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

List of Figures and Tables

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.