

**Identification of Extrachromosomal Elements from Whole
Genome Sequences of the Human Gut Microbiome to
Investigate the Gut Mobilome and Resistome**



**UNIVERSITY OF
CAMBRIDGE**



Tapoka Thulisile Mkandawire
USN:303635303

University of Cambridge
Murray Edwards College

*This dissertation is submitted for the degree Master of Philosophy in Biological Science @
The Sanger Institute*

August, 2017

Table of Contents

1. Introduction	
1.1 The Gut Microbiome	5
1.2 Residents of the Gut Microbiota	5
1.3 The Role of the Microbiota in Health and Disease	7
1.4 Plasmid Biology	8
1.5 Plasmid Replication and Maintenance	10
1.6 Horizontal Gene Transfer	11
1.7 Frequency of and Challenges to HGT	13
1.8 The Utility of Plasmids as Tools for Biotechnology	14
1.9 Plasmids and the Gut Microbiome	15
1.10 Plasmid Functions in the Gut Microbiome	16
1.11 HGT in the Gut Microbiome	17
1.12 Plasmid Capture Strategies	18
1.13 Challenges to Plasmid Capture	19
1.14 Hypothesis	20
1.15 Project Aims	20
1.16 Project Rationale and Identified Challenges	20
2. Methods	
2.1 Computational Plasmid Isolation, Phenotypic Predictions, and Phylogeny	22
2.2 Bacterial Culture and Growth Curve Plotting	23
2.3 Growth Medium	24
2.4 Species Validation by 16S rRNA gene PCR	24
2.5 Plasmid Extractions and Digests	25
2.6 Plasmid Visualisation by Gel Electrophoresis	25
2.7 Antimicrobial Sensitivity Testing	25
2.8 Transformation	25
2.9 Conjugation	26
3. Bioinformatics Results	
3.1 Frequency of Plasmid Detection Across Human Gut Microbiota Phyla	27
3.2 Size and Coverage and Distribution of Predicted Plasmids	29
3.3 Plasmid Classification and Phylogeny	30
3.4 AMR Gene Distribution	35
4. Experimental Results	
4.1 Plasmid Isolation	37
4.2 Species Validation	37
4.3 Plasmid Digests	38
4.4 Antimicrobial Screening	39
4.5 Strain Growth Monitoring	40
4.6 Plasmid Transfer	41
5. Discussion	
5.1 Distribution of Plasmids in the Human Gut Microbiome Culture Collection	45
5.2 Genomic Inference of Biological Functions: Plasmid Classification and AMR Distribution	46
5.3 Transformation of Culture Collection Isolates	48
5.4 Applications of Identified Plasmid Sequences	49
5.5 Plasmids and Plasmid Sequence Data as Genetic Tools	49
5.6 Strategies to Engineer the Microbiota	51
5.7 Challenges to Bacteriotherapy	53
5.8 Non-Health Associated Uses of Engineered Microbiota	54
5.9 Strategies for Investigating Engineered Microbiota	54
5.10 Summary	

List of Figures and Tables

1. Introduction	
1.1. Figure: Distribution of bacterial phyla and environmental composition along the GIT	6
1.2. Figure: Phylogenetic tree of the Lawley lab human gut microbiota culture collection	6
1.3. Figure: Diagram of the conformations of plasmids	9
1.4. Figure: Summary of the process of rolling circle replication	10
1.5. Figure: Figure summarising the methods of horizontal gene transfer	11
1.6. Figure: Phylogenetic tree displaying the distribution of bacteria in environmental water sources	13
1.7. Figure: Chart displaying the frequency with which a variety of vectors are used in gene therapy trials	15
1.8. Figure: Graph displaying the distribution of RelE genes in the gut microbiome	16
1.9. Figure: Flow diagram outlining the variety of processing routes of stool samples for plasmid analysis	19
2. Methods	
2.1. Figure: Workflow diagram summarising the strategy taken to probe the plasmids of the human gut microbiome	22
3. Bioinformatic Results	
3.1. Figure: The distribution of predicted plasmids in the 653 genomes of the Lawley Laboratory culture collection	28
3.2. Figure: Lawley Lab culture collection phylogeny with plasmid prediction annotations	29
3.3. Figure: Detected plasmid coverage compared to chromosomal median coverage and plasmid distribution by size	30
3.4. Figure: Phylogenetic tree constructed using the <i>repE</i> gene	31
3.5. Figure: Phylogenetic tree constructed using the <i>repA</i> gene	32
3.6. Figure: Chord diagram displaying links between genomes containing closely related <i>repE</i> genes.	33
3.7. Figure: Chord diagram displaying links between genomes containing closely related <i>repA</i> genes	34
3.8. Figure: Percentage bar graph displaying the proportions of predicted resistance	36
4. Experimental Results	
4.1. Figure: Agarose gel displaying the isolated plasmids greater than 10kb	37
4.1. Table: Results of the 16S species validation	38
4.2. Figure: Agarose gel displaying small plasmids in native (U) and digested (D) form	39
4.2. Table: Antibiotic resistance levels as observed with Etest strips	39
4.3. Figure: Agarose gel displaying mid size and large plasmids in native (U) and digested (D) form.	39
4.4. Figure: OD growth curves for selected isolates	41
4.5. Figure: Results of the conjugation between the donors (D) and naladixic acid engineered recipient (R)	42
4.6. Figure: Results of the conjugation between the donors (D) and erm-resistant bacteria (R)	43
4.7. Figure: Results of the conjugation between donor H5_29 and aerobic bacterial recipients	44
5. Discussion	
5.1. Figure: The three key synthetic circuits: dynamic, logic, and communication circuits	52

Acknowledgements

I'd like to thank my supervisor Dr Trevor Lawley and the members of Team 162: Anne, Sam, Nitin, Lindsay, Yan *et al.*, for all the support and input. As well as my thesis committee for their invaluable insights: Dr David Summers, Prof Nick Thomson.

Abstract

We are covered with beneficial microbial communities, termed microbiota, that play important roles in our health, sustenance and well-being. Pathological imbalances in our microbiota caused by things like antibiotic exposure and poor diet can directly cause, or predispose us to, a variety of diseases and metabolic syndromes. Our intestinal microbiota contains hundreds of uncultured bacterial species and the genetic complement of our microbiota, our microbiome, contains 150 times more genes than the human genome. The human microbiome is dynamic due to bacteria's ability to horizontally transfer large blocks of DNA between distantly related species. Horizontal gene transfer by mobile elements such as transposons and plasmids plays a central role in the evolution and functions of well-studied intestinal bacteria such as *Escherichia coli* and *Lactobacillus*. However, most of the human microbiota has never been cultured and characterized so the composition of mobile elements in the majority of the human microbiota remains poorly defined. For my MPhil Thesis, I analysed a large scale human commensal reference genome collection of 653 genomes, representing the phylogenetic diversity of the human intestinal microbiota, for the presence of extrachromosomal elements like plasmids and transposons. A combination of bioinformatics analyses and wet-lab validation methods have been employed to identify, isolate, and characterize elements from gut bacteria creating a catalogue of known and novel sequences. 240 genomes were predicted to contain extrachromosomal DNA, these elements are primarily small and high coverage, and predicted to contain resistance genes. A small number of megaplasmids were also detected, and a phylogeny was built to identify any plasmids with a broad host range. This database can be used as a reference for the computational isolation of extrachromosomal elements from whole genome sequencing data and metagenomic datasets. They will also form the foundation for developing tools for genetic manipulation of the novel and uncharacterized gut commensal microbes of the culture collection.