CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	An in silico analysis of bacteriocin production in the human microbiota and its relationship with health
Authors	Walsh, Calum J.
Publication date	2017
Original Citation	Walsh, C. 2017. An in silico analysis of bacteriocin production in the human microbiota and its relationship with health. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2017, Calum Walsh. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-10-27 08:33:23
Item downloaded from	https://hdl.handle.net/10468/4708

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

An *in silico* analysis of bacteriocin production in the human microbiota and its relationship with health

A thesis presented to the National University of Ireland for the degree of

Doctor of Philosophy

By

Calum Walsh M.Sc.

Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland

Department of Microbiology, University College Cork, Cork, Ireland

May 2017

Research supervisors: Dr. Paul Cotter, Professor Paul O'Toole and Professor Paul Ross

Table of Contents

Declaration …….…… i

Abstract ……. ii

Publications …….. iv

List of Abbreviations ……..…….. v

List of Figures ……..…….. viii

List of Tables ……..…….. xi

Chapter 1

Beneficial modulation of the gut microbiota

Abstract ……... 2

Introduction …….……3

Role of the Gut Microbiota in Health and Disease ….…… 8

Inflammatory Bowel Disease …….…… 8

Irritable Bowel Syndrome …….……10

Obesity …….……. 11

Type 2 Diabetes …… 13

Colorectal Cancer ……. 14

Modulation of the Gut Microbiota …….…….. 16

Modulation by Diet ……..……. 16

Modulation by Antimicrobials ……. 25

Identification of antimicrobials with the potential to modulate the gut microbiota ……. 31

Modulation by Probiotics …….…… 35

Modulation by Faecal Microbiota Transplantation …….……. 37

Concluding Remarks …… 39

References …… 40

Chapter 2

In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database

Abstract ……. 74

Introduction …….……. 75

Methods ……..…….. 78

Initial screening of reference genomes for bacteriocin gene clusters …….. 78

Further investigation of individual gene clusters ……. 78

Results and Discussion …….…… 79

In silico screen for putative bacteriocin-encoding gene clusters ….……. 79

Further analysis of PBGCs of particular interest ……..……87

Identification of novel PBGCs in bifidobacteria …….……. 87

Identification of novel PBGCs in Bacteroides spp. ……..…… 95

Identification of novel PBGCs in Ruminococcus spp. …….…… 99

Identification of a novel PBGC in Roseburia spp. …….….101

Conclusions ……... 102

References …….. 105

Chapter 3

A Profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome

Abstract ……... 147

Introduction …… 148

Methods …….……. 152

Data Collection …….…….. 152

Building and Validating the new Profile Hidden Markov Model …….. 152

Target Sequence Translation …….. 153

Metagenomic Screen …….. 153

Manual Examination of Randomly Selected Gene Neighbourhoods …… 157

Genomic Screen …….……. 157

Taxonomic Classification of LanB-encoding contigs …… 157

Statistical Analysis ……. 157

Results …… 158

Validation of the Profile Hidden Markov Model …….….. 158

Metagenomic Screen …….…. 161

Manual Examination of Selected Gene Neighbourhoods …….. 166

Genomic Screen …….…. 171

Taxonomic Classification of LanB-encoding contigs …… 172

Discussion ……... 177

Conclusions …… 183

References …….…. 184

Chapter 4

The relationship between subclass I lantibiotic gene density and the composition and functional potential of the gut microbiota

Abstract …….…….. 193

Introduction …… 194

Methods …….……. 195

Study population …….…… 195

Acquisition of clinical, exercise and dietary data …….…….. 195

Preparation of metagenomic libraries …… 195

Statistical and bioinformatics analysis …….…….. 196

Results …… 198

LanB density differs between athletes and high BMI controls ….……. 198

Taxonomic classification of LanB-encoding assemblies ……... 202

An inverse relationship with LanB density becomes apparent as lean mass increases …….…… 202

LanB density is associated with the functional capacity of the gut microbiota in athletes …… 207

LanB density shows distinct relationships with host diet and gut microbiota function in athletes and non-athletes ……..…… 210

A machine learning-based approach identified variables with the ability to predict LanB density …….....…. 210

Discussion ……... 213

References ……..…… 218

Chapter 5

The effect of probiotic feeding on metabolic health and the gut microbiota in a diet-induced obesity (DIO) mouse model

Abstract ……... 222

Introduction …… 223

Methods …….……. 224

Bacterial strains …….……. 224

Diet-induced obesity (DIO) mouse model ……. 224

Experimental design ……... 225

Weight determination and tissue sampling ……. 227

Measurement of metabolic markers …….….. 227

Total DNA extraction …….… 228

Amplicon Sequencing …… 229

Bioinformatic and statistical analysis ……. 230

Identification of Group-specific microbial biomarkers …….. 232

Predictive Modelling …….. 232

Results …… 233

High-fat feeding induced obesity in C57BL/6J mice: comparison of LFD and HFD control groups …… 233

Probiotic treatment tended to reduce fat mass and significantly reduced corresponding fat pads in a strain-specific manner …… 233

Strain-specific effects on metabolic health in DIO mice …… 238

Diet and strain specific probiotic supplementation influenced overall metabolic phenotype …….…….. 238

Diet and strain specific probiotic supplementation influenced the overall composition and diversity of the murine faecal microbiota ……... 239

Diet and strain specific probiotic supplementation influenced individual taxa within the murine faecal microbiota …….…….. 242

Identification of Group-specific microbial biomarkers …….. 246

Machine learning can accurately predict diet and probiotic supplementation status based on murine faecal microbiota composition ……..... 248

Diet and strain specific probiotic supplementation influenced the overall functional potential of the murine faecal microbiota ……. 248

Diet and strain specific probiotic supplementation influenced individual functional pathways within the murine faecal microbiota ……. 250

Examination of the relationship between microbiota composition and metabolic phenotype ……..……. 251

Examination of the relationship between microbiota-encoded functions and metabolic phenotype …….. 251

Examination of the relationship between the gut microbiota and host hepatic total cholesterol and hepatic triglyceride levels ……. 253

Discussion …….……. 255

References …….……. 262

General Discussion …….…….. 269

References …….. 273

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study, leading to the award of PhD is entirely my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed:

Student Number: 107577061

Date:

Abstract

Advances in computing power and metagenomic sequencing have facilitated the sourcing of a wealth of evidence to support the long-held belief that the complex community of microorganisms inhabiting the human gastrointestinal tract has significant influence on the health of the host. Equally, disruption of the composition and function of this population has been associated with many disease states. Therefore, it stands to reason that selective modulation of the gut microbiota, through antimicrobials, probiotics, or diet, presents an attractive therapeutic approach to the treatment of these diseases. One category of antimicrobials with this potential is the bacteriocins – antimicrobial peptides produced by many lineages of bacteria that kill or inhibit the growth of specific competitors. This thesis focuses primarily on the first step toward harnessing the bacteriocin-producing capacity of the human microbiota to bring about desired changes, i.e., identification of potential bacteriocin-producing bacteria using *in silico* genomic and metagenomic screening approaches. It highlights numerous putative bacteriocin-producing bacteria, including many from genera either not previously associated with bacteriocin-production or recently under consideration as the next generation of probiotics. It also investigates the relationship between diet, the gut microbiota, and bacteriocin gene cluster density in elite athletes and healthy controls, in addition to evaluating the effect of probiotic supplementation of the gut microbiota and overall health in a diet-induced obesity mouse model. Taken together these results show that there is a vast reservoir of putative bacteriocin-producing bacteria in the human microbiota with the potential to impact human health in a beneficial manner and highlights diet as a major factor in influencing the density and distribution of these putative producers.

Publications

Walsh, C.J., Guinane, C.M., O'Toole, P.W. and Cotter, P.D., 2014. Beneficial modulation of the gut microbiota. FEBS letters, 588(22), pp.4120-4130.

Walsh, C.J., Guinane, C.M., Hill, C., Ross, R.P., O'Toole, P.W. and Cotter, P.D., 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. BMC microbiology, 15(1), p.183.

Walsh, C.J., Guinane, C.M., O'Toole, P.W. and Cotter, P.D., 2017. A Profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome. PeerJ, 5, p.e3254.

List of Abbreviations

ANOVA	Analysis of variance
AOI	Area of interest
ACT	Artemis comparison tool
ABC	ATP-binding cassette
BAGEL	Bacteriocin genome mining tool
BLAST	Basic local alignment search tool
BMI	Body mass index
CDI	*Clostridium difficile* infection
CFU	Colony forming unit
CRC	Colorectal cancer
CD	Crohn's disease
DNA	Deoxyribonucleic acid
DIO	Diet-induced obesity
EMBL	European Molecular Biology Laboratory
EPS	Exopolysaccharide
FMT	Faecal microbiota transplantation
FDR	False-discovery rate
FLASH	Fast length adjustment of short reads to improve genome assemblies
GI	Gastrointestinal
GIT	Gastrointestinal tract
GO	Gene ontology
GWAS	Genome-wide association studies
GCV	Gross calorific value

HMM	Hidden Markov model
HTS	High throughput sequencing technologies
HDL	High-density lipoprotein
HFD	High-fat diet
HMP	Human Microbiome Project
HMASM	Human Microbiome Project's Illumina whole genome shotgun assemblies
HMRGD	Human Microbiome Project's reference genome data
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IDBA	Iterative de Bruijn Graph de novo Assembler
JGI	Join Genome Institute
KO	KEGG orthology
KEGG	Kyoto encyclopedia or genes and gneomes
LAB	Lactic acid bacteria
LAP	Linear azol(in)e-containing peptide
LEFSe	Linear discriminant analysis effect size
LOESS	Local polynomial regression model
LDL	Low-denisty lipoprotein
LFD	Low-fat diet
MEM	Minimum exact matches
MDS	Multidimensional scaling
NCBI	National Centre for Biotechnology Information
NEFA	Non-esterified fatty acids
NMDS	Non-metric multidimensional scaling
OTU	Operational Taxonomic Unit

PBMC	Peripheral blood monocyte cytokine
PERMANOVA	Permutation multivariate analysis of variance
PICRUSt	Phylogenetic investigation of communities by reconstruction of unobserved states
PCR	Polymerase chain reaction
PCoA	Principal components analysis
PBGC	Putative bacteriocin gene cluster
QIIME	Quantitative insights in microbial ecology
RF	Random forests
RPK	Reads per kilobase
RNA	Ribonucleic acid
SAM	Sequence alignment map
SCFAs	Short chain fatty acids
SIBO	Small intestinal bacterial overgrowth
NaOH	Sodium Hydroxide
STAT	Subtherapeutic antibiotic treatment
TC	Total cholesterol
TG	Total triglycerides
TNFα	Tumour necrosis factor alpha
T2D	Type 2 diabetes
UC	Ulcerative colitis
UDP	Uridine diphosphate
UDP-MurNAc	Uridine diphosphate-N-acetylmuramic acid
VLDL	Very low-density lipiprotein
WPI	Whey protein isolate

List of Figures

Chapter 1		**Page No.**
Figure 1	Potential strategies for manipulation of the gut microbiota.	7
Chapter 2		
Figure 1	Frequency of bacteriocin class and producing phylum among the 74 PBGCs identified.	86
Figure 2	Diagrammatic representation of remaining PBGCs identified in the Actinobacteria, Fusobacteria, and Synergistetes phyla.	88
Figure 3	Diagrammatic representation of remaining PBGCs identified in the Firmicutes phylum.	89
Figure 4	Diagrammatic representation of remaining PBGCs identified in the Proteobacteria phylum.	91
Figure 5	Diagrammatic representation of PBGCs deemed of particular interest	93
Chapter 3		
Figure 1	Flowchart depicting the step involved in building and validation of, and screening using, a profile HMM	155
Figure 2	BAGEL3 output of putative bacteriocin gene clusters identified in the positive controls used for validation of our new profile HMM. Each predicted open reading frame is colour-coded based on the role it plays in lantibiotic biosynthesis.	159
Figure 3	Barchart depicting the distribution of lanthionine dehydratase protein numbers identified by our new profile HMM in metagenomic samples from the stool and oral microbiota.	162
Figure 4	Venn diagram illustrating the numbers of lanthionine dehydratase proteins reported in stool and oral metagenomic data by single and multiple methods	163
Figure 5	Comparison of lanthionine dehydratase density by body site reported by all three methods.	164
Figure 6	Graphical representation of hits randomly selected for manual examination	169

Figure 7	BAGEL3 output of three putative bacteriocin gene clusters identified from the gastrointestinal tract subset of the Human Microbiome Project's reference genome database by our new profile HMM.	175
Figure 8	Proportion of hits identified in metagenomic stool samples by our model that were also identified by other methods.	182

Chapter 4

Figure 1	Density plots illustrating the distribution of LanB density when separated by group	201
Figure 2	Pie charts depicting taxonomy of LanB producers classified by Kaiju at Phylum and Genus levels in athletes and non-athletes.	204
Figure 3	Proportion of LanB-producing scaffolds classified as selected taxa by Kaiju in athletes and non-athletes.	205
Figure 4	Scatterplot of lean mass versus LanB density.	206
Figure 5	Illustration of diversity measures used to compare athletes to non-athletes based on gene families data.	208
Figure 6	Illustration of diversity measures used to compare athletes, low BMI controls, and high BMI controls based on gene families data.	209
Figure 7	Dotplot of variables that significantly correlated with LanB density across all samples, athlete samples only, and non-athlete samples only.	212
Figure 8	Scatterplot of the four variables identified by Boruta as possessing power to predict LanB density across all samples.	214

Chapter 5

Figure 1	Effect of *L. casei* AH0077 and *L. plantarum* AH0315 on body weight	234
Figure 2	Effect of *L. casei* AH0077 and *L. plantarum* AH0315 on fat mass.	235
Figure 3	Effect of diet, *L. casei* AH0077 and *L. plantarum* AH0315 on weights of subcutaneous fat, brown adipose tissue, epipidymal fat and retroperitoneal fat.	236

Figure 4	Effect of diet, *L. casei* AH0077 and *L. plantarum* AH0315 on hepatic total cholesterol and hepatic triglyceride levels	237
Figure 5	Boxplots showing alpha diversity (Shannon and Simpson indices) of the faeces of the low-fat diet (LFD) control, high-fat diet (HFD) control, HFD + *L. casei* AH0077 and HFD + *L. plantarum* AH0315 groups.	240
Figure 6	Two-dimensional NMDS plots of host metabolic phenotype, faecal microbiota composition, and faecal microbiota function based on Bray-Curtis dissimilarity.	241
Figure 7	Cladogram depicting group-specific microbial biomarkers as identified by LEFSe.	243
Figure 8	Predictive model based on genus-level relative abundances using Random Forests	249
Figure 9	Correlations between the composition of the gut microbiota and host physiology in the low-fat diet (LFD) control, high-fat diet (HFD) control, HFD + *L. casei* AH0077 and HFD + *L. plantarum* AH0315 groups.	252
Figure 10	Correlations between lipid metabolism functionality of the gut microbiota and host physiology in the low-fat diet (LFD) control, high-fat diet (HFD) control, HFD + *L. casei* AH0077 and HFD + *L. plantarum* AH0315 groups.	254
Figure 11	Effect of diet and probiotic supplementation on food intake and energy excretion.	258

List of Tables

Chapter 1		**Page No.**
Table 1	Some examples of studies assessing the influence of diet on the microbiota and health of the host.	20
Table 2	Some examples of studies assessing the influence of antimicrobials on the gut microbiota and, where relevant, the host.	26
Chapter 2		
Table 1	130 unique putative producers identified by BAGEL3	80
Table 2	Additional information on PBGCs and whether the initial identification of the AOI by BAGEL3 was based on the presence of bacteriocin-associated genes (context) or a specific bacteriocin structural gene.	84
Chapter 3		
Table 1	Controls used in validation of the profile HMM, listing the lantibiotic produced, whether the strain was included as a positive or negative control, the subclass of modification protein responsible for lanthionine dehydration, and whether there was a significant hit identified by each method	154
Table 2	Number of metagenomic samples per body site screened.	156
Table 3	Comparison of lanthionine dehydratase proteins identified by each method and their density based on metagenome size.	165
Table 4	Sample and contig identifiers of all randomly selected hits that underwent manual annotation. The table also states whether the hit was identified by Pfam and BlastP approaches.	168
Table 5	Manual annotation of the putative BlastP-identified biosynthetic gene cluster on scaffold 39304 from stool metagenome SRS014923	173
Table 6	Detailed information of all lanthionine dehydratase proteins identified in the gastrointestinal tract subset of the Human Microbiome Project's reference genome database using our profile HMM.	174
Table 7	Genus-level classification of putative LanB-encoding scaffolds identified by our model.	176

Chapter 4

Table 1	Number and density of LanB proteins identified in each metagenome.	199
Table 2	Number of LanB producers classified by Kaiju at Phylum and Genus levels in athletes and non-athletes.	203
Table 3	Variables that significantly correlated with LanB density across all samples, athlete samples only, and non-athlete samples only.	211

Chapter 5

Table 1	Experimental DIO mouse groups and associated diet and treatment regimens.	226
Table 2	Relative abundance (%) at bacterial phylum and family level in the faeces of the low-fat diet (LFD) control, high-fat diet (HFD) control, HFD + *L. casei* AH0077 and HFD + *L. plantarum* AH0315 groups.	244
Table 3	Relative abundance (%) at genus level in the faeces of the low-fat diet (LFD) control, high-fat diet (HFD) control, HFD + *L. casei* AH0077 and HFD + *L. plantarum* AH0315 groups.	247
Table 4	Genera whose significant enrichment (*) or reduction ($) by *L. casei* AH0077 feeding was either reversed or not significant (n.s.) in the *L. plantarum* AH0315-fed group.	259
Table 5	Microbial-encoded functions whose significant enrichment (*) or reduction ($) by *L. casei* AH0077 feeding was either reversed or not significant (n.s.) in the *L. plantarum* AH0315-fed group.	260

Chapter 1

Beneficial modulation of the gut microbiota

Updated for this thesis since publication in FEBS Letters

(doi: 10.1016/j.febslet.2014.03.035)

Abstract

The human gut microbiota has a significant effect on many aspects of human physiology including metabolism, nutrient absorption and immune function. Disruption of this population has been implicated in many conditions and diseases, including obesity, inflammatory bowel disease, colorectal cancer and even mental health. A logical extension of these observations suggests that manipulation of the gut microbiota can be employed to prevent or treat these conditions. This literature review highlights a variety of options, including the use of changes in diet (including the use of prebiotics), antimicrobial-based intervention (and discovery thereof), probiotics, and faecal microbiota transplantation, and discuss their relative merits with respect to modulating the intestinal community in a beneficial way.

Introduction

Humans are now thought of as “superorganisms” or “holobionts” on the basis of the genetic potential encoded within our resident microbial populations in addition to our own genome. It is widely accepted that our microbiota develops with us and alters its own composition and gene expression in response to changing environmental conditions (Ley et al., 2008), but it has also been suggested that our microbiota changed in parallel though our evolutionary history in a relationship termed “phylosymbiosis” (Brooks et al., 2016). While almost all niches in the human body contain their own bespoke microbial communities, the largest and most varied of these exists in the gastrointestinal (GI) tract.

It has been estimated that the human microbiota comprises approximately 100 trillion bacterial cells, outnumbering our own cells by a factor of 10 or more (Bäckhed et al., 2005) and plays an integral role in human health and disease (Clemente et al., 2012, Flint et al., 2012). A recent publication, however, has argued that the ratio is actually more likely to be one-to-one, with the numbers being similar enough that each defecation event may alter the ratio to favour human cells over bacteria (Sender et al., 2016). Regardless of absolute numbers, this community is thought to contain 100–1,000 phylotypes (Faith et al., 2013, Qin et al., 2010), the most abundant being members of the Firmicutes and Bacteroidetes phyla, with smaller numbers representing the Proteobacteria, Fusobacteria, Cyanobacteria, Verrucomicrobia and Actinobacteria, amongst others (Qin et al., 2010), and exhibit robust temporal stability (Belstrøm et al., 2016, Jeffery et al., 2016). Of greater consequence than bacterial numbers, however, is the collection of genes encoded in this metagenome, thought to be approximately 150 times larger than that of the human genome, with a functional

potential far broader than that of its host (Qin et al., 2010). It has been noted that, although there is great inter-individual variation in the composition of the gut microbiota, this functional potential is highly conserved between individuals, suggesting that it is the functionality of the microbiota, rather than its composition, that is of greatest importance to the host. The functions and pathways encoded in this "core microbiome" are thought to confer the greatest benefit to the host and are probably essential for the correct functioning of the gut. Some well-studied benefits include protection against potential pathogens, digestion of polysaccharides, production of essential vitamins, stimulation of angiogenesis, regulation of fat storage and modulation of the host's immune system (Sekirov et al., 2010). Recent studies have also shown that the gut microbiota influences the gut-brain axis and shapes stress-related symptoms such as anxiety and pain tolerance (Cryan and O'Mahony, 2011). In addition, bacteriophage are present in the gut microbiota and may be involved in controlling its composition and functionality, although these studies are relatively recent and the community is not as well characterised as its bacterial counterpart (Manrique et al., 2016).

Advances in high throughput sequencing technologies (HTS) and tools enabling comparative analysis of the large amount of data that are generated by these technologies have led to a better understanding of what constitutes a 'healthy" gut microbiota. One of the most interesting observations drawn from the data generated is that the resident microbiota encodes > 100 fold more genes than the human genome (The Human Microbiome Project Consortium, 2012). The genes present in the microbiome are responsible for many functions essential to host survival but which are not encoded within the human genome. Due to the range and importance of the

metabolic and biochemical processes carried out by the microbiome it has been referred to as “our hidden organ” (O'Hara and Shanahan, 2006).

While the “healthy” gut microbiota is seen to be a stable community (Jeffery et al., 2016), there are stages within the life cycle of humans during which there can be dramatic alterations in the structure and function of this population. These “natural” changes begin with initial colonisation immediately following birth and subsequent development of the microbiota over the first two years of life. The earliest colonizers are usually members of the enterococci and enterobacteria followed by strict anaerobes such as *Bifidobacterium*, *Clostridium* and *Bacteroides* spp. once the initial oxygen supply present has been depleted (Adlerberth and Wold, 2009). Despite this general pattern, it is important to appreciate that the method of delivery and subsequent feeding type have a profound effect on the initial populations (Dominguez-Bello et al., 2010). Once the infant reaches two years of age the microbiota has already begun to transform into its adult form, which is thought to be relatively stable before it undergoes a final shift when entering old age (Palmer et al., 2007). Indeed, with respect to the latter phenomenon, a study by Claesson and colleagues that compared the gut microbiota of individuals ages 65 or older to 9 younger control subjects has highlighted significant changes in community structure associated with ageing, specifically an increase in the abundance of *Bacteroides* spp. and distinct shifts within the *Clostridium* genus (Claesson et al., 2011). It has been hypothesised that alterations in the elderly microbiota are due to physiological changes in the elderly gastrointestinal tract such as chronic low-grade inflammation, in addition to dietary habits and antibiotic use (Franceschi, 2007, Jeffery et al., 2016).

An interesting recent development is that exercise also has a significant influence on the composition and function of the gut microbiota. It was noted that physical activity, as measured through plasma creatine kinase levels, impacted the use of dietary nutrients by the gut microbiota beyond the changes that would be expected by a simple change in diet (Barton et al., 2017, Clarke et al., 2014).

It has been well established that the human gut microbiota is integral to human health, and, as will be discussed below, it also plays an important role in gastrointestinal disease. It is therefore reasonable to assume that modulation of the gut microbiota can be used as a therapeutic approach to treating chronic gastrointestinal diseases. Thus, this review is focussed primarily on the methods that can be employed to modulate the gut microbiota, some of which are illustrated in Figure 1, while highlighting the benefit of guiding community structure towards a more desirable state.

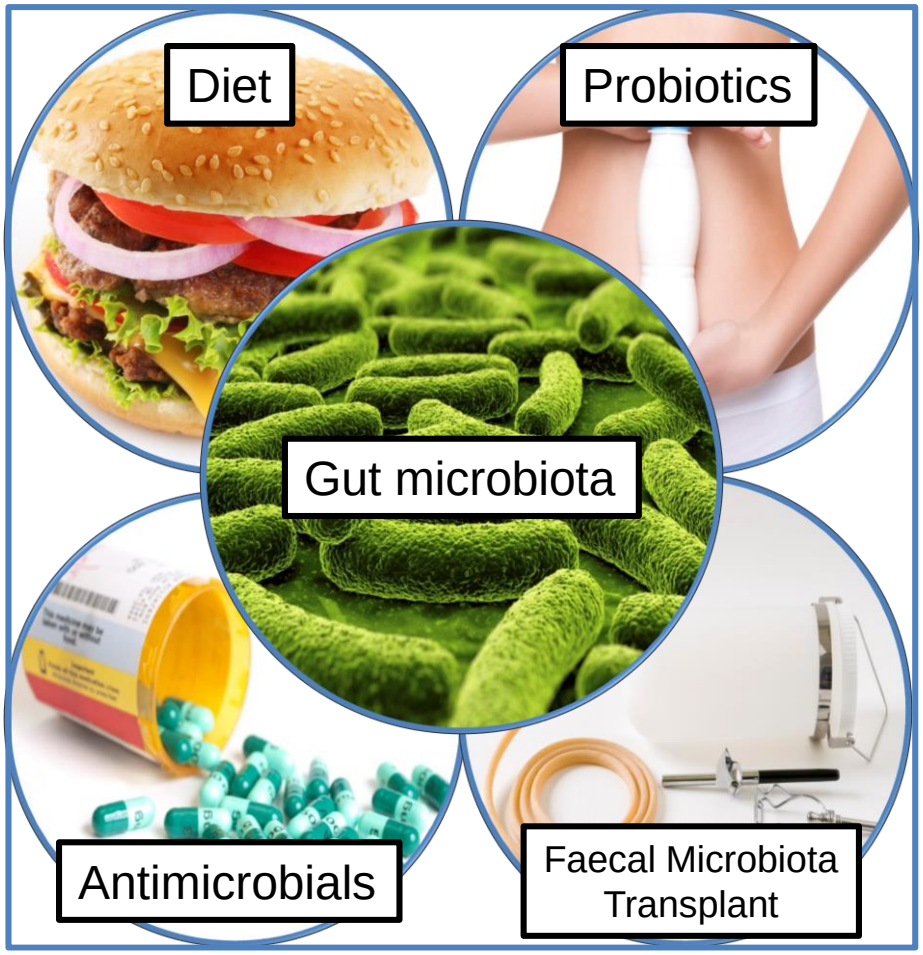


Figure 1. Potential strategies for manipulation of the gut microbiota.

Role of the Gut Microbiota in Health and Disease

There are a growing number of gastrointestinal conditions that have been linked with alterations in the gut microbiota. To properly implement strategies to modulate the gut microbiota as a therapeutic tool, it is first necessary to understand the role of the gut microbiome in specific GI, and other, diseases. Given the recent rapid expansion in the number of disease states that have been linked with alterations in the gut microbiota, it is not possible to address the issue in depth, so a number of relevant and well-studied examples were selected to highlight these links.

Inflammatory Bowel Disease

Inflammatory Bowel Disease (IBD) is a relapsing disorder characterized by chronic inflammation of the GI tract, and of the colon in particular. The two major types of IBD are Crohn's disease (CD) and ulcerative colitis (UC). Evidence suggests that IBD is a complex disease arising from a combination of genetic and environmental factors. From a genetic perspective, genome-wide association studies (GWAS) and subsequent meta-analyses have identified a total of 240 genetic risk loci for IBD (Anderson et al., 2011, Franke et al., 2010, Jostins et al., 2012, de Lange et al., 2017). A German twin cohort study confirmed the strong genetic element to IBD by observing that monozygotic twins are significantly more likely to be concordant for the disease than dizygotic twins (Spehlmann et al., 2008). However, concordance rates between monozygotic twins are nonetheless low (35% for CD and 16% for UC), highlighting that environmental triggers do indeed play an important role in both diseases, and in UC in particular. A similar study of Norwegian siblings reported a

higher risk of CD concordance in dizygotic twins than in ordinary siblings, suggesting the importance of a shared environment *in utero* or in childhood (Bengtson et al., 2010).

It is notable that murine studies have revealed that the presence of commensal enteric bacteria is necessary for the development of spontaneous colitis and immune system activation (Sellon et al., 1998) and, indeed, transferring colitogenic gut microbiota into healthy mice can induce spontaneous colitis (Garrett et al., 2010). Similarly, it has consistently been observed that patients suffering from IBD harbour an altered gut microbiota (Frank et al., 2007, Sokol and Seksik, 2010), specifically reduced bacterial diversity and changes within the Firmicutes phylum (Elson and Cong, 2012). The changes in microbiota composition appear to be somewhat different between UC and CD. For example, decreased abundance of the butyrate-producing bacteria *Roseburia hominis* and *Faecalibacterium prausnitzii* have been observed in UC patients relative to controls (Machiels et al., 2013), while the opposite has been observed in CD patients who possessed increased *F. prausnitzii* levels in addition to a reduced overall diversity (Hansen et al., 2012). Although these microbial changes could be a result of increased inflammation, evidence suggests that it is more likely that shifts in the microbiota are involved in the disease's pathogenesis, either due to an intolerance to a specific group of commensals or due to an imbalance between protective and harmful members of the population (Frank et al., 2007, Lepage et al., 2011, Elson and Cong, 2012).

Irritable Bowel Syndrome

Irritable Bowel Syndrome (IBS) is a chronic GI disorder that presents with symptoms including abdominal pain, bloating and altered bowel function. IBS is divided into several subtypes based on stool characteristics; diarrhoea, constipated or mixed. Its cause, as of yet, is not fully known and although the aetiology is thought to be a combination of a number of factors, it is hypothesised that perturbations in the normal microbial microbiota play a role in the syndrome's characteristic low-grade inflammation (Ohman and Simren, 2013). Indeed, Rajiić-Stojanović et al. used qPCR and phylogenetic microarrays to show that the gut microbiota of IBS patients differed significantly from healthy controls, with IBS sufferers having a 2-fold higher Firmicutes to Bacteroidetes ratio and correlation analysis implicating several groups of Firmicutes and Proteobacteria in IBS pathogenesis (Rajilic-Stojanovic et al., 2011). Contrastingly, Jalanka-Tuovinen and colleagues observed that the faeces of diarrhoea-predominant IBS sufferers harboured 12-fold higher levels of several Bacteroidetes members. This group also noted that healthy controls have 35-fold higher numbers of uncultured clostridia (Jalanka-Tuovinen et al., 2013). Interestingly, these alterations in the microbiota correlated with changes in expression of host genes involved in amino acid synthesis, cell junction integrity and inflammatory response, suggesting impaired epithelial barrier function in IBS patients. An interesting note is that patients suffering from IBS exhibit greater temporal instability with respect to their microbiota than their healthy counterparts, suggesting that the varying symptoms of this disease may result from fluctuations (Mättö et al., 2005). Small intestinal bacterial overgrowth (SIBO), which is characterized by excessive bacteria in the small intestine, has also been put forward as a possible factor in IBS aetiology (Lin, 2004). Bacterial

overgrowth can result in overproduction of gas in the small intestine by degradation of carbohydrates, contributing to the symptoms of IBS (Riordan and Kim, 2006). The most commonly isolated bacteria from SIBO patients are *Escherichia coli*, *Streptococcus*, *Lactobacillus*, *Bacteroides* and *Enterococcus* species (Bouhnik et al., 1999). However it is not fully understood if any of these microorganisms play a specific role in IBS progression. It should also be recognised that differences between studies may be due to the causative microorganisms or imbalances differing between IBS subtypes. Regardless, a bacterial role in IBS onset would seem to be clear, as further evidenced by the disease's response to antibiotic therapy (Pimentel et al., 2011) and differential expression levels of Toll-like receptors in colonic biopsies of patients with IBS (Brint et al., 2011).

Obesity

Obesity is a complex disease resulting from a prolonged imbalance of energy input and energy expenditure. Modern dietary and exercise habits are major contributing factors but it is now understood that the composition and function of the gut microbiome plays an important role through a variety of mechanisms (Backhed et al., 2004). A number of comprehensive reviews focussing on the association between the microbiota and obesity have been published (Cani, 2013, Clarke et al., 2012, Khan et al., 2016). Differences in the gut microbiota between obese and lean individuals have been the subject of great scrutiny. A range of different murine models have been used to this end, including genetically obese (Ley et al., 2005, Turnbaugh et al., 2006), diet-induced obese (Turnbaugh et al., 2008) and humanized (Turnbaugh et al., 2009) mice. Although a number of studies have reported an increased ratio of Firmicutes to

Bacteroidetes in obese mice compared to their lean counterparts, these findings continue to be the subject of much debate in relation to human studies, which have revealed a number of microbial populations that have been associated with obesity (Clarke et al., 2012). Recently, it has become apparent that diversity of the microbiota is of greater relevance to obesity than specific compositional profiles with a less diverse gut microbiota observed in obesity (Turnbaugh et al., 2009), as well as diseases such as type 1 diabetes (Patterson et al., 2015) and rheumatoid arthritis (Chen et al., 2016). Notably, transplanting the faecal microbiota of obese humans into germ-free mice brought about significant increases in the fat-mass of, and obesity-related metabolic phenotypes in, these mice relative to those which occurred when the corresponding faecal microbiota from lean monozygotic twins was transplanted (Ridaura et al., 2013). Furthermore, a second trial showed that cohousing mice harbouring these two microbial communities prevented development of the obese phenotype, a trend correlating with invasion of specific Bacteroidetes members from lean to obese microbiota (Ridaura et al., 2013). A bacterium gaining a lot of attention in the areas of obesity and type 2 diabetes is the mucin-degrader *Akkermansia muciniphila*. *A. muciniphila* abundance was decreased in obese and type 2 diabetic mice and normalised by prebiotic feeding, which in turn correlated with an improved metabolic profile. Orally administered *A. muciniphila* also reversed high-fat diet induced metabolic disorders in these mice (Everard et al., 2013). Recently, it was reported that pasteurized *A. muciniphila*, that was initially grown in a synthetic medium safe for human consumption, possessed the ability to reduce fat mass development, insulin resistance, and dyslipidemia in mice (Plovier et al., 2017). The results of these, and other studies, make it apparent that the microbiota plays a role in

obesity but the specific changes associated with the phenotype are complex and remain unclear.

Type 2 Diabetes

Type 2 diabetes (T2D) is a metabolic disorder with both genetic and environmental influences. It is a major health concern throughout the western world, arising particularly as a result of increasing obesity-related insulin resistance (Wellcome Trust Case Control Consortium, 2007, Scott et al., 2007). It is evident from a number of studies that the gut microbiome is altered in patients suffering from T2D (Larsen et al., 2010, Musso et al., 2011, Qin et al., 2012), although, as with many obesity-related associations, it is not clear whether these changes are a cause or simply a consequence of the disorder. Nonetheless, it was an interesting development when, in 2010 it was reported that the proportions of Firmicutes, and in particular species of clostridia, were significantly reduced in T2D sufferers compared to healthy individuals (Larsen et al., 2010). A subsequent, and much larger, metagenome-wide association study of 345 Chinese individuals showed that the gut microbiota of patients with T2D was characterized by a moderate degree of microbial dysbiosis, lower levels of butyrate-producing bacteria and an enrichment of microbial functions relating to sulphate reduction and resistance to oxidative stress (Qin et al., 2012). Almost all of the microbial genes enriched in T2D patients were from opportunistic pathogens, including genes from several *Clostridium* spp. as well as *Bacteroides caccae* (Qin et al., 2012). These results provided a number of markers that were assessed to determine if they could successfully identify patients with T2D on the basis of an analysis of faecal samples. Notably, this method successfully identified the T2D

disease state with 81% accuracy (Qin et al., 2012), i.e. a greater success rate than using a combination of clinical risk factors and genetic information (Lyssenko et al., 2008). Another notable study using microbiome data, in addition to clinical measurements, allowed a machine-learning algorithm to design personalized diets which significantly lowered postprandial blood glucose levels in type 2 diabetics (Zeevi et al., 2015).

Colorectal Cancer

Colorectal cancer (CRC) is the third most common cause of cancer mortality in the world (Jemal et al., 2011). It is becoming apparent that, even though a single causative microorganism has not been explicitly identified, the gut microbiota plays a role in CRC (Arthur et al., 2012, Plottel and Blaser, 2011). Wang and colleagues noted that there was a clear segregation between the microbiota of CRC patients and healthy volunteers, particularly, as was the case for T2D, a decrease in the abundance of butyrate producers and an increase in the incidence of opportunistic pathogens in CRC patients (Wang et al., 2012b). Increased intake of dietary fibre appears to play a key role in decreasing CRC risk *via* the 'fibre-microbiota-butyrate axis' whereby it is fermented by colonic bacteria into butyrate, a potent histone deacetylate inhibitor, that upregulates tumour-supressing genes in CRC cells and anti-inflammatory genes in immune cells (Bultman, 2017, Bultman and Jobin, 2014). Dietary fibre also speeds colonic transit, thereby limiting the exposure of colonic epithelial cells to potential carcinogens. Members of the *Fusobacterium* genus have also been recently identified as potential causative agents after it was observed that they were enriched in colorectal carcinomas (Kostic et al., 2012), a pattern also noted in other studies (Wang et al., 2012b, Castellarin et al., 2012, McCoy et al., 2013, Tahara et al., 2014). The authors

hypothesised that *Fusobacterium* spp. may contribute to tumourigenesis by an inflammatory-mediated mechanism, a hypothesis supported by a follow-up study which showed that members of fusobacteria could generate a proinflammatory microenvironment through the recruitment of tumour-infiltrating immune cells (Kostic et al., 2013). *E. coli* has also been linked with CRC in a number of studies. Arthur *et al.* observed that *E. coli* levels were ~100-fold higher in the microbiota of the colitis-susceptible *IL10* $^{-/-}$ mouse strain compared to the wild type (Arthur et al., 2012). They went on to show that *E. coli* NC101 mono-association significantly promoted development of invasive mucinous adenocarcinomas in azoxymethane treated, *IL10* $^{-/-}$ mice and that deletion of the polyketide synthase (*pks*) genotoxic island from this *E. coli* strain decreased tumour multiplicity and invasion (Arthur et al., 2012). While further investigations are required, these results suggest that colitis promotes tumourigenesis in mice by altering the composition of the gut microbiota and selecting for members with genotoxic capabilities.

Ultimately, identification of microorganisms, microbial populations or microbial functionalities involved in GI disease is fundamental to developing novel therapies. It is evident that the gut microbiota plays a large role in intestinal health and disease, and therefore manipulation or modulation of this community, is a clinical option that merits serious consideration.

Modulation of the Gut Microbiota

Modulation by Diet

Environmental factors, including dietary intake, can shape the composition of the intestinal microbial community. Indeed, a number of recent studies have highlighted the links between diet and distinct microbial profiles and, in turn, overall gut health (Bäckhed et al., 2005, Claesson et al., 2012, David et al., 2013, Turnbaugh et al., 2008, Wu et al., 2011, Duncan et al., 2007). Having an understanding of how diet influences microbial communities will be of critical importance with respect to employing food to beneficially alter the gut microbiota.

The amount, type and balance of the three main dietary components, i.e. protein, carbohydrates and fat, have a profound impact on the gut microbiota. Short-chain fatty acids (SCFAs), primarily butyrate, propionate and acetate, are the major end products from the microbial degradation of carbohydrates and protein in the gut. SCFAs have a diverse range of physiological effects on the host, with perhaps the most important being their oxidation by mucosal cells to provide energy. An excellent review of the benefits of SCFAs on the host has been published by Macfarlane & Macfarlane (Macfarlane and Macfarlane, 2012) and a study involving native Africans and African Americans illustrated that a relatively short two-week dietary intervention had a profound impact on the gut microbiota and metabolites including butyrate and mucosal biomarkers of cancer risk (O'Keefe et al., 2015). The majority of microbial protein degradation occurs in the distal colon where the pH is neutral and conditions are favourable for the growth of proteolytic bacteria such as *Bacteroides* spp., *Propionibacterium* spp. and *Clostridium perfringens* (Walker et al., 2005, Macfarlane

et al., 1986). The main pathway of protein degradation by this population is deamination of amino acids to the aforementioned SCFAs and ammonia (Cummings and Macfarlane, 1991), high concentrations of the latter have been shown to act as tumour promoters in rats (Clinton et al., 1988). The range of end products generated by protein digestion is broader than that of carbohydrates (see below) and also includes branched-chain amino acids, phenols, indoles and amines (Hamer et al., 2012). The majority of studies examining the effect of dietary protein on the gut microbiota have focussed primarily on the detection of altered fermentation products in the cecum (Lhoste et al., 1998) and faeces (Magee et al., 2000). However, the effects of whey protein isolate on the microbiota have been the topic of some scrutiny in recent years as it has been indicated that dairy products can alleviate several disorders relating to metabolic syndrome (Elwood et al., 2008). One such study noted significantly increased counts of bifidobacteria and lactobacilli in the faeces of rats whose diets included cheese whey protein isolate or casein supplemented with either threonine or cysteine (Sprong et al., 2010). Whey protein isolate (WPI) has also been observed to alter the composition of the gut microbiota of mice in a dose-dependent manner (McAllan et al., 2014). All mice whose high fat diet was supplemented with WPI had significantly increased proportions of *Lactobacillaceae* and significantly decreased proportions of *Clostridiaceae* compared to high-fat fed controls, and increasing the amount of total energy derived from WPI caused a more profound shift in the microbiota (McAllan et al., 2014). Certain components of the normal human dietary intake of carbohydrates cannot by digested directly by the host and act as the major diet-derived energy source for microorganisms in the gut (Cummings and Englyst, 1991). This fraction, comprised largely of resistant starches and non-starch polysaccharides, is degraded by microbial fermentation to a mixture of gasses and the

aforementioned SCFAs. Many such carbohydrates are also referred to as prebiotics. The term prebiotic was introduced by Gibson and Roberfroid in 1995 (Gibson and Roberfroid, 1995) and are defined as "selectively fermented ingredients that allow specific changes, both in the composition and/or activity in the gastrointestinal microflora that confer benefits upon host well-being and health" (Gibson et al., 2004). Prebiotics have most frequently been employed with a view to stimulating the growth of either lactobacilli or bifidobacteria, with many studies focussing on inulin (Costabile et al., 2010, Koleva et al., 2012, Ramnani et al., 2010), oligofructose (Lewis et al., 2005, Waligora-Dupriet et al., 2007) or fructooligosaccharides (Bouhnik et al., 2004, Respondek et al., 2008). There is a substantial body of evidence linking prebiotic consumption to human health benefits through modulation of the gut microbiota, with research in this area having been the subject of a number of recent reviews (Slavin, 2013, Roberfroid et al., 2010, Vieira et al., 2013). In one particularly notable recent study, it was observed that supplementing the murine diet with SCFAs or fructooligosaccharides caused a shift in microbiota composition which strongly correlated with beneficial changes in body weight, adiposity and glucose control. These physiological changes were brought about via butyrate- and propionate-mediated activation of intestinal gluconeogenesis (De Vadder et al., 2014).

The majority of dietary fat is absorbed in the human small intestine but it has been shown that a substantial amount survives digestion and can be recovered in faeces (Gabert et al., 2011). The undigested portion passes through the colon where it can have a profound effect on the intestinal microbiota. Murphy *et al.* observed that high-fat feeding caused a greater compositional change in the gut microbiota than genetically induced obesity (Murphy et al., 2010), in accordance with a previous study which showed that, when fed a high-fat diet, RELMβ knockout mice showed a

significantly altered gut community while staying lean. RELMβ knockout mice were employed as they are known to stay relatively lean when fed a high-fat diet. The authors could therefore conclude that the change in diet, as opposed to the obese state, was responsible for the observed changes in the microbiota (Hildebrandt et al., 2009). Many studies have established that mice fed a high-fat diet have significantly dissimilar microbial populations in the gut compared to mice fed on normal chow (Ley et al., 2005, Turnbaugh et al., 2008, Daniel et al., 2013).

However, life-long calorie restriction significantly alters the gut microbiota in mice fed on both high-fat and low-fat diets (Zhang et al., 2013). This implies that not only the fat content of the diet, but also the number of calories consumed, has the potential to influence the bacterial communities present in the GI tract. The study also linked changes in the gut microbiota to claims that calorie restriction promotes healthy-ageing and increases lifespan in various animal models as the healthiest and longest living mice were those that were fed a low fat diet with calorie restriction (Zhang et al., 2013). Interestingly, another investigation involving calorie restriction, this time in humans, suggested that increased levels of *A. muciniphila* were associated with a healthier metabolic status pre-restriction and better clinical outcomes post-restriction (Dao et al., 2016), implying that the gut microbiota may play a role in the success of dietary interventions.

This specific combination of dietary components can vary according to geographic location, food availability, cultural practices and age and can have a profound impact on the conditions within the gut and the requirements of the microbiota (Table 1 highlights some studies which have investigated this impact). In one instance, the faecal microbiota of European children and children from an African

Diet	Effect on microbiota	Effect on host
Rich in plant-derived polysaccharides (De Filippo et al., 2010, Wu et al., 2011).	Increased Bacteroidetes, decreased Firmicutes (De Filippo et al., 2010). Associated with *Prevotella*-rich enterotype (Wu et al., 2011).	Faster gut transit time compared to high protein and animal fat diet (Wu et al., 2011).
Omnivorous compared to vegetarian and lacto-vegetarian (Kabeerdoss et al., 2012, Liszt et al., 2009, Matijasic et al., 2013).	Increased *Clostridium* clusters IV and XIVa (Kabeerdoss et al., 2012, Liszt et al., 2009, Matijasic et al., 2013).	Not reported
High-fat, simple carbohydrate "Western" diet (Ley et al., 2005, Turnbaugh et al., 2008).	Increased Firmicutes, decreased Bacteroidetes (Ley et al., 2005, Turnbaugh et al., 2008).	Diet-induced obesity. Subsequent transplantation of obese microbiota to germ free mice increased adiposity (Turnbaugh et al., 2008).
Reduced carbohydrate intake (Duncan et al., 2007).	Reduced *Bifidobacterium*, *Roseburia* spp. and *Eubacterium rectale* (Duncan et al., 2007).	Not reported
Animal product-based (David et al., 2013). High protein and animal fat (Wu et al., 2011).	Increased β-diversity and bile-tolerant bacteria, including *Bacteroides*, decreased Firmicutes (David et al., 2013). Associated with *Bacteroides*-rich enterotype (Wu et al., 2011).	Decreased weight independent of calories consumed (David et al., 2013).
Less fruit, vegetables and fish (Cotillard et al., 2013).	Reduced microbial gene richness (Cotillard et al., 2013).	Increased insulin resistance, fasting serum triglyceride levels, LDL cholesterol and inflammation (Cotillard et al., 2013).
Reduced variety due to long-stay care (Claesson et al., 2012).	Increased Bacteroidetes and reduced overall diversity (Claesson et al., 2012).	Increased frailty and poorer general health (Claesson et al., 2012).
Changed from a vegetarian diet to an animal-based diet (David et al., 2013).	Decreased *Prevotella*, increased *Bacteroides* (David et al., 2013).	Not reported

Table 1. Some examples of studies assessing the influence of diet on the microbiota and health of the host.

village in Burkina Faso, whose diets differed considerably, was investigated. The diet of the African children was predominately vegetarian; high in starch, fibre and plant polysaccharides and low in fat and animal protein. This diet correlated with a significant increase in the Bacteroidetes:Firmicutes ratio in addition to an abundance of *Prevotella* and *Xylanibacter* when compared to the microbiota of the children consuming a carbohydrate-rich European diet (De Filippo et al., 2010). The *Xylanibacter* genus, which was absent in European children, is known to contain genes for xylan and cellulose hydrolysis and so it was hypothesised that the gut microbiota coevolved with the polysaccharide-rich diet of the Burkina Faso children, allowing them to increase the energy extracted from dietary fibre while also conferring protection from inflammation and non-infectious colonic disease (De Filippo et al., 2010). The comparatively high abundance of *Prevotella* in the faecal microbiota of the African children and the fact that it coincides with a carbohydrate-rich diet is consistent with the observations of Wu *et al.* who found that the overall composition of the microbiota was strongly associated with long-term diet (Wu et al., 2011). Specifically, a diet rich in protein and animal fat was associated with higher proportions of *Bacteroides* while *Prevotella* were more abundant when the diet was enriched with plant-derived carbohydrates (Wu et al., 2011). A recent study by De Filippo *et al.* took these investigations a step further by focussing specifically on the effect of diets composed entirely of animal or plants products on the gut microbiota (David et al., 2013). It revealed that an animal-based diet increased the numbers of bile-tolerant microorganisms present and decreased the numbers of plant polysaccharide degrading Firmicutes. Interestingly, the respective diets brought about a transcriptional response among the gut microbiota that was consistent with previously reported differences in gene abundances between herbivorous and

carnivorous animals (David et al., 2013). In other studies, members of the *Clostridium* clusters IV and XIVa have been found to be enriched in the faeces of omnivores compared to vegetarians and lacto-vegetarians, who generally consume higher proportions of carbohydrates as part of their diet (Liszt et al., 2009, Kabeerdoss et al., 2012, Matijasic et al., 2013). These clusters of bacteria are noted for their ability to convert dietary fibre to SCFAs.

The overall dietary patterns in the De Filippo study above are similar to a study in mice where conventionalised mice were switched from a low-fat diet rich in complex plant polysaccharides (CHO) to an obesity-inducing high-fat/simple carbohydrate "Western" diet (Turnbaugh et al., 2008). Mice fed on the "Western" diet had a significantly lower level of bacterial diversity, a characteristic seen to be an indicator of an unhealthy microbiota (Bäckhed et al., 2005). These mice possessed a significantly higher relative proportion of Firmicutes and lower relative proportions of Bacteroidetes compared to littermates which remained on the CHO diet. This population shift is similar to what is seen in the *ob/ob* mouse model of obesity (Ley et al., 2005) but differs in that the Firmicutes shift in the genetically-induced obesity model is division-wide whereas the dietary intervention above caused a bloom in a single uncultured clade within the Mollicutes class. A subsequent microbiota transplantation from these diet-induced obese mice into germ-free recipients promoted greater adiposity than transplants from lean donor (Ley et al., 2005). A further study by the same group showed that this response of the microbiome to dietary intervention is rapid and can occur within 24 hours (Turnbaugh et al., 2009), a phenomenon also observed by Wu *et al*., (Wu et al., 2011).

A gut microbiota with decreased diversity has been linked with increased frailty and poorer general health in elderly subjects (Claesson et al., 2012). In this study, clustering of subjects by diet, residence location and by microbial groupings was apparent. Ultimately, it was evident that subjects that were living in the community had a healthier and more varied diet than subjects in long-term residential care, which gave rise to a more diverse gut microbiota with significant changes being noted at phylum and family levels. Differences were also apparent at the genus level with long-stay subjects possessing higher levels of *Parabacteroides, Eubacterium, Anaerotructus, Lactonifactor* and *Coprobacillus,* while *Coprococcus* and *Roseburia* (both members of the *Lachnospiraceae* family) were more abundant in community-dwelling subjects (Claesson et al., 2012). The data also linked microbiota composition to the duration spent in long-stay care. The longer the subject stayed in residential care (and consumed a less varied diet), the more dissimilar their microbiota became to the microbiota of healthy community-dwelling subjects (Claesson et al., 2012). Another study investigating the temporal relationship between food intake, gut microbiota and metabolic and inflammatory phenotypes reported that individuals with reduced microbial gene richness present more pronounced dys-metabolism and low-grade inflammation than their richer counterparts (Cotillard et al., 2013). This microbiota-associated phenotype was suggested to be a result of long-term dietary habits as it was noted that these subjects seemed to consume less fruits, vegetables and fish than their high gene richness equivalents, i.e. a pattern consistent with that reported by Claesson *et al* (Claesson et al., 2012). More specifically, the initial sampling of the cohort (49 obese or overweight subjects) showed that subjects with lower gene richness in the gut microbiota presented with increased obesity-associated phenotypes such as higher insulin resistance and increased levels of fasting serum triglyceride, LDL cholesterol

and inflammation. Dietary intervention (6 week energy-restricted high-protein diet) increased gene richness significantly in individuals that originally had a low gene count. This increased gene richness remained after the subjects were switched to a 6 week weight-maintenance diet suggesting that dietary intervention as the potential to, at least partially, correct a loss of richness in the microbiota (Cotillard et al., 2013).

Given the complexity of the relationship between diet and the gut microbiota, there would seem to be merit in developing and utilising models that allow one to elucidate the specific relationship between specific dietary components and microorganisms. An elegant strategy to facilitate this was provided by Faith *et al.* when they introduced a model community of ten human gut bacteria into gnotobiotic mice and developed a relatively simple statistical model which predicted over 60% of the species variations that occurred in response to changes in diet (Faith et al., 2011). The amount of casein in the diet was observed to be significantly associated with the abundances of all 10 microbial species and highly correlated with the total biomass of the community. Interestingly, *E. coli* and *Clostridium symbosium* were the only two species that had a second dietary variable significantly associated with their abundance, sucrose and starch respectively. The statistical model was subsequently able to determine 61% of the variation of the community members when the host was fed a new, previously unseen diet (Faith et al., 2011). These results represent a significant step towards tailoring diet to address chronic microbiota-associated illnesses and a potential evolution of research within the field.

It is clear that microbial composition varies between groups living on different long-term diets. Recent investigations that suggest that short-term dietary changes can also alter the composition, and result in changes to the metabolic activity of the

microbiome as a whole, are noteworthy but further investigations are required to determine how best to take advantage of these observations.

Modulation by Antimicrobials

The manipulation of the gut microbiota by antimicrobials is emerging as an attractive therapeutic strategy (Table 2). The success of this approach is likely to ultimately depend on the target specificity of the antimicrobials in question, especially as the undesirable consequences of the overuse of broad-spectrum antimicrobials have become ever more apparent in recent years. For quite some time broad-spectrum antibiotics have been commonly used by clinicians as they can be used in the treatment of a wide range of infections or when the causative bacterium has not been formally identified. However, due to the frequent use of these antibiotics, the spread of antibiotic resistance is now posing a serious problem in health care settings. In addition, antibiotic therapies not only affect the target microorganism but can also perturb the host gut microbial communities. The extent of this damage has recently become more evident through the application of high throughput DNA-based sequencing technologies to assess the composition of gut microbial populations (for review see Cotter et al. 2012) (Cotter et al., 2012). Here we provide just a few examples of the negative consequences of the use of broad-spectrum antibiotics on the gut microbiota and, in turn, health.

The widespread use of broad-spectrum antibiotics, such as amoxicillin, to treat childhood infections has been linked to a dramatic decrease in *Helicobacter pylori* carriage (Blaser, 2011). However, studies indicate that those who did not acquire *H.*

Antimicrobial	Effect on Microbiota	Physiological effect on host
Thuricin CD	Eliminated *C. difficile* without impacting overall microbiota composition (Rea et al., 2011).	Not examined – distal colon model
Abp118	Protection against *Listeria monocytogenes* infection (Corr et al., 2007). Increased Bacteroidetes and Proteobacteria, decreased Actinobacteria (Walsh et al., 2008).	Temporarily reduced weight gain in pigs (Corr et al., 2007).
Vancomycin	Decreased Firmicutes and Bacteroidetes, increased Proteobacteria (Murphy et al., 2013).	Decrease in weight gain, fasting blood glucose, plasma TNFα and triglyceride levels in DIO mice (Murphy et al., 2013).
Sub-therapeutic antibiotic therapy*	Increased Firmicutes, especially *Lachnospiraceae*, relative to Bacteroidetes (Cho et al., 2012).	Increased adiposity and bone mineral density in mice (Cho et al., 2012).
5 strain probiotic mixture**	Reduced shedding of *Samonella enterica* serovar Typhimurium in pigs(Casey et al., 2007).	Reduced incidence, severity and duration of diarrhoea in pigs. Also increased weight gain (Casey et al., 2007).
Lactobacillus gasseri SBT2055, producer of gassericin T bacteriocin	Not reported	Decreased abdominal adiposity, body weight, BMI, waist circumference and hip circumference in human adults (Kadooka et al., 2010). Lower triglyceride levels and reduced expression of lipogenic and pro-inflammatory genes in DIO mice (Miyoshi et al., 2013).

* Penicillin, vancomycin, penicillin plus vancomycin, and chlortetracycline
** *Lactobacillus murinus* DPC6002, *Lactobacillus murinus* DPC6003, *Lactobacillus pentosus* DPC6004, *Lactobacillus salivarius* DPC6005, and *Pediococcus pentosaceus* DPC6006

Table 2. Some examples of studies assessing the influence of antimicrobials on the gut microbiota and, where relevant, the host.

pylori in childhood were more likely to subsequently develop asthma, hay fever and skin allergies (Chen and Blaser, 2007), while other investigations suggest that *H. pylori* infection has a protective effect with respect to the development of allergic asthma in mouse models (Arnold et al., 2011). The use of some broad-spectrum antibiotics, including clindamycin, ampicillin, amoxicillin, cephalosporins and flouroquinolones, can also result in *Clostridium difficile* overgrowth by impacting the resident gut microbiota, followed by antibiotic-associated diarrhoea, pseudomembranous colitis and, potentially, life-threatening complications such as toxic megacolon (Rea et al., 2010, Warren and Guerrant, 2011). Low doses of antibiotics have also been used as growth promoters in agriculture since the 1950's despite an unclear understanding of the mechanisms at work. A recent investigation into this effect revealed subtherapeutic antibiotic treatment (STAT) of various antibiotics increased adiposity and hormones related to metabolism in young mice compared to untreated controls (Cho et al., 2012). Analysis of the composition and function of the gut microbiota of these animals made it apparent that STAT exposure selected for microbial species that were able to extract more calories from dietary complex carbohydrates that were otherwise indigestible in the control group (Cho et al., 2012).

When considering these results, it is important to be aware that different broad-spectrum antibiotics differ with respect to their impact on the gut microbiota. Changes to the gut microbiota can also be either long- or short-term. In one instance this was highlighted through murine studies which established that mice treated with a cocktail of amoxicillin, metronidazole and bismuth [3.0 mg, 0.69 mg and 0.185 mg, respectively] daily for 10 days had largely recovered their baseline microbial community structure 2 weeks post-treatment but that treatment with cefoperazone [0.5

mg/ml of drinking water] had long-term effects on community structure and reduced overall diversity (Antonopoulos et al., 2009). The effect of an antibiotic on the gut microbiota is influenced by several factors including its antimicrobial effect (bactericidal or bacteriostatic), its mode of action, the structure of the microbiota and the distribution of antibiotic resistance genes among this population (Perez-Cobas et al., 2013).

In light of this greater appreciation of the impact of broad spectrum antimicrobials on the gut microbiota, it is apparent that there is value in utilising antimicrobials with a narrow spectrum of inhibition. In addition to existing repositories of narrow spectrum antimicrobials that were not previously commercialised, it is worth noting that the gut microbiota is considered a rich, but yet relatively, underutilised source of antimicrobial-producing, and in particular bacteriocin-producing, bacteria. Bacteriocins are ribosomally synthesised peptides to which the producer has a specific immunity gene and can have either a narrow or broad spectrum of activity (Cotter et al., 2005a). Many bacteriocins have a number of desirable traits, including low toxicity, high potency and, in the case of gut associated strains, the possibility of *in situ* antimicrobial. production This combination of traits makes them attractive alternatives to traditional antibiotic therapies. Despite being, as stated above, a relatively underutilised source of antimicrobials, a number of bacteriocins have previously been isolated from mammalian gut microbes (O'Shea et al., 2009, Rea et al., 2011, Casey et al., 2004, Lakshminarayanan et al., 2013). Indeed, for example, screening of faecal samples from 266 elderly Irish subjects identified 13 bacteriocin producing strains (Lakshminarayanan et al., 2013) while a further study lead to the isolation of 23 distinct bacteriocin-producing strains from a range of mammalian gastrointestinal sources (O'Shea et al., 2009). Given that, for a bacteriocin to be

produced and be active in the gut, the producer needs to be able to survive in and colonize the human gut and the associated antimicrobial needs to be active in the gut environment, it has been argued that the gut is an ideal source of bacteriocin producers with the potential to alter the gut microbiota (Saarela et al., 2000). There have already been a number of studies which have highlighted the merits of employing gut-associated bacteriocins, several of which we refer to here. In a distal colon model, the narrow spectrum bacteriocin thuricin CD has been observed to inhibit the growth of *C. difficile* without having any significant additional impact on the other components of the gut microbiota (Rea et al., 2011). This contrasted with the significant shift in the relative proportions of the dominant bacterial populations that were observed when the broad-spectrum antimicrobials lacticin 3147, metronidazole and vancomycin, respectively, were employed. Notably, thuricin CD also exhibited a potency comparable to that of the control antimicrobials (Rea et al., 2011), thereby establishing that thuricin CD has potential as an alternative to the conventional antimicrobial strategies employed to treat *C. difficile* infection, especially as it is less likely to impact negatively on the commensal gut microbiota and, thus, is more likely to prevent recurrent *C. difficile* infections. While, in the above example, thuricin CD, rather than the associated *Bacillus thuringiensis* producer (Rea et al., 2010), was employed, there are other examples that have highlighted the merits of using the bacteriocin-producing strain itself. In one such instance, ingestion of the bacteriocin producing probiotic strain *Lactobacillus salivarius* UCC118 provided significant protection against infection by *Listeria monocytogenes* in mice (Corr et al., 2007). Production of the Abp118 bacteriocin by UCC118, which has previously been shown to be capable of altering the intestinal microbiota of pigs and mice (Riboulet-Bisson et al., 2012), proved to be the key protective factor as a non-bacteriocin producing mutant failed to

confer the same protection. This protective effect was also lost when infection was with a bacteriocin-immune *L. monocytogenes* mutant, thereby confirming that the mode of action was direct antagonism by Abp118 rather than *via* some other indirect effect (Corr et al., 2007). In another instance a combination of 5 probiotic strains were employed to control *Salmonella* Typhimurium-induced diarrhoea in pigs (Casey et al., 2007). It was subsequently established that the only bacteriocin-producing strain, *L. salivarius* DPC6005, was the dominant member of the cocktail in both the ileum digesta and in the mucosa. It could not be established, however, if bacteriocin production was directly responsible for anti-*Salmonella* activity (Walsh et al., 2008).

In addition to the control of pathogens, antimicrobials have also been investigated with a view to altering metabolic health in diet-induced obese mice (Murphy et al., 2013). Supplementation of a high-fat diet with vancomycin caused a significant decrease in Firmicutes and Bacteroidetes populations with a corresponding increase in Proteobacteria. This compositional shift was accompanied by a marked decrease in weight gain, fasting blood glucose, plasma TNFα and triglyceride levels compared to the diet-induced obese controls. Although supplementation of the high-fat diet with the bacteriocin-producing probiotic *L. salivarius* UCC118 did not produce any significant changes in the metabolic profiles of the mice, it did result in an increase in relative proportions of Bacteroidetes and Proteobacteria with a corresponding decrease in Actinobacteria. The authors concluded that antimicrobial strategies have the potential to alter both the composition of the gut microbiota and the metabolic health of the host. However, it was noted that care must be taken when choosing the antimicrobial to be used so as to bring about extended beneficial impacts on metabolic health.

As with diet, the vast majority of work concerning modulation of the microbiota by antimicrobials has taken place in mouse models. Nevertheless, the results are encouraging and suggest that carefully selected antimicrobials represent a viable option with respect to intelligently altering the bacterial populations within the human gut. The first step, however, is identifying antimicrobials with this potential.

Identification of antimicrobials with the potential to modulate the gut microbiota

Almost all classes of life are known to produce antimicrobial peptides as a feature of their natural defence (De Smet and Contreras, 2005, Bishop et al., 2017, Bahar and Ren, 2013). The variety of strategies to identify novel antimicrobial producing bacteria can be broadly divided into traditional, culture-based approaches and modern, *in silico*-based strategies. Culture-based approaches, particularly those focused on the mammalian gastrointestinal tract, have identified a large number of antimicrobials with potential importance for both the food and healthcare industries (Rea et al., 2010, Lakshminarayanan et al., 2013, O'Shea et al., 2009, Drissi et al., 2015, Mullane et al., 2014). One such example is bactofencin A, a bacteriocin showing much potential to subtly modulate the gut microbiota (Guinane et al., 2016), which was originally isolated from the porcine gastrointestinal tract (O'Shea et al., 2013).

Culture-based techniques have the limitation however of only selecting easy-to-cultivate microbes and therefore the antimicrobial-producing activity of all of the components of the community are not sampled. To overcome this, culture-independent techniques have been developed. The first advance to address this was functional metagenomics, whereby total DNA from an environment is expressed in a suitable host and this clonal bank is then screened for the desirable trait. Many novel antibiotics

and antibacterial peptides have been identified by these methods (Banik and Brady, 2010) but also has drawbacks such as insert sizes which may be too small to encompass some of the larger bacteriocin gene clusters or the non-expression of the gene clusters in specific hosts.

In silico mining overcomes the challenges associated with culture-dependent approaches, taking advantage of the continually growing volume of data generated by genome and metagenome sequencing projects and exploiting the highly conserved nature of many features of bacteriocin gene clusters. Sequencing consortiums, particularly those centred on the human microbiome, such as the Human Microbiome Project and MetaHIT (The Human Microbiome Project Consortium, 2012, Qin et al., 2010) have provided large quantities of sequencing data from which novel antimicrobial gene clusters can be identified. The robust temporal stability and mean-reverting behaviour of the microbiota (Belstrøm et al., 2016, Jeffery et al., 2016, Gibbons et al., 2017) is perhaps due, in part, to the protection against invading bacteria conferred by bacteriocins and other antimicrobials produced *in situ* (Corr et al., 2007, Moroni et al., 2006, Rea et al., 2011). It, therefore, stands to reason that the mining of the human microbiome merits consideration with respect to the production of a new generation of therapeutics for diseases such as obesity and type 2 diabetes (Walsh et al., 2014).

Until relatively recently, sequence-homology based approaches, such as BLAST, were the primary method of identifying novel putative bacteriocin-encoding gene clusters (PBGCs) and were used with great success (Marsh et al., 2010, Murphy et al., 2011, Singh and Sareen, 2014). For example, the two-peptide lantibiotic, lichenicindin, was originally identified by mining the NCBI bacterial genome database for sequences similar to the lacticin 3147-associated modification protein LtnM1 (Begley et al.,

2009). However, an abundance of tools and algorithms have, in recent years, become available for mining genomic and metagenomic data for bacteriocins and other biosynthetic gene clusters. BAGEL and antiSMASH are two such tools whose user-friendly, online applications are routinely used to identify bacteriocins, antibiotics, and other secondary metabolic biosynthetic gene clusters in genome sequence data (van Heel et al., 2013, Weber et al., 2015). An approach combining both of these tools identified 1814 putative antimicrobial gene clusters from 328 members of the Bacillales order, including bacteriocins, non-ribosomal synthesised peptides and polyketides (Zhao and Kuipers, 2016). One aspect of BAGEL3's functionality is a sequence similarity approach using a manually curated database of bacteriocin prepeptide protein sequences. This database was utilized by Zheng and colleagues to identify PBGCs in genomic and metagenomic data representative of the human gastrointestinal microbiota (Zheng et al., 2014). However, the true power of BAGEL comes from indirect mining for bacteriocin-associated genes, a tactic particularly useful for identifying peptides which undergo significant post-translational modification such as lantibiotics and sactibiotics, as seen in Chapter 2 of this thesis (Walsh et al., 2015).

As mentioned previously, the conserved nature of bacteriocin gene clusters make them ideal candidates for *in silico* mining. The posttranslational modification enzymes, which provide many bacteriocins with structures beyond those possible by ribosomal translation alone, are particularly important in this regard. These modifications are typically key to the peptide's functionality, stability and target recognition and so, unlike bacteriocin prepeptides which show substantial heterogeneity even within the same subclass, these modification proteins possess highly conserved functional domains which have been exploited to identify many novel lantibiotic (Begley et al.,

2009, Lawton et al., 2007, Marsh et al., 2010, Singh and Sareen, 2014, McClerren et al., 2006) and sactibiotic (Murphy et al., 2011) gene clusters.

To date, the greatest proportion of bacteriocin mining efforts appear to focus on lantibiotics, as they are particularly well-characterised, possess highly conserved biosynthetic machinery, and are produced by many food-grade bacteria (van Kraaij et al., 1999). The vast majority of lantibiotics are produced by members of the Firmicutes phylum (Li and O'Sullivan, 2012), a pattern also reflected by investigations of their distribution in the gut microbiota (Walsh et al., 2015, Zheng et al., 2014, Drissi et al., 2015, Walsh et al., 2017). A similar trend is also apparent in the recently characterised sactibiotic class (Murphy et al., 2011). *In silico* mining is not limited to these classes however, as exploration of genomes representative of the human microbiota using the ClusterFinder algorithm led to the discovery and characterisation of a thiopeptide produced by a *Lactobacillus gasseri* vaginal isolate (Donia et al., 2014). This was the first thiopeptide of human origin to be experimentally characterised, although a semisynthetic member of this class of antimicrobial peptides, LFF571, is already showing promise for the treatment of *Clostridium difficile* infections in human trials with outcomes comparable to vancomycin (Mullane et al., 2014). This is an example of overcoming the second major hurdle in *in silico* mining for antimicrobials, i.e., confirmation and characterisation of production *in vitro*. There are many factors influencing bacteriocin production (Guinane et al., 2015), and as production of these compounds is costly in terms of energy and resources, it is ulinkely that all bacteriocins are expressed in pure culture where antimicrobial production is not advantageous. However, haloduracin (McClerren et al., 2006) and lichenicidin (Begley et al., 2009), in addition to the thiopeptide mentioned above, are examples of *in silico*-identified antimirobals which have been characterised i*n vitro.*

Taken together, these results are encouraging from a public health viewpoint as they suggest intelligent mining of the human microbiota's extensive bioactive potential may be the first step in relieving the pressure on antibiotics and overcoming the persistent threat of resistance.

Modulation by Probiotics

The World Health Organization defines probiotics as "live microorganisms which when administered in adequate amounts confer a health benefit on the host" (Pineiro and Stanton, 2007). Probiotics are becoming increasingly popular and are generally marketed as functional foods or dietary supplements. As it has been recognised that changes in the gut microbiota play a role in GI disease then it is not surprising that probiotics are an attractive option with respect to modulation of the gut microbiome. For a probiotic to successfully exert its benefit on the host's gut microbiota it should be able to remain viable during storage and also be capable of surviving, and potentially colonizing, the host's intestinal environment (Vyas and Ranganathan, 2012). The majority of probiotics currently used are members of lactic acid bacteria (LAB) and, more specifically, strains from the genera *Lactobacillus* and *Bifidobacterium* are most commonly used in commercial probiotics. Mixtures of these strains are becoming increasingly popular as researchers gain a deeper understanding of increasing efficacy via possible additive or synergistic effects (Chapman et al., 2011). Rijkers *et al.* categorised the benefit of probiotics into three levels based on location and method; 1) interference with the growth or survival of pathogenic microorganisms in the gut lumen, 2) improvement of mucosal barrier function or mucosal immune system and 3) influence beyond the gut through the systemic immune system and other organs (Rijkers et al., 2010). A study undertaken by Park et al. found

that DIO mice treated with the probiotic strains *Lactobacillus curvatus* HY7601 and *Lactobacillus plantarum* KY1032 experienced reduced body weight gain and fat accumulation in addition to lowered plasma insulin, leptin, total-cholesterol and liver toxicity biomarkers compared to a group on the same diet supplemented with a placebo (Park et al., 2013). Supplementation with these probiotic strains also resulted in down-regulation of pro-inflammatory genes in adipose tissue, up-regulation of fatty acid oxidation-related genes in the liver and significant alterations in the diversity and function of the gut microbiota. Similar results were observed by Yadav *et al.*, who found that administration of the probiotic VSL#3 prevented and treated obesity and diabetes in a number of different murine models through modulation of the gut microbiota. In particular, an increase in the number of butyrate-producing bacteria was linked with enhanced secretion of the hunger-reducing hormone GLP-1 as well as upregulation of genesinvolved in GLP-1 synthesis and excretion (Yadav et al., 2013). McNulty *et al.* observed that, in gnotobiotic mice harbouring a 15-member model human gut microbial community, introduction of 5 probiotic strains isolated from a fermented milk product did not significantly alter the composition of the intestinal microbiota but instead increased the expression of microbial genes involved in carbohydrate and nucleotide metabolism while decreasing expression of genes involved in the metabolism of lipids and amino acids (McNulty et al., 2011). These metatranscriptomic changes were also apparent in the microbiota of human monozygotic twins when fed the same fermented milk product, primarily upregulation of genes involved in carbohydrate metabolism. In addition to their investigation with a view to contributing to the prevention/treatment of obesity and T2D, it should be noted that probiotics are thought to have the potential to treat a wide range of other conditions such as IBS, allergies, *C. difficile* infection, IBD and others by modulation

of the gut microbiota as highlighted in a number of recent manuscripts (Andreasen et al., 2010, Aronsson et al., 2010, Kadooka et al., 2010, Dai et al., 2013, Fitzpatrick, 2013, Veerappan et al., 2012, Miyoshi et al., 2013). As improvements in metagenome sequencing and computational biology permit us to learn more about other gut microbes and their role in human health, there are indications that the future may lie in the development and application of next-generation probiotics such as *Akkermansia muciniphila, Faecalibacterium prausnitzii*, and members of the *Bacteroides* (O'Toole et al., 2017).

Modulation by Faecal Microbiota Transplantation

Following on from the probiotics principle, but on a community rather than strain level, faecal microbiota transplantation (FMT) is the process of transplanting faecal bacterial communities from a healthy individual to a recipient whose microbiota has been disrupted or altered. Although still somewhat in its infancy, FMT is becoming more commonly used as an approach to replenish the gut microbiota in order to alleviate the symptoms of disease. To date, FMT has most commonly been used to treat recurrent *C. difficile* infection (CDI) by replacing populations of commensal bacteria which have been wiped out by antibiotic therapy. Khoruts and colleagues used terminal-restriction fragment length polymorphism and 16S rRNA approaches to compare the bacterial component of a CDI patient's microbiota before and after FMT intervention (Khoruts et al., 2010) and found that, before intervention, the microbiota was deficient in both Bacteroidetes and Firmicutes but 14 days post-transplantation the microbiota was changed to closely resemble the donor's microbiota and was dominated by *Bacteroides* spp. (Khoruts et al., 2010). These results are similar to

findings by Tvede and Rask-Madsen who reported *Bacteroides* spp. were absent in CDI patients but were replenished after FMT intervention (Tvede and Rask-Madsen, 1989). The composition of the donor's microbiota is the key factor in determining the efficacy of this treatment, as shown by Grehan et al. who collected faecal samples from patients undergoing FMT at 4 time points; pre-treatment and at intervals of 4, 8 and 24 weeks post-treatment to determine the effect of FMT on its microbial content (Tvede and Rask-Madsen, 1989). Using a molecular approach they found that the microbiota was altered by FMT intervention and that at 4, 8 and 24 weeks the community of the recipient was composed predominately of bacteria derived from the healthy donor's samples. Crucially, in addition to bringing about desirable microbiota-related changes, FMT has in a high frequency of cases been successful in controlling CDI. In one such study it was revealed that only 1 of 16 patients treated with FMT experienced a recurrence of colitis during the 90 day follow-up period (Aas et al., 2003). Indeed, when many such studies were combined in a systematic literature review by Gough *et al.*, i.e. to examine the effect of FMT on 317 CDI patients across 27 case studies, it was revealed that disease was resolved in 92% of cases (Gough et al., 2011). An interesting development in the application of FMT is the use of synthetic microbial communities in place of undefined mixtures from donors (for review see de Vos et al (de Vos, 2013)). The synthetic mixtures have the advantage of being controlled, tested extensively for the presence of viruses or pathogens and have the potential to be reproducibly manufactured. Petrof *et al.* showed that a defined mixture of 33 isolates, when administered during a colonoscopy, cured the CDI of 2 patients who had previously failed to respond to antibiotic treatment (Petrof et al., 2013). 16S rRNA analysis showed that taxa found in the stool substitute were rare in the patient's gut microbiota before intervention, however following treatment these taxa accounted

for over 25% of sequences recovered from the gut microbiota. The current consensus supports FMT for use in CDI as a safe, well-tolerated, effective treatment with few adverse events (Kelly et al., 2015) and such is the effectiveness of FMT that a technique known as autologous restoration of gastrointestinal flora (ARGF) proposes that an individual's healthy gut microbiota should be encapsulated and stored indefinitely so that their original microbiota can be reintroduced in the event of disruption by antibiotic treatment (Martin, 2009). Although FMT has been most extensively studied with a view to CDI treatment, it has, however, also been investigated as a potential treatment option for a range of microbiota-associated diseases including IBD, IBS, obesity, idiopathic thrombocytopaenic purpura and even multiple sclerosis. A recently published review by Gupta *et al.* summarises the current state of research and possible future directions of the technique (Gupta et al., 2016).

Concluding Remarks

It is well established that the gut microbiota influences host metabolism, nutrient absorption and immune function, and that disruption of this balanced community can have very serious health implications. As we gain a deeper understanding of the specific relationships between the gut microbiota and disease, we expose potential therapeutic targets. Intelligent modulation of the intestinal community is a topic that had gained considerable interest and has the possibility to be extremely beneficial for human health, particularly as advances in metagenomic sequencing and computational analysis permit the identification of novel antimicrobials and probiotics with targeted therapeutic potential.

References

AAS, J., GESSERT, C. E. & BAKKEN, J. S. 2003. Recurrent Clostridium difficile colitis: case series involving 18 patients treated with donor stool administered via a nasogastric tube. *Clin Infect Dis,* 36**,** 580-5.

ADLERBERTH, I. & WOLD, A. E. 2009. Establishment of the gut microbiota in Western infants. *Acta Paediatr,* 98**,** 229-38.

ANDERSON, C. A., BOUCHER, G., LEES, C. W., FRANKE, A., D'AMATO, M., TAYLOR, K. D., LEE, J. C., GOYETTE, P., IMIELINSKI, M. & LATIANO, A. 2011. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics,* 43**,** 246-252.

ANDREASEN, A. S., LARSEN, N., PEDERSEN-SKOVSGAARD, T., BERG, R. M., MOLLER, K., SVENDSEN, K. D., JAKOBSEN, M. & PEDERSEN, B. K. 2010. Effects of Lactobacillus acidophilus NCFM on insulin sensitivity and the systemic inflammatory response in human subjects. *Br J Nutr,* 104**,** 1831-8.

ANTONOPOULOS, D. A., HUSE, S. M., MORRISON, H. G., SCHMIDT, T. M., SOGIN, M. L. & YOUNG, V. B. 2009. Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation. *Infection and Immunity,* 77**,** 2367-2375.

ARNOLD, I. C., DEHZAD, N., REUTER, S., MARTIN, H., BECHER, B., TAUBE, C., XFC & LLER, A. 2011. Helicobacter pylori infection prevents allergic asthma in mouse models through the induction of regulatory T cells. *The Journal of Clinical Investigation,* 121**,** 3088-3093.

ARONSSON, L., HUANG, Y., PARINI, P., KORACH-ANDRE, M., HAKANSSON, J., GUSTAFSSON, J. A., PETTERSSON, S., ARULAMPALAM, V. & RAFTER, J. 2010. Decreased fat storage by Lactobacillus paracasei is associated with increased levels of angiopoietin-like 4 protein (ANGPTL4). *PLoS One,* 5.

ARTHUR, J. C., PEREZ-CHANONA, E., MUHLBAUER, M., TOMKOVICH, S., URONIS, J. M., FAN, T. J., CAMPBELL, B. J., ABUJAMEL, T., DOGAN, B., ROGERS, A. B., RHODES, J. M., STINTZI, A., SIMPSON, K. W., HANSEN, J. J., KEKU, T. O., FODOR, A. A. & JOBIN, C. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science,* 338**,** 120-3.

BACKHED, F., DING, H., WANG, T., HOOPER, L. V., KOH, G. Y., NAGY, A., SEMENKOVICH, C. F. & GORDON, J. I. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A,* 101**,** 15718-23.

BÄCKHED, F., LEY, R. E., SONNENBURG, J. L., PETERSON, D. A. & GORDON, J. I. 2005. Host-Bacterial Mutualism in the Human Intestine. *Science,* 307**,** 1915-1920.

BAHAR, A. A. & REN, D. 2013. Antimicrobial Peptides. *Pharmaceuticals,* 6**,** 1543-1575.

BANIK, J. J. & BRADY, S. F. 2010. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr Opin Microbiol,* 13**,** 603-9.

BARTON, W., PENNEY, N. C., CRONIN, O., GARCIA-PEREZ, I., MOLLOY, M. G., HOLMES, E., SHANAHAN, F., COTTER, P. D. & O'SULLIVAN, O.

2017. The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. *Gut*.

BEGLEY, M., COTTER, P. D., HILL, C. & ROSS, R. P. 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl Environ Microbiol,* 75**,** 5451-60.

BELSTRØM, D., HOLMSTRUP, P., BARDOW, A., KOKARAS, A., FIEHN, N.-E. & PASTER, B. J. 2016. Temporal Stability of the Salivary Microbiota in Oral Health. *PLoS ONE,* 11**,** e0147472.

BENGTSON, M. B., AAMODT, G., VATN, M. H. & HARRIS, J. R. 2010. Concordance for IBD among twins compared to ordinary siblings--a Norwegian population-based study. *J Crohns Colitis,* 4**,** 312-8.

BISHOP, B. M., JUBA, M. L., RUSSO, P. S., DEVINE, M., BARKSDALE, S. M., SCOTT, S., SETTLAGE, R., MICHALAK, P., GUPTA, K., VLIET, K., SCHNUR, J. M. & VAN HOEK, M. L. 2017. Discovery of Novel Antimicrobial Peptides from Varanus komodoensis (Komodo Dragon) by Large-Scale Analyses and De-Novo-Assisted Sequencing Using Electron-Transfer Dissociation Mass Spectrometry. *J Proteome Res*.

BLASER, M. 2011. Antibiotic overuse: Stop the killing of beneficial bacteria. *Nature,* 476**,** 393-394.

BOUHNIK, Y., ALAIN, S., ATTAR, A., FLOURIE, B., RASKINE, L., SANSON-LE PORS, M. J. & RAMBAUD, J. C. 1999. Bacterial populations contaminating the upper gut in patients with small intestinal bacterial overgrowth syndrome. *Am J Gastroenterol,* 94**,** 1327-31.

BOUHNIK, Y., RASKINE, L., SIMONEAU, G., VICAUT, E., NEUT, C., FLOURIÉ, B., BROUNS, F. & BORNET, F. R. 2004. The capacity of nondigestible carbohydrates to stimulate fecal bifidobacteria in healthy humans: a double-blind, randomized, placebo-controlled, parallel-group, dose-response relation study. *The American journal of clinical nutrition,* 80**,** 1658-1664.

BRINT, E. K., MACSHARRY, J., FANNING, A., SHANAHAN, F. & QUIGLEY, E. M. 2011. Differential expression of toll-like receptors in patients with irritable bowel syndrome. *Am J Gastroenterol,* 106**,** 329-36.

BROOKS, A. W., KOHL, K. D., BRUCKER, R. M., VAN OPSTAL, E. J. & BORDENSTEIN, S. R. 2016. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biology,* 14**,** e2000225.

BULTMAN, S. J. 2017. Interplay between diet, gut microbiota, epigenetic events, and colorectal cancer. *Molecular Nutrition & Food Research,* 61**,** 1500902-n/a.

BULTMAN, S. J. & JOBIN, C. 2014. Microbial-derived butyrate: an oncometabolite or tumor-suppressive metabolite? *Cell Host Microbe,* 16**,** 143-5.

CANI, P. D. 2013. Gut microbiota and obesity: lessons from the microbiome. *Brief Funct Genomics,* 12**,** 381-7.

CASEY, P. G., CASEY, G. D., GARDINER, G. E., TANGNEY, M., STANTON, C., ROSS, R. P., HILL, C. & FITZGERALD, G. F. 2004. Isolation and characterization of anti-Salmonella lactic acid bacteria from the porcine gastrointestinal tract. *Letters in Applied Microbiology,* 39**,** 431-438.

CASEY, P. G., GARDINER, G. E., CASEY, G., BRADSHAW, B., LAWLOR, P. G., LYNCH, P. B., LEONARD, F. C., STANTON, C., ROSS, R. P. & FITZGERALD, G. F. 2007. A five-strain probiotic combination reduces

pathogen shedding and alleviates disease signs in pigs challenged with Salmonella enterica serovar Typhimurium. *Applied and Environmental Microbiology,* 73**,** 1858-1863.

CASTELLARIN, M., WARREN, R. L., FREEMAN, J. D., DREOLINI, L., KRZYWINSKI, M., STRAUSS, J., BARNES, R., WATSON, P., ALLEN-VERCOE, E., MOORE, R. A. & HOLT, R. A. 2012. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res,* 22**,** 299-306.

CHAPMAN, C., GIBSON, G. & ROWLAND, I. 2011. Health benefits of probiotics: are mixtures more effective than single strains? *European journal of nutrition,* 50**,** 1-17.

CHEN, J., WRIGHT, K., DAVIS, J. M., JERALDO, P., MARIETTA, E. V., MURRAY, J., NELSON, H., MATTESON, E. L. & TANEJA, V. 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Medicine,* 8**,** 43.

CHEN, Y. & BLASER, M. J. 2007. Inverse associations of helicobacter pylori with asthma and allergy. *Archives of Internal Medicine,* 167**,** 821-827.

CHO, I., YAMANISHI, S., COX, L., METHE, B. A., ZAVADIL, J., LI, K., GAO, Z., MAHANA, D., RAJU, K., TEITLER, I., LI, H., ALEKSEYENKO, A. V. & BLASER, M. J. 2012. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature,* 488**,** 621-6.

CLAESSON, M. J., CUSACK, S., O'SULLIVAN, O., GREENE-DINIZ, R., DE WEERD, H., FLANNERY, E., MARCHESI, J. R., FALUSH, D., DINAN, T., FITZGERALD, G., STANTON, C., VAN SINDEREN, D., O'CONNOR, M., HARNEDY, N., O'CONNOR, K., HENRY, C., O'MAHONY, D.,

FITZGERALD, A. P., SHANAHAN, F., TWOMEY, C., HILL, C., ROSS, R. P. & O'TOOLE, P. W. 2011. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4586-91.

CLAESSON, M. J., JEFFERY, I. B., CONDE, S., POWER, S. E., O/'CONNOR, E. M., CUSACK, S., HARRIS, H. M. B., COAKLEY, M., LAKSHMINARAYANAN, B., O/'SULLIVAN, O., FITZGERALD, G. F., DEANE, J., O/'CONNOR, M., HARNEDY, N., O/'CONNOR, K., O/'MAHONY, D., VAN SINDEREN, D., WALLACE, M., BRENNAN, L., STANTON, C., MARCHESI, J. R., FITZGERALD, A. P., SHANAHAN, F., HILL, C., ROSS, R. P. & O/'TOOLE, P. W. 2012. Gut microbiota composition correlates with diet and health in the elderly. *Nature,* 488**,** 178-184.

CLARKE, S. F., MURPHY, E. F., NILAWEERA, K., ROSS, P. R., SHANAHAN, F., O'TOOLE, P. W. & COTTER, P. D. 2012. The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes,* 3**,** 186-202.

CLARKE, S. F., MURPHY, E. F., O'SULLIVAN, O., LUCEY, A. J., HUMPHREYS, M., HOGAN, A., HAYES, P., O'REILLY, M., JEFFERY, I. B., WOOD-MARTIN, R., KERINS, D. M., QUIGLEY, E., ROSS, R. P., O'TOOLE, P. W., MOLLOY, M. G., FALVEY, E., SHANAHAN, F. & COTTER, P. D. 2014. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*.

CLEMENTE, JOSE C., URSELL, LUKE K., PARFREY, LAURA W. & KNIGHT, R. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell,* 148**,** 1258-1270.

CLINTON, S. K., BOSTWICK, D. G., OLSON, L. M., MANGIAN, H. J. & VISEK, W. J. 1988. Effects of ammonium acetate and sodium cholate on N-methyl-N′-nitro-N-nitrosoguanidine-induced colon carcinogenesis of rats. *Cancer research,* 48**,** 3035-3039.

CORR, S. C., LI, Y., RIEDEL, C. U., O'TOOLE, P. W., HILL, C. & GAHAN, C. G. M. 2007. Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118. *Proceedings of the National Academy of Sciences,* 104**,** 7617-7621.

COSTABILE, A., KOLIDA, S., KLINDER, A., GIETL, E., BÄUERLEIN, M., FROHBERG, C., LANDSCHÜTZE, V. & GIBSON, G. R. 2010. A double-blind, placebo-controlled, cross-over study to establish the bifidogenic effect of a very-long-chain inulin extracted from globe artichoke (Cynara scolymus) in healthy human subjects. *British Journal of Nutrition,* 104**,** 1007.

COTILLARD, A., KENNEDY, S. P., KONG, L. C., PRIFTI, E., PONS, N., LE CHATELIER, E., ALMEIDA, M., QUINQUIS, B., LEVENEZ, F., GALLERON, N., GOUGIS, S., RIZKALLA, S., BATTO, J.-M., RENAULT, P., CONSORTIUM, A. N. R. M., DORE, J., ZUCKER, J.-D., CLEMENT, K., EHRLICH, S. D. & MEMBERS, A. N. R. M. C. 2013. Dietary intervention impact on gut microbial gene richness. *Nature,* 500**,** 585-588.

COTTER, P. D., HILL, C. & ROSS, R. P. 2005. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol,* 3**,** 777-88.

COTTER, P. D., STANTON, C., ROSS, R. P. & HILL, C. 2012. The impact of antibiotics on the gut microbiota as revealed by high throughput DNA sequencing. *Discovery medicine,* 13**,** 193.

CRYAN, J. F. & O'MAHONY, S. M. 2011. The microbiome-gut-brain axis: from bowel to behavior. *Neurogastroenterol Motil,* 23**,** 187-92.

CUMMINGS, J. & MACFARLANE, G. 1991. The control and consequences of bacterial fermentation in the human colon. *Journal of Applied Microbiology,* 70**,** 443-459.

CUMMINGS, J. H. & ENGLYST, H. N. 1991. What is dietary fibre? *Trends in Food Science & Technology,* 2**,** 99-103.

DAI, C., ZHENG, C. Q., JIANG, M., MA, X. Y. & JIANG, L. J. 2013. Probiotics and irritable bowel syndrome. *World J Gastroenterol,* 19**,** 5973-80.

DANIEL, H., GHOLAMI, A. M., BERRY, D., DESMARCHELIER, C., HAHNE, H., LOH, G., MONDOT, S., LEPAGE, P., ROTHBALLER, M., WALKER, A., BOHM, C., WENNING, M., WAGNER, M., BLAUT, M., SCHMITT-KOPPLIN, P., KUSTER, B., HALLER, D. & CLAVEL, T. 2013. High-fat diet alters gut microbiota physiology in mice. *ISME J*.

DAO, M. C., EVERARD, A., ARON-WISNEWSKY, J., SOKOLOVSKA, N., PRIFTI, E., VERGER, E. O., KAYSER, B. D., LEVENEZ, F., CHILLOUX, J., HOYLES, L., DUMAS, M. E., RIZKALLA, S. W., DORE, J., CANI, P. D. & CLEMENT, K. 2016. Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut,* 65**,** 426-36.

DAVID, L. A., MAURICE, C. F., CARMODY, R. N., GOOTENBERG, D. B., BUTTON, J. E., WOLFE, B. E., LING, A. V., DEVLIN, A. S., VARMA, Y. & FISCHBACH, M. A. 2013. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*.

DE FILIPPO, C., CAVALIERI, D., DI PAOLA, M., RAMAZZOTTI, M., POULLET, J. B., MASSART, S., COLLINI, S., PIERACCINI, G. & LIONETTI, P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences,* 107**,** 14691-14696.

DE LANGE, K. M., MOUTSIANAS, L., LEE, J. C., LAMB, C. A., LUO, Y., KENNEDY, N. A., JOSTINS, L., RICE, D. L., GUTIERREZ-ACHURY, J., JI, S.-G., HEAP, G., NIMMO, E. R., EDWARDS, C., HENDERSON, P., MOWAT, C., SANDERSON, J., SATSANGI, J., SIMMONS, A., WILSON, D. C., TREMELLING, M., HART, A., MATHEW, C. G., NEWMAN, W. G., PARKES, M., LEES, C. W., UHLIG, H., HAWKEY, C., PRESCOTT, N. J., AHMAD, T., MANSFIELD, J. C., ANDERSON, C. A. & BARRETT, J. C. 2017. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet,* 49**,** 256-261.

DE SMET, K. & CONTRERAS, R. 2005. Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnol Lett,* 27**,** 1337-47.

DE VOS, W. M. 2013. Fame and future of faecal transplantations--developing next-generation therapies with synthetic microbiomes. *Microb Biotechnol,* 6**,** 316-25.

DE VADDER, F., KOVATCHEVA-DATCHARY, P., GONCALVES, D., VINERA, J., ZITOUN, C., DUCHAMPT, A., BÄCKHED, F. & MITHIEUX, G. 2014. Microbiota-Generated Metabolites Promote Metabolic Benefits via Gut-Brain Neural Circuits. *Cell,* 156**,** 84-96.

DOMINGUEZ-BELLO, M. G., COSTELLO, E. K. & KNIGHT, R. 2010. Reply to Putignani et al.: Vagina as a major source of natural inoculum for the newborn. *Proceedings of the National Academy of Sciences,* 107**,** E160.

DRISSI, F., BUFFET, S., RAOULT, D. & MERHEJ, V. 2015. Common occurrence of antibacterial agents in human intestinal microbiota. *Frontiers in Microbiology,* 6.

DUNCAN, S. H., BELENGUER, A., HOLTROP, G., JOHNSTONE, A. M., FLINT, H. J. & LOBLEY, G. E. 2007. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Applied and environmental microbiology,* 73**,** 1073-1078.

ELSON, C. O. & CONG, Y. 2012. Host-microbiota interactions in inflammatory bowel disease. *Gut Microbes,* 3**,** 332-44.

ELWOOD, P. C., GIVENS, D. I., BESWICK, A. D., FEHILY, A. M., PICKERING, J. E. & GALLACHER, J. 2008. The survival advantage of milk and dairy consumption: an overview of evidence from cohort studies of vascular diseases, diabetes and cancer. *J Am Coll Nutr,* 27**,** 723s-34s.

EVERARD, A., BELZER, C., GEURTS, L., OUWERKERK, J. P., DRUART, C., BINDELS, L. B., GUIOT, Y., DERRIEN, M., MUCCIOLI, G. G., DELZENNE, N. M., DE VOS, W. M. & CANI, P. D. 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A,* 110**,** 9066-71.

FAITH, J. J., GURUGE, J. L., CHARBONNEAU, M., SUBRAMANIAN, S., SEEDORF, H., GOODMAN, A. L., CLEMENTE, J. C., KNIGHT, R.,

HEATH, A. C., LEIBEL, R. L., ROSENBAUM, M. & GORDON, J. I. 2013. The Long-Term Stability of the Human Gut Microbiota. *Science,* 341.

FAITH, J. J., MCNULTY, N. P., REY, F. E. & GORDON, J. I. 2011. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science,* 333**,** 101-104.

FITZPATRICK, L. R. 2013. Probiotics for the treatment of Clostridium difficile associated disease. *World J Gastrointest Pathophysiol,* 4**,** 47-52.

FLINT, H. J., SCOTT, K. P., LOUIS, P. & DUNCAN, S. H. 2012. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol,* 9**,** 577-589.

FRANCESCHI, C. 2007. Inflammaging as a major characteristic of old people: can it be prevented or cured? *Nutr Rev,* 65**,** S173-6.

FRANK, D. N., AMAND, A. L. S., FELDMAN, R. A., BOEDEKER, E. C., HARPAZ, N. & PACE, N. R. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences,* 104**,** 13780-13785.

FRANKE, A., MCGOVERN, D. P., BARRETT, J. C., WANG, K., RADFORD-SMITH, G. L., AHMAD, T., LEES, C. W., BALSCHUN, T., LEE, J., ROBERTS, R., ANDERSON, C. A., BIS, J. C., BUMPSTEAD, S., ELLINGHAUS, D., FESTEN, E. M., GEORGES, M., GREEN, T., HARITUNIANS, T., JOSTINS, L., LATIANO, A., MATHEW, C. G., MONTGOMERY, G. W., PRESCOTT, N. J., RAYCHAUDHURI, S., ROTTER, J. I., SCHUMM, P., SHARMA, Y., SIMMS, L. A., TAYLOR, K. D., WHITEMAN, D., WIJMENGA, C., BALDASSANO, R. N., BARCLAY, M., BAYLESS, T. M., BRAND, S., BUNING, C., COHEN, A., COLOMBEL, J. F., COTTONE, M., STRONATI, L., DENSON, T., DE VOS, M., D'INCA,

R., DUBINSKY, M., EDWARDS, C., FLORIN, T., FRANCHIMONT, D., GEARRY, R., GLAS, J., VAN GOSSUM, A., GUTHERY, S. L., HALFVARSON, J., VERSPAGET, H. W., HUGOT, J. P., KARBAN, A., LAUKENS, D., LAWRANCE, I., LEMANN, M., LEVINE, A., LIBIOULLE, C., LOUIS, E., MOWAT, C., NEWMAN, W., PANES, J., PHILLIPS, A., PROCTOR, D. D., REGUEIRO, M., RUSSELL, R., RUTGEERTS, P., SANDERSON, J., SANS, M., SEIBOLD, F., STEINHART, A. H., STOKKERS, P. C., TORKVIST, L., KULLAK-UBLICK, G., WILSON, D., WALTERS, T., TARGAN, S. R., BRANT, S. R., RIOUX, J. D., D'AMATO, M., WEERSMA, R. K., KUGATHASAN, S., GRIFFITHS, A. M., MANSFIELD, J. C., VERMEIRE, S., DUERR, R. H., SILVERBERG, M. S., SATSANGI, J., SCHREIBER, S., CHO, J. H., ANNESE, V., HAKONARSON, H., DALY, M. J. & PARKES, M. 2010. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet,* 42**,** 1118-25.

GABERT, L., VORS, C., LOUCHE-PÉLISSIER, C., SAUVINET, V., LAMBERT-PORCHERON, S., DRAI, J., LAVILLE, M., DÉSAGE, M. & MICHALSKI, M. C. 2011. 13C tracer recovery in human stools after digestion of a fat-rich meal labelled with [1, 1, 1-13C3] tripalmitin and [1, 1, 1-13C3] triolein. *Rapid Communications in Mass Spectrometry,* 25**,** 2697-2703.

GARRETT, W. S., GALLINI, C. A., YATSUNENKO, T., MICHAUD, M., DUBOIS, A., DELANEY, M. L., PUNIT, S., KARLSSON, M., BRY, L., GLICKMAN, J. N., GORDON, J. I., ONDERDONK, A. B. & GLIMCHER, L. H. 2010. Enterobacteriaceae Act in Concert with the Gut Microbiota to Induce

Spontaneous and Maternally Transmitted Colitis. *Cell Host & Microbe,* 8**,** 292-300.

GIBBONS, S. M., KEARNEY, S. M., SMILLIE, C. S. & ALM, E. J. 2017. Two dynamic regimes in the human gut microbiome. *PLOS Computational Biology,* 13**,** e1005364.

GIBSON, G. R., PROBERT, H. M., VAN LOO, J., RASTALL, R. A. & ROBERFROID, M. B. 2004. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev,* 17**,** 259-275.

GIBSON, G. R. & ROBERFROID, M. B. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr,* 125**,** 1401-12.

GOUGH, E., SHAIKH, H. & MANGES, A. R. 2011. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection. *Clin Infect Dis,* 53**,** 994-1002.

GUINANE, C. M., LAWTON, E. M., O'CONNOR, P. M., O'SULLIVAN, O., HILL, C., ROSS, R. P. & COTTER, P. D. 2016. The bacteriocin bactofencin A subtly modulates gut microbial populations. *Anaerobe,* 40**,** 41-9.

GUINANE, C. M., PIPER, C., DRAPER, L. A., O'CONNOR, P. M., HILL, C., ROSS, R. P. & COTTER, P. D. 2015. Impact of Environmental Factors on Bacteriocin Promoter Activity in Gut-Derived Lactobacillus salivarius. *Applied and Environmental Microbiology,* 81**,** 7851-7859.

GUPTA, S., ALLEN-VERCOE, E. & PETROF, E. O. 2016. Fecal microbiota transplantation: in perspective. *Therap Adv Gastroenterol,* 9**,** 229-39.

HAMER, H. M., DE PRETER, V., WINDEY, K. & VERBEKE, K. 2012. Functional analysis of colonic bacterial metabolism: relevant to health? *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 302**,** G1-G9.

HANSEN, R., RUSSELL, R. K., REIFF, C., LOUIS, P., MCINTOSH, F., BERRY, S. H., MUKHOPADHYA, I., BISSET, W. M., BARCLAY, A. R., BISHOP, J., FLYNN, D. M., MCGROGAN, P., LOGANATHAN, S., MAHDI, G., FLINT, H. J., EL-OMAR, E. M. & HOLD, G. L. 2012. Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. *Am J Gastroenterol,* 107**,** 1913-1922.

HILDEBRANDT, M. A., HOFFMANN, C., SHERRILL–MIX, S. A., KEILBAUGH, S. A., HAMADY, M., CHEN, Y. Y., KNIGHT, R., AHIMA, R. S., BUSHMAN, F. & WU, G. D. 2009. High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. *Gastroenterology,* 137**,** 1716-1724.e2.

JALANKA-TUOVINEN, J., SALOJÄRVI, J., SALONEN, A., IMMONEN, O., GARSED, K., KELLY, F. M., ZAITOUN, A., PALVA, A., SPILLER, R. C. & DE VOS, W. M. 2013. Faecal microbiota composition and host–microbe cross-talk following gastroenteritis and in postinfectious irritable bowel syndrome. *Gut*.

JEFFERY, I. B., LYNCH, D. B. & O'TOOLE, P. W. 2016. Composition and temporal stability of the gut microbiota in older persons. *ISME J,* 10**,** 170-182.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. 2011. Global cancer statistics. *CA Cancer J Clin,* 61**,** 69-90.

JOSTINS, L., RIPKE, S., WEERSMA, R. K., DUERR, R. H., MCGOVERN, D. P., HUI, K. Y., LEE, J. C., PHILIP SCHUMM, L., SHARMA, Y., ANDERSON, C. A., ESSERS, J., MITROVIC, M., NING, K., CLEYNEN, I., THEATRE, E., SPAIN, S. L., RAYCHAUDHURI, S., GOYETTE, P., WEI, Z., ABRAHAM, C., ACHKAR, J.-P., AHMAD, T., AMININEJAD, L., ANANTHAKRISHNAN, A. N., ANDERSEN, V., ANDREWS, J. M., BAIDOO, L., BALSCHUN, T., BAMPTON, P. A., BITTON, A., BOUCHER, G., BRAND, S., BUNING, C., COHAIN, A., CICHON, S., D/'AMATO, M., DE JONG, D., DEVANEY, K. L., DUBINSKY, M., EDWARDS, C., ELLINGHAUS, D., FERGUSON, L. R., FRANCHIMONT, D., FRANSEN, K., GEARRY, R., GEORGES, M., GIEGER, C., GLAS, J., HARITUNIANS, T., HART, A., HAWKEY, C., HEDL, M., HU, X., KARLSEN, T. H., KUPCINSKAS, L., KUGATHASAN, S., LATIANO, A., LAUKENS, D., LAWRANCE, I. C., LEES, C. W., LOUIS, E., MAHY, G., MANSFIELD, J., MORGAN, A. R., MOWAT, C., NEWMAN, W., PALMIERI, O., PONSIOEN, C. Y., POTOCNIK, U., PRESCOTT, N. J., REGUEIRO, M., ROTTER, J. I., RUSSELL, R. K., SANDERSON, J. D., SANS, M., SATSANGI, J., SCHREIBER, S., SIMMS, L. A., SVENTORAITYTE, J., TARGAN, S. R., TAYLOR, K. D., TREMELLING, M., VERSPAGET, H. W., DE VOS, M., WIJMENGA, C., WILSON, D. C., WINKELMANN, J., XAVIER, R. J., ZEISSIG, S., ZHANG, B., ZHANG, C. K., ZHAO, H., SILVERBERG, M. S., ANNESE, V., HAKONARSON, H., BRANT, S. R., RADFORD-SMITH, G., MATHEW, C. G., RIOUX, J. D., SCHADT, E. E., et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature,* 491**,** 119-124.

KABEERDOSS, J., DEVI, R. S., MARY, R. R. & RAMAKRISHNA, B. S. 2012. Faecal microbiota composition in vegetarians: comparison with omnivores in a cohort of young women in southern India. *Br J Nutr,* 108**,** 953-7.

KADOOKA, Y., SATO, M., IMAIZUMI, K., OGAWA, A., IKUYAMA, K., AKAI, Y., OKANO, M., KAGOSHIMA, M. & TSUCHIDA, T. 2010. Regulation of abdominal adiposity by probiotics (Lactobacillus gasseri SBT2055) in adults with obese tendencies in a randomized controlled trial. *Eur J Clin Nutr,* 64**,** 636-43.

KELLY, C. R., KAHN, S., KASHYAP, P., LAINE, L., RUBIN, D., ATREJA, A., MOORE, T. & WU, G. 2015. Update on Fecal Microbiota Transplantation 2015: Indications, Methodologies, Mechanisms, and Outlook. *Gastroenterology,* 149**,** 223-37.

KHAN, M. J., GERASIMIDIS, K., EDWARDS, C. A. & SHAIKH, M. G. 2016. Role of Gut Microbiota in the Aetiology of Obesity: Proposed Mechanisms and Review of the Literature. *Journal of Obesity,* 2016**,** 27.

KHORUTS, A., DICKSVED, J., JANSSON, J. K. & SADOWSKY, M. J. 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol,* 44**,** 354-60.

KOLEVA, P. T., VALCHEVA, R. S., SUN, X., GÄNZLE, M. G. & DIELEMAN, L. A. 2012. Inulin and fructo-oligosaccharides have divergent effects on colitis and commensal microbiota in HLA-B27 transgenic rats. *British Journal of Nutrition,* 108**,** 1633-1643.

KOSTIC, A. D., CHUN, E., ROBERTSON, L., GLICKMAN, J. N., GALLINI, C. A., MICHAUD, M., CLANCY, T. E., CHUNG, D. C., LOCHHEAD, P., HOLD,

G. L., EL-OMAR, E. M., BRENNER, D., FUCHS, C. S., MEYERSON, M. & GARRETT, W. S. 2013. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe,* 14**,** 207-15.

KOSTIC, A. D., GEVERS, D., PEDAMALLU, C. S., MICHAUD, M., DUKE, F., EARL, A. M., OJESINA, A. I., JUNG, J., BASS, A. J., TABERNERO, J., BASELGA, J., LIU, C., SHIVDASANI, R. A., OGINO, S., BIRREN, B. W., HUTTENHOWER, C., GARRETT, W. S. & MEYERSON, M. 2012. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res,* 22**,** 292-8.

LAKSHMINARAYANAN, B., GUINANE, C. M., O'CONNOR, P. M., COAKLEY, M., HILL, C., STANTON, C., O'TOOLE, P. W. & ROSS, R. P. 2013. Isolation and characterization of bacteriocin-producing bacteria from the intestinal microbiota of elderly Irish subjects. *Journal of Applied Microbiology,* 114**,** 886-898.

LARSEN, N., VOGENSEN, F. K., VAN DEN BERG, F. W., NIELSEN, D. S., ANDREASEN, A. S., PEDERSEN, B. K., AL-SOUD, W. A., SORENSEN, S. J., HANSEN, L. H. & JAKOBSEN, M. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One,* 5**,** e9085.

LAWTON, E. M., COTTER, P. D., HILL, C. & ROSS, R. P. 2007. Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile Bacillus halodurans C-125. *FEMS Microbiol Lett,* 267**,** 64-71.

LEPAGE, P., HÄSLER, R., SPEHLMANN, M. E., REHMAN, A., ZVIRBLIENE, A., BEGUN, A., OTT, S., KUPCINSKAS, L., DORÉ, J. & RAEDLER, A.

2011. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology,* 141**,** 227-236.

LEWIS, S., BRAZIER, J., BEARD, D., NAZEM, N. & PROCTOR, D. 2005. Effects of metronidazole and oligofructose on faecal concentrations of sulphate-reducing bacteria and their activity in human volunteers. *Scandinavian journal of gastroenterology,* 40**,** 1296-1303.

LEY, R. E., BÄCKHED, F., TURNBAUGH, P., LOZUPONE, C. A., KNIGHT, R. D. & GORDON, J. I. 2005. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 11070-11075.

LEY, R. E., HAMADY, M., LOZUPONE, C., TURNBAUGH, P. J., RAMEY, R. R., BIRCHER, J. S., SCHLEGEL, M. L., TUCKER, T. A., SCHRENZEL, M. D. & KNIGHT, R. 2008. Evolution of mammals and their gut microbes. *Science,* 320**,** 1647-1651.

LHOSTE, E. F., MOUZON, B., ANDRIEUX, C., GUEUGNEAU, A.-M., FISZLEWICZ, M., LE, E., CORRING, T. & SZYLIT, O. 1998. Physiological effects of a pea protein isolate in gnotobiotic rats: comparison with a soybean isolate and meat. *Annals of nutrition and metabolism,* 42**,** 44-54.

LI, X. & O'SULLIVAN, D. J. 2012. Contribution of the Actinobacteria to the growing diversity of lantibiotics. *Biotechnol Lett,* 34**,** 2133-45.

LIN, H. C. 2004. Small intestinal bacterial overgrowth: a framework for understanding irritable bowel syndrome. *Jama,* 292**,** 852-8.

LISZT, K., ZWIELEHNER, J., HANDSCHUR, M., HIPPE, B., THALER, R. & HASLBERGER, A. G. 2009. Characterization of bacteria, clostridia and

Bacteroides in faeces of vegetarians using qPCR and PCR-DGGE fingerprinting. *Ann Nutr Metab,* 54**,** 253-7.

LYSSENKO, V., JONSSON, A., ALMGREN, P., PULIZZI, N., ISOMAA, B., TUOMI, T., BERGLUND, G., ALTSHULER, D., NILSSON, P. & GROOP, L. 2008. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med,* 359**,** 2220-32.

MACFARLANE, G., CUMMINGS, J. & ALLISON, C. 1986. Protein degradation by human intestinal bacteria. *Journal of general microbiology,* 132**,** 1647-1656.

MACFARLANE, G. T. & MACFARLANE, S. 2012. Bacteria, colonic fermentation, and gastrointestinal health. *Journal of AOAC International,* 95**,** 50-60.

MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., EECKHAUT, V., BALLET, V., CLAES, K., VAN IMMERSEEL, F., VERBEKE, K., FERRANTE, M., VERHAEGEN, J., RUTGEERTS, P. & VERMEIRE, S. 2013. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut*.

MAGEE, E. A., RICHARDSON, C. J., HUGHES, R. & CUMMINGS, J. H. 2000. Contribution of dietary protein to sulfide production in the large intestine: an in vitro and a controlled feeding study in humans. *The American journal of clinical nutrition,* 72**,** 1488-1494.

MANRIQUE, P., BOLDUC, B., WALK, S. T., VAN DER OOST, J., DE VOS, W. M. & YOUNG, M. J. 2016. Healthy human gut phageome. *Proceedings of the National Academy of Sciences,* 113**,** 10400-10405.

MARSH, A. J., O'SULLIVAN, O., ROSS, R. P., COTTER, P. D. & HILL, C. 2010. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics,* 11**,** 679.

MARTIN, W. 2009. Encapsulated Medicines for Iatrogenic Diseases. *British Patent Application: GB0916335,* 3.

MATIJASIC, B. B., OBERMAJER, T., LIPOGLAVSEK, L., GRABNAR, I., AVGUSTIN, G. & ROGELJ, I. 2013. Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur J Nutr*.

MÄTTÖ, J., MAUNUKSELA, L., KAJANDER, K., PALVA, A., KORPELA, R., KASSINEN, A. & SAARELA, M. 2005. Composition and temporal stability of gastrointestinal microbiota in irritable bowel syndrome – a longitudinal study in IBS and control subjects. *FEMS Immunology & Medical Microbiology,* 43**,** 213-222.

MCALLAN, L., SKUSE, P., COTTER, P. D., O' CONNOR, P., CRYAN, J. F., ROSS, R. P., FITZGERALD, G., ROCHE, H. M. & NILAWEERA, K. N. 2014. Whey protein isolate at varying doses differentially influences energy balance and the composition of the gut microbiota in high-fat fed mice. *PLoS One***,** (in print).

MCCLERREN, A. L., COOPER, L. E., QUAN, C., THOMAS, P. M., KELLEHER, N. L. & VAN DER DONK, W. A. 2006. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 17243-17248.

MCCOY, A. N., ARAUJO-PEREZ, F., AZCARATE-PERIL, A., YEH, J. J., SANDLER, R. S. & KEKU, T. O. 2013. Fusobacterium is associated with colorectal adenomas. *PLoS One,* 8**,** e53653.

MCNULTY, N. P., YATSUNENKO, T., HSIAO, A., FAITH, J. J., MUEGGE, B. D., GOODMAN, A. L., HENRISSAT, B., OOZEER, R., COOLS-PORTIER, S., GOBERT, G., CHERVAUX, C., KNIGHTS, D., LOZUPONE, C. A., KNIGHT, R., DUNCAN, A. E., BAIN, J. R., MUEHLBAUER, M. J., NEWGARD, C. B., HEATH, A. C. & GORDON, J. I. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med,* 3**,** 106ra106.

MIYOSHI, M., OGAWA, A., HIGURASHI, S. & KADOOKA, Y. 2013. Anti-obesity effect of Lactobacillus gasseri SBT2055 accompanied by inhibition of pro-inflammatory gene expression in the visceral adipose tissue in diet-induced obese mice. *Eur J Nutr*.

MORONI, O., KHEADR, E., BOUTIN, Y., LACROIX, C. & FLISS, I. 2006. Inactivation of Adhesion and Invasion of Food-Borne Listeria monocytogenes by Bacteriocin-Producing Bifidobacterium Strains of Human Origin. *Applied and Environmental Microbiology,* 72**,** 6894-6901.

MULLANE, K., LEE, C., BRESSLER, A., BUITRAGO, M., WEISS, K., DABOVIC, K., PRAESTGAARD, J., LEEDS, J. A., BLAIS, J. & PERTEL, P. 2014. A multi-center, randomized clinical trial to compare the safety and efficacy of LFF571 and vancomycin for Clostridium difficile infections. *Antimicrobial Agents and Chemotherapy*.

MURPHY, E. F., COTTER, P. D., HEALY, S., MARQUES, T. M., O'SULLIVAN, O., FOUHY, F., CLARKE, S. F., O'TOOLE, P. W., QUIGLEY, E. M., STANTON, C., ROSS, P. R., O'DOHERTY, R. M. & SHANAHAN, F. 2010. Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut,* 59**,** 1635-1642.

MURPHY, E. F., COTTER, P. D., HOGAN, A., O'SULLIVAN, O., JOYCE, A., FOUHY, F., CLARKE, S. F., MARQUES, T. M., O'TOOLE, P. W., STANTON, C., QUIGLEY, E. M. M., DALY, C., ROSS, P. R., O'DOHERTY, R. M. & SHANAHAN, F. 2013. Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. *Gut,* 62**,** 220-226.

MURPHY, K., O'SULLIVAN, O., REA, M. C., COTTER, P. D., ROSS, R. P. & HILL, C. 2011. Genome Mining for Radical SAM Protein Determinants Reveals Multiple Sactibiotic-Like Gene Clusters. *PLoS ONE,* 6**,** e20852.

MUSSO, G., GAMBINO, R. & CASSADER, M. 2011. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu Rev Med,* 62**,** 361-80.

O'HARA, A. M. & SHANAHAN, F. 2006. The gut flora as a forgotten organ. *EMBO Rep,* 7**,** 688-93.

O'SHEA, E. F., GARDINER, G. E., O'CONNOR, P. M., MILLS, S., ROSS, R. P. & HILL, C. 2009. Characterization of enterocin- and salivaricin-producing lactic acid bacteria from the mammalian gastrointestinal tract. *FEMS Microbiology Letters,* 291**,** 24-34.

O'SHEA, E. F., O'CONNOR, P. M., O'SULLIVAN, O., COTTER, P. D., ROSS, R. P. & HILL, C. 2013. Bactofencin A, a New Type of Cationic Bacteriocin with Unusual Immunity. *mBio,* 4.

O'KEEFE, S. J. D., LI, J. V., LAHTI, L., OU, J., CARBONERO, F., MOHAMMED, K., POSMA, J. M., KINROSS, J., WAHL, E., RUDER, E., VIPPERLA, K., NAIDOO, V., MTSHALI, L., TIMS, S., PUYLAERT, P. G. B., DELANY, J., KRASINSKAS, A., BENEFIEL, A. C., KASEB, H. O., NEWTON, K.,

NICHOLSON, J. K., DE VOS, W. M., GASKINS, H. R. & ZOETENDAL, E. G. 2015. Fat, fibre and cancer risk in African Americans and rural Africans. *Nature Communications,* 6**,** 6342.

O'TOOLE, P. W., MARCHESI, J. R. & HILL, C. 2017. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nature Microbiology,* 2**,** 17057.

OHMAN, L. & SIMREN, M. 2013. Intestinal microbiota and its role in irritable bowel syndrome (IBS). *Curr Gastroenterol Rep,* 15**,** 323.

PALMER, C., BIK, E. M., DIGIULIO, D. B., RELMAN, D. A. & BROWN, P. O. 2007. Development of the human infant intestinal microbiota. *PLoS Biol,* 5**,** e177.

PARK, D. Y., AHN, Y. T., PARK, S. H., HUH, C. S., YOO, S. R., YU, R., SUNG, M. K., MCGREGOR, R. A. & CHOI, M. S. 2013. Supplementation of Lactobacillus curvatus HY7601 and Lactobacillus plantarum KY1032 in diet-induced obese mice is associated with gut microbial changes and reduction in obesity. *PLoS One,* 8**,** e59470.

PATTERSON, E., MARQUES, T. M., O'SULLIVAN, O., FITZGERALD, P., FITZGERALD, G. F., COTTER, P. D., DINAN, T. G., CRYAN, J. F., STANTON, C. & ROSS, R. P. 2015. Streptozotocin-induced type-1-diabetes disease onset in Sprague–Dawley rats is associated with an altered intestinal microbiota composition and decreased diversity. *Microbiology,* 161**,** 182-193.

PEREZ-COBAS, A. E., ARTACHO, A., KNECHT, H., FERRUS, M. L., FRIEDRICHS, A., OTT, S. J., MOYA, A., LATORRE, A. & GOSALBES, M. J. 2013. Differential effects of antibiotic therapy on the structure and function of human gut microbiota. *PLoS One,* 8**,** e80201.

PETROF, E. O., GLOOR, G. B., VANNER, S. J., WEESE, S. J., CARTER, D., DAIGNEAULT, M. C., BROWN, E. M., SCHROETER, K. & ALLEN-VERCOE, E. 2013. Stool substitute transplant therapy for the eradication of Clostridium difficile infection:'RePOOPulating'the gut. *Microbiome,* 1**,** 1-12.

PIMENTEL, M., LEMBO, A., CHEY, W. D., ZAKKO, S., RINGEL, Y., YU, J., MAREYA, S. M., SHAW, A. L., BORTEY, E. & FORBES, W. P. 2011. Rifaximin therapy for patients with irritable bowel syndrome without constipation. *N Engl J Med,* 364**,** 22-32.

PINEIRO, M. & STANTON, C. 2007. Probiotic Bacteria: Legislative Framework—Requirements to Evidence Basis. *The Journal of Nutrition,* 137**,** 850S-853S.

PLOTTEL, C. S. & BLASER, M. J. 2011. Microbiome and malignancy. *Cell Host Microbe,* 10**,** 324-35.

PLOVIER, H., EVERARD, A., DRUART, C., DEPOMMIER, C., VAN HUL, M., GEURTS, L., CHILLOUX, J., OTTMAN, N., DUPARC, T., LICHTENSTEIN, L., MYRIDAKIS, A., DELZENNE, N. M., KLIEVINK, J., BHATTACHARJEE, A., VAN DER ARK, K. C. H., AALVINK, S., MARTINEZ, L. O., DUMAS, M.-E., MAITER, D., LOUMAYE, A., HERMANS, M. P., THISSEN, J.-P., BELZER, C., DE VOS, W. M. & CANI, P. D. 2017. A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat Med,* 23**,** 107-113.

QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., MANICHANH, C., NIELSEN, T., PONS, N., LEVENEZ, F., YAMADA, T., MENDE, D. R., LI, J., XU, J., LI, S., LI, D., CAO, J., WANG, B., LIANG, H., ZHENG, H., XIE, Y., TAP, J., LEPAGE, P., BERTALAN, M., BATTO, J.-M., HANSEN,

T., LE PASLIER, D., LINNEBERG, A., NIELSEN, H. B., PELLETIER, E., RENAULT, P., SICHERITZ-PONTEN, T., TURNER, K., ZHU, H., YU, C., LI, S., JIAN, M., ZHOU, Y., LI, Y., ZHANG, X., LI, S., QIN, N., YANG, H., WANG, J., BRUNAK, S., DORE, J., GUARNER, F., KRISTIANSEN, K., PEDERSEN, O., PARKHILL, J., WEISSENBACH, J., BORK, P., EHRLICH, S. D. & WANG, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464**,** 59-65.

QIN, J., LI, Y., CAI, Z., LI, S., ZHU, J., ZHANG, F., LIANG, S., ZHANG, W., GUAN, Y., SHEN, D., PENG, Y., ZHANG, D., JIE, Z., WU, W., QIN, Y., XUE, W., LI, J., HAN, L., LU, D., WU, P., DAI, Y., SUN, X., LI, Z., TANG, A., ZHONG, S., LI, X., CHEN, W., XU, R., WANG, M., FENG, Q., GONG, M., YU, J., ZHANG, Y., ZHANG, M., HANSEN, T., SANCHEZ, G., RAES, J., FALONY, G., OKUDA, S., ALMEIDA, M., LECHATELIER, E., RENAULT, P., PONS, N., BATTO, J. M., ZHANG, Z., CHEN, H., YANG, R., ZHENG, W., LI, S., YANG, H., WANG, J., EHRLICH, S. D., NIELSEN, R., PEDERSEN, O., KRISTIANSEN, K. & WANG, J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature,* 490**,** 55-60.

RAJILIC-STOJANOVIC, M., BIAGI, E., HEILIG, H. G., KAJANDER, K., KEKKONEN, R. A., TIMS, S. & DE VOS, W. M. 2011. Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology,* 141**,** 1792-801.

RAMNANI, P., GAUDIER, E., BINGHAM, M., VAN BRUGGEN, P., TUOHY, K. M. & GIBSON, G. R. 2010. Prebiotic effect of fruit and vegetable shots

containing Jerusalem artichoke inulin: a human intervention study. *Br J Nutr,* 104**,** 233-240.

REA, M. C., DOBSON, A., O'SULLIVAN, O., CRISPIE, F., FOUHY, F., COTTER, P. D., SHANAHAN, F., KIELY, B., HILL, C. & ROSS, R. P. 2011. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences,* 108**,** 4639-4644.

REA, M. C., SIT, C. S., CLAYTON, E., O'CONNOR, P. M., WHITTAL, R. M., ZHENG, J., VEDERAS, J. C., ROSS, R. P. & HILL, C. 2010. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *Proceedings of the National Academy of Sciences,* 107**,** 9352-9357.

RESPONDEK, F., SWANSON, K. S., BELSITO, K. R., VESTER, B. M., WAGNER, A., ISTASSE, L. & DIEZ, M. 2008. Short-chain fructooligosaccharides influence insulin sensitivity and gene expression of fat tissue in obese dogs. *The Journal of nutrition,* 138**,** 1712-1718.

RIBOULET-BISSON, E., STURME, M. H. J., JEFFERY, I. B., O'DONNELL, M. M., NEVILLE, B. A., FORDE, B. M., CLAESSON, M. J., HARRIS, H., GARDINER, G. E., CASEY, P. G., LAWLOR, P. G., O'TOOLE, P. W. & ROSS, R. P. 2012. Effect of Lactobacillus salivarius Bacteriocin Abp118 on the Mouse and Pig Intestinal Microbiota. *PLoS ONE,* 7**,** e31113.

RIDAURA, V. K., FAITH, J. J., REY, F. E., CHENG, J., DUNCAN, A. E., KAU, A. L., GRIFFIN, N. W., LOMBARD, V., HENRISSAT, B., BAIN, J. R., MUEHLBAUER, M. J., ILKAYEVA, O., SEMENKOVICH, C. F., FUNAI, K., HAYASHI, D. K., LYLE, B. J., MARTINI, M. C., URSELL, L. K.,

CLEMENTE, J. C., VAN TREUREN, W., WALTERS, W. A., KNIGHT, R., NEWGARD, C. B., HEATH, A. C. & GORDON, J. I. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science,* 341**,** 1241214.

RIJKERS, G. T., BENGMARK, S., ENCK, P., HALLER, D., HERZ, U., KALLIOMAKI, M., KUDO, S., LENOIR-WIJNKOOP, I., MERCENIER, A., MYLLYLUOMA, E., RABOT, S., RAFTER, J., SZAJEWSKA, H., WATZL, B., WELLS, J., WOLVERS, D. & ANTOINE, J. M. 2010. Guidance for substantiating the evidence for beneficial effects of probiotics: current status and recommendations for future research. *J Nutr,* 140**,** 671s-6s.

RIORDAN, S. M. & KIM, R. 2006. Bacterial overgrowth as a cause of irritable bowel syndrome. *Curr Opin Gastroenterol,* 22**,** 669-73.

ROBERFROID, M., GIBSON, G. R., HOYLES, L., MCCARTNEY, A. L., RASTALL, R., ROWLAND, I., WOLVERS, D., WATZL, B., SZAJEWSKA, H., STAHL, B., GUARNER, F., RESPONDEK, F., WHELAN, K., COXAM, V., DAVICCO, M. J., LEOTOING, L., WITTRANT, Y., DELZENNE, N. M., CANI, P. D., NEYRINCK, A. M. & MEHEUST, A. 2010. Prebiotic effects: metabolic and health benefits. *Br J Nutr,* 104 Suppl 2**,** S1-63.

SAARELA, M., MOGENSEN, G., FONDÉN, R., MÄTTÖ, J. & MATTILA-SANDHOLM, T. 2000. Probiotic bacteria: safety, functional and technological properties. *Journal of Biotechnology,* 84**,** 197-215.

SCOTT, L. J., MOHLKE, K. L., BONNYCASTLE, L. L., WILLER, C. J., LI, Y., DUREN, W. L., ERDOS, M. R., STRINGHAM, H. M., CHINES, P. S., JACKSON, A. U., PROKUNINA-OLSSON, L., DING, C. J., SWIFT, A. J., NARISU, N., HU, T., PRUIM, R., XIAO, R., LI, X. Y., CONNEELY, K. N.,

RIEBOW, N. L., SPRAU, A. G., TONG, M., WHITE, P. P., HETRICK, K. N., BARNHART, M. W., BARK, C. W., GOLDSTEIN, J. L., WATKINS, L., XIANG, F., SARAMIES, J., BUCHANAN, T. A., WATANABE, R. M., VALLE, T. T., KINNUNEN, L., ABECASIS, G. R., PUGH, E. W., DOHENY, K. F., BERGMAN, R. N., TUOMILEHTO, J., COLLINS, F. S. & BOEHNKE, M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science,* 316**,** 1341-5.

SEKIROV, I., RUSSELL, S. L., ANTUNES, L. C. M. & FINLAY, B. B. 2010. Gut Microbiota in Health and Disease. *Physiological Reviews,* 90**,** 859-904.

SELLON, R. K., TONKONOGY, S., SCHULTZ, M., DIELEMAN, L. A., GRENTHER, W., BALISH, E., RENNICK, D. M. & SARTOR, R. B. 1998. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection and immunity,* 66**,** 5224-5231.

SENDER, R., FUCHS, S. & MILO, R. 2016. Revised estimates for the number of human and bacteria cells in the body. *bioRxiv*.

SINGH, M. & SAREEN, D. 2014. Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One,* 9**,** e91352.

SLAVIN, J. 2013. Fiber and Prebiotics: Mechanisms and Health Benefits. *Nutrients,* 5**,** 1417-1435.

SOKOL, H. & SEKSIK, P. 2010. The intestinal microbiota in inflammatory bowel diseases: time to connect with the host. *Curr Opin Gastroenterol,* 26**,** 327-31.

SPEHLMANN, M. E., BEGUN, A. Z., BURGHARDT, J., LEPAGE, P., RAEDLER, A. & SCHREIBER, S. 2008. Epidemiology of inflammatory bowel disease in

a German twin cohort: results of a nationwide study. *Inflammatory Bowel Diseases,* 14**,** 968-976.

SPRONG, R., SCHONEWILLE, A. & VAN DER MEER, R. 2010. Dietary cheese whey protein protects rats against mild dextran sulfate sodium–induced colitis: Role of mucin and microbiota. *Journal of dairy science,* 93**,** 1364-1371.

TAHARA, T., YAMAMOTO, E., SUZUKI, H., MARUYAMA, R., CHUNG, W., GARRIGA, J., JELINEK, J., YAMANO, H. O., SUGAI, T., AN, B., SHUREIQI, I., TOYOTA, M., KONDO, Y., ESTECIO, M. R. & ISSA, J. P. 2014. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res*.

THE HUMAN MICROBIOME PROJECT CONSORTIUM 2012. Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-14.

TURNBAUGH, P. J., BÄCKHED, F., FULTON, L. & GORDON, J. I. 2008. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe,* 3**,** 213-223.

TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. & GORDON, J. I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* 444**,** 1027-31.

TURNBAUGH, P. J., RIDAURA, V. K., FAITH, J. J., REY, F. E., KNIGHT, R. & GORDON, J. I. 2009. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine,* 1**,** 6ra14.

TVEDE, M. & RASK-MADSEN, J. 1989. Bacteriotherapy for chronic relapsing Clostridium difficile diarrhoea in six patients. *Lancet,* 1**,** 1156-60.

VAN HEEL, A. J., DE JONG, A., MONTALBAN-LOPEZ, M., KOK, J. & KUIPERS, O. P. 2013. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res,* 41**,** W448-53.

VAN KRAAIJ, C., M. DE VOS, W., J. SIEZEN, R. & P. KUIPERS, O. 1999. Lantibiotics: biosynthesis, mode of action and applications. *Natural Product Reports,* 16**,** 575-587.

VEERAPPAN, G. R., BETTERIDGE, J. & YOUNG, P. E. 2012. Probiotics for the treatment of inflammatory bowel disease. *Curr Gastroenterol Rep,* 14**,** 324-33.

VIEIRA, A. T., TEIXEIRA, M. M. & MARTINS, F. S. 2013. The Role of Probiotics and Prebiotics in Inducing Gut Immunity. *Front Immunol,* 4**,** 445.

VYAS, U. & RANGANATHAN, N. 2012. Probiotics, prebiotics, and synbiotics: gut and beyond. *Gastroenterol Res Pract,* 2012**,** 872716.

WALIGORA-DUPRIET, A.-J., CAMPEOTTO, F., NICOLIS, I., BONET, A., SOULAINES, P., DUPONT, C. & BUTEL, M.-J. 2007. Effect of oligofructose supplementation on gut microflora and well-being in young children attending a day care centre. *International journal of food microbiology,* 113**,** 108-113.

WALKER, A. W., DUNCAN, S. H., LEITCH, E. C. M., CHILD, M. W. & FLINT, H. J. 2005. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Applied and environmental microbiology,* 71**,** 3692-3700.

WALSH, C. J., GUINANE, C. M., HILL, C., ROSS, R. P., O'TOOLE, P. W. & COTTER, P. D. 2015. In silico identification of bacteriocin gene clusters in

the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. *BMC Microbiology,* 15**,** 183.

WALSH, C. J., GUINANE, C. M., O'TOOLE, P. W. & COTTER, P. D. 2014. Beneficial modulation of the gut microbiota. *FEBS Lett,* 588**,** 4120-30.

WALSH, C. J., GUINANE, C. M., PW, O. T. & COTTER, P. D. 2017. A Profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome. *PeerJ,* 5**,** e3254.

WALSH, M. C., GARDINER, G. E., HART, O. M., LAWLOR, P. G., DALY, M., LYNCH, B., RICHERT, B. T., RADCLIFFE, S., GIBLIN, L. & HILL, C. 2008. Predominance of a bacteriocin-producing Lactobacillus salivarius component of a five-strain probiotic in the porcine ileum and effects on host immune phenotype. *FEMS microbiology ecology,* 64**,** 317-327.

WANG, T., CAI, G., QIU, Y., FEI, N., ZHANG, M., PANG, X., JIA, W., CAI, S. & ZHAO, L. 2012. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *Isme j,* 6**,** 320-9.

WARREN, C. A. & GUERRANT, R. L. 2011. Pathogenic C difficile is here (and everywhere) to stay. *The Lancet,* 377**,** 8-9.

WEBER, T., BLIN, K., DUDDELA, S., KRUG, D., KIM, H. U., BRUCCOLERI, R., LEE, S. Y., FISCHBACH, M. A., MÜLLER, R., WOHLLEBEN, W., BREITLING, R., TAKANO, E. & MEDEMA, M. H. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research,* 43**,** W237-W243.

WELLCOME TRUST CASE CONTROL CONSORTIUM 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661-78.

WU, G. D., CHEN, J., HOFFMANN, C., BITTINGER, K., CHEN, Y.-Y., KEILBAUGH, S. A., BEWTRA, M., KNIGHTS, D., WALTERS, W. A., KNIGHT, R., SINHA, R., GILROY, E., GUPTA, K., BALDASSANO, R., NESSEL, L., LI, H., BUSHMAN, F. D. & LEWIS, J. D. 2011. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science,* 334**,** 105-108.

YADAV, H., LEE, J. H., LLOYD, J., WALTER, P. & RANE, S. G. 2013. Beneficial metabolic effects of a probiotic via butyrate-induced GLP-1 hormone secretion. *J Biol Chem,* 288**,** 25088-97.

ZEEVI, D., KOREM, T., ZMORA, N., ISRAELI, D., ROTHSCHILD, D., WEINBERGER, A., BEN-YACOV, O., LADOR, D., AVNIT-SAGI, T., LOTAN-POMPAN, M., SUEZ, J., MAHDI, J. A., MATOT, E., MALKA, G., KOSOWER, N., REIN, M., ZILBERMAN-SCHAPIRA, G., DOHNALOVA, L., PEVSNER-FISCHER, M., BIKOVSKY, R., HALPERN, Z., ELINAV, E. & SEGAL, E. 2015. Personalized Nutrition by Prediction of Glycemic Responses. *Cell,* 163**,** 1079-94.

ZHANG, C., LI, S., YANG, L., HUANG, P., LI, W., WANG, S., ZHAO, G., ZHANG, M., PANG, X., YAN, Z., LIU, Y. & ZHAO, L. 2013. Structural modulation of gut microbiota in life-long calorie-restricted mice. *Nat Commun,* 4.

ZHAO, X. & KUIPERS, O. P. 2016. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species. *BMC Genomics,* 17**,** 882.

ZHENG, J., GÄNZLE, M. G., LIN, X. B., RUAN, L. & SUN, M. 2014. Diversity and dynamics of bacteriocins from human microbiome. *Environmental Microbiology*, 2133–2143.

Chapter 2

In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database

Included as published in BMC Microbiology (doi: 10.1186/s12866-015-0515-4)

Abstract

The human gut microbiota comprises approximately 100 trillion microbial cells which significantly impact many aspects of human physiology - including metabolism, nutrient absorption and immune function. Disturbances in this population have been implicated in many conditions and diseases, including obesity, type-2 diabetes and inflammatory bowel disease. This suggests that targeted manipulation or shaping of the gut microbiota has potential as a therapeutic tool for the prevention or treatment of these conditions. The gut is a rich source of antimicrobial/bacteriocin-producers with the potential to alter the intestinal communities in a beneficial way for human health. With this in mind, several studies have used traditional culture-dependent approaches to successfully identify bacteriocin-producers from the mammalian gut. *In silico*-based approaches to identify novel gene clusters are now also being utilised to take advantage of the vast amount of data currently being generated by next generation sequencing technologies. In this study, we employed an *in silico* screening approach to mine potential bacteriocin clusters in genome-sequenced isolates from the gastrointestinal tract (GIT). More specifically, the bacteriocin genome-mining tool BAGEL3 was used to identify potential bacteriocin producers in the genomes of the GIT subset of the Human Microbiome Project's reference genome database. Each of the identified gene clusters were manually annotated and potential bacteriocin-associated genes were evaluated.

We identified 74 clusters of note from 59 unique members of the Firmicutes, Bacteroidetes, Actinobacteria, Fusobacteria and Synergistetes. The most commonly identified class of bacteriocin was the >10 kDa class, formerly known as bacteriolysins, followed by lantibiotics and sactipeptides.

Multiple bacteriocin gene clusters were identified in a dataset representative of the human gut microbiota. Interestingly, many of these were associated with species and genera which are not typically associated with bacteriocin production.

Introduction

Bacteriocins are ribosomally synthesized antimicrobial peptides produced by bacteria that are active against other bacteria, either within the same species (narrow spectrum) or across genera (broad spectrum), and to which the producing organism is immune by a specific immunity protein(s) (Cotter et al., 2005a). Some bacteriocins, most notably nisin, have a long history of use as preservatives in the food industry (Deegan et al., 2006) and these antimicrobials are also receiving increased attention as potential alternatives to antibiotics (Piper et al., 2009).

The intestinal microbiota comprises a dynamic community with 100-1000 phylotypes (Qin et al., 2010, Faith et al., 2013) playing an integral role in gastrointestinal (GI) health and disease (Clemente et al., 2012, Flint et al., 2012). As a consequence of advances in DNA sequencing technologies, there is now a clearer understanding of the composition of the GI microbiota and of associations between specific taxa with health and disease (Clemente et al., 2012, Karlsson et al., 2013). This knowledge can potentially be utilised through the modulation of the gut microbiota to address certain GI disorders (Kadooka et al., 2010, Xiao et al., 2014). Bacteriocins are ideal candidates with respect to the targeting of undesirable populations due to their generally low toxicity, high potency and, particularly in the case of gut-associated isolates, the possibility of *in situ* production (Cotter et al., 2013). There have been some notable proof of concept studies, such as the use of a representative of the sactibiotic group of bacteriocins, thuricin CD, to specifically

inhibit *Clostridium difficile* in a distal colon model, without significantly impacting on other members of the microbiota (Rea et al., 2011). Similarly, bacteriocin production by the probiotic *Lactobacillus salivarius* UCC118 was shown to be directly responsible for significantly protecting mice against *Listeria monocytogenes* infection (Corr et al., 2007). Bacteriocin production has also been investigated to assess the extent to which it can control weight gain as a consequence of changing the composition of the gut microbiota (Murphy et al., 2013, Riboulet-Bisson et al., 2012). There are a variety of strategies by which novel bacteriocin producers can be identified (Marsh et al., 2010). These can be broadly divided into traditional, culture-based approaches and newer, *in silico*-based, strategies. The latter take advantage of the vast amount of data generated by genome and metagenome sequencing projects and the fact that many features of bacteriocin gene clusters, and especially bacteriocin modification genes, are highly conserved. These modification genes encode enzymes responsible for the post-translational modification of Class 1 bacteriocins into their active forms. Other features common to bacteriocin gene clusters include specific immunity genes, ABC transporters for bacteriocin export, and leader cleavage peptidases for removing the leader sequence from the structural prepeptide (for a review see Arnison *et al*. (Arnison et al., 2013)). To date, *in silico* bacteriocin screening strategies have led to the identification of many novel lantibiotic (Begley et al., 2009, Lawton et al., 2007, Marsh et al., 2010, Singh and Sareen, 2014, McClerren et al., 2006), microcin (Scholz et al., 2011) and sactibiotic (Murphy et al., 2011) gene clusters of interest. While in a number of instances standard BLAST-based approaches have been employed to identify such clusters, the BAGEL web-based bacteriocin mining tool (http://bagel.molgenrug.nl/) has been a particularly valuable resource (van Heel et al., 2013). BAGEL combines direct mining for the structural gene with indirect

mining for bacteriocin-associated genes. The latter is particularly useful for identifying peptides which undergo significant post-translational modification such as those observed in lantibiotics. The most recent iteration of this tool, BAGEL3 (van Heel et al., 2013), was recently used to evaluate the density and diversity of bacteriocins in the human microbiome (Zheng et al., 2014). A previous version of this software was, for example, used in the identification of the novel, two-peptide lantibiotic lichenicidin (Begley et al., 2009) and 24 putative novel lantibiotics from genomic data (Singh and Sareen, 2014). BAGEL3 classifies clusters in a manner consistent with the generally accepted approach of dividing bacteriocins on the basis of whether they are modified (class I) or unmodified/minimally modified (class II) (Cotter et al., 2005a, Cotter et al., 2013). The former can be sub-divided into a number of subclasses including the lantibiotics, sactibiotics, some microcins, bottromycins, and linear azol(in)e-containing peptides (LAPs) (Cotter et al., 2013, Arnison et al., 2013). In addition, it also identifies antimicrobial proteins larger than 10 kDa in size (i.e. bacteriolysins, previously referred to as Class III).

Among the large databases of microbiota data that can be screened using *in silico* approaches are those generated by the Human Microbiome Project (HMP). The HMP was established with the goals of characterising the human microbiome, elucidating its role in health and disease, and developing new tools and databases to aid researchers. Among the data generated by the HMP is a reference genome database, which is a collection of genome-sequences from species/strains isolated from a variety of human body sites (http://www.hmpdacc.org/). The gastrointestinal tract (GIT) subset of this reference genome database was chosen as the focus of this study, which aimed to find bacteriocin-producers with the potential to alter the composition of the gut microbiota *in situ*. Indeed, previous culture-based approaches

have shown the human gut is a rich reservoir of bacteriocin-producers (Rea et al., 2010, Lakshminarayanan et al., 2013, O'Shea et al., 2009). Here we employ the bacteriocin genome-mining tool BAGEL3 to screen the GIT subset of the HMP reference genome database and identify 74 putative bacteriocin-encoding gene clusters (PBGCs) from 59 unique producers.

Methods

Initial screening of reference genomes for bacteriocin gene clusters

The GIT subset (382 available sequences as of 20/11/2014) of the HMP's reference genome database (http://www.hmpdacc.org/HMRGD/) was downloaded in multi-FASTA format and both complete and draft genomes were screened for putative bacteriocin gene clusters using the web-version of BAGEL3 (http://bagel2.molgenrug.nl/index.php/bagel3).

Further investigation of individual gene clusters

Approximately 20 kb of sequence data containing the gene/genes identified as being of potential interest by BAGEL3 were extracted and the sequences were manually annotated using the Artemis software (Rutherford et al., 2000). Predicted coding regions were analysed using the BlastP web server on NCBI (http://www.ncbi.nlm.nih/BLAST) and the nr database. The coding regions were also analysed for the presence of conserved domains and, where applicable, compared to previously described gene clusters using the Artemis Comparison Tool (ACT) (Carver et al., 2005).

Results and Discussion

In silico screen for putative bacteriocin-encoding gene clusters

The GIT subset of the HMP reference genome database contained 382 fully sequenced genomes. The bacteriocin mining software tool BAGEL3 initially identified 217 areas of interest (AOIs) from 130 unique putative producers (Table 1). Subsequent manual annotation and Blast analysis determined that 74 of these were PBGCs (Table 2). The remaining AOIs were eliminated following manual annotation due to the absence of key bacteriocin associated genes. However, we accept the possibility that these gene products may work in concert with other novel bacteriocin-related genes encoded elsewhere on the genome. Selection of the 74 PBGCs was achieved based on the presence of bacteriocin-associated genes, arrangement of those genes in the AOI, and by overall similarity to previously described gene clusters. An overall breakdown of the 74 PBGCs according to phylum and predicted bacteriocin type can be seen in Figures 1A and 1B, respectively. The vast majority of PBGCs belonged to members of the Firmicutes and Proteobacteria phyla, and, in the latter case, *Escherichia coli* strains in particular. PBGCs were also identified in the Bacteroidetes, Actinobacteria, Fusobacteria and Synergistetes phyla. The most commonly identified clusters were > 10 kDa bacteriolysins followed by lantibiotics and sactipeptides (Figure 1).

Producer
Bifidobacterium longum subsp. *infantis* JCM 1222
Bifidobacterium sp. 12 1 47BFAA
Collinsella stercoris DSM 13279
Eggerthella sp. HGA1
Propionibacterium sp. 5 U 42AFAA
Bacteroides dorei DSM 17855
Bacteroides fragilis 3 1 12
Bacteroides sp. 2 1 16
Bacteroides sp. 2 1 56FAA
Bacteroides sp. 9 1 42FAA
Bacteroides uniformis ATCC 8492
Odoribacter laneus YIT 12061
Anaerofustis stercorihominis DSM 17244
Anaerotruncus colihominis DSM 17241
Bacillus sp. 7 6 55CFAA CT2
Blautia hansenii DSM 20583
Butyrivibrio crossotus DSM 2876
Catenibacterium mitsuokai DSM 15897
Clostridiales sp. SSC 2
Coprobacillus sp. 29 1
Coprobacillus sp. 8 2 54BFAA
Coprobacillus sp. D6
Coprococcus catus GD 7
Desulfitobacterium hafniense DP7
Dorea formicigenerans 4 6 53AFAA
Dorea longicatena DSM 13814
Enterococcus faecalis PC1.1
Enterococcus faecalis TX0104
Enterococcus faecalis TX1302
Enterococcus faecalis TX1341
Enterococcus faecalis TX1342
Enterococcus faecalis TX1346
Enterococcus faecalis TX1467
Enterococcus faecalis TX2134
Enterococcus faecalis TX2137
Enterococcus faecalis TX4244
Enterococcus faecium PC41
Enterococcus faecium TX1330
Erysipelotrichaceae bacterium 3 1 53
Erysipelotrichaceae bacterium 5 2 54FAA

Erysipelotrichaceae bacterium 6 1 45
Eubacterium hallii DSM 3353
Eubacterium siraeum DSM 15702
Flavonifractor plautii ATCC 29863
Holdemania filiformis DSM 12042
Lachnospiraceae bacterium 2 1 46FAA
Lachnospiraceae bacterium 2 1 58FAA
Lachnospiraceae bacterium 3 1 57FAA CT1
Lachnospiraceae bacterium 5 1 63FAA
Lachnospiraceae bacterium 7 1 58FAA
Lactobacillus acidophilus ATCC 4796
Lactobacillus antri DSM 16041
Lactobacillus brevis subsp. *gravesensis* ATCC 27305
Lactobacillus delbrueckii subsp. *lactis* DSM 20072
Lactobacillus helveticus DSM 20075
Lactobacillus plantarum subsp. *plantarum* ATCC 14917
Lactobacillus reuteri CF48 3A
Lactobacillus reuteri JCM 1112
Lactobacillus reuteri MM2 3
Lactobacillus reuteri MM4 1A
Lactobacillus reuteri SD2112
Lactobacillus rhamnosus ATCC 21052
Lactobacillus ultunensis DSM 16047
Listeria innocua ATCC 33091
Marvinbryantia formatexigens DSM 14469
Paenibacillus sp. HGF7 contig00140
Paenibacillus sp. HGF7 contig00161
Paenibacillus sp. HGF7 contig00230
Parvimonas micra ATCC 33270
Pediococcus acidilactici 7 4
Roseburia intestinalis L1 82
Ruminococcus obeum A2 162
Ruminococcus obeum ATCC 29174
Ruminococcus sp. 5 1 39B FAA
Ruminococcus sp. SR1 5
Streptococcus anginosus 1 2 62CV
Streptococcus infantarius subsp. *infantarius* ATCC BAA 102
Streptococcus sp. 2 1 36FAA
Fusobacterium sp. D12
Fusobacterium ulcerans ATCC 49185
Fusobacterium varium ATCC 27725

Arcobacter butzleri JV22
Campylobacter upsaliensis JV21
Citrobacter freundii 4 7 47CFAA
Citrobacter sp. 30 2
Citrobacter youngae ATCC 29220
Desulfovibrio sp. 3 1 syn3
Desulfovibrio sp. 6 1 46AFAA
Edwardsiella tarda ATCC 23685
Enterobacter cancerogenus ATCC 35316
Enterobacter cloacae subsp. *cloacae* NCTC 9394
Enterobacteriaceae bacterium 9 2 54FAA
Escherichia coli 4 1 47FAA
Escherichia coli MS 107 1
Escherichia coli MS 110 3
Escherichia coli MS 115 1
Escherichia coli MS 116 1
Escherichia coli MS 117 3
Escherichia coli MS 119 7
Escherichia coli MS 124 1
Escherichia coli MS 146 1
Escherichia coli MS 153 1
Escherichia coli MS 16 3
Escherichia coli MS 175 1
Escherichia coli MS 185 1
Escherichia coli MS 187 1
Escherichia coli MS 196 1
Escherichia coli MS 198 1
Escherichia coli MS 200 1
Escherichia coli MS 21 1
Escherichia coli MS 45 1
Escherichia coli MS 57 2
Escherichia coli MS 69 1
Escherichia coli MS 78 1
Escherichia coli MS 79 10
Escherichia coli MS 85 1
Escherichia coli SE11
Escherichia sp. 1 1 43
Escherichia sp. 3 2 53FAA
Escherichia sp. 4 1 40B
Helicobacter bilis ATCC 43879
Klebsiella sp. MS 92 3

Oxalobacter formigenes OXCC13	
Proteus penneri ATCC 35198	
Providencia alcalifaciens DSM 30120	
Providencia rettgeri DSM 1131	
Providencia rustigianii DSM 4541	
Yokenella regensburgei ATCC 43003	
Anaerobaculum hydrogeniformans ATCC BAA 1850	
Synergistes sp. 3 1 syn1	

Table 1. 130 unique putative producers identified by BAGEL3.

Potential Producer	Phylum	Class	BAGEL3 prediction
Bifidobacterium longum subsp. *infantis* JCM 1222	Actinobacteria	Lantibiotic	Context
Bifidobacterium sp. 12 1 47BFAA	Actinobacteria	Lantibiotic	BLD_1648
Eggerthella sp. HGA1	Actinobacteria	>10 kDa	Linocin M18
Bacteroides dorei DSM 17855	Bacteroidetes	Sactipeptide	Context
Bacteroides fragilis 3 1 12	Bacteroidetes	Sactipeptide	Context
Bacteroides sp. 2 1 16	Bacteroidetes	Lantibiotic	Manual
Bacteroides sp. 2 1 56FAA	Bacteroidetes	Unmodified	Manual
Bacteroides sp. 9 1 42FAA	Bacteroidetes	Sactipeptide	Context
Bacteroides uniformis ATCC 8492	Bacteroidetes	Sactipeptide	Context
Anaerofustis stercorihominis DSM 17244	Firmicutes	>10 kDa	Linocin M18
Bacillus sp. 7 6 55CFAA CT2	Firmicutes	>10 kDa	Colicin E9
Bacillus sp. 7 6 55CFAA CT2	Firmicutes	Lantibiotic	Haloduracin
Dorea formicigenerans 4 6 53AFAA	Firmicutes	Sactipeptide	Context
Enterococcus faecalis PC1.1	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX1302	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX1302	Firmicutes	Lantibiotic	Context
Enterococcus faecalis TX1341	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX1342	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX1342	Firmicutes	Lantibiotic	Context
Enterococcus faecalis TX1467	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX1467	Firmicutes	Lantibiotic	Context
Enterococcus faecalis TX2137	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX2137	Firmicutes	Lantibiotic	Context
Enterococcus faecalis TX4244	Firmicutes	>10 kDa	Enterolysin A
Enterococcus faecalis TX4244	Firmicutes	>10 kDa	Enterolysin A
Holdemania filiformis DSM 12042	Firmicutes	>10 kDa	Linocin M18
Lactobacillus acidophilus ATCC 4796	Firmicutes	>10 kDa	Enterolysin A
Lactobacillus acidophilus ATCC 4796	Firmicutes	>10 kDa	Helveticin J
Lactobacillus antri DSM 16041	Firmicutes	>10 kDa	Enterolysin A
Lactobacillus brevis subsp. *gravesensis* ATCC 27305	Firmicutes	Unmodified	Plantaricin NC8
Lactobacillus delbrueckii subsp. *lactis* DSM 20072	Firmicutes	>10 kDa	Enterolysin A
Lactobacillus helveticus DSM 20075	Firmicutes	>10 kDa	Helveticin J
Lactobacillus helveticus DSM 20075	Firmicutes	>10 kDa	Helveticin J
Lactobacillus ultunensis DSM 16047	Firmicutes	>10 kDa	Helveticin J
Lactobacillus ultunensis DSM 16047	Firmicutes	>10 kDa	Enterolysin A
Lactobacillus ultunensis DSM 16047	Firmicutes	>10 kDa	Helveticin J
Listeria innocua ATCC 33091	Firmicutes	LAP	Context
Marvinbryantia formatexigens DSM 14469	Firmicutes	Sactipeptide	Context
Roseburia intestinalis L1 82	Firmicutes	Sactipeptide	Context

Ruminococcus obeum A2 162	Firmicutes	Lantibiotic	Context
Ruminococcus sp. 5 1 39B FAA	Firmicutes	Lantibiotic	Context
Streptococcus anginosus 1 2 62CV	Firmicutes	Unmodified	Multiple
Streptococcus infantarius subsp. *infantarius* ATCC BAA 102	Firmicutes	Unmodified	Multiple
Streptococcus sp. 2 1 36FAA	Firmicutes	Class IIc	Context
Fusobacterium ulcerans ATCC 49185	Fusobacteria	>10 kDa	Linocin M18
Fusobacterium varium ATCC 27725	Fusobacteria	>10 kDa	Linocin M18
Arcobacter butzleri JV22	Proteobacteria	>10 kDa	Colicin E9
Escherichia coli MS 110 3	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 110 3	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 119 7	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 124 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 146 1	Proteobacteria	>10 kDa	Linocin M18
Escherichia coli MS 153 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 16 3	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 16 3	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 16 3	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 185 1	Proteobacteria	>10 kDa	Colicin E9
Escherichia coli MS 196 1	Proteobacteria	>10 kDa	Colicin-10
Escherichia coli MS 200 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 21 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 45 1	Proteobacteria	Microcin	Microcin H47
Escherichia coli MS 45 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 57 2	Proteobacteria	>10 kDa	Colicin E9
Escherichia coli MS 78 1	Proteobacteria	>10 kDa	Colicin
Escherichia coli MS 85 1	Proteobacteria	>10 kDa	Colicin E9
Escherichia coli SE11	Proteobacteria	>10 kDa	Colicin
Escherichia sp. 3 2 53FAA	Proteobacteria	>10 kDa	Colicin
Klebsiella sp. MS 92 3	Proteobacteria	>10 kDa	Klebicin B
Providencia rettgeri DSM 1131	Proteobacteria	>10 kDa	Colicin A
Yokenella regensburgei ATCC 43003	Proteobacteria	Microcin	Context
Anaerobaculum hydrogeniformans ATCC BAA 1850	Synergistetes	>10 kDa	Linocin M18
Synergistes sp. 3 1 syn1	Synergistetes	>10 kDa	Linocin M18

Table 2. Additional information on PBGCs and whether the initial identification of the AOI by BAGEL3 was based on the presence of bacteriocin-associated genes (context) or a specific bacteriocin structural gene.

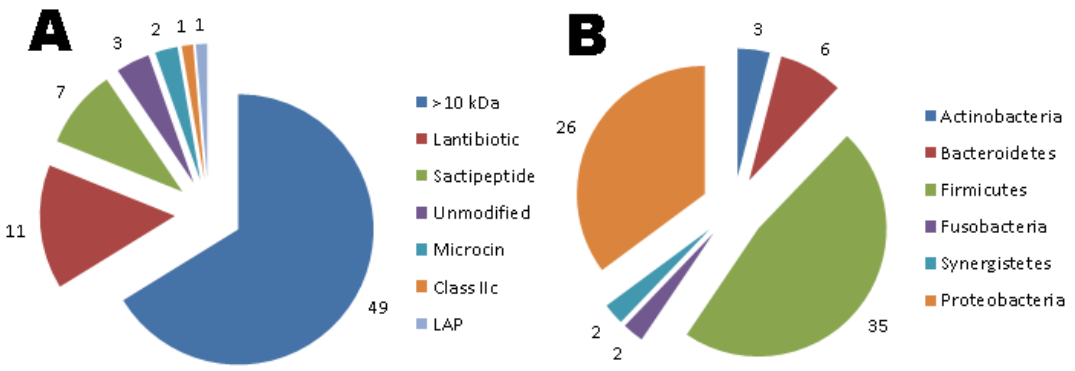


Figure 1. Frequency of (A) bacteriocin class and (B) producing phylum among the 74 PBGCs identified.

Further analysis of PBGCs of particular interest

63 PBGCs are described in the supplementary text (available online at https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-015-0515-4) and depicted in Figures 2, 3, 4. 11 PBGCs from 3 different phyla were deemed of particular interest and were selected for further *in silico* analysis based on the relative rarity with which bacteriocin production has been associated with the corresponding genus (*Bacteroides* and *Roseburia*), or on the probiotic potential of strains from the genus (*Bifidobacterium*) or due to the importance/perceived importance of the genus in a gut environment (*Bacteroides, Roseburia, Ruminococcus*) (Figure 5).

Identification of novel PBGCs in bifidobacteria

Bifidobacteria are an important group of human gut commensal bacteria, accounting for between 3 and 7% of the gut microbiota in adults and up to 91% in newborns (Cheikhyoussef et al., 2009). Members of this genus have a long history of use as health-promoting/probiotic strains due to traits such as the regulation of intestinal microbial homeostasis, the inhibition of pathogens, the modulation of local and systemic immune responses, the maintenance of gastrointestinal barrier function, the production of vitamins and the bioconversion of a number of dietary compounds into bioactive molecules (Mayo and van Sinderen, 2010). Bifidobacteria have the potential to suppress the growth of both Gram-negative and Gram-positive bacteria but, to date, this activity has been more often attributed to the inhibitory action of organic acids rather than bacteriocin production (Martinez et al., 2013, Fukuda et al., 2011). For a review of the relatively rare examples of bacteriocin production by bifidobacteria see

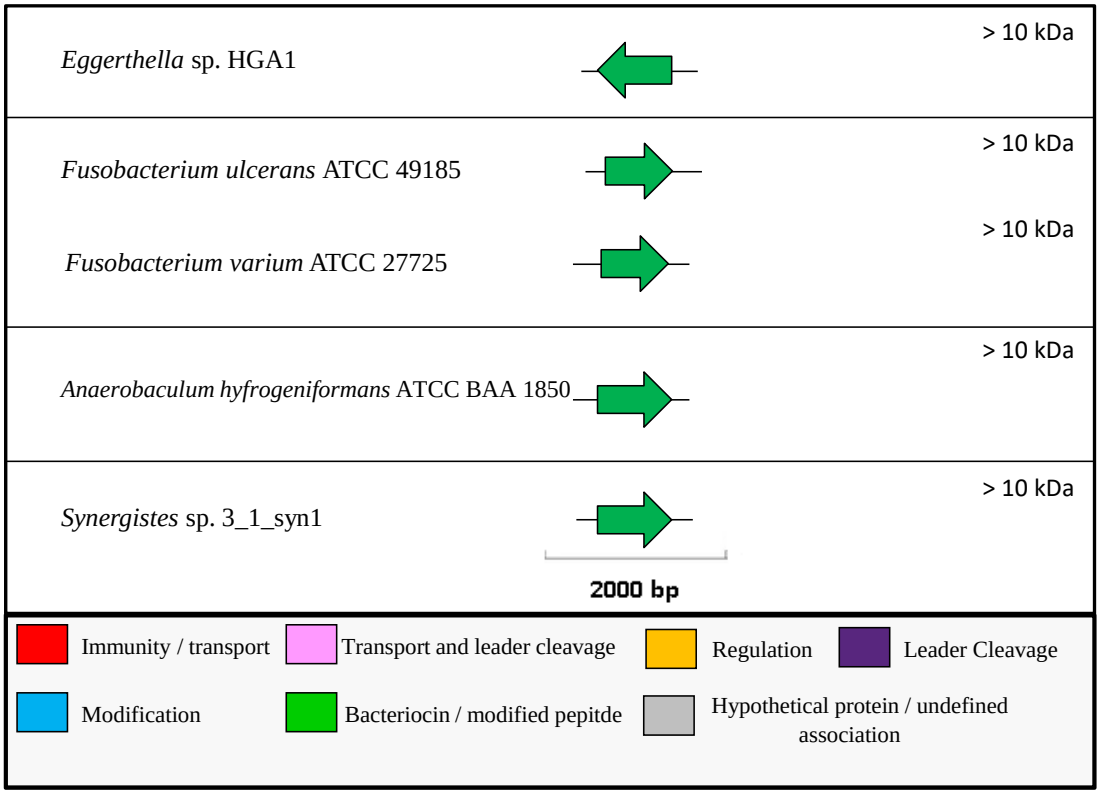
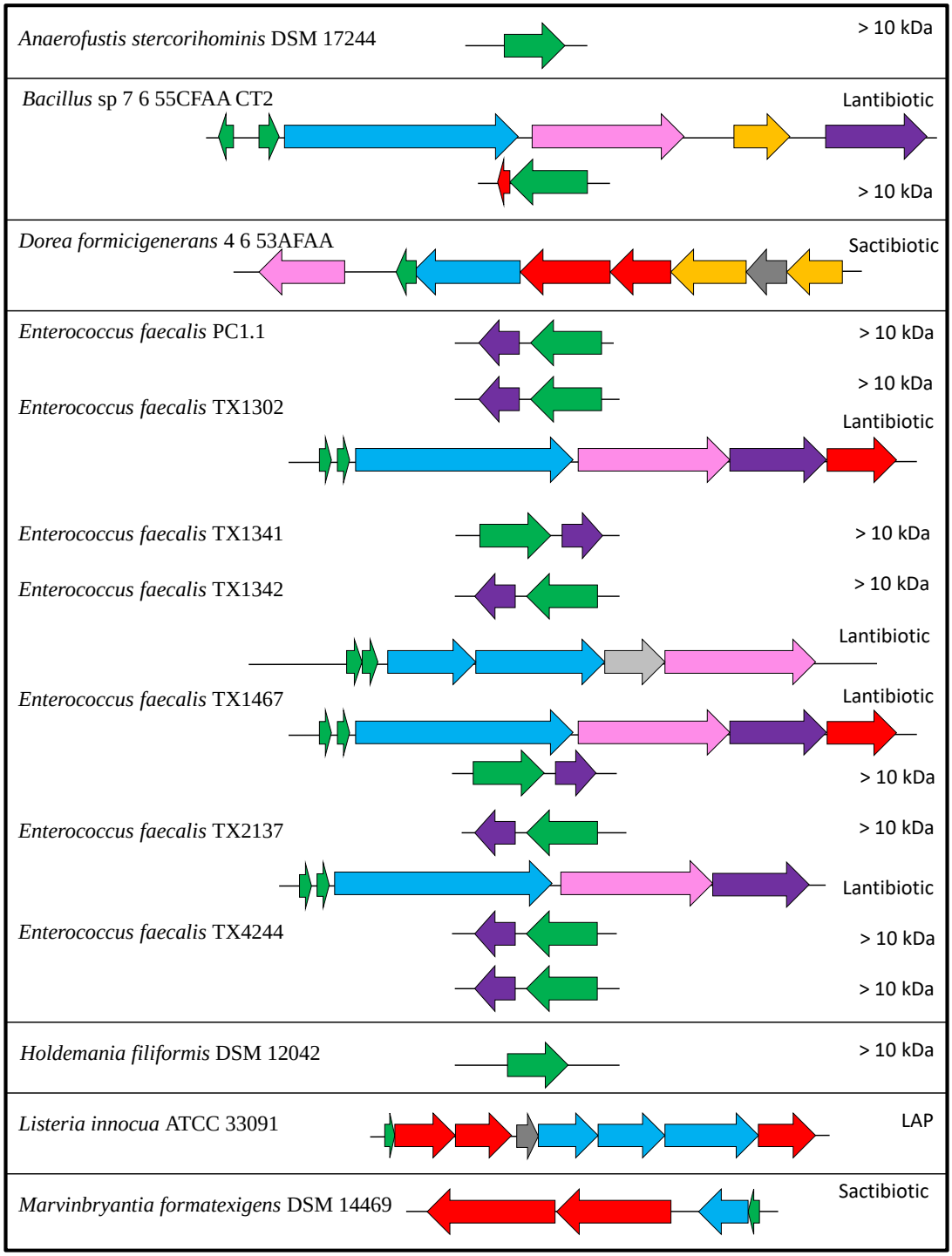


Figure 2. Diagrammatic representation of remaining PBGCs identified in the Actinobacteria, Fusobacteria and Synergistetes phyla.


Anaerofustis stercorihominis DSM 17244
> 10 kDa
Bacillus sp 7 6 55CFAA CT2
Lantibiotic
> 10 kDa
Dorea formicigenerans 4 6 53AFAA
Sactibiotic
Enterococcus faecalis PC1.1
> 10 kDa
Enterococcus faecalis TX1302
> 10 kDa
Lantibiotic
Enterococcus faecalis TX1341
> 10 kDa
Enterococcus faecalis TX1342
> 10 kDa
Lantibiotic
Enterococcus faecalis TX1467
Lantibiotic
> 10 kDa
Enterococcus faecalis TX2137
> 10 kDa
Lantibiotic
Enterococcus faecalis TX4244
> 10 kDa
> 10 kDa
Holdemania filiformis DSM 12042
> 10 kDa
Listeria innocua ATCC 33091
LAP
Marvinbryantia formatexigens DSM 14469
Sactibiotic

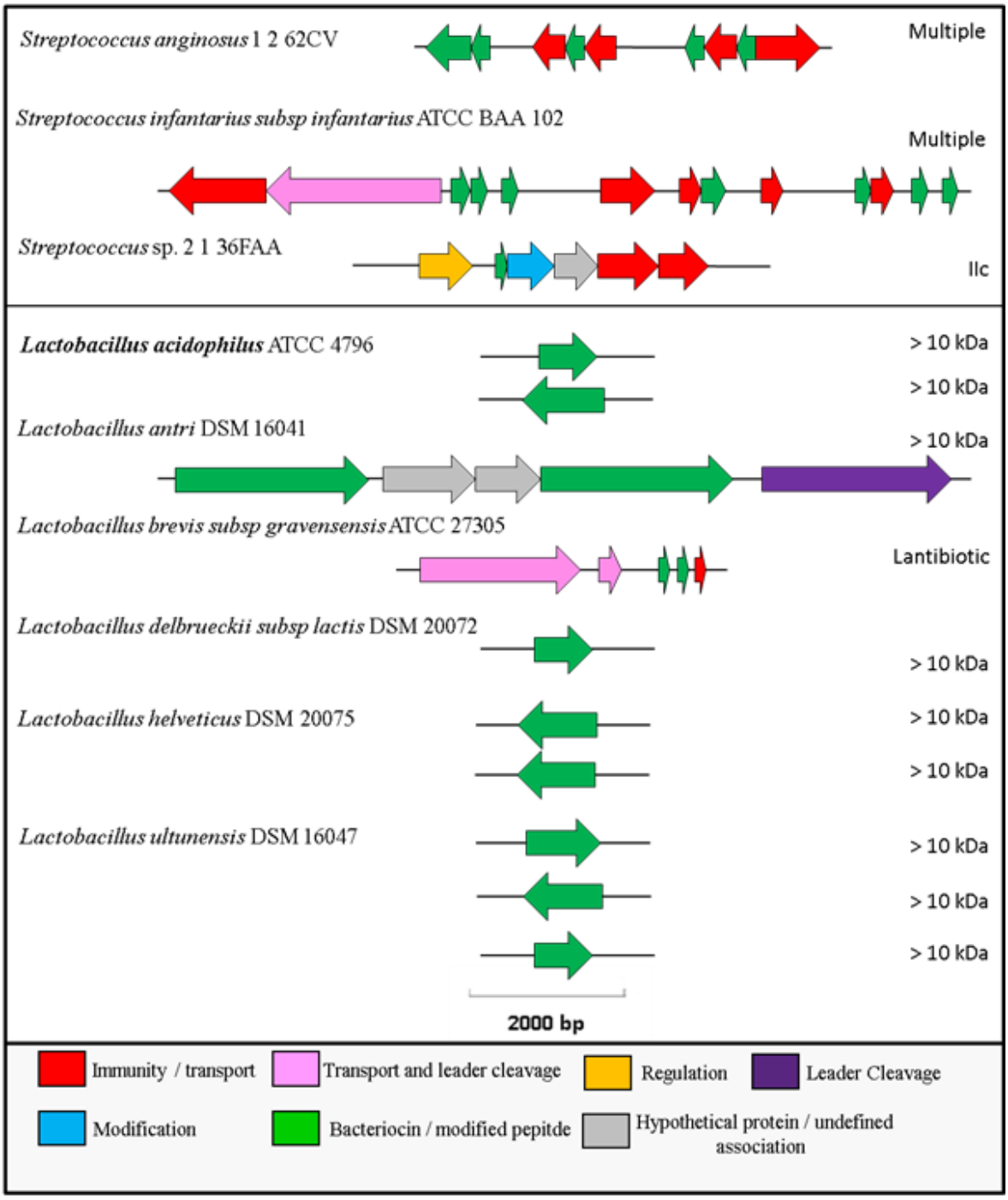
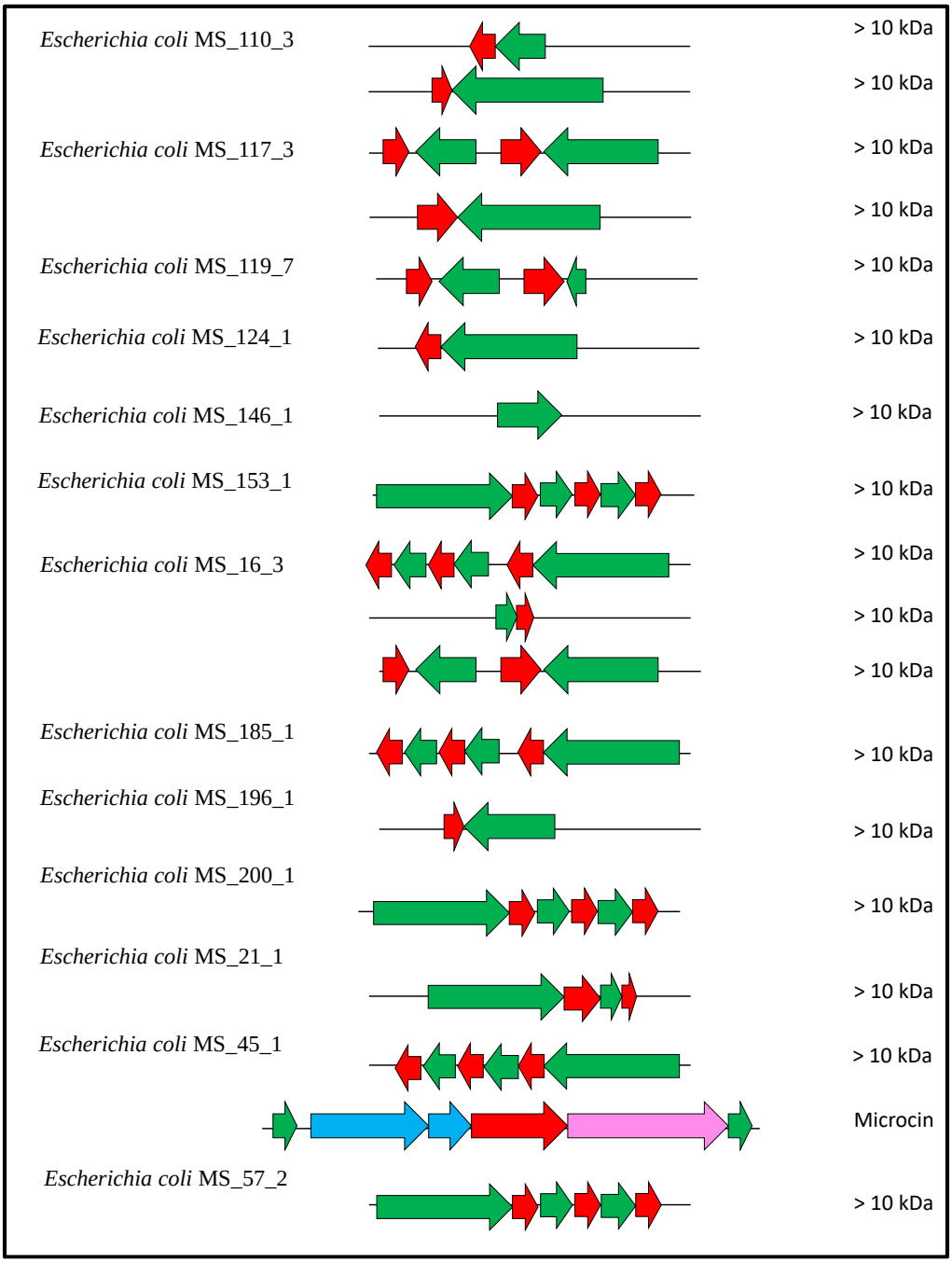


Figure 3. Diagrammatic representation of remaining PBGCs identified in the Firmicutes phylum.


Escherichia coli MS_110_3
> 10 kDa
> 10 kDa
Escherichia coli MS_117_3
> 10 kDa
> 10 kDa
Escherichia coli MS_119_7
> 10 kDa
Escherichia coli MS_124_1
> 10 kDa
Escherichia coli MS_146_1
> 10 kDa
Escherichia coli MS_153_1
> 10 kDa
Escherichia coli MS_16_3
> 10 kDa
> 10 kDa
> 10 kDa
Escherichia coli MS_185_1
> 10 kDa
Escherichia coli MS_196_1
> 10 kDa
Escherichia coli MS_200_1
> 10 kDa
Escherichia coli MS_21_1
> 10 kDa
Escherichia coli MS_45_1
> 10 kDa
Microcin
Escherichia coli MS_57_2
> 10 kDa

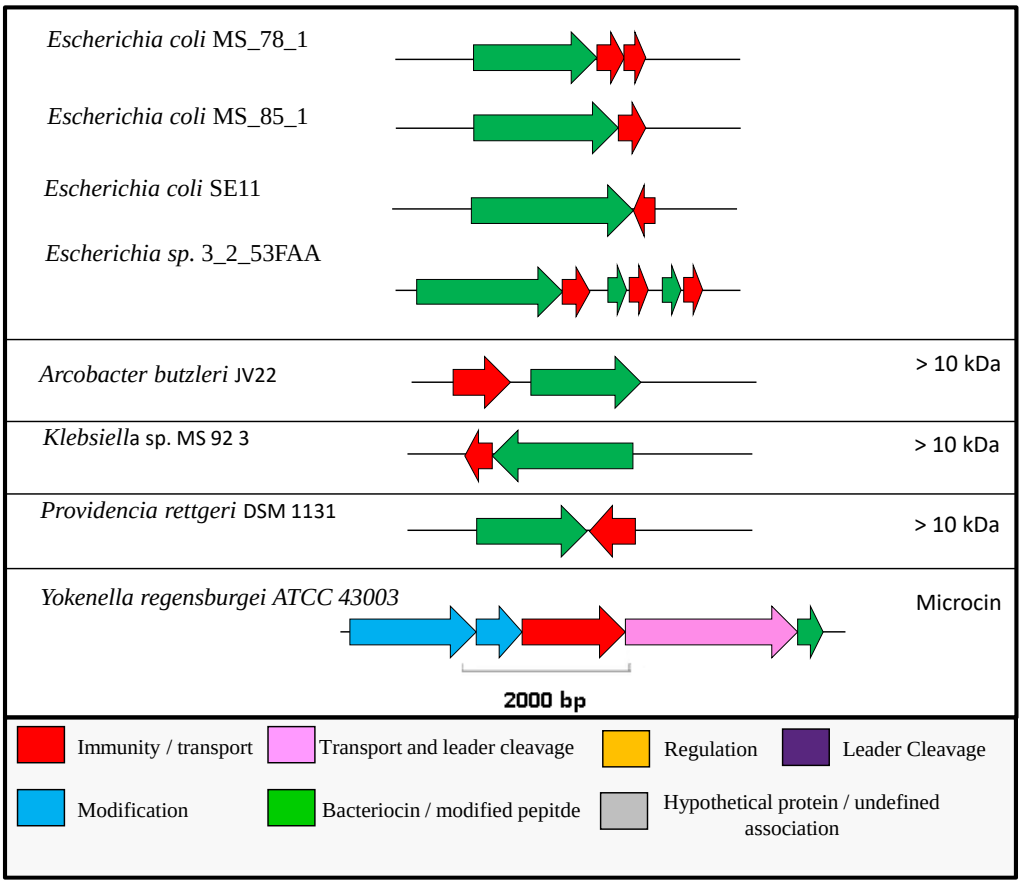


Figure 4. Diagrammatic representation of remaining PBGCs identified in the Proteobacteria phylum

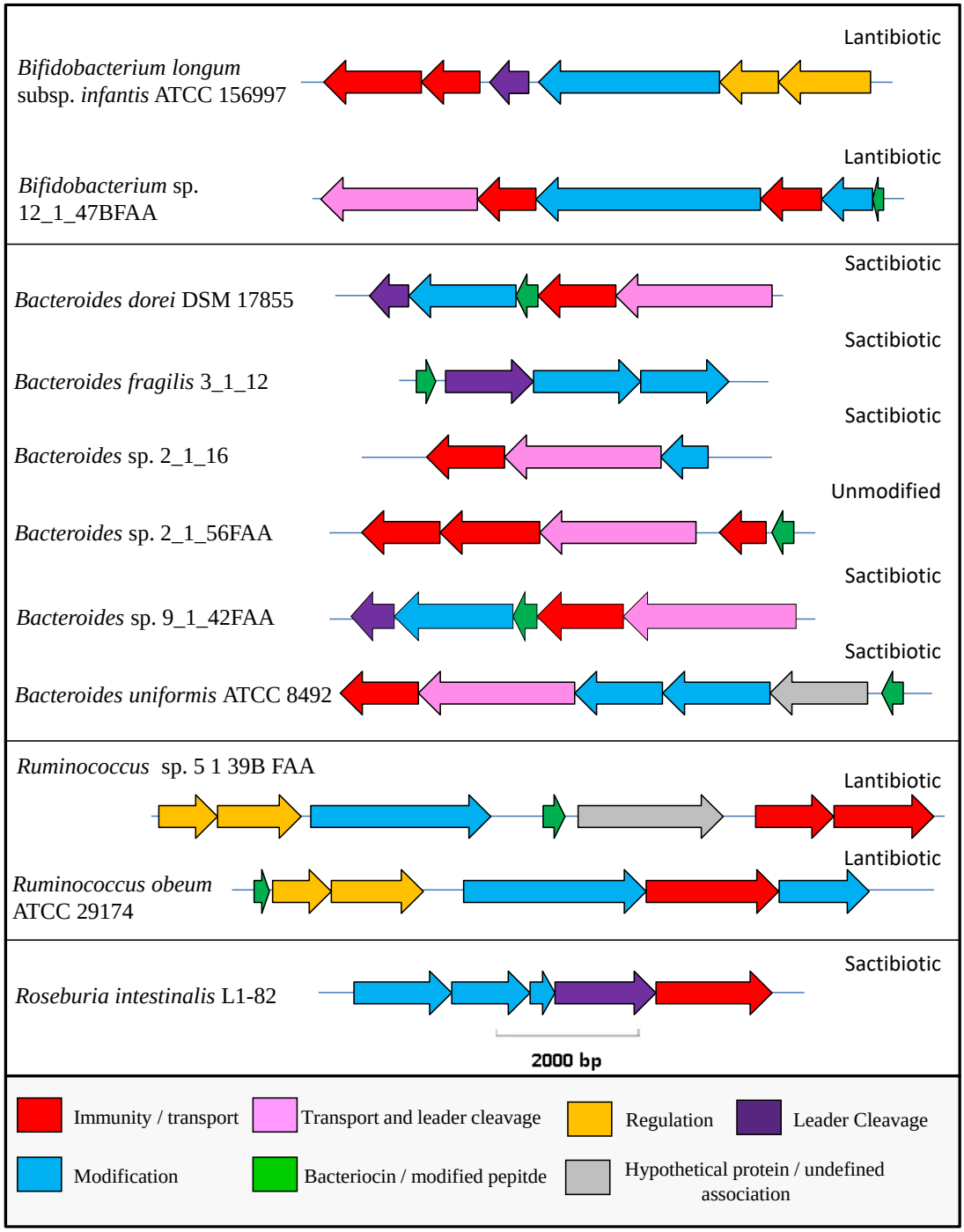


Figure 5. Diagrammatic representation of PBGCs deemed of particular interest.

Martinez *et al*. (Martinez et al., 2013). Our *in silico* screen identified PBGCs of note in *Bifidobacterium longum* subsp. *infantis* ATCC 15697 and *Bifidobacterium* sp. 12_1_47BFA (Figure 5).

Bifidobacterium longum subsp. *infantis* ATCC 15697 was isolated from human infant faeces and sequenced by the Joint Genome Institute (JGI) (Sela et al., 2008, Reuter, 1971). A previous study has shown that this strain has the ability to reduce the levels of plasma endotoxins via modulation of the gut microbiota. However the authors concluded that the effect was mediated by increased levels of faecal organic acids (Rodes et al., 2014). The cluster of six genes identified are predicted to encode a LanL-type lantipeptide based on the presence of a LanL-type lanthionine synthetase gene. More specifically, the 8,139 bp cluster contains several lantibiotic-related genes including a putative lanthionine synthetase (conserved domain pfam05147 3.10e-10), a putative oligopeptidase (conserved domain pfam00326 5.24e-08) and a putative ABC transporter containing ATP-binding and permease subunits (conserved domains cd03255 and pfam02867 respectively). The cluster also contained a two-component regulatory system consisting of a putative histidine kinase (conserved domain CGO4585 6.70e-18) and a putative transcriptional response regulator (conserved domain COG2197 8.85e-57).

Bifidobacterium sp. 12_1_47BFA was recovered from inflamed biopsy tissue from a 25-year-old female patient with Crohn's disease and its genome was found to contain a 7,996 bp lantibiotic cluster comprising six genes (Figure 5). A putative lantibiotic prepeptide LanA was found to be similar to BLD_1648 (BAGEL3 bacteriocin I database 4e-43), a feature that was further supported by manual

annotation (conserved domain TIGR03893 6.47e-9). Also present in the area of interest was a putative LanM lantibiotic biosynthesis protein (conserved domain cd04792 0.0), a putative multidrug ABC transporter ATP-binding protein putatively involved in lantibiotic immunity (conserved domain cd03230 8.53e-42) and an ABC-type bacteriocin/lantibiotic exporter (conserved domain COG2274 7.59e-145) significantly similar (BlastP 4e-117) to the *crnT* protein responsible for transport and leader cleavage of the bacteriocin carnolysin (Tulini et al., 2014). The area of interest also contained a FMN-dependent reductase (conserved domain pfam03358 5.13e-09) similar to that located within the carnolysin-associated *crnJ* protein (Tulini et al., 2014). This family of proteins has been suggested to be an atypical lantibiotic post-translational modification protein (Cotter et al., 2005b, Singh and Sareen, 2014).

Identification of novel PBGCs in Bacteroides spp.

Bacteroides are Gram-negative, non-spore-forming, obligate anaerobes and near universal constituents of the human gut microbiota, especially prevalent in those individuals whose long-term diets are rich in protein and animal fat (Wu et al., 2011). Translocation from the GIT can however result, in some cases, in bacteraemia and abscess formation (Wexler, 2007). Weight loss in obese humans subjected to dietary or surgical intervention has been associated with increased relative abundance in the phylum Bacteroidetes, with specific members including *Bacteroides* spp., *Bacteroides-Prevotella* spp. or the *Bacteroides fragilis* group bacteria having been associated with this phenomenon (Furet et al., 2010, Nadal et al., 2009, Santacruz et al., 2009, Santacruz et al., 2010). Despite their importance as a human gut commensal, there have been relatively few reports of bacteriocin production by members of the

Bacteroides to date (Avelar et al., 1999, Booth et al., 1977, Mossie et al., 1979, Nakano et al., 2006). In this study, six PBGCs were identified in *Bacteroides* strains that possessed features typical of sactipeptide (4), lantibiotic (1) or unmodified bacteriocin (1) clusters.

Bacteroides dorei has been observed to be common in patients with active coeliac disease and it has also been proposed that the species be used as an indicator of water contamination by human faecal material (Shanks et al., 2014, Sánchez et al., 2010). *B. dorei* DSM 17855 was isolated from a healthy, 23 year old, Japanese male (Bakir et al., 2006) and its genome was found to contain a five gene, 5,711 bp sactipeptide-like gene cluster (Figure 5). The cluster contained genes encoding a putative ABC-type transporter ATP-binding protein (BlastP 0.0, conserved domain COG2274 3.02e-34), a putative hemolysin secretion protein HlyD (BlastP 0.0), a structural gene belonging to pfam family pf10439 (Bacteriocin class II with double-glycine leader peptide), a radical SAM domain-containing protein hypothesised to be involved in peptide modification (conserved domain TIGR03962 1.46e-06) and a putative bacteriocin-associated C39 family peptidase (conserved domain pfam03412 1.13e-11). The latter may be involved in transport across the membrane in addition to leader cleavage, either alone or in conjunction with HlyD.

Bacteroides fragilis-produced metabolites are important in the activation and regulation of the T-cell-dependent immune response (Mazmanian et al., 2008, Wexler, 2007) and its administration as a therapeutic has been proposed for gastrointestinal and behavioural symptoms associated with human neurodevelopmental disorders (Hsiao et al., 2013). The genome of *B. fragilis* 3_1_12 found to contain a four gene,

4267 bp sactipeptide-like cluster (Figure 5). The putative structural gene belongs to pfam family PF14406 (Ribosomally synthesized peptide in Bacteroidetes) and BlastP identified it as a putative bacteriocin-type signal sequence containing a predicted leader sequence associated with peptide modification (conserved domain TIGR04149 1.34e-12). Immediately downstream is a putative lipoprotein belong to pfam family PF08139 followed by a pair of putative radical SAM proteins, predicted to be involved in peptide modification. These radical SAM proteins, members of families TIGR04085 and TIGR04150, respectively, are known to occur in cassettes together with the bacteriocin signal sequence noted above (Haft and Basu, 2011).

Bacteroides sp. 2_1_16 was isolated from a healthy biopsy of the descending colon of a 58-year old female patient undergoing colonoscopy its genome was found to contain a 4,167 bp, three-gene cluster predicted to be sactipeptide-encoding based on the presence of a SacCD homolog (Figure 5). However, manual annotation also revealed a cluster of several genes with homology with those typically associated with lantibiotic production. Specifically, the cluster contained a putative LanC-like lanthionine synthetase (conserved domain cd04793 6.02e-08), a putative ABC transporter predicted to be a bacteriocin/lantibiotic transporter based on conserved domains (COG2274 0.0) and a putative ABC transporter secretion protein closely related to hemolysin secretors (conserved domain TIGR01843 1.86e-22). However, a putative structural peptide-encoding gene could not be identified in this gene cluster.

The genome of *Bacteroides* sp. 2_1_56FAA was found to possess a 6,069 bp cluster containing five genes of note (Figure 5). Manual annotation revealed a gene predicted to encode a ribosomally synthesised peptide (pfam PF14406 0.00024 (Iyer

et al., 2009)), located immediately upstream of a putative CAAX protease self-immunity family determinant (conserved domain pfam02517 8.17e-11). A gene encoding a putative ABC transporter containing a C39B peptidase domain (COG2274 7.75e-159), predicted to be responsible for transport and leader cleavage, was also present. Two additional possible transport genes were identified immediately downstream, both putative hemolysin secretion proteins (conserved domain pfam13437 5.74e-09 and conserved domain pfam13437 5.37e-11, respectively). The lack of any bacteriocin-modification genes suggests that this cluster encodes an unmodified bacteriocin.

Bacteroides sp. 9_1_42FAA was isolated from the duodenum of a 47 year old female patient and its genome contained a 5,714 bp area of interest comprised of five genes, This cluster was identified as a potential sactipeptide based on the presence of a SacCD homolog (Figure 5). The structural peptide putatively encoded within this cluster also possesses features associated with pfam family PF10439.4 i.e. unmodified subclass IIc bacteriocins. The area of interest also contains a putative ABC-type bacteriocin/lantibiotic exporter (contains conserved domain COG2274 0.0), a putative hemolysin secretion family protein (conserved domain TIGR01843 3.45e-06), a putative radical SAM peptide modification protein (conserved domain TIGR03962 1.47e-17), and a putative bacteriocin transporter containing an endopeptidase C39 domain (potentially involved in bacteriocin preprocessing; conserved domain pfam03412 1.13e-11) (Havarstein et al., 1995). This sequence exhibited very high (99%) nucleotide identity to the aforementioned gene cluster in *B. dorei* DSM 17855. This similarity includes structural genes with 100% amino acid sequence identity.

It has been previously documented that orally administering *Bacteroides uniformis* (strain CECT 7771) ameliorated high fat diet-induced metabolic and immune dysfunction associated with an altered gut microbiota in adult C57BL-6 mice (Gauffin Cano et al., 2012). Inspection of the genome of *B. uniformis* ATCC 8492 revealed a 7,976 bp, five-gene sactipeptide-like cluster (Figure 5). Manual annotation identified a putative bacteriocin-type signal sequence containing a conserved TIGR04149 domain (7.43e-09). The area of interest also contained a pair of putative peptide-modifying radical SAM proteins (conserved domains TIGR04148 and TIGR04150 respectively) similar to those in *B. fragilis* 3_1_12 that were referred to above, a putative ABC-type bacteriocin exporter (conserved domain COG2274 0.0) and a putative hemolysin secretion protein (conserved domain pfam13437 1.02e-16).

Identification of novel PBGCs in Ruminococcus spp.

Ruminococci are Gram-positive anaerobes commonly found in the human gut, where they have been proposed to play a pivotal role in the fermentation of resistant starch (Ze et al., 2012). There have been several previous reports of bacteriocin production by members of the ruminococci, including a class IIa lantibiotic, ruminococcin A, produced by *Ruminococcus gnavus* E1 and two distinct class III bacteriocins produced by *Ruminococcus albus* 7 (Chen et al., 2004, Dabard et al., 2001, Wang et al., 2012a). We identified two apparently novel *Ruminococcus*-associated PBGCs, from among a total of 35 Firmicutes-associated clusters.

The genome of *Ruminococcus* sp. 5_1_39_B_FAA contained a 13,553 bp lantibiotic-like cluster containing six genes (Figure 5). The cluster contained a putative

response regulator receiver protein (conserved domain COG3279 3.95e-24), a putative histidine kinase (conserved domain pfam14501 3.5e-20), a putative type 2 lantibiotic biosynthesis protein LanM (conserved domain TIGR03897 0.0), a putative UviB-like bacteriocin (BAGEL3 bacteriocin II database 3e-11), a putative ABC transporter ATP-binding protein (conserved domain COG1136 8.20e-111) and a putative efflux ABC transporter permease protein.

Strains of *Ruminococcus obeum* have been shown to restrict *Vibrio cholerae* infection *via* a quorum-sensing-mediated mechanism (Hsiao et al., 2014). *Ruminococcus obeum* ATCC 29174 was isolated from human faeces and sequenced by the Washington University Genome Sequencing Centre. A 8,879 bp lantibiotic-like cluster comprising six genes was identified (Figure 5). The putative structural gene was found to resemble geobacillin I (BAGEL3 bacteriocin I database 5e-12), a nisin homolog isolated from *Geobacillus thermodenitrificans* (Garg et al., 2012). Also present in the area of interest were genes that appear to encode a two-component regulatory system, consisting of a putative histidine kinase (conserved domain COG0642 1.84e-24) and a putative *NisR* homolog containing signal receiver and effector domains (cd00156 and cd00383 respectively). Furthermore, genes potentially enoding a lantibiotic dehydratase similar to the entianin (lantibiotic) modification protein *EtnB* (BlastP 0.0) (Fuchs et al., 2011), an ABC transport protein similar to SpaT (transportation of the lantibiotic subtilin; BlastP 0.0) and a lanthionine synthetase protein similar to SpaC (modification of subtilin; BlastP 6e-117) were identified.

Identification of a novel PBGC in Roseburia spp.

Roseburia is a genus of Gram positive, butyrate-producers found to be negatively associated with type 2 diabetes and ulcerative colitis (Machiels et al., 2014, Qin et al., 2012). It has also been linked with ameliorating high-fat diet induced metabolic alterations in mice (Neyrinck et al., 2012). The only *Roseburia*-associated bacteriocin-producer to have been identified to date is *Roseburia faecis* M72/1 (Hatziioanou et al., 2013). *Roseburia intestinalis* L1-82, the type strain, was found to contain a five gene, 6078 bp sactipeptide-like cluster (Figure 5). The area of interest contained a putative bacteriocin-associated radical SAM protein (conversed domain TIGR04068 0.0), a putative peptide maturation system protein (conserved domain TIGR04066 8.58e-165), a putative peptide maturation system acyl carrier-related protein (conserved domain TIGR04069 1.15e-29), a subtilase family serine protease (conserved domain cd07492 7.11e-40) and a putative ABC transporter (conserved domain cd03228 5.95e-65). However, there were no immediately obvious bacteriocin structural or immunity genes in the area of interest and so it is particularly unclear if this cluster has the potential to produce an antimicrobial.

Conclusions

The large number of fully sequenced genomes available in public repositories means that genome-mining approaches are increasingly valuable with respect to the identification of novel genes and gene clusters (Eustaquio et al., 2011, Velásquez and van der Donk, 2011, Chen et al., 2014). As it has already been established that *in silico* approaches can be applied to the human microbiome for the purpose of identifying antimicrobial-producing microorganisms (Zheng et al., 2014, Donia et al., 2014), and that bacteriocins identified in this manner can be produced *in vitro* (Begley et al., 2009), it is apparent that there are considerable potential benefits in screening for and harnessing putative bacteriocin gene clusters from such databases.

It is commonly reported that between 30 and 99% of bacteria have the potential to produce at least one bacteriocin (Klaenhammer, 1988, Riley, 1998). It is thus notable that this *in silico*-based study identified just 59 genomes encoding probable PBGCs from 382 reference genomes, a frequency of just 15.4%. It is unclear whether this low number is representative of bacteriocin-production in the human GIT or an underestimation due to biases in identification of gene clusters. In support of the former of these theories, a recent study on the human microbiome by Zheng *et al*. reported that the gut contained the lowest density of putative bacteriocin genes of all body sites investigated (Zheng et al., 2014). That study identified 123 putative lantibiotic, 56 putative class II bacteriocin and 148 putative class III bacteriocin gene clusters in the gut environment. However sactipeptide-like clusters were not reported by Zheng *et al*. (Zheng et al., 2014). The discrepancy between the results reported by this study and those reported by Zheng *et al.* can be explained by differences in methodology. This method used BAGEL3 for the initial analysis while Zheng *et al.* performed a PSI-BLAST-based approach using the amino acid sequences from the

BAGEL3 bacteriocin database as driver sequences. Furthermore, we manually annotated the potential clusters returned initially, resulting in a dramatic decrease in reported PBGCs. It is noteworthy that *in silico* screens are limited by their dependence on similarity to previously described bacteriocin-associated genes, meaning that is it possible to overlook completely novel bacteriocin clusters.

The vast majority of known/characterised lantibiotics are produced by members of the Firmicutes (Li and O'Sullivan, 2012). Similarly, of the 11 lantibiotics PBGCs identified in this study, seven were found in the genomes of Firmicutes, with two associated with bifidobacteria (Actinobacteria) and two by *Bacteroides* spp. (Bacteroidetes). While these clusters typically contained features that are common to lantibiotic-associated gene clusters, two putative lantibiotic clusters (in *Bifidobacterium* sp. 12_1_47BFAA and *Enterococcus faecalis* TX1342 (Figures 5 and 3 respectively)) contained predicted FMN reductase genes in addition to those more traditionally associated with lantibiotic modification.

It is apparent that the *in silico* screen identified gene clusters representative of some classes of bacteriocin more frequently than others. Clusters resembling those associated with the production of bacteriolysins (formerly referred to as class III bacteriocins) were most common. The large numbers of colicin-like and enterolysin A-like clusters was possibly due to the overrepresentation of *E. coli* in the reference genome database and the relative ease of detection. It appears that enterolysin A does not possess a specific immunity gene; instead, resistance results from the absence of specific binding receptors (Mendez-Vilas and Antonio, 2011), making this single gene potentially easier to detect than a multi-gene operon. On the other hand, the relatively low frequency of class II bacteriocins (three unmodified and one class IIc) cannot be explained in a similar manner. It is unclear whether this paucity is due to the

methodology or an actual scarcity of class II bacteriocin producers in the gut microbiota. Comparatively, Zheng *et al*. identified 56 class II bacteriocin structural genes from gut-associated strains (Zheng et al., 2014) suggesting that either this is an overestimation due to the lack of manual annotation or the approach used in this study is not ideal for the identification of Class II bacteriocins.

This comprehensive *in silico* study led to the identification of PBGCs in species that not previously associated with bacteriocin production, for example *Bacteroides uniformis* and *Roseburia intestinalis.* We also identified potential bacteriocin gene clusters in two *Bifidobacterium* species, a genus which has long been thought of as beneficial to the human host. It is not possible, by *in silico* methods alone, to state conclusively if these bacteriocins are produced *in vitro*. However, if even a portion of these gene clusters are responsible for bacteriocin production in the corresponding strain, it could greatly expand the arsenal of bacteriocins available for use in food and healthcare. Such investigations will be the focus of our future investigations.

References

AAS, J., GESSERT, C. E. & BAKKEN, J. S. 2003. Recurrent Clostridium difficile colitis: case series involving 18 patients treated with donor stool administered via a nasogastric tube. *Clin Infect Dis,* 36**,** 580-5.

ADLERBERTH, I. & WOLD, A. E. 2009. Establishment of the gut microbiota in Western infants. *Acta Paediatr,* 98**,** 229-38.

ANDERSON, C. A., BOUCHER, G., LEES, C. W., FRANKE, A., D'AMATO, M., TAYLOR, K. D., LEE, J. C., GOYETTE, P., IMIELINSKI, M. & LATIANO, A. 2011. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics,* 43**,** 246-252.

ANDREASEN, A. S., LARSEN, N., PEDERSEN-SKOVSGAARD, T., BERG, R. M., MOLLER, K., SVENDSEN, K. D., JAKOBSEN, M. & PEDERSEN, B. K. 2010. Effects of Lactobacillus acidophilus NCFM on insulin sensitivity and the systemic inflammatory response in human subjects. *Br J Nutr,* 104**,** 1831-8.

ANTONOPOULOS, D. A., HUSE, S. M., MORRISON, H. G., SCHMIDT, T. M., SOGIN, M. L. & YOUNG, V. B. 2009. Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation. *Infection and Immunity,* 77**,** 2367-2375.

ARNISON, P. G., BIBB, M. J., BIERBAUM, G., BOWERS, A. A., BUGNI, T. S., BULAJ, G., CAMARERO, J. A., CAMPOPIANO, D. J., CHALLIS, G. L., CLARDY, J., COTTER, P. D., CRAIK, D. J., DAWSON, M., DITTMANN, E., DONADIO, S., DORRESTEIN, P. C., ENTIAN, K.-D., FISCHBACH, M. A., GARAVELLI, J. S., GORANSSON, U., GRUBER, C. W., HAFT, D. H.,

HEMSCHEIDT, T. K., HERTWECK, C., HILL, C., HORSWILL, A. R., JASPARS, M., KELLY, W. L., KLINMAN, J. P., KUIPERS, O. P., LINK, A. J., LIU, W., MARAHIEL, M. A., MITCHELL, D. A., MOLL, G. N., MOORE, B. S., MULLER, R., NAIR, S. K., NES, I. F., NORRIS, G. E., OLIVERA, B. M., ONAKA, H., PATCHETT, M. L., PIEL, J., REANEY, M. J. T., REBUFFAT, S., ROSS, R. P., SAHL, H.-G., SCHMIDT, E. W., SELSTED, M. E., SEVERINOV, K., SHEN, B., SIVONEN, K., SMITH, L., STEIN, T., SUSSMUTH, R. D., TAGG, J. R., TANG, G.-L., TRUMAN, A. W., VEDERAS, J. C., WALSH, C. T., WALTON, J. D., WENZEL, S. C., WILLEY, J. M. & VAN DER DONK, W. A. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports,* 30**,** 108-160.

ARNOLD, I. C., DEHZAD, N., REUTER, S., MARTIN, H., BECHER, B., TAUBE, C., XFC & LLER, A. 2011. Helicobacter pylori infection prevents allergic asthma in mouse models through the induction of regulatory T cells. *The Journal of Clinical Investigation,* 121**,** 3088-3093.

ARONSSON, L., HUANG, Y., PARINI, P., KORACH-ANDRE, M., HAKANSSON, J., GUSTAFSSON, J. A., PETTERSSON, S., ARULAMPALAM, V. & RAFTER, J. 2010. Decreased fat storage by Lactobacillus paracasei is associated with increased levels of angiopoietin-like 4 protein (ANGPTL4). *PLoS One,* 5.

ARTHUR, J. C., PEREZ-CHANONA, E., MUHLBAUER, M., TOMKOVICH, S., URONIS, J. M., FAN, T. J., CAMPBELL, B. J., ABUJAMEL, T., DOGAN, B., ROGERS, A. B., RHODES, J. M., STINTZI, A., SIMPSON, K. W.,

HANSEN, J. J., KEKU, T. O., FODOR, A. A. & JOBIN, C. 2012. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science,* 338**,** 120-3.

AVELAR, K., PINTO, L., ANTUNES, L., LOBO, L., BASTOS, M., DOMINGUES, R. & DE SOUZA FERREIRA, M. 1999. Production of bacteriocin by Bacteroides fragilis and partial characterization. *Letters in applied microbiology,* 29**,** 264-268.

BACKHED, F., DING, H., WANG, T., HOOPER, L. V., KOH, G. Y., NAGY, A., SEMENKOVICH, C. F. & GORDON, J. I. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A,* 101**,** 15718-23.

BÄCKHED, F., LEY, R. E., SONNENBURG, J. L., PETERSON, D. A. & GORDON, J. I. 2005. Host-Bacterial Mutualism in the Human Intestine. *Science,* 307**,** 1915-1920.

BAHAR, A. A. & REN, D. 2013. Antimicrobial Peptides. *Pharmaceuticals,* 6**,** 1543-1575.

BAKIR, M. A., SAKAMOTO, M., KITAHARA, M., MATSUMOTO, M. & BENNO, Y. 2006. Bacteroides dorei sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol,* 56**,** 1639-43.

BANIK, J. J. & BRADY, S. F. 2010. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr Opin Microbiol,* 13**,** 603-9.

BARTON, W., PENNEY, N. C., CRONIN, O., GARCIA-PEREZ, I., MOLLOY, M. G., HOLMES, E., SHANAHAN, F., COTTER, P. D. & O'SULLIVAN, O. 2017. The microbiome of professional athletes differs from that of more

sedentary subjects in composition and particularly at the functional metabolic level. *Gut*.

BEGLEY, M., COTTER, P. D., HILL, C. & ROSS, R. P. 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl Environ Microbiol*, 75**,** 5451-60.

BELSTRØM, D., HOLMSTRUP, P., BARDOW, A., KOKARAS, A., FIEHN, N.-E. & PASTER, B. J. 2016. Temporal Stability of the Salivary Microbiota in Oral Health. *PLoS ONE*, 11**,** e0147472.

BENGTSON, M. B., AAMODT, G., VATN, M. H. & HARRIS, J. R. 2010. Concordance for IBD among twins compared to ordinary siblings--a Norwegian population-based study. *J Crohns Colitis*, 4**,** 312-8.

BISHOP, B. M., JUBA, M. L., RUSSO, P. S., DEVINE, M., BARKSDALE, S. M., SCOTT, S., SETTLAGE, R., MICHALAK, P., GUPTA, K., VLIET, K., SCHNUR, J. M. & VAN HOEK, M. L. 2017. Discovery of Novel Antimicrobial Peptides from Varanus komodoensis (Komodo Dragon) by Large-Scale Analyses and De-Novo-Assisted Sequencing Using Electron-Transfer Dissociation Mass Spectrometry. *J Proteome Res*.

BLASER, M. 2011. Antibiotic overuse: Stop the killing of beneficial bacteria. *Nature*, 476**,** 393-394.

BOOTH, S., JOHNSON, J. & WILKINS, T. 1977. Bacteriocin production by strains of Bacteroides isolated from human feces and the role of these strains in the bacterial ecology of the colon. *Antimicrobial agents and chemotherapy*, 11**,** 718-724.

BOUHNIK, Y., ALAIN, S., ATTAR, A., FLOURIE, B., RASKINE, L., SANSON-LE PORS, M. J. & RAMBAUD, J. C. 1999. Bacterial populations

contaminating the upper gut in patients with small intestinal bacterial overgrowth syndrome. *Am J Gastroenterol,* 94**,** 1327-31.

BOUHNIK, Y., RASKINE, L., SIMONEAU, G., VICAUT, E., NEUT, C., FLOURIÉ, B., BROUNS, F. & BORNET, F. R. 2004. The capacity of nondigestible carbohydrates to stimulate fecal bifidobacteria in healthy humans: a double-blind, randomized, placebo-controlled, parallel-group, dose-response relation study. *The American journal of clinical nutrition,* 80**,** 1658-1664.

BRINT, E. K., MACSHARRY, J., FANNING, A., SHANAHAN, F. & QUIGLEY, E. M. 2011. Differential expression of toll-like receptors in patients with irritable bowel syndrome. *Am J Gastroenterol,* 106**,** 329-36.

BROOKS, A. W., KOHL, K. D., BRUCKER, R. M., VAN OPSTAL, E. J. & BORDENSTEIN, S. R. 2016. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biology,* 14**,** e2000225.

BULTMAN, S. J. 2017. Interplay between diet, gut microbiota, epigenetic events, and colorectal cancer. *Molecular Nutrition & Food Research,* 61**,** 1500902-n/a.

BULTMAN, S. J. & JOBIN, C. 2014. Microbial-derived butyrate: an oncometabolite or tumor-suppressive metabolite? *Cell Host Microbe,* 16**,** 143-5.

CANI, P. D. 2013. Gut microbiota and obesity: lessons from the microbiome. *Brief Funct Genomics,* 12**,** 381-7.

CARVER, T. J., RUTHERFORD, K. M., BERRIMAN, M., RAJANDREAM, M.-A., BARRELL, B. G. & PARKHILL, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics (Oxford, England),* 21**,** 3422-3423.

CASEY, P. G., CASEY, G. D., GARDINER, G. E., TANGNEY, M., STANTON, C., ROSS, R. P., HILL, C. & FITZGERALD, G. F. 2004. Isolation and

characterization of anti-Salmonella lactic acid bacteria from the porcine gastrointestinal tract. *Letters in Applied Microbiology,* 39**,** 431-438.

CASEY, P. G., GARDINER, G. E., CASEY, G., BRADSHAW, B., LAWLOR, P. G., LYNCH, P. B., LEONARD, F. C., STANTON, C., ROSS, R. P. & FITZGERALD, G. F. 2007. A five-strain probiotic combination reduces pathogen shedding and alleviates disease signs in pigs challenged with Salmonella enterica serovar Typhimurium. *Applied and Environmental Microbiology,* 73**,** 1858-1863.

CASTELLARIN, M., WARREN, R. L., FREEMAN, J. D., DREOLINI, L., KRZYWINSKI, M., STRAUSS, J., BARNES, R., WATSON, P., ALLEN-VERCOE, E., MOORE, R. A. & HOLT, R. A. 2012. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res,* 22**,** 299-306.

CHAPMAN, C., GIBSON, G. & ROWLAND, I. 2011. Health benefits of probiotics: are mixtures more effective than single strains? *European journal of nutrition,* 50**,** 1-17.

CHEIKHYOUSSEF, A., POGORI, N., CHEN, H., TIAN, F., CHEN, W., TANG, J. & ZHANG, H. 2009. Antimicrobial activity and partial characterization of bacteriocin-like inhibitory substances (BLIS) produced by Bifidobacterium infantis BCRC 14602. *Food Control,* 20**,** 553-559.

CHEN, D., FENG, J., HUANG, L., ZHANG, Q., WU, J., ZHU, X., DUAN, Y. & XU, Z. 2014. Identification and Characterization of a New Erythromycin Biosynthetic Gene Cluster in Actinopolyspora erythraea YIM90600, a Novel Erythronolide-Producing Halophilic Actinomycete Isolated from Salt Field. *PLoS One,* 9**,** e108129.

CHEN, J., STEVENSON, D. M. & WEIMER, P. J. 2004. Albusin B, a bacteriocin from the ruminal bacterium Ruminococcus albus 7 that inhibits growth of Ruminococcus flavefaciens. *Appl Environ Microbiol,* 70**,** 3167-70.

CHEN, J., WRIGHT, K., DAVIS, J. M., JERALDO, P., MARIETTA, E. V., MURRAY, J., NELSON, H., MATTESON, E. L. & TANEJA, V. 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Medicine,* 8**,** 43.

CHEN, Y. & BLASER, M. J. 2007. Inverse associations of helicobacter pylori with asthma and allergy. *Archives of Internal Medicine,* 167**,** 821-827.

CHO, I., YAMANISHI, S., COX, L., METHE, B. A., ZAVADIL, J., LI, K., GAO, Z., MAHANA, D., RAJU, K., TEITLER, I., LI, H., ALEKSEYENKO, A. V. & BLASER, M. J. 2012. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature,* 488**,** 621-6.

CLAESSON, M. J., CUSACK, S., O'SULLIVAN, O., GREENE-DINIZ, R., DE WEERD, H., FLANNERY, E., MARCHESI, J. R., FALUSH, D., DINAN, T., FITZGERALD, G., STANTON, C., VAN SINDEREN, D., O'CONNOR, M., HARNEDY, N., O'CONNOR, K., HENRY, C., O'MAHONY, D., FITZGERALD, A. P., SHANAHAN, F., TWOMEY, C., HILL, C., ROSS, R. P. & O'TOOLE, P. W. 2011. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4586-91.

CLAESSON, M. J., JEFFERY, I. B., CONDE, S., POWER, S. E., O/'CONNOR, E. M., CUSACK, S., HARRIS, H. M. B., COAKLEY, M., LAKSHMINARAYANAN, B., O/'SULLIVAN, O., FITZGERALD, G. F., DEANE, J., O/'CONNOR, M., HARNEDY, N., O/'CONNOR, K.,

O/'MAHONY, D., VAN SINDEREN, D., WALLACE, M., BRENNAN, L., STANTON, C., MARCHESI, J. R., FITZGERALD, A. P., SHANAHAN, F., HILL, C., ROSS, R. P. & O/'TOOLE, P. W. 2012. Gut microbiota composition correlates with diet and health in the elderly. *Nature,* 488**,** 178-184.

CLARKE, S. F., MURPHY, E. F., NILAWEERA, K., ROSS, P. R., SHANAHAN, F., O'TOOLE, P. W. & COTTER, P. D. 2012. The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes,* 3**,** 186-202.

CLARKE, S. F., MURPHY, E. F., O'SULLIVAN, O., LUCEY, A. J., HUMPHREYS, M., HOGAN, A., HAYES, P., O'REILLY, M., JEFFERY, I. B., WOOD-MARTIN, R., KERINS, D. M., QUIGLEY, E., ROSS, R. P., O'TOOLE, P. W., MOLLOY, M. G., FALVEY, E., SHANAHAN, F. & COTTER, P. D. 2014. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*.

CLEMENTE, JOSE C., URSELL, LUKE K., PARFREY, LAURA W. & KNIGHT, R. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell,* 148**,** 1258-1270.

CLINTON, S. K., BOSTWICK, D. G., OLSON, L. M., MANGIAN, H. J. & VISEK, W. J. 1988. Effects of ammonium acetate and sodium cholate on N-methyl-N′-nitro-N-nitrosoguanidine-induced colon carcinogenesis of rats. *Cancer research,* 48**,** 3035-3039.

CORR, S. C., LI, Y., RIEDEL, C. U., O'TOOLE, P. W., HILL, C. & GAHAN, C. G. M. 2007. Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118. *Proceedings of the National Academy of Sciences,* 104**,** 7617-7621.

COSTABILE, A., KOLIDA, S., KLINDER, A., GIETL, E., BÄUERLEIN, M., FROHBERG, C., LANDSCHÜTZE, V. & GIBSON, G. R. 2010. A double-blind, placebo-controlled, cross-over study to establish the bifidogenic effect of a very-long-chain inulin extracted from globe artichoke (Cynara scolymus) in healthy human subjects. *British Journal of Nutrition,* 104**,** 1007.

COTILLARD, A., KENNEDY, S. P., KONG, L. C., PRIFTI, E., PONS, N., LE CHATELIER, E., ALMEIDA, M., QUINQUIS, B., LEVENEZ, F., GALLERON, N., GOUGIS, S., RIZKALLA, S., BATTO, J.-M., RENAULT, P., CONSORTIUM, A. N. R. M., DORE, J., ZUCKER, J.-D., CLEMENT, K., EHRLICH, S. D. & MEMBERS, A. N. R. M. C. 2013. Dietary intervention impact on gut microbial gene richness. *Nature,* 500**,** 585-588.

COTTER, P. D., HILL, C. & ROSS, R. P. 2005a. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol,* 3**,** 777-88.

COTTER, P. D., O'CONNOR, P. M., DRAPER, L. A., LAWTON, E. M., DEEGAN, L. H., HILL, C. & ROSS, R. P. 2005b. Posttranslational conversion of l-serines to d-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 18584-18589.

COTTER, P. D., ROSS, R. P. & HILL, C. 2013. Bacteriocins - a viable alternative to antibiotics? *Nat Rev Micro,* 11**,** 95-105.

COTTER, P. D., STANTON, C., ROSS, R. P. & HILL, C. 2012. The impact of antibiotics on the gut microbiota as revealed by high throughput DNA sequencing. *Discovery medicine,* 13**,** 193.

CRYAN, J. F. & O'MAHONY, S. M. 2011. The microbiome-gut-brain axis: from bowel to behavior. *Neurogastroenterol Motil,* 23**,** 187-92.

CUMMINGS, J. & MACFARLANE, G. 1991. The control and consequences of bacterial fermentation in the human colon. *Journal of Applied Microbiology,* 70**,** 443-459.

CUMMINGS, J. H. & ENGLYST, H. N. 1991. What is dietary fibre? *Trends in Food Science & Technology,* 2**,** 99-103.

DABARD, J., BRIDONNEAU, C., PHILLIPE, C., ANGLADE, P., MOLLE, D., NARDI, M., LADIRE, M., GIRARDIN, H., MARCILLE, F., GOMEZ, A. & FONS, M. 2001. Ruminococcin A, a new lantibiotic produced by a Ruminococcus gnavus strain isolated from human feces. *Appl Environ Microbiol,* 67**,** 4111-8.

DAI, C., ZHENG, C. Q., JIANG, M., MA, X. Y. & JIANG, L. J. 2013. Probiotics and irritable bowel syndrome. *World J Gastroenterol,* 19**,** 5973-80.

DANIEL, H., GHOLAMI, A. M., BERRY, D., DESMARCHELIER, C., HAHNE, H., LOH, G., MONDOT, S., LEPAGE, P., ROTHBALLER, M., WALKER, A., BOHM, C., WENNING, M., WAGNER, M., BLAUT, M., SCHMITT-KOPPLIN, P., KUSTER, B., HALLER, D. & CLAVEL, T. 2013. High-fat diet alters gut microbiota physiology in mice. *ISME J.*

DAO, M. C., EVERARD, A., ARON-WISNEWSKY, J., SOKOLOVSKA, N., PRIFTI, E., VERGER, E. O., KAYSER, B. D., LEVENEZ, F., CHILLOUX, J., HOYLES, L., DUMAS, M. E., RIZKALLA, S. W., DORE, J., CANI, P. D. & CLEMENT, K. 2016. Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut,* 65**,** 426-36.

DAVID, L. A., MAURICE, C. F., CARMODY, R. N., GOOTENBERG, D. B., BUTTON, J. E., WOLFE, B. E., LING, A. V., DEVLIN, A. S., VARMA, Y.

& FISCHBACH, M. A. 2013. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*.

DE FILIPPO, C., CAVALIERI, D., DI PAOLA, M., RAMAZZOTTI, M., POULLET, J. B., MASSART, S., COLLINI, S., PIERACCINI, G. & LIONETTI, P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences*, 107**,** 14691-14696.

DE LANGE, K. M., MOUTSIANAS, L., LEE, J. C., LAMB, C. A., LUO, Y., KENNEDY, N. A., JOSTINS, L., RICE, D. L., GUTIERREZ-ACHURY, J., JI, S.-G., HEAP, G., NIMMO, E. R., EDWARDS, C., HENDERSON, P., MOWAT, C., SANDERSON, J., SATSANGI, J., SIMMONS, A., WILSON, D. C., TREMELLING, M., HART, A., MATHEW, C. G., NEWMAN, W. G., PARKES, M., LEES, C. W., UHLIG, H., HAWKEY, C., PRESCOTT, N. J., AHMAD, T., MANSFIELD, J. C., ANDERSON, C. A. & BARRETT, J. C. 2017. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*, 49**,** 256-261.

DE SMET, K. & CONTRERAS, R. 2005. Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnol Lett*, 27**,** 1337-47.

DE VOS, W. M. 2013. Fame and future of faecal transplantations--developing next-generation therapies with synthetic microbiomes. *Microb Biotechnol*, 6**,** 316-25.

DE VADDER, F., KOVATCHEVA-DATCHARY, P., GONCALVES, D., VINERA, J., ZITOUN, C., DUCHAMPT, A., BÄCKHED, F. & MITHIEUX, G. 2014.

Microbiota-Generated Metabolites Promote Metabolic Benefits via Gut-Brain Neural Circuits. *Cell,* 156**,** 84-96.

DEEGAN, L. H., COTTER, P. D., HILL, C. & ROSS, P. 2006. Bacteriocins: Biological tools for bio-preservation and shelf-life extension. *International Dairy Journal,* 16**,** 1058-1071.

DOMINGUEZ-BELLO, M. G., COSTELLO, E. K. & KNIGHT, R. 2010. Reply to Putignani et al.: Vagina as a major source of natural inoculum for the newborn. *Proceedings of the National Academy of Sciences,* 107**,** E160.

DONIA, MOHAMED S., CIMERMANCIC, P., SCHULZE, CHRISTOPHER J., WIELAND BROWN, LAURA C., MARTIN, J., MITREVA, M., CLARDY, J., LININGTON, ROGER G. & FISCHBACH, MICHAEL A. 2014. A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. *Cell,* 158**,** 1402-1414.

DRISSI, F., BUFFET, S., RAOULT, D. & MERHEJ, V. 2015. Common occurrence of antibacterial agents in human intestinal microbiota. *Frontiers in Microbiology,* 6.

DUNCAN, S. H., BELENGUER, A., HOLTROP, G., JOHNSTONE, A. M., FLINT, H. J. & LOBLEY, G. E. 2007. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Applied and environmental microbiology,* 73**,** 1073-1078.

ELSON, C. O. & CONG, Y. 2012. Host-microbiota interactions in inflammatory bowel disease. *Gut Microbes,* 3**,** 332-44.

ELWOOD, P. C., GIVENS, D. I., BESWICK, A. D., FEHILY, A. M., PICKERING, J. E. & GALLACHER, J. 2008. The survival advantage of milk and dairy

consumption: an overview of evidence from cohort studies of vascular diseases, diabetes and cancer. *J Am Coll Nutr,* 27**,** 723s-34s.

EUSTAQUIO, A. S., NAM, S. J., PENN, K., LECHNER, A., WILSON, M. C., FENICAL, W., JENSEN, P. R. & MOORE, B. S. 2011. The discovery of salinosporamide K from the marine bacterium "Salinispora pacifica" by genome mining gives insight into pathway evolution. *Chembiochem,* 12**,** 61-4.

EVERARD, A., BELZER, C., GEURTS, L., OUWERKERK, J. P., DRUART, C., BINDELS, L. B., GUIOT, Y., DERRIEN, M., MUCCIOLI, G. G., DELZENNE, N. M., DE VOS, W. M. & CANI, P. D. 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A,* 110**,** 9066-71.

FAITH, J. J., GURUGE, J. L., CHARBONNEAU, M., SUBRAMANIAN, S., SEEDORF, H., GOODMAN, A. L., CLEMENTE, J. C., KNIGHT, R., HEATH, A. C., LEIBEL, R. L., ROSENBAUM, M. & GORDON, J. I. 2013. The Long-Term Stability of the Human Gut Microbiota. *Science,* 341.

FAITH, J. J., MCNULTY, N. P., REY, F. E. & GORDON, J. I. 2011. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science,* 333**,** 101-104.

FITZPATRICK, L. R. 2013. Probiotics for the treatment of Clostridium difficile associated disease. *World J Gastrointest Pathophysiol,* 4**,** 47-52.

FLINT, H. J., SCOTT, K. P., LOUIS, P. & DUNCAN, S. H. 2012. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol,* 9**,** 577-589.

FRANCESCHI, C. 2007. Inflammaging as a major characteristic of old people: can it be prevented or cured? *Nutr Rev,* 65**,** S173-6.

FRANK, D. N., AMAND, A. L. S., FELDMAN, R. A., BOEDEKER, E. C., HARPAZ, N. & PACE, N. R. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences,* 104**,** 13780-13785.

FRANKE, A., MCGOVERN, D. P., BARRETT, J. C., WANG, K., RADFORD-SMITH, G. L., AHMAD, T., LEES, C. W., BALSCHUN, T., LEE, J., ROBERTS, R., ANDERSON, C. A., BIS, J. C., BUMPSTEAD, S., ELLINGHAUS, D., FESTEN, E. M., GEORGES, M., GREEN, T., HARITUNIANS, T., JOSTINS, L., LATIANO, A., MATHEW, C. G., MONTGOMERY, G. W., PRESCOTT, N. J., RAYCHAUDHURI, S., ROTTER, J. I., SCHUMM, P., SHARMA, Y., SIMMS, L. A., TAYLOR, K. D., WHITEMAN, D., WIJMENGA, C., BALDASSANO, R. N., BARCLAY, M., BAYLESS, T. M., BRAND, S., BUNING, C., COHEN, A., COLOMBEL, J. F., COTTONE, M., STRONATI, L., DENSON, T., DE VOS, M., D'INCA, R., DUBINSKY, M., EDWARDS, C., FLORIN, T., FRANCHIMONT, D., GEARRY, R., GLAS, J., VAN GOSSUM, A., GUTHERY, S. L., HALFVARSON, J., VERSPAGET, H. W., HUGOT, J. P., KARBAN, A., LAUKENS, D., LAWRANCE, I., LEMANN, M., LEVINE, A., LIBIOULLE, C., LOUIS, E., MOWAT, C., NEWMAN, W., PANES, J., PHILLIPS, A., PROCTOR, D. D., REGUEIRO, M., RUSSELL, R., RUTGEERTS, P., SANDERSON, J., SANS, M., SEIBOLD, F., STEINHART, A. H., STOKKERS, P. C., TORKVIST, L., KULLAK-UBLICK, G., WILSON, D., WALTERS, T., TARGAN, S. R., BRANT, S. R., RIOUX, J. D., D'AMATO, M., WEERSMA, R. K., KUGATHASAN, S., GRIFFITHS, A. M., MANSFIELD, J. C., VERMEIRE, S., DUERR, R. H., SILVERBERG, M. S.,

SATSANGI, J., SCHREIBER, S., CHO, J. H., ANNESE, V., HAKONARSON, H., DALY, M. J. & PARKES, M. 2010. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet,* 42**,** 1118-25.

FUCHS, S. W., JASKOLLA, T. W., BOCHMANN, S., KÖTTER, P., WICHELHAUS, T., KARAS, M., STEIN, T. & ENTIAN, K.-D. 2011. Entianin, a Novel Subtilin-Like Lantibiotic from Bacillus subtilis subsp. spizizenii DSM 15029(T) with High Antimicrobial Activity. *Applied and Environmental Microbiology,* 77**,** 1698-1707.

FUKUDA, S., TOH, H., HASE, K., OSHIMA, K., NAKANISHI, Y., YOSHIMURA, K., TOBE, T., CLARKE, J. M., TOPPING, D. L., SUZUKI, T., TAYLOR, T. D., ITOH, K., KIKUCHI, J., MORITA, H., HATTORI, M. & OHNO, H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature,* 469**,** 543-547.

FURET, J. P., KONG, L. C., TAP, J., POITOU, C., BASDEVANT, A., BOUILLOT, J. L., MARIAT, D., CORTHIER, G., DORE, J., HENEGAR, C., RIZKALLA, S. & CLEMENT, K. 2010. Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes,* 59**,** 3049-57.

GABERT, L., VORS, C., LOUCHE-PÉLISSIER, C., SAUVINET, V., LAMBERT-PORCHERON, S., DRAI, J., LAVILLE, M., DÉSAGE, M. & MICHALSKI, M. C. 2011. 13C tracer recovery in human stools after digestion of a fat-rich meal labelled with [1, 1, 1-13C3] tripalmitin and [1, 1, 1-13C3] triolein. *Rapid Communications in Mass Spectrometry,* 25**,** 2697-2703.

GARG, N., TANG, W., GOTO, Y., NAIR, S. K. & VAN DER DONK, W. A. 2012. Lantibiotics from Geobacillus thermodenitrificans. *Proceedings of the National Academy of Sciences,* 109**,** 5241-5246.

GARRETT, W. S., GALLINI, C. A., YATSUNENKO, T., MICHAUD, M., DUBOIS, A., DELANEY, M. L., PUNIT, S., KARLSSON, M., BRY, L., GLICKMAN, J. N., GORDON, J. I., ONDERDONK, A. B. & GLIMCHER, L. H. 2010. Enterobacteriaceae Act in Concert with the Gut Microbiota to Induce Spontaneous and Maternally Transmitted Colitis. *Cell Host & Microbe,* 8**,** 292-300.

GAUFFIN CANO, P., SANTACRUZ, A., MOYA, Á. & SANZ, Y. 2012. Bacteroides uniformis CECT 7771 Ameliorates Metabolic and Immunological Dysfunction in Mice with High-Fat-Diet Induced Obesity. *PLoS ONE,* 7**,** e41079.

GIBBONS, S. M., KEARNEY, S. M., SMILLIE, C. S. & ALM, E. J. 2017. Two dynamic regimes in the human gut microbiome. *PLOS Computational Biology,* 13**,** e1005364.

GIBSON, G. R., PROBERT, H. M., VAN LOO, J., RASTALL, R. A. & ROBERFROID, M. B. 2004. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev,* 17**,** 259-275.

GIBSON, G. R. & ROBERFROID, M. B. 1995. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr,* 125**,** 1401-12.

GOUGH, E., SHAIKH, H. & MANGES, A. R. 2011. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent Clostridium difficile infection. *Clin Infect Dis,* 53**,** 994-1002.

GUINANE, C. M., LAWTON, E. M., O'CONNOR, P. M., O'SULLIVAN, O., HILL, C., ROSS, R. P. & COTTER, P. D. 2016. The bacteriocin bactofencin A subtly modulates gut microbial populations. *Anaerobe,* 40**,** 41-9.

GUINANE, C. M., PIPER, C., DRAPER, L. A., O'CONNOR, P. M., HILL, C., ROSS, R. P. & COTTER, P. D. 2015. Impact of Environmental Factors on Bacteriocin Promoter Activity in Gut-Derived Lactobacillus salivarius. *Applied and Environmental Microbiology,* 81**,** 7851-7859.

GUPTA, S., ALLEN-VERCOE, E. & PETROF, E. O. 2016. Fecal microbiota transplantation: in perspective. *Therap Adv Gastroenterol,* 9**,** 229-39.

HAFT, D. H. & BASU, M. K. 2011. Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J Bacteriol,* 193**,** 2745-55.

HAMER, H. M., DE PRETER, V., WINDEY, K. & VERBEKE, K. 2012. Functional analysis of colonic bacterial metabolism: relevant to health? *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 302**,** G1-G9.

HANSEN, R., RUSSELL, R. K., REIFF, C., LOUIS, P., MCINTOSH, F., BERRY, S. H., MUKHOPADHYA, I., BISSET, W. M., BARCLAY, A. R., BISHOP, J., FLYNN, D. M., MCGROGAN, P., LOGANATHAN, S., MAHDI, G., FLINT, H. J., EL-OMAR, E. M. & HOLD, G. L. 2012. Microbiota of De-Novo Pediatric IBD: Increased Faecalibacterium Prausnitzii and Reduced Bacterial Diversity in Crohn's But Not in Ulcerative Colitis. *Am J Gastroenterol,* 107**,** 1913-1922.

HATZIIOANOU, D., MAYER, M. J., DUNCAN, S. H., FLINT, H. J. & NARBAD, A. 2013. A representative of the dominant human colonic Firmicutes,

Roseburia faecis M72/1, forms a novel bacteriocin-like substance. *Anaerobe,* 23**,** 5-8.

HAVARSTEIN, L. S., DIEP, D. B. & NES, I. F. 1995. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol,* 16**,** 229-40.

HILDEBRANDT, M. A., HOFFMANN, C., SHERRILL–MIX, S. A., KEILBAUGH, S. A., HAMADY, M., CHEN, Y. Y., KNIGHT, R., AHIMA, R. S., BUSHMAN, F. & WU, G. D. 2009. High-Fat Diet Determines the Composition of the Murine Gut Microbiome Independently of Obesity. *Gastroenterology,* 137**,** 1716-1724.e2.

HSIAO, A., AHMED, A. M., SUBRAMANIAN, S., GRIFFIN, N. W., DREWRY, L. L., PETRI, W. A., HAQUE, R., AHMED, T. & GORDON, J. I. 2014. Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. *Nature*.

HSIAO, ELAINE Y., MCBRIDE, SARA W., HSIEN, S., SHARON, G., HYDE, EMBRIETTE R., MCCUE, T., CODELLI, JULIAN A., CHOW, J., REISMAN, SARAH E., PETROSINO, JOSEPH F., PATTERSON, PAUL H. & MAZMANIAN, SARKIS K. 2013. Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders. *Cell,* 155**,** 1451-1463.

IYER, L. M., ABHIMAN, S., BURROUGHS, A. M. & ARAVIND, L. 2009. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Molecular Biosystems,* 5**,** 1636-1660.

JALANKA-TUOVINEN, J., SALOJÄRVI, J., SALONEN, A., IMMONEN, O., GARSED, K., KELLY, F. M., ZAITOUN, A., PALVA, A., SPILLER, R. C. & DE VOS, W. M. 2013. Faecal microbiota composition and host–microbe cross-talk following gastroenteritis and in postinfectious irritable bowel syndrome. *Gut*.

JEFFERY, I. B., LYNCH, D. B. & O'TOOLE, P. W. 2016. Composition and temporal stability of the gut microbiota in older persons. *ISME J,* 10**,** 170-182.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. 2011. Global cancer statistics. *CA Cancer J Clin,* 61**,** 69-90.

JOSTINS, L., RIPKE, S., WEERSMA, R. K., DUERR, R. H., MCGOVERN, D. P., HUI, K. Y., LEE, J. C., PHILIP SCHUMM, L., SHARMA, Y., ANDERSON, C. A., ESSERS, J., MITROVIC, M., NING, K., CLEYNEN, I., THEATRE, E., SPAIN, S. L., RAYCHAUDHURI, S., GOYETTE, P., WEI, Z., ABRAHAM, C., ACHKAR, J.-P., AHMAD, T., AMININEJAD, L., ANANTHAKRISHNAN, A. N., ANDERSEN, V., ANDREWS, J. M., BAIDOO, L., BALSCHUN, T., BAMPTON, P. A., BITTON, A., BOUCHER, G., BRAND, S., BUNING, C., COHAIN, A., CICHON, S., D/'AMATO, M., DE JONG, D., DEVANEY, K. L., DUBINSKY, M., EDWARDS, C., ELLINGHAUS, D., FERGUSON, L. R., FRANCHIMONT, D., FRANSEN, K., GEARRY, R., GEORGES, M., GIEGER, C., GLAS, J., HARITUNIANS, T., HART, A., HAWKEY, C., HEDL, M., HU, X., KARLSEN, T. H., KUPCINSKAS, L., KUGATHASAN, S., LATIANO, A., LAUKENS, D., LAWRANCE, I. C., LEES, C. W., LOUIS, E., MAHY, G., MANSFIELD, J., MORGAN, A. R., MOWAT, C., NEWMAN, W., PALMIERI, O., PONSIOEN, C. Y., POTOCNIK, U., PRESCOTT, N. J., REGUEIRO, M.,

ROTTER, J. I., RUSSELL, R. K., SANDERSON, J. D., SANS, M., SATSANGI, J., SCHREIBER, S., SIMMS, L. A., SVENTORAITYTE, J., TARGAN, S. R., TAYLOR, K. D., TREMELLING, M., VERSPAGET, H. W., DE VOS, M., WIJMENGA, C., WILSON, D. C., WINKELMANN, J., XAVIER, R. J., ZEISSIG, S., ZHANG, B., ZHANG, C. K., ZHAO, H., SILVERBERG, M. S., ANNESE, V., HAKONARSON, H., BRANT, S. R., RADFORD-SMITH, G., MATHEW, C. G., RIOUX, J. D., SCHADT, E. E., et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature,* 491**,** 119-124.

KABEERDOSS, J., DEVI, R. S., MARY, R. R. & RAMAKRISHNA, B. S. 2012. Faecal microbiota composition in vegetarians: comparison with omnivores in a cohort of young women in southern India. *Br J Nutr,* 108**,** 953-7.

KADOOKA, Y., SATO, M., IMAIZUMI, K., OGAWA, A., IKUYAMA, K., AKAI, Y., OKANO, M., KAGOSHIMA, M. & TSUCHIDA, T. 2010. Regulation of abdominal adiposity by probiotics (Lactobacillus gasseri SBT2055) in adults with obese tendencies in a randomized controlled trial. *Eur J Clin Nutr,* 64**,** 636-43.

KARLSSON, F., TREMAROLI, V., NIELSEN, J. & BACKHED, F. 2013. Assessing the human gut microbiota in metabolic diseases. *Diabetes,* 62**,** 3341-9.

KELLY, C. R., KAHN, S., KASHYAP, P., LAINE, L., RUBIN, D., ATREJA, A., MOORE, T. & WU, G. 2015. Update on Fecal Microbiota Transplantation 2015: Indications, Methodologies, Mechanisms, and Outlook. *Gastroenterology,* 149**,** 223-37.

KHAN, M. J., GERASIMIDIS, K., EDWARDS, C. A. & SHAIKH, M. G. 2016. Role of Gut Microbiota in the Aetiology of Obesity: Proposed Mechanisms and Review of the Literature. *Journal of Obesity,* 2016**,** 27.

KHORUTS, A., DICKSVED, J., JANSSON, J. K. & SADOWSKY, M. J. 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol,* 44**,** 354-60.

KLAENHAMMER, T. R. 1988. Bacteriocins of lactic acid bacteria. *Biochimie,* 70**,** 337-349.

KOLEVA, P. T., VALCHEVA, R. S., SUN, X., GÄNZLE, M. G. & DIELEMAN, L. A. 2012. Inulin and fructo-oligosaccharides have divergent effects on colitis and commensal microbiota in HLA-B27 transgenic rats. *British Journal of Nutrition,* 108**,** 1633-1643.

KOSTIC, A. D., CHUN, E., ROBERTSON, L., GLICKMAN, J. N., GALLINI, C. A., MICHAUD, M., CLANCY, T. E., CHUNG, D. C., LOCHHEAD, P., HOLD, G. L., EL-OMAR, E. M., BRENNER, D., FUCHS, C. S., MEYERSON, M. & GARRETT, W. S. 2013. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe,* 14**,** 207-15.

KOSTIC, A. D., GEVERS, D., PEDAMALLU, C. S., MICHAUD, M., DUKE, F., EARL, A. M., OJESINA, A. I., JUNG, J., BASS, A. J., TABERNERO, J., BASELGA, J., LIU, C., SHIVDASANI, R. A., OGINO, S., BIRREN, B. W., HUTTENHOWER, C., GARRETT, W. S. & MEYERSON, M. 2012. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res,* 22**,** 292-8.

LAKSHMINARAYANAN, B., GUINANE, C. M., O'CONNOR, P. M., COAKLEY, M., HILL, C., STANTON, C., O'TOOLE, P. W. & ROSS, R. P. 2013. Isolation and characterization of bacteriocin-producing bacteria from the intestinal microbiota of elderly Irish subjects. *Journal of Applied Microbiology,* 114**,** 886-898.

LARSEN, N., VOGENSEN, F. K., VAN DEN BERG, F. W., NIELSEN, D. S., ANDREASEN, A. S., PEDERSEN, B. K., AL-SOUD, W. A., SORENSEN, S. J., HANSEN, L. H. & JAKOBSEN, M. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One,* 5**,** e9085.

LAWTON, E. M., COTTER, P. D., HILL, C. & ROSS, R. P. 2007. Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile Bacillus halodurans C-125. *FEMS Microbiol Lett,* 267**,** 64-71.

LEPAGE, P., HÄSLER, R., SPEHLMANN, M. E., REHMAN, A., ZVIRBLIENE, A., BEGUN, A., OTT, S., KUPCINSKAS, L., DORÉ, J. & RAEDLER, A. 2011. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology,* 141**,** 227-236.

LEWIS, S., BRAZIER, J., BEARD, D., NAZEM, N. & PROCTOR, D. 2005. Effects of metronidazole and oligofructose on faecal concentrations of sulphate-reducing bacteria and their activity in human volunteers. *Scandinavian journal of gastroenterology,* 40**,** 1296-1303.

LEY, R. E., BÄCKHED, F., TURNBAUGH, P., LOZUPONE, C. A., KNIGHT, R. D. & GORDON, J. I. 2005. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 11070-11075.

LEY, R. E., HAMADY, M., LOZUPONE, C., TURNBAUGH, P. J., RAMEY, R. R., BIRCHER, J. S., SCHLEGEL, M. L., TUCKER, T. A., SCHRENZEL, M. D. & KNIGHT, R. 2008. Evolution of mammals and their gut microbes. *Science,* 320**,** 1647-1651.

LHOSTE, E. F., MOUZON, B., ANDRIEUX, C., GUEUGNEAU, A.-M., FISZLEWICZ, M., LE, E., CORRING, T. & SZYLIT, O. 1998. Physiological effects of a pea protein isolate in gnotobiotic rats: comparison with a soybean isolate and meat. *Annals of nutrition and metabolism,* 42**,** 44-54.

LI, X. & O'SULLIVAN, D. J. 2012. Contribution of the Actinobacteria to the growing diversity of lantibiotics. *Biotechnol Lett,* 34**,** 2133-45.

LIN, H. C. 2004. Small intestinal bacterial overgrowth: a framework for understanding irritable bowel syndrome. *Jama,* 292**,** 852-8.

LISZT, K., ZWIELEHNER, J., HANDSCHUR, M., HIPPE, B., THALER, R. & HASLBERGER, A. G. 2009. Characterization of bacteria, clostridia and Bacteroides in faeces of vegetarians using qPCR and PCR-DGGE fingerprinting. *Ann Nutr Metab,* 54**,** 253-7.

LYSSENKO, V., JONSSON, A., ALMGREN, P., PULIZZI, N., ISOMAA, B., TUOMI, T., BERGLUND, G., ALTSHULER, D., NILSSON, P. & GROOP, L. 2008. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med,* 359**,** 2220-32.

MACFARLANE, G., CUMMINGS, J. & ALLISON, C. 1986. Protein degradation by human intestinal bacteria. *Journal of general microbiology,* 132**,** 1647-1656.

MACFARLANE, G. T. & MACFARLANE, S. 2012. Bacteria, colonic fermentation, and gastrointestinal health. *Journal of AOAC International,* 95**,** 50-60.

MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., EECKHAUT, V., BALLET, V., CLAES, K., VAN IMMERSEEL, F., VERBEKE, K., FERRANTE, M., VERHAEGEN, J., RUTGEERTS, P. & VERMEIRE, S. 2013. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut*.

MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., EECKHAUT, V., BALLET, V., CLAES, K., VAN IMMERSEEL, F., VERBEKE, K., FERRANTE, M., VERHAEGEN, J., RUTGEERTS, P. & VERMEIRE, S. 2014. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63**,** 1275-83.

MAGEE, E. A., RICHARDSON, C. J., HUGHES, R. & CUMMINGS, J. H. 2000. Contribution of dietary protein to sulfide production in the large intestine: an in vitro and a controlled feeding study in humans. *The American journal of clinical nutrition,* 72**,** 1488-1494.

MANRIQUE, P., BOLDUC, B., WALK, S. T., VAN DER OOST, J., DE VOS, W. M. & YOUNG, M. J. 2016. Healthy human gut phageome. *Proceedings of the National Academy of Sciences,* 113**,** 10400-10405.

MARSH, A. J., O'SULLIVAN, O., ROSS, R. P., COTTER, P. D. & HILL, C. 2010. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics,* 11**,** 679.

MARTIN, W. 2009. Encapsulated Medicines for Iatrogenic Diseases. *British Patent Application: GB0916335,* 3.

MARTINEZ, F. A., BALCIUNAS, E. M., CONVERTI, A., COTTER, P. D. & DE SOUZA OLIVEIRA, R. P. 2013. Bacteriocin production by Bifidobacterium spp. A review. *Biotechnol Adv,* 31**,** 482-8.

MATIJASIC, B. B., OBERMAJER, T., LIPOGLAVSEK, L., GRABNAR, I., AVGUSTIN, G. & ROGELJ, I. 2013. Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *Eur J Nutr*.

MÄTTÖ, J., MAUNUKSELA, L., KAJANDER, K., PALVA, A., KORPELA, R., KASSINEN, A. & SAARELA, M. 2005. Composition and temporal stability of gastrointestinal microbiota in irritable bowel syndrome – a longitudinal study in IBS and control subjects. *FEMS Immunology & Medical Microbiology,* 43**,** 213-222.

MAYO, B. & VAN SINDEREN, D. 2010. Bifidobacteria: Genomics and Molecular Aspects. Caister Academic.

MAZMANIAN, S. K., ROUND, J. L. & KASPER, D. L. 2008. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature,* 453**,** 620-625.

MCALLAN, L., SKUSE, P., COTTER, P. D., O' CONNOR, P., CRYAN, J. F., ROSS, R. P., FITZGERALD, G., ROCHE, H. M. & NILAWEERA, K. N. 2014. Whey protein isolate at varying doses differentially influences energy balance and the composition of the gut microbiota in high-fat fed mice. *PLoS One***,** (in print).

MCCLERREN, A. L., COOPER, L. E., QUAN, C., THOMAS, P. M., KELLEHER, N. L. & VAN DER DONK, W. A. 2006. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 17243-17248.

MCCOY, A. N., ARAUJO-PEREZ, F., AZCARATE-PERIL, A., YEH, J. J., SANDLER, R. S. & KEKU, T. O. 2013. Fusobacterium is associated with colorectal adenomas. *PLoS One,* 8**,** e53653.

MCNULTY, N. P., YATSUNENKO, T., HSIAO, A., FAITH, J. J., MUEGGE, B. D., GOODMAN, A. L., HENRISSAT, B., OOZEER, R., COOLS-PORTIER, S., GOBERT, G., CHERVAUX, C., KNIGHTS, D., LOZUPONE, C. A., KNIGHT, R., DUNCAN, A. E., BAIN, J. R., MUEHLBAUER, M. J., NEWGARD, C. B., HEATH, A. C. & GORDON, J. I. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med,* 3**,** 106ra106.

MENDEZ-VILAS, A. & ANTONIO, M. 2011. *Science and Technology Against Microbial Pathogens: Research, Development and Evaluation, Proceedings of the International Conference on Antimicrobial Research (ICAR2010), Valladolid, Spain 3 - 5 November 2010,* World Scientific Publishing Company Pte Limited.

MIYOSHI, M., OGAWA, A., HIGURASHI, S. & KADOOKA, Y. 2013. Anti-obesity effect of Lactobacillus gasseri SBT2055 accompanied by inhibition of pro-inflammatory gene expression in the visceral adipose tissue in diet-induced obese mice. *Eur J Nutr*.

MORONI, O., KHEADR, E., BOUTIN, Y., LACROIX, C. & FLISS, I. 2006. Inactivation of Adhesion and Invasion of Food-Borne Listeria monocytogenes by Bacteriocin-Producing Bifidobacterium Strains of Human Origin. *Applied and Environmental Microbiology,* 72**,** 6894-6901.

MOSSIE, K., JONES, D., ROBB, F. & WOODS, D. 1979. Characterization and mode of action of a bacteriocin produced by a Bacteroides fragilis strain. *Antimicrobial agents and chemotherapy,* 16**,** 724-730.

MULLANE, K., LEE, C., BRESSLER, A., BUITRAGO, M., WEISS, K., DABOVIC, K., PRAESTGAARD, J., LEEDS, J. A., BLAIS, J. & PERTEL, P. 2014. A multi-center, randomized clinical trial to compare the safety and efficacy of LFF571 and vancomycin for Clostridium difficile infections. *Antimicrobial Agents and Chemotherapy*.

MURPHY, E. F., COTTER, P. D., HEALY, S., MARQUES, T. M., O'SULLIVAN, O., FOUHY, F., CLARKE, S. F., O'TOOLE, P. W., QUIGLEY, E. M., STANTON, C., ROSS, P. R., O'DOHERTY, R. M. & SHANAHAN, F. 2010. Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut,* 59**,** 1635-1642.

MURPHY, E. F., COTTER, P. D., HOGAN, A., O'SULLIVAN, O., JOYCE, A., FOUHY, F., CLARKE, S. F., MARQUES, T. M., O'TOOLE, P. W., STANTON, C., QUIGLEY, E. M. M., DALY, C., ROSS, P. R., O'DOHERTY, R. M. & SHANAHAN, F. 2013. Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. *Gut,* 62**,** 220-226.

MURPHY, K., O'SULLIVAN, O., REA, M. C., COTTER, P. D., ROSS, R. P. & HILL, C. 2011. Genome Mining for Radical SAM Protein Determinants Reveals Multiple Sactibiotic-Like Gene Clusters. *PLoS ONE,* 6**,** e20852.

MUSSO, G., GAMBINO, R. & CASSADER, M. 2011. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu Rev Med,* 62**,** 361-80.

NADAL, I., SANTACRUZ, A., MARCOS, A., WARNBERG, J., GARAGORRI, J. M., MORENO, L. A., MARTIN-MATILLAS, M., CAMPOY, C., MARTI, A., MOLERES, A., DELGADO, M., VEIGA, O. L., GARCIA-FUENTES, M., REDONDO, C. G. & SANZ, Y. 2009. Shifts in clostridia, bacteroides and immunoglobulin-coating fecal bacteria associated with weight loss in obese adolescents. *Int J Obes (Lond),* 33**,** 758-67.

NAKANO, V., IGNACIO, A., FERNANDES, M. R., FUKUGAITI, M. H. & AVILA-CAMPOS, M. J. 2006. Intestinal Bacteroides and Parabacteroides species producing antagonistic substances. *Microbiology,* 1**,** 61-64.

NEYRINCK, A. M., POSSEMIERS, S., VERSTRAETE, W., DE BACKER, F., CANI, P. D. & DELZENNE, N. M. 2012. Dietary modulation of clostridial cluster XIVa gut bacteria (Roseburia spp.) by chitin-glucan fiber improves host metabolic alterations induced by high-fat diet in mice. *J Nutr Biochem,* 23**,** 51-9.

O'HARA, A. M. & SHANAHAN, F. 2006. The gut flora as a forgotten organ. *EMBO Rep,* 7**,** 688-93.

O'SHEA, E. F., GARDINER, G. E., O'CONNOR, P. M., MILLS, S., ROSS, R. P. & HILL, C. 2009. Characterization of enterocin- and salivaricin-producing lactic acid bacteria from the mammalian gastrointestinal tract. *FEMS Microbiology Letters,* 291**,** 24-34.

O'SHEA, E. F., O'CONNOR, P. M., O'SULLIVAN, O., COTTER, P. D., ROSS, R. P. & HILL, C. 2013. Bactofencin A, a New Type of Cationic Bacteriocin with Unusual Immunity. *mBio,* 4.

O'KEEFE, S. J. D., LI, J. V., LAHTI, L., OU, J., CARBONERO, F., MOHAMMED, K., POSMA, J. M., KINROSS, J., WAHL, E., RUDER, E., VIPPERLA, K.,

NAIDOO, V., MTSHALI, L., TIMS, S., PUYLAERT, P. G. B., DELANY, J., KRASINSKAS, A., BENEFIEL, A. C., KASEB, H. O., NEWTON, K., NICHOLSON, J. K., DE VOS, W. M., GASKINS, H. R. & ZOETENDAL, E. G. 2015. Fat, fibre and cancer risk in African Americans and rural Africans. *Nature Communications,* 6**,** 6342.

O'TOOLE, P. W., MARCHESI, J. R. & HILL, C. 2017. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nature Microbiology,* 2**,** 17057.

OHMAN, L. & SIMREN, M. 2013. Intestinal microbiota and its role in irritable bowel syndrome (IBS). *Curr Gastroenterol Rep,* 15**,** 323.

PALMER, C., BIK, E. M., DIGIULIO, D. B., RELMAN, D. A. & BROWN, P. O. 2007. Development of the human infant intestinal microbiota. *PLoS Biol,* 5**,** e177.

PARK, D. Y., AHN, Y. T., PARK, S. H., HUH, C. S., YOO, S. R., YU, R., SUNG, M. K., MCGREGOR, R. A. & CHOI, M. S. 2013. Supplementation of Lactobacillus curvatus HY7601 and Lactobacillus plantarum KY1032 in diet-induced obese mice is associated with gut microbial changes and reduction in obesity. *PLoS One,* 8**,** e59470.

PATTERSON, E., MARQUES, T. M., O'SULLIVAN, O., FITZGERALD, P., FITZGERALD, G. F., COTTER, P. D., DINAN, T. G., CRYAN, J. F., STANTON, C. & ROSS, R. P. 2015. Streptozotocin-induced type-1-diabetes disease onset in Sprague–Dawley rats is associated with an altered intestinal microbiota composition and decreased diversity. *Microbiology,* 161**,** 182-193.

PEREZ-COBAS, A. E., ARTACHO, A., KNECHT, H., FERRUS, M. L., FRIEDRICHS, A., OTT, S. J., MOYA, A., LATORRE, A. & GOSALBES, M.

J. 2013. Differential effects of antibiotic therapy on the structure and function of human gut microbiota. *PLoS One,* 8**,** e80201.

PETROF, E. O., GLOOR, G. B., VANNER, S. J., WEESE, S. J., CARTER, D., DAIGNEAULT, M. C., BROWN, E. M., SCHROETER, K. & ALLEN-VERCOE, E. 2013. Stool substitute transplant therapy for the eradication of Clostridium difficile infection:'RePOOPulating'the gut. *Microbiome,* 1**,** 1-12.

PIMENTEL, M., LEMBO, A., CHEY, W. D., ZAKKO, S., RINGEL, Y., YU, J., MAREYA, S. M., SHAW, A. L., BORTEY, E. & FORBES, W. P. 2011. Rifaximin therapy for patients with irritable bowel syndrome without constipation. *N Engl J Med,* 364**,** 22-32.

PINEIRO, M. & STANTON, C. 2007. Probiotic Bacteria: Legislative Framework—Requirements to Evidence Basis. *The Journal of Nutrition,* 137**,** 850S-853S.

PIPER, C., COTTER, P. D., ROSS, R. P. & HILL, C. 2009. Discovery of medically significant lantibiotics. *Curr Drug Discov Technol,* 6**,** 1-18.

PLOTTEL, C. S. & BLASER, M. J. 2011. Microbiome and malignancy. *Cell Host Microbe,* 10**,** 324-35.

PLOVIER, H., EVERARD, A., DRUART, C., DEPOMMIER, C., VAN HUL, M., GEURTS, L., CHILLOUX, J., OTTMAN, N., DUPARC, T., LICHTENSTEIN, L., MYRIDAKIS, A., DELZENNE, N. M., KLIEVINK, J., BHATTACHARJEE, A., VAN DER ARK, K. C. H., AALVINK, S., MARTINEZ, L. O., DUMAS, M.-E., MAITER, D., LOUMAYE, A., HERMANS, M. P., THISSEN, J.-P., BELZER, C., DE VOS, W. M. & CANI, P. D. 2017. A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat Med,* 23**,** 107-113.

QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., MANICHANH, C., NIELSEN, T., PONS, N., LEVENEZ, F., YAMADA, T., MENDE, D. R., LI, J., XU, J., LI, S., LI, D., CAO, J., WANG, B., LIANG, H., ZHENG, H., XIE, Y., TAP, J., LEPAGE, P., BERTALAN, M., BATTO, J.-M., HANSEN, T., LE PASLIER, D., LINNEBERG, A., NIELSEN, H. B., PELLETIER, E., RENAULT, P., SICHERITZ-PONTEN, T., TURNER, K., ZHU, H., YU, C., LI, S., JIAN, M., ZHOU, Y., LI, Y., ZHANG, X., LI, S., QIN, N., YANG, H., WANG, J., BRUNAK, S., DORE, J., GUARNER, F., KRISTIANSEN, K., PEDERSEN, O., PARKHILL, J., WEISSENBACH, J., BORK, P., EHRLICH, S. D. & WANG, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464**,** 59-65.

QIN, J., LI, Y., CAI, Z., LI, S., ZHU, J., ZHANG, F., LIANG, S., ZHANG, W., GUAN, Y., SHEN, D., PENG, Y., ZHANG, D., JIE, Z., WU, W., QIN, Y., XUE, W., LI, J., HAN, L., LU, D., WU, P., DAI, Y., SUN, X., LI, Z., TANG, A., ZHONG, S., LI, X., CHEN, W., XU, R., WANG, M., FENG, Q., GONG, M., YU, J., ZHANG, Y., ZHANG, M., HANSEN, T., SANCHEZ, G., RAES, J., FALONY, G., OKUDA, S., ALMEIDA, M., LECHATELIER, E., RENAULT, P., PONS, N., BATTO, J. M., ZHANG, Z., CHEN, H., YANG, R., ZHENG, W., LI, S., YANG, H., WANG, J., EHRLICH, S. D., NIELSEN, R., PEDERSEN, O., KRISTIANSEN, K. & WANG, J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature,* 490**,** 55-60.

RAJILIC-STOJANOVIC, M., BIAGI, E., HEILIG, H. G., KAJANDER, K., KEKKONEN, R. A., TIMS, S. & DE VOS, W. M. 2011. Global and deep

molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology,* 141**,** 1792-801.

RAMNANI, P., GAUDIER, E., BINGHAM, M., VAN BRUGGEN, P., TUOHY, K. M. & GIBSON, G. R. 2010. Prebiotic effect of fruit and vegetable shots containing Jerusalem artichoke inulin: a human intervention study. *Br J Nutr,* 104**,** 233-240.

REA, M. C., DOBSON, A., O'SULLIVAN, O., CRISPIE, F., FOUHY, F., COTTER, P. D., SHANAHAN, F., KIELY, B., HILL, C. & ROSS, R. P. 2011. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences,* 108**,** 4639-4644.

REA, M. C., SIT, C. S., CLAYTON, E., O'CONNOR, P. M., WHITTAL, R. M., ZHENG, J., VEDERAS, J. C., ROSS, R. P. & HILL, C. 2010. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *Proceedings of the National Academy of Sciences,* 107**,** 9352-9357.

RESPONDEK, F., SWANSON, K. S., BELSITO, K. R., VESTER, B. M., WAGNER, A., ISTASSE, L. & DIEZ, M. 2008. Short-chain fructooligosaccharides influence insulin sensitivity and gene expression of fat tissue in obese dogs. *The Journal of nutrition,* 138**,** 1712-1718.

REUTER, G. 1971. Designation of Type Strains for Bifidobacterium Species. *International Journal of Systematic Bacteriology,* 21**,** 273-275.

RIBOULET-BISSON, E., STURME, M. H. J., JEFFERY, I. B., O'DONNELL, M. M., NEVILLE, B. A., FORDE, B. M., CLAESSON, M. J., HARRIS, H., GARDINER, G. E., CASEY, P. G., LAWLOR, P. G., O'TOOLE, P. W. &

ROSS, R. P. 2012. Effect of Lactobacillus salivarius Bacteriocin Abp118 on the Mouse and Pig Intestinal Microbiota. *PLoS ONE,* 7**,** e31113.

RIDAURA, V. K., FAITH, J. J., REY, F. E., CHENG, J., DUNCAN, A. E., KAU, A. L., GRIFFIN, N. W., LOMBARD, V., HENRISSAT, B., BAIN, J. R., MUEHLBAUER, M. J., ILKAYEVA, O., SEMENKOVICH, C. F., FUNAI, K., HAYASHI, D. K., LYLE, B. J., MARTINI, M. C., URSELL, L. K., CLEMENTE, J. C., VAN TREUREN, W., WALTERS, W. A., KNIGHT, R., NEWGARD, C. B., HEATH, A. C. & GORDON, J. I. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science,* 341**,** 1241214.

RIJKERS, G. T., BENGMARK, S., ENCK, P., HALLER, D., HERZ, U., KALLIOMAKI, M., KUDO, S., LENOIR-WIJNKOOP, I., MERCENIER, A., MYLLYLUOMA, E., RABOT, S., RAFTER, J., SZAJEWSKA, H., WATZL, B., WELLS, J., WOLVERS, D. & ANTOINE, J. M. 2010. Guidance for substantiating the evidence for beneficial effects of probiotics: current status and recommendations for future research. *J Nutr,* 140**,** 671s-6s.

RILEY, M. A. 1998. Molecular mechanisms of bacteriocin evolution. *Annual review of genetics,* 32**,** 255-278.

RIORDAN, S. M. & KIM, R. 2006. Bacterial overgrowth as a cause of irritable bowel syndrome. *Curr Opin Gastroenterol,* 22**,** 669-73.

ROBERFROID, M., GIBSON, G. R., HOYLES, L., MCCARTNEY, A. L., RASTALL, R., ROWLAND, I., WOLVERS, D., WATZL, B., SZAJEWSKA, H., STAHL, B., GUARNER, F., RESPONDEK, F., WHELAN, K., COXAM, V., DAVICCO, M. J., LEOTOING, L., WITTRANT, Y., DELZENNE, N. M.,

CANI, P. D., NEYRINCK, A. M. & MEHEUST, A. 2010. Prebiotic effects: metabolic and health benefits. *Br J Nutr,* 104 Suppl 2**,** S1-63.

RODES, L., SAHA, S. & TOMARO-DUCHESNEAU, C. 2014. Microencapsulated Bifidobacterium longum subsp. infantis ATCC 15697 favorably modulates gut microbiota and reduces circulating endotoxins in F344 rats. 2014**,** 602832.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-5.

SAARELA, M., MOGENSEN, G., FONDÉN, R., MÄTTÖ, J. & MATTILA-SANDHOLM, T. 2000. Probiotic bacteria: safety, functional and technological properties. *Journal of Biotechnology,* 84**,** 197-215.

SÁNCHEZ, E., DONAT, E., RIBES-KONINCKX, C., CALABUIG, M. & SANZ, Y. 2010. Intestinal Bacteroides species associated with coeliac disease. *Journal of clinical pathology,* 63**,** 1105-1111.

SANTACRUZ, A., COLLADO, M. C., GARCIA-VALDES, L., SEGURA, M. T., MARTIN-LAGOS, J. A., ANJOS, T., MARTI-ROMERO, M., LOPEZ, R. M., FLORIDO, J., CAMPOY, C. & SANZ, Y. 2010. Gut microbiota composition is associated with body weight, weight gain and biochemical parameters in pregnant women. *Br J Nutr,* 104**,** 83-92.

SANTACRUZ, A., MARCOS, A., WARNBERG, J., MARTI, A., MARTIN-MATILLAS, M., CAMPOY, C., MORENO, L. A., VEIGA, O., REDONDO-FIGUERO, C., GARAGORRI, J. M., AZCONA, C., DELGADO, M., GARCIA-FUENTES, M., COLLADO, M. C. & SANZ, Y. 2009. Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity (Silver Spring),* 17**,** 1906-15.

SCHOLZ, R., MOLOHON, K. J., NACHTIGALL, J., VATER, J., MARKLEY, A. L., SUSSMUTH, R. D., MITCHELL, D. A. & BORRISS, R. 2011. Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42. *J Bacteriol,* 193**,** 215-24.

SCOTT, L. J., MOHLKE, K. L., BONNYCASTLE, L. L., WILLER, C. J., LI, Y., DUREN, W. L., ERDOS, M. R., STRINGHAM, H. M., CHINES, P. S., JACKSON, A. U., PROKUNINA-OLSSON, L., DING, C. J., SWIFT, A. J., NARISU, N., HU, T., PRUIM, R., XIAO, R., LI, X. Y., CONNEELY, K. N., RIEBOW, N. L., SPRAU, A. G., TONG, M., WHITE, P. P., HETRICK, K. N., BARNHART, M. W., BARK, C. W., GOLDSTEIN, J. L., WATKINS, L., XIANG, F., SARAMIES, J., BUCHANAN, T. A., WATANABE, R. M., VALLE, T. T., KINNUNEN, L., ABECASIS, G. R., PUGH, E. W., DOHENY, K. F., BERGMAN, R. N., TUOMILEHTO, J., COLLINS, F. S. & BOEHNKE, M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science,* 316**,** 1341-5.

SEKIROV, I., RUSSELL, S. L., ANTUNES, L. C. M. & FINLAY, B. B. 2010. Gut Microbiota in Health and Disease. *Physiological Reviews,* 90**,** 859-904.

SELA, D. A., CHAPMAN, J., ADEUYA, A., KIM, J. H., CHEN, F., WHITEHEAD, T. R., LAPIDUS, A., ROKHSAR, D. S., LEBRILLA, C. B., GERMAN, J. B., PRICE, N. P., RICHARDSON, P. M. & MILLS, D. A. 2008. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A,* 105**,** 18964-9.

SELLON, R. K., TONKONOGY, S., SCHULTZ, M., DIELEMAN, L. A., GRENTHER, W., BALISH, E., RENNICK, D. M. & SARTOR, R. B. 1998.

Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection and immunity,* 66**,** 5224-5231.

SENDER, R., FUCHS, S. & MILO, R. 2016. Revised estimates for the number of human and bacteria cells in the body. *bioRxiv*.

SHANKS, O. C., PEED, L., SIVAGANESAN, M., HAUGLAND, R. A. & CHERN, E. C. 2014. Human Fecal Source Identification with Real-Time Quantitative PCR. *Environmental Microbiology.* Springer.

SINGH, M. & SAREEN, D. 2014. Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One,* 9**,** e91352.

SLAVIN, J. 2013. Fiber and Prebiotics: Mechanisms and Health Benefits. *Nutrients,* 5**,** 1417-1435.

SOKOL, H. & SEKSIK, P. 2010. The intestinal microbiota in inflammatory bowel diseases: time to connect with the host. *Curr Opin Gastroenterol,* 26**,** 327-31.

SPEHLMANN, M. E., BEGUN, A. Z., BURGHARDT, J., LEPAGE, P., RAEDLER, A. & SCHREIBER, S. 2008. Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflammatory Bowel Diseases,* 14**,** 968-976.

SPRONG, R., SCHONEWILLE, A. & VAN DER MEER, R. 2010. Dietary cheese whey protein protects rats against mild dextran sulfate sodium–induced colitis: Role of mucin and microbiota. *Journal of dairy science,* 93**,** 1364-1371.

TAHARA, T., YAMAMOTO, E., SUZUKI, H., MARUYAMA, R., CHUNG, W., GARRIGA, J., JELINEK, J., YAMANO, H. O., SUGAI, T., AN, B., SHUREIQI, I., TOYOTA, M., KONDO, Y., ESTECIO, M. R. & ISSA, J. P.

2014. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res*.

THE HUMAN MICROBIOME PROJECT CONSORTIUM 2012. Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-14.

TULINI, F. L., LOHANS, C. T., BORDON, K. C. F., ZHENG, J., ARANTES, E. C., VEDERAS, J. C. & DE MARTINIS, E. C. P. 2014. Purification and characterization of antimicrobial peptides from fish isolate Carnobacterium maltaromaticum C2: Carnobacteriocin X and carnolysins A1 and A2. *International Journal of Food Microbiology,* 173**,** 81-88.

TURNBAUGH, P. J., BÄCKHED, F., FULTON, L. & GORDON, J. I. 2008. Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe,* 3**,** 213-223.

TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. & GORDON, J. I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* 444**,** 1027-31.

TURNBAUGH, P. J., RIDAURA, V. K., FAITH, J. J., REY, F. E., KNIGHT, R. & GORDON, J. I. 2009. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine,* 1**,** 6ra14.

TVEDE, M. & RASK-MADSEN, J. 1989. Bacteriotherapy for chronic relapsing Clostridium difficile diarrhoea in six patients. *Lancet,* 1**,** 1156-60.

VAN HEEL, A. J., DE JONG, A., MONTALBAN-LOPEZ, M., KOK, J. & KUIPERS, O. P. 2013. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res,* 41**,** W448-53.

VAN KRAAIJ, C., M. DE VOS, W., J. SIEZEN, R. & P. KUIPERS, O. 1999. Lantibiotics: biosynthesis, mode of action and applications. *Natural Product Reports,* 16**,** 575-587.

VEERAPPAN, G. R., BETTERIDGE, J. & YOUNG, P. E. 2012. Probiotics for the treatment of inflammatory bowel disease. *Curr Gastroenterol Rep,* 14**,** 324-33.

VELÁSQUEZ, J. E. & VAN DER DONK, W. A. 2011. Genome mining for ribosomally synthesized natural products. *Current Opinion in Chemical Biology,* 15**,** 11-21.

VIEIRA, A. T., TEIXEIRA, M. M. & MARTINS, F. S. 2013. The Role of Probiotics and Prebiotics in Inducing Gut Immunity. *Front Immunol,* 4**,** 445.

VYAS, U. & RANGANATHAN, N. 2012. Probiotics, prebiotics, and synbiotics: gut and beyond. *Gastroenterol Res Pract,* 2012**,** 872716.

WALIGORA-DUPRIET, A.-J., CAMPEOTTO, F., NICOLIS, I., BONET, A., SOULAINES, P., DUPONT, C. & BUTEL, M.-J. 2007. Effect of oligofructose supplementation on gut microflora and well-being in young children attending a day care centre. *International journal of food microbiology,* 113**,** 108-113.

WALKER, A. W., DUNCAN, S. H., LEITCH, E. C. M., CHILD, M. W. & FLINT, H. J. 2005. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Applied and environmental microbiology,* 71**,** 3692-3700.

WALSH, C. J., GUINANE, C. M., HILL, C., ROSS, R. P., O'TOOLE, P. W. & COTTER, P. D. 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. *BMC Microbiology,* 15**,** 183.

WALSH, C. J., GUINANE, C. M., O'TOOLE, P. W. & COTTER, P. D. 2014. Beneficial modulation of the gut microbiota. *FEBS Lett,* 588**,** 4120-30.

WALSH, C. J., GUINANE, C. M., PW, O. T. & COTTER, P. D. 2017. A Profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome. *PeerJ,* 5**,** e3254.

WALSH, M. C., GARDINER, G. E., HART, O. M., LAWLOR, P. G., DALY, M., LYNCH, B., RICHERT, B. T., RADCLIFFE, S., GIBLIN, L. & HILL, C. 2008. Predominance of a bacteriocin-producing Lactobacillus salivarius component of a five-strain probiotic in the porcine ileum and effects on host immune phenotype. *FEMS microbiology ecology,* 64**,** 317-327.

WANG, H. T., CHEN, I. H. & HSU, J. T. 2012a. Production and characterization of a bacteriocin from ruminal bacterium Ruminococcus albus 7. *Biosci Biotechnol Biochem,* 76**,** 34-41.

WANG, T., CAI, G., QIU, Y., FEI, N., ZHANG, M., PANG, X., JIA, W., CAI, S. & ZHAO, L. 2012b. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *Isme j,* 6**,** 320-9.

WARREN, C. A. & GUERRANT, R. L. 2011. Pathogenic C difficile is here (and everywhere) to stay. *The Lancet,* 377**,** 8-9.

WEBER, T., BLIN, K., DUDDELA, S., KRUG, D., KIM, H. U., BRUCCOLERI, R., LEE, S. Y., FISCHBACH, M. A., MÜLLER, R., WOHLLEBEN, W., BREITLING, R., TAKANO, E. & MEDEMA, M. H. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research,* 43**,** W237-W243.

WELLCOME TRUST CASE CONTROL CONSORTIUM 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661-78.

WEXLER, H. M. 2007. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews,* 20**,** 593-621.

WU, G. D., CHEN, J., HOFFMANN, C., BITTINGER, K., CHEN, Y.-Y., KEILBAUGH, S. A., BEWTRA, M., KNIGHTS, D., WALTERS, W. A., KNIGHT, R., SINHA, R., GILROY, E., GUPTA, K., BALDASSANO, R., NESSEL, L., LI, H., BUSHMAN, F. D. & LEWIS, J. D. 2011. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science,* 334**,** 105-108.

XIAO, S., FEI, N., PANG, X., SHEN, J., WANG, L., ZHANG, B., ZHANG, M., ZHANG, X., ZHANG, C., LI, M., SUN, L., XUE, Z., WANG, J., FENG, J., YAN, F., ZHAO, N., LIU, J., LONG, W. & ZHAO, L. 2014. A gut microbiota-targeted dietary intervention for amelioration of chronic inflammation underlying metabolic syndrome. *FEMS Microbiology Ecology,* 87**,** 357-367.

YADAV, H., LEE, J. H., LLOYD, J., WALTER, P. & RANE, S. G. 2013. Beneficial metabolic effects of a probiotic via butyrate-induced GLP-1 hormone secretion. *J Biol Chem,* 288**,** 25088-97.

ZE, X., DUNCAN, S. H., LOUIS, P. & FLINT, H. J. 2012. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J,* 6**,** 1535-1543.

ZEEVI, D., KOREM, T., ZMORA, N., ISRAELI, D., ROTHSCHILD, D., WEINBERGER, A., BEN-YACOV, O., LADOR, D., AVNIT-SAGI, T., LOTAN-POMPAN, M., SUEZ, J., MAHDI, J. A., MATOT, E., MALKA, G., KOSOWER, N., REIN, M., ZILBERMAN-SCHAPIRA, G., DOHNALOVA,

L., PEVSNER-FISCHER, M., BIKOVSKY, R., HALPERN, Z., ELINAV, E. & SEGAL, E. 2015. Personalized Nutrition by Prediction of Glycemic Responses. *Cell,* 163**,** 1079-94.

ZHANG, C., LI, S., YANG, L., HUANG, P., LI, W., WANG, S., ZHAO, G., ZHANG, M., PANG, X., YAN, Z., LIU, Y. & ZHAO, L. 2013. Structural modulation of gut microbiota in life-long calorie-restricted mice. *Nat Commun,* 4.

ZHAO, X. & KUIPERS, O. P. 2016. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species. *BMC Genomics,* 17**,** 882.

ZHENG, J., GÄNZLE, M. G., LIN, X. B., RUAN, L. & SUN, M. 2014. Diversity and dynamics of bacteriocins from human microbiome. *Environmental Microbiology***,** 2133–2143.

Chapter 3

A Profile Hidden Markov Model to investigate the distribution and frequency of LanB-encoding lantibiotic modification genes in the human oral and gut microbiome

Included as published in PeerJ (doi: 10.7717/peerj.3254)

Abstract

The human microbiota plays a key role in health and disease, and bacteriocins, which are small, bacterially produced, antimicrobial peptides, are likely to have an important function in the stability and dynamics of this community. Here we examined the density and distribution of the subclass I lantibiotic modification protein, LanB, in human oral and stool microbiome datasets using a specially constructed profile Hidden Markov Model (HMM).

The model was validated by correctly identifying known *lanB* genes in the genomes of known bacteriocin producers more effectively than other methods, while being sensitive enough to differentiate between different subclasses of lantibiotic modification proteins. This approach was compared with two existing methods to screen both genomic and metagenomic datasets obtained from the Human Microbiome Project (HMP).

Of the methods evaluated, the new profile HMM identified the greatest number of putative LanB proteins in the stool and oral metagenome data while BlastP identified the fewest. In addition, the model identified more LanB proteins than a pre-existing Pfam lanthionine dehydratase model. Searching the gastrointestinal tract subset of the HMP reference genome database with the new HMM identified seven putative subclass I lantibiotic producers, including two members of the *Coprobacillus* genus.

These findings establish custom profile HMMs as a potentially powerful tool in the search for novel bioactive producers with the power to benefit human health, and reinforce the repertoire of apparent bacteriocin-encoding gene clusters that may have been overlooked by culture-dependent mining efforts to date.

Introduction

Bacteriocins are ribosomally synthesised peptides produced by bacteria that inhibit the growth of other bacteria. Some classes of bacteriocins are post-translationally modified to provide structures beyond those possible by ribosomal translation alone. These modifications are typically key to the peptide's functionality, stability and target recognition (Arnison et al. 2013). Class I bacteriocins, also known as lantibiotics, are one such class of small (<5 kDa) modified bacteriocins, possessing the characteristic thioester amino acids lanthionine or methyllanthionine (Perez et al. 2014). Lantibiotics form a subgroup within the larger lantipeptide family, which also includes peptides that lack antimicrobial activity. Lantipeptides can be divided into four different subclasses (I – IV) based on the distinct biosynthetic enzymes responsible for their posttranslational modification (Arnison et al. 2013).

The most commonly studied lantibiotic, Nisin, is a subclass I lantibiotic, meaning that the linear prepeptide is processed by a LanBC modification system (Arnison et al. 2013). The core peptide undergoes a two-step posttranslational modification catalysed by two distinct enzymes - the dehydratase LanB and the cyclase LanC (Xie & van der Donk 2004). The leader sequence, necessary for recognition by the modification enzymes in the two previous steps, is then removed by the protease LanP to produce the active lantibiotic (Xie & van der Donk 2004). The operon-encoded nature of bacteriocins and bacteriocin-like peptides makes them ideal candidates for genome mining. In the case of modified bacteriocins, the structural prepeptide coding sequence often appears alongside the genes encoding proteins responsible for its modification and export from the cell. However, as more bacteriocins are discovered, the heterogeneous nature of these prepeptides is becoming ever more apparent. This diversity, coupled with their small sequence length, makes

bacteriocin prepeptides much more difficult to detect using sequence-homology based searches like BLAST (Altschul et al. 1990). In an effort to address these obstacles, shifting the focus to the detection of bacteriocin-associated proteins opens up more avenues of discovery than simply searching for prepeptide homologs. This provides opportunities to better determine the frequency with which specific types of bacteriocin gene clusters can be found in different environmental niches, such as the human microbiota, through the investigation of metagenomic data.

It has been estimated that the human microbiota comprises approximately 100 trillion bacterial cells, outnumbering our own cells by a factor of 10 or more (Bäckhed et al. 2005). A recent publication, however, has argued that the ratio is actually more likely to be one-to-one, with the numbers being similar enough that each defecation event may alter the ratio to favour human cells over bacteria (Sender et al. 2016). Of greater consequence than bacterial numbers, however, is the collection of genes encoded in this metagenome, thought to be approximately 150 times larger than that of the human genome, with a functional potential far broader than that of its host (Qin et al. 2010). Regardless of absolute numbers, this dynamic community is thought to contain 100-1000 phylotypes (Faith et al. 2013; Qin et al. 2010) and play an integral role in human health and disease (Clemente et al. 2012; Flint et al. 2012). The human microbiota exhibits robust temporal stability (Belstrøm et al. 2016; Jeffery et al. 2016) perhaps due, in part, to the protection against invading bacteria conferred by bacteriocins and other antimicrobials produced *in situ* (Corr et al. 2007; Moroni et al. 2006; Rea et al. 2011a). As such, investigation of the density and diversity of bacteriocins produced in the microbiome of healthy individuals may shed light on beneficial and harmful members of this community, and key organisms for maintaining typical, i.e. health-associated, microbiota composition.

Mining the human microbiota, especially for antimicrobial compounds, has become a popular area of research in recent years (Donia et al. 2014; Walsh et al. 2015). Due to the availability of metagenomic data generated by large public funding initiatives such as the Human Microbiome Project in the U.S. (The Human Microbiome Project Consortium 2012) and the European MetaHIT consortium (Dusko Ehrlich 2010), *in silico* mining of data has emerged as a new tool that has the potential to identify antimicrobial-producing probiotics that can modulate the gut microbiota (Erejuwa et al. 2014; Walsh et al. 2014), or address the increasingly serious threat to public health caused by antimicrobial resistance. There are many available tools for mining the microbiome for antimicrobials, including BAGEL3 (van Heel et al. 2013), antiSMASH (Weber et al. 2015), and traditional sequence-based approaches like BLAST (Altschul et al. 1990). A feature commonly integrated into these tools are Hidden Markov Models (HMM) (Morton et al. 2015; van Heel et al. 2013; Weber et al. 2015), a statistical method often used to model biological data such as speech recognition, disease interaction and changes in gene expression in cancer (Gales & Young 2007; Seifert et al. 2014; Sherlock et al. 2013). Profile HMMs, a specific subset of HMMs, represent the patterns, motifs and other properties of a multiple sequence alignment by applying a statistical model to estimate the true frequency of a nucleotide or amino acid at a given position in the alignment from its observed frequency (Yoon 2009). Profile HMMs differ from general HMMs as they move strictly from left to right along the alignment and do not contain any cycles, a feature that makes them suitable for modelling nucleotide and protein sequence data, and have been notably utilized to detect viral protein sequences in metagenomic sequence data (Skewes-Cox et al. 2014). For each column in the multiple sequence alignment, the profile uses one of three types of hidden state - a *match state,* an insert state, or a delete state, to

describe residue frequencies, insertions, and deletions, respectively (Yoon 2009). Profile HMMs are potentially more sensitive than sequence homology approaches for identifying more distantly related proteins as they focus on function-dependent conserved motifs that are theoretically slower-evolving, as opposed to focusing on overall sequence similarity. Indeed, profile HMMs are known to typically outperform pairwise sequence comparison methods (such as BLAST) in the detection of distant homologs (Park et al. 1998), at the cost of greater computational requirements – particularly in alignment scoring and E-value calculation (Madera & Gough 2002). Correspondingly, speed is the main advantage of BLAST over profile HMMs, however, as it is a heuristic algorithm it does not guarantee identification of the optimal alignment between query and database sequences.

In this study we designed, validated and implemented a Profile HMM to search for putative subclass I lantibiotic gene clusters in the HMP metagenomes and compared its performance to some of the tools mentioned above.

Methods

Data Collection

HMASM (HMP Illumina WGS Assemblies) and HMRGD (HMP Reference Genomes Data) were downloaded from the Data Analysis and Coordination Centre for the HMP. 835 bacterial RefSeq protein sequences annotated as "lantibiotic dehydratase" were downloaded from NCBI Protein website (13 Apr 2015) in FASTA format (list available online at https://doi.org/10.7717/peerj.3254/supp-1).

Building and Validating the new Profile Hidden Markov Model

A global multiple sequence alignment was generated in the aligned-FASTA format using MUSCLE (v3.8.31) (Edgar 2004), and a profile HMM was built from the MSA aligned-FASTA file using the HMMER tool hmmbuild (v3.1b1 May 2013). For comparison of the new model's performance, HMMER3's hmmsearch tool was used, with default parameters, to search the Pfam lantibiotic dehydratase model PF04738 against the same stool and oral HMASM assemblies (the sequences used to build this model are available online at https://doi.org/10.7717/peerj.3254/supp-2). Positive and negative controls (listed in Table 1) were used to evaluate the model's ability to 1) accurately identify LanB protein sequences, and 2) distinguish LanB protein sequences from other, related, lantibiotic modification proteins (i.e. LanM, LanKC, and LanL. The controls were also screened using the PF04738 model, the web-based bacteriocin genome mining tool BAGEL3 (van Heel et al. 2013), and a traditional BlastP using the nisin-associated lanthionine dehydratase, NisB, as the driver sequence (GenBank accession number CAA79468.1) to compare the sensitivity

and specificity of each approach. A flowchart of the steps involved in building, validating and applying a profile HMM is depicted in Figure 1.

Target Sequence Translation

The HMMER3 hmmsearch tool only accepts protein sequences as targets for comparison to protein profile HMMs so a python script was created to translate the nucleotide sequences of the assembled genomes and metagenomes to be screened into protein sequences. The DNA nucleotide sequences were translated in six frames using the standard genetic code.

Metagenomic Screen

The HMMER3 tool hmmsearch was used, with default parameters, to search both the new LanB profile HMM and the Pfam PF04738 profile HMM (Punta et al. 2012) against the stool and oral subsets of the Human Microbiome Project's whole metagenomic shotgun sequencing assemblies (HMASM). 139 stool communities and 382 communities from eight different body sites within the oral cavity were screened from the HMP database. These are listed in Table 2. As an additional comparison of performance, a traditional BlastP screen was performed on the same metagenomic samples using the previously mentioned nisin-associated lanthionine dehydratase, NisB, driver sequence. A significance cutoff of $E \leq 1x10^{-5}$ was chosen for both profile HMM and BlastP methods.

Strain	Bacteriocin	Control	Subclass	pHMM	Pfam	BlastP	Bagel3
Lactococcus lactis ssp. *lactis* S0 [a,b,c,d]	Nisin Z	Positive	LanB	+	+	+	+
Lactococcus lactis ssp. *lactis* CV56 [a,b,c,d]	Nisin A	Positive	LanB	+	+	+	+
Lactococcus lactis ssp. *lactis* IO-1 [a,b,c,d]	Nisin Z	Positive	LanB	+	+	+	+
Bacillus subtilis subsp. *spizizienii* ATCC 6633 [a,b,c,d]	Subtilin	Positive	LanB	+	+	+	+
Staphylococcus aureus subsp. *aureus* USA300_FPR3757 [a,c,d]	Bsa	Positive	LanB	+	-	+	+
Streptococcus mutans CH43 [a,b,d]	Mutacin I	Positive	LanB	+	+	-	+
Streptococcus mutans UA787 [a,b,d]	Mutacin III	Positive	LanB	+	+	-	+
Streptococcus pyogenes [a,b,c,d]	Streptin	Positive	LanB	+	+	+	+
Staphylococcus epidermidis [a,b,c,d]	Pep5	Positive	LanB	+	+	+	+
Lactococcus lactis subsp. *lactis* KF147 [c]	None	Negative	-	-	-	+	-
Streptococcus mutans GS-5	Mutacin GS-5	Negative	LanM	-	-	-	-
Lactococcus lactis subsp. *lactis* plasmid pES2	Lacticin 481	Negative	LanM	-	-	-	-
Streptomyces cinnamoneus cinnamoneus DSM 4005	Cinnamycin	Negative	LanM	-	-	-	-
Bacillus paralichenformis APC 1576	Formicin	Negative	LanM	-	-	-	-
Streptococcus salivarius plasmid pSsal-K12	Salivaricin B	Negative	LanM	-	-	-	-
Streptomyces venezuelae ATCC 10712 [d]	Venezuelin	Negative	LanL	-	-	-	+

Table 1. Controls used in validation of the profile HMM, listing the lantibiotic produced, whether the strain was included as a positive or negative control, the subclass of modification protein responsible for lanthionine dehydration, and whether there was a significant hit identified by each method.

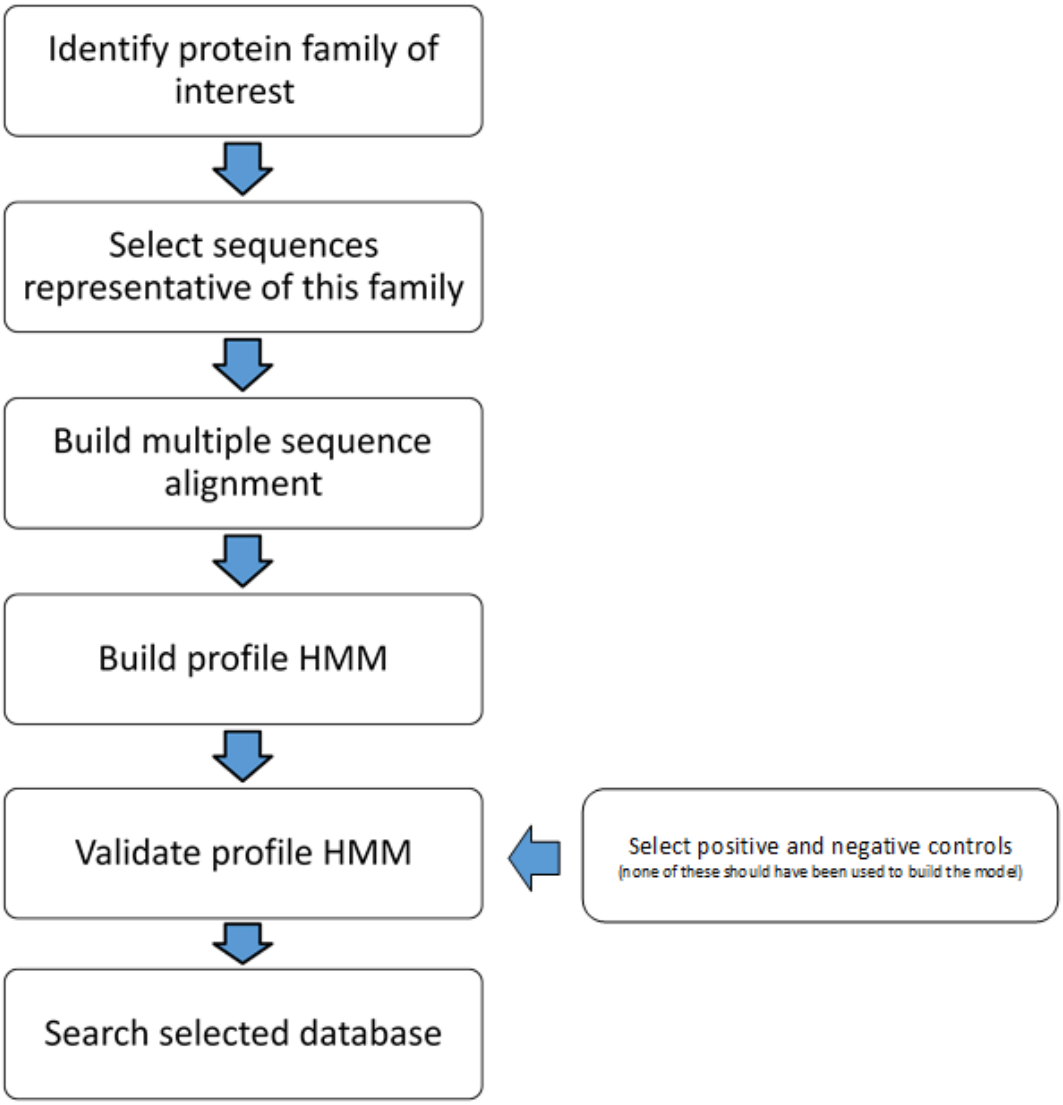


Figure 1. Flowchart depicting the step involved in building and validation of, and screening using, a profile HMM

Site	Number of Samples
Attached Keratinized Gingiva	6
Buccal Mucosa	107
Palatine Tonsils	6
Saliva	3
Stool	139
Subgingival Plaque	7
Supragingival Plaque	118
Throat	7
Tongue Dorsum	128

Table 2. Number of metagenomic samples per body site screened.

Manual Examination of Randomly Selected Gene Neighbourhoods

A subset of sixty hits were randomly selected and the surrounding region examined to identify other proteins involved in lantibiotic biosynthesis. Open Reading Frames were identified using Glimmer v3.02 (Delcher et al. 1999), which were then visualised using Artemis (Carver et al. 2012) and blasted against the nr database using BlastP (significance accepted as $E \leq 10^{-5}$).

Genomic Screen

HMMER3's hmmsearch tool was used, with default parameters, to search the new profile HMM against the draft genomes comprising the gastrointestinal tract subset of the Human Microbiome Project's reference genome database.

Taxonomic Classification of LanB-encoding contigs

Taxonomy was assigned to LanB-encoding contigs, as assigned by our pHMM, using Kaiju (Menzel et al. 2016). Analysis was performed in MEM run mode using default parameters and the NCBI non-redundant protein database.

Statistical Analysis

Statistical analysis was performed in R (v. 3.1.3) (R Core Team 2015).

Results

Validation of the Profile Hidden Markov Model

The ability of the newly developed profile HMM and the Pfam lantibiotic dehydratase model PF04738 to detect LanB-encoding genes were compared using the positive and negative controls listed in Table 1. The positive controls were selected based on a relevant book chapter (Rea et al. 2011b) and all are previously characterised bacteriocin producers for which the sequence of the relevant biosynthetic gene cluster was available. None of the positive control sequences were used in the building of the model and a graphical representation of these clusters is presented in Figure 2. *Lactococcus lactis* subsp. *lactis* KF147 was chosen as a negative control because it is of the same subspecies as three of the positive controls (*Lactococcus lactis* subsp. lactis S0, *Lactococcus lactis* subsp. *lactis* CV56 and *Lactococcus lactis* subsp. *lactis* IO-1) but does not produce a bacteriocin. *Streptococcus mutans* GS-5, *Streptomyces cinnamoneus cinnamoneus* DSM 4005, the *Lactococcus lactis* subsp. *lactis* IL1835 plasmid pES2, the *Streptococcus salivarious* plasmid pSsal-K12, and the newly characterised formicin producer *Bacillus paralicheniformis* APC 1576 were chosen as negative controls to evaluate the ability of the model to differentiate between LanB (subclass I) proteins and the LanM proteins-from these strains, which perform a similar, but distinct, function in the posttranslational modification of subclass II lantibiotics. *Streptomyces venezuelae* ATCC 10712 was chosen as the final negative control as it has been reported to produce a LanL-type lantipeptide (Goto et al. 2010). Examination of the ATCC 10712 genome using BAGEL3 identified several other orphan lantibiotic modification genes, including those encoding putative LanL, LanM, LanD and LanB proteins. The genome also appeared to encode a subclass III

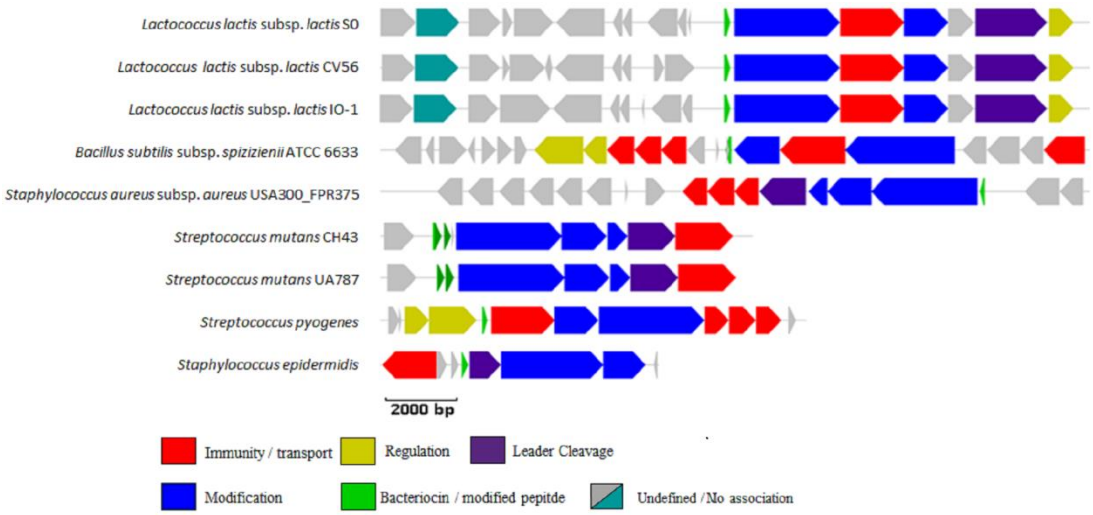


Figure 2. BAGEL3 output of putative bacteriocin gene clusters identified in the positive controls used for validation of our new profile HMM. Each predicted open reading frame is colour-coded based on the role it plays in lantibiotic biosynthesis.

lantipeptide cluster comprised of genes potentially encoding a structural protein, two ABC-type transporters and a LanKC modification protein. Notably, there have been no reports of subclass I lantibiotic production by ATCC 10712 despite an in-depth investigation into the strain's lantipeptide producing capability (Goto et al. 2010), and BAGEL3 identified no other lantibiotic-related genes in the area of interest leading us to determine that this was a false positive.

The newly developed LanB profile HMM correctly identified the LanB protein in all nine positive controls, while the PF04738 profile HMM correctly identified the LanB protein in eight of the nine positive controls, failing to detect the Bsa-associated LanB protein in *Staphylococcus aureus* subsp. *aureus* USA300_FPR3757. Both the LanB and PF04738 profile HMMs returned no false positives when searched against the seven negative controls used in this study.

The web version of BAGEL3 correctly identified the lantibiotic modification proteins in all positive and negative controls, excepting the aforementioned ATCC 10712-encoded LanB concluded to be a false positive. Interestingly, examination of these controls with the BlastP method described previously, failed to correctly identify the LanB proteins encoded by *Streptococcus mutans* CH43 and *Streptococcus mutans* UA787, although the former ($E = 3x10^{-4}$) fell just short of the significance cutoff ($E \leq 1x10\text{-}^{5}$). BlastP also incorrectly identified a LanB protein in the negative control *Lactococcus lactis* subsp. *lactis* KF147.

Metagenomic Screen

A search with the newly developed profile HMM against the HMASM database identified 399 hits from the stool metagenomes and 1169 hits from the oral metagenomes. In contrast, the PF04738 model identified 288 hits from the stool metagenomes and 686 from the oral metagenomes. Our model reported at least one putative lantibiotic gene cluster in 81% of oral metagenomes and 86% of stool metagenomes, compared to 73% and 76%, respectively, identified by the Pfam model. The distribution of hits per sample is presented in Figure 3. BlastP identified 231 hits from the stool metagenomes and 374 hits from the oral metagenomes. The results of these three approaches were compared to ascertain what proportion of significant hits was common to more than one search method. The results of this comparison are summarised in Figure 4 and show that the newly developed profile HMM identified the greatest number of lantibiotic modification genes in datasets from both body sites, while the BlastP approach identified the fewest.

The overall results of these combined screening approaches, illustrated in Figure 5 and summarised in Table 3, show a higher number and density of hits in the oral metagenomes than in the stool metagenomes (Wilcoxon rank sum test, p = 1.399e–05) and they also reveal a large variation in density of hits between the different sites within the oral metagenomes. This pattern was also reflected in four of the Oral subsites, namely Buccal Mucosa, Subgingival Plaque, Supragingival Plaque and Tongue Dorsum, all of which had a significantly higher LanB density than the Stool metagenomes (Wilcoxon rank sum test, (p = 0.0258,0.0014, 6.7e–09, and 9.4e–06, respectively). Within the Oral samples, our model revealed a large variation in density of hits between different subsites. The throat metagenomes had the lowest LanB density, and exhibited significantly lower densities than Buccal Mucosa

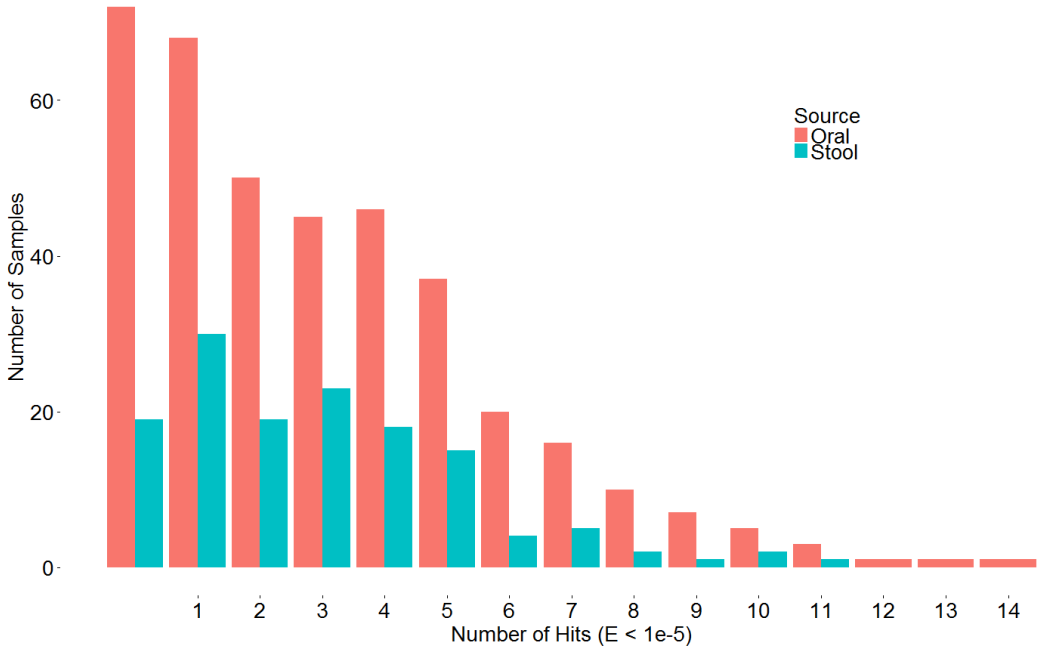


Figure 3. Barchart depicting the distribution of lanthionine dehydratase protein numbers identified by our new profile HMM in metagenomic samples from the stool and oral microbiota.

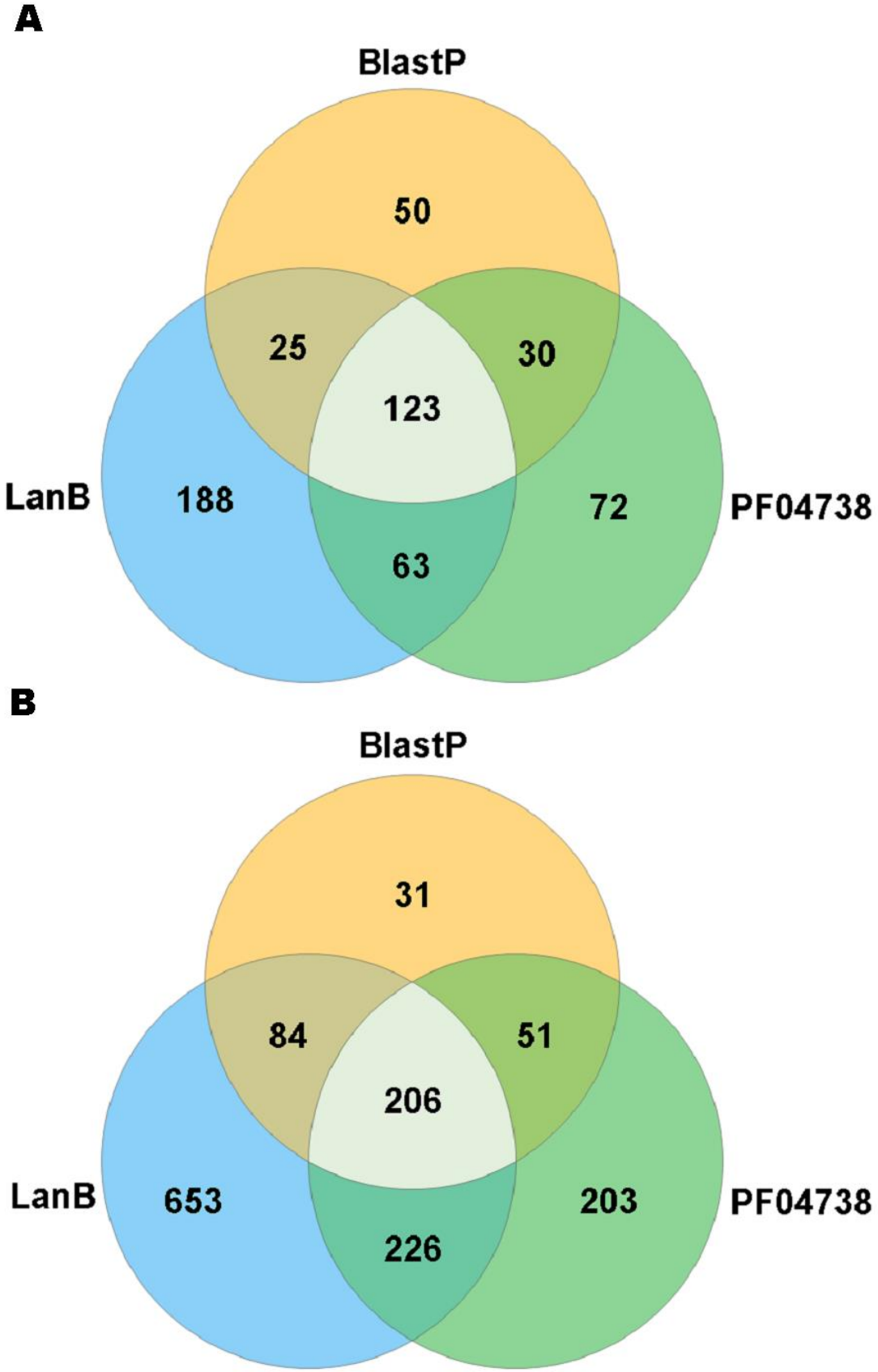


Figure 4. Venn diagram illustrating the numbers of lanthionine dehydratase proteins reported in stool (A) and oral (B) metagenomic data by single and multiple methods.

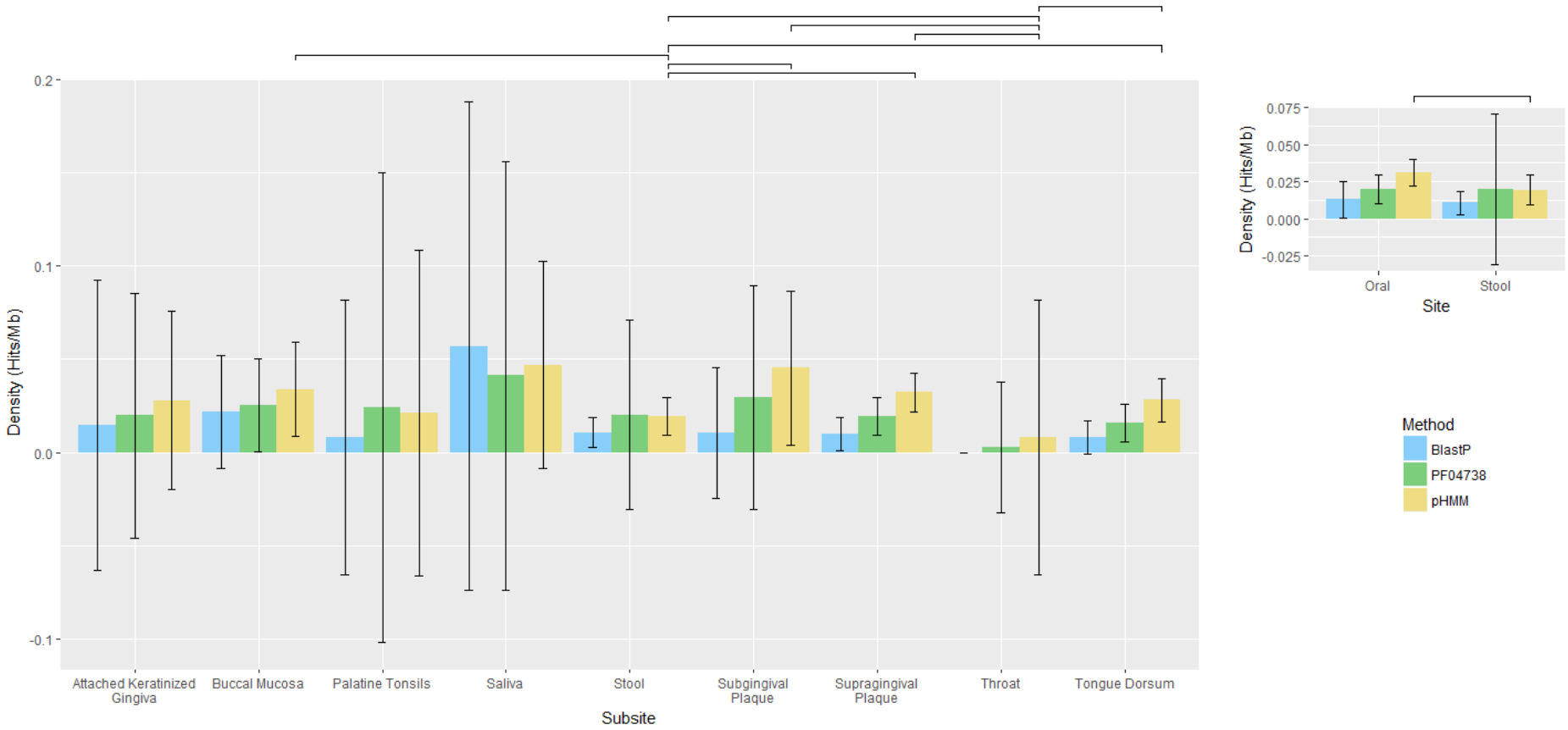


Figure 5. Comparison of lanthionine dehydratase density by body site reported by all three methods. Insert shows overall comparison between stool and oral environments. Significant differences in LanB density between subsites, as identified by our model, are highlighted.

	pHMM Hits	PF04738 Hits	BlastP Hits	pHMM Hit/Mb	PF04738 Hit/Mb	BlastP Hit/Mb
Attached Keratinized Gingiva	7	4	2	0.0244	0.0139	0.0070
Buccal Mucosa	96	74	64	0.0326	0.0251	0.0217
Palatine Tonsils	8	6	2	0.0204	0.0153	0.0051
Saliva	6	5	7	0.0472	0.0393	0.0550
Subgingival Plaque	33	22	9	0.0438	0.0292	0.0120
Supragingival Plaque	528	306	148	0.0309	0.0179	0.0087
Throat	5	2	0	0.0108	0.0043	0.0000
Tongue Dorsum	486	267	142	0.0276	0.0151	0.0081
Stool	399	288	231	0.0189	0.0136	0.0109

Table 3. Comparison of lanthionine dehydratase proteins identified by each method and their density based on metagenome size.

(Wilcoxon rank sum test, $p = 0.0287$), Subgingival Plaque (Wilcoxon rank sum test, $p = 0.009$), Supragingival Plaque (Wilcoxon rank sum test, $p = 0.0016$), and Tongue Dorsum (Wilcoxon rank sum test, $p = 0.0031$) subsites.

Manual Examination of Selected Gene Neighbourhoods

Sixty hits, listed in Table 4, were randomly selected from those identified by the new profile HMM, 45% (27/60) of which were identified by at least one of the other two methods, and manually examined to determine if a bacteriocin gene cluster could be identified. 42% (25/60) of these were not further analysed because the often relatively short regions assembled from the shotgun data prevented the identification of a full lantibiotic gene cluster. However, of the 35 remaining clusters, 28 (80%) appeared to encode multiple genes involved in the biosynthesis of bacteriocins and thiopeptides. These genes encode proteins involved in posttranslational modification, bacteriocin transport, leader cleavage and regulation (Figure 6).

81 hits identified by BlastP were missed by both profile HMM approaches. 50 of these originated in the stool metagenomes and were selected for manual annotation to determine if an overall structure or similarity could be observed. 29 of these 50 were part of clusters whose components showed relatively low sequence identity (39-50%) with proteins responsible for the biosynthesis of thiopeptides and lantibiotics, including a putative lanthionine dehydratase, a radical SAM/SPASM domain-containing protein, a thiopeptide-type bacteriocin biosynthesis domain-containing protein, an S41 family peptidase, and a protein of unknown function (DUF4932)

Body Site	Sample	Contig	Also Identified by Pfam	Also Identified by BlastP
Oral	SRS011098	contig-100_15454.15454	No	No
Oral	SRS013533	contig-100_13993.137974	No	No
Oral	SRS013949	contig-100_16.127362	No	No
Oral	SRS014476	contig-100_19698.19698	No	No
Oral	SRS014476	contig-100_33982.33982	No	No
Oral	SRS014578	contig-100_25195.25195	No	No
Oral	SRS014684	15756	No	No
Oral	SRS014691	contig-100_6081.124540	No	No
Oral	SRS015040	629	Yes	Yes
Oral	SRS015060	9087	No	No
Oral	SRS015158	contig-100_15337.100130	No	No
Oral	SRS015215	24967	No	Yes
Oral	SRS015278	contig-100_10111.10111	No	No
Oral	SRS015470	2052	No	Yes
Oral	SRS016002	contig-100_476.158552	Yes	No
Oral	SRS017080	contig-100_5941.5941	No	No
Oral	SRS017439	contig-100_1672.143182.143182	Yes	Yes
Oral	SRS018591	contig-100_9765.9765	No	No
Oral	SRS019029	23866	No	No
Oral	SRS019219	C2945900	No	No
Oral	SRS019974	contig-100_20140.20140.20140	Yes	No
Oral	SRS021960	contig-100_10619.10619	No	No
Oral	SRS023358	contig-100_2141.2141	No	No
Oral	SRS023938	5276	No	No
Oral	SRS024081	SRS024081_LANL_scaffold_14492	Yes	Yes
Oral	SRS042131	3502	Yes	No
Oral	SRS045127	11735	No	No
Oral	SRS056622	17271	Yes	No
Oral	SRS058808	33685	No	No
Oral	SRS063999	contig-100_1693.183379	No	No

Body Site	Sample	Contig		Also Identified by Pfam	Also Identified by BlastP
Stool	SRS011271	contig-100_502.150577		Yes	Yes
Stool	SRS013476		35183	No	Yes
Stool	SRS014287		9924	Yes	Yes
Stool	SRS015133	C3441160		No	No
Stool	SRS015217	11881		Yes	Yes
Stool	SRS015578	contig-100_12750.137163		No	No
Stool	SRS015782	C3311357		No	No
Stool	SRS016954		19735	No	Yes
Stool	SRS017521	contig-100_1016.215526		Yes	Yes
Stool	SRS018656	contig-100_2624.65167		No	Yes
Stool	SRS018817		12490	Yes	No
Stool	SRS019267	contig-100_30866.30866		No	No
Stool	SRS019397		467	No	No
Stool	SRS019910	contig-100_13822.36985		No	Yes
Stool	SRS020233		53970	Yes	Yes
Stool	SRS022071		5852	No	Yes
Stool	SRS024625	contig-100_26018.53974		Yes	No
Stool	SRS043411		11272	No	No
Stool	SRS047014	contig-100_7652.187852		No	No
Stool	SRS048164		2584	Yes	Yes
Stool	SRS048870	contig-100_3433.99515		Yes	Yes
Stool	SRS049712	contig-100_19642.88407		No	No
Stool	SRS049995		22549	Yes	Yes
Stool	SRS050925	contig-100_30127.70203		No	No
Stool	SRS052697	contig-100_723.255260		Yes	Yes
Stool	SRS054590	contig-100_344.344		No	No
Stool	SRS062427		8699	Yes	Yes
Stool	SRS065504		22874	No	No
Stool	SRS065504	contig-100_509.193795		No	No

Table 4. Sample and contig identifiers of all randomly selected hits that underwent manual annotation. The table also states whether the hit was identified by Pfam and BlastP approaches.

a. Stool

Identifier	Gene cluster	Predicted annotation
SRS047014 contig-100_7652.187852	1 2 3 4	1. Accessory gene regulator B – membrane protein 2. Gramicidin dehydratase 3. 4'-phosphopantetheinyl transferase family protein 4. AMP-dependent synthetase
SRS019397 467	1 2 3 4 5 6	1. Dicarboxylate symporter 2. Serine acetyltransferase 3. Gramicidin dehydrogenase 4. Phosphopantetheinyl 4. transferase 5. AMP-dependent synthetase 6. Non- ribosomal peptide synthase
SRS015578 contig- 100_12750.137163	1	1. Putative thioesterase
SRS011271 4525	1	1. Putative thioesterase
SRS054590 contig-100_344.344	1 2 3 4	1. Serine acetyltransferase 2. 4'-phosphopantetheinyl transferase 3. family protein 3. Gramicidin dehydrogenase 4. SagB-type dehydrogenase
SRS065504 22874	1 2 3 4 5	1. Peptidase 2. ABC transport-related protein 3. ABC transport-related protein 4. Glycosyl transferase 5. Bacillorin synthetase B
SRS018656 contig-100_2624.65167	1 2	1. Lanthionine dehydratase domain protein 2. Lanthionine synthetase C-like protein
SRS062427 8699	1 2 3 4	1. Transposase 2. Putative LanC protein 3. Transposase 4. Putative LanB protein
SRS14287 9924	1 2 3 4 5 6	1. Putative Lan C 2. Transposase 3. Putative LanB 4. ABC transporter 5. ABI protein 6. LanC-like protein
SRS020233 53970	1 2 3 4	1. Lantibiotic dehydratase 2. Peptidase S41 3. Lantibiotic ABC transporter protein 4. Lanthionine synthetase
SRS022071 5852	1 2	1. Putative LanB 2. Putative LanC
SRS016954 19735	1	1. Subtilin biosynthesis protein SpaB
SRS043411 11272	1 2 3 4	1. Putative polyketide synthase 2. Radical SAM protein 3. Radical SAM protein 4. Long chain fatty acid CoAligase
SRS052697 contig-100_723.255260	1 2 3	1. Lantibiotic dehydratase 2. Lantibiotic dehydratase 3. LanC-like protein
SRS048164 2584	1	1. Lantibiotic dehydratase
SRS049995 22549	1 2 3 4 5	1. Transposase 2. Lanthionine synthetase 3. Lanthionine dehydratase 4. Subtilin biosynthesis proteins SpaB 5. Putative alcohol dehydratase
SRS018817 12490	1 2	1. Putative lantibiotic dehydratase 2. Subtilin biosynthesis protein
SRS013476 35183	1 2 3 4	1. Peptidase 2. Lanthionine synthetase C-like protein 3. Lantibiotic dehydratase 4. Thiopeptide-type bacteriocin synthesis domain protein
SRS065504 contig-100_509.193795	1 2 3	1. Lantibiotic biosynthesis protein 2. SagB-type dehydrogenase 3. YcaO-type kinase domain protein

Figure 6a. Graphical representation of stool hits randomly selected for manual examination. The full contig was analysed in each instance but only the area immediately surrounding the predicted LanB protein is illustrated.

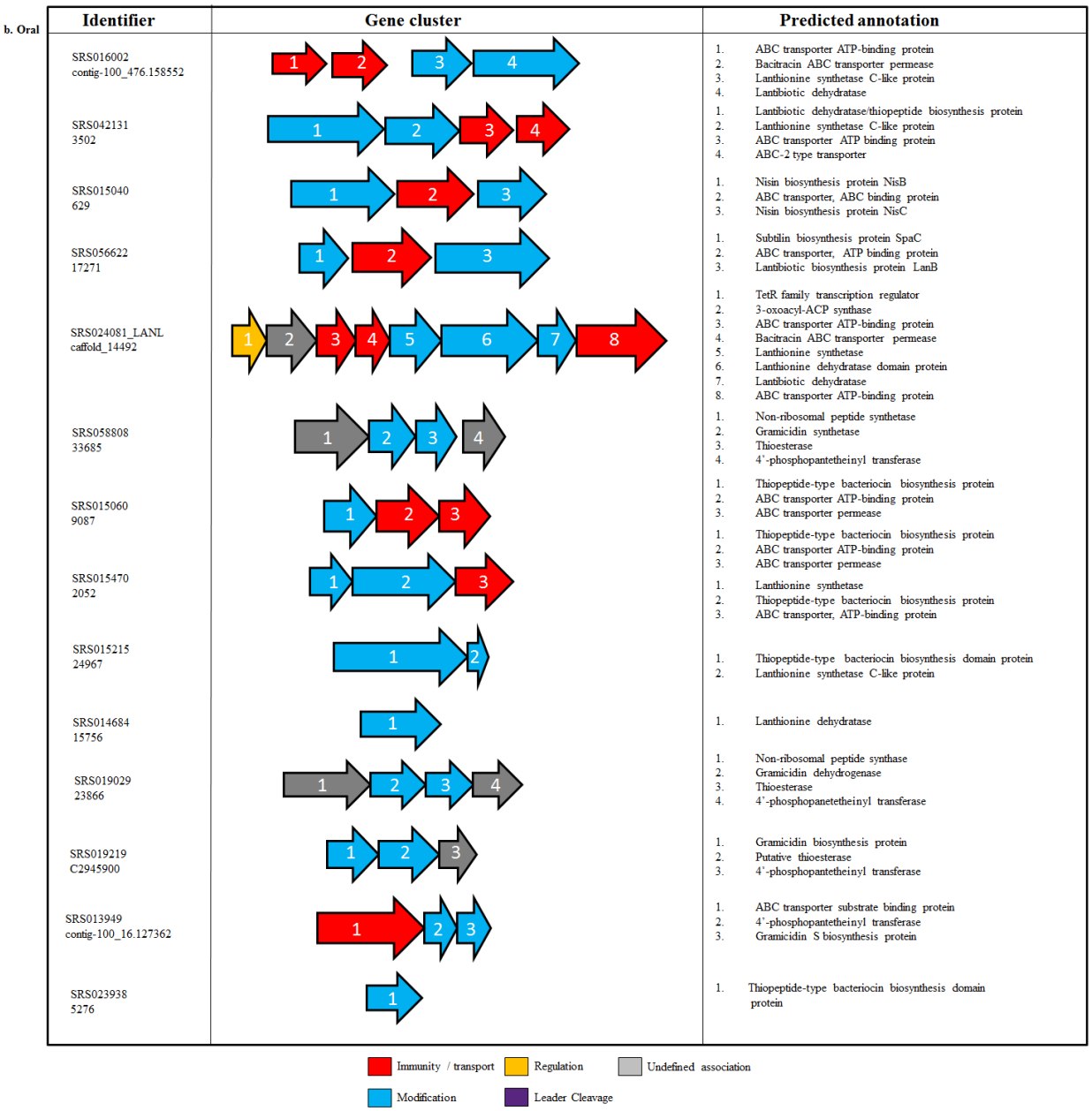


Figure 6b. Graphical representation of oral hits randomly selected for manual examination. The full contig was analysed in each instance but only the area immediately surrounding the predicted LanB protein is illustrated.

predicted to be a putative metalloprotease. All 50 manually annotated gene clusters are available in GENBANK format (https://doi.org/10.6084/m9.figshare.4797790.v1) and an example of this cluster architecture is summarised in Table 5.

Genomic Screen

The draft genomes of the gastrointestinal tract subset of the HMRGD were also used as a database and searched using the new profile HMM. This resulted in the identification of seven hits, including two strains of *Coprobacillus*, a potentially probiotic genus (Stein et al. 2013; Yan et al. 2012) (Table 6). From these seven genomes, only three lantibiotic gene clusters were identified by BAGEL3, these are illustrated in Figure 7. Although this low frequency of lanthionine dehydratase proteins in the genomic dataset (0.006 hits/Mb) contrasts with the findings of the metagenome screen reported above, it is in agreement with previous reports of relatively low subclass I lantibiotic density within the human microbiota (Walsh et al. 2015; Zheng et al. 2014). A possible explanation for this significantly lower gene density (Welch's two sample t-test, p = 1.232e-10) is that the subclass I lantibiotic clusters identified in the metagenomic data by the new profile HMM are present in the genomes of rarer members of the gut microbiota, which are not represented in the HMP reference genome database.

Taxonomic Classification of LanB-encoding contigs

The MEM run mode of Kaiju works by searching for exact matches of given length between the query and database sequences, in the case of multiple hits of the same length in different taxa, a lowest common ancestor is inferred. Kaiju classified 378 of 399 LanB-encoding contigs. Of these, 232 were classified to the species level - however 68 were removed as their exact species was ambiguous. Of the remaining 164 classified contigs, 66 (40.2%) were represented at the species-level in the previously screened HMRGD database. The most abundant genus was *Alistipes*, accounting for 14.03% of LanB-encoding contigs identified by our model, followed by *Blautia* (7.77%), *Clostridium* (4.51%), and *Bacteroides* (3.76%). Genus-level classification is summarised in Table 7.

Annotation	Coverage (%)	E-value	Identitiy (%)
lantibiotic dehydratase	98	4.00E-177	39
thiopeptide-type bacteriocin biosynthesis domain-containing protein	93	6.00E-62	40
lanthionine synthetase C-like protein	97	2.00E-75	39
peptidase S41	97	5.00E-126	39
ABC Transporter (COG2274)	99	0.00E+00	98
hypothetical protein	77	6.00E-23	53
DUF4932 domain-containing protein	95	2.00E-145	46
radical SAM/SPASM domain-containing protein	96	6.00E-177	55
HlyD family secretion protein	98	0.00E+00	65

Table 5. Manual annotation of the putative BlastP-identified biosynthetic gene cluster on scaffold 39304 from stool metagenome SRS014923.

Accession	Strain	E Value
JH414709	*Bacillus* sp. 7_6_55CFAA_CT2	9.0E-16
GL636578	*Coprobacillus* sp. 29_1	3.7E-67
AKCB01000002	*Coprobacillus* sp. D6	4.5E-68
JH126516	*Dorea formicigenerans* 4_6_53AFAA	2.3E-81
ACEP01000029	*Eubacterium hallii* DSM3353	9.4E-27
KI391961	*Fusobacterium nucleatum* subsp. *animalis* 3_1_33	2.2E-09
GG657999	*Fusobacterium* sp. 4_1_13	7.1E-09

Table 6. Detailed information of all lanthionine dehydratase proteins identified in the gastrointestinal tract subset of the Human Microbiome Project's reference genome database using our profile HMM.

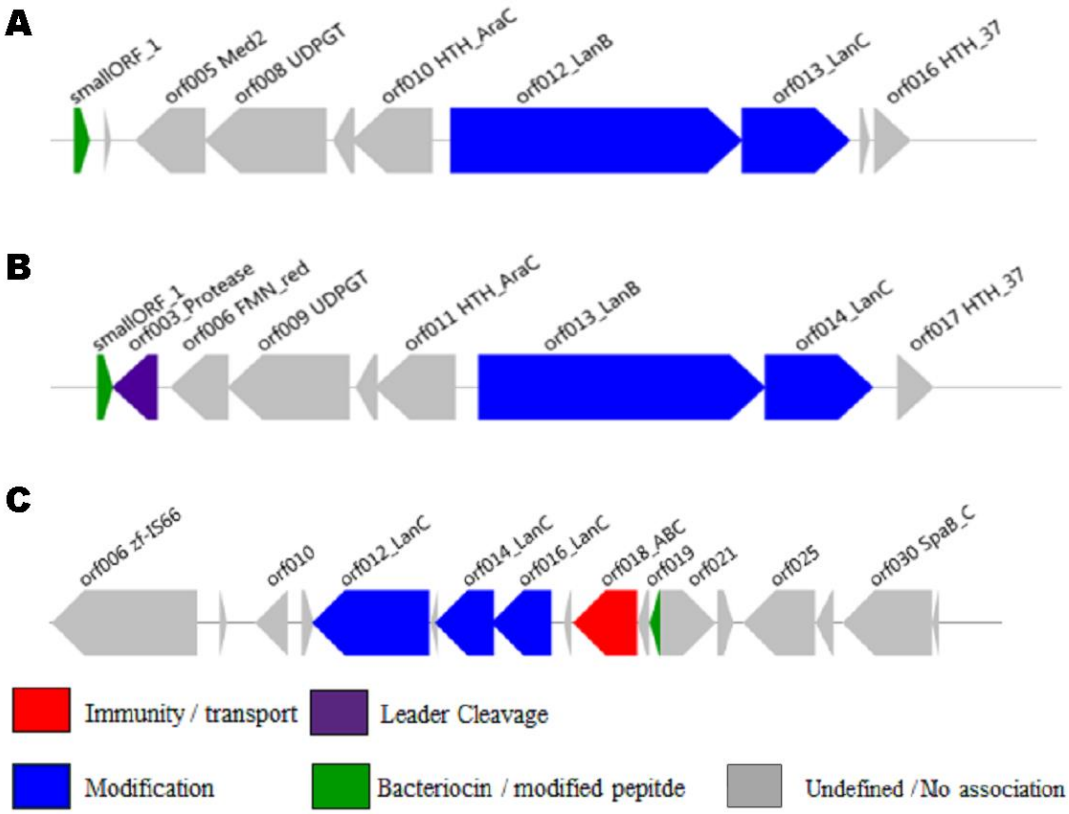


Figure 7. BAGEL3 output of three putative bacteriocin gene clusters identified from the gastrointestinal tract subset of the Human Microbiome Project's reference genome database by our new profile HMM.

A – *Coprobacillus* sp. D6

B – *Coprobacillus* sp. 29_1

C – *Dorea formicigenerans* 4_6_53AFAA

Each predicted open reading frame is colour-coded based on the role it plays in lantibiotic biosynthesis.

%	Reads	Genus
14.03509	56	*Alistipes*
7.769423	31	*Blautia*
4.511278	18	*Clostridium*
3.759398	15	*Bacteroides*
2.506266	10	*Ruminococcus*
2.506266	10	*Parabacteroides*
2.255639	9	*Coprobacillus*
1.503759	6	*Bacillus*
1.503759	6	*Eubacterium*
1.503759	6	*Odoribacter*
1.253133	5	*Chitinophaga*
1.002506	4	*Streptococcus*
1.002506	4	*Janibacter*
1.002506	4	*Faecalibacterium*
0.75188	3	*Prevotella*
0.75188	3	*Paenibacillus*
0.75188	3	*Chryseobacterium*
0.75188	3	*Tenacibaculum*
0.75188	3	*Leisingera*
0.75188	3	*Pseudoflavonifractor*
0.75188	3	*Erysipelatoclostridium*
0.75188	3	*Lachnoclostridium*
0.75188	3	*Tyzzerella*
0.501253	2	*Dorea*
0.501253	2	*Tannerella*
0.501253	2	*Subdoligranulum*
0.501253	2	*Lysinibacillus*
0.501253	2	*Eisenbergiella*
0.250627	1	*Nocardia*
0.250627	1	*Saccharopolyspora*
0.250627	1	*Streptomyces*
0.250627	1	*Coprococcus*
0.250627	1	*Eggerthella*
0.250627	1	*Enterovibrio*
0.250627	1	*Candidatus Accumulibacter*
0.250627	1	*Oscillibacter*
0.250627	1	*Celeribacter*
0.250627	1	*Flavonifractor*
0.250627	1	*Allokutzneria*
36.59148	146	Classified above rank Genus
5.263158	21	Unclassified

Table 7. Genus-level classification of putative LanB-encoding scaffolds identified by our model.

Discussion

Bacteriocin production enhances the competitiveness of bacteria living in complex communities and has the potential to be harnessed for the benefit of human health. The goal of this study was to develop a profile HMM and to assess its ability, in comparison with several other approaches, to detect putative subclass I lantibiotic gene clusters in human metagenomic datasets. Through this process it was also possible to evaluate the potential frequency and distribution of these bacteriocin gene clusters in the human microbiota.

To validate the model, nine positive controls and five negative controls were selected to evaluate its sensitivity and specificity. These controls were selected based on reported bacteriocin production; all positive controls were known producers of subclass I lantibiotics while the negative controls produced either different subclasses of lantibiotics or none at all. Following validation, genomic and metagenomic data corresponding to two niches within the human microbiome were chosen as the focus of this study. The first of these niches was human stool and was selected as the corresponding samples were most likely to yield bacteriocin producers with the potential to modulate undesirable microbiota profiles associated with obesity, colorectal cancer, type 2 diabetes or inflammatory bowel diseases due to their ability to survive and colonise this environment. Secondly, human oral communities were examined as a previous study showed that they contained, by far, the greatest proportion of bacteriocin structural genes across a number of human metagenome samples (Zheng et al. 2014). Zheng *et al.* reported that 80% of class I bacteriocins (lantibiotics) and 89% of all bacteriocins identified using their method originated in the oral metagenomes, while the stool metagenomes contained just 15% and 7%, respectively. The same study reported that 88% of samples from the oral cavity and

73% of samples from the gut contained at least one bacteriocin (regardless of class), while the new profile HMM reported these statistics as 81% and 83%, respectively for subclass I lantibiotics alone. The *in silico* screen carried out with the profile HMM is consistent with the observation by Zheng *et al.* (Zheng et al. 2014) by yielding a higher number and density of hits from the oral, compared to the stool, metagenomic data. Furthermore, the large variation in density of hits between sites within the oral environment suggests that lantibiotic production confers a greater advantage in buccal musoca, subgingival plaque, supragingival plaque, and tongue dorsum communities compared to communities from the throat. This may be due to the direct benefits of antimicrobial activity but could also involve the intra- and interspecies signalling roles attributed to lantibiotic peptides (Upton et al. 2001), particularly in the intensely competitive microbial biofilm environment of dental plaque.

One of the most interesting observations from the study was the large variation in the numbers of *lanB* genes reported by the three different approaches. The BlastP approach identified, by far, the lowest number of significant hits overall and the lowest in every body site examined, except for the saliva microbiome. Our model identified more than double the number of hits provided by the BlastP-based approach, in line with the aforementioned knowledge that profile HMMs can detect as much as three times as many distant homologs than pairwise methods (Park et al. 1998). Our model also identified a greater number of LanB proteins than the Pfam PF04738 model when used to search the same data using the same parameters. While the PF04738 model relates to the C-terminus of the lanthionine dehydratase protein, responsible for the glutamate elimination step of lantibiotic modification (Ortega et al. 2015), the newly developed profile HMM takes the full length of the LanB protein into consideration, thereby providing greater predictive power. The BlastP approach used here utilizes a

single driver sequence while both profile HMMs are representative of a much larger number of sequences – 835 and 573 protein sequences were used to build our model and the Pfam model respectively - perhaps explaining the differing number of hits reported by each approach. It is possible that an altered BlastP approach, pooling the unique hits reported by a standard BlastP search using each of the 835 driver sequences used to build our profile HMM, would result in a comparable number of proteins identified.

Our model, in addition to identifying more potential LanB proteins, was the only method evaluated that did not identify any false positive or false negative in the control dataset. As stated above, profile HMMs are already known to be particularly sensitive, the validation step, however, also suggests that they are more specific than the other methods evaluated as they were the only approach which did not return any false positives. When selecting the controls used to examine the performance of the different approaches, greater consideration was given to the quality of these controls than their quantity. Only controls with experimentally characterised lantibiotic production were included in the validation dataset. This relatively small control group means that, although the results of the validation step may explain the contrasting numbers of LanB proteins reported by our model and the PF04738 model, it cannot be said for certain that our model performed better.

Zheng *et al*., using the same metagenomic data that was the focus of this study, identified 17 potential subclass I lantibiotics from stool samples and 76 from oral samples, a much lower frequency of detection than in this study, probably due to the different methodologies used. That study focused on searching for proteins similar to those in BAGEL3's manually curated database, an approach which likely lost sensitivity because bacteriocin precursor peptides can differ considerably at primary

sequence level. Furthermore, the screen employed a BLAST-based approach which, as demonstrated here, exhibited the lowest number of significant hits reported.

To investigate the areas surrounding the LanB-encoding genes identified by our model we randomly selected thirty positive hits from the oral and stool metagenome screens for manual examination. This approach revealed that several of the hits were on scaffolds that were either too small to contain a full gene or did not contain the gene's start codon. This was most likely as a consequence of the fragmented nature of the metagenomic data, as opposed the identification of true false positives by the model and would probably occur regardless of the method employed. 42% (25/60) of hits selected for manual examination were discarded based on these criteria. It also revealed that a considerable number of hits exhibited low (~30%) similarity to putative thioesterases in the nr protein sequence database, highlighting that lanthionine dehydratases are relatively-closely related to proteins involved in the posttranslational modification of thiopeptides, most likely those responsible for dehydration of serine and threonine residues (Garg et al. 2013). The similarity between these dehydratase proteins suggests a possible common ancestor protein (Kelly et al. 2009). Another possible explanation relates to the fact that all of the proteins annotated as thiopeptide modification proteins are putative annotations and none, to our knowledge, have been confirmed as such *in vitro*. It is possible, therefore, that these may simply be lanthionine dehydratases which have been incorrectly annotated due to automatic software and incomplete/under-curated databases. The majority of clusters identified contained genes encoding both LanB and LanC modification proteins, while many also contained a leader cleavage and activation peptidase and/or ABC transporter proteins for export of the mature peptide, suggesting that these have the potential to encode a functional lantibiotic.

To evaluate the model's performance in a genomic context we applied it to the gastrointestinal tract subset of the HMP's reference genome database and compared the results to our previously published study which used the online bacteriocin genome mining tool BAGEL3 to screen this same database (Walsh et al. 2015). The results of the two screens were startlingly different and served to highlight the variation in results that can arise from applying different methods to the same data. Interestingly, the gastrointestinal tract reference genomes encoded a significantly lower frequency of LanB hits than the stool metagenomic samples. Taxonomic classification of the 399 LanB-encoding contigs identified by our new model from the stool metagenomes revealed that only 40.2% of these potential lantibiotic producing strains were represented in the reference genome database, suggesting that the majority of these lantibiotics were encoded by rarer members of the gut microbiota or those that have not previously been identified as important. Taxonomic classification of these LanB-encoding contigs also served to highlight patterns in the results of the three approaches used (Figure 8), for example our model identified *Allokutzneria*, *Coprococcus*, *Enterovibrio*, *Paenibacillus*, and *Tenicibaculum*-encoded LanB proteins that were completely missed by the Pfram and BlastP approaches.

Figure 8. Proportion of hits identified in metagenomic stool samples by our model that were also identified by other methods. Illustrates that the proportion also identified by A) Pfam and B) BlastP approaches varies by producing genus. The black line shows the overall proportion of hits identified by each method.

Conclusions

Across the oral and stool communities examined, this study identified 2007 unique putative subclass I lantibiotic biosynthetic gene clusters by three different methods, further emphasising the tremendous potential that the human microbiota has as a source of therapeutic compounds. As this study was performed entirely *in* silico, the next challenge lies in experimentally identifying and characterising these putative bacteriocins to identify those with the ability to desirably modulate the microbiota for the treatment of disease.

References

ALTSCHUL SF, GISH W, MILLER W, MYERS EW, AND LIPMAN DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-410. 10.1016/s0022-2836(05)80360-2

ARNISON PG, BIBB MJ, BIERBAUM G, BOWERS AA, BUGNI TS, BULAJ G, CAMARERO JA, CAMPOPIANO DJ, CHALLIS GL, CLARDY J, COTTER PD, CRAIK DJ, DAWSON M, DITTMANN E, DONADIO S, DORRESTEIN PC, ENTIAN K-D, FISCHBACH MA, GARAVELLI JS, GORANSSON U, GRUBER CW, HAFT DH, HEMSCHEIDT TK, HERTWECK C, HILL C, HORSWILL AR, JASPARS M, KELLY WL, KLINMAN JP, KUIPERS OP, LINK AJ, LIU W, MARAHIEL MA, MITCHELL DA, MOLL GN, MOORE BS, MULLER R, NAIR SK, NES IF, NORRIS GE, OLIVERA BM, ONAKA H, PATCHETT ML, PIEL J, REANEY MJT, REBUFFAT S, ROSS RP, SAHL H-G, SCHMIDT EW, SELSTED ME, SEVERINOV K, SHEN B, SIVONEN K, SMITH L, STEIN T, SUSSMUTH RD, TAGG JR, TANG G-L, TRUMAN AW, VEDERAS JC, WALSH CT, WALTON JD, WENZEL SC, WILLEY JM, AND VAN DER DONK WA. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports* 30:108-160. 10.1039/C2NP20085F

BÄCKHED F, LEY RE, SONNENBURG JL, PETERSON DA, AND GORDON JI. 2005. Host-Bacterial Mutualism in the Human Intestine. *Science* 307:1915-1920. 10.1126/science.1104816

BELSTRØM D, HOLMSTRUP P, BARDOW A, KOKARAS A, FIEHN N-E, AND PASTER BJ. 2016. Temporal Stability of the Salivary Microbiota in Oral Health. *PLoS ONE* 11:e0147472. 10.1371/journal.pone.0147472

CARVER T, HARRIS SR, BERRIMAN M, PARKHILL J, AND MCQUILLAN JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464-469. 10.1093/bioinformatics/btr703

CLEMENTE JOSE C, URSELL LUKE K, PARFREY LAURA W, AND KNIGHT R. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* 148:1258-1270. http://dx.doi.org/10.1016/j.cell.2012.01.035

CORR SC, LI Y, RIEDEL CU, O'TOOLE PW, HILL C, AND GAHAN CGM. 2007. Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118. *Proceedings of the National Academy of Sciences* 104:7617-7621. 10.1073/pnas.0700440104

DELCHER AL, HARMON D, KASIF S, WHITE O, AND SALZBERG SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636-4641.

DONIA MOHAMED S, CIMERMANCIC P, SCHULZE CHRISTOPHER J, WIELAND BROWN LAURA C, MARTIN J, MITREVA M, CLARDY J, LININGTON ROGER G, AND FISCHBACH MICHAEL A. 2014. A Systematic Analysis of Biosynthetic Gene Clusters in the Human Microbiome Reveals a Common Family of Antibiotics. *Cell* 158:1402-1414. 10.1016/j.cell.2014.08.032

DUSKO EHRLICH S. 2010. [Metagenomics of the intestinal microbiota: potential applications]. *Gastroenterol Clin Biol* 34 Suppl 1:S23-28. 10.1016/s0399-8320(10)70017-8

EDGAR RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797. 10.1093/nar/gkh340

EREJUWA OO, SULAIMAN SA, AND AB WAHAB MS. 2014. Modulation of gut microbiota in the management of metabolic disorders: the prospects and challenges. *Int J Mol Sci* 15:4158-4188. 10.3390/ijms15034158

FAITH JJ, GURUGE JL, CHARBONNEAU M, SUBRAMANIAN S, SEEDORF H, GOODMAN AL, CLEMENTE JC, KNIGHT R, HEATH AC, LEIBEL RL, ROSENBAUM M, AND GORDON JI. 2013. The Long-Term Stability of the Human Gut Microbiota. *Science* 341. 10.1126/science.1237439

FLINT HJ, SCOTT KP, LOUIS P, AND DUNCAN SH. 2012. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 9:577-589.

GALES M, AND YOUNG S. 2007. The application of hidden Markov models in speech recognition. *Found Trends Signal Process* 1:195-304. 10.1561/2000000004

GARG N, SALAZAR-OCAMPO LMA, AND VAN DER DONK WA. 2013. In vitro activity of the nisin dehydratase NisB. *Proceedings of the National Academy of Sciences of the United States of America* 110:7258-7263. 10.1073/pnas.1222488110

GOTO Y, LI B, CLAESEN J, SHI Y, BIBB MJ, AND VAN DER DONK WA. 2010. Discovery of Unique Lanthionine Synthetases Reveals New Mechanistic and Evolutionary Insights. *PLoS Biol* 8:e1000339. 10.1371/journal.pbio.1000339

JEFFERY IB, LYNCH DB, AND O'TOOLE PW. 2016. Composition and temporal stability of the gut microbiota in older persons. *ISME J* 10:170-182. 10.1038/ismej.2015.88

KELLY WL, PAN L, AND LI C. 2009. Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *J Am Chem Soc* 131:4327-4334. 10.1021/ja807890a

MADERA M, AND GOUGH J. 2002. A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Res* 30:4321-4328.

MENZEL P, NG KL, AND KROGH A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7:11257. 10.1038/ncomms11257

http://www.nature.com/articles/ncomms11257#supplementary-information

MORONI O, KHEADR E, BOUTIN Y, LACROIX C, AND FLISS I. 2006. Inactivation of Adhesion and Invasion of Food-Borne Listeria monocytogenes by Bacteriocin-Producing Bifidobacterium Strains of Human Origin. *Applied and Environmental Microbiology* 72:6894-6901. 10.1128/aem.00928-06

MORTON JT, FREED SD, LEE SW, AND FRIEDBERG I. 2015. A large scale prediction of bacteriocin gene blocks suggests a wide functional spectrum for bacteriocins. *BMC Bioinformatics* 16:1-9. 10.1186/s12859-015-0792-9

ORTEGA MA, HAO Y, ZHANG Q, WALKER MC, VAN DER DONK WA, AND NAIR SK. 2015. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* 517:509-512. 10.1038/nature13888

PARK J, KARPLUS K, BARRETT C, HUGHEY R, HAUSSLER D, HUBBARD T, AND CHOTHIA C. 1998. Sequence comparisons using multiple sequences

detect three times as many remote homologues as pairwise methods. *J Mol Biol* 284:1201-1210. 10.1006/jmbi.1998.2221

PEREZ RH, ZENDO T, AND SONOMOTO K. 2014. Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications. *Microbial Cell Factories* 13:S3-S3. 10.1186/1475-2859-13-S1-S3

PUNTA M, COGGILL PC, EBERHARDT RY, MISTRY J, TATE J, BOURSNELL C, PANG N, FORSLUND K, CERIC G, CLEMENTS J, HEGER A, HOLM L, SONNHAMMER ELL, EDDY SR, BATEMAN A, AND FINN RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290-D301. 10.1093/nar/gkr1065

QIN J, LI R, RAES J, ARUMUGAM M, BURGDORF KS, MANICHANH C, NIELSEN T, PONS N, LEVENEZ F, YAMADA T, MENDE DR, LI J, XU J, LI S, LI D, CAO J, WANG B, LIANG H, ZHENG H, XIE Y, TAP J, LEPAGE P, BERTALAN M, BATTO J-M, HANSEN T, LE PASLIER D, LINNEBERG A, NIELSEN HB, PELLETIER E, RENAULT P, SICHERITZ-PONTEN T, TURNER K, ZHU H, YU C, LI S, JIAN M, ZHOU Y, LI Y, ZHANG X, LI S, QIN N, YANG H, WANG J, BRUNAK S, DORE J, GUARNER F, KRISTIANSEN K, PEDERSEN O, PARKHILL J, WEISSENBACH J, BORK P, EHRLICH SD, AND WANG J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59-65. http://www.nature.com/nature/journal/v464/n7285/suppinfo/nature08821_S1.html

R CORE TEAM. 2015. R: A Language and Environment for Statistical Computing.

REA MC, DOBSON A, O'SULLIVAN O, CRISPIE F, FOUHY F, COTTER PD, SHANAHAN F, KIELY B, HILL C, AND ROSS RP. 2011a. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences* 108:4639-4644. 10.1073/pnas.1001224107

REA MC, ROSS RP, COTTER PD, AND HILL C. 2011b. Classification of Bacteriocins from Gram-Positive Bacteria. In: Drider D, and Rebuffat S, eds. *Prokaryotic Antimicrobial Peptides: From Genes to Applications*. New York, NY: Springer New York, 29-53.

SEIFERT M, ABOU-EL-ARDAT K, FRIEDRICH B, KLINK B, AND DEUTSCH A. 2014. Autoregressive Higher-Order Hidden Markov Models: Exploiting Local Chromosomal Dependencies in the Analysis of Tumor Expression Profiles. *PLoS ONE* 9:e100295. 10.1371/journal.pone.0100295

SENDER R, FUCHS S, AND MILO R. 2016. Revised estimates for the number of human and bacteria cells in the body. *bioRxiv*. 10.1101/036103

SHERLOCK C, XIFARA T, TELFER S, AND BEGON M. 2013. A coupled hidden Markov model for disease interactions. *Journal of the Royal Statistical Society Series C, Applied Statistics* 62:609-627. 10.1111/rssc.12015

SKEWES-COX P, SHARPTON TJ, POLLARD KS, AND DERISI JL. 2014. Profile Hidden Markov Models for the Detection of Viruses within Metagenomic Sequence Data. *PLoS ONE* 9:e105067. 10.1371/journal.pone.0105067

STEIN RR, BUCCI V, TOUSSAINT NC, BUFFIE CG, RÄTSCH G, PAMER EG, SANDER C, AND XAVIER JB. 2013. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Comput Biol* 9:e1003388. 10.1371/journal.pcbi.1003388

THE HUMAN MICROBIOME PROJECT CONSORTIUM. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-214. 10.1038/nature11234

UPTON M, TAGG JR, WESCOMBE P, AND JENKINSON HF. 2001. Intra- and Interspecies Signaling between Streptococcus salivarius and Streptococcus pyogenes Mediated by SalA and SalA1 Lantibiotic Peptides. *J Bacteriol* 183:3931-3938. 10.1128/JB.183.13.3931-3938.2001

VAN HEEL AJ, DE JONG A, MONTALBAN-LOPEZ M, KOK J, AND KUIPERS OP. 2013. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* 41:W448-453. 10.1093/nar/gkt391

WALSH CJ, GUINANE CM, HILL C, ROSS RP, O'TOOLE PW, AND COTTER PD. 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. *BMC Microbiol* 15:183. 10.1186/s12866-015-0515-4

WALSH CJ, GUINANE CM, O'TOOLE PW, AND COTTER PD. 2014. Beneficial modulation of the gut microbiota. *FEBS Lett* 588:4120-4130. 10.1016/j.febslet.2014.03.035

WEBER T, BLIN K, DUDDELA S, KRUG D, KIM HU, BRUCCOLERI R, LEE SY, FISCHBACH MA, MÜLLER R, WOHLLEBEN W, BREITLING R, TAKANO E, AND MEDEMA MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237-W243. 10.1093/nar/gkv437

XIE L, AND VAN DER DONK WA. 2004. Post-translational modifications during lantibiotic biosynthesis. *Curr Opin Chem Biol* 8:498-507. 10.1016/j.cbpa.2004.08.005

YAN X, GURTLER JB, FRATAMICO PM, HU J, AND JUNEJA VK. 2012. Phylogenetic identification of bacterial MazF toxin protein motifs among probiotic strains and foodborne pathogens and potential implications of engineered probiotic intervention in food. *Cell & Bioscience* 2:1-13. 10.1186/2045-3701-2-39

YOON B-J. 2009. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics* 10:402-415. 10.2174/138920209789177575

ZHENG J, GÄNZLE MG, LIN XB, RUAN L, AND SUN M. 2014. Diversity and dynamics of bacteriocins from human microbiome. *Environ Microbiol*:n/a-n/a. 10.1111/1462-2920.12662

Chapter 4

The relationship between subclass I lantibiotic gene density and the composition and functional potential of the gut microbiota

Identification of lantibiotic modification proteins and subsequent analysis was performed by the candidate

DNA extraction, sequencing and preliminary analysis was performed by Wiley Barton of Teagasc Moorepark, Cork.

Abstract

The gut microbiota of elite-athletes has been shown to be more compositionally diverse than that of controls. This dissimilarity is even more pronounced at the functional level. Given the considerable differences between the athlete and control microbiota, and the hypothesised influence of bacteriocin production on the composition and function of the gut microbial community, this data was selected to determine the extent to which the distribution of antimicrobial peptides can vary across different individuals. The profile Hidden Markov Model (HMM) developed in Chapter 3 of this thesis was used to search these data for LanB proteins, which are responsible for the posttranslational modifications characteristic of subclass I lantibiotics – a well-studied type of bacteriocin with potent antimicrobial activity. The sensitive nature of this profile HMM means it is ideally suited to searching cohorts, such as those studied here, which are quite different in respects to their microbiota. Ultimately, eighty-one shotgun-sequenced metagenomes, accompanied by collected metadata, were examined and it was revealed that class I lantibiotic gene cluster levels differed in the microbiota of athletes compared to high BMI sedentary controls. Also apparent was a relationship between the presence of these gene clusters and lean mass. Further investigation revealed that diet and overall functional capacity of the microbiota exhibited the strongest relationship with subclass I lantibiotic gene cluster density and some interesting relationships with the gut microbiota were highlighted for future examination, particularly a seemingly counterintuitive correlation with peptidoglycan biosynthesis.

These results also established that the relationship between subclass I lantibiotics and the host is very different between elite athletes and the sedentary population, most likely due to their extreme diet and exercise regimen.

Introduction

Bacteriocins are a class of ribosomally-synthesised antimicrobial peptides produced by many bacteria. Lantibiotics are a group of post-translationally modified bacteriocins that are distinguished by the modifications they contain, and the proteins responsible for these modifications. The profile Hidden Markov Model (HMM) developed in chapter 3 of this thesis identifies lanthionine dehydratase, LanB, proteins that contribute to the production of subclass I lantibiotics (Arnison et al., 2013). In this study we use LanB proteins as an indicator of the number of unique bacteriocin-producing bacteria in each metagenome. This profile HMM is suited to searching metagenomic data as the increased sensitivity of profile HMMs over sequence homology approaches, such as BLAST, allows detection of more distantly related proteins in relatively under-characterised members of the gut microbiota while also overcoming the limitation of tools tailored to genomic data, such as BAGEL3, which struggle with the fragmented nature of metagenomic assemblies as they rely on identifying multiple proteins related to biosynthesis within a set genomic region.

To assess distribution of LanB proteins across different human gut samples we utilised data obtained from elite-athletes (members of the Irish national rugby team) which, through two previous publications (Barton et al., 2017, Clarke et al., 2014), form part of a growing body of evidence suggesting part of the benefit of exercise is affected through modulation of the gut microbiota (Estaki et al., 2016, Bressa et al., 2017). In addition to high levels of exercise, these elite-athletes have previously been shown to have a significantly different diet to the sedentary populations (Clarke et al., 2014) which, as highlighted previously in this thesis, has tremendous potential to influence the gut microbiota. The elite athlete microbiota differed with respect its compositional and functional diversity from that of control cohorts (Barton et al., 2017, Clarke et al., 2014). The shotgun metagenomic data, and accompanying metadata, generated as part of these previous studies (Barton et al., 2017, Clarke et al., 2014) are ideally suited to

examining whether subclass I lantibiotic production differs between the gut microbiota of athletes and non-athletes using the developed profile HMM, and investigating potential relationships between subclass I lantibiotic production, diet, host physiology, and microbiota composition and function.

Methods

Study population

Elite professional male athletes (the Irish national rugby team; n=40) and healthy controls (n=46) matched for age and gender were enrolled in 2011 (Barton et al., 2017, Clarke et al., 2014). Due to the range of physiques within a rugby team (player position dictates need for a variety of physical constitutions, i.e., forward players tend to have larger BMI values than backs, often in the overweight/ obese range) the recruited control cohort included two groups of low (BMI ≤25.2) and high (BMI ≥26.5) BMI, respectively. Approval for this study was granted by the Cork Clinical Research Ethics Committee.

Acquisition of clinical, exercise and dietary data

Self-reported dietary intake information was accommodated by a research nutritionist within the parameters of a food frequency questionnaire in conjunction with a photographic food atlas as per the initial investigation (Clarke et al., 2014). Fasting blood samples were collected and analysed at the Mercy University Hospital clinical laboratories, Cork. Commercial multispot microplates (Meso Scale Diagnostics) were used to measure cytokines.

Preparation of metagenomic libraries

Upon initial collection, stool and urine samples were stored on ice prior to DNA extraction and purification from the fresh stool using the QIAmp DNA Stool Mini Kit (cat. no. 51504 Qiagen, Crawley, West Sussex, UK), after which samples were stored securely at -80°C.

DNA extraction was carried out in accordance with the manufacturer's protocol with the addition of a zirconia bead (11079101z-BSP, 11079110z-BSP, 11079125z-BSP Stratech Scientific) cell disruption bead-beating step (30s X 3).

Metagenomic library preparation was performed with the Illumina Nextera XT DNA Library Preparation Kit (cat# FC-131-1096, Illumina Inc., USA) in accordance with the manufacturer's protocol (15031942, Illumina). Normalisation of samples to the recommended 0.2 ng/µL per individual library was carried out with the ThermoFisher Qubit 2.0 Flurometric Quantitation system (Q32854, ThermoFisher). Tagmentation and amplification carried out with G-STORM GS1 thermal cycler system. Following the combined enzymatic fragmentation and adapter sequence tagging—tagmentation—and the subsequent amplification of the tagmented DNA, libraries were purified with the AMPure magnetic bead system at a ratio of 1:1.8 (DNA:AMPure) (9A63880, Beckman Coulter). Subsequently, libraries were assessed for appropriate fragment size (~500bp) on the Agilent 2100 Bioanalyzer system. With the libraries passing quality and fragment length requirements, the library preparation was continued on through library normalization, which was met with an additional assessment of suitable molar concentrations (~2 nM) with the KAPA Library Quantification Kit (KK4824, Kapabiosystems) run on a Roche LightCycler 480 (Roche Applied Science). Samples were combined into 8 final pools prior to being shipped on dry ice for sequencing.

Statistical and bioinformatics analysis

Delivered raw FASTQ sequence files (EMBL Nucleotide Sequence Database (ENA) (http://www.ebi.ac.uk/ena/data/), accession number PRJEB15388) were quality checked as follows: contaminating sequences of human origin were first removed through the NCBI Best Match Tagger (BMTagger). Poor-quality and duplicate read removal, as well as trimming was implemented using a combination of SAM (sequence alignment map) and Picard tools.

Processing of raw sequence data produced a total of 2,803,449,392 filtered reads with a mean read count of 32,598,248.74 (±10,639,447 SD) per each of the 86 samples. These refined reads were then subjected to functional profiling by the Human Microbiome Project Unified Metabolic Analysis Network (HUMAnN2 V.0.5.0) pipeline. The functional profiling performed by HUMAnN2 composed tabulated files of microbial metabolic pathway abundance and coverage derived from the Metacyc database, and gene family abundance (reported in reads per kilobase (RPK) to normalize for gene length) derived from the ChocoPhlan and UniRef50 databases. Metagenome assembly was performed by the Iterative de Bruijn Graph de novo assembler IDBA.

The HMMER3 tool hmmsearch was used, with default parameters, to search the profile HMM created in chapter 3 of this thesis against the 81 assembled metagenomes. An identified sequence was deemed to be valid if its associated E-value was less than, or equal to, 1×10^{-5}. All statistical analyses were performed in R (v. 3.2.3). Correlations were performed using the 'psych' package (v. 1.7.21.3). Spearman correlations were calculated and adjusted by FDR. Diversity measures and non-metric multidimensional scaling (NMDS) were calculated using the 'vegan' package (v. 2.3-1) and illustrated with ggplot2. NMDS was performed by the 'metaMDS' function in 2 dimensions using Wisconsin Double Standardised, square-root transformed, Bray-Curtis dissimilarities calculated by 'vegdist'. Random Forests was used to predict LanB density based on diet and microbiota data using the default parameters of the 'randomForest' package in R (v. 4.6-12) with "ntree" set to 2000. The Boruta package for R (v. 5.1) was used with default parameters to identify variables in these data with predictive power.

Results

LanB density differs between athletes and high BMI controls

LanB density was calculated as the number of significant (E < 1e-05) hits identified by the profile HMM divided by the size of the assembled metagenome. In theory, each sequenced region is present exactly once in these assembled metagenomes. As such, LanB density, as used here, represents the number of unique subclass I lantibiotic gene clusters per megabase. The number and density of LanB proteins identified in each metagenome are summarized in Table 1.

No difference in LanB density was observed between athletes and non-athletes (Wilcoxon rank sum test, $p = 0.09768$), athletes and low BMI controls (Wilcoxon rank sum test, $p = 0.56$), or low BMI controls and high BMI controls (Wilcoxon rank sum test, $p = 0.07914$). There was, however, a significantly higher LanB density in high BMI controls relative to athletes (Wilcox rank sum test, $p = 0.03178$). The distribution of LanB density is illustrated in Figure 1.

Sample	Group	LanB Number	LanB Density (Hits/Mb)
APC025-001	Athlete	0	0
APC025-002	Athlete	1	0.033913378
APC025-003	Athlete	1	0.007273414
APC025-004	Athlete	0	0
APC025-005	Athlete	2	0.019967118
APC025-006	Athlete	1	0.011080495
APC025-009	Athlete	1	0.005600408
APC025-012	Athlete	4	0.032608909
APC025-013	Athlete	2	0.013897774
APC025-014	Athlete	2	0.016430241
APC025-015	Athlete	0	0
APC025-016	Athlete	0	0
APC025-017	Athlete	0	0
APC025-018	Athlete	4	0.019281996
APC025-019	Athlete	2	0.01001908
APC025-020	Athlete	2	0.0135753
APC025-021	Athlete	2	0.009669501
APC025-022	Athlete	0	0
APC025-023	Athlete	1	0.007373666
APC025-024	Athlete	0	0
APC025-025	Athlete	4	0.026913032
APC025-026	Athlete	0	0
APC025-027	Athlete	0	0
APC025-028	Athlete	0	0
APC025-029	Athlete	1	0.009199093
APC025-030	Athlete	1	0.006742014
APC025-031	Athlete	0	0
APC025-032	Athlete	0	0
APC025-033	Athlete	2	0.015286124
APC025-034	Athlete	4	0.033107776
APC025-035	Athlete	1	0.007273414
APC025-036	Athlete	0	0
APC025-037	Athlete	0	0
APC025-038	Athlete	2	0.019967118
APC025-039	Athlete	1	0.011080495
APC025-042	Athlete	0	0
APC025-102	Low BMI Control	2	0.012051089
APC025-104	Low BMI Control	1	0.006101681
APC025-105	High BMI Control	8	0.049778401
APC025-106	Low BMI Control	0	0
APC025-107	Low BMI Control	1	0.009217572
APC025-108	High BMI Control	1	0.013442144
APC025-109	High BMI Control	4	0.022023952
APC025-110	Low BMI Control	1	0.006145777
APC025-111	Low BMI Control	2	0.00955504

APC025-112	High BMI Control	6	0.049471412
APC025-113	Low BMI Control	2	0.011476498
APC025-114	High BMI Control	3	0.014647225
APC025-117	Low BMI Control	0	0
APC025-118	Low BMI Control	1	0.010244197
APC025-119	Low BMI Control	1	0.007594878
APC025-120	High BMI Control	4	0.021555498
APC025-121	Low BMI Control	3	0.018519849
APC025-122	Low BMI Control	0	0
APC025-123	Low BMI Control	0	0
APC025-124	High BMI Control	0	0
APC025-125	Low BMI Control	4	0.030515023
APC025-127	High BMI Control	2	0.021437921
APC025-128	Low BMI Control	0	0
APC025-129	Low BMI Control	4	0.035066908
APC025-130	Low BMI Control	1	0.007724202
APC025-131	Low BMI Control	2	0.011663179
APC025-132	Low BMI Control	1	0.007655815
APC025-133	Low BMI Control	0	0
APC025-134	Low BMI Control	2	0.018684523
APC025-136	Low BMI Control	3	0.016121732
APC025-137	High BMI Control	0	0
APC025-138	High BMI Control	1	0.009080728
APC025-139	High BMI Control	0	0
APC025-140	High BMI Control	6	0.086164261
APC025-142	High BMI Control	2	0.014719434
APC025-143	High BMI Control	1	0.007886491
APC025-144	High BMI Control	2	0.019296602
APC025-146	High BMI Control	0	0
APC025-147	High BMI Control	3	0.015211851
APC025-148	High BMI Control	3	0.023917107
APC025-149	High BMI Control	2	0.016430241
APC025-150	High BMI Control	0	0
APC025-151	High BMI Control	3	0.020230598
APC025-152	High BMI Control	3	0.020230598
APC025-154	High BMI Control	0	0

Table 1. Number and density of LanB proteins identified in each metagenome.

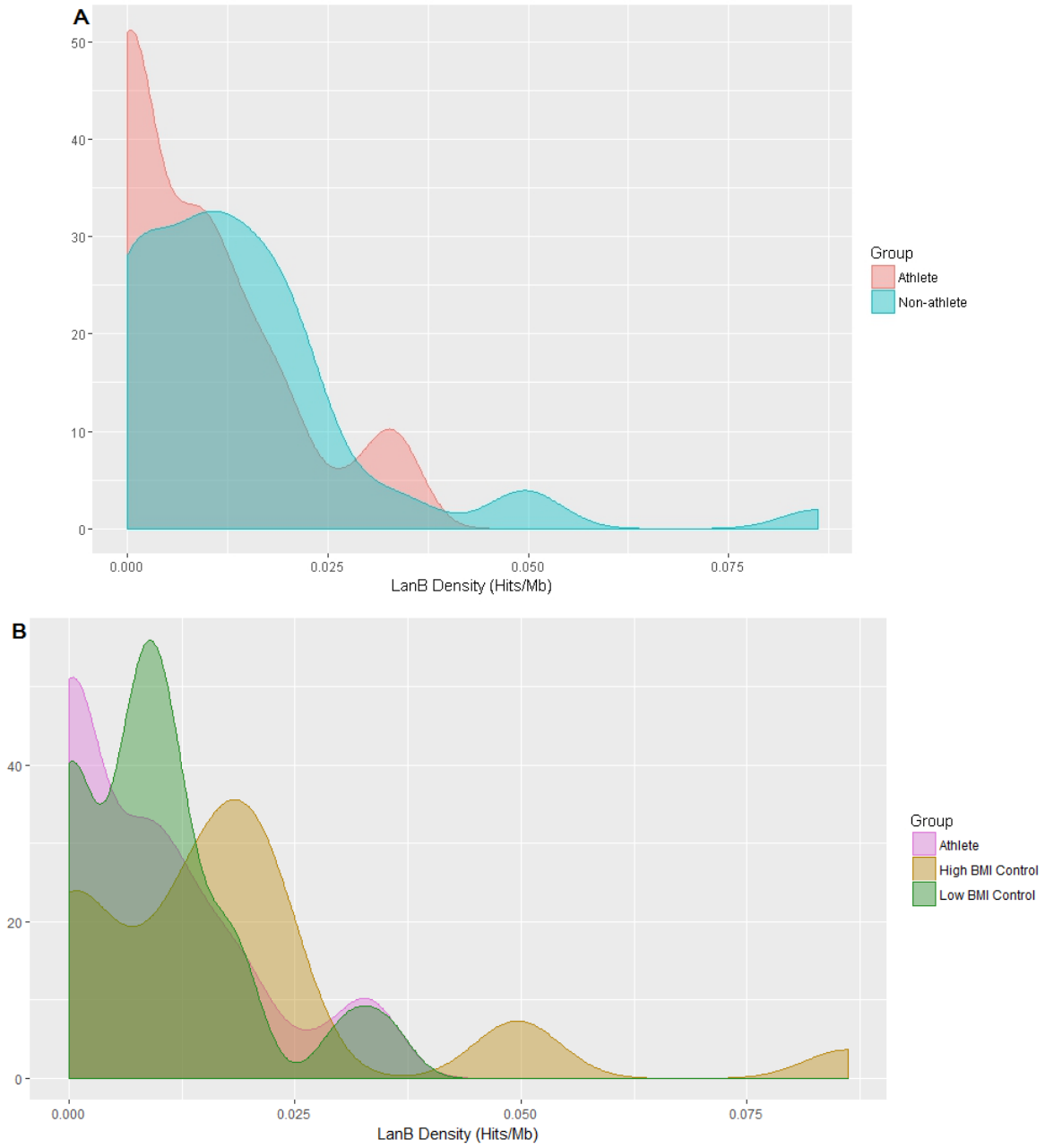


Figure 1. Density plots illustrating the distribution of LanB density when separated into: A) athletes and non-athletes; and B) athletes, Low BMI Controls, and High BMI controls.

Taxonomic classification of LanB-encoding assemblies

MetaPhlAn2 only accepts sequencing reads as input so taxonomy was assigned to the LanB-encoding scaffolds using Kaiju in MEM mode with default parameters. The phylum- and genus-level classifications are listed in Table 1 and illustrated in Figure 2. Interestingly, there was a considerably larger proportion of putative Bacteroidetes producers in athlete samples compared to non-athletes (36.6% vs 22.3%), even though the two groups had comparable Bacteroidetes levels (Wilcoxon rank sum test, $p = 0.2475$). At genus-level, athletes showed a much larger proportion of putative *Chryseobacterium* lantibiotic producers than non-athletes although this genus was not detected by MetaPhlAn2. Athletes also possessed a number of apparent lantibiotic-producing *Robseburia* species which were absent in non-athletes. This genus is known for butyrate-production in the gut microbiota, a function previously highlighted as increased in the microbiota and faecal metabolite profile of athletes (Barton et al., 2017). Finally, athletes also contained a higher proportion of putatively lantibiotic-producing *Dorea*, a genus previously negatively associated with markers for insulin resistance (10.1038/nutd.2015.9). The genus-level classification of putative LanB-encoding scaffolds are described in Table 2 and the relative proportions of the highlighted subclass I lantibiotic producers are illustrated in Figure 3.

An inverse relationship with LanB density becomes apparent as lean mass increases

Linear regression reported a negative relationship between lean mass and LanB density in athletes ($p = 0.005647$) that was not present in non-athletes ($p = 0.8904$). This was most likely due to athletes possessing a higher lean mass than non-athletes (Welch Two Sample T-test, $p = 1.898e-14$) as a local polynomial regression model (LOESS), illustrated in Figure 4, shows this negative relationship begins to appear as the lean mass of non-athletes increases. There was no observed relationship between LanB density and either BMI or waist-hip ratio.

Group	Athlete	Non-athlete	Total
	LanB hits assigned to taxon		
Phylum			
Actinobacteria	3	7	10
Bacteroidetes	15	19	34
Euryarchaeota	0	1	1
Firmicutes	19	45	64
Proteobacteria	2	2	4
Unassigned	0	3	3
Unassigned at Rank	2	8	10
Genus			
Agrobacterium	1	0	1
Alistipes	6	9	15
Bacillus	1	2	3
Bacteroides	2	3	5
Blautia	3	7	10
Chitinophaga	0	1	1
Chryseobacterium	6	1	7
Clostridioides	0	1	1
Clostridium	1	2	3
Desulfomicrobium	1	0	1
Dorea	3	1	4
Erysipelatoclostridium	0	1	1
Eubacterium	0	2	2
Faecalibacterium	0	2	2
Flavonifractor	0	1	1
Frankia	1	1	2
Janibacter	1	1	2
Methanosarcina	0	1	1
Niabella	0	1	1
Paenibacillus	0	2	2
Prevotella	1	0	1
Ralstonia	0	2	2
Roseburia	2	0	2
Ruminococcus	1	1	2
Tyzzerella	2	5	7
Unassigned	0	3	3
Unassigned at Rank	5	32	37
Unclassified *Clostridiales*	0	1	1
Unclassified *Lachnospiraceae*	4	2	6

Table 2. Number of LanB producers classified by Kaiju at Phylum and Genus levels in athletes and non-athletes.

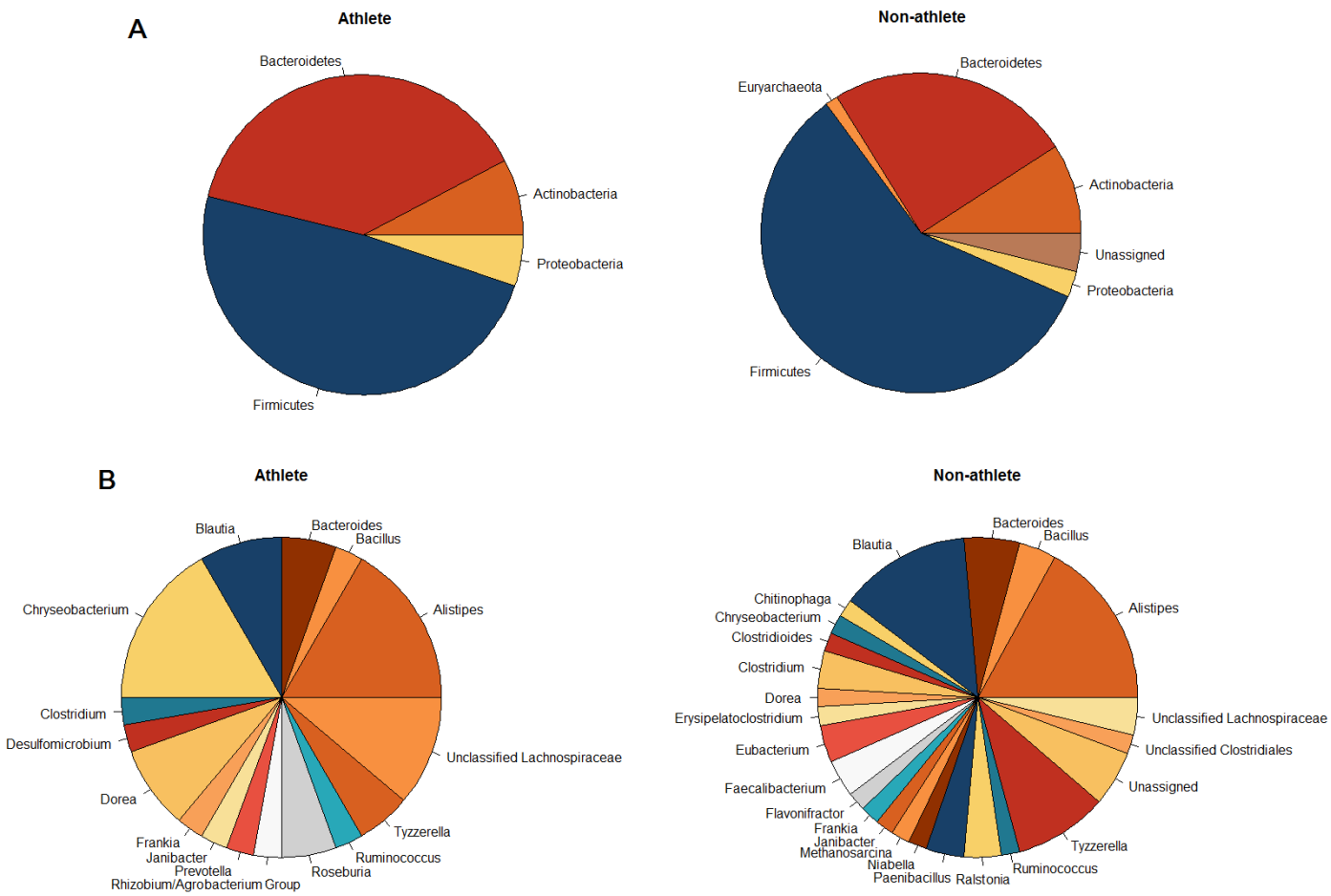


Figure 2. Pie charts depicting taxonomy of LanB producers classified by Kaiju at Phylum (A) and Genus (B) levels in athletes and non-athletes. All sequences 'unclassified at rank' are removed.

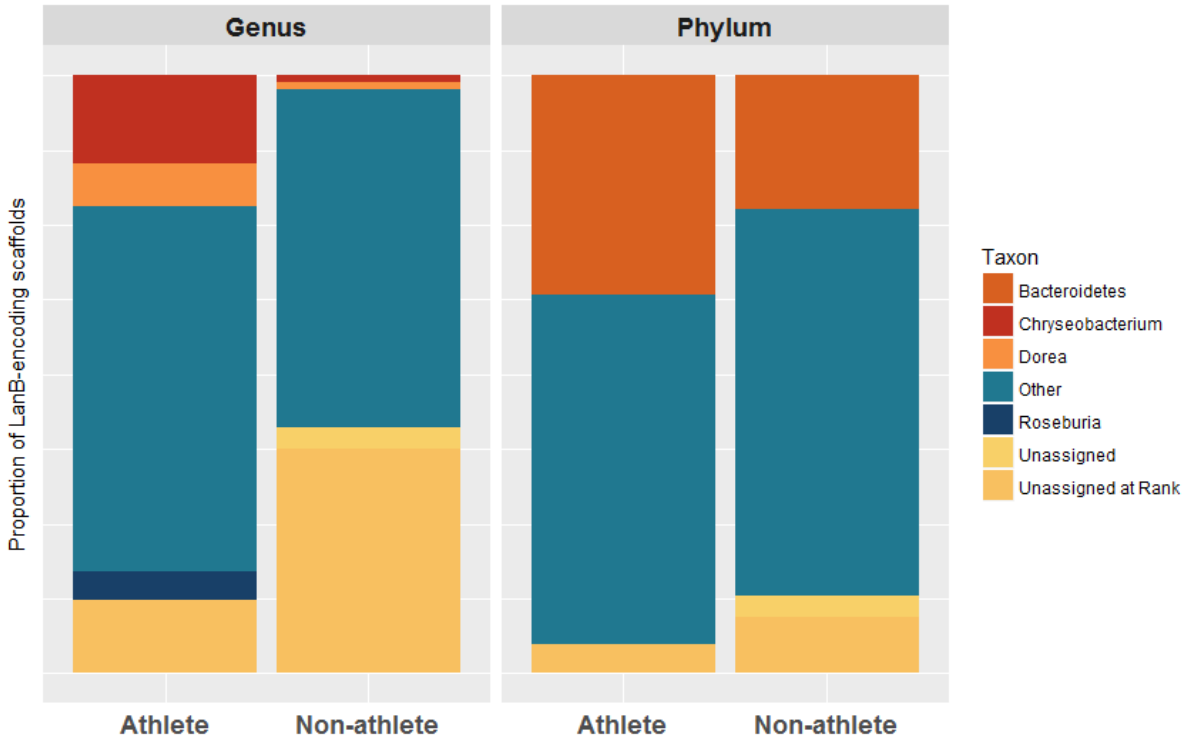


Figure 3. Proportion of LanB-producing scaffolds classified as selected taxa by Kaiju in athletes and non-athletes.

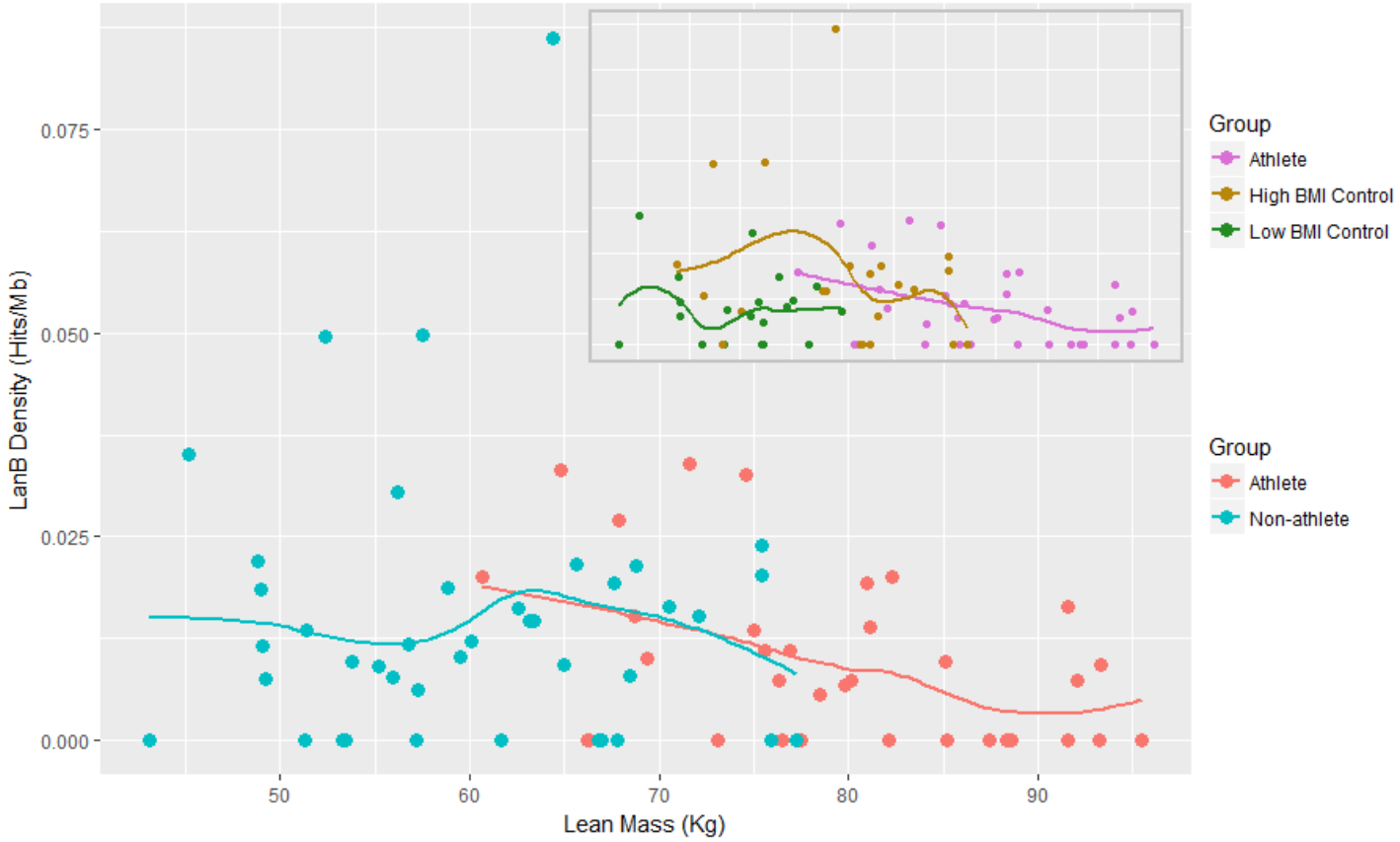


Figure 4. Scatterplot of lean mass versus LanB density. Each point represents a sample and is coloured by group. Lines are local regression (LOESS) lines and are coloured by group. The main plots shows the data when divided into athletes and non-athletes. The insert shows the data when separated into athletes, low BMI controls and high BMI controls.

LanB density is associated with the functional capacity of the gut microbiota in athletes

A negative relationship was observed between LanB density and microbial gene family alpha diversity in athletes (p = 0.03737) but not in non-athletes (p = 0.8943). Interestingly, this was the sole aspect of the overall function and composition of the microbiota, as established by HUMAnN2, where athletes could be separated from non-athletes by alpha diversity (Simpson index increased in non-athletes, Wilcoxon rank sum test, p = 0.0176), beta diversity (PERMANOVA, R^2 = 0.12603, p = 0.001), and within-group variance of beta diversity (increased in athletes, ANOVA, p = 0.0004132). These diversity measures are illustrated in Figure 5. Neither microbiota composition nor microbiota-encoded pathways data significantly separated the groups by the above diversity measures.

Similarly, microbial gene family data possessed the only discriminatory power when the samples were divided into three groups. The negative relationship between LanB density and microbial gene family alpha diversity observed in athletes was not present in either high or low BMI control groups, while low BMI controls had a higher microbial gene family alpha diversity compared to athletes (Simpson index, Wilcoxon rank sum test, p = 0.036). Again, within-group variance of beta diversity was higher in athletes compared to both low BMI and high BMI control groups (Tukey multiple comparisons of means, adjusted p = 0.0038949 and 0.0155110, respectively). These diversity measures are illustrated in Figure 6.

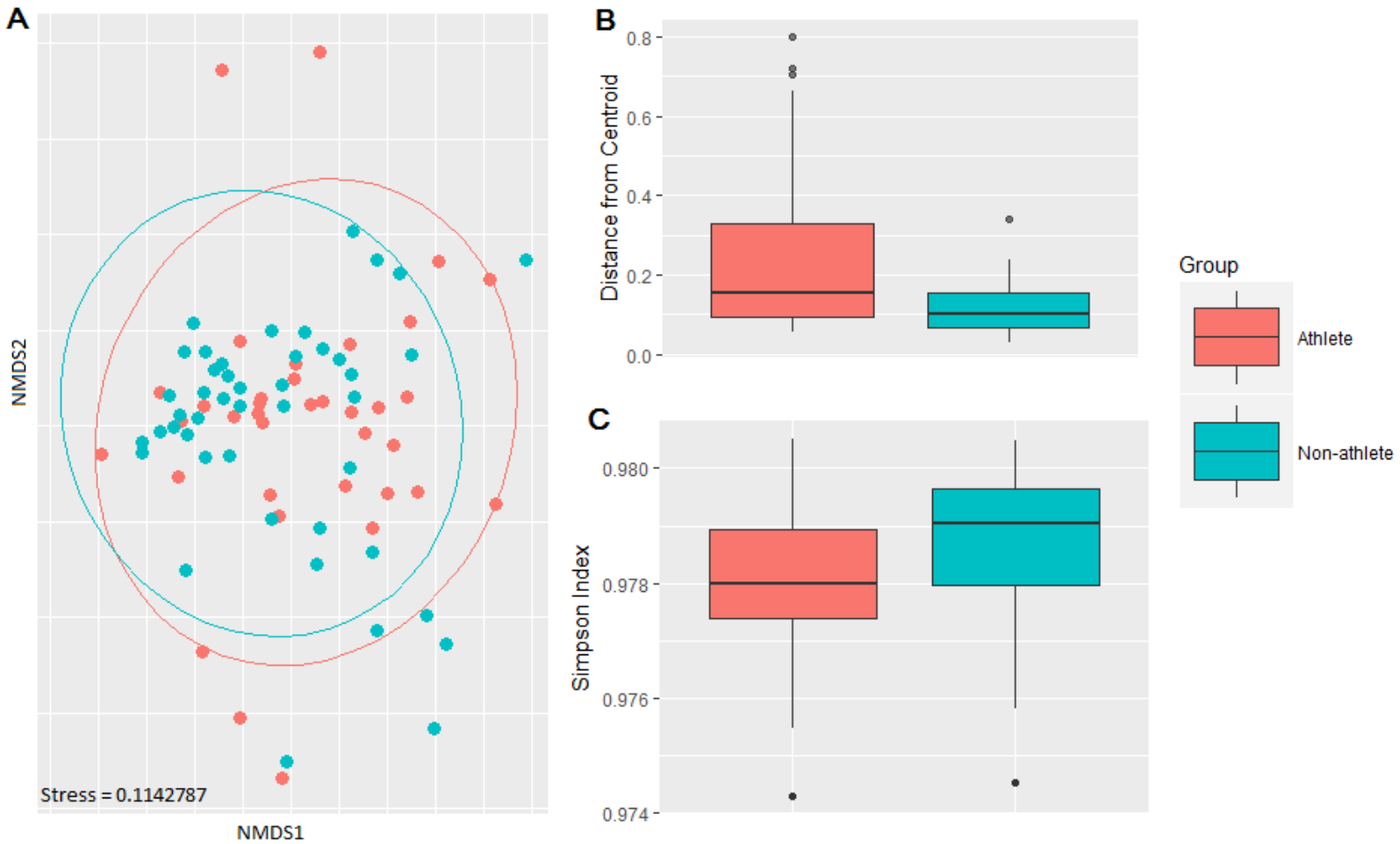


Figure 5. Illustration of diversity measures used to compare athletes to non-athletes based on gene families data. A) NMDS plot with 95% confidence ellipses. B) Within-group variance of beta diversity calculated by distance of each point to the centroid of the group. C) Alpha diversity measured by Simpson's index.

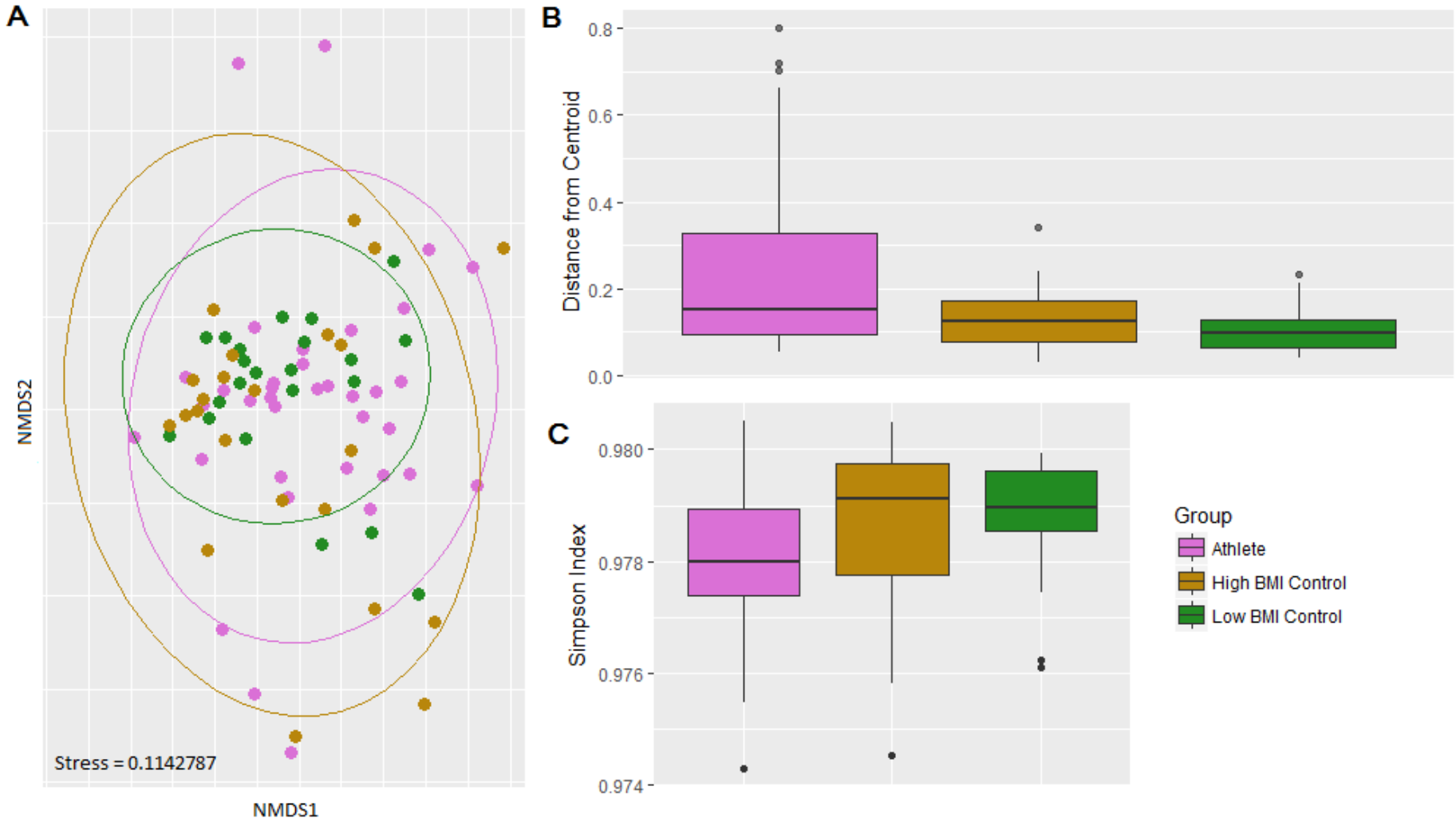


Figure 6. Illustration of diversity measures used to compare athletes, low BMI controls, and high BMI controls based on gene families data. A) NMDS plot with 95% confidence ellipses. B) Within-group variance of beta diversity calculated by distance of each point to the centroid of the group. C) Alpha diversity measured by Simpson's index.

LanB density was observed to have a relationship with the overall function of the gut microbiota in athletes (ANOVA, $p = 0.03896$) but not in non-athletes (ANOVA, $p = 0.9468$). This relationship was identified using the first principal component generated from Bray-Curtis dissimilarities of microbial gene families (Figure 5A).

LanB density shows distinct relationships with host diet and gut microbiota function in athletes and non-athletes

Correlating 2367 variables describing host physiology, diet, microbiota composition and microbiota function with LanB density identified four significant (FDR < 0.1) correlations in athletes and one in non-athletes. Across all samples, ten variables were significantly correlated with LanB density, the subject's height, daily intake of cholesterol, fibre, energy, starch, fat, carbohydrate, protein and sugar, and a microbial pathway involved in cell wall biosynthesis (MetaCyc pathway 6386). All of these correlations are negative with the exception of pathway 6386 and are listed in Table 3 and illustrated in Figure 7.

Cholesterol intake data was not available for non-athletes so it cannot be said for certain that is has no relationship with LanB density in that cohort.

A machine learning-based approach identified variables with the ability to predict LanB density

Random Forests classification and the feature selection algorithm Boruta identified two microbiota-encoded functions and two dietary variables as possessing significant power to predict LanB density across all samples - MetaCyc pathway 6386, involved in cell wall biosynthesis and shown above to be correlated with LanB density in athletes, gene family GO:0000776, responsible for kinetochore structure, daily carbohydrate intake, and daily fibre intake. Multiple linear regression using these four variables to predict LanB density identified a significant relationship ($R^2 = 0.1939$, p

Variable correlated with LanB Density (Hit/Mb)	Correlation Coefficient	FDR-adjusted *p*
All Samples		
Cholesterol (g/day)	-0.418670754	0.020509335
PWY 6386 - UDP N-acetylmuramoyl pentapeptide biosynthesis II (lysine containing)	0.411994762	0.093108528
Height	-0.382731731	0.016012188
Fibre (g/day)	-0.373864145	0.010531227
Energy (KJ/day)	-0.338875963	0.0129467
Starch (g/day)	-0.334036601	0.0129467
Fat (g/day)	-0.315403048	0.016915465
Carbohydrate (g/day)	-0.304987026	0.017568527
Protein (g/day)	-0.29964693	0.017568527
Sugars (g/day)	-0.228857567	0.073583226
Athlete Only		
PWY 6386 - UDP N-acetylmuramoyl pentapeptide biosynthesis II (lysine containing)	0.623529441	0.03349097
Height	-0.496343064	0.078961767
Starch (g/day)	-0.430163152	0.071782674
Cholesterol (g/day)	-0.418670754	0.071782674
Non-athlete Only		
Fibre (g/day)	-0.407337063	0.057472077

Table 3. Variables that significantly correlated with LanB density across all samples, athlete samples only, and non-athlete samples only.

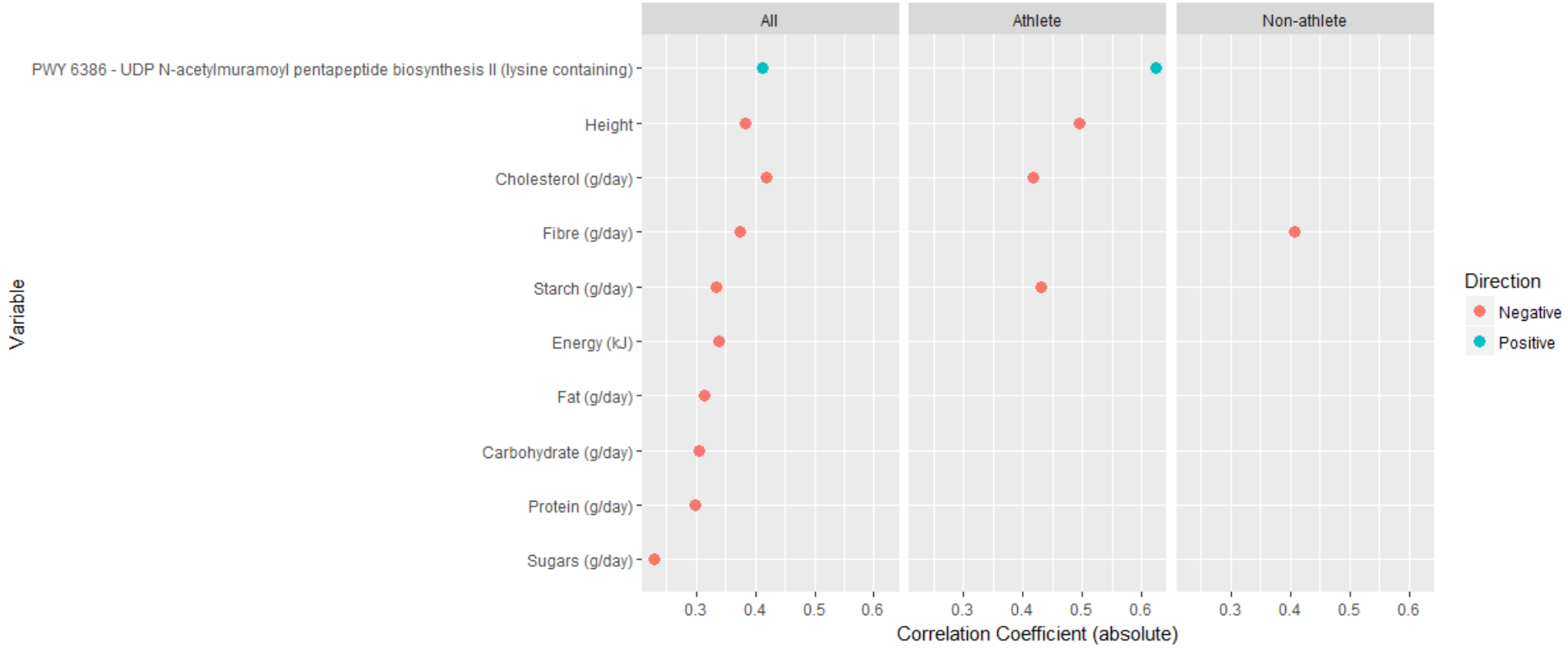


Figure 7. Dotplot of variables that significantly correlated with LanB density across all samples, athlete samples only, and non-athlete samples only. Each dot is position according to the absolute value of the correction coefficient and coloured by direction of the correlation.

= 0.0005796). The individual relationships between these variables and LanB density are illustrated in Figure 8.

GO:0000776 was increased in non-athletes compared to athletes (Wilcoxon rank sum test, p = 0.005326) and the non-athlete group appeared to be responsible for driving the association as the relationship found in that group (p = 0.006237) was absent in athletes (p = 0.6092). Carbohydrate intake and Fibre intake were both increased in athletes compared to non-athletes (Wilcoxon rank sum test, p = 6.026e-05 and 0.006865, respectively). There was no observable difference in MetaCyc pathway 6386 abundance between the two groups.

Discussion

The previously developed profile HMM was employed in this study to assess the distribution of *lanB*-containing gene clusters across individuals known to differ with respect to gut microbiota composition and function.

The main driving force behind this study was to investigate if the extreme diet and exercise regimen of these elite-athletes, two factors known to have a significant impact on the function and composition of the gut microbiota, was reflected in altered subclass I lantibiotic production within this intestinal community. One strategy often employed in identifying potential probiotics is associating the presence or absence of a particular taxa with a desirable phenotype, so any bacteriocin-producers unique to the athlete group could be theorised as playing a role in maintaining this microbiota profile and be highlighted for further investigation. The profile HMM developed previously in this thesis was employed to identify putative subclass I lantibiotic biosynthetic clusters. The profile HMM was used to screen assembled metagenomes as the lanthionine dehydratase protein is 993 amino acids in length (Uniprot ID

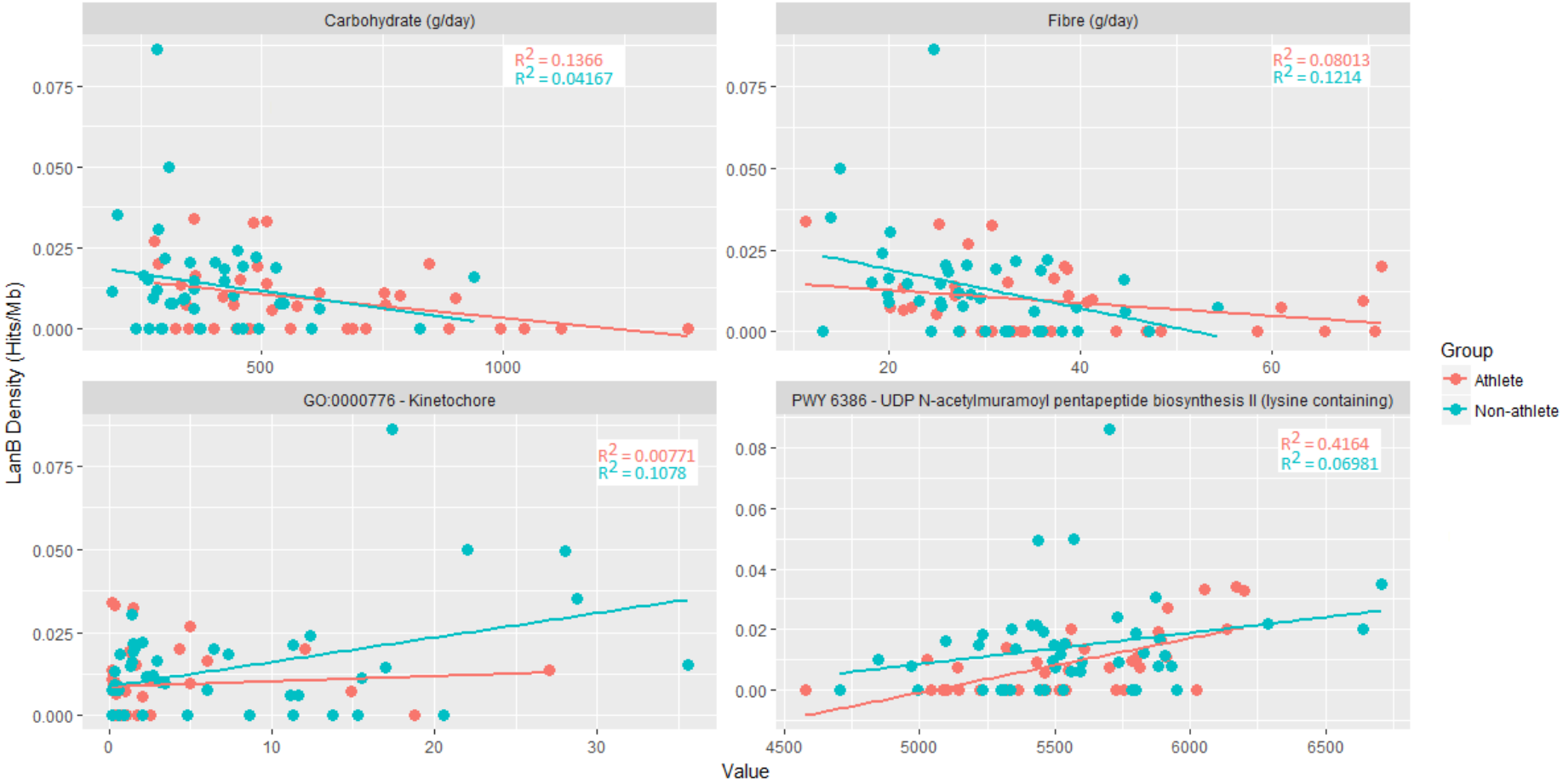


Figure 8. Scatterplot of the four variables identified by Boruta as possessing power to predict LanB density across all samples. Each point represents a sample and is coloured by group. Lines are linear regression lines to show direction of the relationship, and the R-squared value denotes the proportion of variation explained by the linear model, both are coloured by group.

P20103), meaning that only the unique number of LanB proteins in each sample are reported as metagenome assembly results in the loss of quantitative information. In theory, lantibiotic-producers in less diverse samples may be present at a higher relative abundance than would be reported, even after normalization for metagenome size, than producers in more diverse samples. However, as there was no measurable difference in compositional alpha diversity between any of the groups, this was deemed to be of no concern.

While there was no overall difference in LanB density between athletes and non-athletes, it was interesting to note the relationship between subclass I lantibiotic production and lean mass, particularly as this relationship was not present with BMI and waist-hip ratio, two of the most widely used indicators of health. BMI was expected to be of limited use in athletes as it often overestimates body fat in heavily muscled individuals (https://www.cdc.gov/obesity/downloads/bmiforpactitioners.pdf). However, investigation of BMI in non-athletes, and waist-hip ratio in both groups, suggests that there is indeed no relationship between subclass I lantibiotic production and body fat. The plotting of a local regression model (Figure 4) suggests that there is a minimum lean mass below which there is no relationship with LanB density.

The previously published article examining this dataset reported that athletes had the highest mean abundance across 29 of 34 differentially expressed metabolic categories described by MetaCyc pathways (Barton et al., 2017). These pathways (e.g. amino acid and antibiotic biosynthesis, and carbohydrate metabolism), in combination with faecal metabolite data, pointed towards enrichment of microbiota-encoded functions resulting in enhanced muscle turnover and overall health (Barton et al., 2017). A potential explanation for reduced alpha diversity of athletes compared to

non-athletes in microbial gene families data could be specialisation of the microbiota to manage such extreme levels of physical activity.

Correlations showed that diet and microbiota function have the greatest relationship with subclass I lantibiotic production in the human gastrointestinal tract. As mentioned in chapter 1 of this thesis, diet has a substantial impact on the gut microbiota and this is most likely the driving factor behind the observed relationship. In addition, these results mirror repeated observations that the functionality of the microbiota is more important than its composition. An interesting observation taken from these correlations is that there is no overlap in the reported relationships between the two groups, suggesting that athletes possess a different relationship with subclass I lantibiotic production than the rest of the population. There was also some evidence of different taxa of bacteriocin-producers between the two groups. The microbiota of athletes appeared to contain some taxa which have previously been associated with functions beneficial to human health such as short-chain fatty acid biosynthesis and reduced insulin resistance, a feature which may give them a competitive advantage in this population, allowing them to survive and confer these benefits to the host.

It is also interesting to note such a small proportion of the 2367 tested variables were significantly correlated with LanB frequency, a trend also identified by the machine-learning-based approach which identified just four variables with predictive power, implying that the association between subclass I lantibiotic production and the microbiota is relatively subtle but that host diet has the strongest influence on this relationship.

The positive relationship between LanB density and MetaCyc pathway 6386, identified by both correlation and machine learning approaches, is particularly intriguing. This pathway is involved in biosynthesis of cell wall peptidoglycan,

specifically the conversion of UDP-N-acetyl-glucosamine to UDP-MurNAc-pentapeptide (UDPMurNAc) (Munch and Sahl, 2015). UDPMurNAc is a precursor to Lipid II, the target of many antimicrobial peptides, including the subclass I lantibiotics under investigation here. One would expect that increased subclass I lantibiotic density would accompany lower lipid II biosynthesis; however the opposite is evident here. Microbiotas with higher LanB density also tended to contain increased levels of this pathway. A possible explanation could be increased peptidoglycan biosynthesis in the microbiota as a defensive or compensatory response to subclass I lantibiotic production, although metatranscriptomics would be needed to investigate this further. Another possible explanation is that, as this pathway is associated with Gram-positive bacteria (Firmicutes and Actinobacteria) and the largest proportion of identified bacteriocin producers belonged to the Firmicutes phylum, subclass I lantibiotic production may be conferring a competitive advantage to these producers who are intrinsically resistant to their own bacteriocin. Although a definitive mechanism cannot be confirmed, this is an intriguing observation that warrants further investigation to better understand the bacteriocin-production dynamic in the intestinal microbiota.

The work here highlights a potential relationship between subclass I lantibiotic production within the gut microbiota and the host. It also suggests that production of these antimicrobial compounds may affect overall microbiota function as well as individual functions. It is most likely, however, that these observations are a result of specialisation of the gut microbiota by the extreme diet and exercise regime of elite-athletes.

References

ARNISON, P. G., BIBB, M. J., BIERBAUM, G., BOWERS, A. A., BUGNI, T. S., BULAJ, G., CAMARERO, J. A., CAMPOPIANO, D. J., CHALLIS, G. L., CLARDY, J., COTTER, P. D., CRAIK, D. J., DAWSON, M., DITTMANN, E., DONADIO, S., DORRESTEIN, P. C., ENTIAN, K.-D., FISCHBACH, M. A., GARAVELLI, J. S., GORANSSON, U., GRUBER, C. W., HAFT, D. H., HEMSCHEIDT, T. K., HERTWECK, C., HILL, C., HORSWILL, A. R., JASPARS, M., KELLY, W. L., KLINMAN, J. P., KUIPERS, O. P., LINK, A. J., LIU, W., MARAHIEL, M. A., MITCHELL, D. A., MOLL, G. N., MOORE, B. S., MULLER, R., NAIR, S. K., NES, I. F., NORRIS, G. E., OLIVERA, B. M., ONAKA, H., PATCHETT, M. L., PIEL, J., REANEY, M. J. T., REBUFFAT, S., ROSS, R. P., SAHL, H.-G., SCHMIDT, E. W., SELSTED, M. E., SEVERINOV, K., SHEN, B., SIVONEN, K., SMITH, L., STEIN, T., SUSSMUTH, R. D., TAGG, J. R., TANG, G.-L., TRUMAN, A. W., VEDERAS, J. C., WALSH, C. T., WALTON, J. D., WENZEL, S. C., WILLEY, J. M. & VAN DER DONK, W. A. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports,* 30**,** 108-160.

BARTON, W., PENNEY, N. C., CRONIN, O., GARCIA-PEREZ, I., MOLLOY, M. G., HOLMES, E., SHANAHAN, F., COTTER, P. D. & O'SULLIVAN, O. 2017. The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. *Gut*.

BRESSA, C., BAILÉN-ANDRINO, M., PÉREZ-SANTIAGO, J., GONZÁLEZ-SOLTERO, R., PÉREZ, M., MONTALVO-LOMINCHAR, M. G., MATÉ-MUÑOZ, J. L., DOMÍNGUEZ, R., MORENO, D. & LARROSA, M. 2017. Differences in gut microbiota profile between women with active lifestyle and sedentary women. *PLOS ONE,* 12**,** e0171352.

CLARKE, S. F., MURPHY, E. F., O'SULLIVAN, O., LUCEY, A. J., HUMPHREYS, M., HOGAN, A., HAYES, P., O'REILLY, M., JEFFERY, I. B., WOOD-MARTIN, R., KERINS, D. M., QUIGLEY, E., ROSS, R. P., O'TOOLE, P. W., MOLLOY, M. G., FALVEY, E., SHANAHAN, F. & COTTER, P. D. 2014. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*.

ESTAKI, M., PITHER, J., BAUMEISTER, P., LITTLE, J. P., GILL, S. K., GHOSH, S., AHMADI-VAND, Z., MARSDEN, K. R. & GIBSON, D. L. 2016. Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome,* 4**,** 42.

KNOLL, R. L., FORSLUND, K., KULTIMA, J. R., MEYER, C. U., KULLMER, U., SUNAGAWA, S., BORK, P. & GEHRING, S. 2016. Gut microbiota differs between children with Inflammatory Bowel Disease and healthy siblings in taxonomic and functional composition - a metagenomic analysis. *American Journal of Physiology - Gastrointestinal and Liver Physiology*.

MARCHESI, J. R., ADAMS, D. H., FAVA, F., HERMES, G. D. A., HIRSCHFIELD, G. M., HOLD, G., QURAISHI, M. N., KINROSS, J., SMIDT, H., TUOHY, K. M., THOMAS, L. V., ZOETENDAL, E. G. & HART, A. 2015. The gut microbiota and host health: a new clinical frontier. *Gut*.

MUNCH, D. & SAHL, H. G. 2015. Structural variations of the cell wall precursor lipid II in Gram-positive bacteria - Impact on binding and efficacy of antimicrobial peptides. *Biochim Biophys Acta,* 1848**,** 3062-71.

QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., MANICHANH, C., NIELSEN, T., PONS, N., LEVENEZ, F., YAMADA, T., MENDE, D. R., LI, J., XU, J., LI, S., LI, D., CAO, J., WANG, B., LIANG, H., ZHENG, H., XIE, Y., TAP, J., LEPAGE, P., BERTALAN, M., BATTO, J.-M., HANSEN, T., LE PASLIER, D., LINNEBERG, A., NIELSEN, H. B., PELLETIER, E., RENAULT, P., SICHERITZ-PONTEN, T., TURNER, K., ZHU, H., YU, C., LI, S., JIAN, M., ZHOU, Y., LI, Y., ZHANG, X., LI, S., QIN, N., YANG, H., WANG, J., BRUNAK, S., DORE, J., GUARNER, F., KRISTIANSEN, K., PEDERSEN, O., PARKHILL, J., WEISSENBACH, J., BORK, P., EHRLICH, S. D. & WANG, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464**,** 59-65.

REA, M. C., DOBSON, A., O'SULLIVAN, O., CRISPIE, F., FOUHY, F., COTTER, P. D., SHANAHAN, F., KIELY, B., HILL, C. & ROSS, R. P. 2011. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences,* 108**,** 4639-4644.

RICHARDS, J. L., YAP, Y. A., MCLEOD, K. H., MACKAY, C. R. & MARINO, E. 2016. Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases. *Clin Trans Immunol,* 5**,** e82.

SUBIRATS BAYEGO, E., SUBIRATS VILA, G. & SOTERAS MARTINEZ, I. 2012. [Exercise prescription: indications, dosage and side effects]. *Med Clin (Barc),* 138**,** 18-24.

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Walsh, C. 2017. An in silico analysis of bacteriocin production in the human microbiota and its relationship with health. PhD Thesis, University College Cork.

Please note that Chapter 5 (pp.222-269) is unavailable due to a restriction requested by the author.

CORA Cork Open Research Archive http://cora.ucc.ie

General Discussion

As discussed in Chapter 1 of this thesis, many human diseases can be traced to the disrupted function of the gut microbiota, thereby making targeted manipulation of this community, through diet, probiotics, or antimicrobials, a logical intervention for treatment of these conditions. The initial challenge relates to identifying the particular dietary interventions, probiotic species, or antimicrobial compounds that can bring about the desired alterations. With this in mind, chapters 2 and 3 of this thesis focus on the *in silico* analysis of the gut microbiota with a view to identifying bacteriocins, or bacteriocin-producers, within the intestinal community that may have this potential. Regardless of whether the microbiota is regarded as a 'hidden organ' or a co-evolving symbiont, it is in its interest to protect the host, and itself, from individual pathogenic organisms and detrimental shifts in microbiota profile. Therefore, the bacteriocin-producing component of this community is the logical place to search for biotherapeutics to ensure gut health.

The initial, genomics-focused, approach in Chapter 2 employed the widely-used BAGEL3 software to search genomic data, representative of the human gastrointestinal microbiota, which was collected by the Human Microbiome Project. This screen identified 74 putative bacteriocin-encoding gene clusters (PBGCs) from 382 sequenced genomes following manual examination. The study identified potential next-generation probiotics by detecting PBGCs in taxa not classically associated with bacteriocin production such as *Bacteroides* spp. and *Roseburia* spp. Of particular note were putative sactibiotic gene clusters identified in three *Bacteroides* species with members currently under early stages of therapeutic investigation. Preclinical *in vitro* work has suggested that strains of *B. dorei* and *B. fragilis* isolated from the human microbiota have potential applications in cholesterol reduction (Gerard *et al.* 2007) and regulation of the immune system (Deng *et al.* 2016), respectively. A sactibiotic gene cluster was also identified in a *B. uniformis* strain, a species which has recently showed promise in treatment of metabolic and immunological dysfunction in mice

with high-fat-diet inducted obesity (Cano *et al*., 2012). There have been no reports of bacteriocin production by any of these three strains.

Chapter 3 focused on another *in silico* approach centred on designing, validating and implementing a profile HMM to investigate a specific category of bacteriocins, subclass I lantibiotics, in metagenomic data, again using data collected by the Human Microbiome Project. The profile HMM was sensitive enough to identify all positive controls during the validation step while also being specific enough to not return any false positives from similar proteins involved in modification of other lantibiotic subclasses. The model showed that subclass I lantibiotic production is not equal across body sites, implying that bacteriocin production may confer a greater competitive advantage in some microbial niches compared to others. Once taxonomy was assigned to those LanB-encoding scaffolds, it was apparent that the reference genome database searched previously was not fully representative of the human gut microbiota. Just 40% of the putative bacteriocin-producers found through the HMM analysis were present in the genomic database, suggesting that rarer members of the microbiota, or those not previously perceived as important, may be influencing it's functionality, thus highlighting more novel candidates for further therapeutic assessment.

The main motivation for designing a custom profile HMM was that this approach is ideally suited to searching metagenomic data, thereby facilitating a wider search for novel bacteriocin clusters. The increased sensitivity of profile HMMs over sequence-homology approaches, such as a BLAST-based approach, allowed detection of more distantly related proteins in relatively under-characterised members of the gut microbiota, while also overcoming the limitation of tools tailored to genomic data, such as BAGEL3, which struggle with the fragmented nature of metagenomic assemblies as they rely on identifying multiple proteins related to biosynthesis within a set genomic region. These advantages meant that the model was suited to comparing the density and distribution of putative bacteriocin gene

clusters in elite athletes, investigated in chapter 4 of this thesis, with that of a sedentary population, and investigating the factors influencing bacteriocin production in both cohorts. Perhaps unsurprisingly, diet showed the strongest relationship with putative subclass I lantibiotic production within in the gut microbiota. More interesting, however, was the relationship between bacteriocin gene density and the lean mass of the host. Across all subjects, individuals with a higher lean mass tended to have a lower density of unique bacteriocin producers than their lighter counterparts. This was particularly evident in the athletes whose extreme diet and exercise regimens appeared to have focussed their microbiota to support muscle turnover and overall fitness.

The final chapter of this thesis investigated how metabolic health could be altered in a diet-induced obesity mouse model by probiotic feeding, with a particular focus on the gut microbiota through extensive bioinformatics approaches. The outcomes of feeding the two strains under investigation, both members of the *Lactobacillus* genera, were startlingly different. *L. plantarum* AH0315 had a profound impact on the overall composition and function of the gut microbiota without any accompanying change in host metabolic phenotype, while *L. casei* AH0077 significantly improved several metabolic markers without an overall impact on the gut microbiota's overall composition and function. In addition to these observations made from the overall profile of the gut microbiota, there were some alterations by both probiotics of specific taxa and microbial-encoded pathways. The current hypothesised mechanism by which AH0077 brings about these changes is thought to involve EPS-production and lipid-binding but further work is needed to confirm this. Correlation analysis between individual features (both compositional and functional) of the population and metabolic measurements taken from the host supported the current thinking that the function of the gut microbiota is of greater importance to health and disease than its composition. Biomarker discovery using the LEFSe algorithm on the lean control group also

provided further evidence to some taxa gaining attention as next-generation probiotics for the treatment of obesity such as *Akkermansia* and members of the *Christensenellaceae* (Derrien *et al*., 2016, Goodrich *et al*., 2014).

Taken together, this thesis identifies many potential bacteriocin producers from the human microbiota with the potential to influence the structure of this population in the treatment of disease and improvement of overall gastrointestinal and indeed possibly metabolic health. Further work will be needed to confirm functionality and characterise these putative bacteriocin-producers and select those with the desired target spectrum. This work also shows how probiotic feeding can influence the metabolic health of an obese host in a strain-specific manner either independent of the gut microbiota or in a targeted manner – altering specific functions without influencing its overall composition and function.

References

CANO, P.G., SANTACRUZ, A., MOYA, Á. & SANZ, Y., 2012. Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. *PloS one*, 7(7), p.e41079.

DENG, H., LI, Z., TAN, Y., GUO, Z., LIU, Y., WANG, Y., YUAN, Y., YANG, R., BI, Y., BAI, Y. & ZHI, F., 2016. A novel strain of Bacteroides fragilis enhances phagocytosis and polarises M1 macrophages. *Scientific Reports*, 6.

DERRIEN, M., BELZER, C. & DE VOS, W.M., 2016. Akkermansia muciniphila and its role in regulating host functions. *Microbial pathogenesis*, 106, pp. 171-181.

GÉRARD, P., LEPERCQ, P., LECLERC, M., GAVINI, F., RAIBAUD, P. & JUSTE, C., 2007. Bacteroides sp. strain D8, the first cholesterol-reducing bacterium isolated from human feces. *Applied and environmental microbiology*, 73(18), pp.5742-5749.

GOODRICH, J.K., WATERS, J.L., POOLE, A.C., SUTTER, J.L., KOREN, O., BLEKHMAN, R., BEAUMONT, M., VAN TREUREN, W., KNIGHT, R., BELL, J.T. & SPECTOR, T.D., 2014. Human genetics shape the gut microbiome. *Cell*, 159(4), pp.789-799.