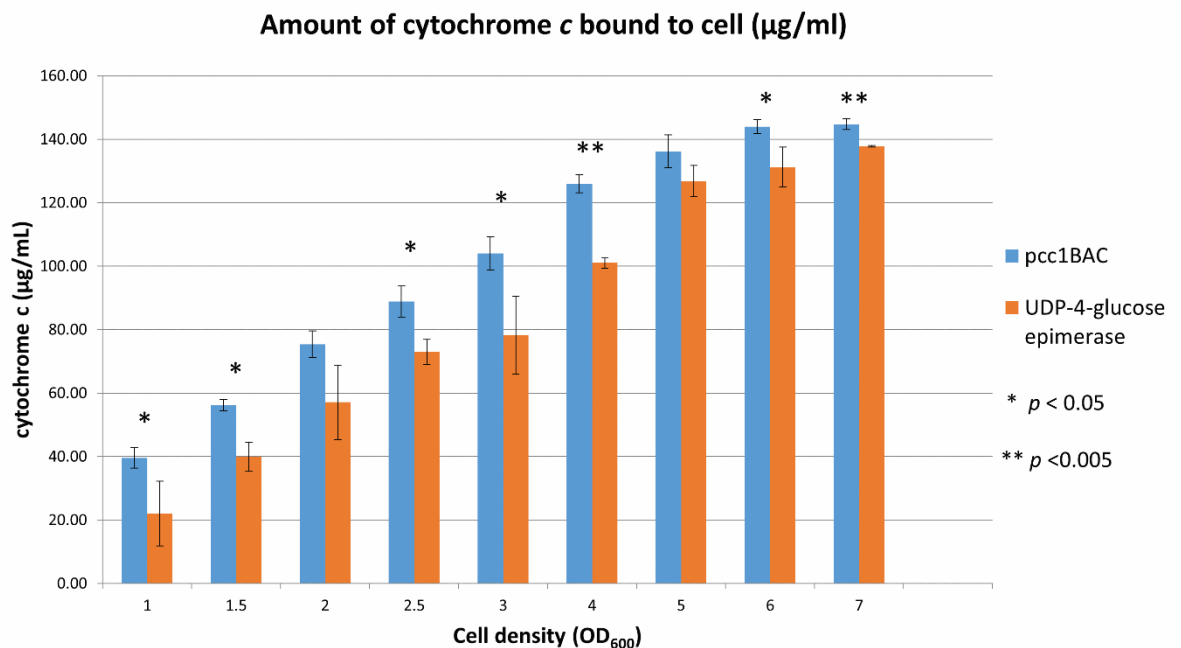


Figure 7-8 The amount of cytochrome *c* bound to *E. coli*. The blue and orange bars represent *E. coli* with pCC1BAC and pCC1BAC::*galE*, respectively. The error bars represent standard deviation from three biological replicates. The asterisks

indicate the statistically significantly difference between the amount of cytochrome *c* bound to *E. coli*::pCC1BAC and *E. coli*::[pCC1BAC::*galE*] at each cell density with the *p*-value <0.05 (*) and <0.005 (**) determined by using t-test. (Retrieved from Tansirichaiya *et al.*, 2017)

7.4 Discussion

Functional-based metagenomics is the approach that has a potential to identify an entirely novel resistance gene, as the screening is based on the phenotypes of interest and no knowledge of known sequences is required. For the human oral metagenome, several novel genes have been isolated by this approach such as the tetracycline resistance gene *tet*(37) and the tetracycline and tigecycline resistance gene *tetAB*(60) (Diaz-Torres *et al.*, 2003, Reynolds *et al.*, 2016). Even though, in this study, the gene conferring the QAC resistance was a known housekeeping *galE* gene, rather than a novel gene, it is the first time that *galE* has been shown to be responsible for QAC resistance.

Both CTAB and CPC are QACs, which are cationic compounds that are commonly used in as surfactants and disinfectants in food preparation industries and consumer products such as cosmetics, hand soaps, oral hygiene products, and disinfectant sprays (Maillard *et al.*, 2013, Gerba, 2015). QACs are widely used as disinfectants due to their broad antimicrobial activity (against bacteria, yeasts, viruses and fungi) and low toxicity. QACs contain a central positive-charged nitrogen surrounding by four alkyl or aryl groups (Figure 7-9) (Gilbert & Moore, 2005). The proposed mode of action of QACs

starts with the absorption and penetration of the compounds into the cell membrane, followed by membrane disorganisation. The intracellular materials are then leaked, leading to the cell lysis by autolytic enzymes (McDonnell & Russell, 1999).

Figure 7-9 Chemical structures of quaternary ammonium compounds (QACs).

A.) A general structure of QACs B.) CTAB C.) CPC (Adapted from Jennings *et al.*, 2015)

Several cases of resistance to QACs have been reported, e.g. in *Listeria* and *Salmonella* (found in food industries) (Holah *et al.*, 2002, Condell *et al.*, 2012, Ortiz *et al.*, 2014). Even though the MIC was increasing in these studies, it was still lower than the concentration used in food processing industries. For example, benzalkonium chloride resistance in *Listeria monocytogenes* was found in different food production chains (meat, fish and poultry productions), which their MIC increased two to eight-fold (8-32 mg/L), but it was still lower than the concentrations used in food production (200-1000 mg/L) (Martínez-Suárez *et al.*, 2016).

Cross-resistance between QACs and antibiotics can occur via efflux pumps and the alteration of the outer membrane (Ishikawa *et al.*, 2002, Soumet *et al.*, 2012, Ortega Morente *et al.*, 2013). The association between CTAB resistance genes and mobile genetic elements has been reported previously. For example, the small multidrug resistance *qrg* gene, encoding efflux protein, was

previously identified as part of conjugative transposon Tn6087 (Ciric *et al.*, 2011), IS 1216-composite transposon and IS 1216-translocatable unit (Chapter 5). The transfer of *qrg* gene carried by Tn6087 to another streptococcal strain has been demonstrated by DNA transformation (Ciric *et al.*, 2011).

UDP-glucose 4-epimerases (GalE) can be found in bacteria, fungi, plants and animals. They catalyse a reversible epimerisation of UDP-glucose and UDP-galactose by inverting the configuration of the hydroxyl (-OH) group on carbon C-4 (Figure 7-10). UDP-glucose can be subsequently converted to UDP-glucuronate by a UDP-galactopyranose mutase, while UDP-galactose can be converted to UDP-glucose dehydrogenase by UDP-galactofuranose (Figure 7-10). All of these UDP-sugars are involved in the LPS biosynthesis pathway as they are components of O-antigen and LPS core. As LPS is a major component, responsible for a negative charge on the outer membrane of Gram-negative bacteria, changes in the structure and quantity of LPS on the cell membrane could result in the alteration of the bacterial cell surface charge. For example, the cell surface charge of *Pseudomonas aeruginosa* with a truncated LPS core (mutation in a *rmIC* gene) was shown to be more negative than the wild-type (Rzhepishevskaja *et al.*, 2013).

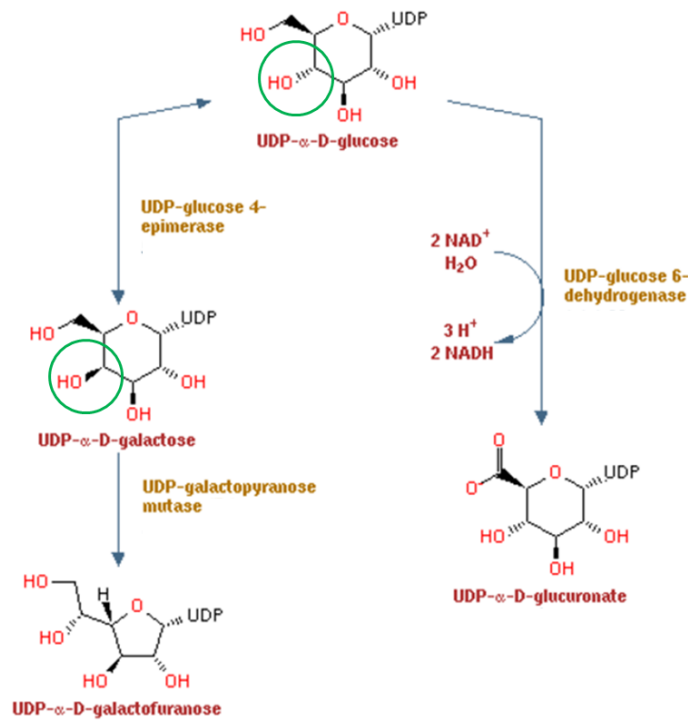


Figure 7-10 The UDP-sugar biosynthesis pathway. GalE catalyses the interconversion of UDP-glucose and UDP-galactose, which can be subsequently converted into UDP-glucuronate and UDP-galactofuranose, respectively. The green circles indicate the OH groups at carbon C-4, which is inverted by GalE (Adapted from Caspi *et al.*, 2012)

The alteration of cell surface charge was previously shown to confer resistance to daptomycin (DAP), a cationic antimicrobial peptide. DAP resistance is conferred by mutations in genes involved in cell wall homeostasis and cell membrane phospholipid metabolism, which resulted in a more positively charged cell surface, repelling and reducing the binding of DAP to the cell membrane (Tran *et al.*, 2013, Tran *et al.*, 2015). For example, daptomycin resistance in *Staphylococcus aureus* was conferred by mutations in *mprF* gene (Yang *et al.*, 2013). MprF catalyses the transfer of lysine residues from lysyl-tRNA to a negatively charged phosphatidylglycerol (PhG), located in the cell membrane, resulting in a positively charged lysyl-PG (L-

PhG). Mutations in *mprF*, resulted in an enhanced L-PhG synthesis, therefore increase the number of L-PhG (positive charge) on the cell surface, repelling the DAP binding to *S. aureus* cells (Yang *et al.*, 2013, Bayer *et al.*, 2015). The depletion of an anionic PhG was also observed in *Bacillus subtilis* with a mutation in *pgsA* gene (Hachmann *et al.*, 2011). As PgsA is required for the first step in PhG synthesis, the reduction in membrane PhG, therefore, diminished the net negative charge of the membrane, weakening the interaction of DAP to bacterial cells.

Mutations in *galE* gene were previously reported to affect the composition of LPS in several studies. For example, in *Campylobacter jejuni*, a mutation in *galE* was shown to reduce the molecular weight of lipid A-core on LPS, and also reduce their ability to adhere and invade human INT 407 (HeLa derivative) cells (Fry *et al.*, 2000). The LPS analysis of *Porphyromonas gingivalis galE* mutant also showed changes in the LPS composition, which had shorter O-antigen chains, compared to the wild-type (Nakao *et al.*, 2006).

In our study, the acquisition of *galE* from *V. parvula* changed the cell surface charge of *E. coli* to be less negative. An alignment of GalE from *V. parvula* and *E. coli* showed the similarity between both enzymes, suggesting that *V. parvula* GalE should be functional in *E. coli*. Introducing an extra copy of *galE* gene from *V. parvula* to *E. coli* will likely result in higher epimerase activity. As the epimerisation catalysed by GalE is a reversible reaction, overexpressing the GalE could affect the balance of substrates for LPS biosynthesis by an altered enzyme kinetic, resulting in the changes on the cell surface charge. The more positive charge on the *E. coli*::[pCC1BAC::*galE*] cell surface will

repel the cationic molecules like QACs, resulting in less binding of QACs to the cells, which is an essential step for the QACs to kill bacteria.

Recently, it has been shown that antimicrobial tolerance (the ability of bacteria to survive without developing resistance) boosts the chances for bacteria to develop resistance because it supports the continued survival of bacteria, which increases the window of opportunity for bacteria to develop rarer mutations (Levin-Reisman *et al.*, 2017). Therefore, the acquisition of a heterologous housekeeping *galE* gene may provide a window of opportunity for bacteria to develop new resistance mechanisms against QACs. The constant exposure to antimicrobials and the highly conducive to horizontal gene transfer of multi-species biofilm in the oral cavity could facilitate such an acquisition event of heterologous housekeeping genes to occur (Roberts & Kreth, 2014).

7.5 Conclusion

In conclusion, we have screened a human oral metagenomic library against various antimicrobials. One clone was identified with QAC resistance phenotype. The resistance was shown to be conferred by the expression of a heterologous housekeeping gene *galE*, which resulted in the alteration of the bacterial cell's surface charge to be less negatively charged. This is the first study that identified QAC resistance gene by the functional metagenomic approach and the first study to demonstrate another possible function of *galE* in conferring QAC resistance. These results have been published in Microbial Drug Resistance (Appendix 5) (Tansirichaiya *et al.*, 2017).

Chapter 8
General Discussion

To understand antimicrobial resistance, the simple model, where pathogens become resistant by mutating to resistance or acquiring resistance genes from other bacteria in their environment, is insufficient to explain the rapidity and spread of resistance. We have seen for example with plasmid-mediated colistin resistance encoded by *mcr-1*, which was first reported in China in 2015 (Liu *et al.*, 2016). Since then, *mcr-1* in *Enterobacteriaceae* has been isolated from animals, animal products, humans and environments (water, hospital sewage and vegetable samples) from over 30 countries (Al-Tawfiq *et al.*, 2017). Therefore, it is complex bacterial interactions, responding to multiple selection pressures and multiple directions of gene flow, which allows bacteria to develop resistance rapidly. Each environment can act as conduits for a global resistance gene flow, shuttling resistance genes into a broader transient bacterial population, mobilised by a myriad of MGEs.

The human oral cavity is also a complex multi-species environment where gene flow is likely to occur. Thus, the human oral microbiome is important, and needs to be understood on a functional level, in order to fully understand the antibiotic resistance crisis. We have shown for the first time that several ARGs in the human oral cavity are associated with integron GCs. These could be primarily maintained in the oral cavity for the bacteria to survive stresses in their environment. However, with a strong selective pressure, they could also have a potential to be selected and transferred to bacterial population in other environments.

In this study, we have gained insights into the fundamental biology of integrons and proven that one of the likely functions of noncoding GCs is to drive

expression of genes within downstream GCs. Our results from an enzymatic assay support the hypothesis that a promoter-containing GC, inserted in the first position in an integron, could act as a genetic clutch, where the expression level of the original first GC can be maintained upon the insertion of a new GC in the first position. This would allow the host bacteria to respond to more than one stress which is the likely situation in the complex physicochemical environment within the human oral cavity.

Sequence analysis on *Vibrio sp.* DAT722 integron GC array also shows that GCs located downstream from promoter GCs (noncoding GCs and TA-containing GCs) were predicted to encode proteins for adaptation such as ribonuclease inhibitor and N-acetyltransferase (possibly conferring aminoglycoside resistance) (Boucher *et al.*, 2006b). The position of promoter GCs and their downstream GCs could also represent a historical record reflecting the episodes of adaptation to stresses, which are sustained over time, from the most recent ones close to the Pc promoter to the earliest further down the array. The disappearance of stresses from the environment would select for cells without the expression of the downstream GC (by shuffling the promoter GCs or the downstream GC down the array). With this, they will have lower metabolic burden or fitness cost, allowing them to be more fit and eventually dominate the population.

A novel resistance mechanism to antiseptic QACs was also shown here for the first time to be associated with a change in cell surface polarity, conferred by a heterologous housekeeping gene *galE* when expressed in *E. coli*. Resistance conferred by a change in cell polarity has been described for

daptomycin resistance but never for cationic antiseptics. The screening of the same human oral metagenomic library also found clones with a reduced susceptibility to triclosan, a common disinfectant, which was conferred by another heterologous housekeeping gene called *fabI*, encoding an enoyl-acyl carrier protein reductase (ENR) (Tansirichaiya *et al.*, 2017). As both resistances were conferred by housekeeping genes, which have main functions in catabolism of D-galactose (*galE*) and fatty acid synthesis (ENR), it suggests that a reservoir of genes conferring resistance, is not only comprised of true ARGs, but it is also can be extended to any gene which may not have a primary role in resistance.

Even though no MGEs were shown to be associated with these heterologous housekeeping genes in our study, they could eventually find their way to be selected and transferred by HGT due to strong selective pressure, provided by the common and widespread use of household disinfectants and hygiene products. This has been shown previously; *fabI* was transferred from *Staphylococcus haemolyticus* to *S. aureus* by IS1272 derived composite transposons, designated as TnSha1 and TnSha2 (Furi *et al.*, 2016).

In conclusion, we have shown that ARGs present on integrons, composite transposons and translocatable units can be readily detected in metagenomic samples derived from the human oral cavity, extending our knowledge about this reservoir of ARGs in the oral microbiome. The biology of integrons has also been expanded, revealing an interesting clutch mechanism for decoupling expression from the primary integron promoter in order to respond to multiple stresses. The techniques and constructs developed in this study

also have the potential to be applied to other metagenomes derived from other environments which will help understand the ebb and flow of MGEs, ARGs and genes emerging as ARGs.

References

Aakre CD, Phung TN, Huang D & Laub MT (2013) A bacterial toxin inhibits DNA replication elongation through a direct interaction with the beta sliding clamp. *Molecular cell* **52**: 617-628.

Abubucker S, Segata N, Goll J, *et al.* (2012) Metabolic reconstruction for metagenomic data and Its application to the human microbiome. *PLOS Computational Biology* **8**: e1002358.

Adhya S & Gottesman M (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**: 939-944.

Akpan A & Morgan R (2002) Oral candidiasis. *Postgraduate Medical Journal* **78**: 455-459.

Al-Shehri SS, Sweeney EL, Cowley DM, Liley HG, Ranasinghe PD, Charles BG, Shaw PN, Vagenas D, Duley JA & Knox CL (2016) Deep sequencing of the 16S ribosomal RNA of the neonatal oral microbiome: a comparison of breast-fed and formula-fed infants. *Scientific reports* **6**: 38309.

Al-Tawfiq JA, Laxminarayan R & Mendelson M (2017) How should we respond to the emergence of plasmid-mediated colistin resistance in humans and animals? *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* **54**: 77-84.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology* **215**: 403-410.

Amarir-Bouhram J, Goin M & Petit M-A (2011) Low efficiency of homology-facilitated illegitimate recombination during conjugation in *Escherichia coli*. *PloS one* **6**: e28876.

Anantham S & Hall RM (2012) pCERC1, a small, globally disseminated plasmid carrying the *dfrA14* cassette in the *strA* gene of the *sul2-strA-strB* gene cluster. *Microbial drug resistance (Larchmont, NY)* **18**: 364-371.

Antonelli A, D'Andrea MM, Vaggelli G, Docquier JD & Rossolini GM (2015) OXA-372, a novel carbapenem-hydrolysing class D β -lactamase from a

Citrobacter freundii isolated from a hospital wastewater plant. *The Journal of antimicrobial chemotherapy* **70**: 2749-2756.

Avery OT, Macleod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *The Journal of experimental medicine* **79**: 137-158.

Azpiroz MF, Bascuas T & Laviña M (2011) Microcin H47 system: an *Escherichia coli* small genomic island with novel features. *PLoS one* **6**: e26179.

Baharoglu Z & Mazel D (2011) *Vibrio cholerae* triggers SOS and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. *Antimicrobial agents and chemotherapy* **55**: 2438-2441.

Baharoglu Z, Bikard D & Mazel D (2010) Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLOS Genetics* **6**: e1001165.

Baharoglu Z, Krin E & Mazel D (2012) Connecting environment and genome plasticity in the characterization of transformation-induced SOS regulation and carbon catabolite control of the *Vibrio cholerae* integron integrase. *Journal of bacteriology* **194**: 1659-1667.

Balbas P, Soberon X, Merino E, Zurita M, Lomeli H, Valle F, Flores N & Bolivar F (1986) Plasmid vector pBR322 and its special-purpose derivatives--a review. *Gene* **50**: 3-40.

Baltrus DA & Guillemin K (2006) Multiple phases of competence occur during the *Helicobacter pylori* growth cycle. *FEMS microbiology letters* **255**: 148-155.

Bartosik D, Sochacka M & Baj J (2003) Identification and characterization of transposable elements of *Paracoccus pantotrophus*. *Journal of bacteriology* **185**: 3753-3763.

Bayer AS, Mishra NN, Chen L, Kreiswirth BN, Rubio A & Yang S-J (2015) Frequency and distribution of single-nucleotide polymorphisms within *mprF* in

methicillin-resistant *Staphylococcus aureus* clinical isolates and their role in cross-resistance to daptomycin and host defense antimicrobial peptides. *Antimicrobial agents and chemotherapy* **59**: 4930-4937.

Bayley DP, Rocha ER & Smith CJ (2000) Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS microbiology letters* **193**: 149-154.

Bebear C, Pereyre S & Peuchant O (2011) *Mycoplasma pneumoniae*: susceptibility and resistance to antibiotics. *Future microbiology* **6**: 423-431.

Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, Pignatelli M & Mira A (2012) The oral metagenome in health and disease. *The ISME journal* **6**: 46-56.

Beloqui A, Nechitaylo TY, Lopez-Cortes N, *et al.* (2010) Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. *Applied and environmental microbiology* **76**: 5934-5946.

Bennett PM (2008) Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology* **153**: S347-S357.

Berezow AB & Darveau RP (2011) Microbial shift and periodontitis. *Periodontology 2000* **55**: 36-47.

Bhatty M, Laverde Gomez JA & Christie PJ (2013) The expanding bacterial type IV secretion lexicon. *Research in Microbiology* **164**: 620-639.

Biasini M, Bienert S, Waterhouse A, *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42**: W252-W258.

Bieber D, Ramer SW, Wu CY, Murray WJ, Tobe T, Fernandez R & Schoolnik GK (1998) Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science (New York, NY)* **280**: 2114-2118.

Biskri L & Mazel D (2003) Erythromycin esterase gene *ere(A)* is located in a functional gene cassette in an unusual class 2 integron. *Antimicrobial agents and chemotherapy* **47**: 3326-3331.

Biskri L, Bouvier M, Guerout AM, Boissard S & Mazel D (2005) Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *Journal of bacteriology* **187**: 1740-1750.

Boucher Y, Labbate M, Koenig JE & Stokes HW (2007) Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends in microbiology* **15**.

Boucher Y, Nesbo CL, Joss MJ, Robinson A, Mabbutt BC, Gillings MR, Doolittle WF & Stokes HW (2006a) Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC evolutionary biology* **6**.

Boucher Y, Nesbø CL, Joss MJ, Robinson A, Mabbutt BC, Gillings MR, Doolittle WF & Stokes HW (2006b) Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC evolutionary biology* **6**: 3-3.

Bouvier M, Demarre G & Mazel D (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *The EMBO journal* **24**: 4356-4367.

Bouvier M, Ducos-Galand M, Loot C, Bikard D & Mazel D (2009) Structural features of single-stranded integron cassette *attC* sites and their role in strand selection. *PLOS Genetics* **5**: e1000632.

Bow EJ (2013) There should be no ESKAPE for febrile neutropenic cancer patients: the dearth of effective antibacterial drugs threatens anticancer efficacy. *The Journal of antimicrobial chemotherapy* **68**: 492-495.

Brantl S (2012) Bacterial type I toxin-antitoxin systems. *RNA biology* **9**: 1488-1490.

Bratzler DW, Dellinger EP, Olsen KM, *et al.* (2013) Clinical practice guidelines for antimicrobial prophylaxis in surgery. *American journal of health-system pharmacy* **70**: 195-283.

Brett PJ, Burtnick MN, Fenno JC & Gherardini FC (2008) *Treponema denticola* TroR is a manganese- and iron-dependent transcriptional repressor. *Molecular microbiology* **70**: 396-409.

Brochet M, Da Cunha V, Couve E, Rusniok C, Trieu-Cuot P & Glaser P (2009) Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Molecular microbiology* **71**: 948-959.

Brooks GF, Carroll KC, Butel JS, Morse SA & Mietzner TA (2012) *Jawetz, Melnick, & Adelberg's Medical Microbiology, Twenty-Fifth Edition*. McGraw-Hill Education.

Brown Kav A, Benhar I & Mizrahi I (2013) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *Journal of microbiological methods* **95**: 272-279.

Brown L, Wolf JM, Prados-Rosales R & Casadevall A (2015) Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nature Reviews Microbiology* **13**: 620-630.

Brown NL & Evans LR (1991) Transposition in prokaryotes: transposon Tn501. *Research in Microbiology* **142**: 689-700.

Burrows LL (2012) *Pseudomonas aeruginosa* twitching motility: type IV pili in action. *Annual review of microbiology* **66**: 493-520.

Bzymek M & Lovett ST (2001) Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 8319-8325.

Calero-Cáceres W & Muniesa M (2016) Persistence of naturally occurring antibiotic resistance genes in the bacteria and bacteriophage fractions of wastewater. *Water Research* **95**: 11-18.

Callen BP, Shearwin KE & Egan JB (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular cell* **14**: 647-656.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML & Brussow H (2003) Phage as agents of lateral gene transfer. *Current opinion in microbiology* **6**: 417-424.

Card RM, Warburton PJ, MacLaren N, Mullany P, Allan E & Anjum MF (2014) Application of microarray and functional-Based screening methods for the detection of antimicrobial resistance genes in the microbiomes of healthy humans. *PloS one* **9**: e86428.

Carlsson J (1997) Bacterial metabolism in dental biofilms. *Advances in dental research* **11**: 75-80.

Carocho M, Barreiro MF, Morales P & Ferreira ICFR (2014) Adding molecules to food, pros and cons: a review on synthetic and natural food additives. *Comprehensive Reviews in Food Science and Food Safety* **13**: 377-399.

Caspi R, Altman T, Dreher K, *et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research* **40**: D742-D753.

Chalmers R, Guhathakurta A, Benjamin H & Kleckner N (1998) IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93**: 897-908.

Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A & Dewhirst FE (2010) The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database : the journal of biological databases and curation* **2010**: baq013.

Chen X-L, Tang D-J, Jiang R-P, He Y-Q, Jiang B-L, Lu G-T & Tang J-L (2011) sRNA-Xcc1, an integron-encoded transposon- and plasmid-transferred trans-acting sRNA, is under the positive control of the key virulence regulators HrpG

and HrpX of *Xanthomonas campestris* pathovar *campestris*. *RNA Biology* **8**: 947-953.

Chiang P & Burrows LL (2003) Biofilm formation by hyperpilated mutants of *Pseudomonas aeruginosa*. *Journal of bacteriology* **185**: 2374-2378.

Chow VCY, Hawkey PM, Chan EWC, Chin ML, Au TK, Fung DKC & Chan RCY (2007) High-level gentamicin resistance mediated by a Tn4001-like transposon in seven nonclonal hospital isolates of *Streptococcus pasteurianus*. *Antimicrobial agents and chemotherapy* **51**: 2508-2513.

Chung CT, Niemela SL & Miller RH (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 2172-2175.

Ciric L, Mullany P & Roberts AP (2011) Antibiotic and antiseptic resistance genes are linked on a novel mobile genetic element: Tn6087. *The Journal of antimicrobial chemotherapy* **66**: 2235-2239.

Ciric L, Brouwer MSM, Mullany P & Roberts AP (2014) Minocycline resistance in an oral *Streptococcus infantis* isolate is encoded by *tet(S)* on a novel small, low copy number plasmid. *FEMS Microbiology Letters* **353**: 106-115.

Clewell DB, Weaver KE, Dunny GM, Coque TM, Francia MV & Hayes F (2014) *Extrachromosomal and mobile elements in enterococci: transmission, maintenance, and epidemiology*. Boston, MA.

Cochetti I, Tili E, Mingoia M, Varaldo PE & Montanari MP (2008) *erm(B)*-carrying elements in tetracycline-resistant Pneumococci and correspondence between Tn1545 and Tn6003. *Antimicrobial agents and chemotherapy* **52**: 1285-1290.

Coleman N, Tetu S, Wilson N & Holmes A (2004) An unusual integron in *Treponema denticola*. *Microbiology (Reading, England)* **150**: 3524-3526.

Collis CM & Hall RM (1992) Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Molecular microbiology* **6**: 2875-2885.

Collis CM & Hall RM (1995) Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrobial agents and chemotherapy* **39**: 155-162.

Collis CM, Grammaticopoulos G, Briton J, Stokes HW & Hall RM (1993) Site-specific insertion of gene cassettes into integrons. *Molecular microbiology* **9**: 41-52.

Collis CM, Recchia GD, Kim MJ, Stokes HW & Hall RM (2001) Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *Journal of bacteriology* **183**: 2535-2542.

Condell O, Iversen C, Cooney S, Power KA, Walsh C, Burgess C & Fanning S (2012) Efficacy of biocides used in the modern food industry to control *Salmonella enterica*, and links between biocide tolerance and resistance to clinically relevant antimicrobial compounds. *Applied and environmental microbiology* **78**: 3087-3097.

Connell SR, Tracz DM, Nierhaus KH & Taylor DE (2003) Ribosomal protection proteins and their mechanism of tetracycline resistance. *Antimicrobial agents and chemotherapy* **47**: 3675-3681.

Corby PM, Biesbrock A, Bartizek R, Corby AL, Monteverde R, Ceschin R & Bretz WA (2008) Treatment outcomes of dental flossing in twins: molecular analysis of the interproximal microflora. *Journal of periodontology* **79**: 1426-1433.

Cox G & Wright GD (2013) Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *International journal of medical microbiology : IJMM* **303**: 287-292.

Coyne S, Guigon G, Courvalin P & Perichon B (2010) Screening and quantification of the expression of antibiotic resistance genes in *Acinetobacter*

baumannii with a microarray. *Antimicrobial agents and chemotherapy* **54**: 333-340.

Craig JW, Chang FY, Kim JH, Obiajulu SC & Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and environmental microbiology* **76**: 1633-1641.

Craig NL (1991) Tn7: a target site-specific transposon. *Molecular microbiology* **5**: 2569-2573.

Csörgő B, Fehér T, Tímár E, Blattner FR & Pósfai G (2012) Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microbial Cell Factories* **11**: 11-11.

d'Aubenton Carafa Y, Brody E & Thermes C (1990) Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *Journal of molecular biology* **216**: 835-858.

da Fonseca EL & Vicente AC (2012) Functional characterization of a Cassette-specific promoter in the class 1 integron-associated *qnrVC1* gene. *Antimicrobial agents and chemotherapy* **56**: 3392-3394.

Davies J & Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR* **74**: 417-433.

Davis BD (1950) Nonfiltrability of the agents of genetic recombination in *Escherichia coli*. *Journal of bacteriology* **60**: 507-508.

De Meirsmen C, Croes C, Desair J, Verreth C, Van Gool A & Vanderleyden J (1989) Identification of insertion sequence element IS427 in pTiT37 plasmid DNA of an *Agrobacterium tumefaciens* T37 isolate. *Plasmid* **21**: 129-137.

de Vries J & Wackernagel W (2002) Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate

recombination. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2094-2099.

De Vries J, Herzfeld T & Wackernagel W (2004) Transfer of plastid DNA from tobacco to the soil bacterium *Acinetobacter* sp. by natural transformation. *Molecular microbiology* **53**: 323-334.

de Vries LE, Valles Y, Agero Y, Vaishampayan PA, Garcia-Montaner A, Kuehl JV, Christensen H, Barlow M & Francino MP (2011) The gut as reservoir of antibiotic resistance: microbial diversity of tetracycline resistance in mother and infant. *PloS one* **6**: e21644.

Deak D, Outterson K, Powers JH & Kesselheim AS (2016) Progress in the fight against multidrug-resistant bacteria? a review of u.s. food and drug administration–approved antibiotics, 2010–2015. *Annals of Internal Medicine* **165**: 363-372.

Deich RA & Smith HO (1980) Mechanism of homospecific DNA uptake in *Haemophilus influenzae* transformation. *Molecular Genetics and Genomics* **177**: 369-374.

Del Grosso M, Camilli R, Libisch B, Füzi M & Pantosti A (2009) New composite genetic element of the Tn916 family with dual macrolide resistance genes in a *Streptococcus pneumoniae* isolate belonging to clonal complex 271. *Antimicrobial agents and chemotherapy* **53**: 1293-1294.

Delano WL (2002) The PyMOL molecular graphics system version 1.8. *Schrödinger, LLC*.

Demain AL (2009) Antibiotics: natural products essential to human health. *Medicinal Research Reviews* **29**: 821-842.

Deng X, Sun F, Ji Q, Liang H, Missiakas D, Lan L & He C (2012) Expression of multidrug resistance efflux pump gene *norA* is iron responsive in *Staphylococcus aureus*. *Journal of bacteriology* **194**: 1753-1762.

Depardieu F, Podglajen I, Leclercq R, Collatz E & Courvalin P (2007) Modes and modulations of antibiotic resistance gene expression. *Clinical microbiology reviews* **20**: 79-114.

Devos S, Van Oudenhove L, Stremersch S, Van Putte W, De Rycke R, Van Driessche G, Vitse J, Raemdonck K & Devreese B (2015) The effect of imipenem and diffusible signaling factors on the secretion of outer membrane vesicles and associated Ax21 proteins in *Stenotrophomonas maltophilia*. *Frontiers in microbiology* **6**: 298.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A & Wade WG (2010) The human oral microbiome. *Journal of bacteriology* **192**: 5002-5017.

Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M & Mullany P (2003) Novel tetracycline resistance determinant from the oral metagenome. *Antimicrobial agents and chemotherapy* **47**: 1430-1432.

Diaz PI, Chalmers NI, Rickard AH, Kong C, Milburn CL, Palmer RJ, Jr. & Kolenbrander PE (2006) Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Applied and environmental microbiology* **72**: 2837-2848.

Dillon Laura W, Kumar P, Shibata Y, Wang Y-H, Willcox S, Griffith Jack D, Pommier Y, Takeda S & Dutta A (2015) Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Reports* **11**: 1749-1759.

Domingues S & Nielsen KM (2017) Membrane vesicles and horizontal gene transfer in prokaryotes. *Current opinion in microbiology* **38**: 16-21.

Dubnau D (1999) DNA uptake in bacteria. *Annual review of microbiology* **53**: 217-244.

Dupuy B & Sonenshein AL (1998) Regulated transcription of *Clostridium difficile* toxin genes. *Molecular microbiology* **27**: 107-120.

Elsaied H, Stokes HW, Nakamura T, Kitamura K, Fuse H & Maruyama A (2007) Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environmental microbiology* **9**: 2298-2312.

Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y & Maruyama A (2011) Marine integrons containing novel integrase genes, attachment sites, *attI*, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *The ISME journal* **5**: 1162-1177.

Epshtein V, Toulme F, Rahmouni AR, Borukhov S & Nudler E (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *The EMBO journal* **22**: 4719-4727.

Escudero JA, Loot C, Nivina A & Mazel D (2015) The integron: adaptation on demand. *Microbiology spectrum* **3**: MDNA3-0019-2014.

Farnham PJ & Platt T (1981) Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic acids research* **9**: 563-577.

Ferguson GP, Totemeyer S, MacLean MJ & Booth IR (1998) Methylglyoxal production in bacteria: suicide or survival? *Archives of microbiology* **170**: 209-218.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH & Caporaso JG (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 21390-21395.

Flensburg J & Skold O (1987) Massive overproduction of dihydrofolate reductase in bacteria as a response to the use of trimethoprim. *European journal of biochemistry* **162**: 473-476.

Fong KP, Goh CB & Tan HM (2000) The genes for benzene catabolism in *Pseudomonas putida* ML2 are flanked by two copies of the insertion element

IS1489, forming a class-I-type catabolic transposon, Tn5542. *Plasmid* **43**: 103-110.

Foster TJ, Davis MA, Roberts DE, Takeshita K & Kleckner N (1981) Genetic organization of transposon Tn10. *Cell* **23**: 201-213.

Freestone P, Nystrom T, Trinei M & Norris V (1997) The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *Journal of molecular biology* **274**: 318-324.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW & DeLong EF (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 3805-3810.

Frost LS, Leplae R, Summers AO & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* **3**: 722-732.

Fry BN, Feng S, Chen YY, Newell DG, Coloe PJ & Korolik V (2000) The *galE* gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence. *Infection and immunity* **68**: 2594-2601.

Fukui K (2010) DNA mismatch repair in eukaryotes and bacteria. *Journal of Nucleic Acids* **2010**: 16.

Fulsundar S, Harms K, Flaten GE, Johnsen PJ, Chopade BA & Nielsen KM (2014) Gene transfer potential of outer membrane vesicles of *Acinetobacter baylyi* and effects of stress on vesiculation. *Applied and environmental microbiology* **80**: 3469-3483.

Furi L, Haigh R, Al Jabri ZJH, Morrissey I, Ou H-Y, León-Sampedro R, Martinez JL, Coque TM & Oggioni MR (2016) Dissemination of novel antimicrobial resistance mechanisms through the insertion sequence mediated spread of metabolic genes. *Frontiers in Microbiology* **7**.

Furuya EY & Lowy FD (2006) Antimicrobial-resistant bacteria in the community setting. *Nature Reviews Microbiology* **4**: 36-45.

Gaida SM, Sandoval NR, Nicolaou SA, Chen Y, Venkataramanan KP & Papoutsakis ET (2015) Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nature Communications* **6**: 7045.

Gay P, Le Coq D, Steinmetz M, Berkelman T & Kado CI (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of bacteriology* **164**: 918-921.

Gerba CP (2015) Quaternary ammonium biocides: efficacy in application. *Applied and environmental microbiology* **81**: 464-469.

Gilbert P & Moore LE (2005) Cationic antiseptics: diversity of action under a common epithet. *Journal of applied microbiology* **99**: 703-715.

Gillings M, Boucher Y, Labbate M, Holmes A, Krishnan S, Holley M & Stokes HW (2008) The evolution of class 1 integrons and the rise of antibiotic resistance. *Journal of bacteriology* **190**: 5095-5100.

Gillings MR (2014) Integrons: past, present, and future. *Microbiology and molecular biology reviews : MMBR* **78**: 257-277.

Gillings MR, Holley MP & Stokes HW (2009) Evidence for dynamic exchange of *qac* gene cassettes between class 1 integrons and other integrons in freshwater biofilms. *FEMS Microbiology Letters* **296**: 282-288.

Gillings MR, Holley MP, Stokes HW & Holmes AJ (2005) Integrons in *Xanthomonas*: a source of species genome diversity. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4419-4424.

Goeders N, Chai R, Chen B, Day A & Salmond GP (2016) Structure, evolution, and functions of bacterial type III toxin-antitoxin systems. *Toxins* **8**.

Goessweiner-Mohr N, Arends K, Keller W & Grohmann E (2014) Conjugation in Gram-positive bacteria. *Microbiology spectrum* **2**: Plas-0004-2013.

Goffic FL, Martel A, Capmau ML, Baca B, Goebel P, Chardon H, Soussy CJ, Duval J & Bouanchaud DH (1976) New plasmid-mediated nucleotidylation of aminoglycoside antibiotics in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* **10**: 258-264.

Gomez-Alvarez V, Revetta RP & Santo Domingo JW (2012) Metagenomic analyses of drinking water receiving different disinfection treatments. *Applied and environmental microbiology* **78**: 6095-6102.

Gorecki RK, Koryszewska-Baginska A, Golebiewski M, Zylinska J, Grynberg M & Bardowski JK (2011) Adaptive potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. *PloS one* **6**: e22238.

Goussard S, Sougakoff W, Mabilat C, Bauernfeind A & Courvalin P (1991) An IS1-like element is responsible for high-level synthesis of extended-spectrum β -lactamase TEM-6 in *Enterobacteriaceae*. *Journal of general microbiology* **137**: 2681-2687.

Graupner S, Weger N, Sohni M & Wackernagel W (2001) Requirement of novel competence genes *pilT* and *pilU* of *Pseudomonas stutzeri* for natural transformation and suppression of *pilT* deficiency by a hexahistidine tag on the type IV pilus protein PilA1. *Journal of bacteriology* **183**: 4694-4701.

Gravel A, Fournier B & Roy PH (1998) DNA complexes obtained with the integrase IntI1 at the *attI1* site. *Nucleic acids research* **26**: 4347-4355.

Griffith F (1928) The significance of Pneumococcal types. *The Journal of hygiene* **27**: 113-159.

Guerin E, Jove T, Tabesse A, Mazel D & Ploy MC (2011) High-level gene cassette transcription prevents integrase expression in class 1 integrons. *Journal of bacteriology* **193**: 5675-5682.

Guerin E, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, Da Re S, Gonzalez-Zorn B, Barbe J, Ploy MC & Mazel D (2009) The SOS response controls integron recombination. *Science (New York, NY)* **324**: 1034.

Guerout AM, Iqbal N, Mine N, Ducos-Galand M, Van Melderen L & Mazel D (2013) Characterization of the *phd-doc* and *ccd* toxin-antitoxin cassettes from *Vibrio* superintegrons. *Journal of bacteriology* **195**: 2270-2283.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D & Andersson DI (2011) Selection of resistant bacteria at very low antibiotic concentrations. *PLOS Pathogens* **7**: e1002158.

Hachmann A-B, Sevim E, Gaballa A, Popham DL, Antelmann H & Helmann JD (2011) Reduction in membrane phosphatidylglycerol content leads to daptomycin resistance in *Bacillus subtilis*. *Antimicrobial agents and chemotherapy* **55**: 4326-4337.

Hamilton-Miller JM (1988) Reversal of activity of trimethoprim against gram-positive cocci by thymidine, thymine and 'folates'. *The Journal of antimicrobial chemotherapy* **22**: 35-39.

Hancock RE & Speert DP (2000) Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and impact on treatment. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* **3**: 247-255.

Hansson K, Sköld O & Sundström L (1997) Non-palindromic *attI* sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Molecular microbiology* **26**: 441-453.

Hardman DJ (1991) Biotransformation of halogenated compounds. *Critical reviews in biotechnology* **11**: 1-40.

Harmer CJ & Hall RM (2015) IS26-Mediated Precise Excision of the IS26-*aphA1a* Translocatable Unit. *mBio* **6**: e01866-01815.

Harmer CJ & Hall RM (2016) IS26-Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* **1**.

Harmer CJ, Moran RA & Hall RM (2014) Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *mBio* **5**: e01801-01814.

Hayes F (2003) Transposon-Based Strategies for Microbial Functional Genomics and Proteomics. *Annual review of genetics* **37**: 3-29.

He J, Li Y, Cao Y, Xue J & Zhou X (2015) The oral microbiome diversity and its relation to human diseases. *Folia Microbiologica* **60**: 69-80.

Heaton BE, Herrou J, Blackwell AE, Wysocki VH & Crosson S (2012) Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *The Journal of biological chemistry* **287**: 12098-12110.

Holah JT, Taylor JH, Dawson DJ & Hall KE (2002) Biocide use in the food industry and the disinfectant resistance of persistent strains of *Listeria monocytogenes* and *Escherichia coli*. *Journal of Applied Microbiology* **92**: 111S-120S.

Holmes AJ, Gillings MR, Nield BS, Mabbutt BC, Nevalainen KMH & Stokes HW (2003) The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environmental microbiology* **5**: 383-394.

Huang X (1992) A contig assembly program based on sensitive detection of fragment overlaps. *Genomics* **14**: 18-25.

Iqbal HA, Craig JW & Brady SF (2014) Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS microbiology letters* **354**: 19-26.

Ishikawa S, Matsumura Y, Yoshizako F & Tsuchido T (2002) Characterization of a cationic surfactant-resistant mutant isolated spontaneously from *Escherichia coli*. *Journal of applied microbiology* **92**: 261-268.

Jacquier H, Zaoui C, Sanson-le Pors MJ, Mazel D & Bercot B (2009) Translation regulation of integrons gene cassette expression by the *attC* sites. *Molecular microbiology* **72**: 1475-1486.

Jager W, Schafer A, Kalinowski J & Puhler A (1995) Isolation of insertion elements from Gram-positive *Brevibacterium*, *Corynebacterium* and *Rhodococcus* strains using the *Bacillus subtilis* *sacB* gene as a positive selection marker. *FEMS microbiology letters* **126**: 1-6.

Jain R, Rivera MC, Moore JE & Lake JA (2002) Horizontal gene transfer in microbial genome evolution. *Theoretical population biology* **61**: 489-495.

Jellen-Ritter AS & Kern WV (2001) Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in *Escherichia coli* mutants selected with a fluoroquinolone. *Antimicrobial agents and chemotherapy* **45**: 1467-1472.

Jennings MC, Minbiole KPC & Wuest WM (2015) Quaternary ammonium compounds: an antimicrobial mainstay and platform for innovation to address bacterial resistance. *ACS infectious diseases* **1**: 288-303.

Johnston C, Martin B, Fichant G, Polard P & Claverys J-P (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* **12**: 181-196.

Jørgensen TS, Kiil AS, Hansen MA, Sørensen SJ & Hansen LH (2014) Current strategies for mobilome research. *Frontiers in Microbiology* **5**: 750.

Jové T, Da Re S, Denis F, Mazel D & Ploy M-C (2010) Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLOS Genetics* **6**: e1000793.

Juers DH, Matthews BW & Huber RE (2012) LacZ β -galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Science : A Publication of the Protein Society* **21**: 1792-1807.

Katayama Y, Ito T & Hiramatsu K (2000) A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* **44**: 1549-1555.

Koenig JE, Boucher Y, Charlebois RL, Nesbø C, Zhaxybayeva O, Baptiste E, Spencer M, Joss MJ, Stokes HW & Doolittle WF (2008) Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. *Environmental microbiology* **10**: 1024-1038.

Kolenbrander PE, Palmer Jr RJ, Periasamy S & Jakubovics NS (2010) Oral multispecies biofilm development and the key role of cell–cell distance. *Nature Reviews Microbiology* **8**: 471-480.

Koo B-M, Rhodius VA, Campbell EA & Gross CA (2009) Mutational analysis of *Escherichia coli* σ^{28} and its target promoters reveal recognition of a composite –10 region, comprised of an “extended –10 motif” and a core-10 element. *Molecular microbiology* **72**: 830-843.

Koonin EV & Tatusov RL (1994) Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *Journal of molecular biology* **244**: 125-132.

Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B & Roeselers G (2014) Shaping the oral microbiota through intimate kissing. *Microbiome* **2**: 41.

Kreth J, Merritt J & Qi F (2009) Bacterial and host interactions of oral streptococci. *DNA and cell biology* **28**: 397-403.

Kvint K, Nachin L, Diez A & Nyström T (2003) The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology* **6**: 140-145.

Lacks SA & Greenberg B (2001) Constitutive competence for genetic transformation in *Streptococcus pneumoniae* caused by mutation of a transmembrane histidine kinase. *Molecular microbiology* **42**: 1035-1045.

Lamberte LE, Baniulyte G, Singh SS, Stringer AM, Bonocora RP, Stracy M, Kapanidis AN, Wade JT & Grainger DC (2017) Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nature Microbiology* **2**: 16249.

Lancaster H, Bedi R, Wilson M & Mullany P (2005) The maintenance in the oral cavity of children of tetracycline-resistant bacteria and the genes encoding such resistance. *The Journal of antimicrobial chemotherapy* **56**: 524-531.

Lederberg J & Tatum EL (1946) Gene recombination in *Escherichia coli*. *Nature* **158**: 558.

Lee C & Park C (2017) Bacterial responses to glyoxal and methylglyoxal: reactive electrophilic species. *International Journal of Molecular Sciences* **18**: 169.

Leelaporn A, Yodkamol K, Waywa D & Pattanachaiwit S (2008) A novel structure of Tn4001-truncated element, type V, in clinical enterococcal isolates and multiplex PCR for detecting aminoglycoside resistance genes. *International journal of antimicrobial agents* **31**: 250-254.

Lejard V, Rebours E, Meersseman C & Rocha D (2014) Construction and validation of a novel dual reporter vector for studying mammalian bidirectional promoters. *Plasmid* **74**: 1-8.

Leplae R, Geeraerts D, Hallez R, Guglielmini J, Dreze P & Van Melderen L (2011) Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic acids research* **39**: 5513-5525.

Lévesque C, Brassard S, Lapointe J & Roy PH (1994) Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integron. *Gene* **142**: 49-54.

Levin-Reisman I, Ronin I, Gefen O, Braniss I, Shores N & Balaban NQ (2017) Antibiotic tolerance facilitates the evolution of resistance. *Science (New York, NY)* **355**: 826-830.

- Lewis K (2001) Riddle of biofilm resistance. *Antimicrobial agents and chemotherapy* **45**: 999-1007.
- Lewis K (2013) Platforms for antibiotic discovery. *Nature reviews drug discovery* **12**: 371-387.
- Lif Holgerson P, Harnevik L, Hernell O, Tanner ACR & Johansson I (2011) Mode of birth delivery affects oral microbiota in infants. *Journal of dental research* **90**: 1183-1188.
- Ligon BL (2004) Penicillin: its discovery and early development. *Seminars in Pediatric Infectious Diseases* **15**: 52-57.
- Lim D & Strynadka NCJ (2002) Structural basis for the β -lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. *Nature structural & molecular biology* **9**: 870-876.
- Limberger RJ, Slivienski LL, Izard J & Samsonoff WA (1999) Insertional inactivation of *Treponema denticola tap1* results in a nonmotile mutant with elongated flagellar hooks. *Journal of bacteriology* **181**: 3743-3750.
- Liu YY, Wang Y, Walsh TR, *et al.* (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious diseases* **16**: 161-168.
- Loberto JCS, Martins CAAdP, Santos SSFd, Cortelli JR & Jorge AOC (2004) *Staphylococcus* spp. in the oral cavity and periodontal pockets of chronic periodontitis patients. *Brazilian Journal of Microbiology* **35**: 64-68.
- Lyon BR, May JW & Skurray RA (1984) Tn4001: a gentamicin and kanamycin resistance transposon in *Staphylococcus aureus*. *Molecular and General Genetics MGG* **193**: 554-556.
- Ma NJ, Moonan DW & Isaacs FJ (2014) Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nature Protocols* **9**: 2285-2300.

Mahillon J & Chandler M (1998) Insertion sequences. *Microbiology and molecular biology reviews* : **MMBR 62**: 725-774.

Maillard JY, Bloomfield S, Coelho JR, *et al.* (2013) Does microbicide use in consumer products promote antimicrobial resistance? A critical review and recommendations for a cohesive approach to risk assessment. *Microbial drug resistance (Larchmont, NY)* **19**: 344-354.

Makarova KS, Grishin NV & Koonin EV (2006) The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. *Bioinformatics* **22**: 2581-2584.

Mantengoli E & Rossolini GM (2005) Tn5393d, a complex Tn5393 derivative carrying the PER-1 extended-spectrum beta-lactamase gene and other resistance determinants. *Antimicrobial agents and chemotherapy* **49**: 3289-3296.

Maravic G (2004) Macrolide resistance based on the Erm-mediated rRNA methylation. *Current drug targets Infectious disorders* **4**: 193-202.

Marsh PD (2004) Dental plaque as a microbial biofilm. *Caries research* **38**: 204-211.

Marsh PD (2006) Dental plaque as a biofilm and a microbial community – implications for health and disease. *BMC Oral Health* **6**: S14-S14.

Marsh PD (2010) Controlling the oral biofilm with antimicrobials. *Journal of dentistry* **38 Suppl 1**: S11-15.

Martin FH & Tinoco JI (1980) DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic acids research* **8**: 2295-2300.

Martinez-Freijo P, Fluit AC, Schmitz FJ, Grek VS, Verhoef J & Jones ME (1998) Class I integrons in Gram-negative isolates from different European hospitals and association with decreased susceptibility to multiple antibiotic compounds. *The Journal of antimicrobial chemotherapy* **42**: 689-696.

Martínez-Suárez JV, Ortiz S & López-Alonso V (2016) Potential impact of the resistance to quaternary ammonium disinfectants on the persistence of *Listeria monocytogenes* in food processing environments. *Frontiers in Microbiology* **7**: 638.

Martinez JL & Baquero F (2000) Mutation frequencies and antibiotic resistance. *Antimicrobial agents and chemotherapy* **44**: 1771-1777.

Mastropaolo MD, Thorson ML & Stevens AM (2009) Comparison of *Bacteroides thetaiotaomicron* and *Escherichia coli* 16S rRNA gene expression signals. *Microbiology (Reading, England)* **155**: 2683-2693.

Maxwell ES (1957) The enzymic interconversion of uridine diphosphogalactose and uridine diphosphoglucose. *The Journal of biological chemistry* **229**: 139-151.

McDonnell G & Russell AD (1999) Antiseptics and disinfectants: activity, action, and resistance. *Clinical microbiology reviews* **12**: 147-179.

McDougal LK, Fosheim GE, Nicholson A, Bulens SN, Limbago BM, Shearer JE, Summers AO & Patel JB (2010) Emergence of resistance among USA300 methicillin-resistant *Staphylococcus aureus* isolates causing invasive disease in the United States. *Antimicrobial agents and chemotherapy* **54**: 3804-3811.

Meier MJ, Paterson ES & Lambert IB (2016) Use of substrate-induced gene expression in metagenomic analysis of an aromatic hydrocarbon-contaminated soil. *Applied and environmental microbiology* **82**: 897-909.

Merz AJ & So M (2000) Interactions of pathogenic *Neisseriae* with epithelial cell membranes. *Annual review of cell and developmental biology* **16**: 423-457.

Michael CA & Labbate M (2010) Gene cassette transcription in a large integron-associated array. *BMC genetics* **11**: 82.

Michael CA, Dominey-Howes D & Labbate M (2014) The antimicrobial resistance crisis: causes, consequences, and management. *Frontiers in public health* **2**: 145.

Miller JH (1972) *Experiments in molecular genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Montanaro L, Poggi A, Visai L, Ravaioli S, Campoccia D, Speziale P & Arciola CR (2011) Extracellular DNA in biofilms. *The International journal of artificial organs* **34**: 824-831.

Mortier-Barriere I, Velten M, Dupaigne P, *et al.* (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* **130**: 824-836.

Moura A, Jove T, Ploy MC, Henriques I & Correia A (2012) Diversity of gene cassette promoters in class 1 integrons from wastewater environments. *Applied and environmental microbiology* **78**: 5413-5416.

Mugglestone MA, Murphy MS, Visintin C, Howe DT & Turner MA (2014) Antibiotics for early-onset neonatal infection: a summary of the NICE guideline 2012. *The Obstetrician & Gynaecologist* **16**: 87-92.

Mutschler H, Gebhardt M, Shoeman RL & Meinhart A (2011) A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLOS biology* **9**: e1001033.

Nachin L, Nannmark U & Nystrom T (2005) Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *Journal of bacteriology* **187**: 6265-6272.

Nakao R, Senpuku H & Watanabe H (2006) *Porphyromonas gingivalis galE* Is Involved in Lipopolysaccharide O-Antigen Synthesis and Biofilm Formation. *Infection and Immunity* **74**: 6145-6153.

National Research Council (2007) Designing a successful metagenomics project: best practices and future needs. *The New Science of Metagenomics*:

Revealing the Secrets of Our Microbial Planet,60-84. National Academies Press (US), Washington (DC).

Naville M, Ghuillot-Gaudeffroy A, Marchais A & Gautheret D (2011) ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA biology* **8**: 11-13.

Nicolas E, Lambin M, Dandoy D, Galloy C, Nguyen N, Oger CA & Hallet B (2015) The Tn3-family of replicative transposons. *Microbiol Spectrum* **3**.

Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM & Stokes HW (2001) Recovery of new integron classes from environmental DNA. *FEMS microbiology letters* **195**: 59-65.

Nikaido H (2001) Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria. *Seminars in cell & developmental biology* **12**: 215-223.

Nojiri H, Shintani M & Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Applied microbiology and biotechnology* **64**: 154-174.

Normark BH & Normark S (2002) Evolution and spread of antibiotic resistance. *Journal of internal medicine* **252**: 91-106.

Nystrom T & Neidhardt FC (1996) Effects of overproducing the universal stress protein, UspA, in *Escherichia coli* K-12. *Journal of bacteriology* **178**: 927-930.

Nyström T & Neidhardt FC (1994) Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Molecular microbiology* **11**: 537-544.

O'Neill J (2016) Tackling drug-resistant infections globally: final report and recommendations. The Review on Antimicrobial Resistance.

Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R & Konstantinidis KT (2011) Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Applied and environmental microbiology* **77**: 6000-6011.

Olliver A, Valle M, Chaslus-Dancla E & Cloeckaert A (2004) Role of an *acrR* mutation in multidrug resistance of *in vitro*-selected fluoroquinolone-resistant mutants of *Salmonella enterica* serovar Typhimurium. *FEMS microbiology letters* **238**: 267-272.

Ortega Morente E, Fernandez-Fuentes MA, Grande Burgos MJ, Abriouel H, Perez Pulido R & Galvez A (2013) Biocide tolerance in bacteria. *International journal of food microbiology* **162**: 13-25.

Ortiz S, López V & Martínez-Suárez JV (2014) Control of *Listeria monocytogenes* contamination in an Iberian pork processing plant and selection of benzalkonium chloride-resistant strains. *Food Microbiology* **39**: 81-88.

Ozeki H & Ikeda H (1968) Transduction mechanisms. *Annual review of genetics* **2**: 245-278.

Page R & Peti W (2016) Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology* **12**: 208-214.

Paget MSB & Helmann JD (2003) The σ^{70} family of sigma factors. *Genome Biology* **4**: 203-203.

Palmieri C, Mingoia M & Varaldo PE (2013) Unconventional circularizable bacterial genetic structures carrying antibiotic resistance determinants. *Antimicrobial agents and chemotherapy* **57**: 2440-2441.

Pankey GA & Sabath LD (2004) Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of Gram-positive bacterial infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **38**: 864-870.

Papagiannitsis CC, Tzouvelekis LS & Miriagou V (2009) Relative strengths of the class 1 integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active P2 promoter. *Antimicrobial agents and chemotherapy* **53**: 277-280.

Partridge SR, Tsafnat G, Coiera E & Iredell JR (2009) Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS microbiology reviews* **33**.

Pedersen LC, Benning MM & Holden HM (1995) Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry* **34**: 13305-13311.

Perkins JB & Youngman PJ (1984) A physical and functional analysis of Tn917, a *Streptococcus* transposon in the Tn3 family that functions in *Bacillus*. *Plasmid* **12**: 119-138.

Peschel A, Otto M, Jack RW, Kalbacher H, Jung G & Gotz F (1999) Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *The Journal of biological chemistry* **274**: 8405-8410.

Piddock LJ (1999) Mechanisms of fluoroquinolone resistance: an update 1994-1998. *Drugs* **58 Suppl 2**: 11-18.

Poirel L, Cabanne L, Collet L & Nordmann P (2006) Class II transposon-borne structure harboring metallo- β -lactamase gene *bla*_{VIM-2} in *Pseudomonas putida*. *Antimicrobial agents and chemotherapy* **50**: 2889-2891.

Post V & Hall RM (2009) Insertion sequences in the IS1111 family that target the *attC* recombination sites of integron-associated gene cassettes. *FEMS microbiology letters* **290**: 182-187.

Prescott EM & Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 8796-8801.

Prudhomme M, Libante V & Claverys J-P (2002a) Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2100-2105.

Prudhomme M, Turlan C, Claverys JP & Chandler M (2002b) Diversity of Tn4001 transposition products: the flanking IS256 elements can form tandem dimers and IS circles. *Journal of bacteriology* **184**: 433-443.

Pujol C, Eugène E, Marceau M & Nassif X (1999) The meningococcal PilT protein is required for induction of intimate attachment to epithelial cells following pilus-mediated adhesion. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 4017-4022.

Pushalkar S, Mane SP, Ji X, Li Y, Evans C, Crasta OR, Morse D, Meagher R, Singh A & Saxena D (2011) Microbial diversity in saliva of oral squamous cell carcinoma. *FEMS immunology and medical microbiology* **61**: 269-277.

Qin J, Li R, Raes J, *et al.* (2010) A human gut microbial gene catalog established by metagenomic sequencing. *Nature* **464**: 59-65.

Quintiliani R, Jr. & Courvalin P (1996) Characterization of Tn1547, a composite transposon flanked by the IS16 and IS256-like elements, that confers vancomycin resistance in *Enterococcus faecalis* BM4281. *Gene* **172**: 1-8.

Ramirez MS & Tolmasky ME (2010) Aminoglycoside modifying enzymes. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* **13**: 151-171.

Ready D, Bedi R, Spratt DA, Mullany P & Wilson M (2003) Prevalence, proportions, and identities of antibiotic-resistant bacteria in the oral microflora of healthy children. *Microbial drug resistance (Larchmont, NY)* **9**: 367-372.

Reid CJ, Roy Chowdhury P & Djordjevic SP (2015) Tn6026 and Tn6029 are found in complex resistance regions mobilised by diverse plasmids and

chromosomal islands in multiple antibiotic resistant *Enterobacteriaceae*. *Plasmid* **80**: 127-137.

Reynolds LJ, Roberts AP & Anjum MF (2016) Efflux in the oral metagenome: the discovery of a novel tetracycline and tigecycline ABC transporter. *Frontiers in Microbiology* **7**.

Reyrat J-M, Pelicic V, Gicquel B & Rappuoli R (1998) Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infection and Immunity* **66**: 4011-4017.

Rezania S, Amirmozaffari N, Tabarraei B, Jeddi-Tehrani M, Zarei O, Alizadeh R, Masjedian F & Zarnani AH (2011) Extraction, purification and characterization of lipopolysaccharide from *Escherichia coli* and *Salmonella typhi*. *Avicenna Journal of Medical Biotechnology* **3**: 3-9.

Rice LB, Carias LL, Marshall S, Rudin SD & Hutton-Thomas R (2005) Tn5386, a novel Tn916-like mobile element in *Enterococcus faecium* D344R that interacts with Tn916 to yield a large genomic deletion. *Journal of bacteriology* **187**: 6668-6677.

Robert HS & Thomas DP (1991) Integration of DNA Fragments by Illegitimate Recombination in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 7585-7589.

Roberts AP & Mullany P (2009) A modular master on the move: the Tn916 family of mobile genetic elements. *Trends in microbiology* **17**: 251-258.

Roberts AP & Kreth J (2014) The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Frontiers in cellular and infection microbiology* **4**: 124.

Roberts MC (2005) Update on acquired tetracycline resistance genes. *FEMS Microbiology Letters* **245**: 195-203.

Rodić N & Burns KH (2013) Long Interspersed Element-1 (LINE-1): passenger or driver in human neoplasms? *PLOS Genetics* **9**: e1003402.

Rouch DA, Messerotti LJ, Loo LS, Jackson CA & Skurray RA (1989) Trimethoprim resistance transposon Tn4003 from *Staphylococcus aureus* encodes genes for a dihydrofolate reductase and thymidylate synthetase flanked by three copies of IS257. *Molecular microbiology* **3**: 161-175.

Rowe-Magnus DA, Guerout AM, Biskri L, Bouige P & Mazel D (2003) Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome research* **13**: 428-442.

Russell AD (2003) Biocide use and antibiotic resistance: the relevance of laboratory findings to clinical and environmental situations. *The Lancet Infectious diseases* **3**: 794-803.

Rzhepishevskaya O, Hakobyan S, Ruhel R, Gautrot J, Barbero D & Ramstedt M (2013) The surface charge of anti-bacterial coatings alters motility and biofilm architecture. *Biomaterials Science* **1**: 589-602.

Sabra W, Lunsdorf H & Zeng AP (2003) Alterations in the formation of lipopolysaccharide and membrane vesicles on the surface of *Pseudomonas aeruginosa* PAO1 under oxygen stress conditions. *Microbiology (Reading, England)* **149**: 2789-2795.

Sabree ZL, Rondon MR & Handelsman J (2009) Metagenomics. *Encyclopedia of Microbiology (Third Edition)*, Vol. A2 (Schaechter M, ed.) 622-632. Academic Press, Oxford.

Sakamoto M, Umeda M, Ishikawa I & Benno Y (2000) Comparison of the oral bacterial flora in saliva from a healthy subject and two periodontitis patients by sequence analysis of 16S rDNA libraries. *Microbiology and immunology* **44**: 643-652.

Salmond GP & Fineran PC (2015) A century of the phage: past, present and future. *Nature Reviews Microbiology* **13**: 777-786.

Sanchez C, Hernandez de Rojas A, Martinez B, Arguelles ME, Suarez JE, Rodriguez A & Mayo B (2000) Nucleotide sequence and analysis of pBL1, a

bacteriocin-producing plasmid from *Lactococcus lactis* IPLA 972. *Plasmid* **44**: 239-249.

Schmieder R & Edwards R (2011) Insights into antibiotic resistance through metagenomic approaches. *Future microbiology* **7**: 73-89.

Schnabel EL & Jones AL (1999) Distribution of tetracycline resistance genes and transposons among phylloplane bacteria in Michigan apple orchards. *Applied and environmental microbiology* **65**: 4898-4907.

Schneider D, Faure D, Noirclerc-Savoye M, Barriere AC, Coursange E & Blot M (2000) A broad-host-range plasmid for isolating mobile genetic elements in gram-negative bacteria. *Plasmid* **44**: 201-207.

Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M & Mead D (2008) Assembly of viral metagenomes from yellowstone hot springs. *Applied and environmental microbiology* **74**: 4164-4174.

Schumacher MA, Piro KM, Xu W, Hansen S, Lewis K & Brennan RG (2009) Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. *Science (New York, NY)* **323**: 396-401.

Schwechheimer C & Kuehn MJ (2015) Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nature Reviews Microbiology* **13**: 605-619.

Scordilis GE, Ree H & Lessie TG (1987) Identification of transposable elements which activate gene expression in *Pseudomonas cepacia*. *Journal of bacteriology* **169**: 8-13.

Seier-Petersen MA, Jasni A, Aarestrup FM, Vigre H, Mullany P, Roberts AP & Agerso Y (2014) Effect of subinhibitory concentrations of four commonly used biocides on the conjugative transfer of Tn916 in *Bacillus subtilis*. *The Journal of antimicrobial chemotherapy* **69**: 343-348.

Seville LA, Patterson AJ, Scott KP, Mullany P, Quail MA, Parkhill J, Ready D, Wilson M, Spratt D & Roberts AP (2009) Distribution of tetracycline and

erythromycin resistance genes among human oral and fecal metagenomic DNA. *Microbial drug resistance (Larchmont, NY)* **15**: 159-166.

Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science* **5**: 209.

Shaw L, Ribeiro ALR, Levine AP, Pontikos N, Balloux F, Segal AW, Roberts AP & Smith A (2017) The human oral microbiome is shaped by shared environment rather than genetics: evidence from a large family of closely-related individuals. *bioRxiv*.

Shearwin KE, Callen BP & Egan JB (2005) Transcriptional interference--a crash course. *Trends in genetics : TIG* **21**: 339-345.

Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD & Dutta A (2012) Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science (New York, NY)* **336**: 82-86.

Shiga Y, Sekine Y & Ohtsubo E (1999) Transposition of IS 1 circles. *Genes to Cells* **4**: 551-561.

Sievers F, Wilm A, Dineen D, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539-539.

Siguier P, Goubeyre E & Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS microbiology reviews* **38**: 865-891.

Siguier P, Goubeyre E, Varani A, Ton-Hoang B & Chandler M (2015) Everyman's guide to bacterial insertion sequences. *Microbiology spectrum* **3**: Mdna3-0030-2014.

Simon C, Wiezer A, Strittmatter AW & Daniel R (2009a) Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Applied and environmental microbiology* **75**: 7519-7526.

Simon C, Herath J, Rockstroh S & Daniel R (2009b) Rapid identification of genes encoding DNA polymerases by function-based screening of

metagenomic libraries derived from glacial ice. *Applied and environmental microbiology* **75**: 2964-2968.

Singh RN (1972) Number of deoxyribonucleic acid uptake sites in competent cells of *Bacillus subtilis*. *Journal of bacteriology* **110**: 266-272.

Skinner R, Cundliffe E & Schmidt FJ (1983) Site of action of a ribosomal RNA methylase responsible for resistance to erythromycin and other antibiotics. *The Journal of biological chemistry* **258**: 12702-12706.

Skold O (2001) Resistance to trimethoprim and sulfonamides. *Veterinary research* **32**: 261-273.

Smith GR (1991) Conjugational recombination in *E. coli*: myths and mechanisms. *Cell* **64**: 19-27.

Soge OO, Beck NK, White TM, No DB & Roberts MC (2008) A novel transposon, Tn6009, composed of a Tn916 element linked with a *Staphylococcus aureus mer* operon. *The Journal of antimicrobial chemotherapy* **62**: 674-680.

Soler N, Marguet E, Verbavatz JM & Forterre P (2008) Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order Thermococcales. *Res Microbiol* **159**: 390-399.

Solomon JM & Grossman AD (1996) Who's competent and when: regulation of natural genetic competence in bacteria. *Trends in genetics : TIG* **12**: 150-155.

Solovyev V & Salamov A (2011) Automatic annotation of microbial genomes and metagenomic sequences. *In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, (Li RW, ed.) 61-78. Nova Science Publishers.

Solyga A & Bartosik D (2004) Entrapment vectors--how to capture a functional transposable element. *Polish journal of microbiology / Polskie Towarzystwo Mikrobiologow = The Polish Society of Microbiologists* **53**: 139-144.

Soumet C, Fourreau E, Legrandois P & Maris P (2012) Resistance to phenicol compounds following adaptation to quaternary ammonium compounds in *Escherichia coli*. *Veterinary Microbiology* **158**: 147-152.

Staley JT & Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology* **39**: 321-346.

Stewart EJ (2012) Growing unculturable bacteria. *Journal of bacteriology* **194**: 4151-4160.

Stokes HW & Hall RM (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular microbiology* **3**: 1669-1683.

Stokes HW & Hall RM (1991) Sequence analysis of the inducible chloramphenicol resistance determinant in the Tn1696 integron suggests regulation by translational attenuation. *Plasmid* **26**: 10-19.

Stokes HW, O'Gorman DB, Recchia GD, Parsekhian M & Hall RM (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Molecular microbiology* **26**: 731-745.

Stokes HW, Holmes AJ, Nield BS, Holley MP, Nevalainen KM, Mabbutt BC & Gillings MR (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Applied and environmental microbiology* **67**: 5240-5246.

Su C, Lei L, Duan Y, Zhang KQ & Yang J (2012) Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Applied microbiology and biotechnology* **93**: 993-1003.

Summers DK & Sherratt DJ (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**: 1097-1103.

Sundstrom L, Swedberg G & Skold O (1993) Characterization of transposon Tn5086, carrying the site-specifically inserted gene *dhfrVII* mediating trimethoprim resistance. *Journal of bacteriology* **175**: 1796-1805.

Swartzman A, Kapoor S, Graham AF & Meighen EA (1990) A new *Vibrio fischeri lux* gene precedes a bidirectional termination site for the *lux* operon. *Journal of bacteriology* **172**: 6797-6802.

Szekeres S, Dauti M, Wilde C, Mazel D & Rowe-Magnus DA (2007) Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Molecular microbiology* **63**: 1588-1605.

Szeverenyi I, Hodel A, Arber W & Olsz F (1996) Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes. *Gene* **174**: 103-110.

Szuplewska M & Bartosik D (2009) Identification of a mosaic transposable element of *Paracoccus marcusii* composed of insertion sequence *ISPmar4* (ISAs1 family) and an *IS1247* a-driven transposable module (TMO). *FEMS Microbiology Letters* **292**: 216-221.

Tanner AC, Kent RL, Jr., Holgerson PL, Hughes CV, Loo CY, Kanasi E, Chalmers NI & Johansson I (2011) Microbiota of severe early childhood caries before and after therapy. *Journal of dental research* **90**: 1298-1305.

Tansirichaiya S, Mullany P & Roberts AP (2016a) PCR-based detection of composite transposons and translocatable units from oral metagenomic DNA. *FEMS microbiology letters* **363**.

Tansirichaiya S, Rahman MA, Antepowicz A, Mullany P & Roberts AP (2016b) Detection of novel integrons in the metagenome of human saliva. *PloS one* **11**: e0157605.

Tansirichaiya S, Reynolds LJ, Cristarella G, Wong LC, Rosendahl K & Roberts AP (2017) Reduced susceptibility to antiseptics is conferred by heterologous housekeeping genes. *Microbial drug resistance (Larchmont, NY)*.

Tatum EL & Lederberg J (1947) Gene recombination in the bacterium *Escherichia coli*. *Journal of bacteriology* **53**: 673-684.

The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207-214.

Thoden JB, Frey PA & Holden HM (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. *Biochemistry* **35**: 5137-5144.

Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* **3**: 711-721.

Tillett WS, Cambier MJ & Harris WH (1943) Sulfonamide-fast pneumococci. A clinical report of two cases of pneumonia together with experimental studies on the effectiveness of penicillin and tyrothricin against sulfonamid-resistant strains. *The Journal of clinical investigation* **22**: 249-255.

Tran TT, Munita JM & Arias CA (2015) Mechanisms of drug resistance: daptomycin resistance. *Annals of the New York Academy of Sciences* **1354**: 32-53.

Tran TT, Panesso D, Gao H, *et al.* (2013) Whole-genome analysis of a daptomycin-susceptible *Enterococcus faecium* strain and its daptomycin-resistant variant arising during therapy. *Antimicrobial agents and chemotherapy* **57**: 261-268.

Tribble GD, Parker AC & Smith CJ (1999) Genetic structure and transcriptional analysis of a mobilizable, antibiotic resistance transposon from Bacteroides. *Plasmid* **42**: 1-12.

Tsutsumi Y, Tomita H & Tanimoto K (2013) Identification of novel genes responsible for overexpression of *ampC* in *Pseudomonas aeruginosa* PAO1. *Antimicrobial agents and chemotherapy* **57**: 5987-5993.

Uchiyama T & Miyazaki K (2010) Substrate-induced gene expression screening: a method for high-throughput screening of metagenome libraries. *Metagenomics: Methods and Protocols*,(Streit WR & Daniel R, eds.), 153-168. Humana Press, Totowa, NJ.

Uchiyama T, Abe T, Ikemura T & Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology* **23**: 88-93.

Van Melderen L & Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLOS genetics* **5**: e1000437.

Vander Jagt DL & Hunsaker LA (2003) Methylglyoxal metabolism and diabetic complications: roles of aldose reductase, glyoxalase-I, betaine aldehyde dehydrogenase and 2-oxoaldehyde dehydrogenase. *Chemico-biological interactions* **143-144**: 341-351.

Vartoukian SR, Palmer RM & Wade WG (2010) Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters* **309**: 1-7.

Ventola CL (2015) The antibiotic resistance crisis: part 1: causes and threats. *P & T : a peer-reviewed journal for formulary management* **40**: 277-283.

Vinue L, Jove T, Torres C & Ploy MC (2011) Diversity of class 1 integron gene cassette Pc promoter variants in clinical *Escherichia coli* strains and description of a new P2 promoter variant. *International journal of antimicrobial agents* **38**: 526-529.

Vogelmann J, Ammelburg M, Finger C, Guezguez J, Linke D, Flotenmeyer M, Stierhof YD, Wohlleben W & Muth G (2011) Conjugal plasmid transfer in *Streptomyces* resembles bacterial chromosome segregation by FtsK/SpoIIIE. *The EMBO journal* **30**: 2246-2254.

von Hippel PH & Yager TD (1992) The elongation-termination decision in transcription. *Science (New York, NY)* **255**: 809-812.

Wade WG (2011) Has the use of molecular methods for the characterization of the human oral microbiome changed our understanding of the role of bacteria in the pathogenesis of periodontal disease? *Journal of clinical periodontology* **38 Suppl 11**: 7-16.

Wade WG (2013) The oral microbiome in health and disease. *Pharmacological research* **69**: 137-143.

Wainwright M & Kristiansen JE (2011) On the 75th anniversary of Prontosil. *Dyes and Pigments* **88**: 231-234.

Walsh C (2003) *Antibiotics*. American Society of Microbiology.

Warburton P, Roberts AP, Allan E, Seville L, Lancaster H & Mullany P (2009) Characterization of *tet(32)* genes from the oral metagenome. *Antimicrobial agents and chemotherapy* **53**: 273-276.

Wiegand I, Hilpert K & Hancock REW (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols* **3**: 163-175.

Wild J & Szybalski W (2004) Copy-control pBAC/oriV vectors for genomic cloning. *Methods in molecular biology (Clifton, NJ)* **267**: 145-154.

Wilson GA & Bott KF (1968) Nutritional factors influencing the development of competence in the *Bacillus subtilis* transformation system. *Journal of bacteriology* **95**: 1439-1449.

Wolfgang M, Lauer P, Park HS, Brossay L, Hebert J & Koomey M (1998) PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Molecular microbiology* **29**: 321-330.

Wood TK, Gonzalez Barrios AF, Herzberg M & Lee J (2006) Motility influences biofilm architecture in *Escherichia coli*. *Applied microbiology and biotechnology* **72**: 361-367.

Woodford N & Ellington MJ (2007) The emergence of antibiotic resistance by mutation. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **13**: 5-18.

Woodhouse JN, Fan L, Brown MV, Thomas T & Neilan BA (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *The ISME journal* **7**: 1842-1851.

Woody ST, Fong RS & Gussin GN (1993) Effects of a single base-pair deletion in the bacteriophage λ P_{RM} promoter. Repression of P_{RM} by repressor bound at O_{R2} and by RNA polymerase bound at P_R . *Journal of molecular biology* **229**: 37-51.

World Health Organization (2014) Antimicrobial resistance: global report on surveillance. *WHO: Geneva, Switzerland*.

Wozniak A, Villagra NA, Undabarrena A, Gallardo N, Keller N, Moraga M, Roman JC, Mora GC & Garcia P (2012) Porin alterations present in non-carbapenemase-producing *Enterobacteriaceae* with high and intermediate levels of carbapenem resistance in Chile. *Journal of medical microbiology* **61**: 1270-1279.

Wright GD (2005) Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced Drug Delivery Reviews* **57**: 1451-1470.

Wu W, Espedido B, Feng Y & Zong Z (2016) *Citrobacter freundii* carrying bla_{KPC-2} and bla_{NDM-1} : characterization by whole genome sequencing. **6**: 30670.

Wu Y-W, Rho M, Doak TG & Ye Y (2012) Oral spirochetes implicated in dental diseases are widespread in normal human subjects and carry extremely diverse integron gene cassettes. *Applied and environmental microbiology* **78**.

Xie G, Chain PS, Lo CC, Liu KL, Gans J, Merritt J & Qi F (2010) Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. *Molecular oral microbiology* **25**: 391-405.

- Xiong J, Déraspe M, Iqbal N, Ma J, Jamieson FB, Wasserscheid J, Dewar K, Hawkey PM & Roy PH (2016) Genome and plasmid analysis of *bla*_{IMP-4}-carrying *Citrobacter freundii* B38. *Antimicrobial agents and chemotherapy* **60**: 6719-6725.
- Xu J, Li W, Wu J, Zhang Y, Zhu Z, Liu J & Hu Z (2006) Stability of plasmid and expression of a recombinant gonadotropin-releasing hormone (GnRH) vaccine in *Escherichia coli*. *Applied microbiology and biotechnology* **73**: 780-788.
- Yang F, Zeng X, Ning K, *et al.* (2012) Saliva microbiomes distinguish caries-active from healthy human populations. *The ISME journal* **6**: 1-10.
- Yang S-J, Mishra NN, Rubio A & Bayer AS (2013) Causal role of single nucleotide polymorphisms within the *mprF* gene of *Staphylococcus aureus* in daptomycin resistance. *Antimicrobial agents and chemotherapy* **57**: 5658-5664.
- Yates A, Akanni W, Amode MR, *et al.* (2016) Ensembl 2016. *Nucleic acids research* **44**: D710-D716.
- Zaura E, Keijser BJ, Huse SM & Crielaard W (2009) Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiology* **9**: 259.
- Zechner EL, Lang S & Schildbach JF (2012) Assembly and mechanisms of bacterial type IV secretion machines. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **367**: 1073-1087.
- Zervos MJ & Schaberg DR (1985) Reversal of the in vitro susceptibility of enterococci to trimethoprim-sulfamethoxazole by folinic acid. *Antimicrobial agents and chemotherapy* **28**: 446-448.
- Zgurskaya HI, Lopez CA & Gnanakaran S (2015) Permeability barrier of Gram-negative cell envelopes and approaches to bypass it. *ACS infectious diseases* **1**: 512-522.

Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E & Ebright RH (2012) Structural basis of transcription initiation. *Science (New York, NY)* **338**: 1076-1080.

Zinder ND & Lederberg J (1952) Genetic exchange in *Salmonella*. *Journal of bacteriology* **64**: 679-699.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**: 3406-3415.

Appendices

Appendix 1. Ethics consent form

Information Sheet for *saliva acquisition* in Research Studies

You will be given a copy of this information sheet.

Title of Project: **Understanding oral microbial ecology**

This study has been approved by the UCL Research Ethics Committee (Project ID Number): **5017/001**

Name Adam Roberts

Work Address **UCL Eastman Dental Institute, 256 Gray's Inn Road, London, WC1X 8LD**

Contact Details 02034561044, adam.roberts@ucl.ac.uk

We would like to invite you to participate in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

Details of Study: Bacteria are found in the oral cavity living in communities these can shift to become either associated with oral health or with disease. This project is about understanding what makes up these communities, what is different between health and disease communities and how we can use different antimicrobial agents, prebiotics or probiotics to perturb these communities. The numbers of oral bacteria associated with both health and disease will be monitored using both culture (growth on selective and non selective media) and molecular biology methods (based on bacterial nucleic acid analysis). This will allow the research team to determine whether the components are having any effect on the number of disease-associated and health-associated bacteria in the mouth.

During the bacterial nucleic acid isolation method, there is a possibility that human nucleic acid will also be recovered, however, any such products will not be investigated here or used for further research projects.

Saliva samples would be obtained from volunteers by asking them to expectorate into a sterile plastic tube until a volume of 2 ml is reached.

If you do decide to take part, please let us know beforehand if you have been involved in any other study during the last year or if you have had antibiotic treatment in the 3 months.

It is up to you to decide whether or not to take part. If you choose not to participate it will involve no penalty or loss of benefits to which you are otherwise entitled. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

All data will be collected and stored in accordance with the Data Protection Act 1998.

Please discuss the information above with others if you wish or ask us if there is anything that is not clear or if you would like more information.

It is up to you to decide whether to take part or not; choosing not to take part will not disadvantage you in any way. If you do decide to take part you are still free to withdraw at any time and without giving a reason.

All data will be collected and stored in accordance with the Data Protection Act 1998.

When you have completed your Information Sheet, please DELETE the advice section below from your application form before submitting it to the Committee.

Informed Consent Form for Participants in Research Studies

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Project: **Understanding oral microbial ecology**

This study has been approved by the UCL Research Ethics Committee (Project ID Number): **5017/001**

Thank you for your interest in taking part in this research. Before you agree to take part, the person organising the research must explain the project to you.

If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you to decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

Participant's Statement

I

- have read the notes written above and the Information Sheet, and understand what the study involves.
- understand that if I decide at any time that I no longer wish to take part in this project, I can notify the researchers involved and withdraw immediately.
- consent to the processing of my personal information for the purposes of this research study.
- understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
- agree that the research project named above has been explained to me to my satisfaction and I agree to take part in this study.

Signed:

Date:

Appendix 2. Composition of media and solutions

Media/solutions	Composition
Luria-Bertani (LB) broth	10g/l Tryptone 5 g/l Yeast Extract 5 g/l NaCl
Luria-Bertani (LB) agar	15g/l Agar 10g/l Tryptone 5 g/l Yeast Extract 5 g/l Sodium chloride (NaCl)
SOC	2% Tryptone 0.5% Yeast extract 10 mM Sodium chloride (NaCl) 2.5 mM Potassium chloride (KCl) 10 mM Magnesium chloride (MgCl ₂) 10 mM Magnesium sulphate (MgSO ₄) 20 mM Glucose
Z buffer	8.0 g Disodium hydrogen phosphate heptahydrate (Na ₂ HPO ₄ •7H ₂ O) 2.75 g Sodium dihydrogen phosphate monohydrate (NaH ₂ PO ₄ •H ₂ O) 0.375 g Potassium chloride (KCl) 0.125 g Magnesium sulphate heptahydrate (MgSO ₄ •7H ₂ O) dH ₂ O to 500 ml, pH7 Add 0.14 ml 2-mercaptoethanol in 50 ml Z buffer prior to usage (50 mM)
B-glucuronidase assay stop solution (1 M Na₂CO₃)	5.3 g Na ₂ CO ₃ dH ₂ O to 50 ml
50X Tris-acetate-EDTA (TAE)	242 g Tris Base 57.1 ml Glacial Acetic Acid 18.61 g EDTA dH ₂ O to 1l, pH 8.0

Media/solutions	Composition
Phosphate-buffered saline (PBS)	8.5 g/l Sodium chloride 1.91 g/l Disodium hydrogen phosphate 0.380 g/l Potassium dihydrogen phosphate dH ₂ O to 500 ml, pH 7.2

Appendix 3. Antimicrobial Working/Stock Concentrations.

Antimicrobials (Sigma Aldrich)	Solvent	Stock concentration	Working concentration
Ampicillin	70% Ethanol	100 mg/ml	100 µg/ml
Chloramphenicol	70% Ethanol	25 mg/ml	12.5 µg/ml
Tetracycline	70% Ethanol	10 mg/ml	5 µg/ml
Kanamycin	Steriled water	50 mg/ml	20 µg/ml
Trimethoprim	Dimethyl sulfoxide (DMSO)	20 mg/ml	20 µg/ml
Cetyltrimethylammonium bromide (CTAB)	70% Ethanol	10 mg/ml	8 µg/ml
Cetylpyridinium chloride (CPC)	70% Ethanol	10 mg/ml	6 µg/ml
Copper Sulphate	Steriled water	64 mg/ml	4 mg/mL
Silver Nitrate	Steriled water	20 mg/ml	18 µg/mL
Chlorhexidine	Steriled water	10 mg/ml	1.8 µg/mL
Sodium Benzoate	Steriled water	144 mg/ml	18 mg/mL

Appendix 4. Primers used in this study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
Primers designed based on an unusual <i>Treponema denticola</i> integrons (Accession number: NC_002967)			
SUPA3	CAGGTTGAAGCGGGTGTTAG	<i>attC</i> site (reverse primer)	This study
SUPA4	CCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
SUPA5	CAGVTTGAAGCGRRYGTTAG	<i>attC</i> site (reverse primer)	This study
SUPA6	CCRCAAATGYWGGTYAAGCG	<i>attC</i> site (forward primer)	This study
Flip-SUPA3	CTAACACCCGCTTCAACCTG	<i>attC</i> site (forward primer)	This study
Flip-SUPA4	CGCTTAACCTGCATTTGCGG	<i>attC</i> site (reverse primer)	This study
Primers designed based on normal integrons			
HS286	TCSGCTKGARCGAMTTGTTAGVC	<i>attC</i> site (reverse primer)	Stokes <i>et al.</i> , 2001
HS287	GCSGCTKANCTCVRRRCGTTAGSC	<i>attC</i> site (forward primer)	Stokes <i>et al.</i> , 2001
Flip-HS286	GBCTAACAAKTCGYTCMAGCSGA	<i>attC</i> site (forward primer)	This study
Flip-HS287	GSCTAACGYBYBGAGNTMAGCSGC	<i>attC</i> site (reverse primer)	This study
5'-CS	GGCATCCAAGCAGCAAG	5'-conserved segment (5'-CS)	Martinez-Freijo <i>et al.</i> , 1998

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
3'-CS	AAGCAGACTTGACCTGA	3'-conserved segment (3'-CS)	Martinez-Freijo <i>et al.</i> , 1998
HS298	ACRTGNGTRTADATCATNGT	Conserved C-terminal sequence in <i>intl</i>	Nield <i>et al.</i> , 2001
Primers for the verification of circular gene cassettes			
SSU1-F	AACGGTTGGTGGGCATATTA	SSU1 clone	This study
SSU1-R	TCAATATCTGGATCGCCACA	SSU1 clone	This study
SSU3-F	GGCGGGATATATGATGATCG	SSU3 clone	This study
SSU3-R	CGCACCATCGTCGGCATATC	SSU3 clone	This study
SSU5-F	GCATATAACCACCTCCTGCA	SSU5 clone	This study
SSU5-R	GCCCTAGCAGAAGGAATGAA	SSU5 clone	This study
SSU6-F	CACAGTTAAATGCTTCCGGC	SSU6 clone	This study
SSU6-R	GAACACCATGTGGGAGGAAC	SSU6 clone	This study
SSU7-F	TTATCGGGGTATTGGAGCAG	SSU7 clone	This study
SSU7-R	CAATCCGGCATCTCAAGAAT	SSU7 clone	This study
SSU8-F	ATTGCTTGGGAAGGTTTTCA	SSU8 clone	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
SSU8-R	TTCCGTCTCCAGGAAGAATG	SSU8 clone	This study
SSU11-F	GAAACCGCTCCGTTTATTGA	SSU11 clone	This study
SSU11-R	GCCAATAAATGAGGATTCCA	SSU11 clone	This study
SSU12-F	CACCCTTGGATGTAATTAGTG	SSU12 clone	This study
SSU12-R	TTCGCTTCATTCTCCTGACC	SSU12 clone	This study
SSU15-F	CCCAACAAGGGAACATCTGG	SSU15 clone	This study
SSU15-R	GACCTTCCTTTCGCAATGCC	SSU15 clone	This study
SSU16-F	ACCAATAAAACCGGCTGATAATC	SSU16 clone	This study
SSU16-R	CGACCTTCCTTTCGCAATGC	SSU16 clone	This study
SSU17-F	TCCAGTTACACTTGCGCATC	SSU17 clone	This study
SSU17-R	CTTTCTCTATTTTAAGGTTCGGG	SSU17 clone	This study
SSU18-F	CCCATGCAGAGAGATGATTG	SSU18 clone	This study
SSU18-R	CCGTAGTAGATAAGCCGATG	SSU18 clone	This study
SSU21-F	AGCTGCCAAAACACTATGCTTATG	SSU21 clone	This study
SSU21-R	CATAAACCCCAACCGTCC	SSU21 clone	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
SSU22-F	AGTATGAAGACGGTAGCTATGAG	SSU22 clone	This study
SSU22-R	TTACGCAGATCCGAAACAGG	SSU22 clone	This study
SSU24-F	TGAAGTGCCGTATGATGGTG	SSU24 clone	This study
SSU24-R	TTCAAACCCCAATATATCCCTG	SSU24 clone	This study
SSU25-R	ACTGCTTCTTCACCAGTTTC	SSU25 clone	This study
SSU25-F	CCATGGTTAGTTTGGGGAAA	SSU25 clone	This study
SSU26-F	CGCCATCATTATCGTTTGTG	SSU26 clone	This study
SSU26-R	AACGAAGAACCATCCACACA	SSU26 clone	This study
SSU27-F	AATCGCTTTCACGGTTCCTA	SSU27 clone	This study
SSU27-R	TGCTGTTCCAAAAGCAAGTG	SSU27 clone	This study
SSU28-F	CCTGTGCTTTGGCAACTACA	SSU28 clone	This study
SSU28-R	ATGAGCAGAAAGGTCGTCGT	SSU28 clone	This study
SSU29-R	CTTCGGACGGTCGGGTAAT	SSU29 clone	This study
SSU29-F	GGCACTTTCACGCTTTACCT	SSU29 clone	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
Primers for the functional screening of oral gene cassettes			
SUPA4-<i>Pst</i>I	GCCCCCTGCAGCCGCAAATGCAGGTTAAGCG	Integron <i>attC</i>	This study
SUPA4-<i>Sa</i>I	GCCCCGTGCGACCCGCAAATGCAGGTTAAGCG	Integron <i>attC</i>	This study
SUPA4-<i>Xba</i>I	GCCCCGTCTAGACCGCAAATGCAGGTTAAGCG	Integron <i>attC</i>	This study
SUPA3-<i>Bam</i>HI	GCGGCGGGATCCCAGGTTGAAGCGGGTGTTAG	Integron <i>attC</i>	This study
SUPA3-<i>Sac</i>I	GCGGCGGAGCTCCAGGTTGAAGCGGGTGTTAG	Integron <i>attC</i>	This study
SUPA3-<i>Acc</i>65I	GGCGGCGGTACCCAGGTTGAAGCGGGTGTTAG	Integron <i>attC</i>	This study
Primers for the cloning of GCs into pUC19-<i>gusA</i> constructs			
SUPA4-<i>Kpn</i>I	CGCGCGGGTACCCCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
SUPA3-<i>Eco</i>RI	CGCGCGGAATTCCAGGTTGAAGCGGGTGTTAG	<i>attC</i> site (reverse primer)	This study
SUPA4-<i>Eco</i>RI	GCGGCCGAATTCCCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
SUPA3-<i>Kpn</i>I	CGCGCGGGTACCCAGGTTGAAGCGGGTGTTAG	<i>attC</i> site (reverse primer)	This study
TMB4-Pc-F1-<i>Kpn</i>I	CGGCCGGGTACCTGCGTGCGTTATCCCATTTA	TMB4 Pc promoter (forward primer)	This study
TMB4-Pc-R1-<i>Eco</i>RI	CGCGCGGAATTCCATCTTTTCGACCTTTCCTC	TMB4 Pc promoter (reverse primer)	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
TMB4-GC-F1-KpnI	CGCGCGGGTACCTTAGACAGATGCCTTGCGG	TMB4 gene cassette (forward primer)	This study
TMB4-GC-F1-EcoRI	CGCGCGGAATTCCTTAGACAGATGCCTTGCGG	TMB4 gene cassette (forward primer)	This study
TMB1-F1-KpnI	CGCCGGGGTACCCGATCTTTCTTTTTCCGTT	TMB1 Pc promoter (forward primer)	This study
TMB1-R1-EcoRI	CCCGCCGAATTCCTTCCCTTCGACCCTTCCT	TMB1 Pc promoter (reverse primer)	This study
Pfla-A	CGACTTTTTTCCTAAACCCGCCTTAAA AATAAGCCGAAAATTTATTGAAGTAAC ATAGGATCAATGTATAGGAGGTTTCATG	P _{fia} sense strand oligo	Limberger et al., 1999
Pfla-B	AATTCATGAACCTCCTATACATTGATCCTA TGTTACTTCAATAAATTTTCGGCTTATTTTT AAGGCGGGTTTAGGAAAAAAGTCGGTAC	Pfla antisense strand oligo	Limberger et al., 1999
Tdtrop/O-A	CGGGATTCAGCTTGACATTTTCTTTAT TTTTTATATATTATAATCATAATTTTG ATATATCAAATAGGAGATTTGAAG	P _{TdTro} sense strand oligo	Brett et al., 2008
Tdtrop /O-B	AATTC TTCAAATCTCCTATTTTGATATATC AAAATTATGATTATAATATATAAAAAAATA AAGAAAATGTCAAGCTGAATCCC GGTAC	P _{TdTro} antisense strand oligo	Brett et al., 2008

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
Primers for the cloning of GCs into pCC1BAC-<i>gusA</i> constructs			
<i>gusA</i> -F4- <i>Hind</i> III	CGCGCGAAGCTTAGGCGATTAAGTTGGGTAAC	The GC from pUC19-constructs to put in pCC1BAC	This study
Primers for the construction of pCC1BAC-<i>lacZα</i>-<i>gusA</i> vectors.			
<i>lacZ</i> -F1	CGGCGCGACGTCAGAGAATATAAAAAGCCGGATTATTA ATCCGGCTTTTTTATTATTGTCTCGGGGCTGGCTTA ACTAT	<i>lacZα</i> (Forward primer)	This study
<i>lacZ</i> -R1	GCGGCGATGCATTGTGAGCGGATAACAATTTTC	<i>lacZα</i> (Reverse primer)	This study
<i>gusA</i> -F1	CGCGCGATGCATGCTAGCATCACGAATTCCTGCAGTAA	<i>gusA</i> (Forward primer)	This study
<i>gusA</i> -R1	GGGCGGCCTAGGAAATAATAAAAAAGCCGGATTAAATAA TCCGGCTTTTTATATTCTCTCGCCAGGAGAGTTGTTGATT	<i>gusA</i> (Forward primer)	This study
pCC1BAC-del <i>lacZ</i> -F1	CCCCCCCCTAGGCCGTCGACCAATTCTCATGT	pCC1BAC- <i>lacZα</i> deletion (Forward primer)	This study
pCC1BAC-del <i>lacZ</i> -R1	GCGGCGGACGTCTAGTTAAGCCAGCCCCGACA	pCC1BAC- <i>lacZα</i> deletion (reverse primer)	This study
Primers for the amplification of GCs and promoters into pCC1BAC-<i>lacZα</i>-<i>gusA</i> vectors.			
pUC-GC-F1- <i>Nsi</i> I	CGCGCGATGCATTTGTAATTCGACGGCCAGTG	GC on pUC19-GC- <i>gusA</i> construct (forward primer)	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
pUC-GC-R1-NheI	GGGCGGGCTAGCATTCTCCGCTACTCCAGG	GC on pUC19-GC- <i>gusA</i> construct (reverse primer)	This study
pUC-GC-F1-NheI	CGCGCGGCTAGCTTGTAAATTCGACGGCCAGTG	GC on pUC19-GC- <i>gusA</i> construct (forward primer)	This study
pUC-GC-R1-NsiI	GGGCGGATGCATATTTCTCCGCTACTCCAGG	GC on pUC19-GC- <i>gusA</i> construct (reverse primer)	This study
SUPA4-NsiI	GCGGCCATGCATCCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
SUPA3-NheI	CGCGCGGCTAGCCAGGTTGAAGCGGGTGTTAG	<i>attC</i> site (reverse primer)	This study
SUPA4-NheI	CGCGCGGCTAGCCCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
SUPA3-NsiI	CGCGCGATGCATCAGGTTGAAGCGGGTGTTAG	<i>attC</i> site (reverse primer)	This study
MARS5-NsiI	GCGGCCATGCATCGCAAATGCAGGTTAAGCG	<i>attC</i> site (forward primer)	This study
MARS2-NheI	CGCGCGGCTAGCGCAATGTCAGGTTGAAGC	<i>attC</i> site (reverse primer)	This study
Primers to check the presence of insertion sequences			
IS3-F1	AGCTGGAAAGCCACGTGTAG	IS3	This study
IS3-R1	TATCAGCGCTGCTTGCTTTA	IS3	This study
IS26-F1	CTGCACATGAACCCATTCAA	IS26	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
IS26-R1	CTCTGCTTACCAGGCGCATT	IS26	This study
IS240-F1	GCGTGATCAGGATGGAAAAA	IS240	This study
IS240-R1	CTGATTTTGGGCAGATTGTG	IS240	This study
IS256-F1	CGACAAAAACATACCCAGGA	IS256	This study
IS256-R1	GTCCATAAGAACGGCTCCAA	IS256	This study
IS257-F1	GTTATCACTGTAGCCGTTGG	IS257	This study
IS257-R1	CATGGCGAAAATCCGTAGAT	IS257	This study
IS861-F1	TTGCTTTGATTCGCTTGTTG	IS861	This study
IS861-R1	GGGATGTCCTGGTCTCAGAT	IS861	This study
IS1161-F1	CTGATAGCCATGAATTCGTC	IS1161	This study
IS1161-R1	CACTTATCTGAAGCTGAGCG	IS1161	This study
IS1167-F1	CCAAACTGGACTACGACGCC	IS1167	This study
IS1167-R1	GGCATTGCGCTTGATAAGTT	IS1167	This study
IS1216-F1	AAGAAGGCACTCTCTTCGGG	IS1216	This study
IS1216-R1	CGAAGATAGTAGCCCACGGC	IS1216	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
IS1381-F1	CACATAACGGATCGATAGGG	IS1381	This study
IS1381-R1	GGTGAAAACGGACTTGGAAC	IS1381	This study
IS1485-F1	TAGCTGGAAAGCCACGTGTA	IS1485	This study
IS1485-R1	GCTCATTATCAAGGATTCGC	IS1485	This study
IS1548-F1	ATCTCTTCATCCTTTTGTGC	IS1548	This study
IS1548-R1	GAGGTGACTACGATGATTGA	IS1548	This study
Primers to amplify the composite transposons			
IS26-F2	CTGCACATGAACCCATTCAA	Outward from IS26	This study
IS26-R2	CTCTGCTTACCAGGCGCATT	Outward from IS26	This study
IS257-F2	GTTATCACTGTAGCCGTTGG	Outward from IS257	This study
IS257-R2	CATGGCGAAAATCCGTAGAT	Outward from IS257	This study
IS1161-F2	GACGAATTCATGGCTATCAG	Outward from IS1161	This study
IS1161-R2	CGCTCAGCTTCAGATAAGTG	Outward from IS1161	This study
IS1167-F2	GGCGTCGTAGTCCAGTTTGG	Outward from IS1167	This study
IS1167-R2	AACTTATCAAGCGCAATGCC	Outward from IS1167	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
IS1216-F2	AAGAAGGCACTCTCTTCGGG	Outward from IS1216	This study
IS1216-R2	CGAAGATAGTAGCCACGGC	Outward from IS1216	This study
IS1485-F2	TAGCTGGAAAGCCACGTGTA	Outward from IS1485	This study
IS1485-R2	GCTCATTATCAAGGATTCGC	Outward from IS1485	This study
Primers for the amplification of TUs			
CTA1-F1	TTTCTTGCCACCGAAACTGC	TU form of CTA1	This study
CTA1-R1	GCGAGTTTTTCGTTTTCCAGT	TU form of CTA1	This study
CTA2-F1	TTTCTTGCCACCGAAACTGC	TU form of CTA2	This study
CTA2-R1	CAATGTCTTGCGTGGTCTCG	TU form of CTA2	This study
CTA3-F1	GGCAATCAGTAATGCGGCTC	TU form of CTA3	This study
CTA3-R1	ACAGTGGAGAAGGTGATCCG	TU form of CTA3 and 4	This study
CTA4-F1	TAACGCCATAATCACGCAGA	TU form of CTA4	This study
CTA5-F1	TTTTCCGTTCCCAATTCCAC	TU form of CTA5	This study
CTA5-R1	CATGGCGCATTAAACGGAATA	TU form of CTA5	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
Primers for the construction of BACpAK entrapment vectors			
<i>cl-tetA</i> (F)-XhoI	GCCCCCTCGAGCAGCCAGCAGAGAATTAAGG	<i>cl-tetA</i> on pAK1 entrapment vector	This study
<i>cl-tetA</i> (R)-XhoI	GCGGCGCTCGAGGATAGCCGCTGCTGGTTT	<i>cl-tetA</i> on pAK1 entrapment vector	This study
Primers for the construction of BACpAK-promoterless entrapment vectors			
BACpAK-del-<i>cl-tetA</i>-F1	TGCGGGCTAGCTGACATAGGGCCTATAGGATCTATAGGCCCTTTTTCTTATTtCTTGTTATGCCGGTACTGC	Deletion of <i>cl</i> from BACpAK	This study
BACpAK-del-<i>cl-tetA</i>-R1	GCGGCGACGTCCTCACTTCAACGTAACACCG	Deletion of <i>cl</i> from BACpAK	This study
Tdtro-F1	CGCGCGGACGTC TTGTAATTCGACGGCCAGTG	Tdtro from pCC1BAC- <i>lacZ</i> -Tdtro- <i>gusA</i> plasmid	This study
Tdtro-R1	GCCGCGGCTAGCGCTACTCCAGGAATTCTTCA	Tdtro from pCC1BAC- <i>lacZ</i> -Tdtro- <i>gusA</i> plasmid	This study
Primers to check the insertion within <i>cl-tetA</i> selection cartridge			
ERIS	GCAAGACTGGCATGATAAGG	<i>cl-tetA</i>	Bartosik et al., 2003
<i>cl-tetA</i> -F1	CAGCCAGCAGAGAATTAAGG	<i>cl-tetA</i>	This study
<i>cl-tetA</i> -F3	TCTATCACCGCAAGGGATAA	<i>cl-tetA</i>	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
Primers to detect the presence of TEs in the oral metagenome and <i>E. coli</i> host			
K6-1-F1	CCTCTGGCAACGTGAATGTA	<i>Tn1000</i>	This study
K6-1-R1	TTCTGGCGGGTGAATAAATC	<i>Tn1000</i>	This study
T6-6-F1	AGCTGGGTAATCTGCTGCAT	IS5	This study
T6-6-R1	TTTCTCCCGTAAATGCCTTG	IS5	This study
Primers for the construction of BACpAK-IS5			
IS5-BamHI-F1	CGCGCGGGATCCGGGCAAACCAAGACAGCTAA	IS5 for cloning	This study
IS5-BamHI-R2	GCGGCGGGATCCGACGATGAGCGCATTGTTAG	IS5 for cloning	This study
Primers for the amplification of GATC-containing fragments			
BPPB2-F2-BamHI	GCCCCCGGATCCAGCTAAAGATCTGAGAACGG	BPPB2-BamHI	This study
BPPB2-R2-BamHI	GCGGCGGGATCCAAGCTGATCCTGTCATTATG	BPPB2-BamHI	This study
BPPW3-F1-BamHI	GCCCCCGGATCCGACAGCTAAAGATCCTTTAAGAC	BPPW3-BamHI	This study
BPPW3-R1-BamHI	GCGGCGGGATCCGCTGATCATTGCTGGTG	BPPW3-BamHI	This study
AmpR-F1-BamHI	GCCGGCGGATCCACGCTCAGTGGAACGAAAAC	AmpR on pGEM-T easy vector	This study
AmpR-R1-BamHI	GCGGCGGGATCCCCTGATGCGGTATTTTCTCC	AmpR on pGEM-T easy vector	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
<i>tet(M)</i> -F1- <i>Bam</i> HI	CGCGCC GGATCC ACCAAAGCAACGCAGGTATC	<i>tet(M)</i> on Tn916	This study
<i>tet(M)</i> -R1- <i>Bam</i> HI	GCGGCG GGATCCT CGGACAATAGAGGGGGAAT	<i>tet(M)</i> on Tn916	This study
Primers for amplification of <i>gpi</i>, <i>galE</i> and <i>gpi-galE</i> genes for subcloning			
<i>gpi</i> -F1	CGCGCG AAGCTTTT TATGCAGTGGGGTTTGGTT	<i>gpi</i> , <i>gpi-galE</i>	This study
<i>gpi</i> -R1	CCCGCC GAATTC CCCCCTTCAGGAATAGATTCT	<i>gpi</i>	This study
<i>galE</i> -F1	CGCGCG AAGCTT GCTCGTGCAAAGGATACACA	<i>galE</i>	This study
<i>galE</i> -R1	GCGGCG GAATTC CAAATCAAACCGATTCATGC	<i>galE</i> , <i>gpi-galE</i>	This study
Primers for sequencing			
M13F	GTAAAACGACGGCCAG	M13 forward sequencing (pGEM-T easy and pUC19 vector)	Universal
M13R	CAGGAAACAGCTATGAC	M13 forward sequencing (pGEM-T easy and pUC19 vector)	Universal
<i>lacZ</i> -F2	GTTTTCCAGTCACGACGTT	Insert forward sequencing	This study
<i>gusA</i> -R2	CGGCGAACTGATCGTTAAAA	Insert reverse sequencing	This study
pCC1-F	GGATGTGCTGCAAGGCGATTAAGTTGG	Inserts in pCC1BAC vector	Epicenter, UK
pCC1-R	CTCGTATGTTGTGTGGAATTGTGAGC	Inserts in pCC1BAC vector	Epicenter, UK

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
SeqW	CGACACACTCCAATCTTTCC	Genes franking Entranceposon	Thermo scientific, UK
SeqE	GGTGGCTGGAGTTAGACATC	Genes franking Entranceposon	Thermo scientific, UK
CTA1-F2	CCACATTCGGGACAACAAGT	CTA1	This study
CTA4-F2	TTTAAACCGGCTTCCCATT	CTA4	This study
CTA4-F3	TAAAGACTGCGACGACAAGC	CTA4	This study
CTA4-F4	CCTCAGTGACGGCATTATCA	CTA4	This study
CTA4-R2	CTGGAGCCGCATTAAGT	CTA4	This study
CTA1-F2	CCACATTCGGGACAACAAGT	CTA1	This study
BPPB1-R1	CGATGCTAGGAAGAACTGC	BPPB1	This study
BPPB2-R1	TTCTTGGAGGCTTTGCTCAT	BPPB2	This study
BPPW1-R1	GCCCACCTTGTGACATTTCT	BPPW1	This study
A10F2-F2	GGATCTTGTTGAGAATAAGG	A10F2 insert	This study
A10F2-R2	CTGTTATTACCCCTAGTGC	A10F2 insert	This study
A10F2-F3	GGACAGTGCGGCGAGTG	A10F2 insert	This study

Primer name	Sequence (5'-3') ^{ab}	Gene target	Reference
A10F2-R3	GTCGAACCATAAATTGAAC	A10F2 insert	This study
A10F2-F4	CAATTCCCATCACGGAGTTC	A10F2 insert	This study
A10F2-R4	CCTCTCTATTGGCATCTAAG	A10F2 insert	This study
A10F2-F5	CCTTCAGTACTTCTTTGTC	A10F2 insert	This study
A10F2-R5	GATGGTAGAACCTGACTTAG	A10F2 insert	This study
A10F2-F6	GCTTTGCCACCTCAACAC	A10F2 insert	This study
A10F2-R6	CCACTGCATAAACAGACC	A10F2 insert	This study
A10F2-F7	CAACCACTTAAAGATAAGC	A10F2 insert	This study
A10F2-R7	CATTGGCATATCTAAAGC	A10F2 insert	This study
A10F2-F8	CGGATCGTGCCTATCTTCTC	A10F2 insert	This study
A10F2-R8	ATCTGGTTTTGCCTCACCTG	A10F2 insert	This study

^a Degenerate nucleotides: D = A, G, or T; H = A, C, or T; I or N = A, C, G, or T; K = G or T; M = A or C; R = A or G; S = G or C; V = A, C, or G; Y = C or T

^b Restriction sites and bi-directional terminators were indicated as bold and italic styles, respectively.

Appendix 5. Publications resulting from this study

1. Tansirichaiya S, Rahman MA, Antepowicz A, Mullany P & Roberts AP (2016) Detection of novel integrons in the metagenome of human saliva. *PloS one* **11**: e0157605.
2. Tansirichaiya S, Mullany P & Roberts AP (2016) PCR-based detection of composite transposons and translocatable units from oral metagenomic DNA. *FEMS Microbiology Letters* **363**: fnw195-fnw195.
3. Tansirichaiya S, Reynolds LJ, Cristarella G, Wong LC, Rosendahl K & Roberts AP (2017) Reduced Susceptibility to Antiseptics Is Conferred by Heterologous Housekeeping Genes. *Microbial drug resistance*.