University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Elucidating the Microbiome of Irish Athletes

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

Doctor of Philosophy

By

Ciara M. O' Donovan, B.Sc.

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

APC Microbiome Ireland, University College Cork, Cork, Ireland

School of Microbiology, University College Cork, Cork, Ireland

Submitted February 2020

Head of School: Prof. Paul O' Toole

Supervisors: Prof. Paul Cotter, Dr. Orla O' Sullivan, Dr. Susan Joyce, Dr. Cormac Gahan

Agriculture and Food Development Authority

Microbiome Ireland
Interfacing Food & Medicine

Table of Contents

Declaration ..... **i**

Thesis Abstract ..... **ii**

Publications ..... **iv**

List of Abbreviations ..... **v**

List of Figures ..... **ix**

List of Tables ..... **xi**

List of Supplementary Material ..... **xii**

Chapter 1 ..... **1**

Gut Microbiology - A Relatively Unexplored Domain

1.0 Abstract 2

1.1 Introduction 3

1.2 Diversity of the gut microbiota 4

1.2.1 Differential colonization throughout the gastrointestinal tract 4

1.2.2 Stability of the gut microbiota 10

1.2.2.1 Birth mode 11

1.2.2.2 Early life 13

1.2.2.3 First foods 14

1.2.2.4 Healthy aging 15

1.2.2.5 Geographical effect 17

1.2.3 Factors influencing the microbiota 17

1.2.3.1 Drugs 18

1.2.3.2 Antibiotics 19

1.2.3.3 Exercise 21

1.2.3.4 Diet 23

1.3 Role of the gut microbes on health 26

1.3.1 Role of the normal gut microbiota in immune development 26

1.3.2 Metabolic potential of the gut microbes and its impact on the host 30

1.3.3 Role of the gut microbiota in disease 32

1.3.3.1 Gut-brain axis 32

1.3.3.2 IBD 34

1.3.3.3 Allergies 35

1.3.3.4 Cancer 36

1.3.4 Perturbations to the intestinal microbiota and its relation to disease 37

1.4 Complementation of the gut microbiota 38

1.5 Tools to study the microbiota-host interaction 41

1.5.1 Culture-based techniques 41

1.5.2 Considerations with faecal sample analysis 42

1.5.3 Sequencing technologies 44

1.5.4 Metabolomics 46

1.5.5 Transcriptomics 48

1.5.6 Animal models 49

1.5.7 *In vitro* gut models 50

1.6 Conclusion 52

1.7 References 53

Chapter 2 64

'Microbes in Sport' – The Potential Role of the Gut Microbiota in Athlete Health and Performance

2.0 Abstract 65

2.1 Introduction 66

2.2 The gut microbiota and general athlete health 67

2.3 The gut microbiota and musculoskeletal health 68

2.4 Potential role of the gut microbiota in athletic performance 68

2.5 Conclusions 69

2.6 References 70

Chapter 3 76

Instances of Altered Gut Microbiomes Among Irish Cricketers Over Periods of Travel in the Lead up to the 2016 World Cup: A Sequencing Analysis

3.0 Abstract 77

3.0.1 Background 77

3.0.2 Methods 77

3.0.3 Results 77

3.0.4 Conclusions 77

3.1 Highlights 78

3.2 Introduction 79

3.3 Methods 81

3.3.1 Compliance with ethical standards 81

3.3.2 Experimental design 81

3.3.3 DNA extraction and library preparation 82

3.3.4 Bioinformatic analysis 83

3.4 Results 85

3.4.1 Timing of gastrointestinal distress symptoms may influence stability of the gut microbiome 85

3.4.2 Individual taxa are unique between baseline and post-India time points 88

3.4.3 Individual taxa are associated with gastrointestinal distress symptoms 91

3.4.4 Antibiotic resistance potential is altered as a consequence of travel to India 91

3.5 Discussion 98

3.6 Conclusion 102

3.7 Acknowledgements 104

3.8 References 105

3.9 Supplementary material 113

Chapter 4 130

Distinct Microbiome Composition and Metabolome Exists Across Subgroups of Elite Irish Athletes

4.0 Abstract 131

4.0.1 Objectives 131

4.0.2 Methods 131

4.0.3 Results 131

4.0.4 Conclusions 132

4.1 Practical implications 133

4.2 Introduction 134

4.3 Methods 135

4.3.1 Ethical approval 135

4.3.2 Experimental design 135

4.3.3 DNA extraction from faecal samples 136

4.3.4 Library preparation 136

4.3.5 Bioinformatic analysis 137

4.3.6 Metabolomics analysis 138

4.3.6.1 ^{1}H-NMR spectroscopy metabolic profiling analysis 138

4.3.6.2 GC-MS urinary targeted analysis 139

4.3.6.3 UPLC-MS urinary and faecal analysis 140

4.3.6.4 Data analysis 140

4.4 Results 141

4.5 Discussion 154

4.6 Conclusion 156

4.7 Acknowledgements 158

4.8 References 159

4.9 Supplementary material 164

Chapter 5 205

Exploring Faecal Bile and Fatty Acids, and Bile Acid Modifications in Elite Athletes

5.0 Abstract 206

5.1 Introduction 208

5.2 Methods 210

5.2.1 Experimental design 210

5.2.2 Bile and fatty acid extraction 210

5.2.3 Ultra-performance liquid chromatography tandem mass spectrometry 210

5.2.4 Bile and fatty acid data analysis 211

5.2.5 Functional metagenomic bank creation 211

5.2.6 Functional metagenomic bank screening 213

5.2.7 Bile salt hydrolase profiling 213

5.2.8 Fosmid extraction and sequencing preparation 214

5.2.9 Bioinformatic analysis 215

5.3 Results 217

5.3.1 BA receptor agonists are significantly different between groups, while several individual FAs are also significantly different 218

5.3.2 Investigating the BSH activity of an athlete functional metagenomic bank reveals 5 distinct BSH clusters 226

5.4 Discussion 234

5.5 Acknowledgements 239

5.6 References 240

5.7 Supplementary material 246

Chapter 6 257

The Gut Microbiome of Runners Reflects Maximal Oxygen Consumption (VO_2 max) and the Amount of Training per Week

6.0 Abstract 258

6.1 Introduction 259

6.2 Methods 261

6.2.1 Ethical approval 261

6.2.2 Study design 261

6.2.3 Maximal oxygen consumption test 261

6.2.4 Exercise challenge to induce intestinal permeability 262

6.2.5 Faecal sample collection and DNA extraction 262

6.2.6 DNA library preparation and sequencing 263

6.2.7 Bioinformatic analysis 266

6.3 Results 267

6.3.1 The gut microbiome of runners clusters by maximal oxygen consumption (VO_2 max) and training frequency 267

6.3.2 While a core microbiome exists, a number of different species differ significantly in relative abundance in line with VO_2 max and training frequency 267

6.3.3 Functional potential of the gut microbiome differs significantly between individuals from different VO_2 max and training frequency groups 275

6.4 Discussion 281

6.5 Acknowledgements 285

6.6 References 286

6.7 Supplementary material 290

Chapter 7 421

General Discussion

7.1 References 427

Acknowledgements 430

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 (except as declared, hereafter), and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: Ciara O P

Ciara O' Donovan

Student Number: 115224014

Date: 24/02/2020

Thesis Abstract

Exercise has been established as a factor that has a beneficial impact on several body systems, including the prevention and treatment of obesity, and alleviation of symptoms affecting the cardiovascular system. A variation in the gut microbiome composition and functionality of athletes relative to that of the general population has been identified. Numerous factors have the potential to alter the gut microbiome of athletes, including type of sport, travel for competition, and training intensity. This thesis explores these factors, while also assessing how they impact on microbial and host metabolism.

After reviewing several aspects of the topic in chapters 1 and 2, in chapter 3, we investigate the impact of travel on the athlete gut microbiome using 16S rRNA and shotgun metagenomic sequencing approaches. In chapter 4, we evaluate the differences in the gut microbiome and metabolome between athletes participating in different types of sports. Chapter 5 focuses on profiling the bile and fatty acid content and bile salt hydrolase potential of athletes and controls and, finally, in chapter 6, we employ shotgun metagenomic sequencing to analyse the composition and functional potential of the gut microbiome in an active cohort to determine the impact of activity level and fitness.

Overall, this thesis explores the factors that influence an athlete's gut microbiome, and indicates a role for travel as well as sports type and activity level on gut microbiome composition, functional potential, and metabolism.

Publications

- **Ciara M. O' Donovan**, Brendan Connor, Sharon M. Madigan, Paul D. Cotter, Orla O' Sullivan, *Instances of altered gut microbiomes among Irish cricketers over periods of travel in the lead up to the 2016 World Cup: A sequencing analysis*. Travel Medicine and Infectious Disease, 2020. https://doi.org/10.1016/j.tmaid.2020.101553.
- **Ciara M. O' Donovan**, Sharon M. Madigan, Isabel Garcia-Perez, Alan Rankin, Orla O' Sullivan, Paul D. Cotter, *Distinct microbiome composition and metabolome exists across subgroups of elite Irish athletes*. Journal of Science and Medicine in Sport, 2020. 23(1): 63-68. https://doi.org/10.1016/j.jsams.2019.08.290.
- **Ciara M. O' Donovan**, Orla O' Sullivan, Paul D. Cotter, *Gut Microbiology – A Relatively Unexplored Domain*. Reference Module in Life Sciences, Elsevier, 2018. ISBN: 978-0-12-809633-8, http://dx.doi.org/10.1016/B978-0-12-809633-8.09233-5.
- Alan Rankin, **Ciara O' Donovan**, Sharon M. Madigan, Orla O' Sullivan, Paul D. Cotter, *'Microbes in sport' –The potential role of the gut microbiota in athlete health and performance.* British Journal of Sports Medicine, 2017. 51(9): 698–699. http://dx.doi.org/10.1136/bjsports-2016-097227.

List of Abbreviations

Abbreviation:	Meaning:
°C	Degrees Celsius
1H-NMR	Proton nuclear magnetic resonance
AAD	Antibiotic associated diarrhoea
APCs	Antigen-presenting cells
AST	Aspartate aminotransferase
BA	Bile acids
BH	Benjamini-Hochberg
BMI	Body mass index
BMTagger	Best match tagger
bp	Base pair
BSH	Bile salt hydrolase
CAAD	*Clostridium difficile* associated antibiotic associated diarrhoea
CD	Crohn's disease
cfu mL^{-1}	Colony forming unit per millilitre
cfu	Colony forming unit
cm	Centimetre
CRC	Colorectal cancer
CK	Creatine kinase
CREC	Clinical Research Ethics Committee
CV	Collision voltage
DCs	Dendritic cells
DNA	Deoxyribonucleic acid
e.g.	Example
ENA	European Nucleotide Archive
EPA	Eicosapentaenoic acid
EPIC	European Prospective Investigation into Cancer
ESBLs	Extended spectrum β-lactamases
FA	Fatty acids
FDA	Food and Drug Administration
FDR	False discovery rate

FETA	Food frequency questionnaire European Prospective Investigation into Cancer and nutrition tool for analysis
FFQ	Food frequency questionnaire
FMT	Faecal microbiota transplant
FXR	Farnesoid X receptor
g	Gram
GALT	Gut associated lymphoid tissue
GC-MS	Gas chromatography-mass spectrometry
GDCA	Glycodeoxycholic acid
GI	Gastrointestinal
GIT	Gastrointestinal tract
GM	Gut microbiota
h	Hour(s)
HBD2/DEB2	Human β-defensin 2
HDL	High-density lipoproteins
HUMAnN2	Human microbiome project unified metabolic analysis network
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
ICH-GCP	International conference on harmonization E6 good clinical practice
IECs	Intestinal epithelial cells
IL	Interleukin
kb	Kilobase
km h^{-1}	Kilometres per hour
LB	Luria bertani broth
LCFA	Long chain fatty acid
LDA	Linear discriminant analysis
LDL	Low-density lipoproteins
LEfSe	LDA effect size
LPS	Lipopolysaccharides
L/R	Lactulose/rhamnose permeability test
LXR	Liver X receptor
m	Metre
MAMPs	Microbial-associated molecular patterns

MAPK	Mitogen-activated protein kinase
MCFA	Medium chain fatty acid
MDS	Multidimensional scaling
mg	Milligram
min	Minute(s)
mL	Millilitre
NAFLD	Non-alcoholic fatty liver disease
NCBI	National Centre for Biotechnology Information
NF-KB	Nuclear factor KB
NGS	Next generation sequencing
NMDS	Non-metric multidimensional scaling
NLRs	Nucleotide-binding oligomerization domain-like receptors
OPLS-DA	Orthogonal partial least squares discriminant analysis
OTU	Operational taxonomic unit
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PFGE	Pulse field gel electrophoresis
pH	Power of hydrogen
PPAR	Peroxisome proliferator-activated receptors
PPI	Proton pump inhibitor
PPRs	Pattern recognition receptors
PSA	Polysaccharide A
PXR	Pregnane X receptor
QC	Quality control
QIIME	Quantitative insights into microbial ecology
R^2	Coefficient of determination
RBBC	Repeated bead beating plus column
RLRs	RIG-I-like receptors
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RPKM	Reads per kilobase of reference sequence per million sample reads

rpm	Revolutions per minute
s	Second(s)
SCFA	Short chain fatty acid
SCGs	Sports classification groups
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
T2D	Type 2 diabetes
TD	Traveller's diarrhoea
TDCA	Taurodeoxycholic acid
TGFβ	Transforming growth factor β
TGR5	G-protein-coupled bile acid receptor
TH	T-helper
TH2	T-helper type 2
TLRs	Toll-like receptors
TNF-α	Tumour necrosis factor alpha
Tregs	Regulatory T cells
UAE	United Arab Emirates
UC	Ulcerative colitis
UPLC-MS	Ultra-performance liquid chromatography-tandem mass spectrometry
VDR	Vitamin D receptor
VO_2 max	Maximal oxygen consumption
vVO_2 max	Minimum running velocity required to achieve VO_2 max
v/v	Volume per volume
Vs.	Versus
WMS	Whole metagenome shotgun sequencing
w/v	Weight per volume
xg	Gravitational force
µg	Microgram
µg mL^{-1}	Microgram per millilitre
µL	Microlitre
µM	Micromolar

List of Figures

Chapter 1

Figure 1.1| Factors influencing the gut microbiome 12

Chapter 3

Figure 3.1| Alpha diversity fluctuations occur during travel 86

Figure 3.2| Individual variability in microbiome stability 89

Figure 3.3| Significant differences in composition based on time point 90

Figure 3.4| Shotgun metagenomic sequencing reveals unique *Escherichia coli* strains between participants post-India 92

Figure 3.5| Significant differences in composition based on reported gastrointestinal distress symptoms 94

Figure 3.6| An increase in antibiotic resistance potential is seen in some individuals post-India 96

Chapter 4

Figure 4.1| Diversity and composition of the gut microbiome 143

Figure 4.2| Composition and functional potential of the gut microbiome vary based on sports classification groups 147

Figure 4.3| Metabolomic analysis reveals separation in microbial and human derived metabolites based on sports classification groups 151

Chapter 5

Figure 5.1| Fatty acid but not bile acid profiles are significantly different between groups 219

Figure 5.2| Significantly different bile and fatty acids correlate with each other 220

Figure 5.3| Receptor agonists vary by group 222

Figure 5.4| Bile salt agar plate assay 227

Figure 5.5| Bile salt hydrolase profiling of a fosmid metagenomic bank reveals 5 distinct groupings 228

Figure 5.6| Significant differences in bile acids between fosmid groupings 230

Chapter 6

Figure 6.1| Compositional and functional potential clusters based on VO_2 max and training frequency 268

Figure 6.2| Compositional profile of samples 269

Figure 6.3| Significantly different species based on VO_2 max 271

Figure 6.4| Significantly different species based on training frequency 273

Figure 6.5| Significantly different pathways based on VO_2 max 277

Figure 6.6| Significantly different pathways based on training frequency 279

List of Tables

Chapter 5

Table 5.1| Clones picked for sequencing 216

Table 5.2| Correlations of bile acids with taxa, pathways, and metadata 225

Table 5.3| Classification of sequenced fosmid clones of interest and identification of features of interest 233

Chapter 6

Table 6.1| Participant characteristics 264

Table 6.2| Classification of VO_2 max by sex and age 265

List of Supplementary Material

Chapter 3

Supplementary document 3.1| Participant information and consent form 113

Supplementary document 3.2| Gut health questionnaire 118

Supplementary figure 3.1| Timeline 123

Supplementary figure 3.2| Variations in beta diversity occur in a subset of travelling athletes over time 124

Supplementary figure 3.3| Taxonomic profiles before and after travel to India 126

Supplementary figure 3.4| Virulence markers in the gut microbiome are increased as a consequence of travel to India 128

Chapter 4

Supplementary document 4.1| Consent form 164

Supplementary document 4.2| Gut health questionnaire 170

Supplementary document 4.3| Food frequency questionnaire 175

Supplementary figure 4.1| Alpha diversity 185

Supplementary figure 4.2| Multidimensional scaling (MDS) result from UPLC-MS analysis 186

Supplementary figure 4.3| OPLS-DA analysis of GC-MS urine data 187

Supplementary figure 4.4| OPLS-DA analysis of NMR urine data 189

Supplementary figure 4.5| OPLS-DA analysis of NMR faeces data 190

Supplementary table 4.1| Participant characteristics and data collected 191

Supplementary table 4.2| Responses to gut health questionnaire 192

Supplementary table 4.3| Responses to food frequency questionnaires converted into dietary intake per day using FETA (Food Frequency Questionnaire European Prospective Investigation into Cancer and Nutrition Tool for Analysis) 194

Supplementary table 4.4| Dietary intakes compared using Kruskal-Wallis between sports classification groups 202
Supplementary table 4.5| Significant differences in metabolites overall 204

Chapter 5

Supplementary figure 5.1| Characteristics of participants by group 246
Supplementary figure 5.2| Significantly different inflammatory cytokines between groups 248
Supplementary table 5.1| Bile acid moieties analysed by UPLC-MS in this study 249
Supplementary table 5.2| Fatty acid moieties analysed by UPLC-MS in this study 251
Supplementary table 5.3| Statistics between groups for all bile acids analysed 252
Supplementary table 5.4| Statistics between groups for all fatty acids analysed 255
Supplementary table 5.5| Significant differences between heatmap clusters and controls 256

Chapter 6

Supplementary figure 6.1| Experimental design 290
Supplementary figure 6.2| Gut permeability differs based on VO_2 max, but not training frequency 291
Supplementary figure 6.3| Alpha and beta diversity over time 293
Supplementary figure 6.4| Alpha diversity measures by VO_2 max classification and training frequency 294
Supplementary figure 6.5| Number of significant differences overall 296
Supplementary table 6.1| Inclusion and exclusion criteria 297
Supplementary table 6.2| BMI classifications 299
Supplementary table 6.3| Taxa of interest 300

Supplementary table 6.4| Species significantly different between VO_2 max classification groupings 303

Supplementary table 6.5| Species significantly different between training frequency groupings 379

Supplementary table 6.6| Species trending towards increased or decreased with increasing VO_2 max grouping 405

Supplementary table 6.7| Species trending towards increased or decreased with increasing training frequency 407

Supplementary table 6.8| Pathways significantly different between VO_2 max classification groupings 409

Supplementary table 6.9| Pathways significantly different between training frequency groupings 416

Chapter 1

Gut Microbiology - A Relatively Unexplored Domain

Included as published in *Reference Module in Life Sciences*

(doi: http://dx.doi.org/10.1016/B978-0-12-809633-8.09233-5)

Authors: Ciara M. O' Donovan, Orla O' Sullivan, and Paul D. Cotter

Contributions:

- **Candidate** wrote the review, with guidance from **OOS** and **PDC**.

1.0 Abstract

The human gastrointestinal tract is home to a complex community of microorganisms which interact with the host in a symbiotic relationship. The gut microbiota plays a role in both health and disease in the human host; modulating the immune system, assisting with digestion of complex compounds, production of enzymes and metabolites, and interactions with other organs. Although studies have revealed great interpersonal variation in composition, the functional capacity of the gut microbiome shares commonality between individuals. The use of a variety of techniques in the study of the human gut microbiome, including culture-based techniques, sequencing technologies, metabolomics, and transcriptomics, has allowed for a better understanding of the gut microbiome from a holistic perspective.

1.1 Introduction

The gastrointestinal tract (GIT), whose primary purpose is digestion of food and absorption of nutrients, encompasses the oral cavity, oesophagus, stomach, and the small and large intestines. Although all organs of the GIT are populated with microorganisms [1-4], the colon has been the focus for the majority of microbiome studies owing both to its dense population in comparison to other areas of the GIT (Colon: 10^{12} cfu mL^{-1}; Jejunum: 10^{5} cfu mL^{-1}) [3] and the interactions between the colonic microbiome and the human host. The GIT is colonized by a complex community of microorganisms from all domains of life; Bacteria, Archaea, and Eukaryota, as well as their associated viruses. The majority of studies of the populations within the GIT have focused on the bacterial domain [5]. However, the viral component of the human gut microbiome, which is mainly composed of bacteriophages, is numerous and far more diverse than bacterial counterparts [6].

While the relationship between the gut microbiota and host has often been viewed in terms of a commensal relationship, most members within these communities are more appropriately described as symbionts [7]. A warm and nutrient rich environment is provided by the human host, while the resident gut microbiota assist in the digestion of complex compounds, generate bioactives used by the host within the body, and prevent colonization of the GIT by potentially pathogenic microorganisms [8]. The gut microbiota has coevolved with humans and play roles both within and beyond the GIT. Through assisting in the digestion process, it is possible for the host to obtain nutrients which may not otherwise have been utilized within the food. The breakdown of starch and undigested or unabsorbed sugars into short chain fatty acids (SCFAs) is an important process within the colon and provides 10% of energy intake. The small intestinal microbiota also plays a role in the digestion of food, producing many of the enzymes used in the breakdown of food. Beyond the GIT the gut microbiome also performs many other important functions

including modulating the immune system from birth, vitamin production, interaction with the brain through the gut-brain axis, and the production of important and useful metabolites [9].

An advanced understanding of the interactions between the gut microbiome and the human host has led to much focus on these interactions. While early research focused purely on the compositional differences in the gut microbiota between healthy and diseased populations or across age groups, the rise in the use of 'omics'-based studies has allowed for an understanding from a more holistic approach. The study of bacterial populations through the use of traditional culture-based methods is time consuming, and due to the complex nature of the GIT many have yet to be cultured. For this reason novel techniques have been established; metagenomics, metatranscriptomics, metaproteomics, and metabolomics technologies, used alone or in combination, allow for an understanding of the gut microbiome beyond taxonomic profiling [10]. Gut microbiome research has not only been able to determine the differences between the microbial profile in healthy individuals in comparison to those who are suffering from GIT disorders, such as inflammatory bowel disease (IBD), but also beyond this to diseases which are not associated with the GIT such as asthma and anxiety. Research into these disciplines has allowed not only for a greater understanding but also for the move towards the development of treatments for such disorders.

1.2 Diversity of the gut microbiota

1.2.1 Differential colonization throughout the gastrointestinal tract

All parts of the GIT; the mouth, oesophagus, stomach, and the small and large intestines, are colonized by microorganisms [1-4,11-16]. Food is digested and absorbed as it moves through the GIT, changing nutrient availability, this along with the changing pH and oxygen

availability results in the creation of several distinct habitats and thus the selection for different microbes in the different organs of the GIT.

Within the oral cavity several distinct habitats exist on the different oral structures including the teeth, gingival sulcus, attached gingiva, lip, hard and soft palates, cheek, and tongue. The oral cavity acts as a major entry point to the body with both food and air entering through the mouth. Due to this constant flow of microbes from the surrounding environment it may be difficult to establish which members are resident. Other surrounding environments which are connected to the oral cavity (such as the nasal passages, tonsils, and middle ear) also influence the microbial population here. 96% of the oral microbiome is composed of taxa from 6 phyla (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Spirochaetes, and Fusobacteria). Members of the genus *Streptococcus* are the most abundant in the oral cavity with others including *Abiotrophia*, *Gemella*, and *Granulicatella* also present [1].

The role of the oesophagus is the movement of the bolus, created in the mouth through chewing of food and mixing with saliva, from the mouth into the stomach. The proximal oesophagus closest to the oral cavity contains a microbiome similar to the oral microbiome, while the distal oesophagus closest to the stomach contains a microbiome made up of a mixture of communities from the stomach and oral environments. Due to the invasive nature of sampling, few studies have focused on the microbial characterisation of the oesophagus microbiome outside of studies of diseased populations. However, the gathering of control group samples in these studies has allowed for the establishment of an understanding of the core microbes of the oesophagus. *Streptococcus* is consistently reported in high proportions in studies of the oesophageal microbiome in both culture-dependent and -independent studies using different sampling techniques. *Prevotella*, *Fusobacterium*, and *Veillonella* are also frequently reported, however in lower proportions

[15]. A movement away from this typical oesophageal microbiome has been found in individuals suffering with oesophageal diseases such as eosinophic esophagitis and in individuals with poor muscle function at the junction between the stomach and the oesophagus [2].

The human stomach is the harshest environment within the GIT; with a pH of 2, a thick mucus layer, strong peristalsis, and acid production all occurring in the stomach it is unsurprising that it was long thought to be inhospitable by microbes [4]. Despite the presence of a low bacterial load this organ is home to a diverse bacterial population. Community structure varies between the mucus layer and the gastric fluid. Differences between the different sections of the stomach (cardia, fundus, body, antrum, and pylorus) have also been identified in animal models [11]. Investigations of the human gastric mucosa are difficult and rare in the absence of medical reasoning. Biopsies of the mucus layers in humans require for the biopsy of all sites sequentially and so differences between sites have yet to be fully elucidated. Gastric fluid by comparison is easier to collect, however again it is rarely collected in healthy cohorts [12]. It is hypothesized that the gastric mucosal layer may more appropriately represent microbial stomach residents. This is due to the presence of both oral and jejunum microorganisms in the gastric fluid which may over represent the true stomach microbiome [4]. At phylum level the stomach microbiome is composed of Firmicutes, Proteobacteria, Actinobacteria, Fusobacteria, Bacteroidetes, and Gemmatimonadetes with the gastric mucosa dominated by Firmicutes and Proteobacteria and the gastric fluid dominated by Firmicutes, Bacteroidetes, and Actinobacteria. Although not dominant *Helicobacter pylori* is present in the gastric fluid of approximately half of all individuals. Inflammation is always present with *Helicobacter pylori* with other complications such as gastric ulcers often observed [11,12].

The small intestine encompasses three different regions; the duodenum, the jejunum, and the ileum. The stomach contents empty into the jejunum and flow through the entire length of the small intestine before emptying into the ascending colon. The small intestine is approximately 6 metres long and is the primary site of nutrient digestion and absorption in the body [17]. Much like the other areas of the GIT it is resident to a unique microbiome which produces enzymes to assist in the host's ability to utilize non-digestible carbohydrates and host derived glycoconjugates. The small intestine microbiota is also involved in bile acid modifications, cholesterol reduction, and the biosynthesis of vitamins [18]. The small intestine does not have as high a bacterial load as the large intestine, containing about 10^3-10^5 cfu mL^{-1} in comparison to 10^{11}-10^{12} cfu mL^{-1} in the large intestine. The composition of the jejunum is unique from both the oral cavity and colon. Facultative anaerobes and oxygen tolerant obligate anaerobes are the primary colonizers of the small intestine with the frequent abundance of members of the Enterobacteriaceae family representing the main difference from the oral microbiome. *Streptococcus*, *Prevotella*, *Veillonella*, and *Fusobacterium* are particularly abundant with non-oral genera including *Escherichia*, *Klebsiella*, and *Citrobacter* also present [3].

Although studies of the small intestine are few, potentially due to the invasive nature of sampling, some have been completed. Differences exist between the duodenal microbiome between healthy and obese individuals at both a compositional and functional level [17]. Additionally the small intestine microbiota has been found to be different between diseased individuals and controls in diseases including irritable bowel syndrome (IBS), celiac disease, Crohn's disease (CD), pouchitis, and additionally with small bowel transplants [17-19].

Differential colonization throughout the small intestine is expected due to the functional and structural differences found along the small intestine. In humans the differentiation of

the microbiome composition throughout the small intestine has not been established due to the difficulty in obtaining samples from the different regions which would require repetitive endoscopy. However, it has been established in cattle where the different sections of the GIT, including the different sections of the small intestine, are colonized in distinct ways with considerable variations among the regions in abundance and the genera composing the predominant taxonomic groups [20]. Due to the importance of the small intestine in the digestion and absorption of food, a better understanding of the progression of microbial composition and functionality along the small intestine might be useful in better understanding microbial-food interactions and may lead to important therapeutic strategies.

Both microbial load and species diversity significantly increase at the ileum (> 10^8 cfu g^{-1}) and throughout the colon (> 10^{11} cfu g^{-1}), with communities dominated by obligate anaerobes [21]. The large intestine facilitates the reabsorption of water into the body and the microorganisms present ferment undigested starch, any unabsorbed sugars, plant cell wall polysaccharides, and mucins into the SCFAs; butyrate, acetate, and propionate [17].

The Bacteroidetes and Firmicutes dominate the community structure of the large intestines of adults (approximately 90%) while Actinobacteria, Proteobacteria, and Verrumicrobia make up minor constituents. The community structure of the gut microbiome remains relatively consistent, but the proportions and species present vary considerably between individuals [22]. Previously a suggestion of a core microbiome was made; however, the notion of this has since been weakened. This proposed core suggested three different enterotypes which were designated as predominantly *Prevotella*, *Bacteroides*, or *Ruminococcus*. However, it has since been demonstrated that the most predominant feature of this data established an inverse correlation between *Prevotella* and *Bacteroides*, while some studies have been completely unable to identify these enterotypes [5, 23]. A

core functional capacity might more appropriately describe gut microbiome structure with polysaccharide digestion, immune system development, defence against infections, vitamin synthesis, fat storage, angiogenesis regulation, and behaviour development existing as part of the core function [24].

Archaeal populations are also present in the gut microbiome with *Methanobrevibacter smithii* being the most frequent colonizer. *M. smithii* produces methane from by-products of bacterial fermentation and an accumulation of H_2 stimulates its growth. One study identified the presence of archaea in 44 out of 96 samples studied with *Methanobrevibacter* and *Nitrososphaera* species identified, usually mutually exclusively but co-existing in 6 samples. *Methanobrevibacter* was relatively more abundant in samples where it was detected in comparison to the less abundant *Nitrososphaera*. Analysis of long- and short-term dietary data identified a correlation between carbohydrate intake and *Methanobrevibacter* while long-term vegetable fat and polyunsaturated fat intake correlated with *Nitrososphaera* and the absence of archaea [5].

Next generation sequencing has identified a diverse gut fungal community; however it is difficult to differentiate between fungi of food origin and gut colonising fungi. In the study of 96 healthy individuals referred to previously, fungal sequences were detected in every sample. The most prevalent fungal genus present in this cohort was *Saccharomyces,* followed by *Candida* and *Cladosporium*. An inverse relationship was determined between the presence of *Ascomycota* and *Basidiomycota*. Fungal communities correlated with short-term but not long-term dietary intake. *Candida* abundance correlated with total carbohydrate intake and also held an inverse relationship with total saturated fatty acids. *Aspergillus* abundance was determined to have an inverse relationship with SCFAs [5]. A greater understanding of these interactions and relationships could be incredibly important especially in the manipulation of the gut microbiome for therapeutic strategies.

Both culture-dependent and -independent techniques have identified the microbiome structure variability over time and within populations. Almost every aspect of an individual's life has been found to have an impact on the colonic microbiota composition including mode of delivery, mode of feeding in infancy, antibiotic usage, and short- and long-term dietary habits, each of which will be discussed in more detail in the following sections. Despite a highly variable composition within the colon, the functional capacity is similar between individuals. Inter-individual variation in the compositional and functional components of the gut microbiome is higher in infants than in adults with an increased diversity and stability established at around three years of age. In adulthood sequential samples collected from the same individuals indicate a relatively stable, distinct community is maintained in the absence of perturbations [22]. The inter-individual variation of the gut microbiome is a very important consideration in study design to ensure enough individuals are included to reduce biases from other factors. Additionally, this variation needs to be considered in the development of therapies through these studies as inter-individual variability could react differently to treatments. Perhaps with time specific treatments will be designed appropriately depending on the gut microbiome composition.

1.2.2 Stability of the gut microbiota

Pretty much every aspect of an individual's lifestyle has the potential to impact on the gut microbial composition which is first established at birth (figure 1.1). Although throughout infancy the gut microbiome is relatively unstable, from around two or three years of age adult like stability is established. In the absence of perturbations or diseases, including IBD, obesity, IBS, type 2 diabetes (T2D), or rheumatoid arthritis, the adult microbiome remains relatively stable throughout adulthood until adjusting again in the elderly populations [25].

1.2.2.1 Birth mode

It was previously accepted that the in-utero environment was sterile with the presence of bacteria only observed with infection. However further investigations have detected bacteria in placenta tissue, umbilical cord blood, amniotic fluid, and foetal membranes from healthy new-borns where no indication of infection or inflammation is observed suggesting the potential for in-utero colonisation. However, further studies have not been able to establish in-utero colonization and so further research is needed in this area to possibly understand the implications for in-utero colonization [26].

Birth mode has been determined as a very important factor in the early colonization of infants, as whether an infant is delivered vaginally or via caesarean section could impact on gut microbiome development and even potentially future health. The microbial colonization of the infant gut is known to play a key role in immunological and metabolic pathways impacting on human health. Disruptions to this process could potentially increase disease susceptibility in later life [26]. In infants delivered via caesarean section colonizers of the gut include skin and oral microbes, as well as those from the surrounding environment during delivery such as *Enterobacter hormaechei*, *Staphylococcus saprophylicus*, and *Haemophilus influenzae*. In contrast infants, who were vaginally delivered, remained most stable over time and were enriched with microbes from the genera *Bacteroides*, *Bifidobacterium*, *Parabacteroides*, and *Escherichia/Shigella*. At one week old the gut microbiome of vaginally delivered infants is dominated by high levels of *Bifidobacterium* and *Bacteroides*, while caesarean delivered infants have an increased abundance of *Clostridium*. At this time bacterial diversity is at the lowest and with the emergence of other genera the abundance of these early colonizers begins to diminish. Although as infants age the impact of delivery mode is seen to decrease, babies delivered via caesarean section remain different even after 12 months. Infants delivered via

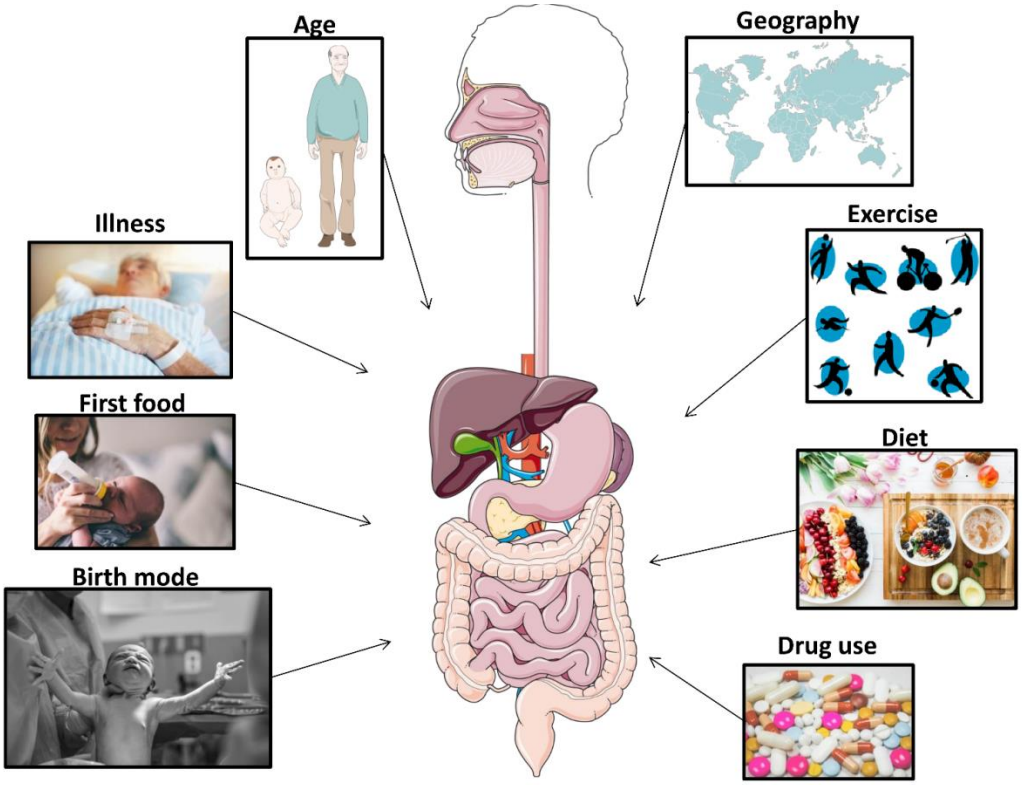


Figure 1.1| Factors influencing the gut microbiome

(Illustration created using images from Servier Medical Art and Unsplash)

caesarean section are also found to have fewer microbes shared with their own mother, likely as an impact of exposure to the delivery environment instead of an initial coating with mothers microbes from the birth canal [27, 28].

Gestational age at birth has also been determined as being a strong influencer of the early gut microbiome following birth. The effect of caesarean section was seen to be more significant on preterm infants. This is likely due to the many additional challenges which are faced by preterm infants including feeding difficulties, maternal and neonatal antibiotic exposure, gut and immune immaturity, and an increased stay in hospital. This impact seems to diminish over time as by 24 weeks of age no significant difference can be found between full term caesarean and preterm caesarean infants [28].

One study of infant and mother pairs identified strains including those from *Bifidobacterium bifidum*, *Coprococcus comes*, and *Ruminococcus bromii* that were present in samples from an infant-mother pair while being clearly distinct from those carried by other pairs indicating vertical transmission occurs between mothers and their infants. The level of vertical transmission varied between pairs and the reason for this could not be determined in such a small study [29].

1.2.2.2 Early life

The gut microbiome of an infant is constantly developing with the predominant Enterobacteriaceae, Bifidobacteriaceae, and Clostridaceae found at 2 months of age seen to decrease by 18 months, while Lachnospiraceae and Ruminococcaceae, which are barely present at 2 months, increase over the first 12 months of life. Despite this constant flux, the gut microbiome of a child in this time remains more similar to adjacent samples from the same child than to samples from age matched unrelated children [30].

Longitudinal study of the development of the infant gut microbiome has also established the greater similarity of the infants to the mother's microbiome at 12 months of age indicating progression towards a stable adult like microbiome. From a functional capacity perspective, during the first year of life the relatively simple new-born microbiome evolves into a more complex adult like configuration. At 1 year of age, infants have an increased functional similarity to mothers with less inter-individual variation existing between infants. Differences still remain between the functional capacity of the microbiome at this age between the infants and mothers, likely due to the differences in substrates available with pathways associated with carbohydrate uptake found to be enriched in new-borns while mothers have an increased fermentative capacity [27].

Several other factors play an important role in the development of the infant gut microbiome including genetics and prenatal stress. Family members share an increased number of bacterial phenotypes [26], while twin pairs share a more similar microbiome than non-twin pairs [28]. Whether this is just as a result of shared environmental exposures is yet to be established. In a rat model prenatal stress was found to impact not only on the gut microbial composition but also on the stress responses of adult offspring [31]. Similar results have been established in humans, where prenatal stress resulted in an altered gut microbiota and impacted on early life health [32].

1.2.2.3 First foods

Breastfeeding is thought to be important in colonization of human infants because we are one of the few mammals that produce milk-specific polysaccharides that are resistant to digestion in the small intestine of the infant and likely provide substrates to the developing colonic community. Both the type and quantities of breast milk polysaccharides are individual specific and are related to the Lewis blood group factor [33].

A clear difference can be observed in the compositional structure between bottle and breastfed infants at 4 months, however, the overall functional difference observed as a result of different types of feeding is small [27]. The length of breastfeeding has an impact on the microbiome of full-term caesarean delivered infants but not full-term vaginally delivered infants. Five genera were found to be more abundant in infants who were breastfed longer than 4 months while four genera were found to be more abundant in infants who were breastfed for less than 4 months. Interestingly levels of *Bifidobacterium* were not significantly different based on duration of breastfeeding [28].

The introduction of solid foods and the cessation of breastfeeding results in changes in microbial composition. Lactobacillaceae, Bifidobacteriaceae, Enterococcacea, and Enterobacteriaceae decrease while Lachnospiraceae, Ruminococcaceae, and Bacteroidaceae species increase during the period of weaning from milk onto solid foods. During the transition to solid food, an increase in alpha diversity is observed in both bottle- and breast-fed infants with both richness and evenness affected. Exclusive breastfeeding in infancy selects efficiently for specific species capable of degrading human milk oligosaccharides resulting in low diversity and *Bifidobacterium* domination which is likely to be beneficial for a child's health. The movement onto a combination of solid foods and milk, and a further transition during complete weaning results in an increased diversity and adult associated microbes belonging to Lachnospiraceae and Ruminococcaceae [34]. To fully appreciate possible health outcomes later in life, it may be beneficial to determine the optimal age at which solid foods should be introduced and which foods should be introduced first to increase diversity and assist with immune maturation.

1.2.2.4 Healthy aging

In elderly individuals, the possible reduction in number of teeth, chewing ability, taste, and intestinal transit times may have an impact on dietary choices and digestion, which in turn

can have a direct or indirect impact on the gut microbiota. One in-depth study of the gut microbiota of the elderly across different dwellings established a correlation between the healthy food diversity index, microbial diversity, and overall health status. Elderly patients in long-stay facilities are distinct with respect to their microbiome composition from community dwelling elderly individuals with the day hospital and rehabilitation individuals clustering between these two groups. The healthiest people were found to live in a community setting; they also eat differently and have distinct microbial profiles from those in long-term care. Although associations between a certain microbial profile and residence location can be established in the elderly, living in a certain location does not absolutely define microbial composition. Upon entry into long-term care facilities the elderly gut microbiome gradually changes for at least 18 months over which time the patients lose community associated genera and gain long-stay type microbiota. The long-stay individuals have an increased abundance of Bacteroidetes while those in community setting have an increased abundance of Firmicutes. A shift in the functional capacity was also found between the community and long-stay dwellers with fewer genes predicted in long-stay patients [25, 35].

Possible explanations for these changes may be related to diet- and digestion-related changes in the elderly. For example, stomach acid concentrations decrease during aging, intestinal transit time is often much longer, and there are also reports of a reduction in the ability to digest protein, which may increase nitrogen availability in the colon [36].

A greater understanding of the gradual changes in the gut microbiome during aging is important and might result in the establishment of therapies promoting the gut microbiome for healthy aging.

1.2.2.5 Geographical effect

Geographical differences in the gut microbiome can be established even in infancy. In Europe a significant difference between Finnish and German infants and also Estonian and Swedish infants has been established. Northern based European infants have been found to have an increased abundance of *Bifidobacterium* species, some *Clostridium* species, and *Atopobium* species, while southern born infants have increased abundance of *Eubacteria*, *Lactobacillus*, and *Bacteroides* [26].

In a study of 531 individuals, across 151 families from villages of Malawi and Venezuela, and the USA inter-individual variation was found to be greater among children than in adults. A significantly different compositional and functional profile was found between individuals living in the different countries studied with an especially pronounced separation found between the USA in comparison to Malawian and Amerindian individuals at all stages of life studied. A separation was observed between the USA and non-USA locations as well as between Malawi and Amerindian individuals, however, no separation was observed between the different villages or regions. Although microbial diversity was observed to increase with age overall, the microbiota of the USA adults was found to be the least diverse [37]. Geographical differences are likely to be as a result of cultural and dietary differences. These geographical differences need to be appreciated, especially in the research of the gut microbiome in relation to disease, as this effect could have implications in these studies.

1.2.3 Factors influencing the microbiota

The factors influencing the gut microbiome extend beyond life stage, associated differences observed in infancy and the elderly, and geographical differences establish the longitudinal profile of the gut microbiome. Other shorter-term influencers such as drug use, antibiotic use, exercise, and diet have all been found to influence the gut microbiome at both a

compositional and functional level. These factors may alter the stable adult microbiome in shorter periods of time.

1.2.3.1 Drugs

As drugs or their breakdown components pass through the GIT, they have the potential to elicit a response in the gut microbiome; such drugs include proton pump inhibitors. Proton pump inhibitors (PPIs) are drugs which are used to suppress gastric acid production and are used in the treatment of disorders of the GIT such as peptic ulcers and gastro-oesophageal reflux. These drugs are considered low risk and thus widely used, often unnecessarily. However, despite this, some recent studies have linked the use of PPIs with an increased risk of nutritional deficiencies, bone fractures, and infections, including enteric infections. PPIs, which only become active in the stomach, work by inhibiting hydrogen-potassium pumps which in turn increases the stomach pH. A study of the impact of PPI use on the gut microbiota of twins identified 22 OTUs (operational taxonomic unit) from the phylum Firmicutes as being significantly reduced in PPI users while 32 OTUs were increased; 20 belonging to the order Bacteroidales and 7 assigned to *Streptococcus* genus. Although antibiotic use was significantly correlated with PPI use, these changes were also observed when the cohort was subset to remove those individuals who were also on antibiotics. Interestingly those families which were increased with PPI use were those which were more often found in the mouth/throat, skin/nose, and vaginal sites indicating the impact of an increased gastric pH on the passage of these microbes from other sites into the intestines. Additionally, PPI use is correlated with an increase in abundance of *Streptococcus* and a decrease in abundance in Lachnospiraceae, both of which have been found to predispose colonization of *Clostridium difficile*, which may explain why PPI users have an increased risk of such infections [38].

In addition to the impact of drugs on the gut microbial composition, the gut microbiome can alter the disposition, efficacy, and toxicity of administered drugs through a variety of mechanisms. Enzymes produced by microbes may activate or inactivate drugs, bacteria may directly bind to drugs resulting in their removal from the GIT, or microbial and host metabolites of a drug may directly compete for a host enzyme [39]. Depending on the drug-microbiome interactions this may be beneficial, such as where modification by the gut microbiota allows for enhanced absorption, or detrimental, such as where modification leads to the production of undesirable products. Important consideration needs to be given to these potential interactions during drug development.

1.2.3.2 Antibiotics

The use of antibiotics is associated with a range of side effects including an impact on the gut microbiome, with effects dependent on factors including type of antibiotic administered [40]. Antibiotic use in early life has been correlated with an increased risk of metabolic and immunological diseases with suggestions that a disrupted microbiome plays a role in this. Antibiotics are the most commonly used drug in paediatric patients and early life use increases the risk of the development of conditions in later life including IBD, obesity, and asthma. In a study of Finnish children, antibiotic use in early life had an impact on the gut microbiota. More specifically, both penicillin and macrolide use resulted in a decrease in microbial richness with the use of the later not showing any recovery even 12-24 months after use. Macrolide use in childhood was found to be associated with long-term distortions in composition, function, and antibiotic resistance profile of the intestinal microbiota. The use of penicillin type antibiotics was not associated with large microbiome wide functional or compositional changes in the microbiome. Macrolide type antibiotics seemed to have a significant effect on the children beyond the microbiome also with frequent use in the first 2 years of life associated with asthma and a high BMI [41].

The development of necrotising enterocolitis, an intestinal disease which primarily affects premature infants, has been associated with early life antibiotic and H_2 blocker use and early colonization by *Clostridium*. A study of short- and long-term antibiotic use and the influence this has on the gut microbiome in susceptible infants was completed in an attempt to understand disease progression. A reduction in alpha diversity was observed at day 10 in infants on longer-term antibiotic exposure but this is no longer seen at day 30. This indicates that the gut continues to acquire more microbes despite antibiotic pressure. This study identified a more pronounced change in patients who received longer antibiotic treatment, which is to be expected. In this study no evidence suggests prolonged impact of early life antibiotic exposure on the gut microbiota, however longitudinal studies need to be completed for this to be completely established [42]. Further study of the impact of early antibiotic use on the gut microbiome established a modest reduction in richness with antibiotic use which was only observed after the first year of life. Additionally, treatment with antibiotics resulted in a less stable community but this recovered within one month [30].

In adults the oral administration of antibiotics has been found to impact on the gut microbiota, resulting in a decrease in diversity and an altered composition [41] with recovery from a single antibiotic exposure usually observed within two weeks however, multiple treatments can cause the time frame to expand considerably [30, 43]. Additional side effects on the GIT, such as antibiotic associated diarrhoea (AAD), which will be discussed in later sections, can occur as a result of antibiotic administration [40]. In contrast with the prolonged impact on the colonic microbiome, the oral microbiome has been found to be more resilient to antibiotics, with only short-term and minor changes experienced [14]. This difference is likely due to the difference in the length of exposure between the two sites and the likely encapsulation of the antibiotic before reaching the stomach.

The gut microbiome resistome can also be altered through antibiotic administration with these effects potentially outlasting any compositional alterations observed [43]. The impact of antibiotic use on antibiotic resistance is dependent upon the location of the resistance genes. For chromosomally encoded genes, a rapid increase of antibiotic resistance genes is observed during treatment with a quick decline after withdrawal of the antibiotics. For antibiotic resistance genes that are encoded episomally, an increase is still observed with antibiotic treatment but no sharp decline is observed upon withdrawal and instead antibiotic resistance genes continue for long periods of time [30].

1.2.3.3 Exercise

Exercise has been determined to have impacts both within the GIT and systemically, with effects on the immune system and brain-gut axis identified. With the implications of exercise in modulating the gut microbiota there is a potential application for this as a cheap and simple to implement therapy for the treatment of disorders associated with changes to the gut microbiome.

A comparison of male elite professional rugby players with low and high BMI controls found lower inflammatory markers and improved metabolic markers in the athlete group. Athletes in this study were found to have a more diverse gut microbiome and an increased abundance of several taxa. The very different dietary profile, and in particular dietary protein intake, was found to correlate with differences in alpha diversity [44]. Further investigation of these samples using shotgun metagenomic sequencing allowed for a greater understanding of the taxonomic changes by delving down to species level. The use of this technology also allowed for the identification of the functional capacity of the athletes in comparison to controls. Athletes were again found to have a more diverse microbiome than controls, this time from a functional perspective. A total of 98 pathways were significantly different between the athletes and controls, identifying the massive

impact of exercise on not only the compositional structure but also the functional capacity of the gut microbiome. Additionally, the athlete group were found to have a more diverse SCFA profile, which has previously been associated with health benefits [45]. Although differences are identified between athletes and controls, due to the major diet differences between the groups it is difficult to determine if a combination of the two are resulting in this alteration. Further investigations need to be completed in order to further understand how exercise alone impacts on the gut microbiome.

Investigations of the impact of exercise on the gut microbiome have extended beyond rugby into other sports, such as cycling. An investigation of the gut microbiome of competitive cyclists demonstrated some similarities with the above study such as the low abundance of *Bacteroides* in these athletes. Additionally, this study established the impact of a variation in hours spent training; with an increased training load (>11 hours/week) associated with a relatively high abundance of *Prevotella*. Although in depth nutritional analysis was not completed, this association could potentially be the result of increased calorie intake in response to an increased training schedule [46].

Important consideration must be given to the translation of these observations in athletes to the general public as a study of the impact of exercise and activity load on the gut microbiome established a correlation between sedentary parameters (such as sedentary time) with microbial diversity, while the opposite was seen above in athlete groups. Some similarities could be observed between active women (exercising for greater than 3 hours per week) in this study and athletes in the above studies such as the increased abundance of *Akkermansia muciniphila* in these individuals in comparison to controls and sedentary women (less than 30 minutes exercise 3 days per week) [47].

1.2.3.4 Diet

Both first food, as breast milk or formula, and foods introduced during weaning, as previously mentioned, have an impact on the gut microbiome. The impacts of both long- and short-term diets on the gut microbiome continue throughout an adult's life with each of these potentially impacting on host health.

Historically, interest in the composition of the human colonic community became particularly intense during the 1970s, largely in response to observations suggesting the substantially different diets of people living in Africa and England could be linked to the higher prevalence of diseases of the large intestine, including appendicitis, constipation, diverticulosis, and colon cancer in individuals ingesting a Westernized diet. It was suggested that a deficiency in dietary fibre and an increased intake of refined foods, red meats, and fat were major contributors to these diseases [48]. The first successful attempt to assess change within the collective faecal community in response to diet occurred in the early 1990s when Peltonen and colleagues followed change in response to diet in both rodents and in humans based on changes in the profiles of extracted total faecal bacterial fatty acids [49, 50].

Much of the interest in diet-mediated change has been driven by the rapidly expanding market for functional foods with purported gut health benefits and the need to both substantiate health claims and regulate these products. While it has become generally accepted that diet can alter gut communities, whether or not changes in species abundance in response to diet holds substantive value in terms of specific health benefits in the host has yet to be determined.

Much like exercise, diet is a lifestyle factor that is relatively easily modified, and it may potentially be used as a route for therapeutic interventions. Although the original concept of the presence of three enterotypes (dominated by *Bacteroides*, *Prevotella*, or

Ruminococcus) has since been found to not fit all data sets, the inclusion of more data suggests there may be some evidence to support the existence of two such groups, dominated by either *Prevotella* or *Bacteroides*. In one study long-term, but not short-term, food intake is associated with the enterotype observed. The *Bacteroides* enterotype was associated with animal protein, a variety of amino acids and saturated fats, and the *Prevotella* enterotype was associated with a high carbohydrate and simple sugar intake. Inter-individual variation was the largest source of separation in investigations of short-term diet. That said, it should be noted that changes can be observed in faecal populations within as little as 24 hours following the initiation of feeding trials, indicating a rapid change in the gut microbiome, without a shift in enterotype [23].

It has been established that the gut microbiota may be a contributing factor in diet related obesity in both animal models and in humans. Germfree mice are resistant to high sugar and high fat western diet induced obesity while the colonization of these mice with faecal material or microbiota from obese humans or mice results in an increase in fat deposits. Dietary intervention, in children with genetic or simple obesity, with a diet rich in non-digestible carbohydrates induces a shift in the gut microbiota to a healthier structure, with a lower level of bacteria capable of potentially producing toxic metabolites from the fermentation of dietary proteins and fats, and higher levels of bacteria which can produce potentially beneficial products from the fermentation of carbohydrates. This intervention found that almost all relevant biomarkers indicated alleviation of metabolic deteriorations in children with both genetic and simple obesity after 30 days on the diet [51].

Diet does not only have impacts on the microbial composition, but also has metabolic effects. In recent years, NMR based metabolomics has been used as a potential tool to identify dietary biomarkers and also explore the metabolic effects of dietary interventions. A study of the impact of an energy restricted intervention with low- or high-dairy

consumption on obesity was completed using metabolic analysis of faecal, urine, and bloods samples. A significant effect of dairy intake was established on the urine metabolome but not the blood equivalent. However, energy restriction had a significant impact on the blood metabolome. Dairy intake affected microbial related metabolites in the faecal metabolome including the SCFAs acetate, butyrate and propionate, and malonate [52].

In a study of elderly individuals an increased consumption of animal products and a reduced intake of plant-based content were associated with microbial populations, which was in turn associated with frailty in these individuals. The consumption of a high proportion of plant-based foods was correlated with a diverse microbiome, which in turn is associated with healthy aging. Overall diet was established as being a better indicator of health associated microbiota than individual dietary components [25]. A dietary intervention, as opposed to the use of a specific dietary component as a supplement, might provide more information about how much of the microbiome is modulated by diet and to what degree.

Interestingly diet changes seem to long precede microbiome changes. In a study of elderly individual's long-stay residential care patients and community dwelling individuals were compared. Although individuals who were newly admitted to long-stay facility changed to a long-stay diet after 1 month the microbiome took 1 year to be clearly clustered with other long-stay patients. The diets between the two locations was found to be significantly different and was the driver for reduced bacterial diversity identified in the long-stay group [35].

The applications of these results could potentially allow for the development of a simple and easy to implement therapy using diet to modulate the gut microbiome in diseases

where the gut microbiome is implicated. A greater understanding needs to be established for the potential long-term impacts of this type of therapy.

1.3 Role of the gut microbes on health

The human gut microbiome plays a vitally important role in the health of the host from the establishment of a healthy immune system in infancy, to host metabolism, development, and behaviour. The gut microbiome may be implicated in the development of different diseases and thus overall host health [27]. The gut microbiota encodes for proteins which are involved in important functions for host health such as production of enzymes required for the hydrolysis of otherwise indigestible dietary compounds and the synthesis of vitamins [24].

1.3.1 Role of the normal gut microbiota in immune development

The commensal microbiota are intrinsically involved with the development of the gut mucosal immune system and necessary in preventing pathogen colonization through direct interactions between commensal and pathogenic microorganisms, and also by stimulating the immune system from birth [53]. The immune system plays an important role in keeping the body healthy by providing balance between the elimination of invading pathogens and maintaining tolerance to healthy self-tissue. The development of autoimmune disorders occurs when the mechanism to maintain self-tolerance fails and this results in the immune system mistakenly attacking and destroying healthy self-tissue [9].

The immune system operates through a complex interplay between innate and adaptive responses. The innate or nonspecific response reflects the actions of cell types and molecules that react to the presence of invading microorganisms and their components without a high degree of specificity. Key features of the innate response are speed and a lack of immunological memory. The rapidity of the response reflects the ability of macrophages, dendritic cells (DCs), granulocytes, and even intestinal epithelial cells (IECs)

to react rapidly to stimuli, a property essential for immediate host protection. By contrast, the adaptive or specific immune response, while less rapid, possesses immunological memory, and a high degree of specificity. Memory improves the efficiency and speed of the adaptive response, and reflects clonal expansion and activity of specific B and T lymphocytes. Bridging between the innate and adaptive responses are the antigen-presenting cells (APCs), including DCs and macrophages.

Cell types responsible for innate immune responses have the ability to recognize certain molecular patterns on microorganisms, whether these are pathogens or components of the normal gut microbiota. These microbial-associated molecular patterns (MAMPs) are recognized by pattern recognition receptors (PRRs) expressed by DCs, macrophages, and IECs. PRRs are highly conserved and include toll-like receptors (TLRs), nucleotide-binding oligomerization domain-like receptors (NLRs), RIG-I-like receptors (RLRs), and C-type lectin receptors. TLRs recognize components of bacteria, fungi, protozoa, and viruses, with some detecting extracellular MAMPs and some detecting intracellular MAMPs. TLR5 is expressed at the IEC basolateral surface, and serves as a sensor for bacterial flagellin. An indication of the importance of the flagellin-TLR5 interaction is illustrated by the development of spontaneous colitis in TLR5 knockout mice, and the cytoprotective effects of flagellin administration. TLR2 and TLR4 are expressed on the surface of APCs; TLR4 interacts with co-receptors to recognize lipopolysaccharides (LPS), and TLR2 forms heterodimers with other TLR receptors to recognize several microbial components including lipoteichoic acid and yeast zymosan. TLR9 recognizes bacterial DNA containing unmethylated CpG motifs, and is expressed only intracellularly. NLRs detect mainly intracellular bacterial components, while RLRs detect viral components. Together, this array of PRRs provides the innate immune system with a highly competent ability to detect key microbial components and to rapidly respond to their presence. Pathways activated in response to PRRs are also

conserved, and mainly operate through nuclear factor KB (NF-KB), mitogen-activated protein kinase (MAPK), and caspase-dependent routes. Interactions between these PRR families serve to coordinate innate immune responses. The outcomes of PRR-mediated MAMP recognition are frequently inflammatory; however, the degree of the pro-inflammatory signal shows considerable variation, especially between responses to pathogens relative to the normal microbiota. A further degree of complexity in PRR-mediated signalling is illustrated by recent evidence associating the normal gut microbiota with TLR-mediated signalling through NF-KB, leading to protective anti-inflammatory effects [54, 55]. Mice deficient in TLR expression or in MYD88 (a key intermediate in TLR signalling) show colonic inflammation, reflecting the loss of TLR-signalling induced by the gut microbiota [56]. Overall, the current picture of host–microbiota interactions at the gut interface suggests that the microbiota plays an essential role in maintaining immune homeostasis. If these complex interactions fail, autoimmune or autoinflammatory responses can occur within the body to otherwise harmless compounds [53].

The hygiene hypothesis, which was established in 1989, suggested that the absence of exposure to microorganisms in childhood can result in a perturbation of the gut microbiota composition and thus impact on immune responses later in life with the development of atopic diseases in susceptible individuals [53]. Germfree animal models have illustrated the importance of the normal gut microbiota in development of the immune system. Numerous immune abnormalities are seen in the absence of gut microbiota, including greatly decreased serum immunoglobulin levels and reduced densities of lymphoid cells in the Gut-associated lymphoid tissue (GALT). The structure of lymphoid follicles (i.e., locations where cells of the immune system interact to generate responses to challenges) is also abnormal in germfree animals, with greatly reduced or absent germinal centres, reflecting reduced lymphocyte numbers. Germinal centres provide a microenvironment for optimal interactions between B and T lymphocytes, and the antigen-presenting DC. When a

normal gut microbiota is established in germfree animals, immunoglobulin levels increase, due to increased numbers of GALT lymphocytes and germinal centre size. These dramatic changes reveal an essential role for gut microbes in the development of host defences [57, 58].

Members of the normal gut microbiota have dynamic effects on the immune system. *Bacteroides fragilis* administration to germfree animals can restore normal population balances of T-helper (TH) cells, reflecting effects on GALT regulatory signals mediated through *B. fragilis* polysaccharide A (PSA). *B. fragilis* PSA recognized by DC directs them to promote development of regulatory T cells (Tregs), and so down regulates gut mucosa inflammatory responses. Mono-colonization of germfree mice with *B. thetaiotaomicron* has also been reported to restore aspects of intestinal and GALT structure and function. Sampling of microbiota and MAMPs by M cells in Peyer's patches and interdigitating DCs at the gut mucosal surface is believed to promote differentiation of mucosal CD4+ TH cells to Tregs, which produce the immunoregulatory cytokines, interleukin-10 (IL-10) and transforming growth factor β (TGFβ). The microbiota is also involved in suppressing differentiation of the pro-inflammatory Th17 CD4+ Th cell subset, through up regulation of IL-25 and down regulation of IL-23. These immunomodulatory effects of the microbiota on DCs may influence both adaptive and innate host responses, having far-reaching effects on the immune system. Overall, a key role for the microbiota is in promoting DC conditioning to a regulatory phenotype, stimulating Treg differentiation, and promoting production of IgA that serves both a barrier and an anti-inflammatory role [59].

Infant type *Bifidobacterium* play an important role in the maturation of a child's immune system. Low microbial diversity in early life has been associated with some health disorders dependent on the age of reduced diversity, with eczema and asthma associated with low microbial diversity in very early life (1 week to 4 months) and type one diabetes associated with low microbial diversity at 18-24 months of age [34]. It has been suggested that a

decrease in community diversity can limit the education of the immune system, including its ability to recognize commensal bacteria which can result in a less robust community [30]. This may potentially result in the development of diseases associated with the gut microbiome in later life.

1.3.2 Metabolic potential of the gut microbes and its impact on the host

Metabolic products of microbiota are strongly implicated in mediating varied effects on the host, including effects on the immune system [60]. The metabolic products produced by these microbes impact on the host, which can be detected in the faeces, urine, and blood [61].

Many of the metabolic products of the gut microbiota are produced through fermentation of undigested food and host derived products, while fermentation of products from other microorganisms also occurs. Metabolic potential of the gut microbiome is mainly driven by diet, as substrate availability influences microbiome composition, which in turn impacts on metabolites produced [61]. Variation in dairy intake with energy restriction can alter host and microbial metabolite production. High dairy intake with energy restriction results in increased urinary excretion of citrate, creatine, and urea and a reduction in secretion in hippurate and TMAO in comparison with a low dairy diet with energy restriction. Additionally, high dairy intake impacts on the faecal metabolome including SCFAs (acetate, butyrate, and propionate) and malonate [52]. Microbial community is also responsible for metabolic potential, as metabolites produced by one microbe may become a substrate for another microbe [61].

SCFAs are the major metabolic products of microbes of the GIT from carbohydrates. They provide energy for host tissues but also exert anti-inflammatory and anti-apoptotic effects. They have implications for both host health and gut environment homeostasis. Through interactions with host receptors, SCFAs are capable of regulating hormones that affect

satiety [61]. Furthermore, acetate is important for muscle function, heart function, and brain cells, propionate is used in host hepatic neoglucogenic process, and butyrate acts as the main source of energy for enterocytes [62].

Bacteroides are capable of metabolising a wide variety of substrates in the GIT of both human and dietary origin. Being major polysaccharide digesters, *Bacteroides* species have organized gene clusters of polysaccharide utilisation including those involved with hydrolysis, and outer membrane proteins for initial polysaccharide binding and transfer. Firmicutes are the dominant producers of butyrate in the colon and also help maintain gut homeostasis through the maintenance of pH by conversion of lactate to butyrate or propionate. Firmicutes are also involved in polysaccharide degradation, however, with a more specialized approach and so generally carry fewer associated genes [61].

The metabolic potential of the gut microbiome is affected by several disease states, including IBD. These alterations in functional capacity might be more important than taxonomic changes in understanding diseases with gut microbiome associations. For instance up to 12% of metabolic pathways are changed in patients with active IBD in comparison with 2% of genera changes [6]. A reduction in production of butyrate, acetate, and other SCFAs is observed with CD. This is not only a reflection of microbial shifts during the disease state, but these alterations may also be implicated in disease progression. SCFAs affect host immune balance by regulating the number and function of colonic T regulatory cells. The reduction in SCFA production found in CD would therefore lead to an enhancement of the pro-inflammatory state [6]. Treatment of ulcerative colitis (UC) with SCFAs, such as butyrate, can be clinically beneficial as they act as a major energy source for the epithelium, promote T regulatory cell differentiation and reduce inflammation through G-protein-coupled receptor 43 signalling [63]. A variety of other bioactives are generated in the gut from microbial metabolism of proteins, dietary phytochemicals, and drugs [61]. A

number of vitamins (including biotin and phylloquinone) and essential amino acids can also be produced by bacteria [62].

1.3.3 Role of the gut microbiota in disease

Microbial richness and diversity is usually considered as an indicator of the healthy adult status and is reduced in many diseases from GI associated diseases such as CD, UC, IBS, and colorectal cancer (CRC), to metabolic disorders including obesity and T2D, as well as neurological disorders such as autism [24, 34]. The gut microbiome has been found to be altered in many disease states however; it must be stressed that some alterations to the gut microbiota may be a consequence rather than a cause of disorders.

For many years the study of gut microbiome associations with human diseases worked off the basis of the identification of one microorganism related to the disease. However, with increasing investigation and understanding of the human microbiome it is more likely that many microorganisms, or the interactions between microorganisms, are responsible for diseases as opposed to one specific microbe. This understanding might lead to a better understanding of diseases and more importantly the mechanism of treatment of these diseases [1].

1.3.3.1 Gut-brain axis

Accumulating evidence suggests the microbiome plays a role in brain and behaviour. Alterations in the gut microbiota have been implicated in stress related disorders including depression and anxiety, in neurodevelopmental disorders including autism, and in cognitive functioning. Neurogenesis is the birth of new neurons, which plays a role in learning and memory responses to stress and anti-depressants. In a mouse model, an increase in cell proliferation across the total subgranular zone in germfree mice and in germfree mice that were colonized post-weaning was identified. The survival of newly born cells and neurons was also significantly increased in germfree mice whilst also not being prevented by post-

weaning microbial colonisation. Although the underlying mechanisms and implications of this remain unclear, it demonstrates the presence of an early window of opportunity in the life of the mouse where the introduction of microbes may influence neurogenesis [64].

The gut-brain axis is a newly established axis which is of vital importance to many aspects of a human's life. As this is a bidirectional axis which allows the gut and brain to remain connected, this also includes the microbes residing in the gut. The interaction between the microbes and the brain has been implicated in some psychological disorders including depression and anxiety with individuals with these disorders having a unique microbial population. The understanding of these alterations and the implications of these for the gut-brain axis could allow for therapeutic development for the treatment of such disorders.

Serotonin functions as a key neurotransmitter at both terminals of the gut-brain axis. Manipulating the composition of the gut microbiota across the lifespan or altering microbial colonization of the GIT in early life influences the availability of tryptophan. Tryptophan is an essential amino acid available in protein and in breast milk. It is absorbed in the gut and in free circulation in the body. It can cross the blood brain barrier where it is available for serotonin synthesis in the central nervous system. The vast majority of serotonin is in the gut where it is synthesized from tryptophan in enterochromaffin cells of the GIT which are also present in enteric nerves. IBS is a functional GI disorder which is regarded as a gut-brain axis disorder with alterations in stability and diversity of the microbiome frequently noted as being involved along with alterations in the serotonergic system in both the brain and gut. Certain bacterial strains harbour tryptophanase enzyme which produces indole from tryptophan. *B. fragilis* for example has this enzymatic capability which has recently been linked with GI abnormalities in autism spectrum disorders. Initial life exposures to microbes are important in the development of the gut-brain axis and may be important in the understanding of these interactions. Early life

stressors have been found to be correlated with an increase in anxiety and depressive like behaviours as well as GI disorders with a stress component. In animal models of depression or where anxiety has been induced in early life, there are implications for changes to the gut microbiota [65].

1.3.3.2 IBD

IBD comprises a group of chronic relapsing inflammatory disorders involving the GIT which arise from a combination of genetic, immunological, microbial and environmental factors. There are two main clinical presentations; UC and CD. UC involves continuous, superficial inflammation involving the rectum and extending up to the cecum, while CD involves deep ulcerations distributed in a skipped pattern anywhere from the mouth to the anus. CD is associated with manifestations such as arthritis, uveitis, and perianal disease. UC is associated with primary sclerosing cholangitis, and ankylosing spondylitis.

Genetic predisposition can sometimes be involved however in 77% of CD and >80% UC patients no genetic associations can be identified. A theory has been established that the increase in western disorders, such as that of IBD, is likely as a result changes in gut microbiome colonization attributed to increased rates of caesarean section, sanitation, dietary changes, and early and broad use of antibiotics. There are over 160 single nucleotide polymorphisms associated with IBD. Many of these genes are involved in pathways which modulate how the host responds to microbial stimuli potentially implicating the microbiota in this disease.

Large changes in the gut microbiota in patients with IBD have been established. A decrease in diversity and richness is associated with UC with alterations observed to precede the disease development. Patients with IBD have an altered rate of Bacteroidetes, Proteobacteria, and Firmicutes. Other components of the microbiome are also implicated in disease development including bacteriophage, fungi, and archaea. A significantly greater

number of bacteriophages are detected in the non-ulcerated mucosa in comparison to healthy controls. Additionally, changes in the fungal components have been identified as being different between the colonic mucosa, inflamed mucosa, and faecal samples. Differences have also been observed where inflammation is active with diversity of the fungal community correlating with mucosal inflammation. The potential application of nutritional therapy might result in a reduction of inflammation and the return of a gut microbiome comparable to controls as was identified in one study [6, 24].

Breakdown of immunological tolerance to the gut microbiota is a recognized factor in IBD, resulting in abnormal responses to the normal gut microbiota. For example, antiflagellin antibody is present in the serum of IBD patients, suggesting dysregulation of the host response to normal gut microbiota. Individuals with CD often have a mutation in the microbial sensor molecule NOD2, resulting in the loss of this receptor's ability to detect the muramyl dipeptide component of peptidoglycan. Peptidoglycan is an integral bacterial cell wall component, and this outcome of NOD2 mutation implicates loss of the ability to respond normally to the local intestinal microbiota in the overall pathogenesis of CD. Increased CD susceptibility has also been linked to decreased production of human β-defensin 2 (HBD2/DEB2), and antibiotic therapy can help maintain remission of active CD [66].

1.3.3.3 Allergies

The modern hygiene hypothesis implicates the increase in hygiene practices including sanitisation with the prevalence of allergies [67]. Functionally, allergy results from the inappropriate T-helper type 2 (Th2) immune response to generally innocuous proteins. Maturation of Th2 responses that predominate at birth to Th1 predominance in infancy and adulthood requires the presence of the commensal gut microbiome. Gut microbiome composition was found to be capable of predicting allergies. Infants with a higher

abundance of *C. difficile* and *E. coli* have an increased risk of the development of allergies in the future. Low diversity at one month predicted atopic eczema at 2 years while fewer faecal bacterial taxa predicted allergic rhinitis by 6 years. Richness was found to be strongly negatively correlated with all allergies except bee sting, asthma, and eczema. Strongest non-food correlations were with drugs and seasonal allergies while the strongest food correlation was with peanut allergies. There were also specific relationships observed between the relative abundance of certain taxa and allergies [67]. With a greater understanding of allergy development as a consequence of early life gut microbiome development it may be possible to modulate the gut in early life to potentially prevent allergy development. In adults, the main composition difference in the GIT in individuals with allergies is the reduced abundance of *Lactobacillus* and *Bifidobacterium* [53]. Potentially the modulation of the gut microbiome could prevent or control allergy development.

1.3.3.4 Cancer

The gut microbiome can potentially be implicated in not only the presence of GIT cancers, such as colorectal cancer, but also cancers outside the GIT. Alterations of the gut microbiome in mouse models influence the incidence and progression of extra-intestinal cancers including breast and hepatocellular carcinoma presumably through inflammatory and metabolic products. Altering the composition of the gut microbiome influences the incidence and progression of CRC in genetic and carcinogen induced models. Several by-products of the gut microbiota directly target the intestinal epithelial cells and either mediate oncogenic effects, as reported for hydrogen sulphide and *B. fragilis*, or suppress tumorigenesis, as demonstrated for SCFAs. Of equal importance are the interactions between the gut microbiome and drugs used for cancer treatment. The gut microbiome can determine bioavailability and biological effects of drugs, not just those taken orally, which may result in increased toxicity or efficacy of drugs used [8].

1.3.4 Perturbations to the intestinal microbiota and its relation to disease

The increased susceptibility of germfree mice to a variety of enteric pathogens led to the formulation and promulgation of the 'colonization resistance hypothesis' – that the normal gut microbiota directly or indirectly attenuates or prevents disease [68, 69]. Competition for attachment sites, stimulation of increased secretion of mucins by the host, production of bacteriostatic agents such as SCFAs and bacteriocins by the gut microbiota, and enhanced immune activation have all been suggested as possible mechanisms [56]. In addition, cell-to-cell communication (i.e., quorum sensing) has been implicated to be important for colonization resistance. Even in the presence of a 'normal' microbiota, some enteric pathogens retain the ability to colonize the intestine and cause disease. One hypothesis suggests that pathogens incite inflammation to allow infection [70, 71]. A proposed mechanism detailing how pathogens may benefit from inflammation is known as the 'differential killing hypothesis' [68] and predicts that inflammation is accompanied by an increase in the release of antimicrobial substances that target specific gut bacteria responsible for colonization resistance, while the pathogen has evolved strategies for resisting such harmful substances.

Perhaps the best characterized factor shown to cause a perturbation of the microbiota is antibiotic administration. As previously discussed, antibiotic administration can impact on the gut microbiome structure and function. Additionally, antibiotics can result in a range of both short- and long-term side effects, both within and beyond the GIT. Diarrhoea is a common side effect associated with the use of antibiotics with *C. difficile* being implicated with most severe cases of AAD. Although *C. difficile* colonizes the minority of healthy individuals, it is suggested that patients who develop *C. difficile* associated AAD (CAAD) following the administration of antibiotics either allows colonization by *C. difficile* after the ingestion of environmental spores or permits the overgrowth of indigenous *C. difficile*. Standard treatment of CAAD involves the administration of additional antibiotics to

suppress *C. difficile* but infection can reoccur once these are stopped. The treatment of reoccurring CAAD can be difficult and might involve faecal microbiota transplant (FMT). With recurrent CAAD the composition of the microbiome is more variable and is deviated from the normal predominance of Bacteroidetes and Firmicutes. Additionally, patients with reoccurring CAAD but not CAAD have a significantly reduced diversity and richness. In one study an individual with the lowest diversity and richness in the CAAD group developed recurrent-CAAD within 10 days [72].

Antibiotic administration also increases the susceptibility of colonization by other pathogens such as *Salmonella typhimurium* which can result in symptoms such as diarrhoea, fever, and abdominal cramps. The outcome of colonization by this pathogen seems to be dependent on the type of antibiotic therapy the individual is recovering from, with hospitalization observed in some instances [40].

1.4 Complementation of the gut microbiota

As has been established, the gut microbiota is of vital importance to host health and many factors might influence its populations. In order to maintain health, there is merit to assisting the recovery of the gut microbiota post-perturbation in order to prevent disease development.

Although the number and diversity of the bacterial species within an individual's GIT remains relatively stable throughout life, it is possible to stimulate proliferation of specific microorganisms known to have beneficial effects by manipulating the host diet. Prebiotics are non-digestible, plant derived carbohydrates which act as fermentation substrates within the colon stimulating preferential growth and activity of a limited number of microbial species that confer health benefits on the host. Carbohydrates which have been established as having a prebiotic effect include inulin type fructans (inulin, oligofructose, and fructo-oligosaccharides) and galactans (galacto-oligosaccharides) known to promote

proliferation of lactic acid producing species such as *Bifidobacteria* and *Lactobacillus*. The use of prebiotics has been found to be useful in the treatment of gut microbiota related syndromes and diseases, and some limited studies suggest their effects beyond the GIT [73, 74]. Prebiotics have been found to be beneficial in controlling IBD as well as alleviating some of the symptoms of IBS, such as diarrhoea. Additionally prebiotic administration can be beneficial in controlling some of the risk factors of cardiovascular disease, such as increased cholesterol [74]. Prebiotics administration may be potentially used as a simple, safe, and effective intervention in the prevention and treatment of obesity and associated disorders. Prebiotic administration can improve satiety measurements, while adherence to a regimen for greater than two weeks results in a reduction in total energy consumption. Additionally, some evidence supports the use of prebiotics in the reduction of body weight in longer trials (12-17 weeks). Prebiotic consumption reduces postprandial glucose and insulin concentrations. However many of the results here were seen to be conflicted in other studies which establishes the need for longitudinal studies of the impact of prebiotic use on all of these factors before this can be used as a therapeutic intervention [73, 75-77].

Probiotics are live microorganisms which when ingested in significant quantities can have beneficial impacts on the host. Probiotics have been found to be beneficial in the repopulation of beneficial bacteria following a perturbation, such as antibiotic use. There are several mechanisms through which probiotics are thought to be beneficial to the intestinal mucosa including preventing the binding and growth of pathogenic bacteria, improvement of epithelial barrier function and modulation of immune activity of the host. Additionally, probiotics secrete SCFAs, reducing luminal pH, and also bactericidal proteins [53]. Lactic acid bacteria are the most frequently used probiotics, mainly those from *Lactobacillus* and *Bifidobacterium* genera [78]. While some probiotic benefits are general, such as supporting a healthy GIT and immune system, others are more specific and have been linked only to a few strains, such as supporting the health of the reproductive tract,

oral cavity, lungs, skin, and gut-brain axis. Through these general and more specific mechanisms probiotics have been found to be beneficial in protecting against and/or reducing the severity of symptoms in infectious diarrhoea, AAD, UC, and IBS [79-82].

FMT is a technology in which intestinal microbes are transferred from a healthy donor to the patient, with the goal being to introduce or restore stable microbial community in the gut. FMT has achieved successful cure rates in recurrent *C. difficile* infections, with a high proportion of these individuals remaining relapse free [83, 84]. There have been some cases where FMT has shown some impressive results with UC, with CD seen as less responsive [85-87]. Some studies suggest a benefit associated with the treatment of IBS with FMT, with relief from symptoms including bloating and abdominal pain found. However, as others found no benefit of FMT in IBS treatment more investigation into this area is necessary. An additional confounding factor in the use of FMT in the treatment of IBS is the presence of only self-reported data and the absence of any clinical biomarkers [88]. There is also the potential of the use of FMT in the treatment of metabolic disorders. Both insulin sensitivity and levels of butyrate producing intestinal microbiota (*Roseburia intestinalis* and *Eubacterium halli*) were increased after a 6 week infusion with the microbiota from a lean donor with implications of utilising this as a potential strategy to increase insulin sensitivity [84, 89].

The use of each of these therapies in manipulating the gut microbiome is associated with its own difficulties, especially for FMT where there is a potential of the transfer of undesirable microorganisms. For this reason, it is vital that studies into the impacts of these on a longitudinal basis are completed to determine their efficacy in treating both intestinal and extra-intestinal disorders.

1.5 Tools to study the microbiota-host interaction

1.5.1 Culture-based techniques

Traditional studies of the gut microbiota were completed using culture-based techniques and this was advanced by the introduction of anaerobic methods which improved such studies. For a long time, it was said that a large proportion of the gut microbiota was unculturable. However, many have challenged this claiming that all microbes are culturable if the correct techniques are applied. There is no consensus regarding what percentage of the gut microbiota has been cultured, with suggestions ranging from 2.6 to 95% [90-92]. Regardless of the true value, the use of culture-dependent techniques is time consuming and labour intensive, due to the wide range of conditions (media types, pH, temperature, etc.) required to culture all members of the community.

When used, cultivation has the advantage that it provides information not only on substrate utilization and end-product formation but also, when combined with whole genome sequencing, offers the potential to better infer the functional relationships of individual species within these very complex communities, and facilitates subsequent elucidation of their functionality within the intestinal ecosystem (e.g., using animal models). A number of current studies aim to standardize culture-dependent techniques. There are several conditions at play which need to be considered in the creation of isolation techniques from human faecal samples such as the anaerobic nature of the GIT, the presence of bile acids, and the possible symbiotic relationship between different microbes within the GIT which might be required for their isolation. Much of the not yet cultured taxa are strictly anaerobic and so methods need to be carefully developed for their isolation [91, 92].

Culture-independent techniques, such as next generation sequencing (NGS), allow for the identification of microorganisms and microbial communities without the need for initial

growth of the microorganisms in culture. This allows for the identification of community composition in the absence of appropriate methodologies to culture all members of the community. With the combination of both culture-dependent and -independent technologies a better understanding can be established for the growth conditions of not yet cultured taxa which might allow for their isolation. Some studies of this nature have emerged with the combination of many different agars plus high-throughput identification of isolates with mass spectrometry allowing for the identification of novel taxa [91].

1.5.2 Considerations with faecal sample analysis

Much of our current knowledge concerning membership and community structure of the microbiota has been based on the analysis of faeces. However, caution must be exercised when drawing conclusions concerning events that occur within the gut based strictly on the analysis of faecal samples [93]. Although mucosal biopsy samples might be the most appropriate in studying the gut microbiome, faecal samples are used as they are easy to collect and non-invasive. However, due to the variation of community structure throughout the different areas of the GIT the use of faeces may be inappropriate and may lead to inaccurate conclusions. Another problem associated with culture-dependent and -independent studies of the gut microbiota using faecal samples is the inconsistent and variable methodologies used for collection, storage, and DNA extraction (culture-independent studies) across different groups and studies.

Although fresh faecal samples are desirable, their use is not always possible, such as in large cohort studies where the processing of a large number of samples may not be feasible. Additionally, participants may not always be located within close proximity to the study centre and in cases where longitudinal samples need to be collected continuous visits to the study centre may not be possible. For these reasons it may often be necessary to

store samples prior to extraction, which may have an impact on the faecal microbial populations.

Several studies have identified the impact of different storage conditions on faecal microbiota populations observed using both culture-dependent and -independent methods. These studies identified the most important consideration when choosing the methods of collection and storage involves consistency within a study. The use of freezing is a favourable method to preserve faecal microbiota as long as consideration is given to avoid freeze-thaw cycles and ensure the faecal sample is frozen as soon as possible. Cold temperature storage (i.e. 4°C) does not have an impact, if not used for long-term storage, while room temperature storage is not appropriate [94]. The use of collection buffers such as OMNIgene-Gut collection kits (DNA Genotek Inc., Ottawa, Canada) may also be useful especially in studies where participants are located away from a study centre as it reduces the need for transportation on dry ice. However, the use of these buffers may alter the composition of the faecal microbiota over time and so consideration should be taken to process samples promptly [95]. Care should also be taken when using these buffers to use as intended to prevent any deviations from original fresh samples [94]. The use of these methods may allow for the processing of a greater number of samples and the collection of samples from individuals who are unable to visit the study centre. An important consideration is the use of these methodologies in the collection of samples which are to be used for other analysis such as metabolomics as it is not known the impact such storage processes may have on these types of studies. Consideration should also be given to the types of samples used for gut microbiome analysis as low diversity samples, such as samples collected from infants, are more susceptible to changes as a result of storage in a stabilisation buffer [95].

Following the collection of samples, DNA is normally extracted, and the methodologies used here is another important consideration in the study of the gut microbiota. Extraction methods need to be validated, for all gut communities, to ensure an accurate representation of the GIT is obtained [96]. The use of an inappropriate methodology for the extraction of DNA from the faecal matrix may result in problems downstream or inappropriately represent the complex community present. The extraction of DNA from faecal samples may be difficult as other substances may be co-extracted from the sample which could have an impact downstream. Additionally, the DNA from some microbes may be difficult to extract due to incomplete lysis of the bacterial cell. Adjusting the method to accommodate for such microbes may result in DNA shearing which may also implicate downstream applications. Several recent studies have compared the different methodologies available (commercial and non-commercial; mechanical and enzymatic lysis). Similarly, to the study of the different storage methods such studies identify the importance of consistency within a study. The use of mechanical based lysis (e.g. bead beating) was found to be most preferable [96, 97].

1.5.3 Sequencing technologies

With the decreasing cost and increasing speed of high-throughput sequencing approaches these are currently applied much more frequently and much of our knowledge of the gut microbiome comes from such studies. The use of culture-independent studies have revolutionized the study of the gut microbiome allowing for the understanding of microbial populations of the GIT over a lifespan and between populations, studies which would not previously have been feasible using culture-dependent methods [22, 98].

Technological advancements led to the move from original sequencing platforms to NGS. This movement allowed for the reduction in the cost of sequencing while improving read lengths, speed, and throughput. Third generation sequencing platforms have now also

emerged which focus on the use of few biochemicals and the removal of PCR cycles allowing for sequencing from a single DNA strand [99].

There are two culture-independent approaches widely used in studies of the gut microbiome; compositional and shotgun metagenomic sequencing. Compositional sequencing involves target gene amplification, using targets such as 16S or 18S rRNA. These genes are often used as a target for compositional studies due to the nature of ribosomal RNA which is highly conserved and evolutionary stable with differences in the hypervariable regions. 16S rRNA gene sequencing studies, used to study the bacterial component of the microbiome, are the most common. An increasing interest and knowledge in the use of this gene led to an improvement in the associated databases, which has further advanced studies [98]. Compositional sequencing is a targeted, easy, cost effective method used for compositional analysis of faecal samples and other environments, and is particularly useful in large scale studies [24, 100].

While compositional sequencing allows us discover who is present in the community (to genus level), with a shotgun metagenomic sequencing approach we can discover their functional capabilities [101] and get the most accurate estimation of species abundance [62, 98]. This approach has improved studies of the gut microbiome as it allows for the study of all kingdoms of life present in the gut together without the need for specific primer development. Shotgun metagenomic sequencing is based on the extraction of all genomic DNA from a sample and using this to prepare next generation sequencing libraries for downstream high-throughput sequencing. Subsequent data analysis with bioinformatics tools is completed in order to assign reads to host and microbial components and complete genome assemblies. This technique does not require culture or PCR and allows for identification down to species level and virome analysis [24]. Functional profiling and the possibilities for strain level analysis using shotgun metagenomic data make it very useful in

the greater understanding of the gut microbiome. Strain level analysis can allow for a greater understanding of microbial transfer, for instance from mothers to their infants [29]. Profiling the functional capacity of the gut microbiome can reveal patterns in health and disease which cannot be established using compositional profiling alone [100]. With genome assemblies from shotgun metagenomic data the evolution of certain genes and gene families, such as antibiotic resistance genes, can be established [101].

As with the impact of different collection and extraction methods, there are several aspects of a sequencing workflow which can impact on quality and validity of results obtained. Considerations vary depending on whether compositional sequencing or shotgun metagenomic sequencing are used but include primer choice (compositional sequencing only), number of PCR cycles used, platform used (each platform has its own associated error rates and nature of errors), and initial concentration of DNA used [98]. With shotgun metagenomic sequencing assembly might be flawed with less abundant or closely related members, and function assignments could be ambiguous. An incredibly important consideration in culture-independent studies is the use of up to date databases which are appropriate for the study design. During study design phase care should be taken to ensure consistency within the study and between studies where comparisons will be completed [24]. Although shotgun metagenomic sequencing is a very useful tool in establishing community structure in the GIT and elucidating functional capabilities, metatranscriptomics, metaproteomics, and metabolomics complement the understanding of the gut microbiome and are necessary to determine actual functionality and host interactions [22, 101].

1.5.4 Metabolomics

The faecal microbiome only partially represents the gut microbiome and so techniques, such as metabolomics, may be used to investigate possible patterns and behaviour of

metabolites to better understand the gut microbiome. Metabolomics is the comprehensive analysis of all of the metabolites in a biological system with their identification and quantification. There are several different matrices which can be used for metabolomics analysis: blood, urine, faecal, and saliva; each with their own associated challenges. The use of the different matrices singularly or in conjunction with each other is dependent on an interest on metabolites from the host, microbes or their co-metabolism. Metabolomics on faecal samples for the study of the gut microbial metabolism is a promising field as stool is easily accessible and non-invasive to collect with metabolites originating from the host, the gut microbiota, and food components [10, 102].

Individual metabolomics profiles may be useful in the diagnosis of disease states, as metabolic profiles have high resolution power allowing for the observation of distinct separation between healthy and diseased groups. Compositional changes in IBD and IBS result in distinct metabolic profiles, allowing for the discrimination between patient groups. Faecal samples from IBD sufferers contain high levels of alanine, isoleucine, leucine, lysine, valine, phenylalanine, and butyrate [10, 102]. It is possible that the metabolite changes observed with IBD are as a consequence of inflammation and gut microbiome changes as opposed to a direct cause of the disease. The increase in amino acids in the faecal samples of IBD patients could be a side effect of malnutrition as the inflamed intestinal tissue has low absorption potential for nutrients [10]. Different metabolic profiles are also found in obesity, in both animal models and humans. In animal models some indications suggest the metabolome can be passed on to young, impacting on fat deposition in early life [102].

As with other gut microbiome related techniques, the sample collection, sample preparation, and tools used for metabolomics are very important and need to be considered before project commencement. Metabolomics can be either targeted , which involves the analysis of specific classes (e.g. amino acids, fatty acids, lipids, carbohydrates,

or bile acids) or non-targeted, which involves getting an overview of the metabolic profile of a sample with technology which allows for the description of a large number of metabolites. In order to correctly identify the widest range of metabolites, a combination of techniques should be applied [102].

1.5.5 Transcriptomics

Metatranscriptomics can be a very useful tool in the study of the gut microbiome. A core functionality established carbohydrate metabolism, energy production, and synthesis of cellular components as the main functional roles of the gut microbiome. Understanding this core functionality can lead to a better understanding of mechanisms which potentially promote health, and also how alterations from this core in disease might implicate on the host [101]. Gaining insight into metabolically active microbes and comparing activity between individuals or between regions of the GIT provides information surrounding influencing factors, such as diet, on activity [29]. Metatranscriptomics can be used to observe the impact of dietary alterations and drug administration on activity within the gastrointestinal tract, which might lead to an improved appreciation of the implications of such therapies. Only through investigation of functional activity can the potential implications of changes to the gut microbiome on the human host be hypothesized [103].

Metatranscriptomics serves to analyse the entire transcriptome of an environment to obtain a comprehensive view of gene expression profiles and functional data. The disadvantages of this technology are the technical issues associated with the extraction and storage of RNA, low RNA quality, and quantity from extractions, appropriate enrichment procedure required to remove rRNAs, host contamination from extraction, and the tools available still being under development [24].

1.5.6 Animal models

An ideal animal model to study the host–microbiota interaction should closely mimic variables identified in humans, namely physiology and metabolism, susceptibility to metabolic disorders and pathogens, pathophysiology, and cell and tissue tropisms of introduced products. While not phylogenetically close to humans, small rodents have historically been the primary animal model used as a result of their manageable size, resistance and adaptation to captivity, and ease of fecund reproduction. Rodent models also possess a reasonably comparable physiology, neuroanatomy, and immune system, and, in many regards, share many genetic and histological characters of various tissues (e.g., intestinal tissue). Perhaps one of the most important advantages of using rodent models is the ability to limit the confounding effects of variability in the intestinal microbial community structure. Another salient advantage of rodent models is their amenability to being reared under germfree (i.e., axenic) conditions. Germfree animal models are useful especially in the study of the understanding of how the lack of microbes might have a role to play in certain disorders. It also allows for the addition of single microbes in isolation to view the impact. However, this might not necessarily be the best manner in which to test microbes, due to the probable microbial interplay in the gut. Germfree models can be used to reveal the importance of the microbiota in shaping the innate and adaptive immune system [9]. Rodent models possess the added advantage of well-defined genotypes that are commercially available (e.g., 'knock-out' and 'knock-in' genotypes) and it is now relatively easy to manipulate a gene in a mouse that has a homologous role in humans. Importantly, the use of transgenic methods is also proving to be very effective in addressing various aspects of the metabolism of biologically active compounds and for studying the pathophysiology of variety of diseases, including enteric diseases such as IBD [104].

The genetic homogeneity of inbred rodent models is also advantageous since it potentially reduces the variability of biological responses produced through its relatively uniform genome. This can, however, potentially underestimate the broad ranges of responses identified in people, a consequence of humanity's diverse genetic variability. Thus, a researcher must carefully consider the impacts of genetic variation in the test population. For example, considerable heterogeneity exists in the xenobiotic-metabolizing enzymes responsible for drug metabolism in humans. This variability results in very different pharmacokinetics in different individuals. Researchers must also be cognizant that rodent models deviate from humans in a number of tissue-dependent and tissue-nondependent ways. As an example, the digestive metabolism of rodents (e.g., amino acid responses) can deviate substantially from humans [105]. The intestinal microbiota of rodents also differs in composition from people, and the homogeneity of the microbiota among individual inbred rodents may not simulate the complex and diverse responses that occur in humans. These observations emphasize the requirement to validate findings in animal models in humans. While animal models permit researchers to study mechanisms, efficacy evaluations conducted in animal models cannot always correctly predict responses that will occur in human subjects. Genetic diversity, societal and cultural behaviours, and socioeconomic aspects of individuals and various populations of people make it impossible to accurately generalize responses. It is thus essential that hypotheses and experimental designs be carefully considered to address the inherent advantages and disadvantages of the animal model selected. In addition, the model status of a given animal should be constantly challenged in order to identify and address its limitations. A gold standard model for a particular system should be developed wherever possible.

1.5.7 *In vitro* gut models

In vitro gut models offer investigations into gut microbial species and their interactions which each other and exogenous factors [106]. They are advantageous over the use of

faecal samples in analysis of the gut microbiome as composition and activity in different regions of the simulated gut can be investigated at specific time points. Additionally, procedures can be standardized to allow reproducibility and there are no ethical constraints associated so experimental design is virtually unlimited [106, 107].

Several different *in vitro* gut models exist including batch culture, continuous culture, and multistage continuous culture. Short-term batch incubations are the simplest form of *in vitro* gut modelling and while they are valuable for initial studies, as they are closed systems, build-up of metabolites can impact on validity of results obtained [107]. While batch models can be useful in investigating the role of dietary components, such as prebiotics, multistage continuous culture can be used to investigate the consequence of probiotic and prebiotic administration on the proximal, transverse, and distal colon [106]. *In vitro* colon models have been used to determine what drugs are significantly metabolized by bacteria of the GIT. By understanding that simvastatin (a statin), prednisolone (a steroid), and ranitidine (a H_2 antagonist) can be metabolized by the gut microbiome further steps can be taken in order to ensure adequate dosage of the medication despite microbial metabolism [39]. The introduction of a mucosal environment may lead to a more *in vivo* like community and allow for the study of the mucosal microbiome [107]. However, over-simplification of the GIT is a major issue associated with the use of such models as host intestinal factors are only partially studied in some models with appropriate feedback mechanisms completely absent. Additionally, microbial balancing, to ensure an accurate representation of the gut microbiome, which is not impacted by features of the system other than the experimental factors, can be difficult [106, 107].

1.6 Conclusion

The human gut microbiome is a diverse community whose structure is impacted by a variety of factors; both short- and long-term. Although great advances have been made in understanding the composition and function of the gut microbiome, gaps remain. Technological advancements pave the way for further analysis investigations of the gut microbiome, where a focus should be on mechanisms and elucidating the specific consequences of the compositional and functional differences which have been revealed.

1.7 References

1. Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C.R., Yu, W.-H., Lakshmanan, A., and Wade, W.G., *The human oral microbiome.* Journal of Bacteriology, 2010. **192**(19): p. 5002-5017.
2. Harris, J.K., Fang, R., Wagner, B.D., Choe, H.N., Kelly, C., J., Schroeder, S., Moore, W., Stevens, M.J., Yeckes, A., Amsden, K., Kagalwalla, A.F., Zalewski, A., Hirano, I., Gonsalves, N., Henry, L.N., Masterson, J.C., Robertson, C.E., Leung, D.Y., Pace, N.R., Ackerman, S.J., Furuta, G.T., and Fillon, S.A., *Esophageal microbiome in eosinophilic esophagitis.* PLoS One, 2015. **10**(5).
3. Sundin, O.H., Mendoza-Ladd, A., Zeng, M., Diaz-Arevalo, D., Morales, E., Fagan, B.M., Ordonez, J., Velez, P., Antony, N., and McCallum, R.W., *The human jejunum has an endogenous microbiota that differs from those in the oral cavity and colon.* BMC Microbiology, 2017. **17**(1): p. 160.
4. Nardone, G. and Compare, D., *The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases?* United European Gastroenterology Journal, 2015. **3**(3): p. 255-260.
5. Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., Lewis, J.D., and Bushman, F.D., *Archaea and fungi of the human gut microbiome: Correlations with diet and bacterial residents.* PLoS One, 2013. **8**(6).
6. Bellaguarda, E. and Chang, E.B., *IBD and the gut microbiota-from bench to personalized medicine.* Current Gastroenterology Reports, 2015. **17**(4): p. 15.
7. 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., and Gordon, J.I., *Evolution of mammals and their gut microbes.* Science, 2008. **320**(5883): p. 1647-1651.
8. Zitvogel, L., Galluzzi, L., Viaud, S., Vetizou, M., Daillere, R., Merad, M., and Kroemer, G., *Cancer and the gut microbiota: An unexpected link.* Science Translational Medicine, 2015. **7**(271).
9. Wu, H.J. and Wu, E., *The role of gut microbiota in immune homeostasis and autoimmunity.* Gut Microbes, 2012. **3**(1): p. 4-14.
10. Smirnov, K.S., Maier, T.V., Walker, A., Heinzmann, S.S., Forcisi, S., Martinez, I., Walter, J., and Schmitt-Kopplin, P., *Challenges of metabolomics in human gut microbiota research.* International Journal of Medical Microbiology, 2016. **306**(5): p. 266-279.

11. Wu, W.M., Yang, Y.S., and Peng, L.H., *Microbiota in the stomach: New insights.* Journal of Digestive Diseases, 2014. **15**(2): p. 54-61.
12. Yang, I., Nell, S., and Suerbaum, S., *Survival in hostile territory: The microbiota of the stomach.* FEMS Microbiology Reviews, 2013. **37**(5): p. 736-761.
13. Araujo, J.R., Tomas, J., Brenner, C., and Sansonetti, P.J., *Impact of high-fat diet on the intestinal microbiota and small intestinal physiology before and after the onset of obesity.* Biochimie, 2017.
14. Zaura, E., Brandt, B.W., Teixeira de Mattos, M.J., Buijs, M.J., Caspers, M.P., Rashid, M.U., Weintraub, A., Nord, C.E., Savell, A., Hu, Y., Coates, A.R., Hubank, M., Spratt, D.A., Wilson, M., Keijser, B.J., and Crielaard, W., *Same exposure but two radically different responses to antibiotics: Resilience of the salivary microbiome versus long-term microbial shifts in feces.* MBio, 2015. **6**(6): p. e01693-01615.
15. Di Pilato, V., Freschi, G., Ringressi, M.N., Pallecchi, L., Rossolini, G.M., and Bechi, P., *The esophageal microbiota in health and disease.* Annals of the New York Academy of Sciences, 2016. **1381**(1): p. 21-33.
16. El Aidy, S., van den Bogert, B., and Kleerebezem, M., *The small intestine microbiota, nutritional modulation and relevance for health.* Current Opinion in Biotechnology, 2015. **32**: p. 14-20.
17. Angelakis, E., Armougom, F., Carriere, F., Bachar, D., Laugier, R., Lagier, J.C., Robert, C., Michelle, C., Henrissat, B., and Raoult, D., *A metagenomic investigation of the duodenal microbiota reveals links with obesity.* PLoS One, 2015. **10**(9): p. e0137784.
18. Giamarellos-Bourboulis, E., Tang, J., Pyleris, E., Pistiki, A., Barbatzas, C., Brown, J., Lee, C.C., Harkins, T.T., Kim, G., Weitsman, S., Barlow, G.M., Funari, V.A., and Pimentel, M., *Molecular assessment of differences in the duodenal microbiome in subjects with irritable bowel syndrome.* Scandinavian Journal of Gastroenterology, 2015. **50**(9): p. 1076-1087.
19. Cotter, P.D., *Small intestine and microbiota.* Current Opinion Gastroenterology, 2011. **27**: p. 99-105.
20. Mao, S., Zhang, M., Liu, J., and Zhu, W., *Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: Membership and potential function.* Scientific Reports, 2015. **5**.
21. Johansson, M.E., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and Hansson, G.C., *The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of*

bacteria. Proceedings of the National Academy of Sciences, 2008. **105**(39): p. 15064-15069.

22. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-230.
23. 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., and Lewis, J.D., *Linking long-term dietary patterns with gut microbial enterotypes.* Science, 2011. **334**(6052): p. 105-108.
24. D'Argenio, V. and Salvatore, F., *The role of the gut microbiome in the healthy adult status.* Clinica Chimica Acta, 2015. **451**(Pt A): p. 97-102.
25. Jeffery, I.B., Lynch, D.B., and O'Toole, P.W., *Composition and temporal stability of the gut microbiota in older persons.* The ISME Journal, 2016. **10**(1): p. 170-182.
26. Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., Marchesi, J.R., and Collado, M.C., *The composition of the gut microbiota throughout life, with an emphasis on early life.* Microbial Ecology in Health and Disease, 2015. **26**: p. 26050.
27. Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M.T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y.S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Masen, L., Kristiansen, K., Dahlgren, J., and Wang, J., *Dynamics and stabilization of the human gut microbiome during the first year of life.* Cell Host Microbe, 2015. **17**: p. 690-703.
28. Hill, C.J., Lynch, D.B., Murphy, K., Ulaszewska, M., Jeffery, I.B., O' Shea, C.A., Watkins, C., Dempsey, E., Mattivi, F., Tuohy, K., Ross, R.P., Ryan, C.A., O' Toole, P.W., and Stanton, C., *Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET cohort.* Microbiome, 2017. **5**(4).
29. Asnicar, F., Manara, S., Zolfo, M., Truong, D.T., Scholz, M., Armanini, F., Ferretti, P., Gorfer, V., Pedrotti, A., Tett, A., and Segata, N., *Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling.* mSystems, 2017. **2**(1).
30. Yassour, M., Vatanen, T., Siljander, H., Hamalainen, A.M., Harkonen, T., Ryhanen, S.J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E.S., Knip, M., Group, D.S., and Xavier, R.J., *Natural history of the infant gut microbiome and*

impact of antibiotic treatment on bacterial strain diversity and stability. Science Translational Medicine, 2016. **8**(343): p. 343ra381.

31. Golubeva, A.V., Crampton, S., Desbonnet, L., Edge, D., O'Sullivan, O., Lomasney, K.W., Zhdanov, A.V., Crispie, F., Moloney, R.D., Borre, Y.E., Cotter, P.D., Hyland, N.P., O'Halloran, K.D., Dinan, T.G., O'Keeffe, G.W., and Cryan, J.F., *Prenatal stress-induced alterations in major physiological systems correlate with gut microbiota composition in adulthood.* Psychoneuroendocrinology, 2015. **60**: p. 58-74.
32. Zijlmans, M.A., Korpela, K., Riksen-Walraven, J.M., de Vos, W.M., and de Weerth, C., *Maternal prenatal stress is associated with the infant intestinal microbiota.* Psychoneuroendocrinology, 2015. **53**: p. 233-245.
33. Newburg, D.S., Ruiz-Palacios, G.M., Altaye, M., Chaturvedi, P., Meinzen-Derr, J., Guerrero Mde, L., and Morrow, A.L., *Innate protection conferred by fucosylated oligosaccharides of human milk against diarrhea in breastfed infants.* Glycobiology, 2004. **14**(3): p. 253-263.
34. Laursen, M.F., Bahl, M.I., Michaelsen, K.F., and Licht, T.R., *First foods and gut microbes.* Frontiers in Microbiology, 2017. **8**: p. 356.
35. 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., and O'Toole, P.W., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**: p. 178-185.
36. Dicksved, J., Halfvarson, J., Rosenquist, M., Jarnerot, G., Tysk, C., Apajalahti, J., Engstrand, L., and Jansson, J.K., *Molecular analysis of the gut microbiota of identical twins with Crohn's disease.* The ISME Journal, 2008. **2**(7): p. 716-727.
37. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R., and Gordon, J.I., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-227.
38. Jackson, M.A., Goodrich, J.K., Maxan, M.E., Freedberg, D.E., Abrams, J.A., Poole, A.C., Sutter, J.L., Welter, D., Ley, R.E., Bell, J.T., Spector, T.D., and Steves, C.J., *Proton pump inhibitors alter the composition of the gut microbiota.* Gut, 2016. **65**(5): p. 749-756.

39. Swanson, H.I., *Drug metabolism by the host and gut microbiota: A partnership or rivalry?* Drug Metabolism and Disposition, 2015. **43**: p. 1499-1504.
40. Keeney, K.M., Yurist-Doutsch, S., Arrieta, M.C., and Finlay, B.B., *Effects of antibiotics on human microbiota and subsequent disease.* Annual Review of Microbiology, 2014. **68**: p. 217-235.
41. Korpela, K., Salonen, A., Virta, L.J., Kekkonen, R.A., Forslund, K., Bork, P., and de Vos, W.M., *Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children.* Nature Communications, 2016. **7**.
42. Dardas, M., Gill, S.R., Grier, A., Pryhuber, G.S., Gill, A.L., Lee, Y.H., and Guillet, R., *The impact of postnatal antibiotics on the preterm intestinal microbiome.* Pediatric Research, 2014. **76**(2): p. 150-158.
43. Modi, S.R., Collins, J.J., and Relman, D.A., *Antibiotics and the gut microbiota.* Journal of Clinical Investigation, 2014. **124**(10): p. 4212-4218.
44. 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., and Cotter, P.D., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.
45. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.
46. Petersen, L.M., Bautista, E.J., Nguyen, H., Hanson, B.M., Chen, L., Lek, S.H., Sodergren, E., and Weinstock, G.M., *Community characteristics of the gut microbiomes of competitive cyclists.* Microbiome, 2017. **5**(1): p. 98.
47. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).
48. Finegold, S.M., Attebery, H.R., and Sutter, V.L., *Effect of diet on human fecal flora: Comparison of Japanese and American diets.* The American Journal of Clinical Nutrition, 1974. **27**(12): p. 1456-1469.
49. Peltonen, R. and Eerola, E., *Direct automatic bacterial analysis of rat stool samples; the effects of diet and medical treatment studied by computerised gas-liquid*

chromatography of bacterial fatty acids. Microbial Ecology in Health and Disease, 1992. **5**(2).

50. Peltonen, R., Ling, W.H., Hanninen, O., and Eerola, E., *An uncooked vegan diet shifts the profile of human fecal microflora: Computerized analysis of direct stool sample gas-liquid chromatography profiles of bacterial cellular fatty acids.* Applied and Environmental Microbiology, 1992. **58**(11): p. 3660-3666.
51. Zhang, C., Yin, A., Li, H., Wang, R., Wu, G., Shen, J., Zhang, M., Wang, L., Hou, Y., Ouyang, H., Zhang, Y., Zheng, Y., Wang, J., Lv, X., Wang, Y., Zhang, F., Zeng, B., Li, W., Yan, F., Zhao, Y., Pang, X., Weinkove, D., Clement, K., Dore, J., Holmes, E., Xiao, H., Zhao, G., Zhang, X., and Zhao, L., *Dietary modulation of gut microbiota contributes to alleviation of both genetic and simple obesity in children.* EBioMedicine, 2015: p. 968-984.
52. Zheng, H., Lorenzen, J.K., Astrup, A., Larsen, L.H., Yde, C.C., Clausen, M.R., and Bertram, H.C., *Metabolic effects of a 24-week energy-restricted intervention combined with low or high dairy intake in overweight women: An NMR-based metabolomics investigation.* Nutrients, 2016. **8**(3): p. 108.
53. Purchiaroni, F., Tortora, A., Gabrielli, M., Bertucci, F., Gigante, G., Ianiro, G., Ojetti, V., Scarpellini, E., and Gasbarrini, A., *The role of intestinal microbiota and the immune system.* European Review for Medical and Pharmacological Sciences, 2013. **17**(3): p. 323-333.
54. Abreu, M.T., *Toll-like receptor signalling in the intestinal epithelium: How bacterial recognition shapes intestinal function.* Nature Reviews Immunology, 2010. **10**: p. 131-144.
55. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R., *Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis.* Cell, 2004. **118**(2): p. 229-241.
56. Neish, A.S., *Microbes in gastrointestinal health and disease.* Gastroenterology, 2009. **136**(1): p. 65-80.
57. Cebra, J.J., *Influences of microbiota on intestinal immune system development.* American Journal of Clinical Nutrition, 1999. **69**(69): p. 1046S-1051S.
58. Eberl, G. and Lochner, M., *The development of intestinal lymphoid tissues at the interface of self and microbiota.* Mucosal Immunology, 2009. **2**(6): p. 478-485.
59. Coombes, J.L. and Powrie, F., *Dendritic cells in intestinal immune regulation.* Nature Reviews Immunology, 2008. **8**(6): p. 435-446.

60. Kumar, A., Wu, H., Collier-Hyams, L.S., Kwon, Y.M., Hanson, J.M., and Neish, A.S., *The bacterial fermentation product butyrate influences epithelial signaling via reactive oxygen species-mediated changes in cullin-1 neddylation.* Journal of Immunology 2009. **182**(1): p. 538-546.

61. Flint, H.J., Duncan, S.H., Scott, K.P., and Louis, P., *Links between diet, gut microbiota composition and gut metabolism.* Proceedings of the Nutrition Society, 2015. **74**(1): p. 13-22.

62. Qin, J., Li, R., Raea, J., Arumugam, M., Solvsten-Burgdorf, K., 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., Pederson, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., and Wang, J., *A human gut microbial gene catalog established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.

63. Wlodarska, M., Kostic, A.D., and Xavier, R.J., *An integrative view of microbiome-host interactions in inflammatory bowel diseases.* Cell Host Microbe, 2015. **17**(5): p. 577-591.

64. Ogbonnaya, E.S., Clarke, G., Shanahan, F., Dinan, T.G., Cryan, J.F., and O'Leary, O.F., *Adult hippocampal neurogenesis is regulated by the microbiome.* Biological Psychiatry, 2015. **78**(4): p. 7-9.

65. O'Mahony, S.M., Clarke, G., Borre, Y.E., Dinan, T.G., and Cryan, J.F., *Serotonin, tryptophan metabolism and the brain-gut-microbiome axis.* Behavioural Brain Research, 2015. **277**: p. 32-48.

66. Vignal, C., Singer, E., Peyrin-Biroulet, L., Desreumaux, P., and Chamaillard, M., *How NOD2 mutations predispose to Crohn's disease?* Microbes and Infection, 2007. **9**(5): p. 658-663.

67. Hua, X., Goedart, J.J., Pu, A., Yu, G., and Shi, J., *Allergy associations with the adult fecal microbiota: Analysis of the American gut project.* EBioMedicine, 2015: p. 172-179.

68. Stecher, B. and Hardt, W.D., *The role of microbiota in infectious disease.* Trends in Microbiology, 2008. **16**(3): p. 107-114.

69. Vollaard, E.J. and Clasener, H.A., *Colonization resistance.* Antimicrobial Agents and Chemotherapy, 1994. **38**(3): p. 409-414.

70. Lupp, C., Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L., Gaynor, E.C., and Finlay, B.B., *Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae.* Cell Host and Microbe, 2007. **2**(2): p. 119-129.

71. Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M., Barthel, M., Kremer, M., Chaffron, S., Macpherson, A.J., Buer, J., Parkhill, J., Dougan, G., von Mering, C., and Hardt, W.D., *Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota.* PLoS Biology, 2007. **5**(10): p. 2177-2189.

72. Chang, J.Y., Antonopoulos, D.A., Kalra, A., Tonelli, A., Khalife, W.T., Schmidt, T.M., and Young, V.B., *Decreased diversity of the fecal microbiome in recurrent Clostridium difficile-associated diarrhea.* The Journal of Infectious Diseases, 2008. **197**(3): p. 435-438.

73. Kellow, N.J., Coughlan, M.T., and Reid, C.M., *Metabolic benefits of dietary prebiotics in human subjects: A systematic review of randomised controlled trials.* British Journal of Nutrition, 2014. **111**(7): p. 1147-1161.

74. Pandey, K.R., Naik, S.R., and Vakil, B.V., *Probiotics, prebiotics and synbiotics- A review.* Journal of Food Science and Technology, 2015. **52**(12): p. 7577-7587.

75. Parnell, J.A. and Reimer, R.A., *Weight loss during oligofructose supplementation is associated with decreased ghrelin and increased peptide YY in overweight and obese adults.* The American Journal of Clinical Nutrition, 2009. **89**(6): p. 1751-1759.

76. Giacco, R., Clemente, G., Luongo, D., Lasorella, G., Fiume, I., Brouns, F., Bornet, F., Patti, L., Cipriano, P., Rivellese, A.A., and Riccardi, G., *Effects of short-chain fructo-oligosaccharides on glucose and lipid metabolism in mild hypercholesterolaemic individuals.* Clinical Nutrition, 2004. **23**(3): p. 331-340.

77. Cani, P.D., Joly, E., Horsmans, Y., and Delzenne, N.M., *Oligofructose promotes satiety in healthy human: A pilot study.* European Journal of Clinical Nutrition, 2006. **60**(5): p. 567-572.

78. Butel, M.J., *Probiotics, gut microbiota and health.* Medecine et Maladies Infectieuses, 2014. **44**(1): p. 1-8.

79. Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J., Salminen, S., Calder, P.C., and Sanders, M.E., *Expert consensus document. The international scientific association for probiotics and*

prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nature Reviews Gastroenterology & Hepatology, 2014. **11**(8): p. 506-514.

80. Ritchie, M.L. and Romanuk, T.N., *A meta-analysis of probiotic efficacy for gastrointestinal diseases.* PLoS One, 2012. **7**(4): p. e34938.
81. Bernaola Apone, G., Bada Mancilla, C.A., Carreazo, N.Y., and Rojas Galarza, R.A., *Probiotics for treating persistent diarrhoea in children.* Cochrane Database of Systematic Reviews, 2013(8).
82. Allen, S.J., Martinez, E.G., Gregorio, G.V., and Dans, L.F., *Probiotics for treating acute infectious diarrhoea.* Cochrane Database of Systematic Reviews, 2010(11).
83. Rossen, N.G., MacDonald, J.K., de Vries, E.M., D'Haens, G.R., de Vos, W.M., Zoetendal, E.G., and Ponsioen, C.Y., *Fecal microbiota transplantation as novel therapy in gastroenterology: A systematic review.* World Journal of Gastroenterology, 2015. **21**(17): p. 5359-5371.
84. Xu, M.Q., Cao, H.L., Wang, W.Q., Wang, S., Cao, X.C., Yan, F., and Wang, B.M., *Fecal microbiota transplantation broadening its application beyond intestinal disorders.* World Journal of Gastroenterology, 2015. **21**(1): p. 102-111.
85. Kao, D., Hotte, N., Gillevet, P., and Madsen, K., *Fecal microbiota transplantation inducing remission in Crohn's colitis and the associated changes in fecal microbial profile.* Journal of Clinical Gastroenterology, 2014. **48**(7): p. 625-628.
86. Gordon, H. and Harbord, M., *A patient with severe Crohn's colitis responds to faecal microbiota transplantation.* Journal of Crohn's and Colitis, 2014. **8**(3): p. 256-257.
87. Hohmann, E.L., Ananthakrishnan, A.N., and Deshpande, V., *Case records of the Massachusetts general hospital. Case 25-2014. A 37-year-old man with ulcerative colitis and bloody diarrhea.* The New England Journal of Medicine, 2014. **371**(7): p. 668-675.
88. Pinn, D.M., Aroniadis, O.C., and Brandt, L.J., *Is fecal microbiota transplantation (FMT) an effective treatment for patients with functional gastrointestinal disorders (FGID)?* Neurogastroenterology & Motility, 2015. **27**(1): p. 19-29.
89. Vrieze, A., Van Nood, E., Holleman, F., Salojarvi, J., Kootte, R.S., Bartelsman, J.F., Dallinga-Thie, G.M., Ackermans, M.T., Serlie, M.J., Oozeer, R., Derrien, M., Druesne, A., Van Hylckama Vlieg, J.E., Bloks, V.W., Groen, A.K., Heilig, H.G., Zoetendal, E.G., Stroes, E.S., de Vos, W.M., Hoekstra, J.B., and Nieuwdorp, M., *Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome.* Gastroenterology, 2012. **143**(4): p. 913-916 e917.

90. Fouhy, F., Deane, J., Rea, M.C., O'Sullivan, O., Ross, R.P., O'Callaghan, G., Plant, B.J., and Stanton, C., *The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations.* PLoS One, 2015. **10**(3): p. e0119355.

91. Lagkouvardos, I., Overmann, J., and Clavel, T., *Cultured microbes represent a substantial fraction of the human and mouse gut microbiota.* Gut Microbes, 2017.

92. Lagier, J.C., Hugon, P., Khelaifia, S., Fournier, P.E., La Scola, B., and Raoult, D., *The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota.* Clinical Microbiology Reviews, 2015. **28**(1): p. 237-264.

93. Ben-Amor, K., Heilig, H., Smidt, H., Vaughan, E.E., Abee, T., and de Vos, W.M., *Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16s rRNA gene analysis.* Applied and Environmental Microbiology, 2005. **71**(8): p. 4679-4689.

94. Choo, J.M., Leong, L.E., and Rogers, G.B., *Sample storage conditions significantly influence faecal microbiome profiles.* Scientific Reports, 2015. **5**: p. 16350.

95. Hill, C.J., Brown, J.M.R., Lynch, D.B., Jeffery, I.B., Ryan, C.A., Ross, R.P., Stanton, C., and O' Toole, P.W., *Effect of room temperature transport vials on DNA quality and phylogenetic composition of faecal microbiota of elderly adults and infants.* Microbiome, 2016. **4**(19).

96. Claassen, S., du Toit, E., Kaba, M., Moodley, C., Zar, H.J., and Nicol, M.P., *A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples.* Journal of Microbiological Methods, 2013. **94**: p. 103-110.

97. Ferrand, J., Patron, K., Legrand-Frossi, C., Frippiat, J.-P., Merlin, C., Alauzet, C., and Lozniewski, A., *Comparison of seven methods for extraction of bacterial DNA from fecal and cecal samples of mice.* Journal of Microbiological Methods, 2014. **105**: p. 180-185.

98. D'Amore, R., Ijaz, U.Z., Schirmer, M., Kenny, J.G., Gregory, R., Darby, A.C., Shakya, M., Podar, M., Quince, C., and Hall, N., *A comprehensive benchmarking study of protocols and sequencing platforms for 16s rRNA community profiling.* BMC Genomics, 2016. **17**(55).

99. Pareek, C.S., Smoczynski, R., and Tretyn, A., *Sequencing technologies and genome sequencing.* Journal of Applied Genetics, 2011. **52**(4): p. 413-435.

100. Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., and Wong, G.K., *Characterization of the gut microbiome using 16s or shotgun metagenomics.* Frontiers in Microbiology, 2016. **7**: p. 459.

101. Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., and Zheng, S.-S., *Application of metagenomics in the human gut microbiome.* World Journal of Gastroenterology, 2015. **21**(3): p. 803-814.

102. Vernocchi, P., Del Chierico, F., and Putignani, L., *Gut microbiota profiling: Metabolomics based approach to unravel compounds affecting human health.* Frontiers in Microbiology, 2016. **7**: p. 1144.

103. Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., Giannoukos, G., Boylan, M.R., Ciulla, D., Gevers, D., Izard, J., Garrett, W.S., Chan, A.T., and Huttenhower, C., *Relating the metatranscriptome and metagenome of the human gut.* Proceedings of the National Academy of Sciences, 2014. **111**(22): p. E2329-2338.

104. Asquith, M.J., Boulard, O., Powrie, F., and Maloy, K.J., *Pathogenic and protective roles of MyD88 in leukocytes and epithelial cells in mouse models of inflammatory bowel disease.* Gastroenterology, 2010. **139**: p. 519-529.

105. Baker, D.H., *Animal models of human amino acid responses.* Journal of Nutrition, 2004. **134**(Suppl 6): p. 1646S-1650S.

106. Payne, A.N., Zihler, A., Chassard, C., and Lacroix, C., *Advances and perspectives in in vitro human gut fermentation modeling.* Trends in Biotechnology, 2012. **30**(1): p. 17-25.

107. Venema, K. and van den Abbeele, P., *Experimental models of the gut microbiome.* Best Practice & Research Clinical Gastroenterology, 2013. **27**: p. 115-126.

Chapter 2

‘Microbes in Sport’ – The Potential Role of the Gut Microbiota in Athlete Health and Performance

Updated since publication in the *British Journal of Sports Medicine*

(doi: http://dx.doi.org/10.1136/bjsports-2016-097227corr1)

Authors: Alan Rankin, Ciara O’ Donovan, Sharon M. Madigan, Orla O’ Sullivan, and Paul D. Cotter

Contributions:

- **Candidate** contributed to the original writing of the article and updated the current version.
- **AR** and **SMM** contributed to the original editorial manuscript.
- **OOS** and **PDC** provided guidance.

2.0 Abstract

The human intestinal tract is home to approximately 100 trillion microbes, which are involved with an array of functions including digestion, immune system interactions, and contributions to health and disease. Exercise is one of the multiple factors which have the potential to influence the gut microbiome from both a compositional and functional perspective. Athletes, who participate in exercise at the extreme end of a spectrum, have a gut microbiome which is unique from non-athletes. The alteration and modulation of the gut microbiome in athletes has the potential to have an influence over general athlete health, musculoskeletal health, and potentially athletic performance.

2.1 Introduction

The human intestinal tract is home to approximately 100 trillion microbes with the majority of these residing in the colon. The microbes in the human intestinal tract include bacteria, archaea, fungi, protozoans, and a large population of viruses [1-5]. These microbes have a vast array of functions including vitamin production, fibre digestion, interacting with the immune system, and they contribute significantly to health and disease [6-12].

Previously, it was only possible to study the human gut microbiota (GM; microbiota refers to the assemblage of microorganisms present in a defined environment) after first culturing the microbes on agar plates. Gut microbes have an extremely specific set of growth conditions *in vivo* that are difficult to recreate in a lab environment. Advancements in DNA sequencing technology has made it possible to study the genetic material of the microbes present and negates the need for culture [7, 13]. Increased understanding of the role of GM has led to innovative methods to alter gut microbe composition and subsequently GM function. Many factors have been shown to influence the GM including antibiotics, other medication, diet, and exercise [14-19]. Research investigating links between exercise (+/− diet) and the GM in humans, has yielded inconsistent results [20-23].

To date, only a relatively small number of human studies, other than those described in this thesis, have investigated the athlete GM. Rugby players were identified as having greater microbial diversity (a positive indicator of gut health) compared with controls, which correlated with a higher protein intake (dietary factor) and higher creatine kinase levels (physical activity factor) [18, 19]. Exercise load in amateur and professional cyclists was correlated with relative abundance of *Prevotella*, while *Methanobrevibacter smithii* abundance was found to be increased in professional cyclists relative to amateur [24]. Sports type, and the associated diet, was found to be characteristic of the GM [25], while

the participation in an ultra-endurance event by athletes was found to increase alpha diversity and butyrate producing species over time [26].

2.2 The gut microbiota and general athlete health

There are many routes through which the GM can impact on athlete health. An undesirable GM composition has been associated with local inflammatory change leading to gut wall permeability that may permit increased systemic migration of bacterial material [27]. This in turn, may lead to systemic immune and metabolic dysfunctions. These processes are implicated in many chronic diseases [28-30]. From an immunology perspective, compromise of GM and subsequent immunopathology has also been associated with atopic and allergic conditions such as asthma [31]. Allergic conditions are common in athletes whose airways are often exposed to environmental factors such as cold air or chlorine during aquatic sports [28, 32]. Allergic conditions increase the risk of upper respiratory tract infections. It has been shown that probiotic supplementation can reduce the incidence and severity of upper respiratory tract infections in some athletes [28, 33-35]. Gastrointestinal complaints are common in the spectrum of conditions associated with 'relative energy deficiency in sport' and the 'female athlete triad'. In this low energy state, multiple body systems may experience compromise, including metabolic and hormonal functions. Changes in GM composition are likely to occur following a period of relative energy deficiency. If relevant GM changes could be demonstrated, this could prove a useful biomarker of energy deficiency and provide a target for therapeutic intervention. Notably, GM alterations have also been observed following prolonged periods of physical or psychological stress in conditions such as chronic fatigue syndrome and post-traumatic stress disorder [36-38]. Furthermore, changes to the hypothalamic-pituitary axis are influenced by GM [28, 39, 40]. This hormonal axis has also been implicated in overtraining

syndrome. Investigation of GM in athletes in states of overtraining or non-functional over-reaching may aid understanding of these conditions.

2.3 The gut microbiota and musculoskeletal health

Changes to GM have been investigated in chronic musculoskeletal conditions. Multiple studies suggest a relationship between GM and inflammatory conditions such as rheumatoid arthritis, spondyloarthropathies, and gout [41-44]. Bacterial DNA of oral and gut origin has been found in blood serum and joint fluid in these conditions, suggesting a compromised gut barrier is a contributing factor. Low-grade inflammation, GM alterations and bacterial DNA are also present in degenerative musculoskeletal conditions such as osteoarthritis and rotator cuff tendon degeneration [41, 45, 46]. Chronic tendinopathy is more prevalent in persons with metabolic dysfunctions including type 2 diabetes and dyslipidaemia. Changes to GM are present in metabolic dysfunctions [28, 47, 48] and investigation of GM in chronic tendinopathy may provide valuable insights. Ensuring optimal bone health in athletes is important to reduce injury risk and to aid recovery. GM has a proposed regulatory effect on bone mass by altering the skeletal immune system, influencing hormonal regulation of bone metabolism, and by production of bacterial metabolites that act as cellular messengers to the bone [49-51]. GM analysis may provide a biomarker for bone health and target for therapeutic manipulation in cases of acute fracture, stress fracture, and osteoporosis.

2.4 Potential role of the gut microbiota in athletic performance

The optimisation of GM for athlete health and injury treatment will produce indirect benefits to athletic performance. Research investigating performance benefits from alterations in the GM is beginning to emerge. Changes to GM can positively alter body composition through a number of mechanisms [52]. One study [53] has demonstrated that GM composition is associated with performance in a rodent model. This was proposed to

occur through the action of GM on antioxidant enzyme systems. *Veillonella*, a genus identified as increased following marathon running, and specifically *V. atypica* has been found to increase maximum running times, in a mouse model. The increased levels of this species following marathon running, and the associated performance, have been hypothesized to be associated with its capacity to breakdown lactate [54]. Further research in this area would be of value for many competitive sports.

2.5 Conclusions

While research on the potential of GM in sports medicine is in its infancy, there is definite potential to positively impact athlete health, injury and ultimately performance, as greater understanding of the complex microbe-human relationship is developed.

2.6 References

1. Ley, R.E., Peterson, D.A., and Gordan, J.I., *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. **124**: p. 837-848.
2. Whitman, W.B., Coleman, D.C., and Wiebe, W.J., *Prokaryotes: The unseen majority.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6578-6583.
3. Qin, J., Li, R., Raea, J., Arumugam, M., Solvsten-Burgdorf, K., 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., Pederson, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., and Wang, J., *A human gut microbial gene catalog established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.
4. Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., Lewis, J.D., and Bushman, F.D., *Archaea and fungi of the human gut microbiome: Correlations with diet and bacterial residents.* PLoS One, 2013. **8**(6): p. e66019.
5. Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J., *Healthy human gut phageome.* Proc Natl Acad Sci U S A, 2016. **113**(37): p. 10400-10405.
6. Cronin, O., Molloy, M.G., and Shanahan, F., *Exercise, fitness, and the gut.* Curr Opin Gastroenterol, 2016. **32**(2): p. 67-73.
7. Ursell, L.K., Metcalf, J.L., Parfrey, L.W., and Knight, R., *Defining the human microbiome.* Nutr Rev, 2012. **70 Suppl 1**: p. S38-44.
8. Das, P., Babaei, P., and Nielsen, J., *Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome.* BMC Genomics, 2019. **20**(1): p. 208.
9. El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B., *The abundance and variety of carbohydrate-active enzymes in the human gut microbiota.* Nat Rev Microbiol, 2013. **11**(7): p. 497-504.
10. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R., *Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis.* Cell, 2004. **118**(2): p. 229-241.

11. Shreiner, A.B., Kao, J.Y., and Young, V.B., *The gut microbiome in health and in disease.* Curr Opin Gastroenterol, 2015. **31**(1): p. 69-75.
12. Kinross, J.M., Darzi, A.W., and Nicholson, J.K., *Gut microbiome-host interactions in health and disease.* Genome Med, 2011. **3**(3): p. 14.
13. Kuczynski, J., Lauber, C.L., Walters, W.A., Parfrey, L.W., Clemente, J.C., Gevers, D., and Knight, R., *Experimental and analytical tools for studying the human microbiome.* Nat Rev Genet, 2011. **13**(1): p. 47-58.
14. O'Sullivan, O., Cronin, O., Clarke, S.F., Murphy, E.F., Molloy, M.G., Shanahan, F., and Cotter, P.D., *Exercise and the microbiota.* Gut Microbes, 2015. **6**(2): p. 131-136.
15. Jakobsson, H.E., Jernberg, C., Andersson, A.F., Sjolund-Karlsson, M., Jansson, J.K., and Engstrand, L., *Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome.* PLoS One, 2010. **5**(3): p. e9836.
16. Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R., and Nicholson, J.K., *Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism.* Proc Natl Acad Sci U S A, 2009. **106**(34): p. 14728-14733.
17. 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., and Lewis, J.D., *Linking long-term dietary patterns with gut microbial enterotypes.* Science, 2011. **334**(6052): p. 105-108.
18. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.
19. 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., and D., C.P., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.
20. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).

21. Allen, J.M., Mailing, L.J., Niemiro, G.M., Moore, R., Cook, M.D., White, B.A., Holscher, H.D., and Woods, J.A., *Exercise alters gut microbiota composition and function in lean and obese humans.* Medicine & Science in Sports & Exercise, 2018. **50**(4): p. 747-757.
22. Zhao, X., Zhang, Z., Hu, B., Huang, W., Yuan, C., and Zou, L., *Response of gut microbiota to metabolite changes induced by endurance exercise.* Front Microbiol, 2018. **9**: p. 765.
23. Cronin, O., Barton, W., Skuse, P., Penney, N.C., Garcia-Perez, I., Murphy, E.F., Woods, T., Nugent, H., Fanning, A., Melgar, S., Falvey, E.C., Holmes, E., Cotter, P.D., O'Sullivan, O., Molloy, M.G., and Shanahan, F., *A prospective metagenomic and metabolomic analysis of the impact of exercise and/or whey protein supplementation on the gut microbiome of sedentary adults.* mSystems, 2018. **3**(3).
24. Petersen, L.M., Bautista, E.J., Nguyen, H., Hanson, B.M., Chen, L., Lek, S.H., Sodergren, E., and Weinstock, G.M., *Community characteristics of the gut microbiomes of competitive cyclists.* Microbiome, 2017. **5**(1): p. 98.
25. Jang, L.G., Choi, G., Kim, S.W., Kim, B.Y., Lee, S., and Park, H., *The combination of sport and sport-specific diet is associated with characteristics of gut microbiota: An observational study.* J Int Soc Sports Nutr, 2019. **16**(1): p. 21.
26. Keohane, D.M., Woods, T., O'Connor, P., Underwood, S., Cronin, O., Whiston, R., O'Sullivan, O., Cotter, P., Shanahan, F., and Molloy, M.G.M., *Four men in a boat: Ultra-endurance exercise alters the gut microbiome.* J Sci Med Sport, 2019. **22**(9): p. 1059-1064.
27. Gutekunst, K., Kruger, K., August, C., Diener, M., and Mooren, F.C., *Acute exercises induce disorders of the gastrointestinal integrity in a murine model.* Eur J Appl Physiol, 2014. **114**(3): p. 609-617.
28. Bermon, S., Petriz, B., Kajeniene, A., Prestes, J., Castell, L., and Franco, O.L., *The microbiota: An exercise immunology perspective.* Exerc Immunol Rev, 2015. **21**: p. 70-79.
29. Strober, W., *Impact of the gut microbiome on mucosal inflammation.* Trends Immunol, 2013. **34**(9): p. 423-430.
30. 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., and Gordon, J.I.,

Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science, 2013. **341**(6150): p. 1241214.

31. Arrieta, M.C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K.M., Sears, M.R., Kollmann, T., Investigators, C.S., Mohn, W.W., Turvey, S.E., and Finlay, B.B., *Early infancy microbial and metabolic alterations affect risk of childhood asthma.* Sci Transl Med, 2015. **7**(307): p. 307ra152.
32. Helenius, I.J., Tikkanen, H.O., and Haahtela, T., *Association between type of training and risk of asthma in elite athletes.* Thorax, 1997. **52**(2): p. 157-160.
33. Pyne, D.B., West, N.P., Cox, A.J., and Cripps, A.W., *Probiotics supplementation for athletes - clinical and physiological effects.* Eur J Sport Sci, 2015. **15**(1): p. 63-72.
34. Gleeson, M., Bishop, N.C., Oliveira, M., and Tauler, P., *Daily probiotic's (lactobacillus casei shirota) reduction of infection incidence in athletes.* Int J Sport Nutr Exerc Metab, 2011. **21**(1): p. 55-64.
35. Cox, A.J., Pyne, D.B., Saunders, P.U., and Fricker, P.A., *Oral administration of the probiotic lactobacillus fermentum vri-003 and mucosal immunity in endurance athletes.* Br J Sports Med, 2010. **44**(4): p. 222-226.
36. Leclercq, S., Forsythe, P., and Bienenstock, J., *Posttraumatic stress disorder: Does the gut microbiome hold the key?* Can J Psychiatry, 2016. **61**(4): p. 204-213.
37. Hemmings, S.M.J., Malan-Muller, S., van den Heuvel, L.L., Demmitt, B.A., Stanislawski, M.A., Smith, D.G., Bohr, A.D., Stamper, C.E., Hyde, E.R., Morton, J.T., Marotz, C.A., Siebler, P.H., Braspenning, M., Van Criekinge, W., Hoisington, A.J., Brenner, L.A., Postolache, T.T., McQueen, M.B., Krauter, K.S., Knight, R., Seedat, S., and Lowry, C.A., *The microbiome in posttraumatic stress disorder and trauma-exposed controls: An exploratory study.* Psychosom Med, 2017. **79**(8): p. 936-946.
38. Nagy-Szakal, D., Williams, B.L., Mishra, N., Che, X., Lee, B., Bateman, L., Klimas, N.G., Komaroff, A.L., Levine, S., Montoya, J.G., Peterson, D.L., Ramanan, D., Jain, K., Eddy, M.L., Hornig, M., and Lipkin, W.I., *Fecal metagenomic profiles in subgroups of patients with myalgic encephalomyelitis/chronic fatigue syndrome.* Microbiome, 2017. **5**(1): p. 44.
39. Zeng, L., Zeng, B., Wang, H., Li, B., Huo, R., Zheng, P., Zhang, X., Du, X., Liu, M., Fang, Z., Xu, X., Zhou, C., Chen, J., Li, W., Guo, J., Wei, H., and Xie, P., *Microbiota*

modulates behavior and protein kinase c mediated camp response element-binding protein signaling. Sci Rep, 2016. **6**: p. 29998.

40. Dunn, A.J., Ando, T., Brown, R.F., and Berg, R.D., *Hpa axis activation and neurochemical responses to bacterial translocation from the gastrointestinal tract.* Ann N Y Acad Sci, 2003. **992**: p. 21-29.
41. Steves, C.J., Bird, S., Williams, F.M., and Spector, T.D., *The microbiome and musculoskeletal conditions of aging: A review of evidence for impact and potential therapeutics.* J Bone Miner Res, 2016. **31**(2): p. 261-269.
42. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J.Y., Chen, H., Wang, L., Wu, Q.J., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y., and Wang, J., *The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment.* Nat Med, 2015. **21**(8): p. 895-905.
43. Sinkorova, Z., Capkova, J., Niederlova, J., Stepankova, R., and Sinkora, J., *Commensal intestinal bacterial strains trigger ankylosing enthesopathy of the ankle in inbred b10.Br (h-2(k)) male mice.* Hum Immunol, 2008. **69**(12): p. 845-850.
44. Vieira, A.T., Macia, L., Galvao, I., Martins, F.S., Canesso, M.C., Amaral, F.A., Garcia, C.C., Maslowski, K.M., De Leon, E., Shim, D., Nicoli, J.R., Harper, J.L., Teixeira, M.M., and Mackay, C.R., *A role for gut microbiota and the metabolite-sensing receptor gpr43 in a murine model of gout.* Arthritis Rheumatol, 2015. **67**(6): p. 1646-1656.
45. Collins, K.H., Paul, H.A., Reimer, R.A., Seerattan, R.A., Hart, D.A., and Herzog, W., *Relationship between inflammation, the gut microbiota, and metabolic osteoarthritis development: Studies in a rat model.* Osteoarthritis Cartilage, 2015. **23**(11): p. 1989-1998.
46. Patel, A., Calfee, R.P., Plante, M., Fischer, S.A., and Green, A., *Propionibacterium acnes colonization of the human shoulder.* J Shoulder Elbow Surg, 2009. **18**(6): p. 897-902.
47. Sanz, Y., Olivares, M., Moya-Perez, A., and Agostoni, C., *Understanding the role of gut microbiome in metabolic disease risk.* Pediatr Res, 2015. **77**(1-2): p. 236-244.
48. Lam, Y.Y., Ha, C.W., Campbell, C.R., Mitchell, A.J., Dinudom, A., Oscarsson, J., Cook, D.I., Hunt, N.H., Caterson, I.D., Holmes, A.J., and Storlien, L.H., *Increased gut permeability and microbiota change associate with mesenteric fat inflammation*

and metabolic dysfunction in diet-induced obese mice. PLoS One, 2012. **7**(3): p. e34233.

49. Ohlsson, C. and Sjogren, K., *Effects of the gut microbiota on bone mass.* Trends Endocrinol Metab, 2015. **26**(2): p. 69-74.

50. McCabe, L.R., Irwin, R., Schaefer, L., and Britton, R.A., *Probiotic use decreases intestinal inflammation and increases bone density in healthy male but not female mice.* J Cell Physiol, 2013. **228**(8): p. 1793-1798.

51. Sjogren, K., Engdahl, C., Henning, P., Lerner, U.H., Tremaroli, V., Lagerquist, M.K., Backhed, F., and Ohlsson, C., *The gut microbiota regulates bone mass in mice.* J Bone Miner Res, 2012. **27**(6): p. 1357-1367.

52. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I., *An obesity-associated gut microbiome with increased capacity for energy harvest.* Nature, 2006. **444**(7122): p. 1027-1031.

53. Hsu, Y.J., Chiu, C.C., Li, Y.P., Huang, W.C., Huang, Y.T., Huang, C.C., and Chuang, H.L., *Effect of intestinal microbiota on exercise performance in mice.* J Strength Cond Res, 2015. **29**(2): p. 552-558.

54. Scheiman, J., Luber, J.M., Chavkin, T.A., MacDonald, T., Tung, A., Pham, L.D., Wibowo, M.C., Wurth, R.C., Punthambaker, S., Tierney, B.T., Yang, Z., Hattab, M.W., Avila-Pacheco, J., Clish, C.B., Lessard, S., Church, G.M., and Kostic, A.D., *Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism.* Nat Med, 2019. **25**(7): p. 1104-1109.

Chapter 3

Instances of Altered Gut Microbiomes Among Irish Cricketers Over Periods of Travel in the Lead up to the 2016 World Cup: A Sequencing Analysis

Included as published in *Travel Medicine and Infectious Disease*

(doi: https://doi.org/10.1016/j.tmaid.2020.101553)

Authors: Ciara M. O' Donovan, Brendan Connor, Sharon M. Madigan, Paul D. Cotter, and Orla O' Sullivan

Contributions:

- **Candidate** extracted DNA from faecal samples, prepared samples for 16S and shotgun metagenomic sequencing, and completed bioinformatics and statistical analysis.
- **BC** and **SMM** were involved with project conceptualization, sample and metadata collection.
- **PDC** and **OOS** supervised the study.

3.0 Abstract

3.0.1 Background

Changes and stresses experienced during travel have the potential to impact the gut microbiome, with travel implicated in the spread of antibiotic resistance genes across continents. The possibility of gut microbiome-mediated negative impacts arising from travel, and associated consequences for peak performance, would be of particular concern for elite athletes.

3.0.2 Methods

Faecal samples were collected from male (N=14) and female (N=7) cricket players during the build-up to the 2016 Cricket World Cup. Baseline and post-travel samples were collected from all participants and subjected to 16S rRNA amplicon sequencing. Samples from a subset of participants (N=4) were analysed by shotgun metagenomic sequencing.

3.0.3 Results

Analysis revealed a single travel time point as having the potential to have an impact on the gut microbiome. Reductions in alpha diversity following travel were observed, accompanied by shifts in the taxonomic profile of the gut microbiome. Antibiotic resistance and virulence genes were also identified as undergoing changes following travel.

3.0.4 Conclusions

This study reveals that periods of travel, in particular following gastrointestinal distress, may result in gut microbiome disruption. While this analysis was completed in athletes, the findings are applicable to all travelling individuals and considerations should be made surrounding travel in an attempt to reduce these changes.

3.1 Highlights

- This analysis investigates the Irish cricket teams during travel, highlighting potential considerations which should be undertaken by travelling athletes.
- Fluctuations in microbial stability occur with travel, with those who experience gastrointestinal distress potentially being more susceptible to stability fluctuations.
- Antibiotic resistance acquisition travel was identified as potentially associated with both pre-travel stresses (i.e., gastrointestinal distress) and travel destination.

3.2 Introduction

Many aspects of a modern lifestyle including diet, pharmacological agents, exercise, and disease state influence the gut microbiome, which in turn can impact significantly on host health [1-11]. It is hoped that in gaining a better understanding of factors that modulate the gut microbiome, it will be possible to more successfully mitigate undesirable changes and enhance health.

Recently, there has been an increased focus on the gut microbiome of athletes and its contribution to athlete health and performance [4, 5, 12]. Many athletes travel extensively to train and compete, facing changes in time zone, stress, diet, and gastrointestinal (GI) disturbances, which may in turn result in gut microbiome alterations [13-17]. The 2020 Olympics in Japan is an upcoming occasion of travel for a range of athletes worldwide. The consideration of the potential implications of travel is important for travelling athletes and their coaches [18].

Gut microbiome composition prior to travel may also be important in relation to the impact of travel on antibiotic resistance gene acquisition during travel [19, 20]. Some gut microbiota-associated risks that are particularly enhanced as a consequence of travel include an increased exposure to antibiotic resistance [21]. Additionally, traveller's diarrhoea (TD), a disorder commonly experienced by individuals who travel, in particular to Asia, can result in gut microbiota alterations [22, 23].

Anecdotal evidence from Cricket Ireland suggested travel was having a major impact on GI wellbeing, which subsequently impacted performance. This study set out to characterise the gut microbiota of the Cricket Ireland teams during the build-up to the Cricket World Cup of 2016 and, more specifically, to determine if travel was associated with gut microbiota fluctuations. During this period the male team travelled to multiple destinations, i.e., Zambia/Namibia, Australia, United Arab Emirates (UAE), and India, while

the female team travelled to India. Faecal samples were collected before and after each journey and subjected to 16S rRNA gene amplicon sequencing. Additionally, samples from a subset of participants underwent shotgun metagenomic sequencing to allow for greater exploration of compositional, as well as functional, changes. While the approach taken does not offer the possibility of discerning between the relative impacts of the many different potential modulating factors involved in long distance travel, it provides an insight into the net impact of such travel on the athlete gut microbiome.

3.3 Methods

3.3.1 Compliance with ethical standards

Ethical approval for the study was granted by the clinical research ethics committee of the Cork teaching hospitals. All procedures performed in studies involving human participants were in accordance with the ethical standards of the research committee, and with the 1964 Helsinki declaration, and its later amendments or comparable ethical standards. All subjects gave written informed consent prior to the beginning of the study (supplementary document 3.1). COD, SM, BC, PC, and OOS declare that they have no conflict of interest.

3.3.2 Experimental design

Faecal samples were collected from male (N=14) and female (N=7) cricket players during the build-up to the 2016 Cricket World Cup using Omnigene-Gut collection kits (DNA Genotek Inc., Ottawa, Canada) to facilitate the stabilisation of the faecal microbiota of samples at room temperature [24,25]. Baseline faecal samples were collected from female participants before travel (Time point LA; N=7) and following a 30-day tour to India (Time point LB; N=7). For males, baseline faecal samples were also collected before travel commenced (Time point A; N=9) and again following travel to Zambia/Namibia (Time point B; N=11), an extended stay back in Ireland (Time point C; N=12), and following subsequent travel to Australia (Time point D; N=12), UAE (Time point E; N=9), and India (Time point F; N=11) (supplementary figure 3.1). Male players stayed at the various destinations, i.e., Zambia/Namibia, Australia, UAE, and India, for 23, 30, 8, and 18 days, respectively. Baseline samples were collected from all individuals 0 to 14 days before travel (average = 3 days). After travel, all samples were collected within 0 to 20 days following return (average = 4 days). The mean age of the female and male groups was 27 (+/- 4) and 28 (+/- 4) years, respectively. Gut health questionnaires (supplementary document 3.2) were completed at each travel time point for both male and female participants. Based on individual and

group scores for each question, individuals were designated as being with or without GI distress.

3.3.3 DNA extraction and library preparation

DNA extraction was completed using the PowerFecal DNA Isolation kit (MoBio laboratories inc., Carlsbad, California, USA) using the altered manufacturer's instructions for use with the Omnigene-Gut collection kits. In short, 250 µL of the sample suspended in the Omnigene-Gut stabilization liquid was homogenised in a bead beating tube containing garnet beads and bead solution. Inhibitors of PCR analysis were removed, and total genomic DNA was captured on a silica spin column. DNA was washed and eluted. Extracted DNA was stored at -20°C until library preparation was completed. One sample (B009) was excluded as this was the only sample collected from this individual.

All samples were prepared for 16S rRNA sequencing on the Illumina Miseq system according to the 16S library preparation workflow. In brief, the V3 and V4 region of the 16S bacterial rRNA gene was amplified using 16S primers (Sigma Aldrich Ireland ltd., Wicklow, Ireland): Forward 5'TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG and Reverse 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C which created amplicons of approximately 460 bp. The resulting PCR products were cleaned up using Ampure XP beads (Labplan ltd., Co. Kildare, Ireland). Dual indices and Illumina sequencing adapters (Nextera XT Index Kit, Illumina, Inc., San Diego, California, USA) were attached to the amplicons. Following another cleaning step, amplicons were quantified, normalized, and pooled. The resulting pool was once again cleaned before being sequenced at the Teagasc sequencing facility.

A subset of samples, from the 4 male participants who provided samples at all 6 time points (players 1, 2, 3, and 11), were also prepared for shotgun metagenomic sequencing on the Illumina NextSeq system as per modified Nextera XT DNA library preparation

protocol (Illumina). Briefly, 0.2 ng μL^{-1} of sample DNA was tagmented for 7.5 min at 55°C with adapter sequences. Tagmented DNA was amplified and index adapters (i5 and i7) added (Nextera XT Index Kit, Illumina). Agencourt AMPure beads (Labplan ltd.) were used to clean up prepared library DNA. DNA samples were diluted to equimolar concentrations before pooling. Prepared library was sequenced on the NextSeq platform using standard Illumina protocols, at the Teagasc sequencing facility. The datasets generated and analysed during the current study are available in the European Nucleotide Archive under the study accession number PRJEB28338.

3.3.4 Bioinformatic analysis

Resulting FastQ forward and reverse reads from amplicon sequencing were joined using FLASH [26]. A second sample (D004) was excluded based on a low number of reads per sample, most likely as sample was seen to be not suspended in stabilising liquid following collection. Sequences were clustered into Operational Taxonomic Units (OTU's) and chimeras removed using USEARCH (Version 7.0) [27]. Taxonomy was assigned to clustered OTU tables using a blast against Silva123 [28].

Resulting FastQ reads from shotgun metagenomic sequencing were quality checked by first removing contaminating human derived sequences using the NCBI Best Match Tagger (BMTagger). Resulting reads were trimmed and poor quality and duplicate reads were removed using a combination of Picard and SAM tools. Taxonomic classifications of trimmed reads were determined using Kaiju [29]. Strain level profiling was completed using StrainEst [30]. Functional profiling was completed using the Human Microbiome Project Unified Metabolic Analysis Network 2 (HUMAnN2; http://huttenhower.sph.harvard.edu/humann2) [31]. Antibiotic resistance associated genes were determined using the MEGARes pipeline [32]. Reads were also analysed against the 2017 virulence factors marker collection using ShortBRED [33].

Analysis of resulting taxonomic and functional tables was completed in R (R Foundation for Statistical Computing, Vienna, 2012) (version 3.4.1). Alpha diversities were generated in R using the Phyloseq package for amplicon sequencing data [34]. An UPGMA representation of Bray-Curtis dissimilarities was made using QIIME and visualised using iToL for amplicon sequencing data [35, 36]. Bray-curtis distance matrices were computed using the R package vegan for shotgun metagenomic data [37]. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. All p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05.LEfSe was used to determine taxa which characterise specific groups [38]. The significance threshold was set at 0.05 and an LDA effect size threshold of 2 was used for discriminative features. Binary Jaccard distances were calculated using the R package proxy [39]. The Jaccard similarity index is the proportion of intersection between samples divided by the proportion of their union [40]. Therefore, the closer to 1 the binary Jaccard distance is the more distinct samples are, while closer to 0 indicates similarity between samples. Spearman correlations were calculated in R using the RcmdrMisc package (version 1.0-6) and corrected for multiple comparisons using Holms method.

3.4 Results

3.4.1 Timing of gastrointestinal distress symptoms may influence stability of the gut microbiome

Within sample (alpha) diversity was not found to significantly vary between baseline and final (i.e., post-India) time points (figure 3.1 A and B). Subtle shifts were seen in individuals over time with a mean decrease observed, though not significant (p=0.1 - 0.64). Figures 3.1 C (Shannon) and D (Simpson) show shifts in alpha diversity measures across all time points for individuals where shotgun data was available. Fluctuations in alpha diversity measures occurred in all cases, with larger shifts observed within some individuals.

Between group (beta) diversity measured by UPGMA (an unweighted hierarchical cluster method) analysis (amplicon data; supplementary figure 3.2 C and D), and MDS analysis (shotgun data; supplementary figure 3.2 A and B), revealed that samples cluster by individual regardless of when they were collected. However, some exceptions were observed in both datasets, with one example being the samples collected from player 11. More specifically, the composition, of both amplicon and shotgun metagenomic sequencing, and functional potential from player 11 corresponding to the post-India sample diverged from that of other samples collected from the same individual.

To assess stability of the microbiome binary Jaccard distances were computed for each pair of samples from the respective male participants to understand subtle changes to community composition (figure 3.2). From the shotgun dataset, at species level, samples from players 2 and 11 can be seen to be variable. Players 2 and 11 reported GI distress symptoms at 3 and 2 time points, respectively. Both of these individuals reported GI distress symptoms at time point E, leading us to postulate that the timing of GI distress, i.e. before travel to India, could result in instability in the gut microbiome composition.

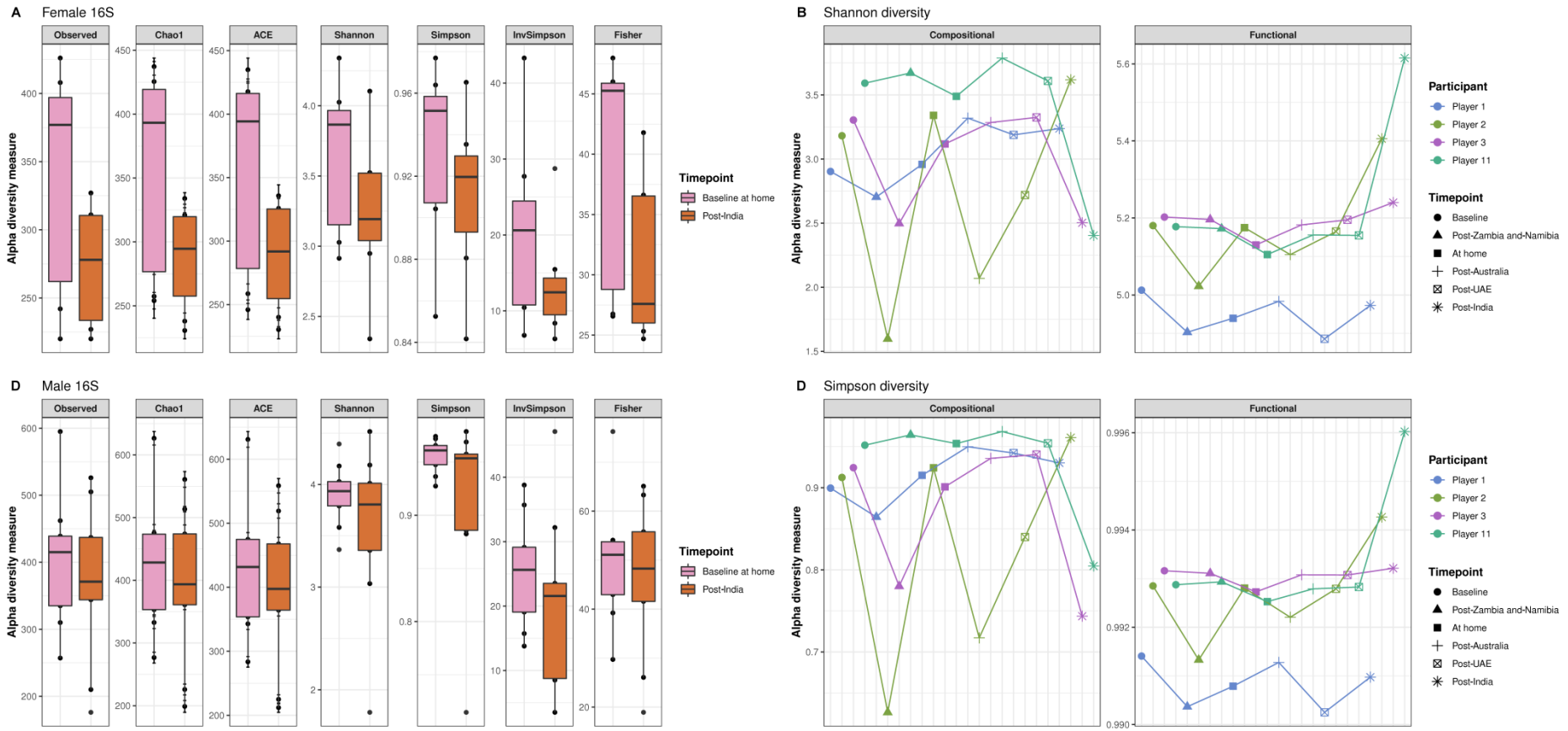


Figure 3.1| Alpha diversity fluctuations occur during travel

Figure 3.1 continued

Alpha diversity measures for (A) female (N=7) and (B) male (N=9) groups at baseline at home (pink) and post-India (orange) time points, from amplicon datasets. (C) Shannon and (D) Simpson alpha diversity measures over time for compositional and functional data for samples sequenced by shotgun metagenomic sequencing (N=4).

3.4.2 Individual taxa are unique between baseline and post-India time points

While individual specific responses to travel could be observed from the amplicon dataset (supplementary figure 3.3), 13 taxa were identified as discriminatory for the baseline (7 taxa) or post-India (6 taxa) time points. Differences based on travel time points were only identified in the amplicon dataset, which may be as a consequence of the small number of samples (N=4) at each of the time points in the shotgun metagenomic sequencing dataset. Discriminatory taxa were only established for the baseline and post-India time points, indicating that perhaps while fluctuations in individual taxa happen during other time points (i.e. post-Zambia and -Namibia, at home, post-Australia, or post-UAE time points) these two time points are the most distinct in terms of composition. This finding may also have been as a consequence of the comparably lower number of samples available for these time points as only male participants travelled to these destinations. *Staphylococcus*, Staphylococcaceae, *Ruminiclostridium* 5, Incertae Sedis Lachnospiraceae, *Erysipelotrichaceae* UCG003, *Ruminococcus* 1, and *Dorea* were found to be associated with baseline samples, while *Enterococcus*, Enterococcaceae, Actinobacteria, Nocardiaceae, Corynebacteriales, and *Rhodococcus* were found to be associated with post-India samples.

Shotgun data allowed a more in-depth analysis of the taxonomic profiles in a subset of samples, with a total of 126 species being identified across these 24 samples (figure 3.4 A). Taxonomic profiles based on shotgun metagenomic sequencing data were found to be individual specific. Of particular interest, *Escherichia coli* was identified in samples F011 and F002 (i.e. post-India samples from players 2 and 11). These samples were of particular interest as they did not cluster with other samples from those individuals on an MDS plot (supplementary figure 3.2 A). To further analyse this result strain level analysis was completed for *E. coli*. Strain level profiles (figure 3.4 B) reveal that while the *E. coli* population (2.4% of overall relative abundance) of sample F002 (from male participant 2) was dominated by *E. coli* DEC15D (88%), the *E. coli* profile of sample F011 (41.5% overall

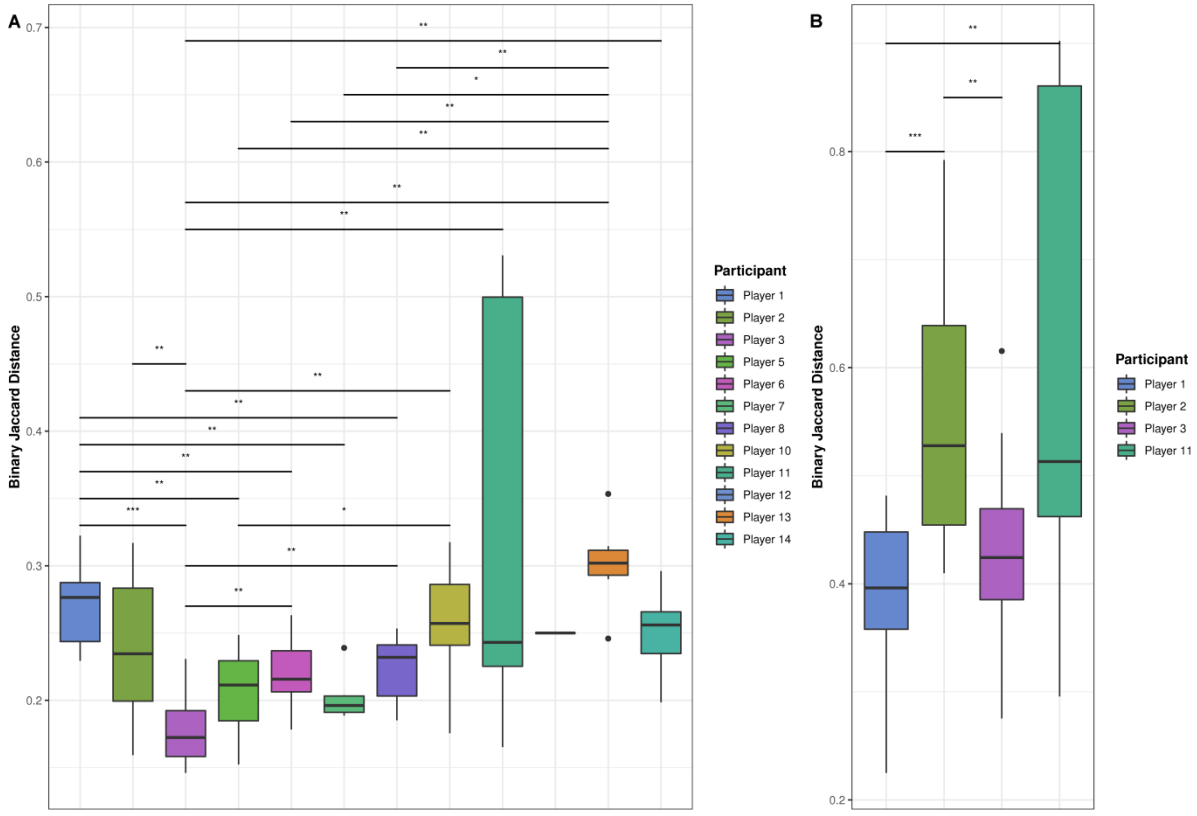


Figure 3.2| Individual variability in microbiome stability

Binary Jaccard distances were calculated between time points available for each individual. Boxplots show distances for each individual in (A) the genera classified in the amplicon sequencing dataset and (B) the species classified in the shotgun metagenomic sequencing dataset. Binary Jaccard distances between individuals were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the players between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. * P<0.05; ** P<0.01; *** P≤0.001.

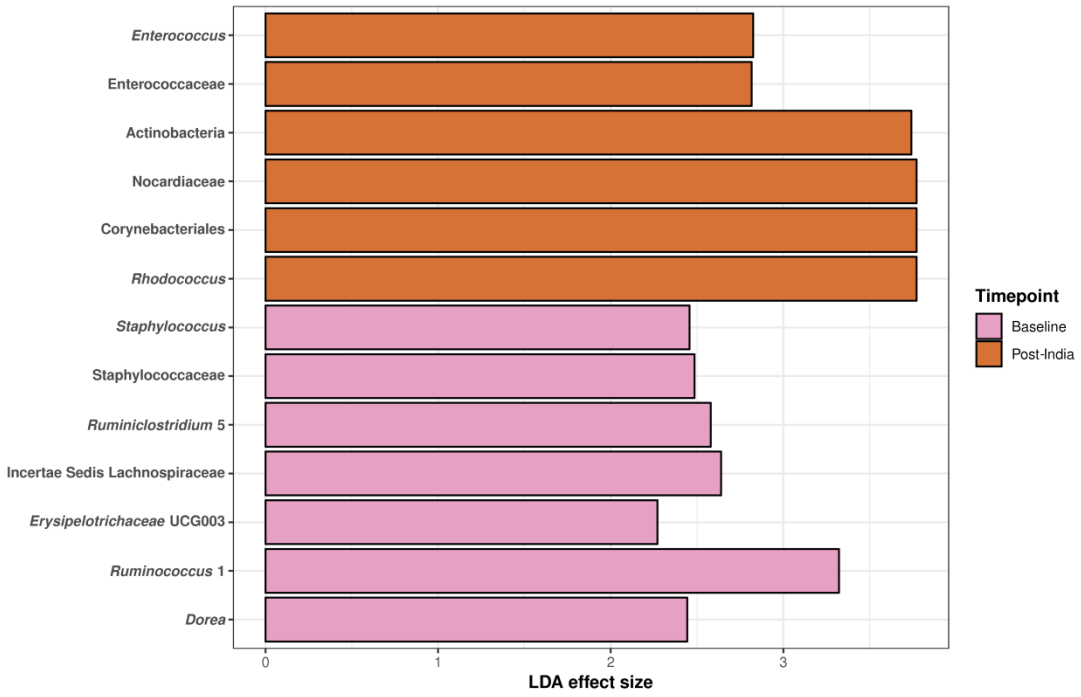


Figure 3.3| Significant differences in composition based on time point

LEfSe plots identifying discriminatory taxa between baseline (N=16) and post-India (N=16) time points where both female and male participants are grouped from amplicon sequencing data. An LDA effect size cut off of 2 was used.

relative abundance) was dominated by *E. coli* B7A (84%), with only one strain (p0305293.7) shared between these two samples, indicating that while the time point of acquisition may have been shared, the source may have varied between these two individuals.

3.4.3 Individual taxa are associated with gastrointestinal distress symptoms

LEfSe [37] revealed 20 taxa which were associated with GI distress (figure 3.5), 7 of these were identified from amplicon analysis (A), while 13 were identified from shotgun metagenomic sequencing analysis (B). *Ruminiclostridium* 9, Enterococcaceae, *Enterococcus*, Lentisphaerae, Victivallaceae, Victivallales, and *Victivallis* were found to be associated with reported GI distress from the amplicon dataset. Interestingly crossover was not observed between those taxa identified as being associated with GI distress from the shotgun metagenomic sequencing dataset; this may be as a result of the variance in the databases used in each of these analyses. Ascomycota, *Clostridium* sp. CAG 245-30-32, Armatimonadetes, *Armatimonadetes* bacterium Cent15-Ar3, *Clostridium* sp. CAG 440, *Klebsiella pneumoniae*, *Klebsiella*, Flavobacteriales, Flavobacteriia, and 4 species of *Prevotella* (namely *P. copri* CAG 164, *P.* sp. CAG 732, *P.* sp. 885, and *P.* sp. Marseille P4119) were identified as being associated with reported GI distress in the shotgun metagenomic sequencing dataset.

3.4.4 Antibiotic resistance potential is altered as a consequence of travel to India

Overall alterations in gut microbiome functional potential were observed with travel. Further investigations analysed antibiotic resistance and virulence genes, as these are of particular relevance to travel medicine. . Players 2 and 11 showed an increase in the number of antibiotic resistance genes identified at the post-India time point (figure 3.6: F002 and F011). Prior to this time point antibiotic resistance remained individual specific across both of these participants. Post-India, resistance potential to 8 antibiotic

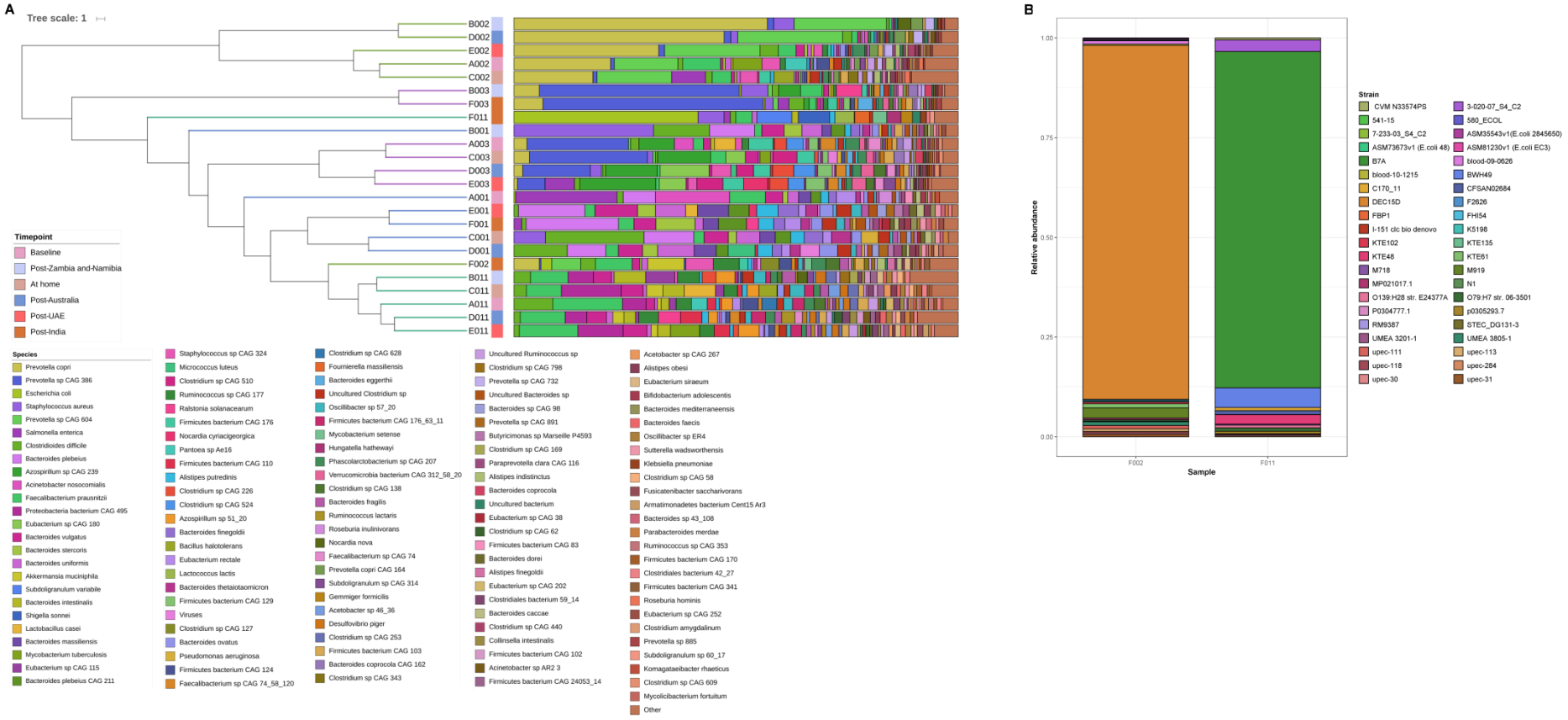


Figure 3.4| Shotgun metagenomic sequencing reveals unique *Escherichia coli* strains between participants post-India

Figure 3.4 continued

(A) Hierarchical cluster analysis, with species level taxonomic profiles of individual players over time for whom shotgun metagenomic sequencing data was available (N=4). Individual bars are coloured by participant, while colours of stacked bar alongside tree represent time point at which sample was collected. Individual numbers relate to one player and letters relate to time point (A: male baseline at home, B: male post-Zambia and -Namibia, C: male at home, D: male post-Australia, E: male post-UAE, F: male post-India (see supplementary figure 3.1 for timeline)). Each bar represents an individual sample. Each colour represents an individual species with only those species which were present at >1% relative abundance shown, with all others grouped as other. (B) Strain level profiles of *Escherichia coli* in participants 2 and 11 at the post-India time point.

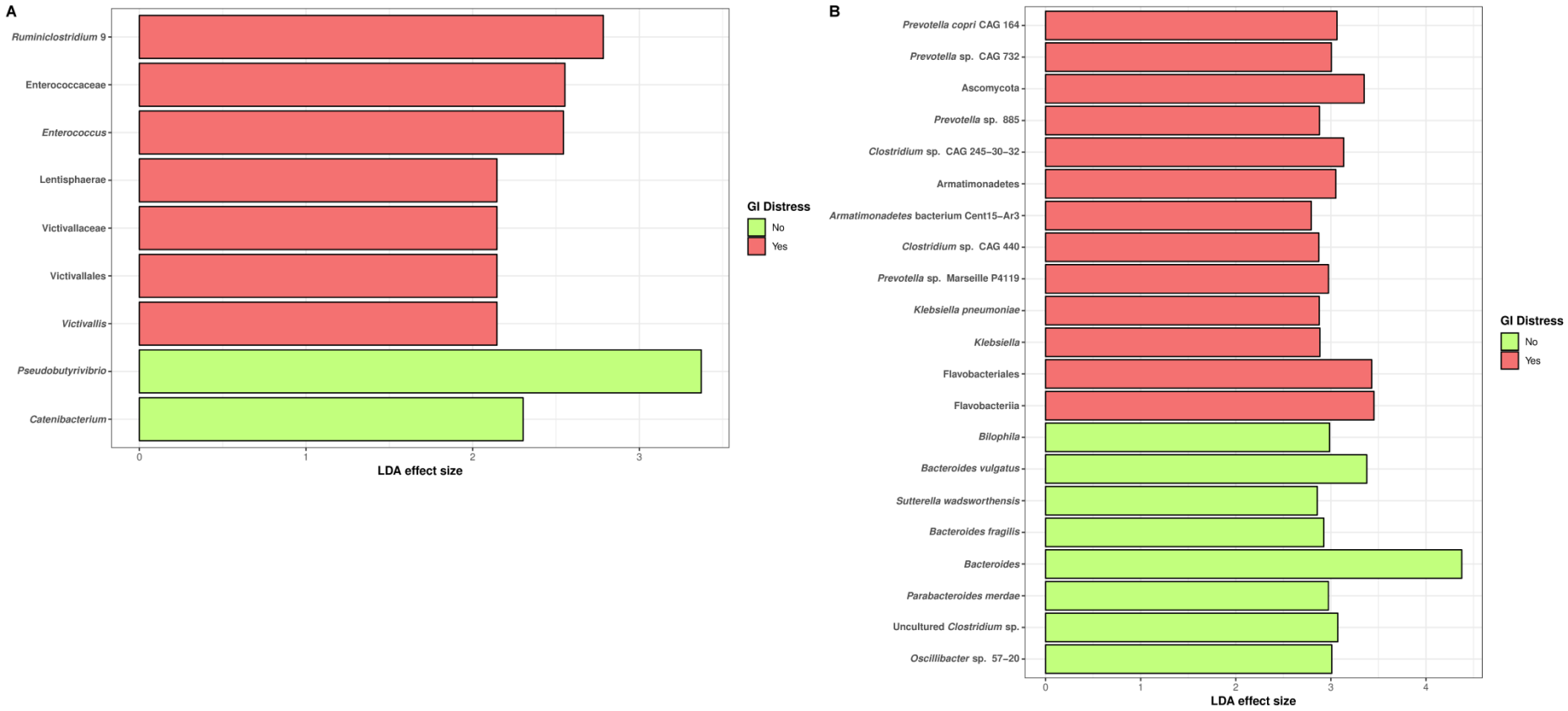


Figure 3.5| Significant differences in composition based on reported gastrointestinal distress symptoms

Figure 3.5 continued

LEfSe plots identifying discriminatory taxa from (A) amplicon analysis and (B) shotgun metagenomic sequencing, between those who reported (A: N=10; B: N=7) and did not report gastrointestinal distress symptoms (A: N=22; B: N=17). Taxonomy was assigned to the 16S dataset using Silva123 and to the shotgun metagenomic sequencing dataset using Kaiju. An LDA effect size cut off of 2 was used.

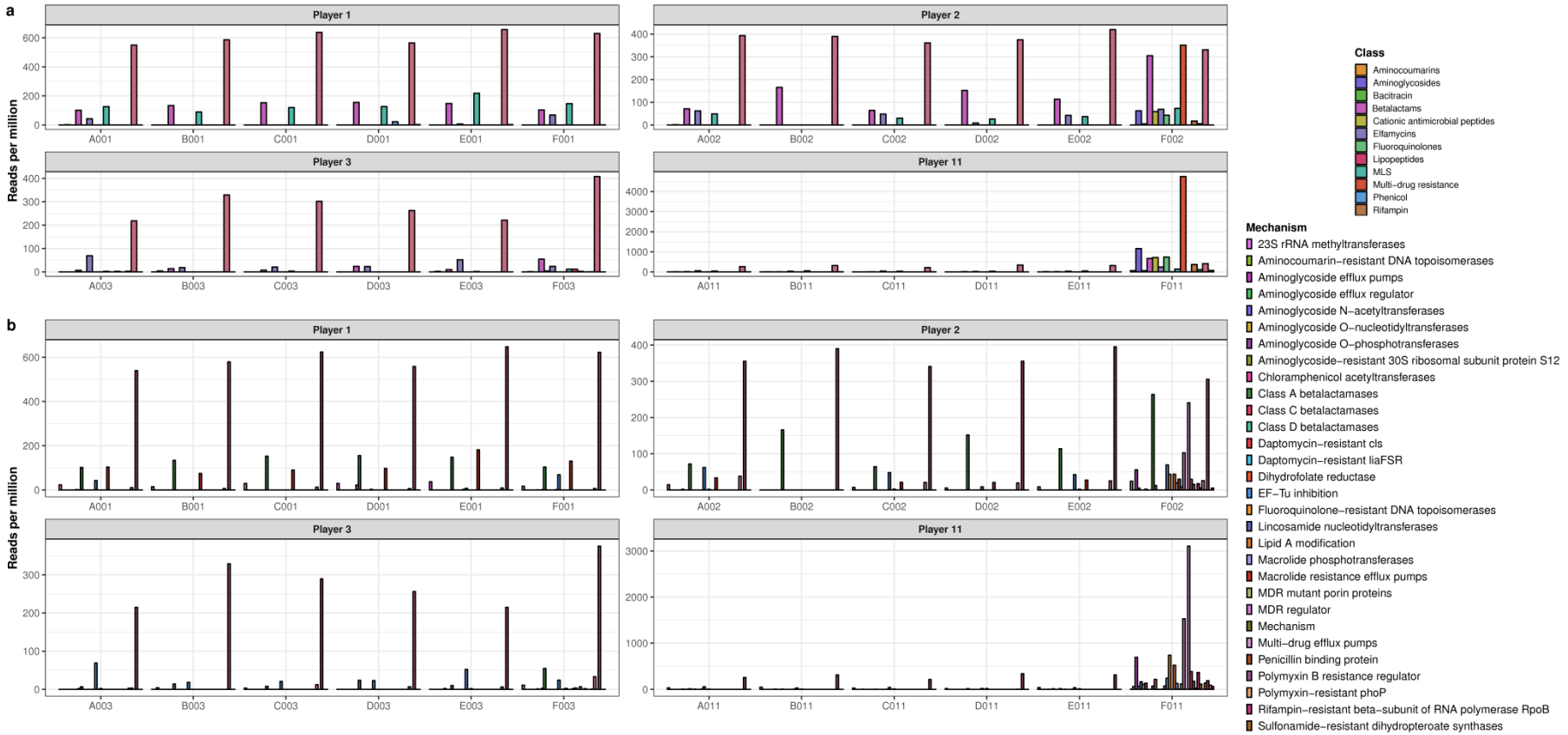


Figure 3.6| An increase in antibiotic resistance potential is seen in some individuals post-India

Antibiotic resistance potential over time, characterised using resistome analyser. (A) Resistance to antibiotic classes by participant. (B) Mechanisms of action of antibiotic resistance potential by participant.

classes was seen to increase in both of these participants, while in addition resistance potential to a further 4 antibiotic classes was seen to increase in player 11 at this time point. Notably, players 2 and 11 reported GI distress symptoms at the post-UAE time point (E). While GI distress symptoms were reported at other time points (player 2 at time points B and D; player 3 at time points D and F; and player 11 at time point E), no increase in antibiotic resistance potential was observed following these events. It is possible that the period of GI distress symptoms combined with other factors encountered during travel to India contributed to the acquisition of antibiotic resistance genes. It also is important to note that player 11 had received antibiotic treatment in the three months prior to the collection of sample F. However, no antibiotic usage was reported for player 2. A corresponding investigation of virulence marker genes also identified changes here, in particular in post-India samples (supplementary figure 3.4).

3.5 Discussion

Individuals, who travel on a regular basis, including athletes, are susceptible to stress, time zone changes, alterations in diet, and symptoms such as travellers' diarrhoea [17, 41, 42]. While the impacts of these factors on performance have been variable, associations between international travel with injury potential and reduced performance have been identified [43-47]. Identification of undesirable changes that occur within the gut microbiota during travel might result in the development of therapeutics to maintain health. To this end, this pilot study of 21 individuals was carried out and a number of notable changes in microbiome composition, diversity, and functional potential were observed.

The main finding of this research was the establishment that a specific event, in this case travel to India, can have a significant impact on gut microbiome composition and functional potential, with associated increases in antibiotic resistance and virulence genes. Other destinations were not identified in this study as having a significant impact on the gut microbiome; however, this finding may be as a result of a larger number of samples available for baseline and post-India time points (i.e., samples from both female and male groups), in comparison to other time points.

Relative stability in the human gut microbiome is expected over time, with perturbations occurring for example with disease states and antibiotic use. However, the microbiome is also resilient allowing for the return to a stable composition in most situations [48]. Here, for the majority, microbial stability is observed, however in certain individuals instability of the gut microbiome was observed. This is consistent with very many studies that have established the stability and resilience of the gut microbiota over time [49-51]. In order to establish an understanding of more subtle, individual shifts in the microbiome over time, binary Jaccard distances (used to estimate stability) were computed. Timing of reported GI

distress symptoms could potentially be an important factor in any instability noted in the gut microbiome, with symptoms reported prior to travel to India potentially related to instability post-India. As such a small number of participants fit into these criteria (2 participants), it is possible that any other number of factors is influencing these results. Other influencing factors which cannot be accounted for here include individual variability and dietary impacts. Potential alterations in stability may be associated with the changes observed in taxa and antibiotic resistance profiles, with gut microbiome instability previously associated with diarrhoea, IBS symptoms, and *Clostridium difficle* associated disease [48].

Although travel was found to only induce minimal shifts in the gut microbiome overall, taxa were identified which were significantly altered at baseline and post-India time points, while another group of taxa were found to be significantly altered with reported GI distress. Actinobacteria, found in this study to be associated with the post-India time point, has previously been identified as increased in IBS [52], potentially indicating this change could put these individuals at risk of post-infectious IBS [53]. Of interest, *Ruminococcus* was identified as being associated previously with the ability to clear antibiotic resistance post-travel, while in this dataset it was associated with baseline samples, a potentially beneficial finding [19]. Such a relationship could not be identified in this dataset as it was not possible to collect samples following an extended period at home post-India.

Noteworthy, from the shotgun dataset, was an increase in *E.* coli identified in two participants following travel to India, with these individuals reporting GI distress at the pre-India time point (i.e. post-UAE). *Escherichia* species are among those most frequently identified as causative for TD and so investigation of this was of particular interest in this cohort [23]. Strain level differences in *E. coli* were established between players 2 and 11, with DEC15D and B7A found to be dominating members, respectively. *E. coli* DEC15D was

sequenced as part of a collection of diarrheagenic *E. coli* strains, which are a major cause of diarrhoea, while *E. coli* B7A is from the B1 subgroup which contains many commensals but also a number of pathogens [54-56]. This indicated that it is possible that these strains may have been related to GI distress symptoms experienced by these individuals.

Klebsiella was identified in the shotgun metagenomic sequencing dataset as being associated with reported GI distress, which has previously been identified in TD [23]. *Prevotella* species were identified as being associated with reported GI distress, with *Prevotella* previously associated with diarrhoea predominant IBS, TD, and more abundant with loose stool [19, 57, 58]. Of interest, *P. copri*, one of the species identified in this dataset as associated with GI distress, has previously been identified as more abundant in those individuals who experienced TD even before travel [19]. This pattern was evident to some extent in our study (i.e. player 2) but would require a shotgun metagenomic investigation of a larger cohort to establish its significance.

Acquisition of antibiotic resistance potential and virulence marker genes were explored in the shotgun data, as previous studies have outlined the potential of acquiring antibiotic resistance during travel, while the presence of virulence marker genes could reveal regions where acquisition of potential pathogens is particularly increased [21, 22, 51, 59]. Acquisition of antibiotic resistance related genes seemed to be individual specific, with players 2 and 11 acquiring resistance to a number of new antibiotic classes and also an array of mechanisms of antibiotic resistance activity. Previous studies into the acquisition of antibiotic resistance during travel have mainly focused on extended spectrum β-lactamases (ESBLs) producers, with an increase seen post-travel across a number of studies [60-64]. Individuals have been found to be at increased risk of acquiring antibiotic resistance genes if travel destination is in Asia (particularly India), antibiotics are used during travel, TD persists after return from travel, and/or as a consequence of consuming

certain foods during travel such as ice-creams and pastry [61-63]. Travel to India, GI distress, and antibiotic usage could have been contributing factors in the acquisition of antibiotic resistance potential in this cohort. Antibiotic resistance is a major global health problem and the acquisition of antibiotic resistance genes during travel could impact on global carriage of antibiotic resistance genes [19]. Carriage of antibiotic resistance genes has the potential to persist for years and have the potential to transfer to pathogenic bacteria, increasing the risk of acquiring and harbouring an antibiotic resistant infection [65].

Analysis of virulence marker genes revealed an increase in the alpha diversity of these marker genes in players 2 and 11 post-India, with a notable shift in beta diversity also observed for these samples. Interestingly, in this subset of individuals travel to India resulted in an increase in nine, individual virulence marker genes. Although the presence of these genes doesn't necessarily represent functionality or pathogenicity, increases in proportions of virulence and antibiotic resistance genes following travel is a concern and merits further investigation in the future. Presence of virulence factors in the GI tract could result in intestinal infections including diarrhoea [56], and thus presence of these virulence genes could be related to reported GI distress.

While this study only includes a small number of participants, owing to the number of travelling members of the Cricket Ireland teams available and willingness to participate, it further explores the influence of travel on the gut microbiome. While in this study post-Zambia and -Namibia, at home, post-Australia, or post-UAE time points were not found to be associated with major alterations this could be as a result of the numbers available and so further studies with a larger number to these destinations may offer differing results. This study identifies travel to India, and in particular travel to India following GI distress symptoms, as associated with disturbances in the gut microbiome from both a

compositional and functional potential perspective. Previously, travel to India has been identified as having a major impact on the gut microbiome, as has GI distress during travel [61-63].

GI disturbances pre-travel to high risk destinations (e.g. India) could potentially place individuals at an increased risk of gut microbiome alterations, including the acquisition of antibiotic resistance genes, and so individuals falling into this category should in particular take care. Probiotics have been shown to be beneficial in the prevention of TD if taken 4 weeks prior to travel, however, as studies vary by strain(s) used it is not yet possible to recommend one overall option [47, 66]. However, none of the investigations to date have been completed in travelling athletes. Although not investigated in this study, diet changes during travel may impact on the gut microbiome. To this end, travelling individuals should take care to avoid those foods associated with the development of TD including but not limited to any raw and unpeeled produce, local water, uncooked meats, and unpasteurised dairy products [47, 67].

3.6 Conclusion

Here we have identified one travel period, travel to India, as having a significant impact on gut microbiome composition, as well as an alteration in the antibiotic resistance and virulence genes present. While individual fluctuations were identified, taxa associated both to pre- and -post-travel time points, as well as experienced GI distress, could be identified. Individual, pre-travel insults to the GI tract, such as reported distress or antibiotic usage, could potentially influence the impact of travel on the gut microbiome. These changes identified during travel are of particular relevance and interest to athletes travelling for competition, such as for the upcoming 2020 Olympics [18]. Consideration should be given to potential gut microbiome changes during travel to reduce potential risk factors and the potential negative impact of these identified changes. Ultimately, it is hoped that a greater

appreciation of the negative impacts of travel on the gut microbiome of athletes, and other travellers, can reveal interventions that can minimise undesirable health consequences.

3.7 Acknowledgements

The authors thank Fiona Crispie, Vicki Murray, and Laura Finnegan for their DNA sequencing. The authors are grateful to Cricket Ireland, Adam Sullivan, Aaron Hamilton, and all those who donated their time, effort, and samples to this study. This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273, 13/SIRG/2160, and 11/PI/1137.

3.8 References

1. Conlon, M.A. and Bird, A.R., *The impact of diet and lifestyle on gut microbiota and human health.* Nutrients, 2015. **7**: p. 17-44.
2. Korpela, K., Salonen, A., Virta, L.J., Kekkonen, R.A., Forslund, K., Bork, P., and de Vos, W.M., *Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children.* Nature Communications, 2016. **7**.
3. Han, N., Oh, M., Park, S.M., Kim, Y.J., Lee, E.J., Kim, T.K., Kim, T.N., Kwon, M.J., Kim, M.K., Lee, S.H., Rhee, B.D., and Park, J.H., *The effect of proton pump inhibitors on glycated hemoglobin levels in patients with type 2 diabetes mellitus.* Canadian Journal of Diabetes, 2015. **39**(1): p. 24-28.
4. 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., and Cotter, P.D., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.
5. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.
6. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).
7. Carroll, I.M., Ringel-Kulka, T., Siddle, J.P., and Ringel, Y., *Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome.* Journal of Neurogastroenterology and Motility, 2012. **24**(6).
8. Flemer, B., Lynch, D.B., Brown, J.M.R., Jeffery, I.B., Ryan, F.J., Claesson, M.J., O' Riordain, M., Shanahan, F., and O' Toole, P.W., *Tumour-associated and non-tumour-associated microbiota in colorectal cancer.* Gut, 2017. **66**: p. 633-643.
9. Fujimura, K.E. and Lynch, S.V., *Microbiota in allergy and asthma and the emerging relationship with the gut microbiome.* Cell Host Microbe, 2015. **17**(5): p. 592-602.
10. Desbonnet, L., Clarke, G., Traplin, A., O'Sullivan, O., Crispie, F., Moloney, R.D., Cotter, P.D., Dinan, T.G., and Cryan, J.F., *Gut microbiota depletion from early*

adolescence in mice: Implications for brain and behaviour. Brain, Behavior and Immunity, 2015. **48**: p. 165-173.

11. Sha, S., Xu, B., Wang, X., Zhang, Y., Wang, H., Kong, X., Zhu, H., and Wu, K., *The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease.* Diagnostic Microbiology and Infectious Disease, 2013. **75**(3): p. 245-251.
12. Petersen, L.M., Bautista, E.J., Nguyen, H., Hanson, B.M., Chen, L., Lek, S.H., Sodergren, E., and Weinstock, G.M., *Community characteristics of the gut microbiomes of competitive cyclists.* Microbiome, 2017. **5**(1): p. 98.
13. Dey, N., Wagner, V.E., Blanton, L.V., Cheng, J., Fontana, L., Haque, R., Ahmed, T., and Gordon, J.I., *Regulators of gut motility revealed by gnotobiotic model of diet-microbiome interactions related to travel.* Cell, 2015. **163**: p. 95-107.
14. Wiley, N.C., Dinan, T.G., Ross, R.P., Stanton, C., Clarke, G., and Cryan, J.F., *The microbiota-gut-brain axis as a key regulator of neural function and the stress response: Implications for human and animal health.* Journal of Animal Science 2017. **95**(7): p. 3225-3246.
15. Poroyko, V.A., Carreras, A., Khalyfa, A., Khalyfa, A.A., Leone, V., Peris, E., Almendros, I., Gileles-Hillel, A., Qiao, Z., Hubert, N., Farre, R., Chang, E.B., and Gozal, D., *Chronic sleep disruption alters gut microbiota, induces systemic and adipose tissue inflammation and insulin resistance in mice.* Scientific Reports, 2016. **6**: p. 35405.
16. Benedict, C., Vogel, H., Jonas, W., Woting, A., Blaut, M., Schurmann, A., and Cedernaes, J., *Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals.* Molecular Metabolism, 2016. **5**(12): p. 1175-1186.
17. Boggess, B.R., *Gastrointestinal infections in the traveling athlete.* Current Sports Medicine Reports, 2007. **6**(2): p. 125-129.
18. Nakamura, S., Wada, K., Yanagisawa, N., and Smith, D.R., Health risks and precautions for visitors to the Tokyo 2020 Olympic and Paralympic games. Travel medicine and infectious disease, 2018. **22**: p. 3-7.
19. Leo, S., Lazarevic, V., Gaia, N., Estellat, C., Girard, M., Matheron, S., Armand-Lefevre, L., Andremont, A., The VOYAG-R study group, Schrenzel, J., and Ruppe, E., *The intestinal microbiota predisposes to traveler's diarrhea and to the carriage of*

multidrug-resistant Enterobacteriaceae after traveling to tropical regions. Gut Microbes, 2019: p. 1-11.

20. Pires, J., Kraemer, J.G., Kuenzli, E., Kasraian, S., Tinguely, R., Hatz, C., Endimiani, A., and Hilty, M., *Gut microbiota dynamics in travelers returning from India colonized with extended-spectrum cephalosporin-resistant Enterobacteriaceae: A longitudinal study.* Travel Medicine and Infectious Disease, 2019. **27**: p. 72-80.
21. Johnning, A., Kristiansson, E., Angelin, M., Marathe, N., Shouche, Y.S., Anders, J., and Larsson, D.G.J., *Quinolone resistance mutations in the faecal microbiota of Swedish travellers to India.* BMC Microbiology, 2015. **15**.
22. Youmans, B.P., Ajami, N.J., Jiang, Z.-D., Campbell, F., Wadsworth, W.D., Petrosino, J.F., DuPont, H.L., and Highlander, S.K., *Characterization of the human gut microbiome during travelers' diarrhea.* Gut Microbes, 2015. **6**(2): p. 110-119.
23. Zhu, Q., Dupont, C.L., Jones, M.B., Pham, K.M., Jiang, Z.D., DuPont, H.L., and Highlander, S.K., *Visualization-assisted binning of metagenome assemblies reveals potential new pathogenic profiles in idiopathic travelers' diarrhea.* Microbiome, 2018. **6**(1): p. 201.
24. Hill, C.J., Brown, J.M.R., Lynch, D.B., Jeffery, I.B., Ryan, C.A., Ross, R.P., Stanton, C., and O' Toole, P.W., *Effect of room temperature transport vials on DNA quality and phylogenetic composition of faecal microbiota of elderly adults and infants.* Microbiome, 2016. **4**(19).
25. Choo, J.M., Leong, L.E., and Rogers, G.B., *Sample storage conditions significantly influence faecal microbiome profiles.* Scientific Reports, 2015. **5**: p. 16350.
26. Magoc, T. and Salzberg, S.L., *FLASH: Fast length adjustment of short reads to improve genome assemblies.* Bioinformatics, 2011. **27**(21): p. 2957-2963.
27. Edgar, R.C., *Search and clustering orders of magnitude faster than blast.* Bioinformatics, 2010. **26**(19): p. 2460-2461.
28. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O., *The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.* Nucleic Acids Research, 2013. **41**(D1): p. D590-S596.
29. Menzel, P., Ng, K.L., and Krogh, A., *Fast and sensitive taxonomic classification for metagenomics with Kaiju.* Nature Communications, 2016. **7**: p. 11257.
30. Albanese, D. and Donati, C., *Strain profiling and epidemiology of bacterial species from metagenomic sequencing.* Nature Communications, 2017. **8**(1): p. 2260.

31. Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., et al., Species-level functional profiling of metagenomes and metatranscriptomes. Nature Methods, 2018. **15**(11): p. 962-8.
32. Lakin, S.M., Dean, C., Noyes, N.R., Dettenwanger, A., Ross, A.S., Doster, E., Rovira, P., Abdo, Z., Jones, K.L., Ruiz, J., Belk, K.E., Morley, P.S., and Boucher, C., *MEGARes: An antimicrobial resistance database for high throughput sequencing.* Nucleic Acids Research, 2017. **45**(D1): p. D574-D580.
33. Kaminski, J., Gibson, M.K., Franzosa, E.A., Segata, N., Dantas, G., and Huttenhower, C., *High-specificity targeted functional profiling in microbial communities with ShortBRED.* PLoS Computational Biology, 2015. **11**(12): p. e1004557.
34. McMurdie, P.J. and Holmes, S., *Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data.* PLoS ONE, 2013. **8**(4).
35. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordan, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R., *QIIME allows analysis of high-throughput community sequencing data.* Nature Methods, 2010. **7**: p. 335-336.
36. Letunic, I. and Bork, P., *Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees.* Nucleic Acids Research, 2016. **44**: p. 242-245.
37. Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O' Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoeca, E., and Wagner, H., Vegan: *Community ecology package*. R package version 2.2-1.2.1-2, 2018.
38. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C., *Metagenomic biomarker discovery and explanation.* Genome Biology, 2011. **12**(6): p. R60.
39. Meyer, D. and Buchta, C., *Proxy*. R package version 0.4-23, 2018.
40. Prokopenko, D., Hecker, J., Silverman, E.K., Pagano, M., Nothen, M.M., Dina, C., Lange, C., and Fier, H.L., *Utilizing the Jaccard index to reveal population stratification in sequencing data: A simulation study and an application to the 1000 genomes project.* Bioinformatics, 2016. **32**(9): p. 1366-1372.

41. Schwellnus, M.P., Derman, W.E., Jordaan, E., Page, T., Lambert, M.I., Readhead, C., Roberts, C., Kohler, R., Collins, R., Kara, S., Morris, M.I., Strauss, O., and Webb, S., *Elite athletes travelling to international destinations >5 time zone differences from their home country have a 2-3-fold increased risk of illness.* British Journal of Sports Medicine, 2012. **46**(11): p. 816-821.
42. Jeukendrup, A.E., *Training the gut for athletes.* Sports Medicine, 2017. **47**(Suppl 1): p. 101-110.
43. McGuckin, T.A., Sinclair, W.H., Sealey, R.M., and Bowman, P., *The effects of air travel on performance measures of elite Australian rugby league players.* European Journal of Sport Science, 2014. **14 Suppl 1**: p. S116-122.
44. Huyghe, T., Scanlan, A.T., Dalbo, V.J., and Calleja-Gonzalez, J., *The negative influence of air travel on health and performance in the National Basketball Association: A narrative review.* Sports, 2018. **6**(3).
45. Bullock, N., Martin, D.T., Ross, A., Rosemond, D., and Marino, F.E., *Effect of long haul travel on maximal sprint performance and diurnal variations in elite skeleton athletes.* British Journal of Sports Medicine, 2007. **41**(9): p. 569-573
46. Chapman, D.W., Bullock, N., Ross, A., Rosemond, D., and Martin, D.T., *Detrimental effects of west to east transmeridian flight on jump performance.* European Journal of Applied Physiology, 2012. **112**(5): p. 1663-1669.
47. Reilly, T., Waterhouse, J., Burke, L.M., Alonso, J.M., and International Association of Athletics Federations, *Nutrition for travel.* Journal of Sports Sciences, 2007. **25 Suppl 1**: p. S125-134.
48. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-230.
49. David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., Erdman, S.E., and Alm, E.J., *Host lifestyle affects human microbiota on daily timescales.* Genome Biology and Evolution., 2014. **15**(7).
50. 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 Sindersen, 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., and O' Toole, P.W., *Composition, variability and temporal stability of the intestinal microbiota of the elderly.*

Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**: p. 4586-4591.

51. Leo, S., Lazarevic, V., Gaia, N., Estellat, C., Girard, M., Matheron, S., Armand-Lefévre, L., Andremont, A., The VOYAG-R study group, Schrenzel, J., and Ruppé, E., *The intestinal microbiota predisposes to traveller's diarrhoea and to the carriage of multidrug-resistant Enterobacteriaceae after travelling to tropical regions.* bioRxiv, 2018.
52. Tap, J., Derrien, M., Tornblom, H., Brazeilles, R., Cools-Portier, S., Dore, J., Storsrud, S., Le Neve, B., Ohman, L., and Simren, M., *Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome.* Gastroenterology, 2017. **152**(1): p. 111-123 e118.
53. Thabane, M. and Marshall, J.K., *Post-infectious irritable bowel syndrome.* World Journal of Gastroenterology, 2009. **15**(29): p. 3591-3596.
54. Hazen, T.H., Sahl, J.W., Redman, J.C., Morris, C.R., Daugherty, S.C., Chibucos, M.C., Sengamalay, N.A., Fraser-Liggett, C.M., Steinsland, H., Whittam, T.S., Whittam, B., Manning, S.D., and Rasko, D.A., *Draft genome sequences of the diarrheagenic Escherichia coli collection.* Journal of Bacteriology, 2012. **194**(11): p. 3026-3027.
55. Gallardo, P., Izquierdo, M., Vidal, R.M., Chamorro-Veloso, N., Rossello-Mora, R., O'Ryan, M., and Farfan, M.J., *Distinctive gut microbiota is associated with diarrheagenic Escherichia coli infections in Chilean children.* Frontiers in Cellular and Infection Microbiology, 2017. **7**: p. 424.
56. Crossman, L.C., Chaudhuri, R.R., Beatson, S.A., Wells, T.J., Desvaux, M., Cunningham, A.F., Petty, N.K., Mahon, V., Brinkley, C., Hobman, J.L., Savarino, S.J., Turner, S.M., Pallen, M.J., Penn, C.W., Parkhill, J., Turner, A.K., Johnson, T.J., Thomson, N.R., Smith, S.G., and Henderson, I.R., *A commensal gone bad: Complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407.* Journal of Bacteriology, 2010. **192**(21): p. 5822-5831.
57. Su, T., Liu, R., Lee, A., Long, Y., Du, L., Lai, S., Chen, X., Wang, L., Si, J., Owyang, C., and Chen, S., *Altered intestinal microbiota with increased abundance of Prevotella is associated with high risk of diarrhea-predominant irritable bowel syndrome.* Gastroenterology Research and Practice, 2018. **2018**: p. 6961783.
58. Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R.Y., Joossens, M., and Raes, J., *Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates.* Gut, 2016. **65**(1): p. 57-62.

59. Belda-Ferre, P., Cabrera-Rubio, R., Moya, A., and Mira, A., *Mining virulence genes using metagenomics.* PLoS One, 2011. **6**(10): p. e24975.

60. Kuenzli, E., *Antibiotic resistance and international travel: Causes and consequences.* Travel Medicine and Infectious Disease, 2016. **14**: p. 595-598.

61. Arcilla, M.S., van Hattem, J.M., Haverkate, M.R., Bootsma, M.C.J., van Genderen, P.J.J., Goorhuis, A., Grobusch, M.P., Lashof, A.M.O., Molhoek, N., Schultz, C., Stobberingh, E.E., Verbrugh, H.A., de Jong, M.D., Melles, D.C., and Penders, J., *Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): A prospective, multicentre cohort study.* The Lancet Infectious Diseases, 2017. **17**(1): p. 78-85.

62. Kuenzli, E., Jaeger, V.K., Frei, R., Neumayr, A., DeCrom, S., Haller, S., Blum, J., Widmer, A.F., Furrer, H., Battegay, M., Endimiani, A., and Hatz, C., *High colonization rates of extended-spectrum β-lactamase (ESBL)-producing Escherichia coli in Swiss travellers to South Asia- a prospective observational multicentre cohort study looking at epidemiology, microbiology and risk factors.* BMC Infectious Diseases, 2014. **14**: p. 528.

63. Tangden, T., Cars, O., Melhus, A., and Lowdin, E., *Foreign travel is a major risk factor for colonization with Escherichia coli producing CTX-M-type extended-spectrum β-lactamases: A prospective study with Swedish volunteers.* Antimicrobial Agents and Chemotherapy, 2010. **54**(9): p. 3564-3568.

64. Ruppe, E., Armand-Lefevre, L., Estellat, C., Consigny, P.H., El Mniai, A., Boussadia, Y., Goujon, C., Ralaimazava, P., Campa, P., Girard, P.M., Wyplosz, B., Vittecoq, D., Bouchaud, O., Le Loup, G., Pialoux, G., Perrier, M., Wieder, I., Moussa, N., Esposito-Farese, M., Hoffmann, I., Coignard, B., Lucet, J.C., Andremont, A., and Matheron, S., *High rate of acquisition but short duration of carriage of multidrug-resistant Enterobacteriaceae after travel to the tropics.* Clinical Infectious Diseases, 2015. **61**(4): p. 593-600.

65. Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K., *Long-term impacts of antibiotic exposure on the human intestinal microbiota.* Microbiology, 2010. **156**(11): p. 3216-3223.

66. McFarland, L.V. and Goh, S., Are probiotics and prebiotics effective in the prevention of travellers' diarrhea: A systematic review and meta-analysis. Travel medicine and infectious disease, 2019. 27: p. 11-9.

67. Riddle, M.S., Connor, B.A., Beeching, N.J., DuPont, H.L., Hamer, D.H., Kozarsky, P., Libman, M., Steffen, R., Taylor, D., Tribble, D.R., Vila, J., Zanger, P., and Ericsson, C.D., *Guidelines for the prevention and treatment of travelers' diarrhea: A graded expert panel report.* Journal of Travel Medicine, 2017. **24**(suppl_1): p. S57-S74.

3.9 Supplementary material

Supplementary document 3.1| Participant information and consent form

SUBJECT INFORMATION SHEET

Study Title: To define the composition of the gut microbiota in three groups of healthy volunteers and correlate it to body condition, diet, and microbial (short chain fatty acids) and host metabolism, and immune status (metabolic and inflammatory markers).

Protocol No.: APC025

Investigators: Prof. Fergus Shanahan (PI), Prof. Michael Molloy, Prof. Eamonn Quigley, Prof. David Kerins, Prof. Charlie Daly, Prof. Paul Ross, Dr. Paul Cotter & Dr. Paul O'Toole, in association with Dr. Conor McCarthy, Dr. Eanna Falvey & Ms. Ruth Wood-Martin (IRFU).

Site of Investigation: Department of Medicine, Cork University Hospital; Clinical Research Facility, Mercy University Hospital, Cork, University College Cork & Teagasc Food Research Centre, Moorepark, Fermoy, Cork.

Why is this study being run?

Recent research has demonstrated that there is a relationship between a person's health and the bacteria that grow in the bowel.

You are invited to participate in a research study where we aim to define the relationship between your body mass index and the variety of bacteria in your bowels (your microbiota). This could reveal new strategies to optimise metabolic health, immune status, and body conditioning. If you agree to participate in the study you may be asked questions about your symptoms and your health.

Study Procedure

You will be advised of the purpose of the study and the procedures which will be undertaken. You will be given a copy of the Subject Information Sheet, which will explain what is required from you. If interested, you will then be requested to read and sign the Informed Consent Form, and receive a signed copy.

You will be given a faecal sample collection kit and instructions on how to collect a sample.

You will be asked some general personal questions, as well as your medical history and current medical status. Details of current or past (within previous 6 months) medication (including probiotics, prebiotics, supplements, as well as over the counter medications) will be collected.

At a later date further faecal samples may be collected.

Where possible a dietary assessment will be performed.

Depending on preliminary results, we may ask you to provide a faecal sample in the future.

What happens if I start the study and change my mind later?

You do not have to take part in the study, participation is entirely voluntary. Refusal to participate, or discontinuing participation at any time, will involve no penalty, loss of benefits or denial of treatment or services by the Cork Teaching Hospitals or the participating doctor.

Will I experience any unpleasant side effects?

You will be requested to provide a sample of faeces.

Funding of trial

There are no cost implications for the Health Services Executive or to you. The management of subjects and investigative tests will comply with current standards of care. Cost of research tests will be incurred by the Alimentary Pharmabiotic Centre, University College Cork, and Teagasc Food Research Centre, Moorepark, Fermoy, Cork.

Confidentiality

All the information gathered from this study will be stored on a computer and in paper files and will be treated confidentially. You will be identified only by a subject number.

In the event of any publication regarding this study, your identity will not be disclosed.

What happens if there is anything I do not understand?

If there is anything you are not sure about, the Doctor will be happy to explain in more detail to yourself or your relatives, guardians (or legal representative if required). The study will be fully explained to you before you decide if you want to take part. If you have any problems or questions after the study has started you may contact:

Professor Fergus Shanahan

INFORMED CONSENT BY SUBJECT FOR PARTICIPATION IN RESEARCH PROTOCOL

Protocol Number: APC025

Subject's Name:

Title of protocol: To define the composition of the gut microbiota in three groups of healthy volunteers and correlate it to body condition, diet, and microbial (short chain fatty acids) and host metabolism, and immune status (metabolic and inflammatory markers).

Investigators: Prof Fergus Shanahan (PI), Prof Michael Molloy, Prof Eamonn Quigley, Prof David Kerins, Prof Charlie Daly, Prof Paul Ross, Dr Paul Cotter & Dr Paul O'Toole, in association with Dr Conor McCarthy, Dr Eanna Falvey & Ms Ruth Wood-Martin (IRFU).

Participation in this study is voluntary and you may withdraw at any time for any reason

The research project and procedure associated with it have been fully explained to me. All experimental procedures have been identified and no guarantees have been given about the possible results. I have had the opportunity to ask questions concerning any and all aspects of the project and any procedures involved. I am aware that participation is voluntary and I may withdraw consent at any time. I am aware that my decision not to participate or to withdraw will not restrict my access to health care services normally available to me. Confidentiality of records concerning my involvement in this project will be maintained in an appropriate manner. I understand that the investigators have such insurance as is required by law in the event of injury resulting from this research.

I, the undersigned, hereby consent to participate as a subject in the above described project conducted at the Cork Teaching Hospitals and University College Cork. I have received a copy of this consent form for my records. I understand that if I have any questions concerning this research, I can contact the Doctor listed below. If I have further

queries concerning my rights in connection with the research, I can contact the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Lancaster Hall, 6 Little Hanover Street, Cork.

After reading the entire consent form, if you have no further questions about giving consent, please sign the where indicated.

Subject's Signature: _____ Date: _____

NAME (BLOCK LETTERS): _____

Witness' Signature: _____ Date: _____

NAME (BLOCK LETTERS): _____

Investigator's Signature: _____ Date: _____

NAME (BLOCK LETTERS): _____

Supplementary document 3.2| Gut health questionnaire

Date: _____

Code: _____

DIRECTIONS this questionnaire asks you to assess how you have been feeling during the last week. All information is held in strict confidence. Take all the time you need to complete this questionnaire.

Section A:

For each question, place the number that best describes your symptoms beside the question:

0 = No or rarely - you have never experienced the symptom or the symptom is familiar to you but you perceive it as insignificant (monthly or less).

1 = 0ccasionally - Symptom comes and goes and is linked in your mind to stress, diet, fatigue, or some identifiable trigger.

2 = Often - Symptom occurs 2-3 times per week and/or with a frequency that bothers you enough that you would like to do something about it.

3 = Frequently - Symptom occurs 4 or more times per week and/or you are aware of the symptom every day, or it occurs with regularity on a monthly or cyclical basis.

1. Do you feel any abdominal pain or discomfort during or after eating? _____
2. Do you feel any abdominal bloating, excessive burping, or belching after meals _____
3. Do you feel any increase flatulence or wind? _____
4. Do you feel any abdominal grumbling or gurgling? _____

5. Do you ever feel the urgency to open your bowels? _____
6. Do you ever feel like you don't fully empty your bowels (An inability to pass your entire stool)? _____
7. Do you ever feel nausea, vomiting, or heartburn? _____
8. Do you ever feel a loss of appetite? _____
9. Do you ever feel tired or lethargic? _____

Section B:

Currently, how often do you pass a bowel action (please tick a box)

Once a week ☐

Once every 4 - 6 days ☐

Once every 2 - 3 days ☐

Once every 1 - 2 days ☐

Once a day ☐

2 - 3 times per day ☐

4 - 6 times per day ☐

Other _____

Section C:

Do you currently have satisfactory relief of your gut symptoms? (Please tick a box)

Yes ☐

No; If not why? ☐ _____

Section D:

Please tick the box that best describes your current stool type?

Bristol Stool Chart

Type 1		Separate hard lumps, like nuts (hard to pass)
Type 2		Sausage-shaped but lumpy
Type 3		Like a sausage but with cracks on its surface
Type 4		Like a sausage or snake, smooth and soft
Type 5		Soft blobs with clear-cut edges (passed easily)
Type 6		Fluffy pieces with ragged edges, a mushy stool
Type 7		Watery, no solid pieces. **Entirely Liquid**

Section E: Medications

1. When did you last take antibiotics?

Over 12 months ago ☐

6-12 months ago ☐

3-6 months ago ☐

Less than 3 months ago ☐

Unsure ☐

Other_____

2. When did you last take antifungal medication? (Ingestion only)

Over 12 months ago ☐

6-12 months ago ☐

3-6 months ago ☐

Less than 3 months ago ☐

Unsure ☐

Other _____

3. Have you taken any other medication in the last 6 months?

Yes ☐

Please specify _____

No ☐

Rather not say ☐

4. Do you/have you ever taken supplements containing probiotics?

No ☐

Yes

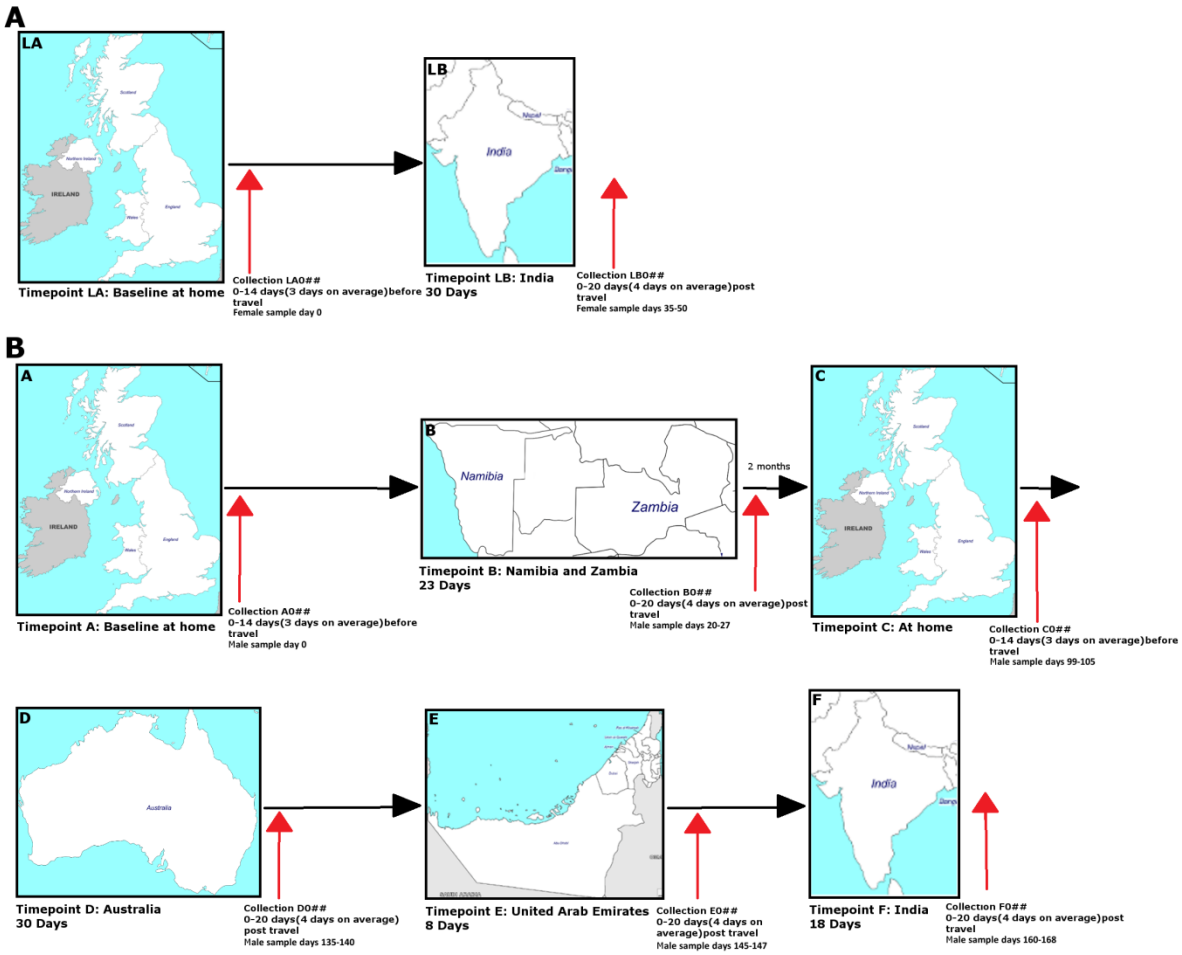
Daily ☐

Weekly ☐

In the last 3 months ☐

Over 3 months ago ☐

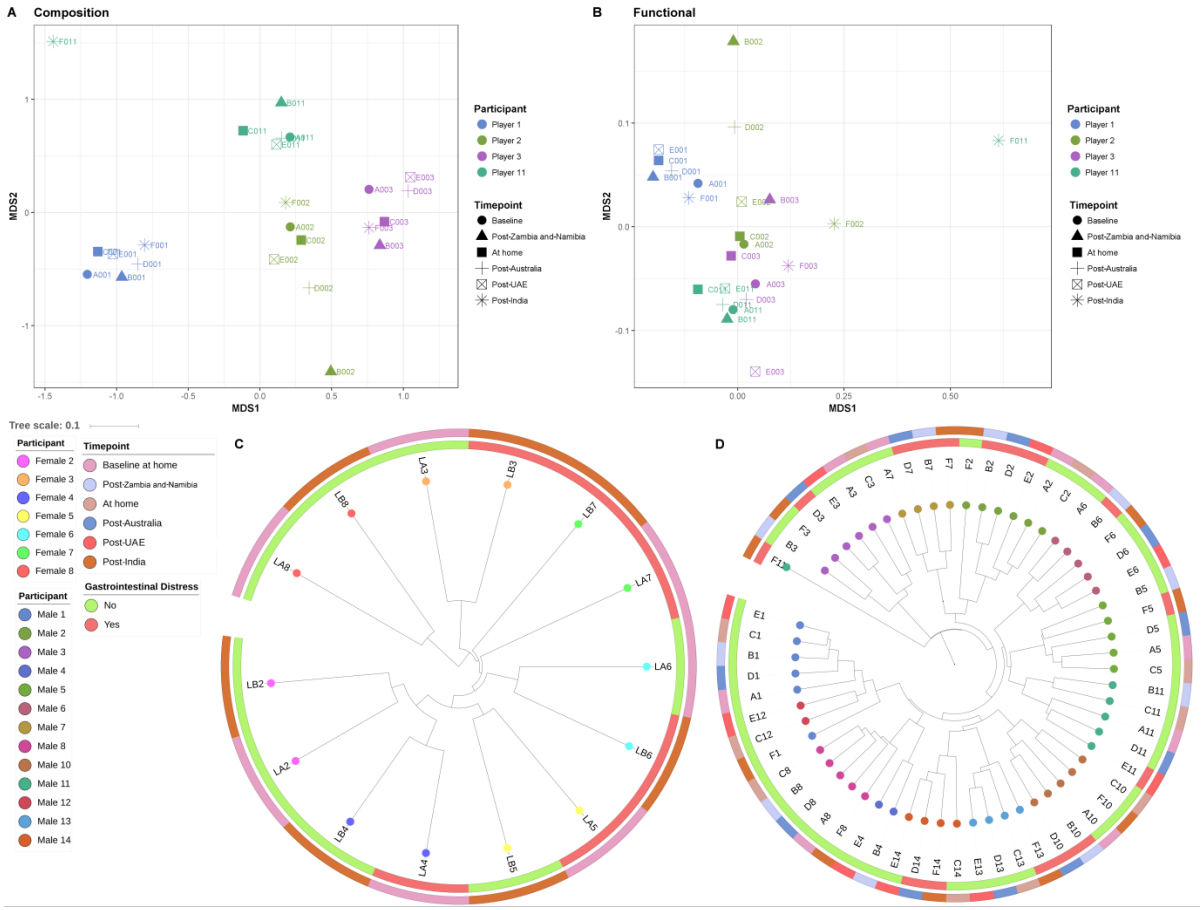
Please provide brand name _____



Supplementary figure 3.1| Timeline

Timeline showing collection points of faecal samples in (A) female and (B) male groups.

Black arrows indicate travel between destinations. Red arrows indicate sample collection.

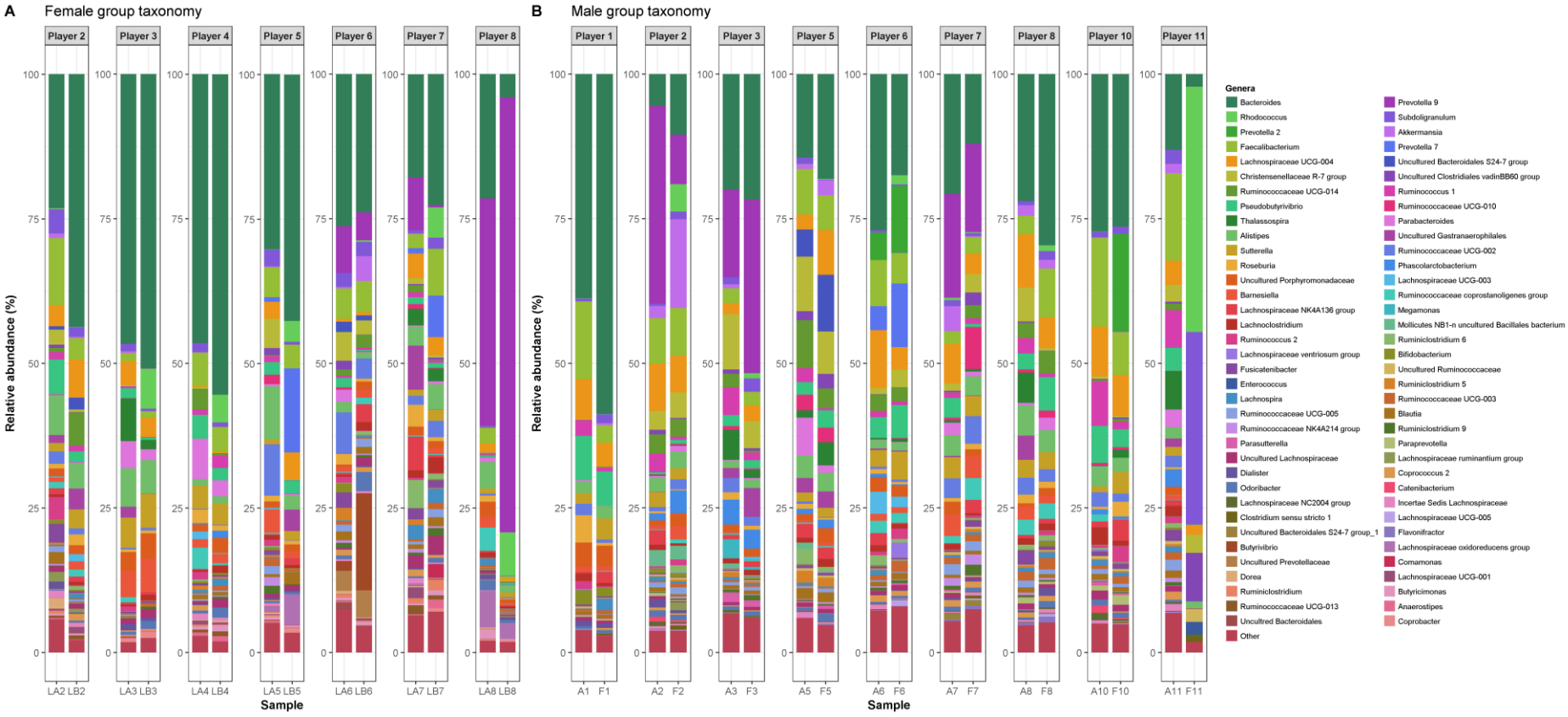


Supplementary figure 3.2| Variations in beta diversity occur in a subset of travelling athletes over time

Supplementary figure 3.2 continued

MDS plots depicting bray-curtis dissimilarities between samples at (A) compositional and (B) functional levels. Colours represent individual participants (N=4) and shapes represent time points. Stress was calculated to be 0.1202777 and 0.06955216, respectively for each plot.

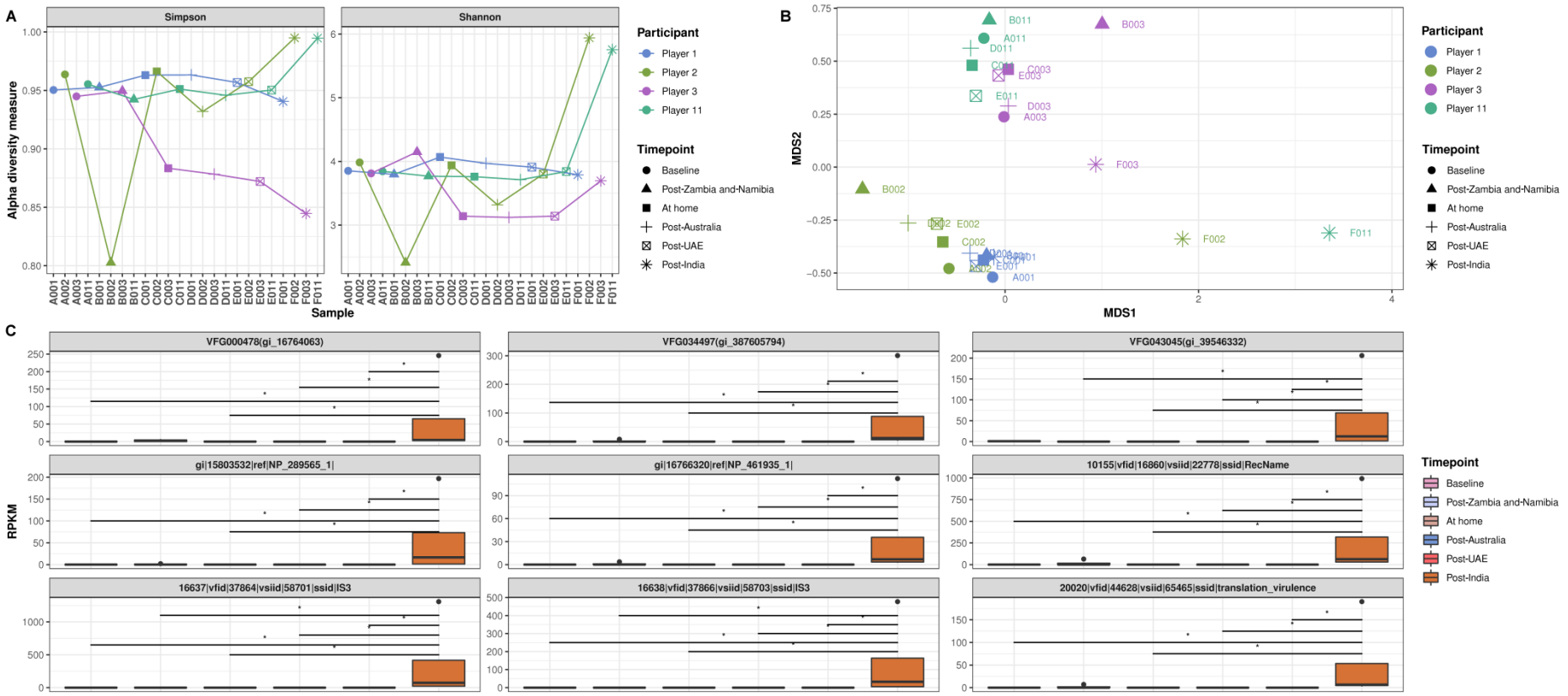
UPGMA phylogenetic trees depicting bray-curtis dissimilarities between samples in (C) female and (D) male groups coloured by the individual sample was from. Individual numbers relate to one player and letters relate to time point (LA: female baseline at home, LB: female post-India, A: male baseline at home, B: male post-Zambia and -Namibia, C: male at home, D: male post-Australia, E: male post-UAE, F: male post-India (see supplementary figure 3.1 for timeline)). Inner circle surrounding trees depicts whether or not an individual had gastrointestinal distress in the week prior to sampling with green being no and red being yes. The outer ring depicts the time point the sample was taken at. Samples cluster based on the individual samples came from.



Supplementary figure 3.3| Taxonomic profiles before and after travel to India

Supplementary figure 3.3 continued

Genus level taxonomic profiles of individual players in the (A) female group (N=7) and (B) male group (N=9) at baseline at home (LA# or A#) and post-India (LB# or F#) time points depicting individual specific variations in taxonomic composition following travel. Each bar represents an individual sample. Each colour represents an individual genus with only those genera which were present at >1% relative abundance shown, with all others grouped as other. 58 and 60 individual genera are shown for the female and male groups, respectively.



Supplementary figure 3.4| Virulence markers in the gut microbiome are increased as a consequence of travel to India

Supplementary figure 3.4 continued

ShortBRED was used to assign metagenomic reads to the virulence factors database. (A) Alpha diversity of virulence genes by participant over time. (B) MDS plot depicting bray-curtis dissimilarities between samples based on reads per kilobase of reference sequence per million sample reads (RPKM) assigned to a virulence marker gene. Samples cluster mostly based on individual the sample derived from, with post-India samples from three participants diverging from this pattern towards each other. Stress was calculated to be 0.09441694. (C) Significant differences between virulence markers by time point. 9 virulence markers were significantly different between the post-India time point and between 4 and 5 of the other time points. RPKM were higher in post-India samples for all cases.

Virulence markers: VFG000478 (gi_16764063): ferric uptake regulator in a Salmonella enterica strain; VFG034497 (gi_387605794): fimbria adhesin EcpD in Escherichia coli; VFG043045 (gi_39546332): flagellar transcriptional regulator FlhC in Salmonella enterica; gi|16766320|ref|NP_461935_1|: pectin-degrading enzyme in a Salmonella enterica strain.

Chapter 4

Distinct Microbiome Composition and Metabolome Exists Across Subgroups of Elite Irish Athletes

Included as published in the *Journal of Science and Medicine in Sport*

(doi: https://doi.org/10.1016/j.jsams.2019.08.290)

Authors: Ciara M. O'Donovan, Sharon M. Madigan, Isabel Garcia-Perez, Alan Rankin, Orla O' Sullivan, and Paul D. Cotter

Contributions

- **Candidate** extracted DNA from faecal samples, prepared samples for shotgun metagenomic sequencing, and completed bioinformatic and statistical analysis.
- **SMM** and **AR** were involved with project conceptualization, sample and metadata collection.
- **IGP** performed metabolomics analysis.
- **OOS** provided guidance for bioinformatics analysis.
- **OOS** and **PDC** supervised the study.

4.0 Abstract

4.0.1 Objectives

The gut microbiome has begun to be characterised in athlete groups, albeit, to date, only across a subset of sports. This study aimed to determine if the gut microbiome and metabolome differed across sports classification groups (SCGs) among elite Irish athletes, many of whom were participating in the 2016 Summer Olympics.

4.0.2 Methods

Faecal and urine samples were collected from 37 international level athletes. Faecal samples were prepared for shotgun metagenomic sequencing and faecal and urine samples underwent metabolomic profiling.

4.0.3 Results

Differences were observed in the composition and functional capacity of the gut microbiome of athletes across SCGs. The microbiomes of athletes participating in sports with a high dynamic component were the most distinct compositionally (greater differences in proportions of species), while those of athletes participating in sports with high dynamic and static components were the most functionally distinct (greater differences in functional potential). Additionally, both microbial (faecal) and human (urine) derived metabolites were found to vary between SCGs. In particular *cis*-aconitate, succinic acid, and lactate, in urine samples, and creatinine, in faeces, were found to be significantly different between groups. These differences were evident despite the absence of significant differences in diet, as determined using food frequency questionnaires, which were translated into nutrient intake values using FETA (Food Frequency Questionnaire European Prospective Investigation into Cancer and Nutrition Tool for Analysis).

4.0.4 Conclusions

A difference in the gut microbiome and metabolome between groups, in the absence of dietary changes, indicates a role for training load or type as a contributory factor. Further exploration of this hypothesis has the potential to benefit athletes, aspiring athletes, and the general public.

4.1 Practical implications

- The gut microbiome and metabolome of athletes differs across sports classification groups.
- Diet is not the driver of differences in the gut microbiome and metabolome across these sports classification groups.
- The difference in microbiomes and metabolome across groups, in the absence of significant dietary differences, suggests a role for exercise type as a contributing factor.
- The identification of a causative factor of these separations was not feasible within this study and potential future studies would need to assess parameters such as detailed exercise tracking in a larger cohort to strive towards the identification of causative factors.

4.2 Introduction

It has been established that a variety of lifestyle and environmental factors impact on the gut microbiome. Modulators include diet and pharmacological agents, including antibiotics [1]. Physical fitness has also been recognised as an influencer of the gut microbiome, although the impact of exercise has been variable across different studies [2, 3]. As individuals who participate in exercise at an extreme level and over an extended period, the gut microbiome of elite-level athletes has been the focus of several studies. The gut microbiome of these athletes has been found to be distinct from both a composition and functional perspective from that of controls, while training load across cyclists was found to be associated with differences in gut microbiome composition [4, 5]. While different types and intensities of training, for example forced versus voluntary wheel running, have been found to impact on the gut microbiome in murine models [6], to date the impact of different training types on the human microbiome has not been determined. Nonetheless, different sports have been found to have distinct impacts on other parameters, including cardiovascular measurements, bone turnover, and muscle mass [7, 8].

This study set out to investigate if the gut microbiome and/or metabolome of elite-level athletes reflect the type of sport in which they participated. To achieve this goal, 37 elite Irish athletes, who competed across 16 different sports, were studied.

4.3 Methods

4.3.1 Ethical approval

Ethical approval for the study was granted by the clinical research ethics committee of the Cork teaching hospitals under project code APC073. This study was carried out in accordance with the ethical standards of the research committee, and with the 1964 Helsinki declaration, and its later amendments. All subjects gave written informed consent prior to the beginning of the study (supplementary document 4.1). The authors declare no conflict of interest.

4.3.2 Experimental design

Irish athletes preparing for and/or participating in major international competition (including the 2016 Summer Olympics) were recruited. Faecal and urine samples were collected from female (N=14) and male (N=23) athletes from 16 different sports (field events (N=1), boxing (N=1), cycling (N=1), fencing (N=2), hockey (N=10), judo (N=2), marathon (N=3), Paralympic cycling (N=2), Paralympic swimming (N=2), pentathlon (N=1), race walking (N=2), rowing (N=2), middle distance running (N=4), swimming (N=2), taekwondo (N=1), and triathlon (N=1); see supplementary table 4.1 for participant characteristics summary). All samples were stored initially on a short-term basis at -20°C before storage at -80°C pre-processing. The mean total storage time of faecal samples (between initial storage and DNA extraction) was 15 (±9) days. The mean age of the athletes was 27 (±5) years. Gut health questionnaires (see supplementary document 4.2 for questionnaire and supplementary table 4.2 for responses) and food frequency questionnaires (see supplementary document 4.3 for questionnaire and supplementary table 4.3 for responses) were completed by all participants. Food frequency questionnaires were converted to nutrient intake values using FETA (Food Frequency Questionnaire European Prospective Investigation into Cancer and Nutrition Tool for Analysis) [9]. A

subset of samples (N=14) were collected during the 2016 Summer Olympics in Rio de Janeiro, Brazil. Two individuals (participant OM008 and OM033) provided samples at more than one time point (2 and 3 time points, respectively).

4.3.3 DNA extraction from faecal samples

Faecal samples were homogenised and DNA extraction was completed using the repeated bead beating plus column (RBBC) method, as previously described [10]. Briefly, 0.25 g of faecal sample was added into a sterile 2 mL screw-cap tube containing 0.25 g of zirconia beads (0.25 g of 0.1 mm, 0.125 g of 1.5 mm, and a single 2.5 mm bead). Following the addition of 1 mL of lysis buffer (500 mM NaCl, 50 mM Tris-HCl, 50 mM EDTA, and 4% sodium dodecyl sulphate (SDS)) samples were homogenized for 3 min on a Mini Beadbeater (BioSpec, Bartlesville, USA). Samples were incubated at 70°C for 15 min to lyse the cells. Following centrifugation (16,000 xg; 5 min; 4°C) supernatant was removed and the bead beating steps were repeated, this time using 200 µL of lysis buffer. Supernatant from both bead beating steps was pooled and 10 M ammonium acetate was added. Samples were centrifuged (16,000 xg; 10 min; 4°C) and resulting supernatant was treated with one volume of isopropanol. DNA was pelleted and washed with 70% ethanol before being dissolved in Tris-EDTA. DNA was RNase and proteinase K treated. DNA was finally washed with 100% ethanol and buffers AW1 and AW2 (QIAamp DNA stool mini kit, Qiagen, England) before being eluted in 200 µL AE buffer (Qiagen). DNA was stored at -20°C before library preparation.

4.3.4 Library preparation

Samples were prepared for Illumina NextSeq shotgun sequencing as per modified Nextera XT DNA library preparation protocol (Illumina, California, USA). Briefly, 0.2 ng μL^{-1} of sample DNA was tagmented for 7.5 min at 55°C. Tagmented DNA was amplified and index adapters (i5 and i7) added. Agencourt AMPure beads (Beckman Coulter, Clare, Ireland) were used to

clean up prepared library DNA. DNA samples were diluted to equimolar concentrations before pooling. Prepared library was sequenced on the NextSeq platform using standard Illumina protocols at the Teagasc sequencing facility. The datasets generated and analysed during the current study are available in the European Nucleotide Archive under the study accession number PRJEB32794.

4.3.5 Bioinformatic analysis

Resulting FastQ reads were quality checked by first removing contaminating human derived reads using NCBI Best Match Tagger (BMTagger). Resulting reads were trimmed and poor quality and duplicate reads removed using a combination of Picard and SAM tools [11, 12]. Taxonomic classifications of trimmed reads were determined using Kraken 2 and Bracken [13, 14]. Functional profiling was completed using the Human Microbiome Project Unified Metabolic Analysis Network 2 (HUMAnN2; http://huttenhower.sph.harvard.edu/humann2) [15, 16].

Analysis of resulting taxonomic and functional tables was completed in R (R Foundation for Statistical Computing, Vienna, 2012) (version 3.5.1). Sports were classified into sports classification groups (SCGs), as previously outlined (figure 1 A) [7]. SCGs used in this study were established based on the dynamic and static components of exercise involved within an individual sport. A highly dynamic exercise involves changes in muscle length and joint movement, which is associated with increased oxygen uptake resulting in increased cardiac output. A highly static exercise involves the development of intramuscular strength with minimal changes in muscular length or joint movement. Static exercise is associated with increased voluntary contraction which results in an increased blood pressure load [7]. Statistical analysis was completed using SCGs with groups B1 and B2 being grouped as B (moderate dynamic component) as only 1 individual could be classified to SCG B2. Data was also analysed for potential confounding factors including sex and location of sample

collection. Alpha and beta diversities were computed in R using vegan (version 2.5-2) [17]. A hclust tree was created in the R package ape (version 5.1) and visualised using iToL [18, 19]. LEfSe was used to determine species and pathways which characterise specific groups [20]. LEfSe allows for the identification of pathways or taxa which characterise groups. The significance threshold was set at 0.05 and an LDA effect size threshold of 3 was used for discriminative features. Spearman correlations were calculated in R using the RcmdrMisc package (version 1.0-10) and corrected for multiple comparisons using Holms method. A heatmap was created to visualise functional pathways from HUMAnN2 using the pheatmap package in R.

4.3.6 Metabolomics analysis

Urine samples underwent GC-MS, UPLC-MS, and ^{1}H-NMR analysis, while faecal samples underwent ^{1}H-NMR and UPLC-MS analysis in Imperial College London, as previously described [21-23].

4.3.6.1 ^{1}H-NMR spectroscopy metabolic profiling analysis

Briefly, urine and faecal water samples were prepared with a pH 7.4 phosphate buffer for ^{1}H-NMR spectroscopy as described previously by Garcia-Perez et al. [5]. Samples were analysed at 300 K on a 600 MHz spectrometer (Bruker BioSpin, Karlsruhe, Germany), following a previously described method [21]. A standard one-dimensional pulse sequence with saturation of the water resonance was utilised: RD – gz, 1 – 90° – t – 90° – tm – gz, 2 – 90° – ACQ, whereby the relaxation delay (RD) was set at 4 s, 90° represents the applied 90° radio frequency pulse, interpulse delay (t_1) was set to an interval of 4 µs, mixing time (tm) was 10 ms, magnetic field gradients (g_{z1} and g_{z2}) were applied for 1 ms and the acquisition period (AQA) was 2.7 s. The receiver gain was set to 90.5 across the board. Suppression of the water signal was achieved through continuous wave irradiation at the water resonance frequency using 25 Hz radio-frequency pulse strength during the RD and tm. Each urine

spectrum was acquired using 4 dummy scans, 32 scans, 64 K time domain points, and with a spectral window set to 20 ppm for urine. Prior to Fourier transformation, the free induction decays were multiplied by an exponential function corresponding to a line broadening of 0.3 Hz. ^{1}H NMR spectra were automatically phased and digitized over the range δ -0.5 to 9.5, and each spectrum was baseline corrected. Spectra were subsequently referenced to the internal chemical shift reference (trimethylsilyl-[2, 2, 3, 3,-2H4]-propionate, TSP) at δ 0.0. Spectral regions corresponding to the internal standard (δ -0.5 to 0.5) and water (δ 4.5 to 5.5) were excluded. Confirmation of metabolite identities in the NMR data was obtained using 1D ^{1}H NMR, 2D ^{1}H-^{1}H NMR, and ^{1}H-^{13}C NMR experiments. In addition, statistical tools such as SubseT Optimization by Reference Matching (STORM) and Statistical Total Correlation SpectroscopY (STOCSY) were also applied [24, 25].

4.3.6.2 GC-MS urinary targeted analysis

Urine samples were analysed by GC coupled to MS using a previously described method [22]. The GC-MS system consisted of an Agilent 7890 B (Agilent Technologies, Palo Alto, CA, USA), equipped with an automatic liquid sampler model 7693 coupled to an Agilent 7000 C single quadruple mass selective detector. Acquisition was done by Mass Hunter workstation software for Q (Agilent Technologies, Palo Alto, CA). The GC was fitted with aDB-5 MS capillary column (30 m 0.25 mm i.d., 0.25 mm film thickness; (5%-phenyl)-methylpolysiloxane bonded and cross-linked; Agilent J&W Scientific, Folsom, CA) and helium was used as carrier gas at 1 mL min^{-1}. 1 µL of each derivatised samples was injected in splitless mode injector temperature of 270°C. A glass liner ultra-inert, split less, single taper and glass wool was used to avoid a possible contamination with non-volatile elements into the column. Chromatograms obtained were uploaded directly into the MassHunter Workstation Software Version B.07.1 (Agilent Technologies Inc.) for data processing as previously described [22].

4.3.6.3 UPLC-MS urinary and faecal analysis

Urine and faecal samples were analysed by Liquid Chromatography coupled to Mass Spectrometry, using a previously described method by Visser et al. [23]. An Acquity UPLC system coupled to a Xevo tandem MS (Waters Co, Milford MA, USA) was used. 5 µL of derivatized urine and 2.5 µL of derivatized faecal sample were injected into a reversed-phase column (Acquity BEH C18 100 mm x 2.1 mm, 1.7 µm). The desolvation gas flow was set at 900 mL min^{-1}, and the cone gas flow at 40 mL min^{-1}. The desolvation temperature was 650°C. Data was collected in multiple reaction monitoring, with an electrospray ionization source in positive mode. The capillary voltage was 2700 V, the cone voltage 18 V and the collision voltage (CV). Spectral acquisitions were performed using the Waters MassLynx v4.1 software.

4.3.6.4 Data analysis

Resulting data sets were imported into R for multivariate and univariate data analysis (R Foundation for Statistical Computing, Vienna, 2012) (version 3.5.1). PCA and OPLS-DA analyses were performed using the ropls package to observe clustering based on sex, location of sample collection, and SCGs. Multidimensional scaling (MDS) plots were computed in R using vegan (version 2.5-2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05.

4.4 Results

This study set out to determine if differences existed in the gut microbiome and metabolome of elite-level athletes participating in different categories of sports. To gain an understanding of the impact of dynamic versus static components of exercise, each sport was classified into a broader sports classification group (SCG) (figure 4.1 A) [7]. Dietary intake was established using food frequency questionnaires which were converted into energy, food group, and nutrient intakes using FETA [9] (supplementary table 4.3). While individual variation in nutrient intakes were observed, no significant differences were found between any nutrients or food groups based on SCGs, gender or location of sample collection following correction for multiple comparisons (supplementary table 4.4). Although microbial diversity within samples was not found to differ based on SCGs (supplementary figure 4.1), microbial diversity between samples, as illustrated with MDS, revealed that samples from athletes within the same SCG generally cluster together (figure 4.1 B and C), from both a compositional (R^2 = 0.20396; p = 0.003; Stress = 0.1663576) and functional potential (R^2 = 0.23542; p = 0.001; Stress = 0.143677) perspective.

Analysis of compositional variation at species level revealed samples were dominated by one or a combination of five species, namely, *Eubacterium rectale*, *Polynucleobacter necessarius*, *Faecalibacterium prausnitzii*, *Bacteroides vulgatus*, and *Gordonibacter massiliensis* and, while many of these species (including *F. prausnitzii*) were present across all samples, the clustering of samples reflected the relative abundance of these species. These five species were found to account for 19-72% relative abundance within these samples. Samples in this study were identified as being dominated by species from the genus *Bacteroides*, with samples comprised of 19.8±11.6% *Bacteroides* species.

Discriminatory species were identified with linear discriminant analysis (LDA) effect size (LEfSe), which allows for the identification of species or pathways which explain differences

between groups [20]. LEfSe analysis revealed a number of species that discriminated between SCGs (figure 4.2 A). *Streptococcus suis*, *Clostridium bolteae*, Lactobacillus phage LfeInf, and *Anaerostipes hadrus* were found to be associated with those groups with a moderate dynamic component (B1 and B2), which includes sports such as fencing. *Bifidobacterium animalis*, *Lactobacillus acidophilus*, *Prevotella intermedia,* and *F. prausnitzii* were found to be associated with the SCG C1 (high dynamic and low static components; including sports such as field hockey), while *Bacteroides caccae* was found to be associated with the SCG C3 (high dynamic and static components; including sports such as rowing). *Bacteroides caccae*, *F. prausnitzii,* and *Anaerostipes hadrus* exhibited greatest discriminatory potential, with LDA effect sizes of >4, with respect to the SCGs C3, C1, and B, respectively. Samples from the group C3 were found to have 4.5 times greater relative abundance of *Bacteroides caccae* than other groups, while *Anaerostipes hadrus* was present at 3 times greater relative abundance in the group B relative to other groups. *F. prausnitzii* was present at 1.5 times greater relative abundance in samples from group C1 in comparison to other groups. Notably, samples from the group C1 were found to have 25 times greater relative abundance of *Lactobacillus acidophilus*. Interestingly, no species were identified as being associated with the SCGs A3 (low dynamic and high static components; including sports such as judo) or C2 (high dynamic and moderate static components; including sports such as swimming) when the LDA effect size cut off was set to 3. Substantial variability exists at the functional level across all groups. A heatmap was used to visualise 25 pathways that were the most variable between samples and samples were separated into five clusters based on the reads per million of these pathways. Samples from individuals OM007 (middle distance running; SCG C2) and OM032 (swimming; SCG C2) were seen to have a higher abundance of many pathways than was apparent in their counterparts, and these samples cluster together, away from all other samples in cluster 1. Samples from cluster 2 were mostly collected in Ireland and can be

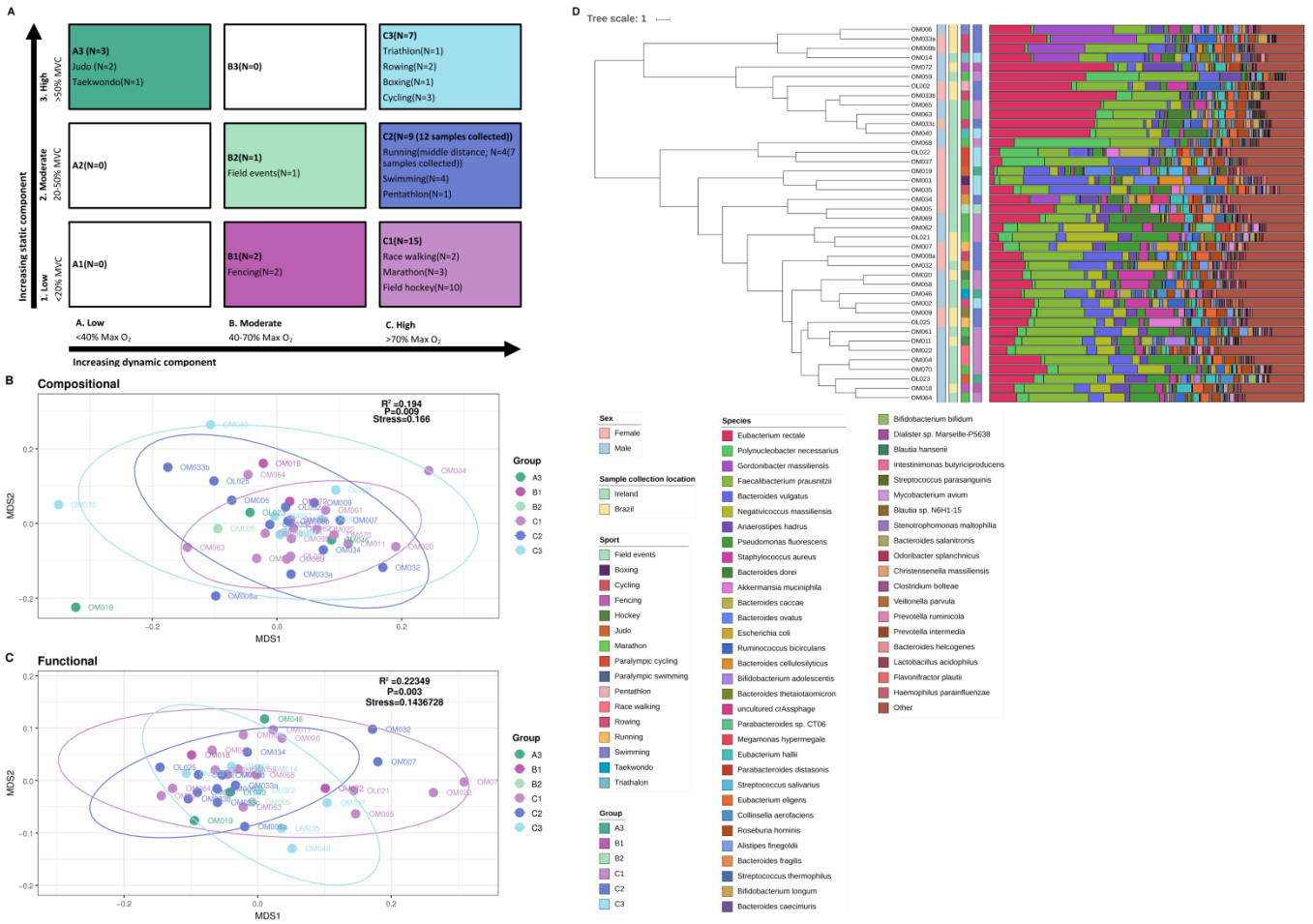


Figure 4.1| Diversity and composition of the gut microbiome

Figure 4.1 continued

(A) Classification of sports based on peak static and dynamic components. Adapted from Mitchell et al., to only include those sports from which samples were collected from during this study and coloured by classification group [7]. With an increasing dynamic component an increase in estimated percent of maximal oxygen uptake (Max O_2) is achieved, resulting in an increased cardiac output. An increasing static component is related to estimated percent of maximal voluntary contraction reached and results in an increased blood pressure load. The lowest cardiovascular demands (cardiac output and blood pressure) are in those sports grouped to B1 (in this analysis), while the highest are grouped to C3. No classification was available for pentathlon and so this was classified to highest score of those sports involved in pentathlon (i.e. swimming and running which are classified as C2).

(B + C) Multidimensional scaling (MDS) result from (B) species level compositional profiles from Bracken following taxonomic classification with Kraken 2 and (C) HUMAnN2 profiles of functional potential, as determined using vegan in R, show the differences between faecal samples based on group (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)). Clustering of samples was observed based on classification group for both compositional and functional data.

(D) Hierarchical cluster analysis, with species level taxonomic profiles. Each bar represents an individual sample. Each colour represents an individual species, as classified with Bracken following taxonomic classification with Kraken 2. Only those species which were present at >1% relative abundance in at

Figure 4.1 continued

least 1 sample are shown, with all other species grouped as other. Colours of stacked bars alongside tree represent sex of participant, location at which sample was collected, sport, and sports classification group. No clustering is observed based on metadata available.

characterised by samples which have greater proportions of 5 pathways (SO4ASSIM-PWY, PWY-821, ARGININE-SYN4-PWY, P162-PWY, and ARGORNPROST-PWY), which are involved with acetate biosynthesis and sulphate degradation. Those samples from cluster 4 contain a lower proportion of reads for these 25 pathways in comparison to other samples. Samples from cluster 5 have greater proportion of 3 pathways (PWY-7007, PWY-7315, and P562-PWY), which are involved with methyl ketone and outer membrane biosynthesis.

In order to identify pathways which were significantly different between groups, LefSe was again utilised (figure 4.2 C). A greater number of pathways (10) were found to be associated with the SCG C3 than any other group and this included pathways involved with folate and amino acid biosynthesis which were found to be 1.5 times greater in this group in comparison to other groups. 5 pathways were found to be associated with the SCG B and this included pathways involved with flavin biosynthesis and fermentation. Notably, the pathway P461-PWY was found to be present in group B in proportions 2 times greater than in other groups. This pathway is responsible for fermentation of sugar alcohols and has been identified in a number of *Streptococcus* species. *S. suis*, a species identified as associated with this group may therefore be responsible for this observation. 5 pathways were found to be associated with the SCG A3 while, interestingly no species were identified as associated with this group. No pathways were identified as being associated with the SCG C2, although 4 species were identified as being significantly increased in this group in comparison to other groups, indicating that while the composition of this group was unique from other groups, the functional potential did not vary. No species or pathways were identified as being significantly different in the SCG C1 in comparison to other groups. None of the collected metadata (gut health or dietary intake) was found to be correlated with the species or pathways identified in this study.

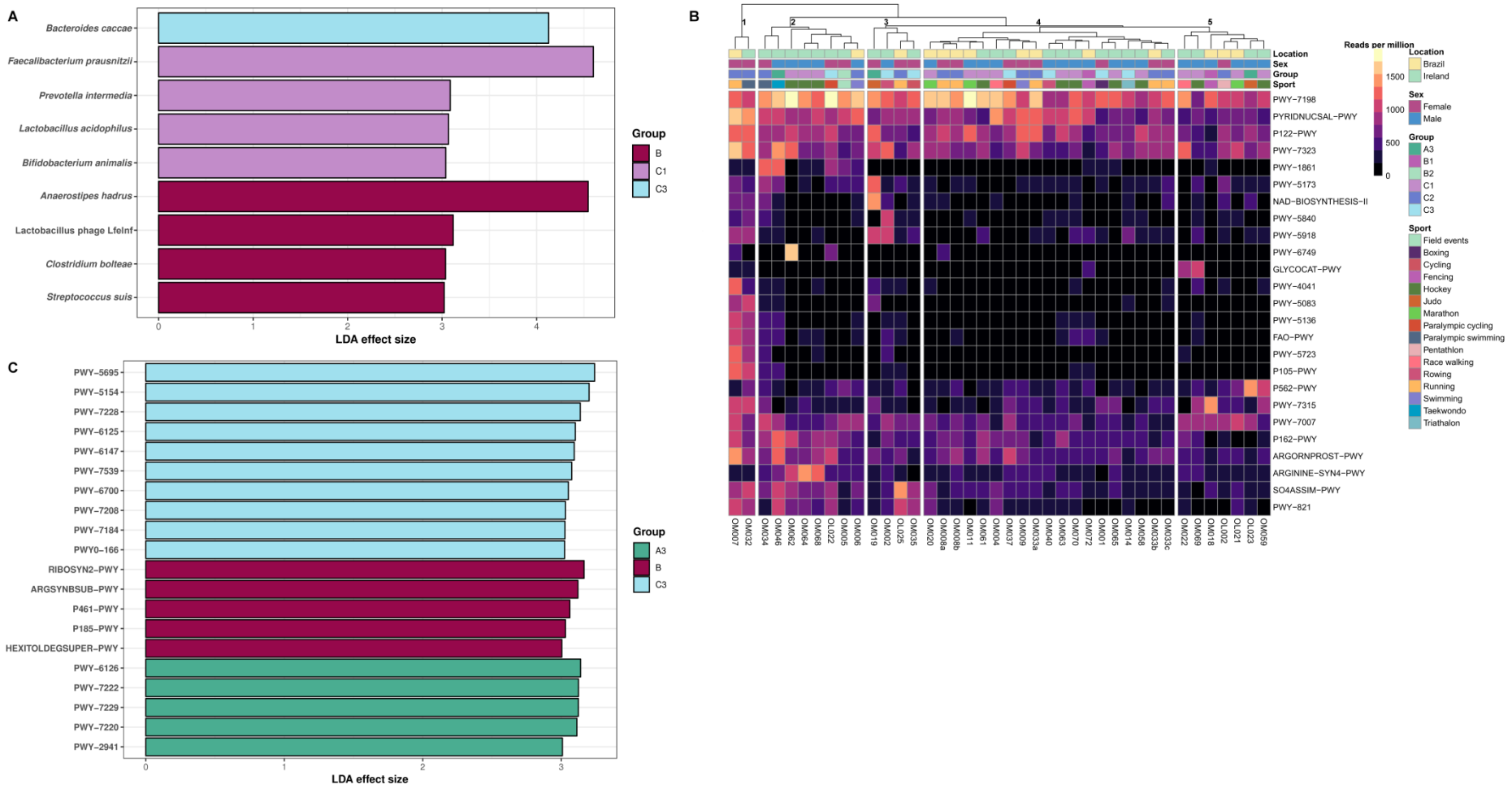


Figure 4.2| Composition and functional potential of the gut microbiome vary based on sports classification groups

Figure 4.2 continued

(A) LEfSe plots identifying discriminatory species, as classified with Bracken following taxonomic classification with Kraken 2, between sports classification groups (A3 (N=3), B (grouped samples with a moderate dynamic component for statistical analysis; N=3), C1 (N=15), C2 (N=11), C3 (N=7)) (See figure 4.1 A for more information regarding groups). An LDA effect size cut off of 3 was used.

(B) HUMAnN2 heat map reporting the 25 most variable pathways that were greater than 1% relative abundance in at least one sample. Clustering completed using Wards method. Annotations below tree represent sex of participant, location at which sample was collected, sport, and sports classification group. Variability in pathway abundance was observed between samples which was not accounted for by metadata available.

Pathways; PWY-4041: Gamma glutamyl cycle, PWY-5136:Fatty acid beta oxidation II (peroxisome), PWY-5840: Superpathway of menaquinol-7 biosynthesis, PWY-7007: Methyl ketone biosynthesis, PWY-5083: NAD/NADH phosphorylation and dephosphorylation, FAO-PWY: Fatty acid beta oxidation I, P122-PWY: Heterolactic fermentation, PWY-5723: Rubisco shunt, PWY-5173: Superpathway of acetyl-CoA biosynthesis, GLYCOCAT-PWY: Glycogen degradation I (bacterial), P562-PWY: Myo-inositol degradation I, P162-PWY: L-glutamate degradation V (via hydroxyglutarate), NAD-BIOSYNTHESIS-II: NAD salvage pathway II, ARGORNPROST-PWY: Arginine, ornithine, and proline interconversion, PWY-5918: Superpathay of heme biosynthesis from glutamate, ARGININE-SYN4-PWY: L-ornithine de novo biosynthesis, SO4ASSIM-PWY: Sulfate reduction I (assimilatory), P105-PWY: TCA cycle IV (2-oxoglutarate decarboxylase), PYRIDNUCSAL-PWY: NAD salvage pathway I, PWY-7315: dTDP-N-acetylthomosamine biosynthesis, PWY-6749: CMP-legionaminate biosynthesis I, PWY-7198: Pyrimidine deoxyribonucleotides de novo biosynthesis IV, PWY-7323: Superpathway of GDP-mannose-

Figure 4.2 continued

derived O-antigen building blocks biosynthesis, PWY-1861: Formaldehyde assimilation II (RuMP Cycle), PWY-821: Superpathway of sulphur amino acid biosynthesis (Saccharomyces cerevisiae).

(C) LEfSe plots identifying discriminatory pathways, as classified with HUMAnN2, between sports classification group (A3 (N=3), B (grouped samples with a moderate dynamic component for statistical analysis; N=3), C1 (N=15), C2 (N=11), C3 (N=7)). Unstratified pathway abundance (reads per million) assignments as determined using HUMAnN2 were analysed. An LDA effect size cut off of 3 was used.

Pathways; PWY-5695: Urate biosynthesis/inosine 5'-phosphate degradation, PWY-5154: L-arginine biosynthesis III (via N-acetyl-L-citrulline), PWY-7228: Superpathway of guanosine nucleotides de novo biosynthesis I, PWY-6125: Superpathway of guanosine nucleotides de novo biosynthesis II, PWY-6147: 6-hydroxymethyl-dihydropterin diphosphate biosynthesis I, PWY-7539: 6-hydroxymethyl-dihydropterin diphosphate biosynthesis III, PWY-7208: Superpathway of pyrimidine nucleobases salvage, PWY-7184: Pyrimidine deoxyribonucleotides de novo biosynthesis I, PWY0-166: Superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis (E. coli), RIBOSYN2-PWY: Flavin biosynthesis I (bacteria and plants), ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle), P461-PWY: Hexitol fermentation to lactate, formate, ethanol and acetate, P185-PWY: Formaldehyde assimilation III (dihydroxyacetone cycle), HEXITOLDEGSUPER-PWY: Superpathway of hexitol degradation (bacteria), PWY-6126: Superpathway of adenosine nucleotides de novo biosynthesis II, PWY-7222: Guanosine deoxyribonucleotides de novo biosynthesis II, PWY-7229: Superpathway of adenosine nucleotides de novo biosynthesis I, PWY-7220: Adenosine deoxyribonucleotides de novo biosynthesis II, PWY-2941: L-lysine biosynthesis II, PWY-6700: queuosine biosynthesis.

Faecal and urine samples underwent metabolomic analysis (NMR and UPLC-MS analysis for faecal samples and NMR, GC-MS, and UPLC-MS analysis for urine samples) to explore variations in microbial and human derived metabolites, respectively. MDS analysis (figure 4.3) revealed significant separation between SCGs based on urine GC-MS (A) (R^2 = 0.471; p = 0.003; Stress = 0.021), urine NMR (B) (R^2 = 0.212; p = 0.033; Stress = 0.158), and faecal NMR (C) (R^2 = 0.246; p = 0.036; Stress = 0.178) analysis. No separation was seen between SCGs based on UPLC-MS analysis (faecal and urine; supplementary figure 4.2). OPLS-DA analysis revealed separation between the SCGs A3 and B1; B1 and C2; and C1 and C2 based on urine GC-MS, between the SCGs A3 and B1; and A3 and C2 based on urine NMR and between the SCGs A3 and B1 based on faecal NMR (supplementary figures 4.3-4.5).

21 metabolites were found to be significantly different between SCGs (supplementary table 4.5), while pairwise analysis between SCGs found that only 4 of these were significantly different between at least 2 SCGs, 3 urine metabolites and 1 faecal metabolite (figure 4.3 D). *Cis*-aconitate and succinic acid were found to be in significantly greater concentrations in GC-MS analysed urine samples from the SCG C2 in comparison to C1. While these metabolites were outside the limit of quantification for some samples, where the metabolites were quantified they were in concentrations 2 and 2.5 times greater than group C1. Lactate was found to be significantly different between the groups C1 and C2, and C1 and C3 from NMR analysis of urine samples. Although concentrations of this metabolite were small, differences observed in the mean concentrations found that the SCG C1 had concentrations of lactate which were 8 and 2 times lower than those found in the SCG's C2 and C3, respectively. Creatinine was the only faecal metabolite identified as being significantly different between SCGs (p=0.017), with this metabolite identified as being in a greater concentration in the group C1 in comparison to the groups A3, C2, and C3 following NMR analysis.

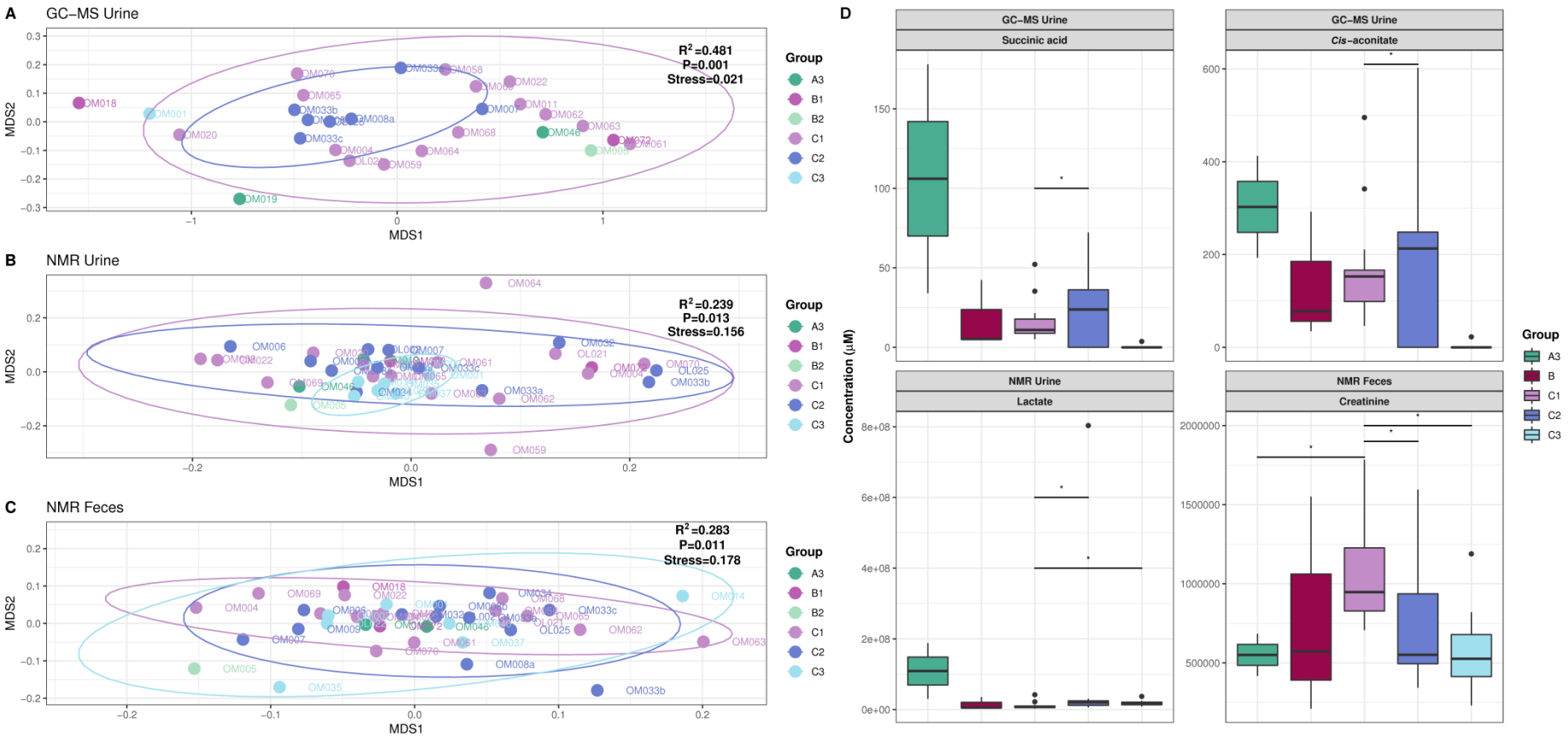


Figure 4.3| Metabolomic analysis reveals separation in microbial and human derived metabolites based on sports classification groups

Figure 4.3 continued

(A + B + C) Multidimensional scaling (MDS) result from (A) GC-MS analysis of urine samples, (B) NMR analysis of urine samples, and (C) NMR analysis of faecal samples, as determined using vegan in R, show the differences between metabolites contained in samples based on sports classification group (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups). Clustering of samples was observed based on sports classification group for both GC-MS urine and NMR analysis of urine and faecal samples.

(D) Significantly different metabolites in urine and faecal samples, as determined with NMR and GC-MS analysis, between sports classification groups (A3 (N=3), B (grouped samples with a moderate dynamic component for statistical analysis; N=3), C1 (N=15), C2 (N=11), C3 (N=7)) (See figure 4.1 A for more information regarding groups). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. * $P<0.05$.

Interestingly, all 4 metabolites identified as significantly different between groups were found to be significantly different between the SCGs C1 and C2, which only differ based on their static component. As these samples have a comparably larger sample size (N=15 and N=12, respectively) in comparison to other groups it is possible that this is the reason for this observation. Notably, the SCG C2 was found to have no significant differences in relation to microbial composition and functional potential when compared to other groups; however variation was observed in this group in relation to metabolites highlighting the importance of a multi-faceted approach utilising both sequencing and metabolic approaches to identify differences between groups. No significant correlations were observed between metabolites and any individual species or pathways. Furthermore, no correlations were identified between any of the metabolites and any available metadata for this study.

4.5 Discussion

Exercise has been found to have a beneficial impact on a number of body systems including the prevention and treatment of obesity, and alleviation of symptoms of depression, anxiety, and disorders affecting the cardiovascular system [26, 27]. In addition, exercise is thought to influence, or correlate with differences in, the gut microbiome across animal and human studies [2, 5, 6]. However, little is known about the variation in the gut microbiome and metabolome between sports. This study set out to gain an initial insight into the possibility that the gut microbiome or metabolome differ across athletes from different sports or SCGs.

SCGs were used to distinguish between sports based on the static and dynamic components of the sport allowing for a comparison of sports based on their impact on cardiorespiratory measures [7]. Blood pressure (static exercise) and VO_2 max (dynamic exercise) have been linked with gut microbiome composition previously, however most of these associations were found in obese cohorts or those with cardiovascular disease [28, 29]. Although cardiorespiratory measures were not taken in the current study, the classification into SCGs, and correlations with gut microbiome composition and functionality as well as metabolome, provides further evidence of a link.

Samples from this study were found to be significantly different from a microbial composition and functional potential perspective between SCGs, while the metabolome was also identified as variable between groups. Individual variability exists across athletes with clustering of the majority of samples driven by the relative abundances of five species, namely *E. rectale*, *P. necessaries*, *F. prausnitzii*, *B. vulgatus,* and *G. massiliensis*, which have previously been identified as common gut inhabitants. Although *E. rectale* has been found to be enriched in participants with a low VO_2 max [29], this finding was not identified in this study, potentially as a consequence of a high maximal voluntary contraction being

associated with the group with the lowest VO_2 max (A3). *F. prausnitzii*, a butyrate producer, has previously been found in greater relative abundance in active women, with an increase seen with exercise training [2]. This species, which was also identified in this study as being associated with the SCG C1, is a dominant taxon in healthy cohorts, with depletion evident in disease-associated cohorts, including in IBD [30], indicating their presence may be a beneficial. Among the species found to discriminate for SCGs, *L. acidophilus* was found to discriminate for the SCG C1. This is not surprising given that eight out of the fifteen participants from this group reported taking a product which contains this species, demonstrating compliance with reported supplementation rather than a reflection of the species of the general SCG. Another species used in probiotic supplements, *B. animalis*, was also identified in this group, although this species was not listed as a component of the reported supplement [31]. Species associated with the SCG B include those with the potential to produce butyrate, a short chain fatty acid found to promote gut homeostasis, specifically *A. hadrus* and *C. bolteae* [32, 33]. Variation in species composition may be as a result of the variance in demands across different types of sports, including duration of activity and training modes. This in turn may result in an alteration in gut microbiome composition as a consequence of the alteration of gut transit time, or some other physiological change [34, 35].

Both the faecal and urine metabolomes were found to be distinct between SCGs indicating differences in microbial and human derived metabolites, respectively. Succinic acid was found to be greater in the SCG C2 relative to the SCG C1. Conversely, succinate was previously found to be lower in rugby players relative to inactive controls [5], while in the current study succinic acid was found to increase with an increased cardiac output. Lactate, produced in the muscles during normal metabolism and exercise [36], was found to be increased in SCGs with an increased static component (i.e., in the SCGs C2 and C3 relative to C1). An increased static component involves the development of intramuscular strength

[7], potentially resulting in an increase in the production of lactate. Creatinine is the breakdown product of creatine phosphate in muscle and is related to muscle turnover [37]. This metabolite was found to be greatest in the SCG C1 relative to A3, C2, and C3 indicating muscle turnover may have been greatest in this cohort with a high dynamic and low static component. Dynamic exercises are involved with changes in muscle length and joint movement [7], potentially resulting in the increase in the production of creatinine. Previously, those sports involving resistance training have been indicated as resulting in the greatest muscle turnover [38], however the extended endurance component involved with the sports in this cohort may give rise to a similar effect.

These differences in microbiomes and metabolomes were observed despite no significant variation in dietary intakes across athletes from different SCGs; suggesting that variations in training loads and competition requirements contributed to these microbiome and metabolome-related patterns. Studies of athlete groups across various laboratories have also highlighted microbiome differences, and while this may be as a result of variation in other methodologies used (e.g., sequencing analysis methodologies), differences in sport may also play a role [4, 5]. Furthermore, those differences identified between sports types in the gut microbiome composition, but in particular in functionality and metabolism, may confer sports specific benefits; however research identifying such benefits would require further work.

4.6 Conclusion

In conclusion, the gut microbiome of the athlete cohorts tested in this study are distinct based on SCGs, with differences also established in the urine and faecal metabolome. While studies have identified variation in the gut microbiome within cyclists based on time per week spent exercising [4], to our knowledge this is the first study identifying a separation in the gut microbiome between different sports. The identification of a

causative factor of these separations was not feasible within this study and potential future studies would need to assess parameters such as detailed exercise tracking in a larger cohort to strive towards the identification of causative factors. The identification of species, pathways, and metabolites driving separations between SCGs could help identify an optimal gut microbiome in the context of the sport an individual is participating in. Additionally, an understanding of the variation in an athlete microbiome could help guide future studies surrounding the impact of exercise on the gut microbiome within general populations. Furthermore, species identified here as increased for individual SCGs could form the basis of future investigations to identify their influence on performance, as was recently determined for *Veillonella* [34].

4.7 Acknowledgements

The authors thank Fiona Crispie and Laura Finnegan for their contributions to DNA sequencing. The authors are grateful to Orla Walsh, Laura Mahony, and all those who donated their time, effort, and samples to this study. This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273, 13/SIRG/2160, and 11/PI/1137. Isabel Garcia-Perez is supported by a NIHR Career development research fellowship (NIHR-CDF-2017-10-032). Infrastructure support was provided by the NIHR Imperial Biomedical Research Centre (BRC) based at Imperial College Healthcare National Health Service (NHS) Trust and Imperial College London.

4.8 References

1. Conlon, M.A. and Bird, A.R., *The impact of diet and lifestyle on gut microbiota and human health.* Nutrients, 2014. **7**(1): p. 17-44.
2. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).
3. Cronin, O., Barton, W., Skuse, P., Penney, N.C., Garcia-Perez, I., Murphy, E.F., Woods, T., Nugent, H., Fanning, A., Melgar, S., Falvey, E.C., Holmes, E., Cotter, P.D., O'Sullivan, O., Molloy, M.G., and Shanahan, F., *A prospective metagenomic and metabolomic analysis of the impact of exercise and/or whey protein supplementation on the gut microbiome of sedentary adults.* mSystems, 2018. **3**(3).
4. Petersen, L.M., Bautista, E.J., Nguyen, H., Hanson, B.M., Chen, L., Lek, S.H., Sodergren, E., and Weinstock, G.M., *Community characteristics of the gut microbiomes of competitive cyclists.* Microbiome, 2017. **5**(1): p. 98.
5. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.
6. Allen, J.M., Berg Miller, M.E., Pence, B.D., Whitlock, K., Nehra, V., Gaskins, H.R., White, B.A., Fryer, J.D., and Woods, J.A., *Voluntary and forced exercise differentially alters the gut microbiome in C57BL/6J mice.* Journal of Applied Physiology, 2015. **118**(8): p. 1059-1066.
7. Mitchell, J.H., Haskell, W., Snell, P., and Van Camp, S.P., *Task force 8: Classification of sports.* Journal of the American College of Cardiology, 2005. **45**(8): p. 1364-1367.
8. Kouvelioti, R., LeBlanc, P., Falk, B., Ward, W.E., Josse, A.R., and Klentrou, P., *Effects of high-intensity interval running versus cycling on sclerostin, and markers of bone turnover and oxidative stress in young men.* Calcified Tissue International, 2019. **104**(6): p. 582-590.
9. Mulligan, A.A., Luben, R.N., Bhaniani, A., Parry-Smith, D.J., O'Connor, L., Khawaja, A.P., Forouhi, N.G., Khaw, K.T., and Study, EPIC-Norfolk FFQ study, *A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability.* BMJ Open, 2014. **4**(3): p. e004503.

10. Yu, Z. and Morrison, M., *Improved extraction of PCR-quality community DNA from digesta and fecal samples.* Biotechniques, 2004. **36**(5): p. 808-812.
11. BroadInstitute. *Picard*. 17.01.2019; Available from: http://broadinstitute.github.io/picard/.
12. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
13. Wood, D.E. and Salzberg, S.L., *Kraken: Ultrafast metagenomic sequence classification using exact alignments.* Genome Biology, 2014. **15**(3): p. R46.
14. Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L., *Bracken: Estimating species abundance in metagenomics data.* PeerJ Computer Science, 2017. **3**: p. e104.
15. Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methe, B., Schloss, P.D., Gevers, D., Mitreva, M., and Huttenhower, C., *Metabolic reconstruction for metagenomic data and its application to the human microbiome.* PLoS Computational Biology, 2012. **8**(6): p. e1002358.
16. Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C., *Species-level functional profiling of metagenomes and metatranscriptomes.* Nature Methods, 2018. **15**(11): p. 962-968.
17. Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O' Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoeca, E., and Wagner, H., Vegan: *Community ecology package*. R package version 2.2-1.2.1-2, 2018.
18. Letunic, I. and Bork, P., *Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees.* Nucleic Acids Research, 2016. **44**: p. 242-245.
19. Paradis, E., Claude, J., and Strimmer, K., *APE: Analyses of phylogenetics and evolution in R language.* Bioinformatics, 2004. **20**(2): p. 289-290.
20. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C., *Metagenomic biomarker discovery and explanation.* Genome Biology, 2011. **12**(6): p. R60.
21. Dona, A.C., Jimenez, B., Schafer, H., Humpfer, E., Spraul, M., Lewis, M.R., Pearce, J.T., Holmes, E., Lindon, J.C., and Nicholson, J.K., *Precision high-throughput proton*

NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Analytical Chemistry, 2014. **86**(19): p. 9887-9894.

22. Zhao, L., Ni, Y., Su, M., Li, H., Dong, F., Chen, W., Wei, R., Zhang, L., Guiraud, S.P., Martin, F.P., Rajani, C., Xie, G., and Jia, W., *High throughput and quantitative measurement of microbial metabolome by gas chromatography/mass spectrometry using automated alkyl chloroformate derivatization.* Analytical Chemistry, 2017. **89**(10): p. 5565-5577.
23. Visser, W.F., Verhoeven-Duif, N.M., Ophoff, R., Bakker, S., Klomp, L.W., Berger, R., and de Koning, T.J., *A sensitive and simple ultra-high-performance-liquid chromatography-tandem mass spectrometry based method for the quantification of D-amino acids in body fluids.* Journal of Chromatography A, 2011. **1218**(40): p. 7130-7136.
24. Cloarec, O., Dumas, M.E., Craig, A., Barton, R.H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J.C., Holmes, E., and Nicholson, J., *Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic 1H NMR data sets.* Analytical Chemistry, 2005. **77**(5): p. 1282-1289.
25. Posma, J.M., Garcia-Perez, I., De Iorio, M., Lindon, J.C., Elliott, P., Holmes, E., Ebbels, T.M., and Nicholson, J.K., *Subset optimization by reference matching (STORM): An optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids.* Analytical Chemistry, 2012. **84**(24): p. 10694-10701.
26. Rawson, R.A., Chudzynski, J., Gonzales, R., Mooney, L., Dickerson, D., Ang, A., Dolezal, B., and Cooper, C.B., *The impact of exercise on depression and anxiety symptoms among abstinent methamphetamine-dependent individuals in a residential treatment setting.* Journal of Substance Abuse Treatment, 2015. **57**: p. 36-40.
27. Golbidi, S. and Laher, I., *Exercise and the cardiovascular system.* Cardiology Research and Practice, 2012. **2012**: p. 210852.
28. Tang, W.H. and Hazen, S.L., *The gut microbiome and its role in cardiovascular diseases.* Circulation, 2017. **135**(11): p. 1008-1010.
29. Yang, Y., Shi, Y., Wiklund, P., Tan, X., Wu, N., Zhang, X., Tikkanen, O., Zhang, C., Munukka, E., and Cheng, S., *The association between cardiorespiratory fitness and gut microbiota composition in premenopausal women.* Nutrients, 2017. **9**(8).

30. Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottiere, H.M., Dore, J., Marteau, P., Seksik, P., and Langella, P., *Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(43): p. 16731-16736.

31. Eskesen, D., Jespersen, L., Michelsen, B., Whorwell, P.J., Muller-Lissner, S., and Morberg, C.M., *Effect of the probiotic strain Bifidobacterium animalis subsp. Lactis, BB-12(r), on defecation frequency in healthy subjects with low defecation frequency and abdominal discomfort: A randomised, double-blind, placebo-controlled, parallel-group trial.* British Journal of Nutrition, 2015. **114**(10): p. 1638-1646.

32. Kant, R., Rasinkangas, P., Satokari, R., Pietila, T.E., and Palva, A., *Genome sequence of the butyrate-producing anaerobic bacterium Anaerostipes hadrus pel 85.* Genome Announcements, 2015. **3**(2).

33. Dehoux, P., Marvaud, J.C., Abouelleil, A., Earl, A.M., Lambert, T., and Dauga, C., *Comparative genomics of Clostridium bolteae and Clostridium clostridioforme reveals species-specific genomic properties and numerous putative antibiotic resistance determinants.* BMC Genomics, 2016. **17**(1): p. 819.

34. Scheiman, J., Luber, J.M., Chavkin, T.A., MacDonald, T., Tung, A., Pham, L.D., Wibowo, M.C., Wurth, R.C., Punthambaker, S., Tierney, B.T., Yang, Z., Hattab, M.W., Avila-Pacheco, J., Clish, C.B., Lessard, S., Church, G.M., and Kostic, A.D., *Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism.* Nature Medicine, 2019. **25**(7): p. 1104-1109.

35. O'Sullivan, O., Cronin, O., Clarke, S.F., Murphy, E.F., Molloy, M.G., Shanahan, F., and Cotter, P.D., *Exercise and the microbiota.* Gut Microbes, 2015. **6**(2): p. 131-136.

36. Kaya, M., Moriwaki, Y., Ka, T., Inokuchi, T., Yamamoto, A., Takahashi, S., Tsutsumi, Z., Tsuzita, J., Oku, Y., and Yamamoto, T., *Plasma concentrations and urinary excretion of purine bases (uric acid, hypoxanthine, and xanthine) and oxypurinol after rigorous exercise.* Metabolism, 2006. **55**(1): p. 103-107.

37. Knitter, A.E., Panton, L., Rathmacher, J.A., Petersen, A., and Sharp, R., *Effects of beta-hydroxy-beta-methylbutyrate on muscle damage after a prolonged run.* Journal of Applied Physiology, 2000. **89**(4): p. 1340-1344.

38. Atherton, P.J. and Smith, K., *Muscle protein synthesis in response to nutrition and exercise.* The Journal of Physiology, 2012. **590**(5): p. 1049-1057.

4.9 Supplementary material

Supplementary document 4.1| Consent form

SUBJECT INFORMATION SHEET

Study Title: **OlympicMet**

Protocol No.: APC073

Principal Investigator: Prof. Fergus Shanahan

Co-investigators: Dr. Orla O'Sullivan

Dr. Paul Cotter

Dr. Sharon Madigan

Dr. Rod McLoughlin

Dr. Sarah Jane Cullen

Dr. Adrian McGoldrick

Ciara O'Donovan

Site of Investigation: APC, Microbiome Institute, University College Cork (UCC); Teagasc Food Research Centre, Fermoy, Co. Cork; Racing Academy and Centre of Education (RACE), The Curragh, Co Kildare; and Irish Sports Institute, National Sports Campus, Dublin 15.

Why is this study being run?

The objective of this study is to establish the microbiome of elite athletes in the preparation for and/or during major competition and to determine any influences of exercise and diet on host metabolism, body composition, or immune function.

If you agree to participate in the study you will be asked questions about your diet and general health and your eligibility to participate in the study will be determined.

The visits will take place at the Irish Sports Institute or the Racing Academy and Centre of Education (RACE), The Curragh, Co Kildare or other affiliated training facility. This study will involve at least 1 study over 12 months with possible follow up visits in the following 2 years. At least 20 healthy elite athletes (males and females) who have qualified for major competition will be enrolled.

Study Procedure

You will be advised of the purpose of the study by an investigator from the study team and the procedures which will be undertaken. You will be given a copy of the subject information sheet, which will explain what is required from you. If interested, you will then be requested to read and sign the informed consent form, and receive a signed copy. Your suitability to participate in the study will be checked and you will be asked some general personal questions. The investigator dealing with you will discuss all samples and that we wish to take. In addition you may be asked to attend follow up visits for up to two years.

Visit Outline

You will be given, in advance, a container, and instructions for collecting a stool sample and urine sample in a quick and easy, hygienic device, in your own home.

At the study visit, you will bring your stool and urine sample, produced within the previous 12 hours. Alternatively you will collect these with special kits provided by the study team and post them to the research centre in Fermoy, Co. Cork (kits and stamped envelopes will be provided by the study team). Details of your current and/or past medications will also be collected. You will be asked to give a blood sample (up to 20 mL, which is 4 teaspoons), which will be analysed for biomarkers of health and disease. Your diet will be recorded using the Food Frequency Questionnaire that will be administered during this visit. Your weight and height will be measured and your perceived gut health will be established with a gut health questionnaire.

Should further visits be required the investigator will discuss and plan these with you at your convenience. DNA, urine, and faecal samples will be stored for further testing should it be required. Samples will be destroyed on request.

What happens if I start the study and change my mind later?

You do not have to take part in the study, participation is entirely voluntary. Refusal to participate, or discontinuing participation at any time, will involve no penalty, loss of benefits or denial of treatment or services by the Cork Teaching Hospital or the participating investigators.

Will I experience any unpleasant side effects?

During the collection of your blood sample or insertion of the cannula, you may experience a slight scratch, which may be uncomfortable for a moment but quickly passes.

Funding of trial

There are no cost implications for the Health Services Executive (HSE) or to you. The management of patients and investigative tests will comply with current standards of care.

Cost of research tests will be incurred by the Alimentary Pharmabiotic Centre, University College Cork.

Confidentiality

Any personal information which you volunteer to this study will be treated in accordance with the Data Protection Act 1988 and Data Protection (Amendment) Act 2003. Accordingly, the study team shall only process such information in line with the purpose for which you provided it. Your information shall not be disclosed to third parties unless you are specifically informed otherwise or we require a third party to process the information on our behalf (in which case the University requires these third parties to comply with its instructions and we require that they do not use your personal information for their own business purposes, unless you have explicitly consented to the use of your personal information in this way). You have the right to review your data and request a copy and/or you have the right to request deletion of your data.

All the information gathered from this study will be stored on a computer and in paper files and will be treated confidentially. You will be identified only by a subject number. In the event of any publication regarding this study, your identity will not be disclosed.

What happens if there is anything I do not understand?

If there is anything you are not sure about, the investigators will be happy to explain in more detail to yourself or your relatives, guardians (or legal representative if required). The study will be fully explained to you before you decide if you want to take part.

Dr. Orla O'Sullivan, Research Fellow

Dr. Paul Cotter, Principal Research Officer

Prof Fergus Shanahan, Consultant Gastroenterologist

CONSENT BY SUBJECT FOR PARTICIPATION IN A HUMAN INTERVENTION STUDY

Protocol Number: APC073

Patient Name: _____

Title of protocol: OlympicMet

Principal Investigator: Prof. Fergus Shanahan

Participation in this study is voluntary and you may withdraw at any time for any reason

The research project and procedure associated with it have been fully explained to me. All experimental procedures have been identified and no guarantees have been given about the possible results. I have had the opportunity to ask questions concerning any and all aspects of the project and any procedures involved. I am aware that participation is voluntary and I may withdraw consent at any time. I am aware that my decision not to participate or to withdraw will not restrict my access to health care services normally available to me. Confidentiality of records concerning my involvement in this project will be maintained in an appropriate manner. I understand that the investigators have such insurance as is required by law in the event of injury resulting from this research.

I, the undersigned, hereby consent to participate as a subject in the above described project conducted at the Cork University Hospital and University College Cork. I have received a copy of this consent form for my records. I understand that if I have any questions concerning this research, I can contact the Doctor listed below. If I have further queries concerning my rights in connection with the research, I can contact the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Lancaster Hall, 6 Little Hanover Street, Cork.

Analyses of all samples and information collected will be conducted in the CUMH and/or APC. On some occasions the analyses may be done in collaboration with third parties, including commercial partners, which may require samples to be shipped to these organisations. Please note that data protection laws in these countries may not be as stringent.

In addition samples and data will be stored and may be used in other research studies. In all cases, samples and data will be coded with anonymised identifier numbers. ☐

(Please tick box if you agree)

Access to samples and/or data will require approval from the Director and/or Executive Management Group of the APC, Microbiome Institute, Cork.

I agree that results and microbial isolates may be used in the development of health promoting products. **(Please tick box if you agree)** ☐

I agree that my contact details may be made available for recruitment to further studies and also allow a follow up check on my status following my participation by the this project or its CREC approved collaborators. **(Please tick box if you agree)** ☐

After reading the entire consent form, if you have no further questions about giving consent, please sign the where indicated.

Subject's Signature: _____ Date: _____

NAME (BLOCK LETTERS): _____

Investigator's Signature: _____ Date: _____

NAME (BLOCK LETTERS) _____

Supplementary document 4.2| Gut health questionnaire

Code: _____

Date: _____

DIRECTIONS:

This questionnaire asks you to assess how you have been feeling **during the last week.** All information is held in strict confidence. Take all the time you need to complete this questionnaire.

Section A:

For each question, place the number that best describes your symptoms beside the question:

0 = No or rarely - you have never experienced the symptom or the symptom is familiar to you but you perceive it as insignificant (monthly or less).

1 = 0ccasionally - Symptom comes and goes and is linked in your mind to stress, diet, fatigue, or some identifiable trigger.

2 = Often - Symptom occurs 2-3 times per week and/or with a frequency that bothers you enough that you would like to do something about it.

3 = Frequently - Symptom occurs 4 or more times per week and/or you are aware of the symptom every day, or it occurs with regularity on a monthly or cyclical basis.

1. Do you feel any abdominal pain or discomfort during or after eating? _____
2. Do you feel any abdominal bloating, excessive burping, or belching after meals? _____

3. Do you feel any increase flatulence or wind? _____
4. Do you feel any abdominal grumbling or gurgling? _____
5. Do you ever feel the urgency to open your bowels? _____
6. Do you ever feel like you don't fully empty your bowels (An inability to pass your entire stool)? _____
7. Do you ever feel nausea, vomiting, or heartburn? _____
8. Do you ever feel a loss of appetite? _____
9. Do you ever feel tired or lethargic? _____

Section B:

Currently, how often do you pass a bowel action (please tick a box)

Once a week ☐

Once every 4-6 days ☐

Once every 2-3 days ☐

Once every 1-2 days ☐

Once a day ☐

2-3 times per day ☐

4-6 times per day ☐

Other _____

Section C:

Do you currently have satisfactory relief of your gut symptoms? (Please tick a box)

Yes ☐

No; If no why ☐ _____

Section D:

Please tick the box that best describes your current stool type?

Bristol Stool Chart

Type 1		Separate hard lumps, like nuts (hard to pass)
Type 2		Sausage-shaped but lumpy
Type 3		Like a sausage but with cracks on its surface
Type 4		Like a sausage or snake, smooth and soft
Type 5		Soft blobs with clear-cut edges (passed easily)
Type 6		Fluffy pieces with ragged edges, a mushy stool
Type 7		Watery, no solid pieces. **Entirely Liquid**

Section E: Medications

1. When did you last take antibiotics?

   Over 12 months ago ☐

   6-12 months ago ☐

   3-6 months ago ☐

   Less than 3 months ago ☐

   Unsure ☐

   Other _____

2. When did you last take antifungal medication? (Ingestion only)

   ☐ Over 12 months ago

   ☐ 6-12 months ago

   ☐ 3-6 months ago

   ☐ Less than 3 months ago

   ☐ Unsure

   Other _____

3. Have you taken any other medication in the last 6 months?

   ☐ Yes; please specify _____

   ☐ No

   ☐ Rather not say

4. Do you/have you ever taken supplements containing probiotics?

No ☐

Yes

Daily ☐

Weekly ☐

In the last 3 months ☐

Over 3 months ago ☐

Please provide brand name _____

Supplementary document 4.3| Food frequency questionnaire

Please fill in the following based on the previous months eating habits

		M = month		W = week		D = Day					
ID:											**Date:**
Meat, fish, poultry - Medium serving (size deck of cards)	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Beef (roast / steak)											
Beef (stew)											
Beef burger (1 burger)											
Pork (roast / chops / escalope's)											
Lamb (roast / chops / stew)											
Chicken or other poultry e.g. turkey (roast)											
Breaded chicken, chicken nuggets, chicken burger											
Bacon/Ham											
Processed meat (Corned beef, Luncheon meats, sausages)											
Savoury pies (e.g. meat, pork, steak & kidney, sausage roll)											
Liver, heart, kidney, pâté											

Fish fried (batter/breadcrumbs), baked, grilled, fingers/cakes											
White fish, fresh or frozen (e.g. cod, haddock, plaice, sole)											
Oily fish, fresh, frozen, or canned (e.g. mackerel, kippers, tuna, salmon, sardines, herring)											
Shellfish (e.g. crab, prawns, mussels)											
Bread and savoury biscuits - One slice or one biscuit	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
White bread and rolls (including ciabatta & panini)											
Brown bread and rolls											
Wholemeal bread and rolls											
Wheat-free bread, Rye bread, spelt bread (specify)											
Cream crackers, cheese biscuits											
Crisp bread, e.g. Ryvita											
Pancakes, muffins, oatcakes											
Scone (white)											
Scone (brown)											

Cereals - One medium sized bowl	**Never**	**<1 /M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Non-ready to eat - Porridge, Readybrek											
High fibre (All Bran/flakes, Weetabix, Shredded wheat)											
Cornflakes, Rice Krispies											
Muesli (e.g. Country Store, Alpen)											
Sugar-coated cereals (e.g. Frosties, Crunchy nut cornflakes)											
Potatoes, rice and pasta - Medium serving = a cupful	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Boiled, instant or jacket potatoes											
Mashed potatoes											
Chips / Roast potatoes											
Potato Salad											
White Rice											
Brown Rice											
White/yellow or green pastas (e.g. spaghetti, noodles)											
Wholemeal pasta											
Lasagne (meat based)											

Lasagne (vegetarian)											
Pizza											
Dairy products and fat	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Cream (tablespoon)											
Full-fat yoghurt or Greek style yoghurt (125g carton)											
Dairy desserts (125g carton)											
Cheddar cheese (medium serving)											
Low-fat cheddar cheese (medium serving)											
Cottage cheese											
Eggs as boiled, fried, scrambled, poached (one)											
Quiche (medium serving)											
Light salad cream or light mayonnaise (tablespoon)											
Salad cream, mayonnaise (tablespoon)											
French dressing (tablespoon)											
Other salad dressing (tablespoon)											
The following on bread or vegetables (teaspoon):											
Butter											

Light Butter (e.g. Dawn Lite, Connacht Gold)											
Sunflower margarine (e.g. Flora)											
Low-fat margarine (e.g. low low)											
Cholesterol Lowering (e.g. Flora Pro-Active, Dairy Gold Heart)											
Cream & Veg. Oil spread (e.g. Kerrymaid, Dairy Gold)											
Olive oil spread (e.g. Golden Olive)											
Fruit - 1 fruit or a medium serving (a cupful)	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Apples											
Pears											
Oranges, satsumas, mandarins											
Grapefruit											
Bananas											
Grapes											
Melon											
Peaches, plums											
Apricots											
Strawberries, raspberries, kiwi fruit											
Blueberries											

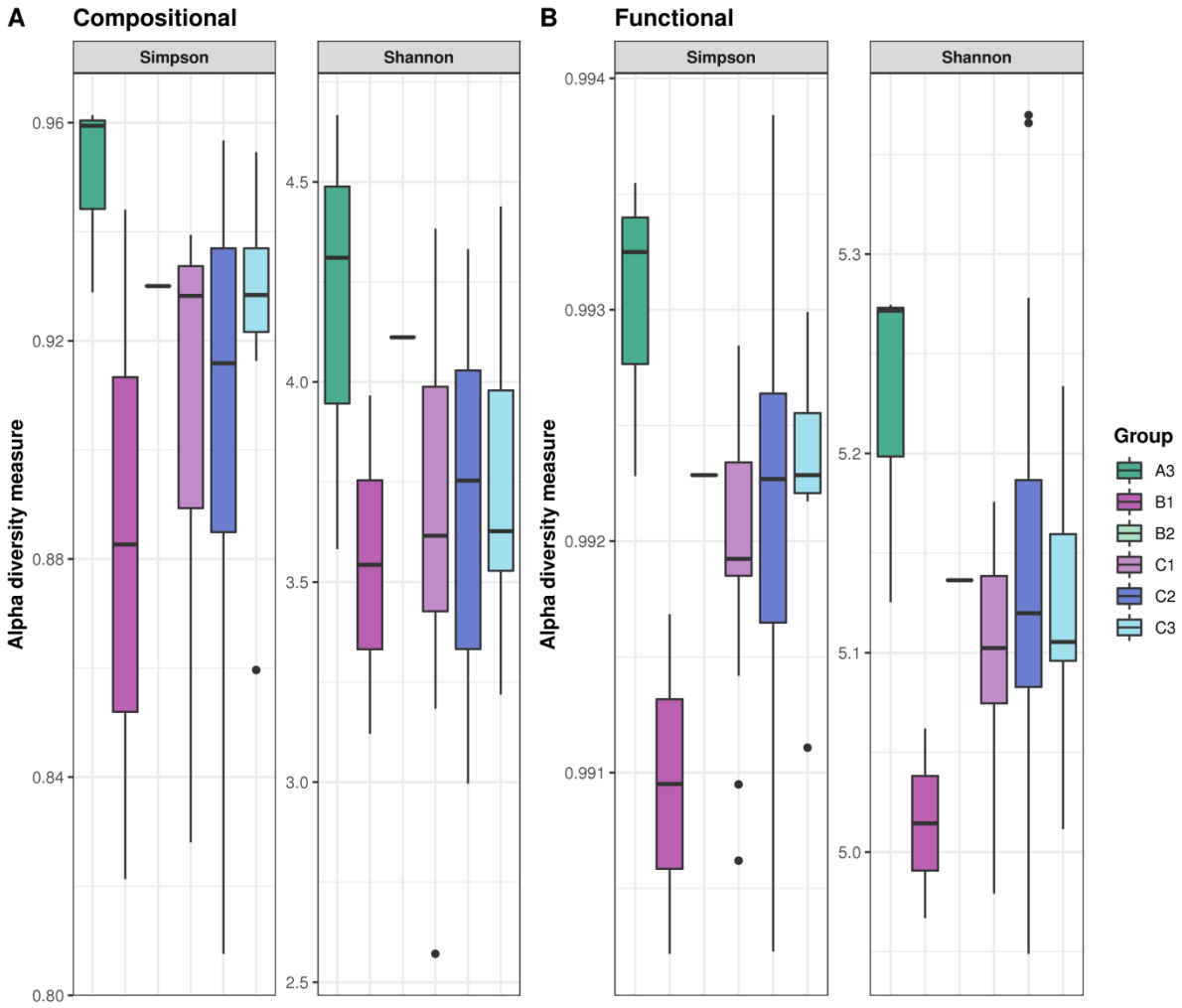
Tinned fruit (specify)											
Dried fruit (e.g. raisins)											
Frozen fruit (specify)											
Vegetables (Fresh, frozen, tinned) - 2 tablespoons	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Carrots											
Spinach											
Broccoli, spring greens, kale											
Brussel sprouts											
Cabbage											
Peas											
Green beans, broad beans, runner beans											
Courgettes											
Cauliflower											
Parsnips, turnips											
Leeks											
Onions											
Garlic											
Mushrooms											

Sweet peppers											
Green salad, lettuce											
Cucumber, celery											
Tomatoes											
Sweetcorn											
Beetroot											
Coleslaw											
Baked beans											
Dried lentils, beans, peas											
Tofu, soya meat, TVP, veggie burger											
Sweets and snacks - medium serving	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Chocolate coated sweet biscuits (e.g. digestive (one))											
Plain biscuit (e.g. marietta, digestives, rich tea (one))											
Cakes (e.g. fruit, sponge)											
Buns, pastries (e.g. croissants, doughnuts)											
Fruit pies, tarts, crumbles											
Sponge puddings											
Milk puddings (e.g. rice, custard, trifle)											

Ice cream, choc ices, frozen desserts											
Chocolates, singles or squares											
Sweets, toffees, mints											
Sugar added to tea, coffee, cereals (teaspoon)											
Sugar substitute (e.g. Canderel (teaspoon))											
Crisps or other packet snacks											
Peanuts or other nuts											
Soups, sauces, and spreads	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Vegetable soups: homemade/fresh (1 bowl)											
Vegetable soups: tinned/packet (1 bowl)											
Meat or cream soups: homemade/fresh (1 bowl)											
Meat or cream soups: tinned/packet (1 bowl)											
Sauces (e.g. white, cheese, gravy (tablespoon))											
Tomato based sauces (e.g. pasta sauces)											
Curry-type sauces											
Pickles, chutney (tablespoon)											
Marmite, Bovril (tablespoon)											
Jam, marmalade, honey, syrup (teaspoon)											

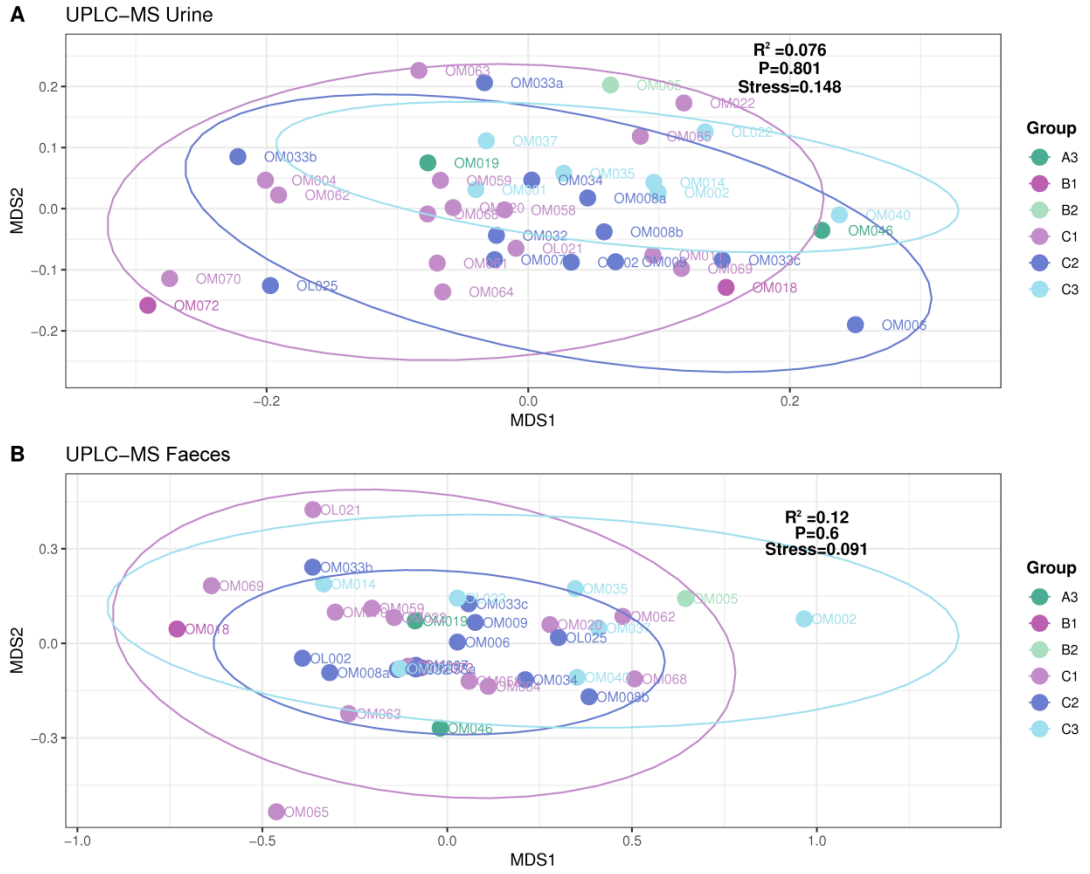
Peanut butter (teaspoon)											
Drinks	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Tea (cup)											
Herbal tea (cup)											
Whole milk (cup) - cow, goat, soya, rice milk - specify											
Semi-skimmed milk (cup)											
Coffee, instant (cup)											
Coffee, ground (cup)											
Coffee, decaffeinated (cup)											
Coffee whitener (e.g. Coffee-mate (teaspoon))											
Cocoa, Hot Chocolate (cup)											
Horlicks, Ovaltine (cup)											
Wine (glass)											
Beer, Larger or Cider (half pint)											
Low alcohol / alcohol free beer / larger (half pint)											
Port, Sherry, Vermouth, liqueurs (glass)											
Spirits (e.g. Gin, Whiskey (Single measure))											
Low calorie or diet soft fizzy drink (glass)											

Fizzy Soft drinks (e.g. Coca Cola (Glass))											
Pure fruit drinks (e.g. orange juice (small glass))											
Fruit squash (small glass)											
Functional Foods / Supplements	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Probiotic Yoghurts / cheese / milk											
Vitamin supplements (details)											
Mineral supplements (details)											
Ready meals	**Never**	**<1 / M**	**1/M**	**1/W**	**2/W**	**5/W**	**1/D**	**2/D**	**4/D**	**6+/D**	**Comment**
Ready meal (specify)											
Takeaway (specify)											



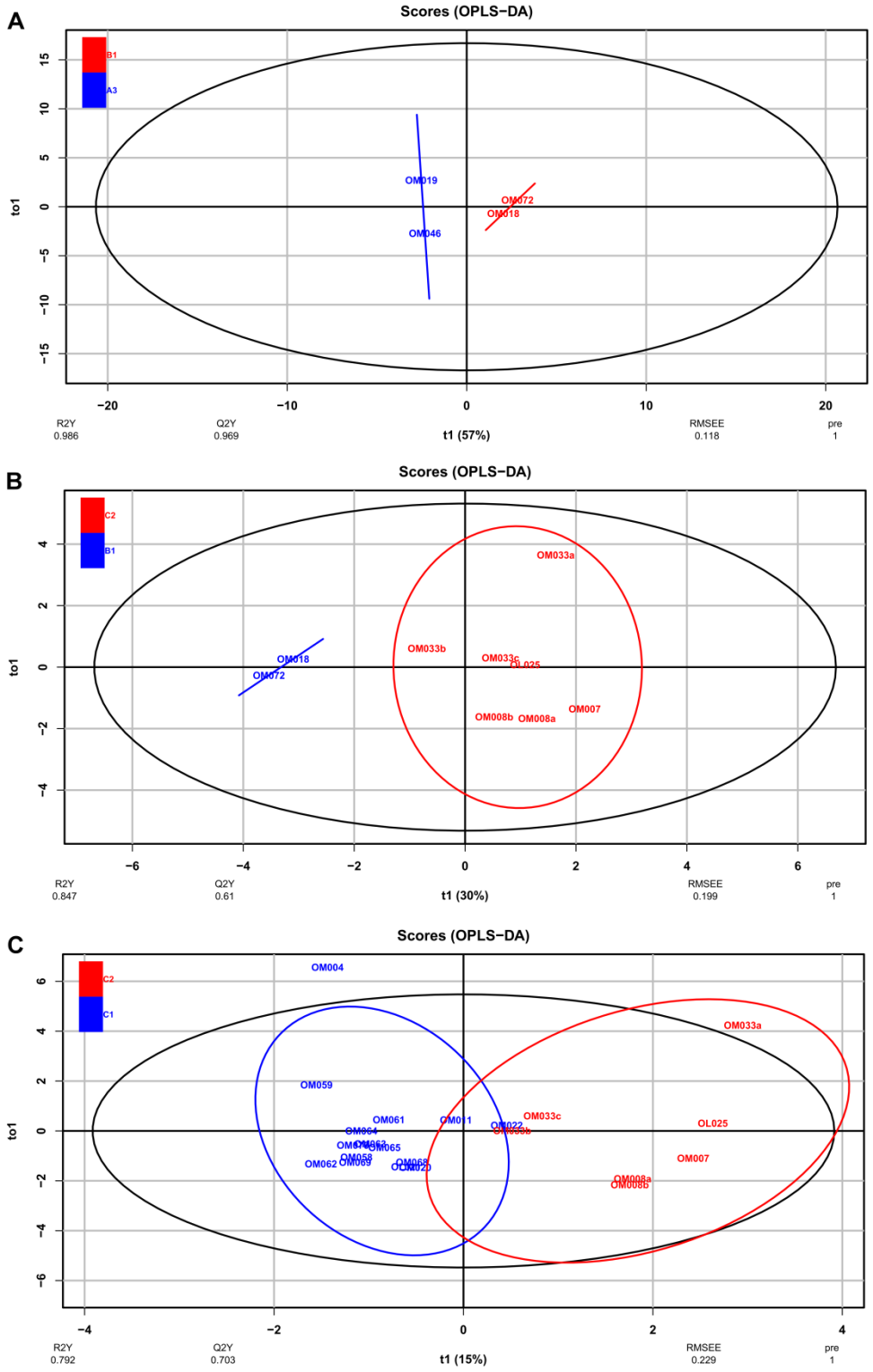
Supplementary figure 4.1| Alpha diversity

Diversity measures of gut microbiome composition (A) and functional potential (B) as determined using vegan in R, coloured by group. Species level assignments from Bracken, following taxonomic classification with Kraken 2, were analysed with vegan to determine Shannon and Simpson diversities for compositional data. Unstratified pathway abundance (reads per million) assignments, as determined using HUMAnN2, were analysed with vegan to determine Shannon and Simpson diversities for functional data. No significant differences were established in alpha diversity by sex, location of sample collection, sport or sports classification group.



Supplementary figure 4.2| Multidimensional scaling (MDS) result from UPLC-MS analysis

Multidimensional scaling (MDS) result from UPLC-MS analysis of (A) urine and (B) faecal samples, as determined using vegan in R, show the differences between metabolites contained in samples based on sports classification group (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups). Clustering of samples was not observed based on sports classification group for UPLC-MS analysis of urine or faecal samples.

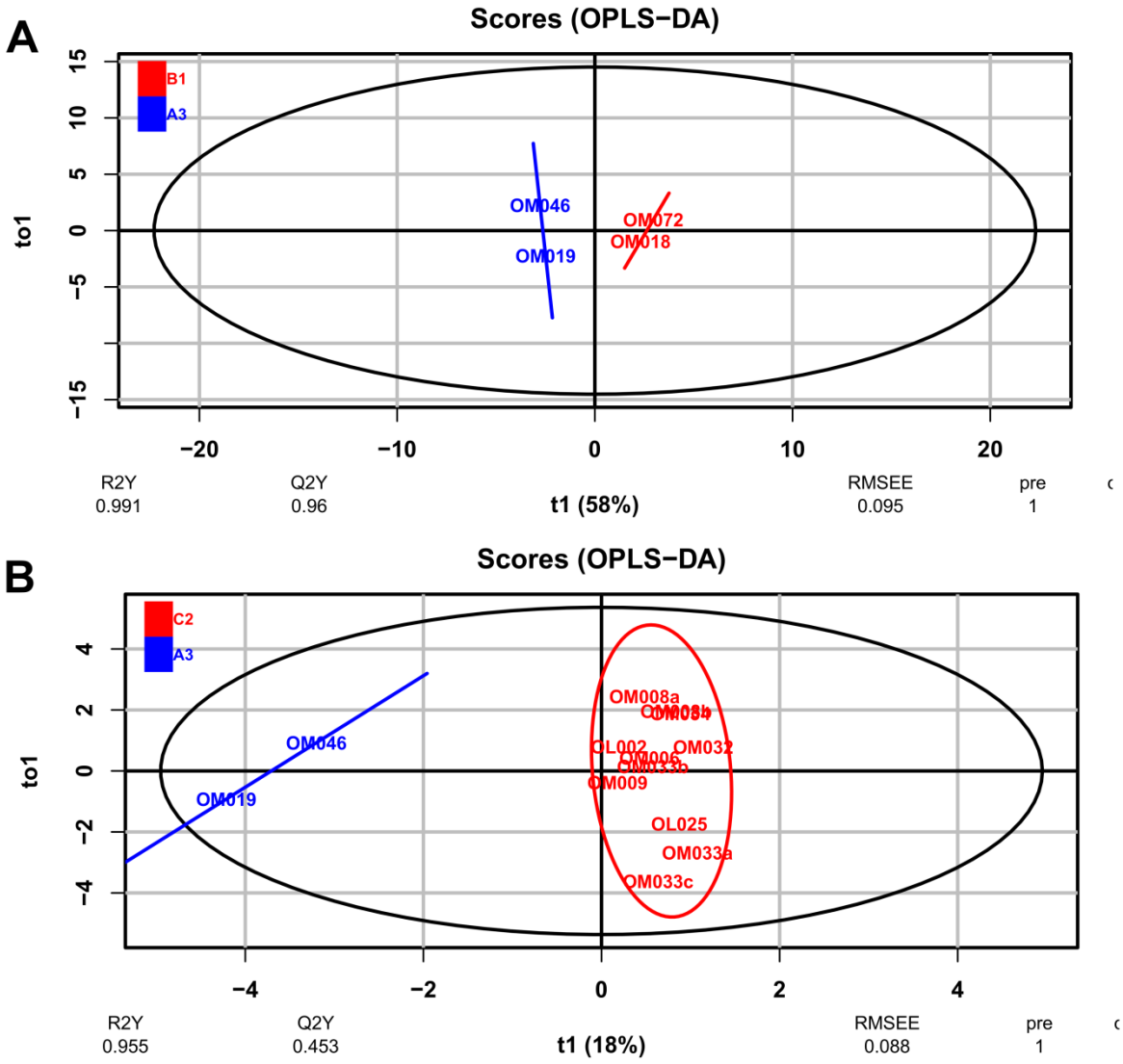


Supplementary figure 4.3| OPLS-DA analysis of GC-MS urine data

OPLS-DA analysis of GC-MS urine data. OPLS-DA analysis was performed using the ropls package in R to analyse the differences between sports classification groups (A3 (N=3), B1

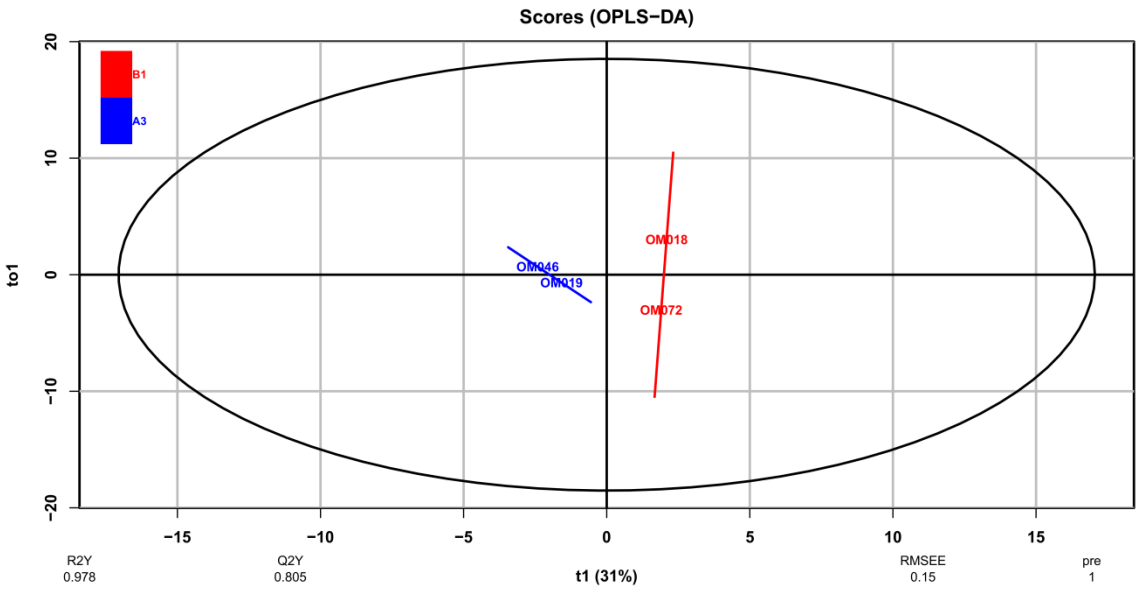
Supplementary figure 4.3 continued

(N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups) using paired analysis.



Supplementary figure 4.4| OPLS-DA analysis of NMR urine data

OPLS-DA analysis of NMR urine data. OPLS-DA analysis was performed using the ropls package in R to analyse the differences between sports classification groups (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups) using paired analysis.



Supplementary figure 4.5| OPLS-DA analysis of NMR faeces data

OPLS-DA analysis of NMR faeces data. OPLS-DA analysis was performed using the ropls package in R to analyse the differences between sports classification groups (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups) using paired analysis.

Supplementary table 4.1| Participant characteristics and data collected

Total participants	37
Total samples	40
Age	27 (±5) years
Sex	
Male	23
Female	14
Location collected	
Brazil	14
Ireland	26
Food frequency questionnaires completed	35 individuals (38 FFQs)
Gut health questionnaires completed	36 individuals (39 GHQs)

Disparity in numbers of individuals involved versus forms collected from arises from the collection of 2 and 3 samples from two participants. Number given for sex is given by participants involved. Numbers given for location is given as samples collected per location.

Supplementary table 4.2| Responses to gut health questionnaire

Sample	SCG	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A.8	A.9	B	C	D	E.1	E.2	E.3	E.4
OL002	C2	1	0	0	0	0	0	2	0	1	6	1	2	2	6	1	2
OL021	C1	1	0	0	0	1	0	0	1	1	6	1	4	4	4	1	1
OL022	C3	1	2	3	1	3	3	0	0	0	0	0	5	3	5	2	4
OL023	A3	1	0	0	1	1	0	0	0	1	6	1	3	2	5	1	4
OL025	C2	1	2	2	3	1	1	0	0	0	5	2	3	1	6	1	2
OM001	C3	0	1	0	1	0	1	0	1	1	6	1	4	1	1	1	4
OM002	C3	0	0	0	0	0	0	0	0	0	6	1	3	6	6	2	2
OM004	C1	0	0	1	1	1	0	0	0	2	5	1	6	1	1	2	1
OM005	B2	0	0	0	0	0	1	0	0	1	6	1	2	1	6	1	1
OM006	C2	0	0	1	0	2	0	0	0	1	6	1	3	4	4	2	1
OM007	C2	2	2	0	1	2	1	0	0	1	6	1	6	1	1	1	2
OM008	C2	1	0	0	3	1	0	1	1	1	5	1	4	1	1	1	1
OM009	C2	0	0	1	1	0	1	1	0	0	4	1	4	1	1	2	5
OM014	C3	0	3	3	0	2	2	0	0	2	7	1	5	4	3	2	5
OM018	B1	0	0	0	0	0	0	0	2	0	5	1	4	0	0	0	2
OM019	A3	2	0	1	1	2	1	1	1	2	6	2	5	1	1	1	1
OM020	C1	0	0	0	0	0	0	0	0	1	6	1	6	1	6	2	1
OM022	C1	0	0	0	2	2	2	0	0	2	5	1	3	1	1	2	1
OM032	C2	0	1	1	1	1	0	0	1	1	1	1	1	2	5	2	1
OM033	C2	0	0	1	1	1	0	0	0	1	6	1	6	1	1	1	2
OM034	C2	0	1	1	1	0	0	0	1	1	3	1	5	4	5	1	1
OM035	C3	1	1	1	1	1	0	0	0	1	5	1	4	3	6	1	4
OM037	C3	2	2	3	3	3	0	3	0	2	6	1	3	1	5	2	4
OM040	C3	0	0	0	0	0	0	0	0	1	5	1	2	1	1	2	2
OM046	A3	0	0	1	1	0	0	0	1	1	6	1	4	2	6	2	4
OM058	C1	0	1	0	0	1	1	0	0	0	5	1	6	2	5	2	2
OM059	C1	1	2	2	1	1	1	0	0	1	6	1	4	2	1	1	2
OM061	C1	0	0	1	1	1	1	0	0	0	6	1	2	1	5	2	3
OM062	C1	1	1	1	0	1	0	0	0	1	6	1	4	1	1	1	1
OM063	C1	0	0	1	1	2	1	0	1	2	6	1	6	2	2	2	5
OM064	C1	0	0	1	0	0	1	0	1	0	6	1	4	3	1	2	1
OM065	C1	0	0	0	0	1	0	1	2	1	5	1	4	1	1	1	2

OM068	C1	0	3	2	1	2	1	1	1	2	6	1	4	4	5	1	2
OM069	C1	1	2	2	2	3	0	1	1	3	6	1	4	5	1	2	2
OM070	C1	1	2	4	2	1	2	0	1	1	6	1	5	5	1	2	2
OM072	B1	0	1	0	1	0	0	0	0	1	5	1	3	4	5	2	1

Supplementary table 4.3| Responses to food frequency questionnaires converted into dietary intake per day using FETA (Food Frequency Questionnaire European Prospective Investigation into Cancer and Nutrition Tool for Analysis)

Sample	Group	Energy (kcal)	Energy (kj)	Protein (g)	Alcohol (g)	Total carbohydrate (g)	Starch (g)	Englyst fibre(g)	Total fat (g)	Monounsaturated fat (g)
OL002	C2	1493.478	6306.804	60.74123	0	225.5353	102.9857	14.65726	45.11437	15.17618
OL021	C1	1715.639	7217.372	93.19006	0	192.3456	99.18692	18.92083	69.21509	27.52418
OL022	C3	1888.233	7942.476	118.7276	0	191.9373	76.02736	34.83842	77.43715	30.41065
OL023	A3	1321.528	5600.232	87.59377	0	175.3467	79.46309	14.60451	35.31568	13.14764
OL025	C2	3990.024	16799.75	140.4593	2.31546	543.2769	372.961	40.6778	153.3358	51.00902
OM001	C3	4291.931	18108.37	210.3562	0.51037	558.7066	240.199	52.09706	150.9348	56.68257
OM002	C3	3512.292	14782.99	185.4709	0	424.2069	230.2033	49.84457	131.3045	47.57845
OM004	C1	1842.056	7748.953	90.47322	6.39354	215.241	116.7682	18.59099	69.92778	27.38079
OM005	B2	1883.538	7910.464	101.4179	0	179.4335	132.3666	16.73356	89.71652	34.84877
OM006	C2	2168.565	9114.684	129.089	0.79296	226.5528	112.0711	19.30676	88.54161	35.41239
OM007	C2	2016.366	8466.752	84.09357	1.55421	219.4016	98.1322	16.74327	94.2642	34.70046
OM008	C2	2834.216	11966.79	129.4316	0	392.7593	199.4686	33.2018	94.10163	30.96027
OM009	C2	1152.377	4855.542	68.98347	0	129.6327	74.31514	10.88483	43.53157	16.50557
OM014	C3	6952.273	29237.11	208.8922	0	955.5956	363.7598	74.18764	282.0124	124.5721
OM018	B1	2618.3	11014.47	141.2109	9.54729	297.1316	144.0235	25.41343	96.95314	36.53338
OM019	A3	2207.022	9335.882	130.4606	1.02074	297.7534	138.5503	29.02582	62.9133	23.52586

OM020	C1	2901.475	12251.65	122.691	9.54729	426.1438	253.4045	26.88654	83.00887	31.15878
OM022	C1	1180.991	4986.754	44.03481	5.63229	172.982	100.5767	11.915	35.42649	13.56054
OM032	C2	3206.134	13376.11	138.2751	0	411.8937	295.9885	64.44134	123.669	35.8003
OM033	C2	3381.732	14218.39	127.4285	0.76125	387.947	197.5658	31.25187	157.5008	54.82821
OM034	C2	906.2678	3844.071	40.24367	0	130.2208	80.59913	31.35441	29.50014	8.614377
OM035	C3	2413.26	10180.79	143.2717	0	305.8168	203.9237	24.66089	77.29722	28.08451
OM040	C3	2343.202	9862.997	115.5013	0	292.4687	122.0018	24.76232	87.27181	31.70655
OM046	A3	1218.72	5140.47	56.57857	0	172.9557	40.328	20.374	38.19818	14.178
OM058	C1	1760.734	7433.998	77.61733	2.06458	258.0672	86.61593	17.60058	52.1544	19.32169
OM059	C1	2736.495	11501.36	182.7983	1.55421	277.4601	166.8093	26.15516	106.173	41.19496
OM061	C1	4836.125	20314.48	234.9095	0.79296	539.0784	255.5545	28.63097	207.5798	76.82948
OM062	C1	1003.021	4194.123	97.39382	0	37.27658	14.58863	7.96673	52.58849	21.49014
OM063	C1	1915.152	8016.182	115.2146	1.30333	163.3399	85.1393	14.86378	92.59298	35.93011
OM064	C1	1914.143	8056.417	97.51229	0	226.4001	86.244	19.14352	75.12152	28.40874
OM065	C1	1250.398	5242.532	48.76549	0.76125	147.8597	49.79852	14.37194	55.08552	19.44113
OM068	C1	2807.31	11789.08	131.6731	0.76125	312.1269	148.1405	23.41961	122.6989	46.25088
OM069	C1	2958.159	12460.98	155.7876	0	388.0921	157.7837	24.481	97.63421	32.59157
OM070	C1	2548.931	10714.37	135.5468	1.55421	280.664	152.1885	18.2877	104.9578	39.86489
OM072	B1	2316.28	9751.198	120.6413	2.09629	260.3463	141.4966	20.07376	93.90436	31.8428

Sample	Polyunsaturated fat (g)	Saturated fat (g)	Calcium (mg)	Iron (mg)	Potassium (mg)	Carotene (µg)	Folate (µg)	Vitamin C (mg)	Vitamin D (µg)	Vitamin E (mg)
OL002	6.607555	19.46861	653.5102	8.42403	2694.161	3261.21	214.1039	123.3001	0.639522	7.636322
OL021	10.24969	25.15792	982.5187	11.48065	3561.007	4525.557	298.5479	161.7157	3.059318	9.208172
OL022	17.44846	22.5751	888.2859	17.47578	4377.074	12151.87	660.7925	311.932	6.697461	14.10344
OL023	7.487339	10.74555	451.0219	8.37457	2963.749	4791.422	249.8603	122.1101	2.54807	5.901533
OL025	27.01639	62.33333	1296.669	25.70848	5466.92	7562.035	527.0272	204.9995	1.768964	19.5079
OM001	27.84383	51.70232	1821.983	27.10793	9262.411	9285.175	820.2434	513.789	7.027051	28.53506
OM002	27.2854	45.0058	1395.486	29.34451	7471.749	9422.008	818.3642	277.0552	4.879898	18.41496
OM004	11.1283	25.4014	684.0874	11.65354	3679.032	3998.233	275.1529	123.5405	2.488532	10.0483
OM005	17.16874	29.82363	616.0178	11.97169	3014.085	4442.587	236.3818	63.33638	9.11267	11.64214
OM006	19.73065	25.50587	833.5338	13.73979	4737.708	6225.453	356.7586	166.3824	6.368729	15.17523
OM007	13.87132	38.03199	713.4339	10.72392	3859.538	3812.31	275.167	123.8617	3.317155	11.23397
OM008	16.33971	37.50063	792.941	18.48298	5578.82	5789.429	410.3697	260.6926	3.219161	15.59915
OM009	7.874536	14.23354	393.5756	8.78868	2089.171	4749.865	239.0707	69.28678	2.69507	5.311583
OM014	48.66951	89.95723	2401.609	42.24465	11701.19	30757.82	1194.599	857.4777	8.64098	50.68468
OM018	17.39745	32.37431	965.3533	18.49497	5090.703	9936.375	494.1035	211.2566	9.516911	16.75223
OM019	14.63945	19.44988	726.7783	16.57099	5375.685	12879.21	521.2582	140.2131	2.17259	15.00347
OM020	14.15107	27.75618	1077.98	18.28193	4087.501	2344.738	356.9241	123.9345	7.242792	10.66825
OM022	7.761302	10.69476	376.1541	8.02955	1791.609	4180.04	162.8951	86.35448	3.268024	7.306001

OM032	34.43753	38.38326	1097.497	28.6844	6775.118	21992.83	846.9921	487.1317	6.132398	27.45328
OM033	26.63095	62.60279	1024.178	20.7709	5459.229	4221.048	419.8544	124.1071	5.247818	16.84357
OM034	6.341829	11.36521	391.4895	10.11827	2462.098	25688	351.7476	98.6288	0.56993	6.942301
OM035	16.8466	23.58857	995.8484	17.9366	4637.702	6940.465	434.2199	136.3891	8.411148	12.82508
OM040	19.42347	29.09645	680.5696	17.42149	4146.711	6633.065	379.6176	141.7186	4.031264	12.01761
OM046	7.574519	12.91393	362.0704	7.06349	3039.443	5620.235	150.7994	167.5329	2.052626	8.668501
OM058	9.156678	19.38207	682.0828	9.84545	3271.417	7276.217	247.2988	173.7731	2.673482	7.597587
OM059	22.19957	32.15194	1186.947	18.82358	5273.034	7739.224	465.437	147.0744	11.00775	15.22003
OM061	39.41893	71.57329	1775.476	22.74964	7023.624	4857.59	450.4408	175.9767	12.37337	31.95709
OM062	10.83323	14.76828	543.5674	8.59315	2593.638	6127.362	263.8016	81.0838	6.858363	7.969218
OM063	20.48513	28.44492	599.2562	10.92873	3337.85	5211.601	284.7291	142.1914	6.46267	13.7092
OM064	16.08325	23.76981	504.4489	11.7861	3733.648	7399.891	257.7122	166.7541	3.762725	13.53674
OM065	12.52557	18.5459	540.0504	7.43028	2492.952	2093.216	177.3326	106.0602	3.101526	10.44662
OM068	22.58592	43.5597	1085.019	16.08986	4734.955	6608.089	381.2853	173.0097	4.698091	17.45539
OM069	12.94129	42.57574	1643.896	14.74132	5931.97	3716.706	395.4653	299.4171	3.821836	10.63415
OM070	19.03106	37.37524	866.1236	15.99932	4339.742	4980.397	313.5483	146.3189	6.022409	12.16616
OM072	15.57822	37.32342	692.7155	13.14678	3846.283	4725.459	297.4924	184.5597	6.778886	11.91808

Sample	Alcoholic beverages (g)	Cereals and cereal products (g)	Eggs and egg dishes (g)	Fats and oils (g)	Fish and fish products (g)	Fruit (g)	Meat and meat products (g)	Milk and milk products (g)	Non-alcoholic beverages (g)	Nuts and seeds (g)
OL002	0	362.85	0	3.08	0	242.45	82.63	216.44	253.68	3.43
OL021	0	346.75	50	0.21	32.06	229.6	133.47	286	388.4	30
OL022	0	252.45	125	4.3	73.52	298.2	110.87	134.16	235.16	42.7
OL023	0	290.41	39.5	1.12	88.93	293.7	135.02	117.18	1046.6	3.43
OL025	37.66	1448.43	21.5	64.34	63.14	93.8	106.22	231.34	1621.38	38.71
OM001	1.61	453.99	39.5	17.05	217.15	846.75	312.18	256.94	1820.9	31.87
OM002	0	797.26	125	10.21	24.01	654	155.25	436	65.2	77.5
OM004	141.34	315	21.5	17.24	24.43	296.05	147.42	179.96	218.7	26.36
OM005	0	397.24	50	2.2	147.91	3.5	161.63	46.22	244.6	26.36
OM006	20.16	249.47	60.5	2.2	112.16	248.25	211.65	203.84	193.04	38.17
OM007	28.91	238.9	39.5	29.09	32.41	234.65	138.79	208.25	599.24	31.87
OM008	0	559.52	21.5	27.94	32.06	742.7	217.07	140.2	886.6	19.21
OM009	0	191.2	125	2.31	0	119.6	104.78	28.84	639.1	10.27
OM014	0	1476.49	42.5	48.33	98.47	711.3	101.53	108.36	1083.84	154
OM018	177.59	320.8	125	38.29	193.5	529.3	164.85	110.74	211.96	8.17
OM019	3.22	412.38	50	10	12.18	544.9	287.5	30.66	1178.76	1.33
OM020	177.59	706.2	50	18.02	137.11	268.45	106.57	162.46	505.6	0

OM022	132.59	404.33	21.5	13.64	15.96	88.15	35.91	48.86	298.4	5.53
OM032	0	759	321	29.72	0	85.4	28.7	4.76	0	0
OM033	8.75	502.65	135.5	47.8	15.96	242	92.45	135.22	1330	89.7
OM034	0	201.11	28	8.74	0	52.5	28.98	0	1140	2.1
OM035	0	790.24	225	0	98.47	266.55	103.45	299.4	1053.4	19.21
OM040	0	511.01	50	18.39	24.01	339	155.76	173.12	401.2	51.7
OM046	0	113.67	10.5	0.91	32.06	845.7	109.02	17.64	537.84	19.21
OM058	30.52	274.73	0	7.49	24.01	351.65	141.44	201.88	448.9	17.11
OM059	28.91	534.55	125	4.9	208.63	325.55	240.17	226.72	172.5	38.71
OM061	20.16	651.92	39.5	30.44	446.35	599.75	315.86	280.12	354.6	15.56
OM062	0	31.64	135.5	0.7	102.67	15.4	139.56	108.36	587.6	21.07
OM063	21.77	193.58	50	12.28	121.99	145.8	185.92	58.1	122.7	31.87
OM064	0	214.38	60.5	18.43	73.81	339.5	155.99	4.76	762.88	27.91
OM065	8.75	140.13	21.5	17.52	36.26	427.6	42.49	124.74	381.9	6.86
OM068	8.75	433.46	71	18.66	36.26	295	196.77	239.5	344.2	38.71
OM069	0	468.18	50	10.49	15.96	956.75	290.09	756	329	6.86
OM070	28.91	417.29	71	26.68	32.41	103	280.93	229.64	913	12.37
OM072	41.93	379.35	50	33.64	61.27	456.05	244.17	70.34	601.98	2.1

Sample	Potatoes (g)	Soups and sauces (g)	Sugars, preserves, and snacks (g)	Vegetables (g)
OL002	52.78	37.8	84.61	225.13
OL021	71.39	40.9	19.89	319.15
OL022	8.75	2.73	38.35	1081.25
OL023	26.39	28	22.48	259.73
OL025	92.18	17.1	67.6	527.55
OM001	137.18	1539.2	23.47	680.21
OM002	330.14	228	16.31	640.18
OM004	129.01	2.1	22.18	293.41
OM005	108.36	49.3	30.4	234.89
OM006	125.86	46.1	24.16	474.13
OM007	162.11	29.05	56.8	279.59
OM008	53.75	51.7	48.22	570.95
OM009	0	14	23.66	250.78
OM014	181.46	86	437.43	1329.28
OM018	52.78	100	38.35	510.9
OM019	312.5	28	87.34	526.14
OM020	52.78	14	49.06	128.12
OM022	17.5	13.53	13.09	165.76
OM032	125.86	4.2	15.89	2399.96

OM033	89.03	28	145.78	428.18
OM034	17.64	32.2	8.12	896.89
OM035	0	30.1	28.69	449.74
OM040	8.75	220.3	72.56	309.34
OM046	0	30.1	62.15	266.14
OM058	52.78	30.1	34.85	242.09
OM059	53.75	44.1	33.68	563.98
OM061	120.26	144.05	137.29	303.87
OM062	0	2.1	0.42	354.99
OM063	63.38	55.9	62.71	398.84
OM064	17.64	253.8	9.84	285.94
OM065	0	72.7	29.12	243.72
OM068	117.11	37	99.07	400.21
OM069	17.64	13.65	18.34	396.89
OM070	47.18	50.45	25.12	341.23
OM072	89.03	66.7	29.45	257.91

Supplementary table 4.4| Dietary intakes compared using Kruskal-Wallis between sports classification groups

Energy, nutrients, and food groups	P value
Energy (kcal)	0.233
Energy (kj)	0.233
Protein (g)	0.249
Alcohol (g)	0.216
Total carbohydrate (g)	0.255
Starch (g)	0.233
Englyst fibre (g)	0.164
Total fat (g)	0.216
Monounsaturated fat (g)	0.216
Polyunsaturated fat (g)	0.216
Saturated fat (g)	0.216
Calcium (mg)	0.216
Iron (mg)	0.164
Potassium (mg)	0.216
Carotene (µg)	0.216
Folate (µg)	0.164
Vitamin C (mg)	0.255
Vitamin D (µg)	0.139
Vitamin E (mg)	0.216
Alcoholic beverages (g)	0.216
Cereals and cereal products (g)	0.216
Eggs and egg dishes (g)	0.588
Fats and oils (g)	0.216
Fish and fish products (g)	0.164
Fruit (g)	0.216
Meat and meat products (g)	0.216
Milk and milk products (g)	0.216
Non-alcoholic beverages (g)	0.216
Nuts and seeds (g)	0.188

Potatoes (g)	0.728
Soups and sauces (g)	0.216
Sugars, preserves, and snacks (g)	0.588
Vegetables (g)	0.164

Dietary intakes compared by sports classification group (A3 (N=3), B (grouped samples with a moderate dynamic component for statistical analysis; N=3), C1 (N=15), C2 (N=11), C3 (N=7)) (See figure 4.1 A for more information regarding groups) using the Kruskal-Wallis test and adjusted using the Benjamini-Hochberg method.

Supplementary table 4.5| Significant differences in metabolites overall

Metabolite	Material	Analysis	P value
Creatinine	Faeces	NMR	0.016979
Pyroglutamic acid	Urine	GC-MS	0.005045
α-hydroxybutyric acid	Urine	GC-MS	0.009892
3-aminoisobutanoic acid	Urine	GC-MS	0.00366
γ-aminobutyric acid	Urine	GC-MS	0.006636
Itaconic acid	Urine	GC-MS	0.007733
4-hydroxybenzouc acid	Urine	GC-MS	0.004581
L-Phenylalanine	Urine	GC-MS	0.006021
L-Tryptophan	Urine	GC-MS	0.003693
3-indolepropionic acid	Urine	GC-MS	0.018406
3-hydroxyphenylacetic acid	Urine	GC-MS	0.004851
Phenyllactic acid	Urine	GC-MS	0.004086
Succinic acid	Urine	GC-MS	0.003413
***Cis*-aconitate**	Urine	GC-MS	0.009654
Succinate	Urine	NMR	0.0072
Hippurate	Urine	NMR	0.007106
Lactate	Urine	NMR	0.00578
Acetate	Urine	NMR	0.0225
Creatinine	Urine	NMR	0.0457
Citrate	Urine	NMR	0.0219
Isoleucine	Urine	UPLC-MS	0.043851

Significantly different metabolites between sports classification groups (A3 (N=3), B1 (N=2), B2 (N=1), C1 (N=15), C2 (N=12), C3 (N=7)) (See figure 4.1 A for more information regarding groups).Statistical analysis was completed between sports classification groups using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied (as shown in figure 4.3 D). All reported p values were adjusted using the Benjamini-Hochberg method.

Chapter 5

Exploring Faecal Bile and Fatty Acids, and Bile Acid Modifications in Elite Athletes

Authors: Ciara M. O' Donovan, Fiona Fouhy, Siobhan F. Clarke, Wiley Barton, Katie O' Connor, Fergus Shanahan, Paul D. Cotter, Susan A. Joyce, and Orla O' Sullivan

Contributions:

- **Candidate** performed functional screening, bile salt hydrolase profiling, plasmid extractions, bile and fatty acid extractions and analysis, and completed bioinformatics and statistical analysis.
- **FF** created the functional fosmid library.
- **SFC** completed original processing of samples and metadata analysis, which has been used for correlations within this chapter.
- **KOC** assisted with the completion of bile salt hydrolase profiling.
- **WB** performed shotgun metagenomic sequencing and bioinformatics analysis on original DNA samples, the data from which has been used for correlations within this chapter.
- **OOS** provided guidance for bioinformatics analysis.
- **SAJ** provided guidance for bile and fatty acid analysis.
- **FS, PDC, SAJ,** and **OOS** supervised the study.

5.0 Abstract

Exercise has been established as a factor that alters the gut microbiome composition and functionality. Functional changes to the microbiome have the potential to impact host metabolism. Bile acids (BAs) are liver produced emulsifiers and enhancers of lipid and lipid-soluble vitamin digestion. The gut microbiome has the capability to modify host produced BAs with the result that BA signatures can be assigned locally and regionally within the GI tract and in enterohepatic circulation. One microbial activity, bile salt hydrolysis is the gateway reaction for microbial modification to produce a palate of BAs. The degree of alteration of these BAs can affect local and global signalling. Many dietary derived medium and long chain fatty acids (FA) are associated with inflammatory responses. The investigation of these host derived and dietary derived metabolites can provide a measure of adjusted microbiome functionality as a consequence of exercise.

This study aimed to examine BA and FA profiles of athletes (N=40) in comparison to low (N=20) and high (N=25) BMI controls using UPLC-MS/MS. This analysis highlighted altered BA and FA profiles in the athletes relative to low and high BMI controls. Overall, conjugated and tauro-conjugated BAs were decreased in the samples from athletes in comparison to controls. Additionally, 3 FAs, namely eicosapentaenoic acid, was found to be significantly increased, and myristic acid and pyruvic acid, were found to be significantly decreased, in athletes compared with controls. In support of these changes, the microbial BA metabolizing ability of a functional fosmid library containing metagenomic DNA from 40 athletes was examined for BA modification ability. A subset of 100 randomly selected BA metabolizing fosmids were independently co-cultured with a BA mixture to identify the spectrum of BSH activity for each. 5 unique clusters of fosmids with similar BSH modification activity were identified and the fosmid insert of a subset (8) of those clones were sequenced to identify the putative genes responsible.

Identification of unique profiles in the BA and FA composition that differ between athletes and controls adds an additional layer of knowledge relating to differences in the metabolic activity of these groups. Additionally, it offers an opportunity to potentially identify desirable athlete BA and FA profiles and their relationship with metabolic and immune functions, and with microbial sequencing data.

5.1 Introduction

Bile acids (BAs) are host produced from cholesterol and are the major component of bile. They are produced in the liver and stored in the gallbladder before being released into the small intestine where they are involved in the digestion and absorption of lipids and lipid soluble vitamins [1]. Additionally, BAs can act as regulatory molecules by activating receptors, including G-protein-coupled BA receptor (TGR5), liver X receptor (LXR), and farnesoid X receptor (FXR) [1, 2]. Differences in BA concentrations relative to controls have been observed in a number of disease states, including diseases of the liver and gallbladder, inflammatory bowel disease, and metabolic syndrome [2, 3]. Exercise has been associated with lower concentrations of BAs both in serum and faeces [4-6], although the mechanism for, and significance of, this is not yet understood.

The majority (95-99%) of BAs released from the gallbladder are reabsorbed for recycling via the enterohepatic circulation system, with the remaining 1-5% entering the colon where they are subjected to microbial modification [6, 7]. Microbial modification of BAs occurs through several processes including hydrolysis, epimerization, esterification, and oxidation [8]. Bile salt hydrolysing (BSH) capabilities are present in species across almost 120 genera in the human microbiome, with seven gut phylotypes of BSH activity identified, mainly driven by an individual's geographical location [9]. While BA modification increases the pool of BAs in the colon, microbial species participate in these modifying functions to reduce toxicity of conjugated BAs to the microbial communities and to obtain carbon, nitrogen, and energy sources from the released amino acid [7]. Freed and secondary BAs act as agonists for receptors, altering host metabolism, and they also have been linked with disease states including colorectal cancer (CRC) [4, 5, 8].

Another group of important metabolites produced through the activity of gut microbes on dietary components are short chain fatty acids (SCFAs) and they have been the focus of

increased attention due their colonic health promoting properties [10, 11]. Medium (MCFAs) and long chain fatty acids (LCFAs) are generally reflective of dietary intake and of microbe-diet interactions, while the proportions and levels of individual fatty acids (FAs) within these groupings has the potential to enhance or moderate inflammatory responses [11, 12].

Here, we investigate the faecal BA and FA composition, reflecting microbial functionality, i.e. microbial host factor modification, and microbial dietary modification, respectively, of athletes previously shown to have a unique gut microbiome composition and functionality [13, 14], relative to controls. In order to more specifically identify microbial activity associated with exercise and BA metabolism, a functional metagenomic bank, created using faecal samples from these athletes, was investigated for its BA modification capabilities.

5.2 Methods

5.2.1 Experimental design

Faecal samples used for bile and fatty acid profiling were those previously collected [13, 14]. Approval was received from the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Cork, Ireland and all volunteers provided written, informed consent. Briefly, faecal samples were collected from male elite professional rugby players (N=40) as well as low (N=20) and high (N=25) BMI controls. The mean age of the athletes and controls was 29 (±4) years and 29 (±6) years, respectively. The athlete group had a mean BMI of 29.1 (±2.9), while the BMI of the low BMI and high BMI controls was ≤25 and >28, respectively. Faecal samples were stored at -80°C before bile and fatty acid profiling.

5.2.2 Bile and fatty acid extraction

Faecal samples were homogenised and 100 mg was weighed and added to sterile microfuge tubes with ceramic beads (Roche, Dublin, Ireland). Samples were softened by the addition of distilled water and subject to bead beading (3x30 s cycles; 4500 xg; cooling at 4°C between cycles). Deuterated standard was added prior to addition of 100% methanol and centrifugation (10,000 xg; 10 min). Resultant supernatant was removed and dried at 45°C for 3 h. Resultant pellet was reconstituted with 5% acetonitrile-formic acid solution before centrifugation (10,000 xg; 10 min). Resultant supernatant was removed and dried at 45°C for 3 h. Dry pellets were reconstituted in 150 µL 50% methanol. The final supernatant containing the extracted acids were transferred to vials which were used for UPLC-MS/MS analysis on Xevo G2 QToF. All samples were run in triplicate.

5.2.3 Ultra-performance liquid chromatography tandem mass spectrometry

All procedures were performed as previously described by Joyce et al. [15]. In short, 5 µL of each sample was injected onto a 50 mm T3 Acquity column (Waters Chromatography Ireland ltd., Dublin, Ireland) and bile and fatty acids were eluted using a 20 min gradient of

100% A to 100% B (A: water, 0.1% formic acid; B: methanol, 0.1% formic acid) at a flow rate of 400 µL min^{-1} and a column temperature of 50°C. Samples were analysed using an Acquity UPLC system (Waters ltd.) coupled online to an LCT Premier mass spectrometer (Waters ltd.) in negative electrospray mode with a scan range of 50–1,000 m/z. Bile acids ionize strongly in negative mode, producing a prominent [M-H]− ion. Capillary voltage was 2.4 Kv, sample cone was 35 V, desolvation temperature was 350°C, source temperature was 120°C, and desolvation gas flow was 900 L h^{-1}. Each analyte was identified according to its mass and retention time. Standard curves were performed using known bile acids and each analyte was quantified according to the standard curve and normalized according to the deuterated internal standards using Masslynx and Targetlynx software (Waters ltd.).

5.2.4 Bile and fatty acid data analysis

Resulting tables were analysed in R (R Foundation for Statistical Computing, Vienna, 2012) (version 3.5.1). Displayed concentrations are means as ng mg^{-1} faecal sample. Multidimensional scaling (MDS) plots were computed in R using vegan (version 2.5-2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Spearman correlations were calculated in R using the RcmdrMisc package (version 1.0-10) and corrected for multiple comparisons using Holms method. Correlations were completed with shotgun metagenomic sequencing data previously generated by Barton et al. [13]. Data visualisations were completed with ggplot2 (version 2.2.1) in R. Heatmaps were created using the pheatmap package in R.

5.2.5 Functional metagenomic bank creation

A functional fosmid library containing athlete faecal metagenomic DNA was previously created, as follows, and used in this study. Forty athlete volunteers recruited were aged 29

(±4) years, free from gastrointestinal disorders, and had not received antibiotic treatment in the previous 6 months. Fresh faecal samples were collected from all volunteers and stored at -80°C before further processing. Faecal samples were weighed, homogenised, and pooled for metagenomic DNA extraction. 250 mg from each sample was taken and pooled to form one sample, from which metagenomic DNA was extracted, using a method optimised for total bacterial genomic DNA. Briefly, the pooled sample was homogenised in PBS and centrifuged at 1000 xg for 5 min. The supernatant was removed and retained and this was repeated 3 times. The supernatant then underwent Nycodenz (Axis Shield, Dundee, UK) density gradient centrifugation separation, to separate out bacterial cells from faecal matter. Bacterial cells underwent enzymatic lysis using lysozyme and mutanolysin (Sigma Aldrich, Dublin, Ireland) followed by protein precipitation using Proteinase K and ammonium acetate (Sigma Aldrich). Bacterial DNA was then precipitated and washed using standard chloroform and ethanol procedures, and eluted in TE buffer. To allow functional screening of the metagenomic DNA, a fosmid metagenomic bank was created using the EpiCentre CopyControl Fosmid Library Production kit with the PCC1FOS vector, as per manufacturer's protocol (Cambio, Cambridge, England). In brief, size selection was performed on the metagenomic DNA using pulse field gel electrophoresis (PFGE) (0.5 ×TAE; 0.1 initial/10 final switch times; 4 V; 17 h; 14°C) and fragments of ~40 Kb were gel extracted from the low melting point agarose (Promega, Medical Supply Company, Dublin, Ireland). DNA fragments were ligated with the pCC1FOS vector as per manufacturer's protocol and packaged into EPI300-T1R *E. coli* plating strain. Cells were then plated onto LB agar plates (Difco, Becton, Dickinson & Co, Oxford, England) containing 12.5 μg mL^{-1} chloramphenicol, IPTG, and Xgal (Sigma Aldrich) and grown overnight aerobically at 37°C. The entire library was plated and then picked and stocked in twenty eight, 384-well plates using the QPix2-XT robotic system (Molecular Devices, Berkshire, UK) and stored at −80°C.

5.2.6 Functional metagenomic bank screening

The twenty eight 384-well plates were screened for bile salt hydrolase activity (BSH), using slight modifications to previously described methods [16-18]. In brief, the library was plated in triplicate onto LB agar plates (Merck, Darmstadt, Germany) containing 12.5 μg mL^{-1} chloramphenicol, 0.037% (weight/volume) calcium chloride, and 0.5% (weight/volume) taurodeoxycholic acid (TDCA; Sigma Aldrich). Plates were incubated aerobically at 37°C and monitored for BSH activity over 4 days. Candidates with halos of precipitated deconjugated bile acids were designated as being positive for BSH activity. Candidates that were positive for BSH activity across three replicate plates were selected and stocked. Overnight cultures of stocked clones were spotted (1 μL) in duplicate onto LB agar as described above in order to verify BSH activity. Subsequently, all clones screened that proved positive for the ability to hydrolyse TDCA were then screened for activity with another bile salt, glycodeoxycholic acid (GDCA; Sigma Aldrich), in triplicate, as described above.

5.2.7 Bile salt hydrolase profiling

A randomly selected subset (100) of the clones that were identified as being positive for BSH activity were carried forward for BSH specific activity profiling. Clones were grown overnight in triplicate in 5 mL of LB broth supplemented with 12.5 μg mL^{-1} chloramphenicol (Sigma Aldrich) and 2 μL mL^{-1} 500x CopyControl fosmid autoinduction solution (Epicentre, Wisconsin, USA). Each overnight culture was normalised to an OD_{600} of 1 and rinsed twice in PBS. 1 mL of bile mixture (LB broth, 0.5% porcine bile spiked with 1 μg mL^{-1} each of TαMCA, TβMCA, and TωMCA) was added to each cell pellet and incubated at 37°C for 2 h. *E. coli* without a fosmid insert was used as a control. Following centrifugation at 16,000 xg for 10 min, the resulting supernatant was split into 2 fresh tubes containing deuterated standard and 100% methanol. Samples were incubated for 30 min before centrifugation at 10,000 xg for 10 min. Supernatant was transferred into fresh tubes and dried at 45°C for 3 h. The resultant pellet was reconstituted with 5% acetonitrile-formic acid solution and

agitated for 1 h. Samples were centrifuged (10,000 xg; 10 min) and the resultant supernatant was removed and dried at 45°C for 3 h. Dry pellets were reconstituted in 150 µL 50% methanol. The final supernatant containing the extracted bile acids were transferred to vials which were used for UPLC-MS/MS analysis on Xevo G2 QTof (Waters ltd.), as described above. All samples were run in triplicate. Each analyte was identified according to its mass and retention time. Standard curves were performed using known bile acids and each analyte was quantified according to the standard curve and normalized according to the deuterated internal standards using Masslynx and Targetlynx software (Waters ltd.). Resulting tables were analysed in R (R Foundation for Statistical Computing, Vienna, 2012) (version 3.5.1). Mean µg mL^{-1} concentrations were calculated and changes from control (*E. coli* EPI300-T1R without a fosmid insert) were calculated and these values were used in further analysis. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Data visualisations were completed with ggplot2 (version 2.2.1) in R. Heatmaps were created using the pheatmap package in R.

5.2.8 Fosmid extraction and sequencing preparation

Eight clones were selected from the 100 clones selected for BSH profiling (table 5.1) for sequencing to identify the fosmid inserts based on their specific activity. Clones were grown shaking overnight (180 rpm) at 37°C in 5 mL of LB broth supplemented with 12.5 µg mL^{-1} chloramphenicol (Sigma Aldrich) and 2 µL mL^{-1} 500x CopyControl fosmid autoinduction solution (Epicentre) to an OD_{600} of between 3 and 4. Fosmids were extracted using the protocol for the FosmidMAX DNA purification kit (Epicentre). Briefly, 1.5 mL of overnight culture was centrifuged at 15,000 xg for 3 min. Resultant pellet was reconstituted in FosmidMAX Solution 1, before the addition of FosmidMAX solutions 2 and 3. Samples were

incubated on ice for 15 min before centrifugation at 15,000 xg for 15 min at 4°C. The resultant supernatant was transferred to a fresh microcentrifuge tube along with 0.6 volumes of room temperature isopropanol before centrifugation at 15,000 xg for 15 min at 4°C. The resultant pellet was air dried before resuspension in TE buffer and the addition of FosmidMAX Solution 4 and incubated on ice for 15 min. Samples were centrifuged at 15,000 xg for 15 min at 4°C. Absolute ethanol was added to the resultant supernatant to precipitate the fosmid DNA. Samples were centrifuged at 15,000 xg for 15 min at 4°C and supernatant was removed. Pellet was dried at room temperature and resuspended in TE buffer. Resultant extract was treated with RNase, as well as Plasmid-Safe ATP-Dependent DNase to remove contaminants. Samples were prepared for Illumina NextSeq shotgun sequencing as per modified Nextera XT DNA library preparation protocol (Illumina, California, USA). Briefly, 0.2 ng μL^{-1} of sample DNA was tagmented for 7.5 min at 55°C. Tagmented DNA was amplified and index adapters (i5 and i7) added. Agencourt AMPure beads (Beckman Coulter, Clare, Ireland) were used to clean up prepared library DNA. DNA samples were diluted to equimolar concentrations before pooling. Prepared library was sequenced on the NextSeq platform using standard Illumina protocols at the Teagasc sequencing facility.

5.2.9 Bioinformatic analysis

Resulting FastQ reads were quality checked to remove poor quality reads, and contaminating human derived and *E. coli* EPI300 reads were removed using NCBI Best Match Tagger (BMTagger). Extracted fosmid inserts were assembled into contigs using Unicycler [19]. Taxonomic classifications were determined using Kraken 2 [20]. Annotations of assembled contigs and identification of putative genes responsible for bile acid modification activity was completed using PATRIC [21].

Table 5.1| Clones picked for sequencing

Clone picked	Cluster on heatmap	Profile
26G21	1	Lower concentrations of conjugated bile acids.
14H18	2	Lower concentrations of most bile acids.
4I22	3	Most similar to control (EPI300 with no fosmid insert).
1D7	4	Around mean concentrations of many bile acids with decreases seen in a few.
2N16	4	Lower concentration of free and secondary bile acids.
1F14	5	Greater concentration of conjugated bile acids.
3L23	5	Greater concentration of most bile acids.
3O10	5	Greater concentration of conjugated bile acids, lower concentration of cholic acid.

The above clones were selected for sequencing as they were found to be representative of the clusters from which they derived (figure 5.5). Reference to lower or greater concentrations of certain BAs is relative to other samples or clusters.

5.3 Results

This study followed on from previous investigations that revealed that the athletes in question, professional rugby players, have a gut microbiome that is distinct from controls, from a compositional, functional potential, and metabolic perspective, including production of higher levels of short chain fatty acids (SCFAs) [13, 14]. In the current study, we aimed to identify differences in the functionality of their microbiota by examining host derived BA and dietary derived FA profiles of athletes from controls, while additionally investigating a functional metagenomic bank created from the athlete faecal samples. This investigation had the potential to identify BA and additional FA signatures that could be representative of elite athletes. It also had the potential to reveal microbial BA modifying genes representative of health and exercise status.

Previously, we established that these athletes have a significantly higher BMI in comparison to low BMI control groups, and a significantly lower BMI in comparison to high BMI controls (supplementary figure 5.1 A). Although athletes had a higher energy intake in comparison to controls and cholesterol intake in comparison to low BMI controls, cholesterol was not significantly different between groups (supplementary figure 5.1 B and C). Blood lipid levels varied between groups with high-density lipoproteins (HDL) significantly higher in the athletes and low BMI control group in comparison to high BMI controls, while triglycerides were significantly higher in the high BMI control group in comparison to athletes and low BMI controls (supplementary figure 5.1 C). Additionally, blood hormone levels measured were significantly different across groups, with adiponectin significantly higher in the low BMI control group in comparison to athletes. Additionally, insulin levels were significantly higher in high BMI controls in comparison to athletes and leptin was significantly higher in the high BMI controls in comparison to both other groups (supplementary figure 5.1 D) [13, 14].

5.3.1 BA receptor agonists are significantly different between groups, while several individual FAs are also significantly different

Multidimensional scaling (MDS) analysis (figure 5.1 A and B) revealed separation of the overall FA (A) (R^2 = 0.05669; P= 0.012; Stress = 0.1583248), but not BA (B) (R^2 = 0.04487; P= 0.092; Stress = 0.02482479) profiles based on groupings. This analysis is based on analysis of 28 BAs and 14 FAs. However, differences in individual BAs were evident. More specifically, the comparison of BA concentrations per group revealed 16 BAs that were significantly different between at least two groups (figure 5.2 A). In all cases concentrations of significantly different BAs were found to be significantly lower in the athlete group relative to at least one of the control groups and, in the case of DHCA, MurCA, THCA, GHCA, GUDCA, and GLCA, and also for the associated amino acid Taurine (see supplementary table 5.1 for abbreviations reference), the concentrations found in the athlete group were significantly lower than those of both control groups.

BAs and FAs are agonists for certain receptors in the human body which regulate their production [1, 22]. We therefore analysed the levels of these BAs and FAs as agonists to reflect the relative levels of receptor activation to investigate whether the driver of the separation between groups could be further established. A heatmap, where BAs and FAs were grouped by their agonist capacity, identified a separation between athletes and controls, but not between the 2 control groups (figure 5.3 A). This separation is driven by a greater concentration of those BAs that target liver X receptor (LXR) and G-protein-coupled BA receptor (TGR5) in the control groups. Specifically, secondary BA, LCA, an agonist for farnesoid X receptor (FXR), LXR, pregnane X receptor (PXR), and vitamin D receptor (VDR), was found at greater concentrations in the high BMI control group in comparison to athletes (figure 5.3 B). Its taurine conjugated form, TLCA, a potent agonist of TGR5, was detected as increased in the high BMI controls in comparison to athletes. Furthermore FA,

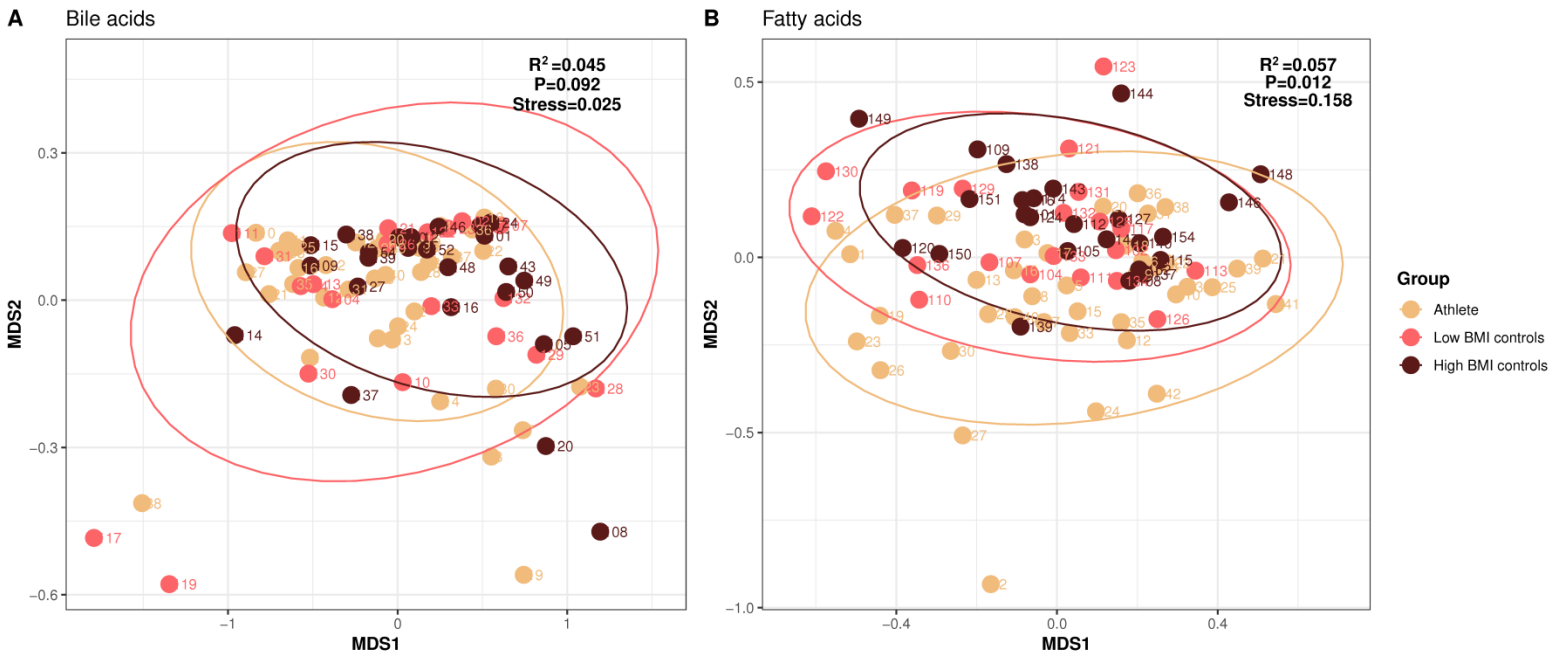


Figure 5.1| Fatty acid but not bile acid profiles are significantly different between groups

Multidimensional Scaling (MDS) analysis depicting bray-curtis dissimilarities between samples, based on bile acid (A) and fatty acid (B) profiles, coloured by group (Athletes (N=40); Low BMI controls (N=20); High BMI controls (N=25)). Bile and fatty acid profiles were determined with UPLC-MS/MS and MDS plots were completed in R (package: vegan). Significant separation was found between groups based on fatty acid (R^2 = 0.05669; P= 0.012; Stress = 0.1583248), but not bile acid profiles (R^2 = 0.04487; P= 0.092; Stress = 0.02482479).

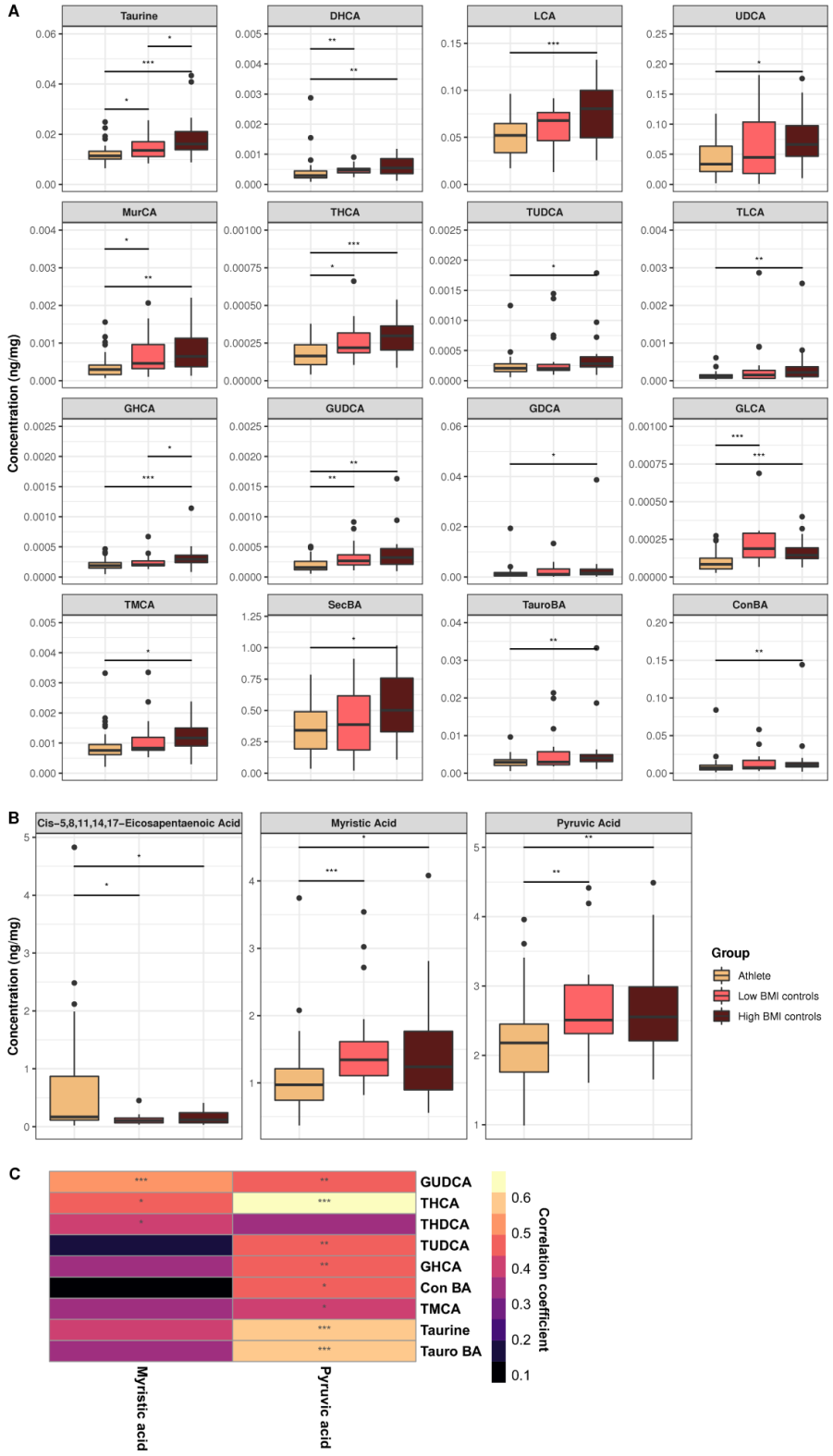


Figure 5.2| Significantly different bile and fatty acids correlate with each other

Bile (A; see supplementary table 5.1 for abbreviations) and fatty (B) acids which are significantly different between at least 2 groups (Athletes vs. low BMI controls, athletes vs.

Figure 5.2 continued

high BMI controls, or high vs. low BMI controls). Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. (C) Heatmap of pairwise correlation values of those bile and fatty acids which were significantly different between at least 2 groups. Spearman correlations were completed in R using the RcmdrMisc package (v2.5-1) and corrected for multiple inferences using Holm's method and only bile and fatty acids with significant correlations ($P \leq 0.05$) are visualised. * $P < 0.05$; ** $P < 0.01$; *** $P \leq 0.001$. Dark colours indicate a correlation coefficient closer to 0 while lighter colours indicate a correlation coefficient closer to 1.

Bile acids; Con BA: conjugated BAs summed, Tauro BA: tauro-conjugated BAs summed, Glyco BA: glyco-conjugated BAs summed, Sec BA: Secondary BAs summed, Pri BA: Primary BAs summed.

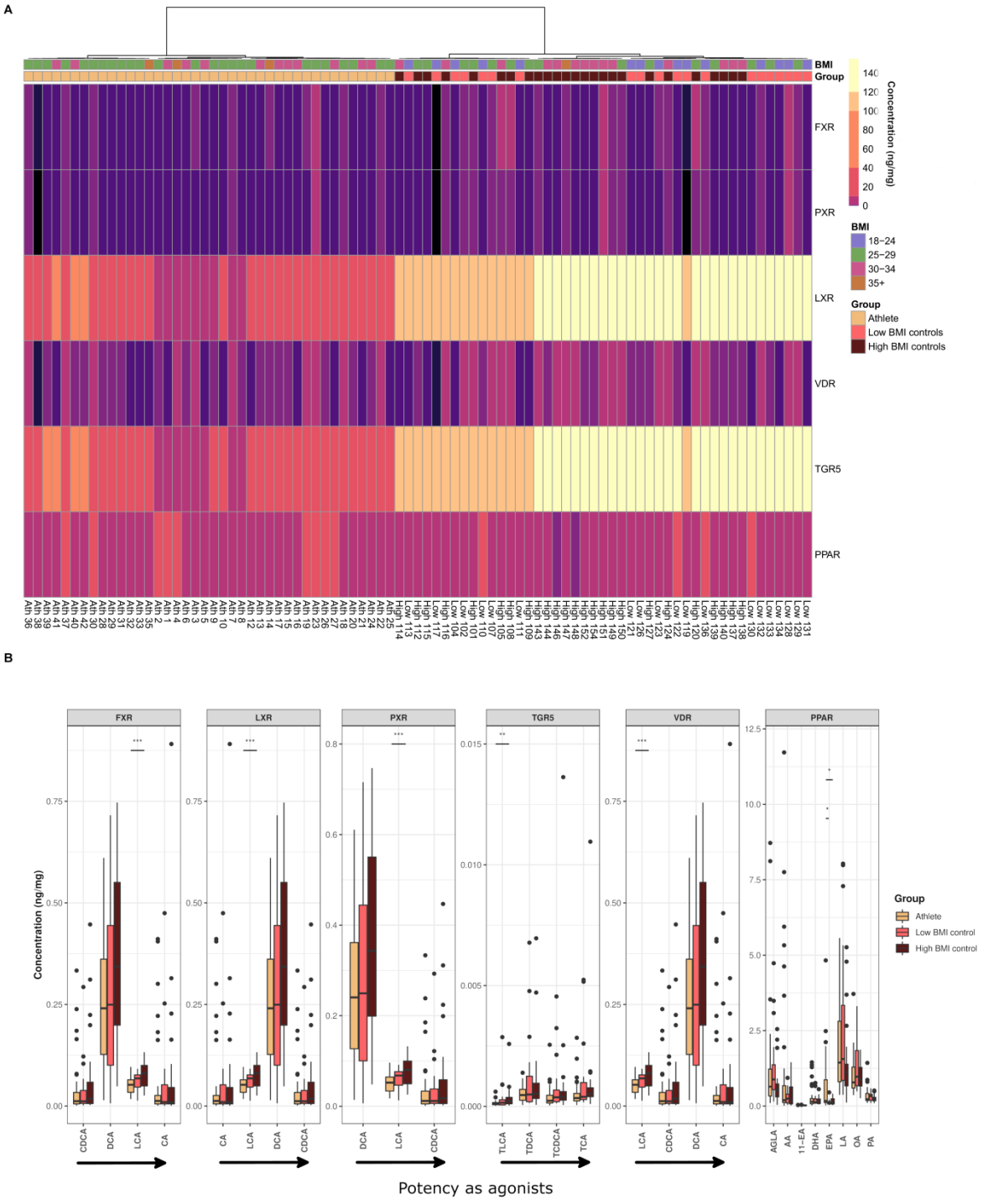


Figure 5.3| Receptor agonists vary by group

(A) Heatmap of bile and fatty acid concentrations summed by the receptor they are agonists for. Athletes (N=40) separated from control groups (Low BMI controls (N=20); High BMI controls (N=25)) based on LXR and TGR5 agonists being in greater amounts in control

Figure 5.3 continued

samples. Control samples did not separate based on grouping. Dark colours indicate lower concentrations (closer to 0) while lighter colours indicate greater concentrations. Heatmaps were drawn in R (package: pheatmap) using Ward clustering.

(B) Bile acid and fatty acid concentrations by group (Athletes (N=40); Low BMI controls (N=20); High BMI controls (N=25)) separated by the receptor they act as agonists for. LCA, an agonist for FXR, LXR, PXR, and VDR, was found to be in increased concentrations in the high BMI control group in comparison to athletes. TLCA, an agonist of TGR5, was found to be increased in the high BMI controls in comparison to athletes. EPA, an agonist of PPAR, was found to be increased in athletes in comparison to both control groups. Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. * $P<0.05$; ** $P<0.01$; *** $P\leq0.001$.

Fatty acids; AGLA: α + γ-Linolenic acid, AA: Arachidonic acid, 11-EA: Cis-11-Eicosenoic acid, DHA: Cis-4,7,10,13,16,19-Docosahexaenoic acid, EPA: Cis-5,8,11,14,17-Eicosapentaenoic acid, LA: Linoleic acid, OA: Oleic acid, PA: Palmitoleic acid.

Bile acids; Con BA: conjugated BAs summed, Tauro BA: tauro-conjugated BAs summed, Glyco BA: glyco-conjugated BAs summed, Sec BA: Secondary BAs summed, Pri BA: Primary BAs summed.

Cis-5,8,11,14,17-Eicosapentaenoic acid (EPA), an agonist of peroxisome proliferator-activated receptors (PPAR), is increased in athletes in comparison to both control groups.

As these receptors are involved with inflammatory responses [23, 24], the inflammatory responses of these groups were of interest. It had previously been determined that inflammatory cytokine levels were significantly different between groups (supplementary figure 5.2). Interleukin 1β (IL-1β) and Interleukin 6 (IL-6) were higher in the high BMI controls compared with athletes. Interleukin 8 (IL-8) was elevated in the athletes compared with both control groups, while tumour necrosis factor α (TNFα) was higher in the high BMI controls compared with both the athletes and low BMI controls [14].

Comparison of individual FA concentrations across groups revealed 3 FAs that differed significantly between at least two groups (figure 5.2 B). Cis-5,8,11,14,17-Eicosapentaenoic acid was detected at significantly greater concentrations in the athlete group relative to samples from both of the other groups. In contrast, myristic and pyruvic acids were present at significantly lower concentrations in athletes relative to both sets of controls. Some, but not all, significantly different BAs and FAs in the datasets correlate with each other (figure 5.2 C). Myristic and pyruvic acid were found to correlate with several conjugated BAs. More specifically, myristic acid correlated with GUDCA, THCA, and THDCA, while pyruvic acid was found to correlate with GUDCA, THCA, TUDCA, GHCA, total conjugated BAs, TMCA, Taurine, and Tauro-conjugated BAs.

From previous shotgun metagenomic analysis of these samples, correlations could be made between individual species and pathways with BAs (table 5.2). GUDCA was correlated with two species; *Methanococcus maripaludis* was significantly negatively correlated with GUDCA (correlation coefficient = -0.49609; p = 0.0295), while *Campylobacter hominis* was significantly positively correlated with GUDCA (correlation coefficient = 0.56465; p =

Table 5.2| Correlations of bile acids with taxa, pathways, and metadata

Bile acid	Taxa/Metadata	Correlation coefficient	P value
GUDCA	*Methanococcus maripaludis*	-0.49609	0.0295
GUDCA	*Campylobacter hominis*	0.56465	0.0319
GLCA	Aspartate transaminase	-0.51221	0.0029
GLCA	Creatine kinase	-0.49635	0.0066

Spearman correlations were calculated in R using the RcmdrMisc package and corrected for multiple comparisons using Holms method. Correlations were completed with shotgun metagenomic sequencing data previously generated by Barton et al. [13] and metadata collected in the study.

0.0319). No correlations were identified between BAs and functional pathways. However, BAs were correlated with other data collected: specifically, aspartate aminotransferase (AST) and creatine kinase (CK) were negatively correlated with GLCA (correlation coefficient = -0.51221; p = 0.0029, and correlation coefficient = -0.49635; p = 0.0066, respectively). No correlations between FAs and any collected metadata were identified.

5.3.2 Investigating the BSH activity of an athlete functional metagenomic bank reveals 5 distinct BSH clusters

Utilising agar based screening methods, 2.1% (227/10,752 clones screened) of an athlete functional metagenomic bank was identified as having the capacity to hydrolase TDCA (tauro-deconjugation), this was determined by the presence of halos of precipitated deconjugated BAs (figure 5.4). Clones that were positive for this activity were then further screened for glyco-deconjugation activity with bile salt, GDCA. This screen identified that 42.7% (97/227) of these tauro-deconjugating candidates also exhibited glyco-deconjugation activity, specifically against GDCA.

High-throughput BSH profiling using a co-culture experiment was completed to develop a profile of a subset of these deconjugaters, namely 100 randomly selected candidates that proved positive for activity against at least TDCA. Five significantly different profile classes, or clusters, of BSH modification were identified from candidate-bile co-cultures analysed following UPLC-MS/MS application (figures 5.5 and 5.6). While all clusters identified were significantly different from each other, cluster 1 and 5 had no BAs significantly different from the controls (supplementary table 5.5), although these clusters had significantly lower or higher concentrations of BAs, respectively, relative to other clusters. It should however be noted that samples from cluster 1 showed alterations to concentrations of several BAs relative to the control, but there was a lack of significance potentially as a consequence of the small number of samples in this cluster (4 samples). Cluster 2 featured reduced

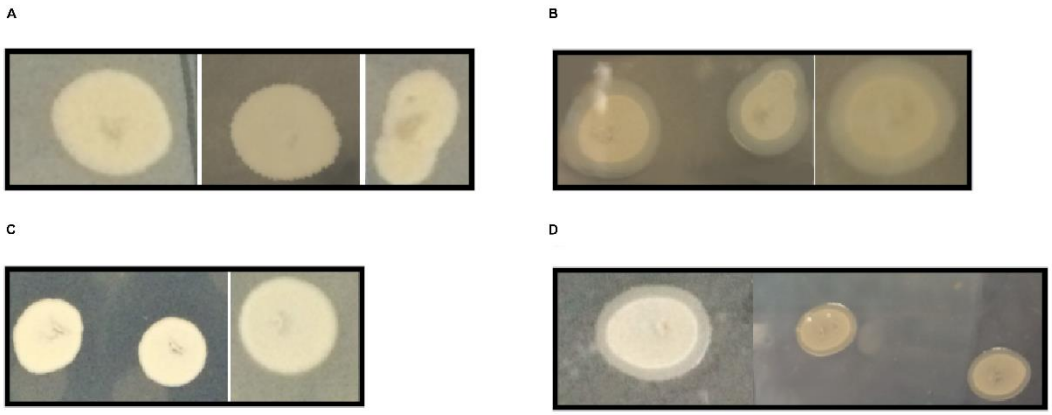


Figure 5.4| Bile salt agar plate assay

A functional metagenomic library was screened for bile salt hydrolase (BSH) activity using taurodeoxycholic acid (TDCA) giving either a negative (A) or positive (B) result for BSH capacity as indicated by halos of precipitated deconjugated bile acids. Clones found to be positive for BSH activity across three replicate plates were subsequently screened for activity with another bile salt, glycodeoxycholic acid (GDCA), again with a negative (C) or positive (D) result.

Figure 5.5| Bile salt hydrolase profiling of a fosmid metagenomic bank reveals 5 distinct groupings

100 clones which were identified as being positive for bile salt hydrolase activity with at least taurodeoxycholic acid (TDCA) were carried forward for bile salt hydrolase profiling, through co-incubation with porcine bile and Tauro α-Muricholic acid (TαMCA), Tauro β-Muricholic acid (TβMCA), and Tauro ω-Muricholic acid (TωMCA). Extracted bile acids from the co-incubation were used for UPLC-MS/MS analysis on Xevo G2 QTol. Mean µg mL^{-1} concentrations were calculated and changes from control (*E. coli* EPI300-T1R without a

Figure 5.5 continued

fosmid insert) were calculated. (A) Heatmap, created using the pheatmap package in R, illustrates five clusters of activity obtained from the co-incubation experiment. Pie charts below the heatmap indicate percentage representation of (B) tauro- and glyco-conjugated bile acids and (C) primary, secondary, and free bile acids within each cluster.

Figure 5.6| Significant differences in bile acids between fosmid groupings

Figure 5.6 continued

Bile salt hydrolase profiling of 100 fosmid clones revealed 5 groupings were present (figure 5.6). All 28 bile acids analysed through this method were significantly different between these fosmid groupings. Significant differences were established between the differences of each fosmid from the control (*E. coli* EPI300-T1R without a fosmid insert). Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. The significance threshold was set at 0.05. * $P<0.05$; ** $P<0.01$; *** $P\leq0.001$.

concentrations of conjugated BAs, relative to the control, with the greatest number of differences between clusters identified between clusters 1 and 2. Samples from cluster 3, the dominant type identified in this analysis (48% of samples), were associated with a greater concentration of free and secondary BAs. Cluster 4 varied most from the control, with increased concentrations of several conjugated BAs, total BAs, and hydrophobic BAs, while hydrophilic BAs were reduced, relative to the control.

At least one clone from each cluster was selected for sequencing analysis (table 5.1). Sequencing analysis indicated that 2 of the resulting fosmid inserts contained *Escherichia coli* DNA and 6 contained DNA from the genus *Staphylococcus*, with 4 specifically assigned as *S. haemolyticus* (Table 5.3). Bile acid 7-α dehydratase BaiE and 2 copies of choloylglycine hydrolase genes were identified in the clones 1D7, 1F14, 2N16, 3L23, and 3O10, when annotated with PATRIC. Investigation of the sequence of these genes revealed 100% identity across clones, with homology also identified between these genes and those identified from other publicly available *Staphylococcus* genomes. However, examination of other regions of these inserts revealed variances between clones. Features of interest related to BSH activity were not annotated in the clones 26G21, 4I22, or 14H18. Additionally, searching for sequences with homology to those identified in the other 5 clones, did not yield any positive hits in the sequences of these clones.

Table 5.3| Classification of sequenced fosmid clones of interest and identification of features of interest

Sample	Taxonomy assigned	Feature(s) of interest identified: PATRIC
26G21	*Escherichia coli*	None identified
14H18	*Staphylococcus*	None identified
4I22	*Escherichia coli*	None identified
1D7	*Staphylococcus haemolyticus*	Choloylglycine hydrolase (EC 3.5.1.24)
		Bile acid 7-α dehydratase BaiE (EC 4.2.1.106)
2N16	*Staphylococcus haemolyticus*	Choloylglycine hydrolase (EC 3.5.1.24)
		Bile acid 7-α dehydratase BaiE (EC 4.2.1.106)
1F14	*Staphylococcus haemolyticus*	Choloylglycine hydrolase (EC 3.5.1.24)
		Bile acid 7-α dehydratase BaiE (EC 4.2.1.106)
3L23	*Staphylococcus haemolyticus*	Choloylglycine hydrolase (EC 3.5.1.24)
		Bile acid 7-α dehydratase BaiE (EC 4.2.1.106)
3O10	*Staphylococcus*	Choloylglycine hydrolase (EC 3.5.1.24)
		Bile acid 7-α dehydratase BaiE (EC 4.2.1.106)

Clones for sequencing selected as representative of the clusters from which they derived (figure 5.5). Taxonomy was assigned with Kraken2. Features were annotated with PATRIC and features of interest related to BA modification are listed.

5.4 Discussion

The gut microbiome of athletes is distinct from more sedentary controls, with differences in the composition and functionality previously identified between the groups that were the focus of this study [13, 14]. Here, we further investigated the variations in the host derived BA and dietary derived FA profiles between athletes and controls. Additionally, we identified the bile altering capacity of an athlete metagenomic bank.

We previously reported that the athletes investigated in this study were classified as clinically obese using BMI, but not their waste:hip ratio, while serum related biomarkers of metabolic health, including HDL and triglycerides, were not identified as being negatively impacted by this status [13, 14, 25]. We identified that although the overall BA profile did not differ between groups, 16 specific BAs were significantly different between groups with the lowest concentrations, of these BAs, identified in the athlete group. In particular, tauro-conjugated BAs were significantly lower in the athlete group. Previously, exercise has been identified as reducing BA concentrations in serum and faeces [4-6, 26], which is further supported here in the athlete cohort. In particular, we identified a reduction in conjugated BAs, an observation that is notable since an increased concentration of these BAs has previously been linked with risk of colorectal cancer (CRC) development and have been found to be greater in patients with type 2 diabetes (T2D) [4, 27]. Furthermore, reduced conjugated BAs, as observed in the current study, have previously been associated with decreased weight gain and reduced cholesterol [15]. Notably the athletes that were the focus of this study showed reduced levels of cholesterol relative to similar BMI controls [14].

BAs have the capacity to act as agonists for a number of receptors including FXR, TGR5, PXR, LXR, and VDR, while FAs have the capacity to act as agonists for PPAR [23, 28-32]. In this study we identified a separation between athletes and controls based on BA

concentrations for those BA which act as agonists for the above receptors. Specifically, a higher concentration of BA agonists for LXR (LCA) and TGR5 (TLCA) were identified in the control groups with respect to the athletes, with a significant difference between athletes and high BMI controls for these BAs. TGR5 is associated with a wide range of functions including glucose homeostasis, conversion of white adipose tissue to brown adipose tissue, and anti-inflammatory responses [33, 34], and other studies have indicated an increase in TGR5 activation in response to exercise [24]. However, a difference in sample collection is noted between these studies. Although BA activation of TGR5 is associated with anti-inflammatory responses, at high physiological concentrations secondary BAs, such as the hydrophobic LCA, can cause oxidative stress and inflammatory responses [35]. As such, the increase in BA agonists of TGR5 observed in the high BMI controls here may be associated, at least in part, with the increase in inflammatory cytokines observed. Similarly, LXR activity, which is also anti-inflammatory and linked to cholesterol efflux from cells [36], has previously been found to be increased with exercise in mouse models [37]. It should be noted that many of the studies relating the influence of exercise on these receptors have been in animal models.

A number of intriguing differences in FAs were also observed. Myristic and pyruvic acids were higher in the high BMI controls compared with low BMI controls and athletes. These FAs have previously been identified as increased in patients with *Clostridium difficile* infections, while increased myristic acid is reported as associated with an increased risk of cardiovascular disease [38, 39]. EPA, identified here in greater concentrations in the athletes compared with both control groups, is a long chain n-3 FA which is commonly identified in high concentrations in individuals who consume a diet rich in fish [40]. Increased concentrations of n-3 FAs can reduce levels of inflammatory cytokines, including TNF-α, indeed reduced TNF-α levels were observed in the athlete cohort in this study, and

this is associated with reduced occurrence of inflammatory diseases such as cardiovascular diseases [40].

GUDCA concentrations were positively correlated with *Campylobacter hominis* and negatively correlated with *Methanococcus maripaludis*. *M. maripaludis* strains have been identified as having a BA sodium symporter [41], potentially supporting their capacity to reduce concentrations of this BA. *C. hominis* has been identified as being capable of secondary BA biosynthesis [42], potentially explaining the correlation observed here.

Concentrations of GLCA, a BA was found to be significantly lower in the athlete cohort in comparison with both control groups, was negatively correlated with aspartate transaminase (AST) and creatine kinase (CK). As faecal BA concentrations have been reported to decrease with increasing exercise [6], a correlation with creatine kinase, a marker of exercise, is not surprising. Similarly, AST, a biomarker for impaired liver function, is reported as increased in athletes compared with non-athletes [43, 44]. Although impaired liver function has previously been associated with an increased concentration of BAs [45], this was not the case here, indicating the further influence of exercise on BA profiles beyond impacts on liver functionality.

Microbial modification of BAs occurs through several processes including hydrolysis, epimerization, esterification, and oxidation [8]. BSH involves the deconjugation of the primary BAs, conjugated to either taurine or glycine, to create secondary BAs [7]. A functional metagenomic bank was employed to investigate the enzymes responsible for BA modification in athletes. It was established, through screening with TDCA, that 2.1% of an athlete functional metagenomic bank had the capacity for BSH modifications. This contrasts with a previous observation that approximately 0.11% of a metagenomic bank was identified as having BSH capacity. However, this older fosmid bank was created using metagenomic DNA from only 1, non-athlete, individual while, in the present study,

metagenomic DNA was sourced from 40 athletes [16]. Here, we identified 5 clusters of clones based on their BSH capacity, varying in their specificity for modifying individual BAs and their activity. The majority of clones investigated in this analysis made small changes to a narrow population of BAs. This finding was similar to the deconjugation capacity identified for several phylotypes by Song et al., where 7 different phylotypes were identified with geography being the main indicator of separation between phylotypes [9]. Interestingly, taurine deconjugation was highly represented (48%) among these fosmids and this activity observed in the faecal BA concentrations separated athletes from control groups. Despite the identification of 5 clusters of clones based on their BA modification capabilities, the 8 representative clones contained inserts that appear to have originated from just 3 sources: *E. coli*, *Staphylococcus*, or *S. haemolyticus*. The introduction of bias may have occurred during the screening of this athlete functional metagenomic bank due to the methodologies used or the capacity of expression of the fosmid insert resulting in the narrowing of the species identified. Higher levels of DCA can be generated by both deconjugation and Bai associated genes, which were found to be present in the fosmids investigated here, which may have resulted in increased visibility of precipitation in the agar assays. Although a reduced taxonomic variability was observed in these clones, the BA modifications observed were varied. While the sequence of the genes was found to be homologous between clones, they were also found to have homology with BA modifying genes from other *Staphylococcus* species, highlighting the homologous nature of these genes which has been previously identified in other species [15, 46, 47]. However, as variance exists in the overall insert sequence for these clones the phenotypic differences may be as a consequence of other genes involved or expression level differences. These taxa have previously been identified as having BA modifying potential, with *Staphylococcus* previously identified to have BSH capabilities, while *E. coli* has previously been identified as having 7α/β-dehydroxylation activity [9, 48, 49]. While bile acid 7-α dehydratase BaiE and

choloylglycine hydrolase genes could be identified in 5 out of 6 *Staphylococcus* clones, no known BA modifying genes were annotated for the remaining *Staphylococcus* clone or the *E. coli* clones, and thus further investigation will be required to identify the relevant contributing genes.

Taken together, this study indicates that the BA and FA markers of metabolic health within the athlete cohort in this study are within healthy ranges. Additionally, the BA profiles differ significantly from the high BMI controls with several differences identified in relation to the low BMI controls also. Specifically, a lower concentration of tauro-conjugated BAs was identified in the athlete cohort, an indication of potential beneficial microbial modifications to reduce BA toxicity within the gut. Overall, this study supports previous findings identifying unique host and microbial patterns with exercise [13, 14]. Future research should focus on identifying the relationship of anti-inflammatory receptor agonists with inflammatory cytokines in athlete cohorts to further understand this relationship. Furthermore, identification of 5 clusters of BSH capacity in the athlete cohort indicates the range of activity possible across individuals. Future research should establish the role of these different phylotypes in relation to health and disease.

5.5 Acknowledgements

The authors thank Fiona Crispie and Laura Finnegan for their contributions to DNA sequencing and Alvaro Lopez Gallardo for his contribution to UPLC-MS/MS. The authors are grateful to all those who donated their time, effort, and samples to this study, in particular staff and players at the Irish Rugby Football Union. This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273, 13/SIRG/2160, 11/PI/1137, and 16/ERA-HDHL/3358.

5.6 References

1. Long, S.L., Gahan, C.G.M., and Joyce, S.A., *Interactions between gut bacteria and bile in health and disease.* Molecular Aspects of Medicine, 2017. **56**: p. 54-65.2. Gerard, P., *Metabolism of cholesterol and bile acids by the gut microbiota.* Pathogens, 2013. **3**(1): p. 14-24.

3. Staley, C., Weingarden, A.R., Khoruts, A., and Sadowsky, M.J., *Interaction of gut microbiota with bile acid metabolism and its influence on disease states.* Applied Microbiology and Biotechnology, 2017. **101**(1): p. 47-64.

4. Hagio, M., Matsumoto, M., Yajima, T., Hara, H., and Ishizuka, S., *Voluntary wheel running exercise and dietary lactose concomitantly reduce proportion of secondary bile acids in rat feces.* Journal of Applied Physiology, 2010. **109**(3): p. 663-668.

5. Danese, E., Salvagno, G.L., Tarperi, C., Negrini, D., Montagnana, M., Festa, L., Sanchis-Gomar, F., Schena, F., and Lippi, G., *Middle-distance running acutely influences the concentration and composition of serum bile acids: Potential implications for cancer risk?* Oncotarget, 2017. **8**(32): p. 52775-52782.

6. Wertheim, B.C., Martinez, M.E., Ashbeck, E.L., Roe, D.J., Jacobs, E.T., Alberts, D.S., and Thompson, P.A., *Physical activity as a determinant of fecal bile acid levels.* Cancer Epidemiology, Biomarkers & Prevention, 2009. **18**(5): p. 1591-1598.

7. Chand, D., Avinash, V.S., Yadav, Y., Pundle, A.V., Suresh, C.G., and Ramasamy, S., *Molecular features of bile salt hydrolases and relevance in human health.* Biochimica et Biophysica Acta, 2017. **1861**(1): p. 2981-2991.

8. Martin, G., Kolida, S., Marchesi, J.R., Want, E., Sidaway, J.E., and Swann, J.R., *In vitro modeling of bile acid processing by the human fecal microbiota.* Frontiers in Microbiology, 2018. **9**: p. 1153.

9. Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., Kalavagunta, P.K., Liao, J., Jin, L., Shang, J., and Li, J., *Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome.* Microbiome, 2019. **7**(1): p. 9.

10. Lordan, C., Thapa, D., Ross, R.P., and Cotter, P.D., *Potential for enriching next-generation health-promoting gut bacteria through prebiotics and other dietary components.* Gut Microbes, 2019: p. 1-20.

11. Wan, J., Hu, S., Ni, K., Chang, G., Sun, X., and Yu, L., *Characterisation of fecal soap fatty acids, calcium contents, bacterial community and short-chain fatty acids in Sprague Dawley rats fed with different sn-2 palmitic triacylglycerols diets.* PLoS One, 2016. **11**(10): p. e0164894.

12. Fritsche, K.L., *The science of fatty acids and inflammation.* Advances in Nutrition, 2015. **6**(3): p. 293S-301S.

13. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.

14. 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., and Cotter, P.D., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.

15. Joyce, S.A., MacSharry, J., Casey, P.G., Kinsella, M., Murphy, E.F., Shanahan, F., Hill, C., and Gahan, C.G., *Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(20): p. 7421-7426.

16. Jones, B.V., Begley, M., Hill, C., Gahan, C.G., and Marchesi, J.R., *Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(36): p. 13580-13585.

17. Christiaens, H., Leer, R.J., Pouwels, P.H., and Verstraete, W., *Cloning and expression of a conjugated bile acid hydrolase gene from Lactobacillus plantarum by using a direct plate assay.* Applied and Environmental Microbiology, 1992. **58**(12): p. 3792-3798.

18. Dashkevicz, M.P. and Feighner, S.D., *Development of a differential medium for bile salt hydrolase-active Lactobacillus spp.* Applied and Environmental Microbiology, 1989. **55**(1): p. 11-16.

19. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E., *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* PLoS Computational Biology, 2017. **13**(6): p. e1005595.

20. Wood, D.E. and Salzberg, S.L., *Kraken: Ultrafast metagenomic sequence classification using exact alignments.* Genome Biology, 2014. **15**(3): p. R46.

21. Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., Gerdes, S., Henry, C.S., Kenyon, R.W., Machi, D., Mao, C., Nordberg, E.K., Olsen, G.J., Murphy-Olson, D.E., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H., and Stevens, R.L., *Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center.* Nucleic Acids Research, 2017. **45**(D1): p. D535-D542.

22. Song, P., Rockwell, C.E., Cui, J.Y., and Klaassen, C.D., *Individual bile acids have differential effects on bile acid signaling in mice.* Toxicology and Applied Pharmacology, 2015. **283**(1): p. 57-64.

23. Varga, T., Czimmerer, Z., and Nagy, L., *PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation.* Biochimica et Biophysica Acta, 2011. **1812**(8): p. 1007-1022.

24. Morville, T., Sahl, R.E., Trammell, S.A., Svenningsen, J.S., Gillum, M.P., Helge, J.W., and Clemmensen, C., *Divergent effects of resistance and endurance exercise on plasma bile acids, FGF19, and FGF21 in humans.* JCI Insight, 2018. **3**(15).

25. Meza, M.N. and Carrillo, J.A.B., *Biomarkers, obesity and cardiovascular diseases*, in *Role of biomarkers in medicine*, Wang, M. and Witzmann, F.A., Editors. 2016. InTech: Croatia.

26. Suarez-Garcia, S., Del Bas, J.M., Caimari, A., Escorihuela, R.M., Arola, L., and Suarez, M., *Impact of a cafeteria diet and daily physical training on the rat serum metabolome.* PLoS One, 2017. **12**(2): p. e0171970.

27. Wewalka, M., Patti, M.E., Barbato, C., Houten, S.M., and Goldfine, A.B., *Fasting serum taurine-conjugated bile acids are elevated in type 2 diabetes and do not change with intensification of insulin.* The Journal of Clinical Endocrinology and Metabolism, 2014. **99**(4): p. 1442-1451.

28. Li, T. and Chiang, J.Y., *Bile acid signaling in metabolic disease and drug therapy.* Pharmacological Reviews, 2014. **66**(4): p. 948-983.

29. Makishima, M., Lu, T.T., Xie, W., Whitfield, G.K., Domoto, H., Evans, R.M., Haussler, M.R., and Mangelsdorf, D.J., *Vitamin D receptor as an intestinal bile acid sensor.* Science, 2002. **296**(5571): p. 1313-1316.

30. Song, C. and Liao, S., *Cholestenoic acid is a naturally occurring ligand for liver X receptor alpha.* Endocrinology, 2000. **141**(11): p. 4180-4184.

31. Xie, W., Radominska-Pandya, A., Shi, Y., Simon, C.M., Nelson, M.C., Ong, E.S., Waxman, D.J., and Evans, R.M., *An essential role for nuclear receptors SXR/PXR in detoxification of cholestatic bile acids.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(6): p. 3375-3380.

32. Parks, D.J., Blanchard, S.G., Bledsoe, R.K., Chandra, G., Consler, T.G., Kliewer, S.A., Stimmel, J.B., Willson, T.M., Zavacki, A.M., Moore, D.D., and Lehmann, J.M., *Bile acids: Natural ligands for an orphan nuclear receptor.* Science, 1999. **284**(5418): p. 1365-1368.

33. Guo, C., Chen, W.D., and Wang, Y.D., *TGR5, not only a metabolic regulator.* Frontiers in Physiology, 2016. **7**: p. 646.

34. Chavez-Talavera, O., Tailleux, A., Lefebvre, P., and Staels, B., *Bile acid control of metabolism and inflammation in obesity, type 2 diabetes, dyslipidemia, and nonalcoholic fatty liver disease.* Gastroenterology, 2017. **152**(7): p. 1679-1694 e1673.

35. Zeng, H., Umar, S., Rust, B., Lazarova, D., and Bordonaro, M., *Secondary bile acids and short chain fatty acids in the colon: A focus on colonic microbiome, cell proliferation, inflammation, and cancer.* International Journal of Molecular Sciences, 2019. **20**(5).

36. Baranowski, M., *Biological role of liver X receptors.* Journal of Physiology and Pharmacology, 2008. **59 Suppl 7**: p. 31-55.

37. Kazeminasab, F., Marandi, S.M., Ghaedi, K., Esfarjani, F., and Moshtaghian, J., *Endurance training and LXR activation increased LXRα and ABCA1 transcripts in Wistar male rats.* Science & Sports, 2018. **33**(3): p. 152-157.

38. Brown, J.R., Flemer, B., Joyce, S.A., Zulquernain, A., Sheehan, D., Shanahan, F., and O'Toole, P.W., *Changes in microbiota composition, bile and fatty acid metabolism, in successful faecal microbiota transplantation for Clostridioides difficile infection.* BMC Gastroenterology, 2018. **18**(1): p. 131.

39. Le Goff, C., Kauz, J.F., Leroy, L., Pincemail, J., Chapelle, J.P., and Cavalier, E., *Fatty acids and associated cardiovascular risk.* Food and Nutrition Sciences, 2013. **4**: p. 188-194.

40. Institute of Medicine (US) Committee on Nutrition, Trauma, and the Brain, *Eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA)*, in *Nutrition and traumatic brain injury: Improving acute and subacute health outcomes in military personnel*, Erdman, J., Oria, M., and Pillsbury, L., Editors. 2011, National Academies Press (US): Washington (DC).

41. UniProt, C., *Uniprot: A worldwide hub of protein knowledge.* Nucleic Acids Research, 2019. **47**(D1): p. D506-D515.

42. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M., *KEGG as a reference resource for gene and protein annotation.* Nucleic Acids Research, 2016. **44**(D1): p. D457-462.

43. Nikolaidis, M.G., Protosygellou, M.D., Petridou, A., Tsalis, G., Tsigilis, N., and Mougios, V., *Hematologic and biochemical profile of juvenile and adult athletes of both sexes: Implications for clinical evaluation.* International Journal of Sports Medicine, 2003. **24**(7): p. 506-511.

44. Raastad, T. and Hallen, J., *Recovery of skeletal muscle contractility after high- and moderate-intensity strength exercise.* European Journal of Applied Physiology, 2000. **82**(3): p. 206-214.

45. Luo, L., Aubrecht, J., Li, D., Warner, R.L., Johnson, K.J., Kenny, J., and Colangelo, J.L., *Assessment of serum bile acid profiles as biomarkers of liver injury and liver disease in humans.* PLoS One, 2018. **13**(3): p. e0193824.

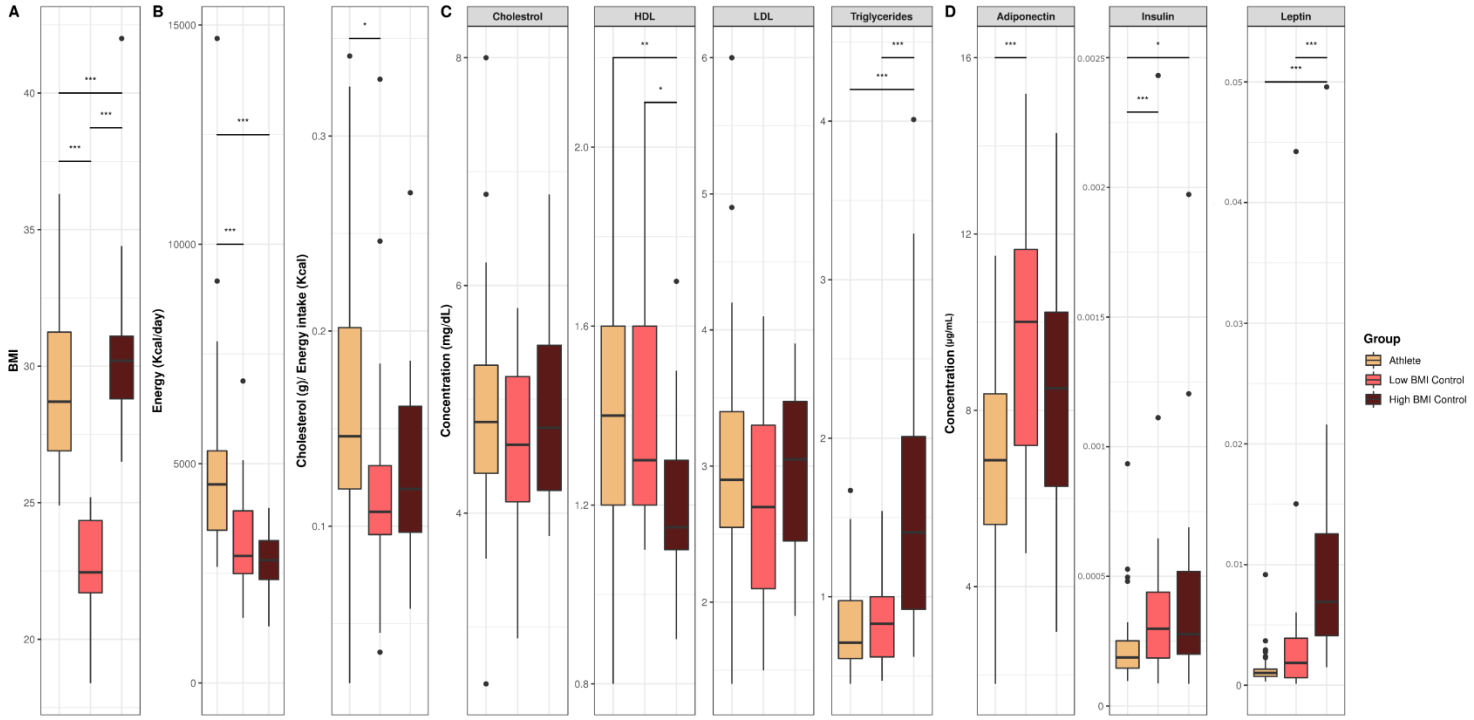
46. Kim, G.B., Miyamoto, C.M., Meighen, E.A., and Lee, B.H., *Cloning and characterization of the bile salt hydrolase genes (BSH) from Bifidobacterium bifidum strains.* Applied and Environmental Microbiology, 2004. **70**(9): p. 5603-5612.

47. Harris, H.M.B., Bourin, M.J.B., Claesson, M.J., and O'Toole, P.W., *Phylogenomics and comparative genomics of Lactobacillus salivarius, a mammalian gut commensal.* Microbial Genomics, 2017. **3**(8): p. e000115.

48. Mukherji, R. and Prabhune, A., *Possible correlation between bile salt hydrolysis and AHL deamidation: Staphylococcus epidermidis RM1, a potent quorum quencher and bile salt hydrolase producer.* Applied Biochemistry and Biotechnology, 2015. **176**(1): p. 140-150.

49. Ridlon, J.M., Kang, D.J., and Hylemon, P.B., *Bile salt biotransformations by human intestinal bacteria.* Journal of Lipid Research, 2006. **47**(2): p. 241-259.

5.7 Supplementary material



Supplementary figure 5.1| Characteristics of participants by group

(A) BMI (Body Mass Index) of participants by group. Athletes (N=40) have a significantly higher BMI in comparison to low BMI (N=20) participants and significantly lower BMI in comparison to high BMI (N=25) participants.

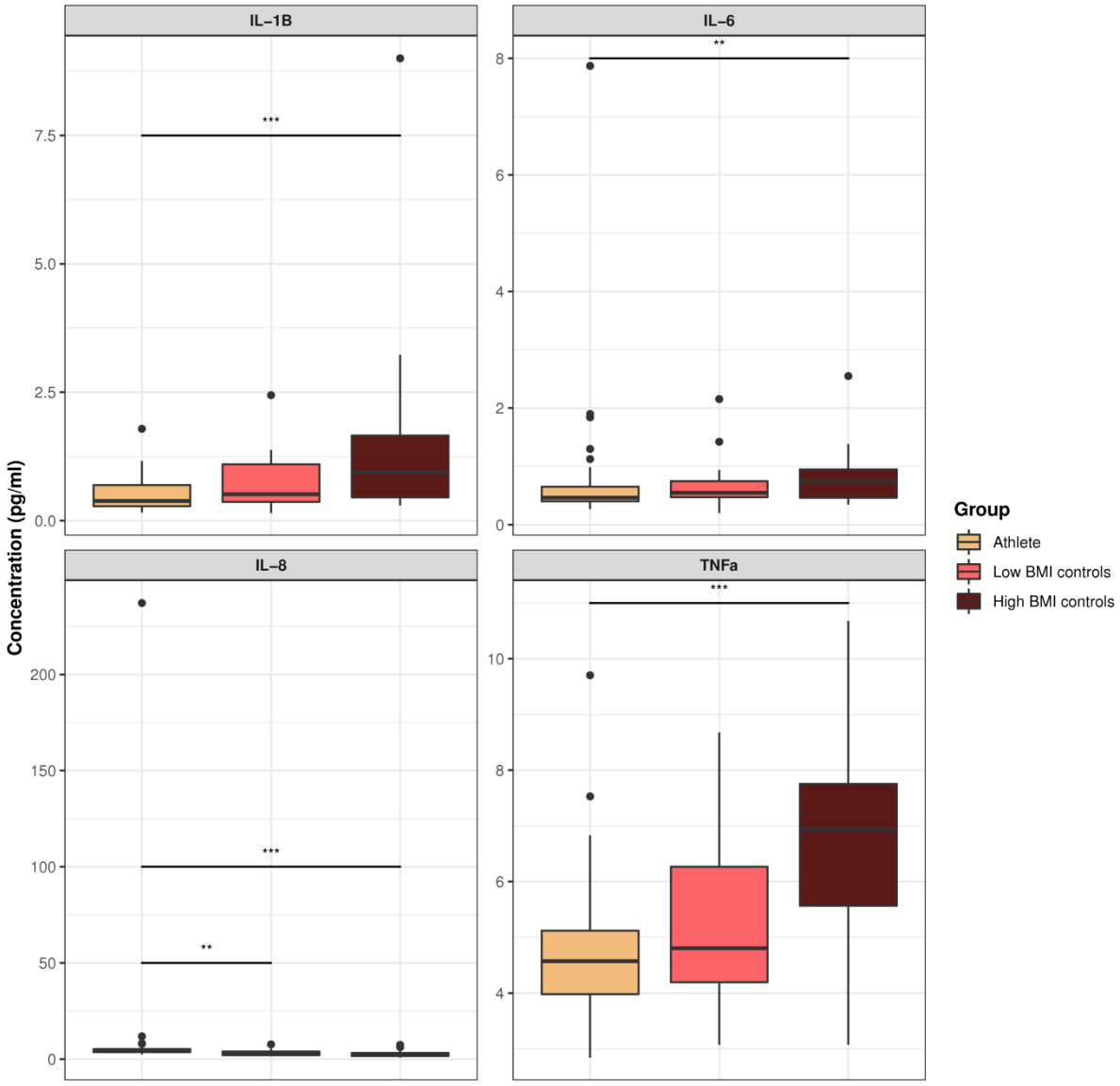
Supplementary figure 5.1 continued

(B) Energy (kcal/day) and cholesterol (g)/Energy (kcal) intake per day by groups. Significantly greater energy and cholesterol intake in athletes in comparison to low BMI controls. Additionally, significantly higher energy intake (kcal) in athletes than high BMI controls.

(C) Cholesterol, high-density lipoproteins (HDL), low-density lipoproteins (LDL), and triglyceride concentrations in blood per group. HDL is significantly higher in athletes and low BMI controls in comparison to high BMI controls. Triglycerides are significantly higher in high BMI controls in comparison to athletes and low BMI controls.

(D) Hormone level concentrations in blood per group. Adiponectin is significantly lower in athletes in comparison to low BMI controls. Insulin is significantly lower in athletes in comparison to high BMI controls. Leptin is significantly lower in athletes in comparison to low and high BMI controls [13, 14].

Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. * $P<0.05$; ** $P<0.01$; *** $P\leq0.001$.



Supplementary figure 5.2| Significantly different inflammatory cytokines between groups

Significantly different inflammatory cytokines, as measured from blood serum. Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. The significance threshold was set at 0.05. * P<0.05; ** P<0.01; *** P≤0.001 [13, 14].

IL-1β: Interleukin 1 beta; IL-6: Interleukin 6, IL-8: Interleukin 8; TNFα: Tumour Necrosis Factor alpha.

Supplementary table 5.1| Bile acid moieties analysed by UPLC-MS in this study

Analyte	Abbreviation	$[MW]^{-H}$	Neutral Formula	RT	R^2
Taurine	Taurine	124.0068	$C_2H_7NO_3S$	0.81	0.986
Dehydrocholic Acid	DHCA	401.2328	$C_{24}H_{34}O_5$	1.77	0.987
Lithocholic Acid	LCA	375.2899	$C_{24}H_{40}O_3$	21.65	0.995
Murocholic Acid	MurCA	391.2848	$C_{24}H_{40}O_4$	6.69	0.997
Ursodeoxycholic Acid	UDCA	391.2848	$C_{24}H_{40}O_4$	8.32	0.998
Hyodeoxycholic Acid	HDCA	391.2848	$C_{24}H_{40}O_4$	10.14	0.995
Chenodeoxycholic Acid	CDCA	391.2848	$C_{24}H_{40}O_4$	17.15	0.998
Deoxycholic Acid	DCA	391.2848	$C_{24}H_{40}O_4$	17.82	0.998
α-Muricholic Acid	α-MCA	407.2797	$C_{24}H_{40}O_5$	5.59	0.998
β-Muricholic Acid	β-MCA	407.2797	$C_{24}H_{40}O_5$	5.99	0.999
ω-Muricholic Acid	ω-MCA	407.2797	$C_{24}H_{40}O_5$	5.59	0.998
Hyocholic Acid	HCA	407.2797	$C_{24}H_{40}O_5$	8.6	0.994
Cholic Acid	CA	407.2797	$C_{24}H_{40}O_5$	11.22	0.998
7-Ketolithocholic Acid	7-KLCA	389.2692	$C_{24}H_{38}O_4$	9.47	0.999
Tauroursodeoxycholic Acid	TUDCA	498.2889	$C_{26}H_{45}NO_6S$	2.89	0.997
Taurohyodeoxycholic Acid	THDCA	498.2889	$C_{26}H_{45}NO_6S$	3.36	0.992
Taurochenodeoxycholic Acid	TCDCA	498.2889	$C_{26}H_{45}NO_6S$	8.15	0.997
Taurodeoxycholic Acid	TDCA	498.2889	$C_{26}H_{45}NO_6S$	9.14	0.996
Tauro α-Muricholic Acid	α-TMCA	514.2838	$C_{26}H_{45}NO_7S$	1.85	0.999
Tauro β-Muricholic Acid	β-TMCA	514.2838	$C_{26}H_{45}NO_7S$	1.85	0.999
Tauro ω-Muricholic Acid	ω-TMCA	514.2838	$C_{26}H_{45}NO_7S$	1.85	0.999
Taurohyocholic Acid	THCA	514.2838	$C_{26}H_{45}NO_7S$	2.83	0.999
Taurocholic Acid	TCA	514.2838	$C_{26}H_{45}NO_7S$	4.64	0.995
Taurolithocholic Acid	TLCA	482.2940	$C_{26}H_{45}NO_5S$	14.19	0.993
Glycoursodeoxycholic Acid	GUDCA	448.3063	$C_{26}H_{43}NO_5$	3.23	0.981
Glycohyodeoxycholic Acid	GHDCA	448.3063	$C_{26}H_{43}NO_5$	3.82	0.998
Glycochenodeoxycholic Acid	GCDCA	448.3063	$C_{26}H_{43}NO_5$	9.01	0.995
Glycodeoxycholic Acid	GDCA	448.3063	$C_{26}H_{43}NO_5$	10.23	0.993
Glycohyocholic Acid	GHCA	464.3012	$C_{26}H_{43}NO_6$	3.19	0.995
Glycocholic Acid	GCA	464.3012	$C_{26}H_{43}NO_6$	5.33	0.998
Glycolithocholic Acid	GLCA	432.3114	$C_{26}H_{43}NO_4$	15.07	0.997

Cholic Acid d4	CA d4	411.3049	$C_{24}H_{36}D_4O_5$	11.22	0
Chenodeoxycholic Acid d4	CDCA d4	395.3099	$C_{24}H_{36}D_4O_4$	17.15	0
Deoxycholic Acid d4	DCA d4	395.3099	$C_{24}H_{36}D_4O_4$	17.78	0

RT: column retention time.

Supplementary table 5.2| Fatty acid moieties analysed by UPLC-MS in this study

Analyte	[MW]-H	Neutral Formula	RT	R^2
Pyruvic acid	87.0082	$C_3H_4O_3$	0.85	0.977
Lactic acid	89.0239	$C_3H_6O_3$	0.83	0.994
Heptanoic acid	129.0916	$C_7H_{14}O_2$	2.73	0.99
Octanoic acid	143.1072	$C_8H_{16}O_2$	4.56	0.992
Myristic acid	227.2011	$C_{14}H_{28}O_2$	22.16	0.835
Palmitoleic acid	253.2168	$C_{16}H_{30}O_2$	22.28	0.978
Palmitic acid	255.2324	$C_{16}H_{32}O_2$	22.56	0.801
ɣ-Linolenic acid	277.2168	$C_{18}H_{30}O_2$	22.13	0.979
α-Linolenic acid	277.2168	$C_{18}H_{30}O_2$	22.13	0.979
Linoleic acid	279.2324	$C_{18}H_{32}O_2$	22.4	0.99
Oleic acid	281.2481	$C_{18}H_{34}O_2$	22.64	0.992
Cis-5,8,11,14,17-Eicosapentaenoic acid	301.2168	$C_{20}H_{30}O_2$	22.21	0.983
Arachidonic acid	303.2324	$C_{20}H_{32}O_2$	22.36	0.991
Cis-11-Eicosenoic acid	309.2794	$C_{20}H_{38}O_2$	22.96	0.889
Cis-4,7,10,13,16,19-Docosahexaenoic acid	327.2324	$C_{22}H_{32}O_2$	22.14	0.993
Cholic acid d4	411.3049	$C_{24}H_{36}D_4O_5$	12.55	0
Chenodeoxycholic acid d4	395.3099	$C_{24}H_{36}D_4O_4$	18.11	0

RT: column (25 m) retention time.

Supplementary table 5.3| Statistics between groups for all bile acids analysed

Bile acid	Athlete vs. High BMI Controls	Athlete vs. Low BMI controls	High BMI vs. Low BMI controls
Taurine	**<0.001**	**0.041**	**0.028**
DHCA	**0.005**	**0.005**	0.337
LCA	**<0.001**	0.203	0.142
HDCA	0.74	0.863	0.74
CDCA	0.829	0.875	0.829
UDCA	**0.02**	0.627	0.425
DCA	0.087	0.913	0.256
MurCA	**0.003**	**0.014**	0.411
CA	0.964	0.964	0.964
HCA	0.631	0.631	0.631
TCA	0.053	0.332	0.424
THCA	**<0.001**	**0.017**	0.315
TCDCA	0.149	0.149	0.873
TUDCA	**0.024**	0.802	0.087
TDCA	0.393	0.648	0.648
THDCA	0.106	0.107	0.767
TLCA	**0.003**	0.224	0.315

GCA	0.425	0.485	0.664
GHCA	**0.001**	0.198	**0.012**
GHDCA	0.439	0.605	0.378
GCDCA	0.073	0.315	0.315
GUDCA	**0.003**	**0.007**	0.508
GDCA	**0.017**	0.406	0.339
GLCA	**<0.001**	**<0.001**	0.201
α + ω MCA	0.925	0.925	0.925
β MCA	0.618	0.618	0.618
MCA	0.978	0.784	0.784
TMCA	**0.012**	0.115	0.126
Total BAs	0.125	0.79	0.245
Pri. BAs	0.863	0.863	0.863
Sec. BAs	**0.023**	0.551	0.245
Free BAs	0.125	0.766	0.256
Tauro. BAs	**0.008**	0.323	0.398
Glyco. Bas	0.084	0.315	0.315
Con. BAs	**0.009**	0.347	0.347

Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Significant differences are

highlighted in bold. Bile acids; Con BA: conjugated BAs summed, Tauro BA: tauro-conjugated BAs summed, Glyco BA: glyco-conjugated BAs summed, Sec BA: Secondary BAs summed, Pri BA: Primary BAs summed.

Supplementary table 5.4| Statistics between groups for all fatty acids analysed

Compound	Athlete vs. High BMI controls	Athlete vs. Low BMI controls	High BMI vs. Low BMI controls
α and γ Linolenic Acid	0.337	0.433	0.141
Arachidonic Acid	0.732	0.732	0.732
Cis-11-Eicosenoic Acid	0.386	0.271	0.749
Cis-4,7,10,13,16,19-Docosahexaenoic Acid	0.861	0.861	0.861
Cis-5,8,11,14,17-Eicosapentaenoic Acid	**0.048**	**0.017**	0.508
Heptanoic Acid	0.127	0.084	0.706
Lactic Acid	0.2	0.465	0.465
Linoleic Acid	0.216	0.707	0.216
Myristic Acid	**0.032**	**0.001**	0.398
Octanoic Acid	0.078	0.058	0.411
Oleic Acid	0.716	0.597	0.597
Palmitic Acid	0.395	0.395	0.891
Palmitoleic Acid	0.676	0.676	0.45
Pyruvic Acid	**0.006**	**0.006**	0.945

Groups were compared using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Significant differences are highlighted in bold.

Supplementary table 5.5| Significant differences between heatmap clusters and controls

Bile acid	Cluster 1 vs. Controls	Cluster 2 vs. Controls	Cluster 3 vs. Controls	Cluster 4 vs. Controls	Cluster 5 vs. Control
DCA	0.096	0.342	**0.038**	0.362	0.9
CA	0.08	**0.043**	**0.033**	0.06	0.151
TCDCA	0.074	**0.039**	**0.039**	**0.039**	0.055
GCA	1	1	1	**0.044**	0.163
GHCA	0.074	**0.037**	**0.037**	**0.035**	0.14
GCDCA	0.078	0.078	0.729	**0.048**	0.078
GUDCA	0.074	**0.039**	0.299	**0.035**	0.14
Total BAs	0.08	0.061	0.255	**0.035**	0.083
Pri. BAs	0.08	0.154	**0.033**	0.06	0.117
Free BAs	1	1	0.345	**0.048**	0.283
Con. BAs	0.08	0.061	0.255	**0.035**	0.083
Tauro. BAs	0.084	**0.043**	0.084	**0.039**	0.09
Glyco. BAs	0.074	0.056	0.4	**0.044**	0.083
CDCA family	0.069	**0.039**	**0.039**	**0.039**	0.068
Hydrophilic BAs	0.074	0.056	0.4	**0.044**	0.109
Hydrophobic BAs	0.074	**0.043**	0.058	**0.039**	0.073

Individual clusters as established with a heatmap (figure 5.5) were compared with controls (E. coli EPI300-T1R without a fosmid insert) using the Kruskal-Wallis test, with a Mann-Whitney U test performed for significant results to determine the groups between which this difference applied. P values were corrected for multiple comparisons using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Significant differences are highlighted in bold. Bile acids; Con BA: conjugated BAs summed, Tauro BA: tauro-conjugated BAs summed, Glyco BA: glyco-conjugated BAs summed, Sec BA: Secondary BAs summed, Pri BA: Primary BAs summed.

Chapter 6

The Gut Microbiome of Runners Reflects Maximal Oxygen Consumption (VO_2 max) and the Amount of Training per Week

Authors: Ciara M. O' Donovan, Fergus Shanahan, Paul D. Cotter, and Orla O' Sullivan

Contributions:

- **Candidate** extracted DNA from collected faecal samples, prepared samples for shotgun metagenomic sequencing, and completed bioinformatics and statistical analysis.
- **OOS** provided guidance for bioinformatics analysis.
- **FS, PDC,** and **OOS** supervised the study.

6.0 Abstract

Gut microbiome composition and functionality can be influenced by a range of environmental and lifestyle factors. There is increasing evidence that exercise and, ultimately, fitness represent such factors. This study employed shotgun metagenomic sequencing to investigate the composition and functionality, on 4 occasions across a 7 week period, of the gut microbiome of amateur, club level, runners (n=62), as well the factors that were associated with features of these microbial communities.

Analysis revealed the composition and functional potential of the gut microbiome remained consistent within individual athletes over time. However, when the athletes were grouped on the basis of their VO_2 max and training frequency, associations with microbial composition and functionality were observed. More specifically, a number of gut species differed significantly between the VO_2 max or exercise frequency groups. *Ruminococcus albus* and several *Prevotella* species levels were proportional to increasing VO_2 max, with *Prevotella* species also reflecting increased training frequency. *Akkermansia muciniphila* was greatest at the highest VO_2 max grouping, while proportions of *Veilonella rodentium* were higher among individuals with a good VO_2 max relative to all other groups. From a functional perspective, a decrease in branched chain amino acid synthesis was found with increased exercise frequency, while a decrease in overall amino acid synthesis was found with increased VO_2 max. Overall, this study reveals that the composition and functionality of the gut microbiome in an active cohort reflects differences in VO_2 max and exercise frequency.

6.1 Introduction

The composition and functionality of an individual's gut microbiome has been found to be relatively stable over time [1-3]. However, this can be influenced by some environmental and lifestyle factors, including exercise [2]. More specifically, the gut microbiome of athletes and those who participate in regular exercise has been found to differ from that of more sedentary individuals [4-6]. In particular, an increased abundance of specific taxa, including *Veillonella*, *Akkermansia*, *Faecalibacterium prausnitzii*, or *Ruminococcus*, has been noted among some athlete cohorts [4-6]. In addition to these patterns, individual taxa, including *Coprococcus* and, more recently, *Veillonella*, have also been found to be increased following specific sporting endeavours, for example following a marathon [7, 8], while exercise frequency was noted to be positively associated with *Prevotella* abundance among a cycling cohort [9].

Cardiorespiratory fitness is measured by maximal oxygen consumption (VO_2 max) and has been found to be predictive of aspects of gut microbial functionality. In particular, increased alpha diversity and butyrate production has been associated with an increasing VO_2 max, in healthy adults [10]. Increasing VO_2 max indicates increased cardiorespiratory fitness, which decreases the risk of several diseases, including cardiovascular related diseases, and is also indicative of performance potential. Several factors are involved with an increased VO_2 max including increased blood volume and capillary density [11, 12], and it may be that these, or other physiological changes associated with increased VO_2 max, contribute to the gut microbiota associations observed. Additionally, an increase in exercise intensity towards an individual's VO_2 max can have an impact on intestinal permeability [13], which in turn can influence gut microbiota composition [14].

To further investigate these phenomena, this study explored the extent to which a variety of factors, including VO_2 max and training frequency, were associated with gut

permeability, gut microbiome composition and functional potential across a cohort of club level runners.

6.2 Methods

6.2.1 Ethical approval

The samples used in this study were collected in accordance with the ethical principles set forth in the current version of the Declaration of Helsinki and the International Conference on Harmonization E6 Good Clinical Practice (ICH-GCP). Approval was received from the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Cork, Ireland and all volunteers provided written, informed consent.

6.2.2 Study design

Faecal samples were collected from individuals (N=62; table 6.1), who participated in at least 4 hours of endurance sports per week, with a minimum 1.5 hours of the training being a running activity, over the course of 7 weeks (4 time points; supplementary figure 6.1). Subjects were instructed to maintain their diet and lifestyle, and refrain from consuming non-steroidal anti-inflammatory drugs and any products that may contain live microbes during the study (see supplementary table 6.1 for inclusion and exclusion criteria). For 2 days before each of the study visits, participants were instructed to refrain from consuming alcohol, spicy foods, caffeine, and artificial sugars, and avoid strenuous exercise. A 2 day food and fluid intake was recorded before the first study visit and participants were instructed to standardize dietary intake the 2 days before each subsequent visit using this record. A standardized dinner and evening snack was provided for the day before the study visit, and a standardized breakfast for the day of the study visit (consumed 2.5 hours before the study visit) following overnight fasting from 22:00. No food was consumed during the 5 hour study visit.

6.2.3 Maximal oxygen consumption test

An incremental test was performed in order to measure maximal oxygen consumption (VO_2 max) and this was utilized to subsequently establish the minimum running velocity required

to achieve VO_2 max (vVO_2 max). The test was performed on a motorised treadmill with an initial speed of 10 km h^{-1}, with an increase of 1.0 km h^{-1} every 2 min until subjects achieved volitional exhaustion [15]. During the entire test treadmill inclination was set at 1% [16]. Respiratory gases were measured with an online respiratory system and VO_2 max was determined as the highest oxygen uptake per minute measured during the test and vVO_2 max was determined as the running speed at the time of the determined VO_2 max peak. Samples were classified based on their VO_2 max into 4 groups (fair, good, excellent, and superior) based on previous classifications (table 6.2) [17,18].

6.2.4 Exercise challenge to induce intestinal permeability

At visit 3 and visit 5 subjects performed a strenuous physical exercise challenge consisting of treadmill running at a speed corresponding to 80% of vVO_2 max for 60 min (1% inclination). At a similar intensity, well-trained runners manage just over 60 min in a time to exhaustion test [19]. If the subject was unable to keep this intensity for the full 60 min, the intensity was lowered in 1 km h^{-1} increments and exercise time was prolonged so that the distance covered was equal to running 60 min at 80% of vVO_2max. The 150 mL of sugar solution for the lactulose/rhamnose (L/R) permeability test was initiated orally during a very short running break 30 min into the exercise challenge.

6.2.5 Faecal sample collection and DNA extraction

Faecal samples were collected by participants in faecal tubes as the first bowel movement post-permeability test (visit 2 and 4) or post-exercise test (visits 3 and 5). If samples were collected at home they were transported on transportable cooling boxes to ensure the samples were frozen until they were returned to the trial site. Samples were stored at -80°C before DNA extraction.

Faecal samples were homogenised and DNA extraction was completed using the repeated bead beating plus column (RBBC) method, as previously described [20]. Briefly, 0.25 g of

faecal sample was added into a sterile 2 mL screw-cap tube containing 0.25 g of zirconia beads (0.25 g of 0.1 mm, 0.125 g of 1.5 mm, and a single 2.5 mm bead). Following the addition of 1 mL of lysis buffer (500 mM NaCl, 50 mM Tris-HCl, 50 mM EDTA, and 4% sodium dodecyl sulphate (SDS)) samples were homogenized for 3 min on a Mini Beadbeater (BioSpec, Bartlesville, USA). Samples were incubated at 70°C for 15 min to lyse the cells. Following centrifugation (16,000 xg; 5 min; 4°C) supernatant was removed and the bead beating steps were repeated, this time using 200 µL of lysis buffer. Supernatant from both bead beating steps was pooled and 10 M ammonium acetate was added. Samples were centrifuged (16,000 xg; 10 min; 4°C) and resulting supernatant was treated with one volume of isopropanol. DNA was pelleted and washed with 70% ethanol before being dissolved in Tris-EDTA. DNA was RNase and proteinase K treated. DNA was finally washed with 100% ethanol and buffers AW1 and AW2 (QIAamp DNA stool mini kit, Qiagen, England) before being eluted in 200 µL AE buffer (Qiagen). DNA was stored at -20°C before library preparation.

6.2.6 DNA library preparation and sequencing

Samples were prepared for Illumina NextSeq shotgun sequencing as per modified Nextera XT DNA library preparation protocol (Illumina, California, USA). Briefly, 0.2 ng μL^{-1} of sample DNA was tagmented for 7.5 min at 55°C. Tagmented DNA was amplified and index adapters (i5 and i7) were added. Agencourt AMPure beads (Beckman Coulter, Clare, Ireland) were used to clean up prepared library DNA. DNA samples were diluted to equimolar concentrations before pooling. Prepared library was sequenced on the NextSeq platform using standard Illumina protocols at the Teagasc sequencing facility.

Table 6.1| Participant characteristics

Total participants	62
Total samples	238
Age	34 (±9) years
Sex	
Male	39
Female	23
VO_2 max	46.8 (±6.6)
BMI (kg/m^2)	25.6 (±3.1)
Height (cm)	174 (±9)

Age, VO_2 max, BMI, and height are given as mean (± SD). BMI: Body Mass Index.

Table 6.2| Classification of VO_2 max by sex and age

	Age	Fair	Good	Excellent	Superior
Male	<20	38.4 - 45.1	45.2 - 50.9	51.0 - 55.9	>55.9
	20-29	36.5 - 42.4	42.5 - 46.4	46.5 - 52.4	>52.4
	30-39	35.5 - 40.9	41.0 - 44.9	45.0 - 49.4	>49.4
	40-49	33.6 - 38.9	39.0 - 43.7	43.8 - 48.0	>48.0
	50-59	31.0 - 35.7	35.8 - 40.9	41.0 - 45.3	>45.3
Female	<20	31.0 - 34.9	35.0 - 38.9	39.0 - 41.9	>41.9
	20-29	29.0 - 32.9	33.0 - 36.9	37.0 - 41.0	>41.0
	30-39	27.0 - 31.4	31.5 - 35.6	35.7 - 40.0	>40.0
	40-49	24.5 - 28.9	29.0 - 32.8	32.9 - 36.9	>36.9
	50-59	22.8 - 26.9	27.0 - 31.4	31.5 - 35.7	>35.7

Classification of VO_2 max based on sex and age. Adapted from Heyward and The Cooper Institute for Aerobics Research [17, 18]. Only those ages and classifications which were used in this study are shown.

6.2.7 Bioinformatic analysis

Resulting FastQ reads were quality checked by first removing contaminating human derived reads using NCBI Best Match Tagger (BMTagger). Resulting reads were trimmed and poor quality and duplicate reads removed using KneadData [21]. Taxonomic classifications of trimmed reads were determined using Kraken 2 and Bracken [22, 23]. Functional profiling was completed using the Human Microbiome Project Unified Metabolic Analysis Network 2 (HUMAnN2; http://huttenhower.sph.harvard.edu/humann2) [24, 25].

Analysis of resulting taxonomic and functional tables was completed in R (R Foundation for Statistical Computing, Vienna, 2012) (version 3.5.1). Alpha and beta diversities were computed in R using vegan (version 2.5-2) [26]. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Data visualisations were completed with ggplot2 (version 2.2.1) in R. A heatmap was created to visualise functional pathways from HUMAnN2 using the pheatmap package in R.

6.3 Results

6.3.1 The gut microbiome of runners clusters by maximal oxygen consumption (VO_2 max) and training frequency

Exercise induced gut permeability, as determined with the lactulose/rhamnose (L/R) permeability test, was found to be significantly greater in runners from the fair VO_2 max group relative to those in either the excellent (p=0.004) or superior (p=0.004) groups (supplementary figure 6.2). No significant difference in gut permeability was apparent across groups with different training frequencies (supplementary figure 6.2).

The gut microbiome of participants was found to remain consistent over time with no change in diversity (supplementary figure 6.3), functional potential, or taxonomic profile over the 4 time points of the study. However, gut microbiomes were found to cluster based on training frequency (Compositional: $R^2 = 0.023$, p = 0.007, Stress = 0.159; Functional: $R^2 = 0.024$, p = 0.006, Stress = 0.161) and VO_2 max (Compositional: $R^2 = 0.035$, p = 0.001, Stress = 0.159; Functional: $R^2 = 0.032$, p = 0.001, Stress = 0.161) groups (figure 6.1).

6.3.2 While a core microbiome exists, a number of different species differ significantly in relative abundance in line with VO_2 max and training frequency

A core microbiome was established across participants by identifying the top 35 most abundant species present in greater than 95% of all samples (figure 6.2). *Eubacterium rectale*, *Bacteroides dorei*, *Faecalibacterium prausnitzii*, *Bacteroides vulgatus,* and *Bifidobacterium adolescentis* were the dominant species identified, with *F. prausnitzii* identified in all samples across all time points.

Additionally, species were identified that were significantly different across athletes from different VO_2 max (supplementary table 6.4) and training frequency (supplementary table

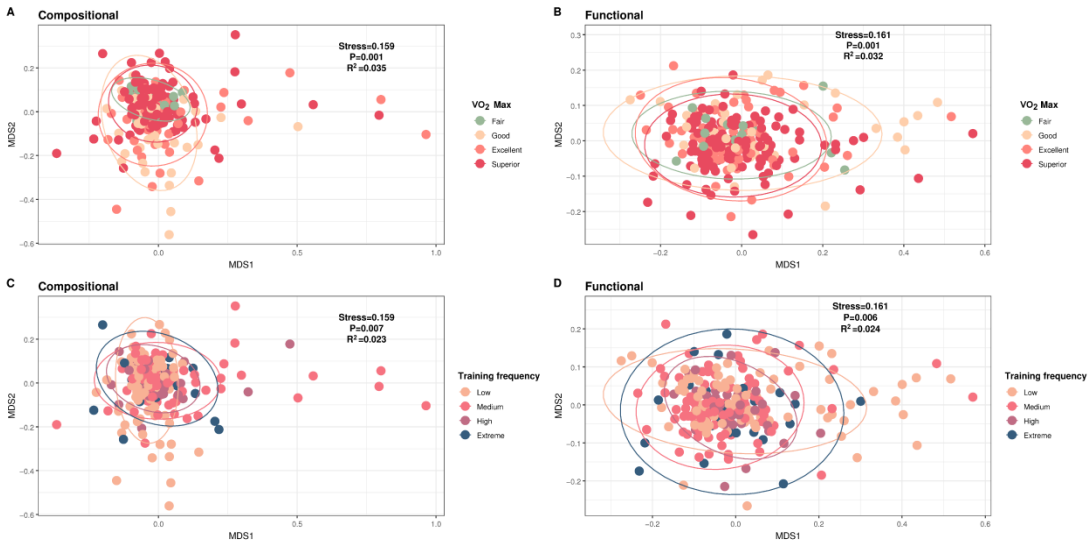


Figure 6.1| Compositional and functional potential clusters based on VO_2 max and training frequency

Multidimensional scaling (MDS) result from (A + C) species level compositional profiles from Bracken following taxonomic classification with Kraken 2 and (B + D) HUMAnN2 profiles of functional potential, as determined using vegan in R, show the differences between faecal samples based on (A + B) VO_2 max classification and (C + D) training frequency. Clustering of samples was observed based on VO_2 max classification group and training frequency for both compositional and functional data. VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Samples were classified into four groupings: fair (N=15), good (N=44), excellent (N=60), and superior (N=118). Samples were classified by training frequency into four groupings: low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23) based on self-reported, average hours of different activities per week.

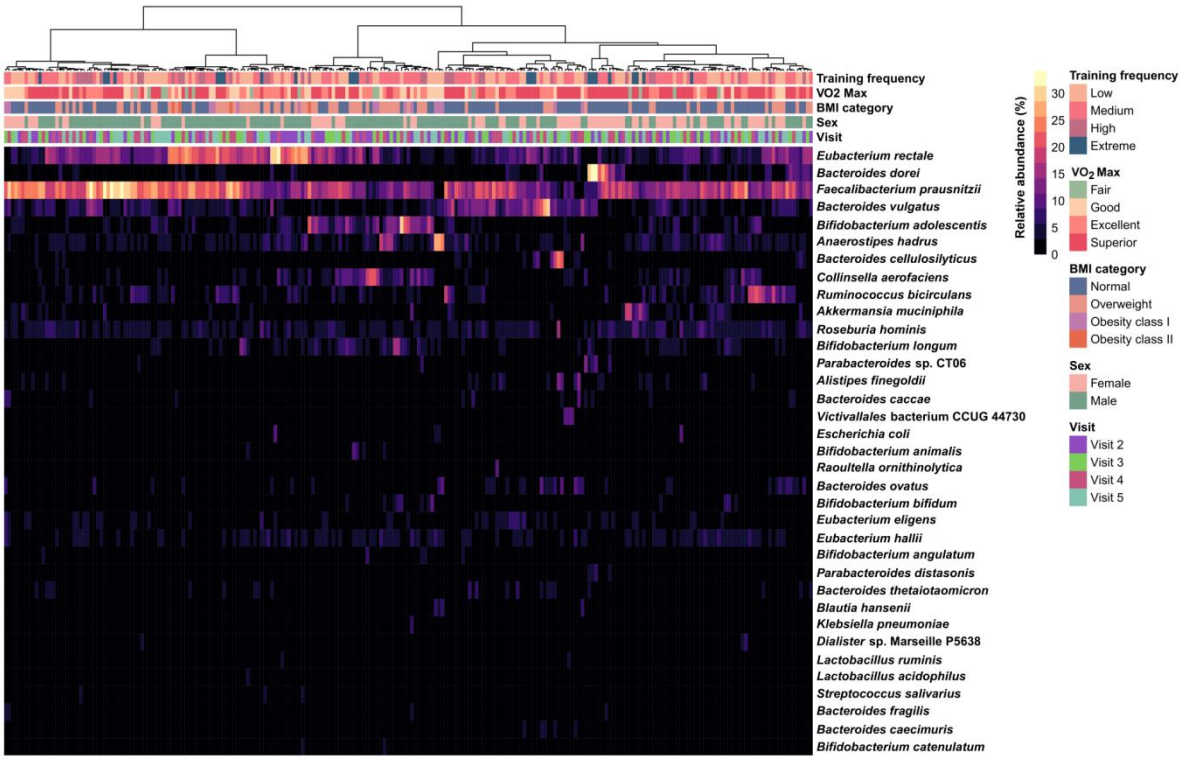


Figure 6.2| Compositional profile of samples

Heatmap reporting the top 35 most abundant species, as established from Bracken following taxonomic classification with Kraken 2, which were present in over 95% of all samples (N=240). Clustering completed using Wards method. Annotations below tree represent training frequency, VO_2 max classification, BMI category (supplementary table 6.2), sex and visit at which the sample was collected. Samples were classified by training frequency into four groupings: low (4-6 hours/week), medium (7-9 hours/week), high (10-12 hours/week), and extreme (13+ hours/week) based on self-reported, average hours of different activities per week.

6.5) groups. 1499 species were identified as significantly differing in relative abundance between at least two VO_2 max classification groups, while 495 species differed significantly between at least two training frequency groups. The number of significantly different species was greatest when the fair and good VO_2 max groups, and the low and medium training frequency groups, were compared (supplementary figure 6.5 A and C).

A particular emphasis was placed on assessing differences among species previously identified as being present in greater abundance in athletes or in response to exercise (supplementary table 6.3), as well as those that have been previously identified as producers of short chain fatty acids (SCFAs), as concentrations of SCFAs have been found to be relatively greater in athletes relative to controls [4-9, 27-32]. Several species that fell into this category (previously identified as increased in athletes or as a result of exercise, or involved with SCFA production) were found to differ significantly in abundance between groups from different VO_2 max classification (figure 6.3) and training frequency (figure 6.4) groups. More specifically, *Akkermansia muciniphila* was greater in those samples with a superior, relative to excellent, VO_2 max. *Vellonella rodentium* was more abundant among individuals with a good VO_2 max relative to those from the other three VO_2 max groups. *Ruminococcus albus* and *Prevotella* species both increased in relative abundance with increasing VO_2 max, with proportions of *Prevotella* also reflecting increased training frequency. *Bifidobacterium animalis* and *Eubacterium eligans* were also found to be increased with an increasing VO_2 max. A pattern was identified across several species whereby a decreased relative abundance was apparent in the good VO_2 max group relative to all others. This pattern was observed in the species *Verrucomicrobia* bacterium IMCC25134, *Coriobacteriaceae* bacterium 68-1-3, *Bifidobacterium thermophilum*, several *Desulfovibrio* species (*D. africanus*, *D. desulfuricans*, *D. fairfieldensis*, *D. magneticus*, *D. piger*, *D.* sp. G11, *D. vulgaris*, and *D. alaskensis*), *Clostridium cellulosi,* and *Christensenella massiliensis*.

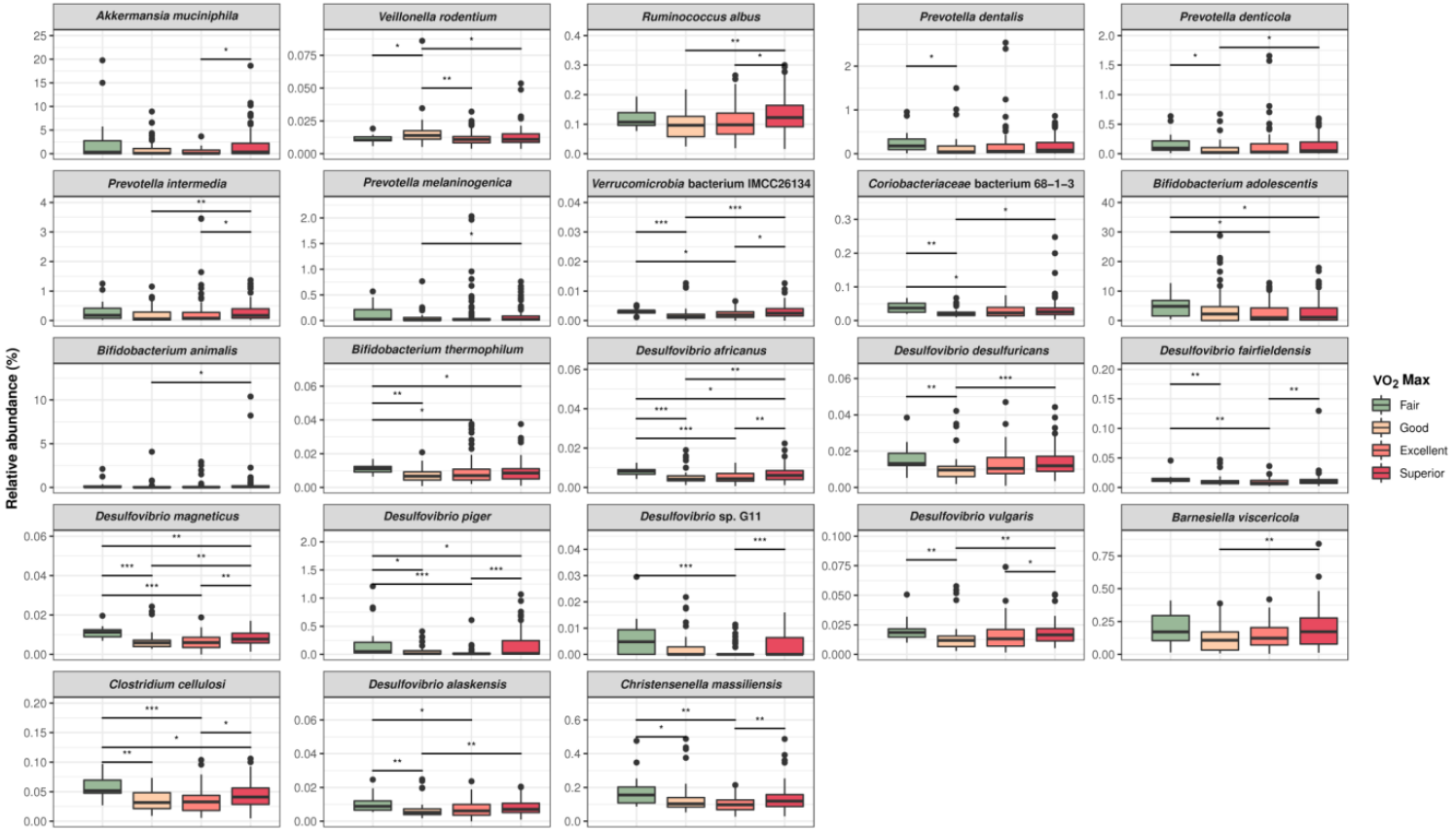


Figure 6.3| Significantly different species based on VO_2 max

Figure 6.3 continued

Significantly different species, as established from Bracken following taxonomic classification with Kraken 2, between VO_2 max classification groupings (fair (N=15), good (N=44), excellent (N=60), and superior (N=118)). VO_2 max classification groupings were classified based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Only species of interest, which had previously been identified as being increased in athlete cohorts or as a consequence of exercise, as well as those species which have been associated with short chain fatty acid (SCFA) production, as SCFA production has been found to be increased in athletes (supplementary table 6.3), are plotted with all other significant differences between groups given as supplementary table 6.4. Additionally, *Christensenella massiliensis* was plotted as it is closely related to *C. minuta*, a SCFA producer. * $P<0.05$, ** $P\leq0.01$, *** $P\leq0.001$.

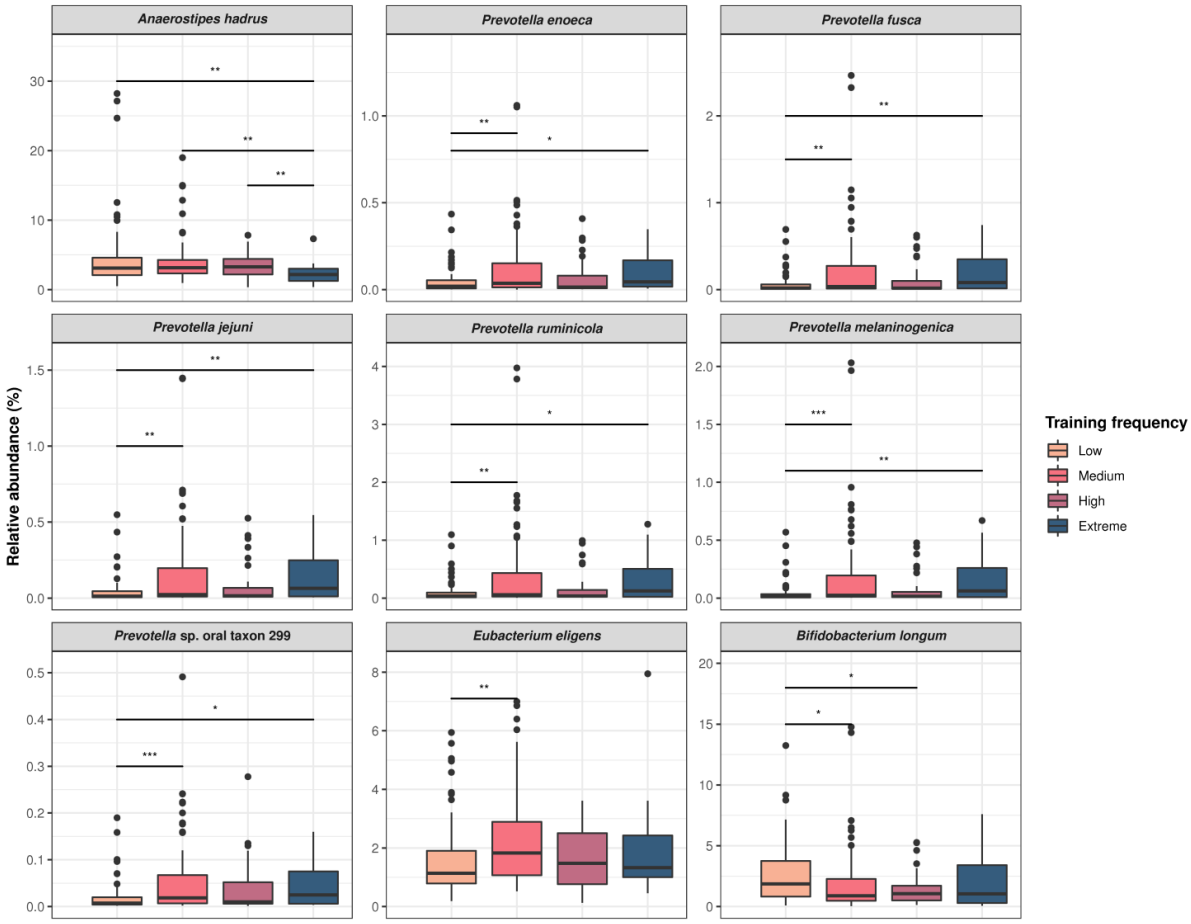


Figure 6.4| Significantly different species based on training frequency

Significantly different species, as established from Bracken following taxonomic classification with Kraken 2, between training frequency groupings (low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23)). Training groups were based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Only species of interest, which had previously been identified as being increased in athlete cohorts or as a

Figure 6.4 continued

consequence of exercise, as well as those species which have been associated with short chain fatty acid (SCFA) production, as SCFA production has been found to be increased in athletes (supplementary table 6.3), are plotted with all other significant differences between groups given as supplementary table 6.5. * $P<0.05$, ** $P\leq0.01$, *** $P\leq0.001$.

In addition to significant differences in microbial composition between groups, a trend was also observed between increases or decreases in particular species and VO_2 max or training frequency. In this regard, 10 species were increased and 1 species was decreased with increased VO_2 max grouping (supplementary table 6.6), while 12 species were increased and 16 species were decreased with an increased training frequency grouping (supplementary table 6.7).

6.3.3 Functional potential of the gut microbiome differs significantly between individuals from different VO_2 max and training frequency groups

94 pathways were identified that were significantly different between at least two VO_2 max classification groups (supplementary table 6.8), while 496 pathways were identified as significantly different between at least two training frequency groups (supplementary table 6.9). The greatest number of significantly different pathways was found between the VO_2 max groups good and superior (54 pathways), and those in the low and medium training frequency groups (36 pathways) (supplementary figure 6.5 B and D).

Those pathways which were found to be significantly different between VO_2 max groups or training groups with p values ≤0.001 are depicted in figures 6.5 and 6.6, respectively. An interesting pattern was observed for VO_2 max groupings in that, for several pathways, an increase in functional pathways was observed from the fair to the good groups, but the number of pathways decreased again in individuals with a higher VO_2 max. This pattern was observed for the pathways ARGSYN-PWY, ARGSYNBSUB-PWY, GLUTORN-PWY, GOLPDLCAT-PWY, NONOXIPENT-PWY, P23-PWY, PWY-5384, PWY-7400, and PWY0-1297, involved with amino acid synthesis. The opposite pattern, i.e. a decrease from the fair to good group with a subsequent increase in the excellent and superior groups, was observed across a few pathways. These pathways were mainly involved with general cell metabolism, but also

included a long chain fatty acid production pathway and factor 420 biosynthesis (PWY-5198). Factor 420 biosynthesis represents an essential component of methanogenesis and a significant decrease was found between the good and superior groups.

From a training frequency perspective, an increased training frequency corresponded with a decrease in several pathways, namely BRANCHED CHAIN AA SYN PWY, PWY-5103, PWY-5189, PWY-6630, PWY-7392, and PWY0-1479. These pathways are generally involved in amino acid biosynthesis, including those amino acids involved in the biosynthesis of tetrapyrrole compounds (which include heme, vitamin B_{12}, and the methanogenic coenzyme F_{430}) and branched chain amino acids, specifically isoleucine. Also several pathways involved with SCFA production were found to be increased with increased exercise frequency (supplementary table 6.9).

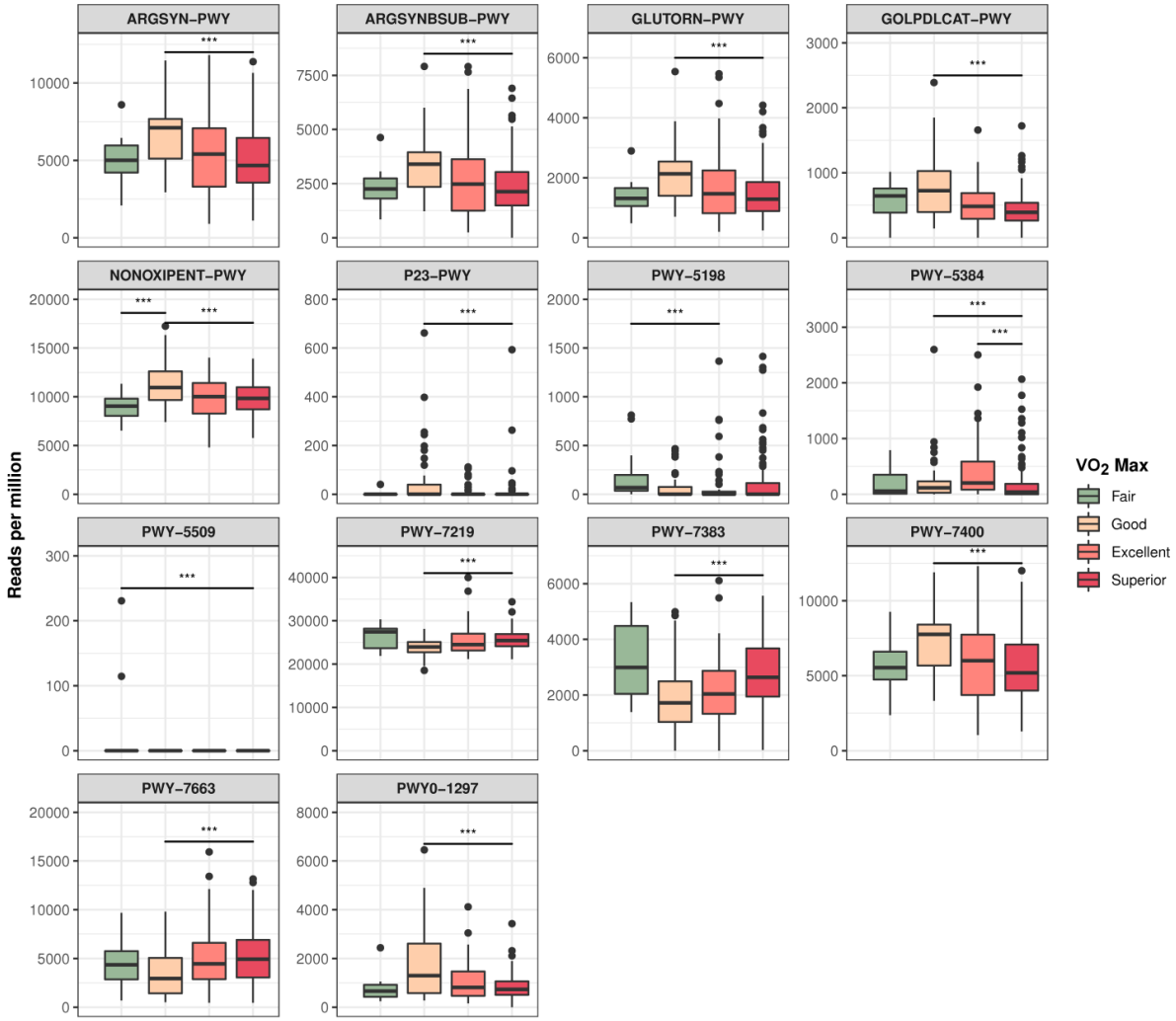


Figure 6.5| Significantly different pathways based on VO_2 max

Significantly different pathways, as established from HUMAnN2, between VO_2 max classification groupings (fair (N=15), good (N=44), excellent (N=60), and superior (N=118)). VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. Only those significant differences with P≤0.001 are plotted. All significant differences between groups are given as supplementary table 6.8. *** P≤0.001.

Figure 6.5 continued

ARGSYN-PWY: L-arginine biosynthesis I (via L-ornithine); ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle); GLUTORN-PWY: L-ornithine biosynthesis; GOLPDLCAT-PWY: superpathway of glycerol degradation to 1,3-propanediol; NONOXIPENT-PWY: pentose phosphate pathway (non-oxidative branch); P23-PWY: reductive TCA cycle I; PWY-5198: factor 420 biosynthesis; PWY-5384: sucrose degradation IV (sucrose phosphorylase); PWY-7219: adenosine ribonucleotides de novo biosynthesis; PWY-7383: anaerobic energy metabolism (invertebrates, cytosol); PWY-7400: L-arginine biosynthesis IV (archaebacteria); PWY-7663: gondoate biosynthesis (anaerobic); PWY0-1297: superpathway of purine deoxyribonucleosides degradation.

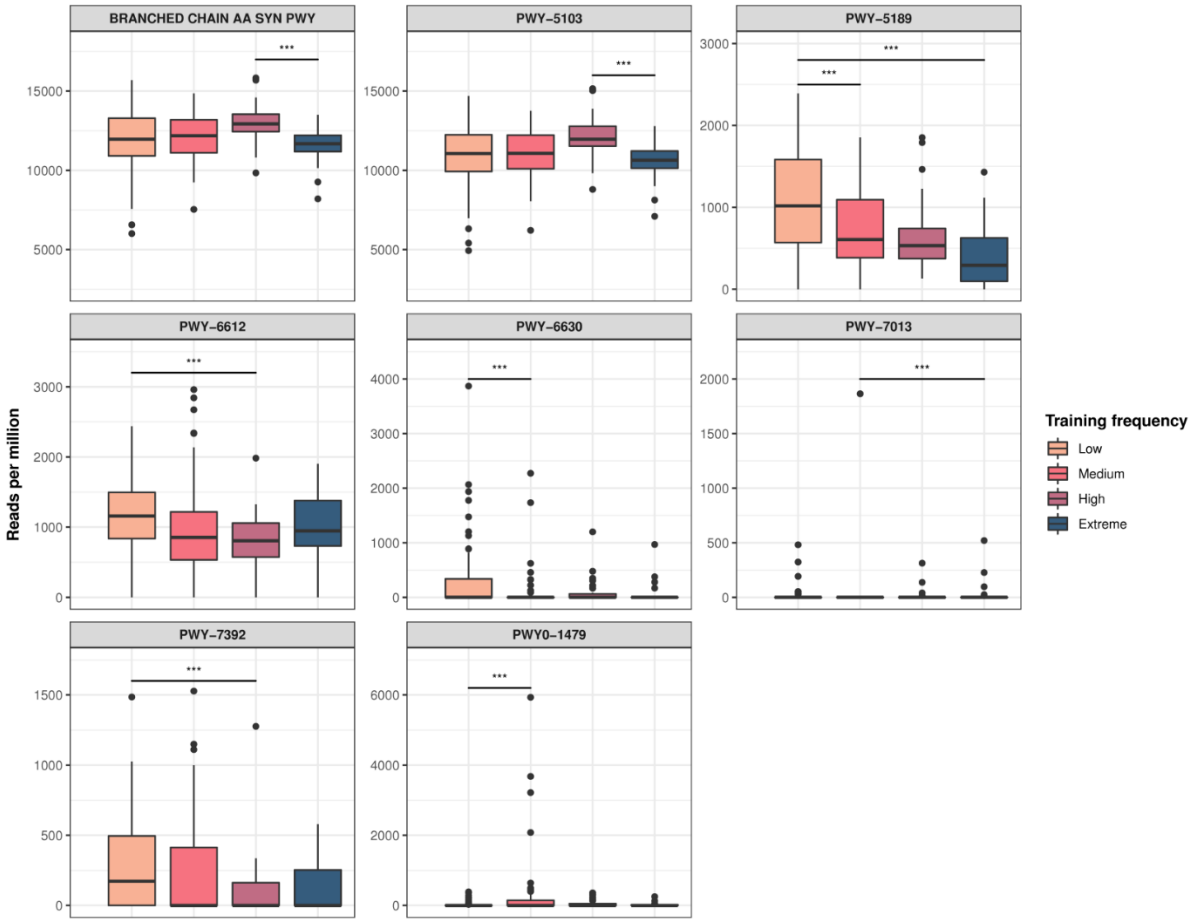


Figure 6.6| Significantly different pathways based on training frequency

Significantly different pathways, as established from HUMAnN2, between training frequency groups (low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23)). Training groups were based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. Only those significant differences with P≤0.001 are plotted. All significant differences between groups are given as supplementary table 6.9. *** P≤0.001.

Figure 6.6 continued

BRANCHED-CHAIN-AA-SYN-PWY: superpathway of branched amino acid biosynthesis; PWY-5103: L-isoleucine biosynthesis III; PWY-5189: tetrapyrrole biosynthesis II (from glycine); PWY-6612: superpathway of tetrahydrofolate biosynthesis; PWY-6630: superpathway of L-tyrosine biosynthesis; PWY-7013: L-1,2-propanediol degradation; PWY-7392: taxadiene biosynthesis; PWY0-1479: tRNA processing.

6.4 Discussion

This study aimed to study the gut microbiome composition and its functional potential in an active cohort of runners. Microbial composition and functional potential were found to remain consistent over the 7 week period of the study but differed depending on the VO_2 max or training frequency group to which the athletes belonged. The stability of the gut microbiome in healthy adults has been previously reported [2, 3] and, while cardiorespiratory fitness, as established with VO_2 max, has previously been found to be associated with specific gut microbiome functions [10], a relationship with microbial composition was not noted in that instance. The results of our study are also consistent with a previous observation that exercise frequency was associated with composition and functionality in a cyclist cohort [9].

To date, several studies have investigated the influence of exercise on the gut microbiome of humans, either through the implementation of an exercise-based intervention or the study of athlete cohorts. These studies have identified differences in the microbiomes of athlete groups relative to controls, although the precise nature of these changes have differed, potentially due to variations in the duration, types, or intensity of the exercise/training involved [4-9, 27, 30-32]. Exercise, in particular prolonged exercise and endurance exercise, has been documented to decrease blood flow to the visceral organs, including the intestines, resulting in an increased incidence of gut permeability [33, 34]. However, these impacts are predominantly acute, occurring during and immediately post exercise [34], as is observed here with a reduction in gut permeability with increased fitness.

Given that many taxa differ across VO_2 max or training frequency groups in this study, we focused specifically on the patterns relating to those, including species involved in SCFA production [4], which have been previously identified as being significantly altered in

athlete groups or as a consequence of exercise. Among these, *Akkermansia* has previously been identified as increased in athletes with respect to controls and increased as a result of the introduction of an exercise regimen [4-6, 9, 27] and, in this study, *A. muciniphila* was found to be enriched in individuals with a superior VO_2 max relative to those with an excellent VO_2 max. We also found *Veillonella rodentium* to be present at significantly higher proportions in the good VO_2 max group relative to all other groupings. This is of note as Veillonellaceae, and more specifically *Veillonella*, have been found to increasefollowing marathon running, with a species from the same serogroup as *V. rodentium*, *V. atypica* found to improve treadmill runtime in mice, reportedly as a consequence of the consumption of lactate [8].

Evidence exists that Prevotellaceae are present in greater proportions among athletes and active individuals, while abundance of *Prevotella* has also been correlated with the amount (time) of reported exercising [5, 6, 9]. A similar pattern was observed here with an increasing abundance of *Prevotella* species observed in line with increased training frequency and VO_2 max. Similarly, here *Ruminococcus albus* was found to be increased with increased VO_2 max, with Ruminococcaceae previously identified in higher abundance in athletes compared with sedentary controls [5]. Ruminococcaceae have previously been identified as increased with reduced gut transit times [35], which is reduced with cardiorespiratory fitness [36]. Several species were found to follow a trend of increasing or decreasing as a consequence of increased VO_2 max or training frequency groupings and, although these trends were not found to be statistically significant when correlations were completed, they might warrant further exploration. The relationship observed here between several species and VO_2 max measures may be contributed to by the variance in gut permeability between these groups.

Notably, another taxon, *Faecalibacterium prausnitzii*, which has been identified as present in an increased abundance with exercise [6, 30, 31], and found in all samples from our study, did not vary with VO_2 max or training frequency, potentially indicating its relationship with active over sedentary individuals was not dictated by fitness (VO_2 max) or training frequency, or that the participants in this study all undertook a level of exercise that ensured its presence, beyond which further changes in abundance did not occur. *Eubacterium eligans*, a SCFA producer [28], was found to increase significantly between a low and medium training frequency, with the relative abundance remaining similar as training frequency further increased potentially indicating a plateau effect of exercise on this species. Overall, SCFA production pathways increased with increasing training frequency, potentially benefitting gut homeostasis in these individuals, with SCFA production previously identified as increased with exercise and in athlete cohorts [4, 8, 31].

Variation in the functional capacity of the microbiome was also established in line with VO_2 max and training frequency. Amino acid synthesis was observed to increase between the fair and good VO_2 max groups, with a subsequent decrease towards a superior VO_2 max. Although amino acid synthesis has been previously found to be increased in athletes relative to controls, decreases were observed with the implementation of an endurance exercise regimen [4, 27]. An important methanogenesis pathway followed the opposite pattern (i.e. a decrease between the fair and good VO_2 max groups, with a subsequent increase towards a superior VO_2 max); although no specific methanogenic taxa were identified as following the same profile. An increased capacity for respiration with increasing VO_2 max may result in an increase in substrates for methanogens and explain this finding. Long chain fatty acid (LCFA) production was found to significantly increase with an increased VO_2 max, in line with a previous report [10]. LCFAs can increase gut motility and also act as an important source of energy in skeletal muscle during exercise [37, 38]. With an increased training frequency, a decrease in branched chain amino acid synthesis

was observed, an observation that was opposite to that previously noted in cyclists [9], potentially as a consequence of a variation in exercise modes between these studies.

Overall, those differences observed in the functional capacity of the microbiome highlight important interactions between microbial populations, host, and energy metabolism. In response to host stimulation, with an increased VO_2 max or training frequency, modified microbial functionality may benefit the athlete host. The increased abundance of specific pathways may be exploited by the athlete for the production of amino acids, LCFA's, and branched chain amino acids, which may benefit the specific energy states of the individual.

Taken together, this study provides evidence that gut microbiome composition and functional potential is influenced by the VO_2 max and training frequency of an individual. Although some of these differences had previously been noted in other studies, others were new to this study of active individuals. This study has limitations, including the absence of longitudinal dietary and intestinal transit time data, which may contribute to these differences. Future studies should focus on the investigation of these factors along with differences in the host and microbial metabolites produced by these individuals.

6.5 Acknowledgements

The authors thank Fiona Crispie and Laura Finnegan for their contributions to DNA sequencing. The authors are grateful to all those who donated their time, effort, and samples to this study. This research was conducted with the financial support of Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273, 13/SIRG/2160, and 11/PI/1137.

6.6 References

1. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-230.
2. Conlon, M.A. and Bird, A.R., *The impact of diet and lifestyle on gut microbiota and human health.* Nutrients, 2015. **7**: p. 17-44.
3. David, L.A., Materna, A.C., Friedman, J., Campos-Baptista, M.I., Blackburn, M.C., Perrotta, A., Erdman, S.E., and Alm, E.J., *Host lifestyle affects human microbiota on daily timescales.* Genome Biology and Evolution., 2014. **15**(7).
4. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.
5. 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., and Cotter, P.D., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.
6. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).
7. Zhao, X., Zhang, Z., Hu, B., Huang, W., Yuan, C., and Zou, L., *Response of gut microbiota to metabolite changes induced by endurance exercise.* Frontiers in Microbiology, 2018. **9**: p. 765.
8. Scheiman, J., Luber, J.M., Chavkin, T.A., MacDonald, T., Tung, A., Pham, L.D., Wibowo, M.C., Wurth, R.C., Punthambaker, S., Tierney, B.T., Yang, Z., Hattab, M.W., Avila-Pacheco, J., Clish, C.B., Lessard, S., Church, G.M., and Kostic, A.D., *Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism.* Nature Medicine, 2019. **25**(7): p. 1104-1109.
9. Petersen, L.M., Bautista, E.J., Nguyen, H., Hanson, B.M., Chen, L., Lek, S.H., Sodergren, E., and Weinstock, G.M., *Community characteristics of the gut microbiomes of competitive cyclists.* Microbiome, 2017. **5**(1): p. 98.

10. Estaki, M., Pither, J., Baumeister, P., Little, J.P., Gill, S.K., Ghosh, S., Ahmadi-Vand, Z., Marsden, K.R., and Gibson, D.L., *Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions.* Microbiome, 2016. **4**(1): p. 42.
11. Bacon, A.P., Carter, R.E., Ogle, E.A., and Joyner, M.J., *VO2max trainability and high intensity interval training in humans: A meta-analysis.* PLoS One, 2013. **8**(9): p. e73182.
12. Gim, M.N. and Choi, J.H., *The effects of weekly exercise time on VO2max and resting metabolic rate in normal adults.* The Journal of Physical Therapy Science, 2016. **28**(4): p. 1359-1363.
13. Pals, K.L., Chang, R.T., Ryan, A.J., and Gisolfi, C.V., *Effect of running intensity on intestinal permeability.* Journal of Applied Physiology, 1997. **82**(2): p. 571-576.
14. Karl, J.P., Margolis, L.M., Madslien, E.H., Murphy, N.E., Castellani, J.W., Gundersen, Y., Hoke, A.V., Levangie, M.W., Kumar, R., Chakraborty, N., Gautam, A., Hammamieh, R., Martini, S., Montain, S.J., and Pasiakos, S.M., *Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2017. **312**(6): p. G559-G571.
15. Billat, L.V. and Koralsztein, J.P., *Significance of the velocity at VO2max and time to exhaustion at this velocity.* Sports Medicine, 1996. **22**(2): p. 90-108.
16. Jones, A.M. and Doust, J.H., *A 1% treadmill grade most accurately reflects the energetic cost of outdoor running.* Journal of Sports Sciences, 1996. **14**(4): p. 321-327.
17. The Cooper Institute, *The physical fitness specialist manual*. 1997. The Cooper Institute for Aerobics Research, Dalla.
18. Heyward, V.H., *Advanced fitness assessment and exercise prescription*. 3rd Edition ed. 1998. Burgess Publishing Company, Leeds.
19. Dittrich, N., de Lucas, R.D., Beneke, R., and Guglielmo, L.G., *Time to exhaustion at continuous and intermittent maximal lactate steady state during running exercise.* International Journal of Sports Physiology and Performance, 2014. **9**(5): p. 772-776.
20. Yu, Z. and Morrison, M., *Improved extraction of PCR-quality community DNA from digesta and fecal samples.* Biotechniques, 2004. **36**(5): p. 808-812.

21. Hsu, T., Joice, R., Vallarino, J., Abu-Ali, G., Hartmann, E.M., Shafquat, A., DuLong, C., Baranowski, C., Gevers, D., Green, J.L., Morgan, X.C., Spengler, J.D., and Huttenhower, C., *Urban transit system microbial communities differ by surface type and interaction with humans and the environment.* mSystems, 2016. **1**(3).
22. Wood, D.E. and Salzberg, S.L., *Kraken: Ultrafast metagenomic sequence classification using exact alignments.* Genome Biology, 2014. **15**(3): p. R46.
23. Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L., *Bracken: Estimating species abundance in metagenomics data.* PeerJ Computer Science, 2017. **3**: p. e104.
24. Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methe, B., Schloss, P.D., Gevers, D., Mitreva, M., and Huttenhower, C., *Metabolic reconstruction for metagenomic data and its application to the human microbiome.* PLoS Computational Biology, 2012. **8**(6): p. e1002358.
25. Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C., *Species-level functional profiling of metagenomes and metatranscriptomes.* Nature Methods, 2018. **15**(11): p. 962-968.
26. Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O' Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoeca, E., and Wagner, H., Vegan: *Community ecology package*. R package version 2.2-1.2.1-2, 2018.
27. Munukka, E., Ahtiainen, J.P., Puigbo, P., Jalkanen, S., Pahkala, K., Keskitalo, A., Kujala, U.M., Pietila, S., Hollmen, M., Elo, L., Huovinen, P., D'Auria, G., and Pekkala, S., *Six-week endurance exercise alters gut metagenome that is not reflected in systemic metabolism in over-weight women.* Frontiers in Microbiology, 2018. **9**: p. 2323.
28. Lordan, C., Thapa, D., Ross, R.P., and Cotter, P.D., *Potential for enriching next-generation health-promoting gut bacteria through prebiotics and other dietary components.* Gut Microbes, 2019: p. 1-20.
29. Baxter, N.T., Schmidt, A.W., Venkataraman, A., Kim, K.S., Waldron, C., and Schmidt, T.M., *Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers.* MBio, 2019. **10**(1).
30. Jang, L.G., Choi, G., Kim, S.W., Kim, B.Y., Lee, S., and Park, H., *The combination of sport and sport-specific diet is associated with characteristics of gut microbiota: An*

observational study. Journal of the International Society of Sports Nutrition, 2019. **16**(1): p. 21.

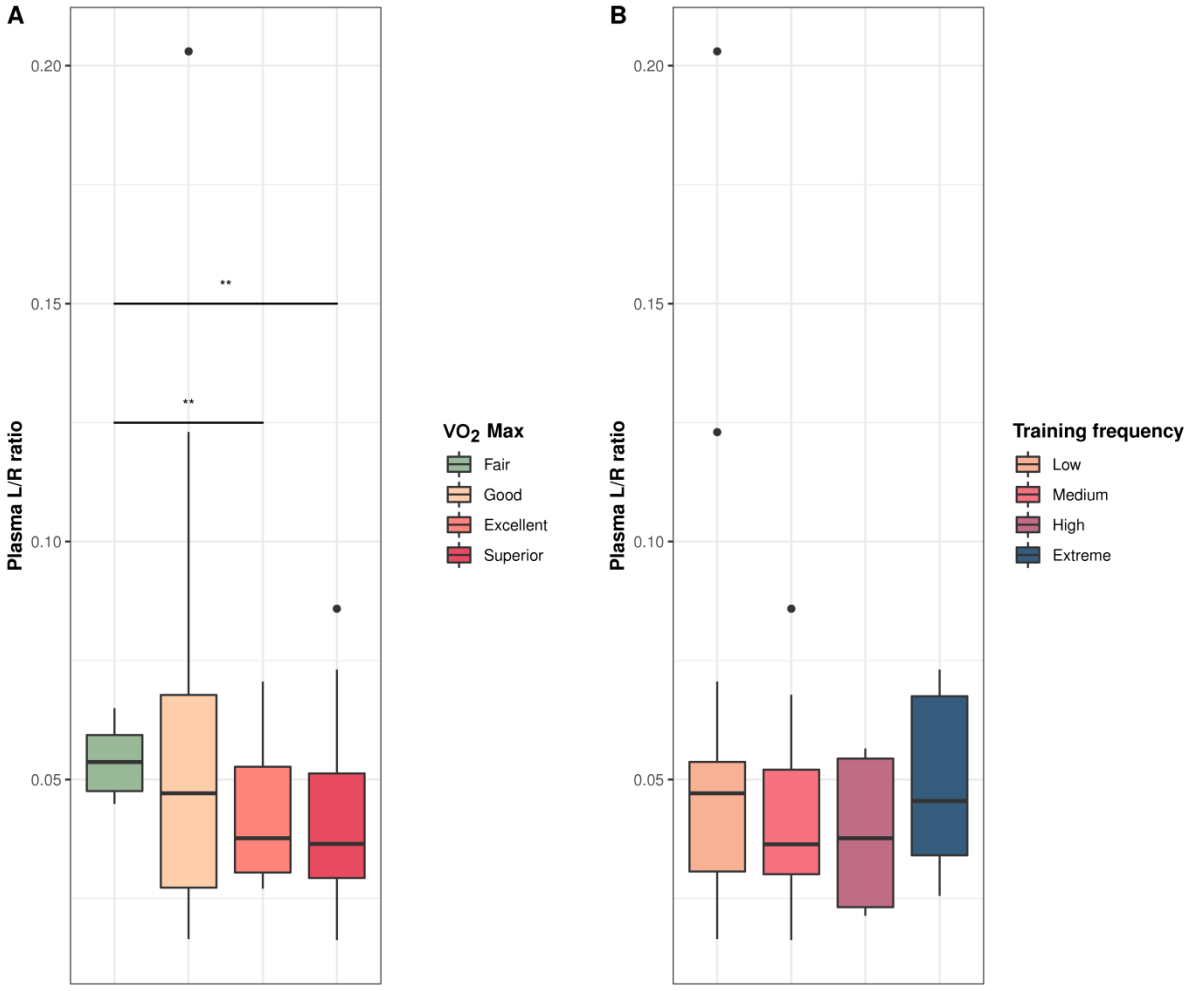
31. Allen, J.M., Mailing, L.J., Niemiro, G.M., Moore, R., Cook, M.D., White, B.A., Holscher, H.D., and Woods, J.A., *Exercise alters gut microbiota composition and function in lean and obese humans.* Medicine & Science in Sports & Exercise, 2018. **50**(4): p. 747-757.
32. Taniguchi, H., Tanisawa, K., Sun, X., Kubo, T., Hoshino, Y., Hosokawa, M., Takeyama, H., and Higuchi, M., *Effects of short-term endurance exercise on gut microbiota in elderly men.* Physiological Reports, 2018. **6**(23): p. e13935.
33. O'Sullivan, O., Cronin, O., Clarke, S.F., Murphy, E.F., Molloy, M.G., Shanahan, F., and Cotter, P.D., *Exercise and the microbiota.* Gut Microbes, 2015. **6**(2): p. 131-136.
34. Monda, V., Villano, I., Messina, A., Valenzano, A., Esposito, T., Moscatelli, F., Viggiano, A., Cibelli, G., Chieffi, S., Monda, M., and Messina, G., *Exercise modifies the gut microbiota with positive health effects.* Oxidative Medicine and Cellular Longevity, 2017. **2017**: p. 3831972.
35. Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R.Y., Joossens, M., and Raes, J., *Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates.* Gut, 2016. **65**(1): p. 57-62.
36. Yang, Y., Shi, Y., Wiklund, P., Tan, X., Wu, N., Zhang, X., Tikkanen, O., Zhang, C., Munukka, E., and Cheng, S., *The association between cardiorespiratory fitness and gut microbiota composition in premenopausal women.* Nutrients, 2017. **9**(8).
37. Zhao, L., Huang, Y., Lu, L., Yang, W., Huang, T., Lin, Z., Lin, C., Kwan, H., Wong, H.L.X., Chen, Y., Sun, S., Xie, X., Fang, X., Yang, H., Wang, J., Zhu, L., and Bian, Z., *Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats.* Microbiome, 2018. **6**(1): p. 107.
38. Kiens, B. and Roepstorff, C., *Utilization of long-chain fatty acids in human skeletal muscle during exercise.* Acta Physiologica Scandinavica, 2003. **178**(4): p. 391-396.
39. Dwyer, J.T., Melanson, K.J., Sriprachy-anunt, U., Cross, P., and Wilson, M., *Dietary treatment of obesity*, in *Endotext*, Feingold K.R., Anawalt B., and Boyce A., Editors. 2000: South Dartmouth (MA).

6.7 Supplementary material

Week	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**	**6**
			n = 62					
Visit	**1**	**2**	**3**				**4**	**5**
	Screening **VO_2 max**	**L/R** **Faecal** **sample**	**L/R +** **exercise** **Faecal** **sample**				**L/R** **Faecal** **sample**	**L/R +** **exercise** **Faecal** **sample**

Supplementary figure 6.1| Experimental design

Timeline of intestinal permeability testing by lactulose/rhamnose (L/R) permeability test and exercise challenge consisting of treadmill running at a speed corresponding to 80% of vVO_2 max for 60 min (1% inclination). Faecal samples analysed in this study were collected at visit 2, 3, 4, and 5. VO_2 max used for VO_2 max groupings in this study were established in week 1 on a motorised treadmill with an initial speed of 10 km h^{-1} (1% incline), with an increase of 1.0 km h^{-1} every 2 min until subjects achieved volitional exhaustion [15, 16]. Respiratory gases were measured and VO_2 max was determined as the highest oxygen uptake per minute.

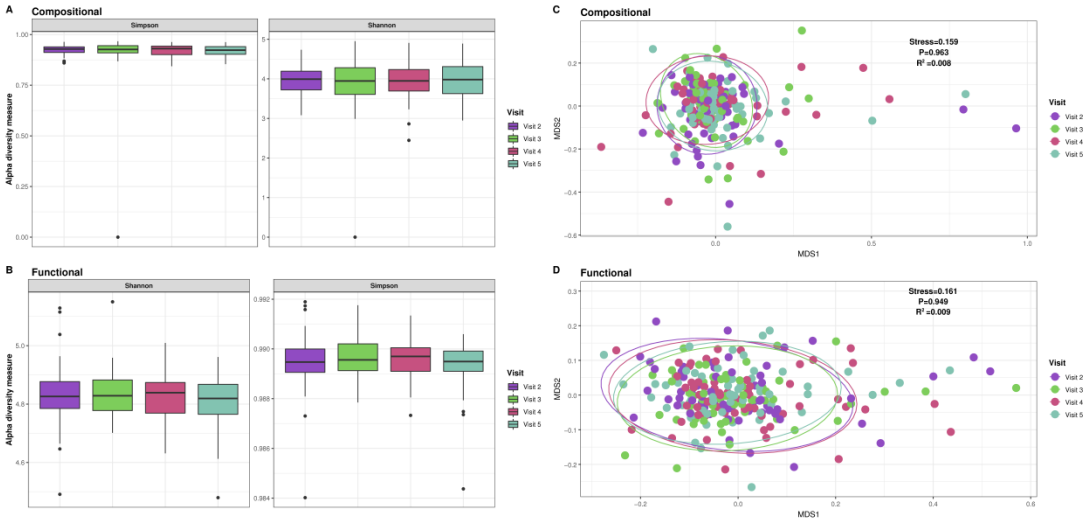


Supplementary figure 6.2| Gut permeability differs based on VO_2 max, but not training frequency

Gut permeability of individuals separated by (A) VO_2 max classification and (B) training frequency based on the lactulose/rhamnose (L/R) permeability test completed during an exercise challenge. Samples were classified into four groupings: fair (N=15), good (N=44), excellent (N=60), and superior (N=118). Samples were classified by training frequency into four groupings: low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23) based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results

Supplementary figure 6.2 continued

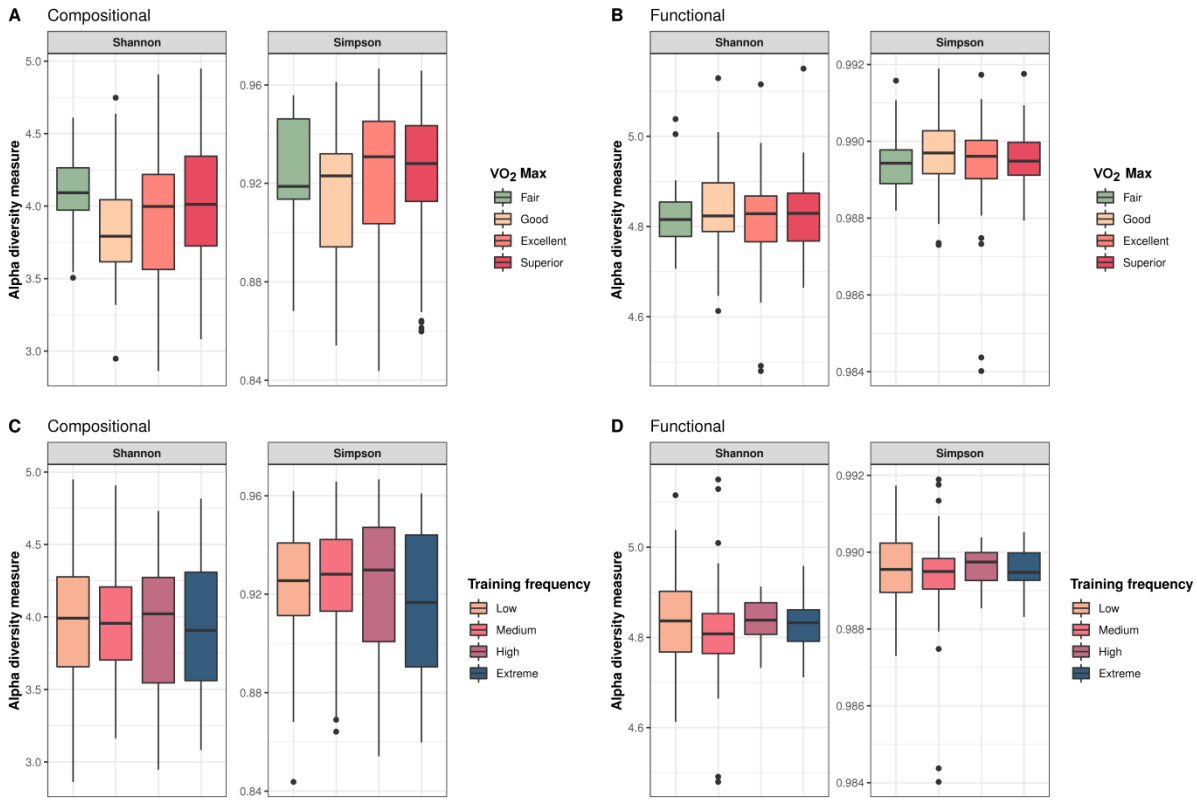
to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05.



Supplementary figure 6.3| Alpha and beta diversity over time

Diversity measures of gut microbiome (A) composition and (B) functional potential as determined using vegan in R coloured by visit (visit 2 (N=61), visit 3 (N=58), visit 4 (N=59), and visit 5 (N=59)). Species level assignments from Bracken following taxonomic classification with Kraken 2 were analysed with vegan to determine Shannon and Simpson diversities, for compositional data. Unstratified pathway abundance (reads per million) assignments as determined using HUMAnN2 were analysed with vegan to determine Shannon and Simpson diversities, for functional data. No significant differences were established in alpha diversity by visit.

Multidimensional scaling (MDS) result from (C) species level compositional profiles from Bracken following taxonomic classification with Kraken 2 and (D) HUMAnN2 profiles of functional potential, as determined using vegan in R, show the differences between faecal samples based on visit. Clustering of samples was not observed based on visit.

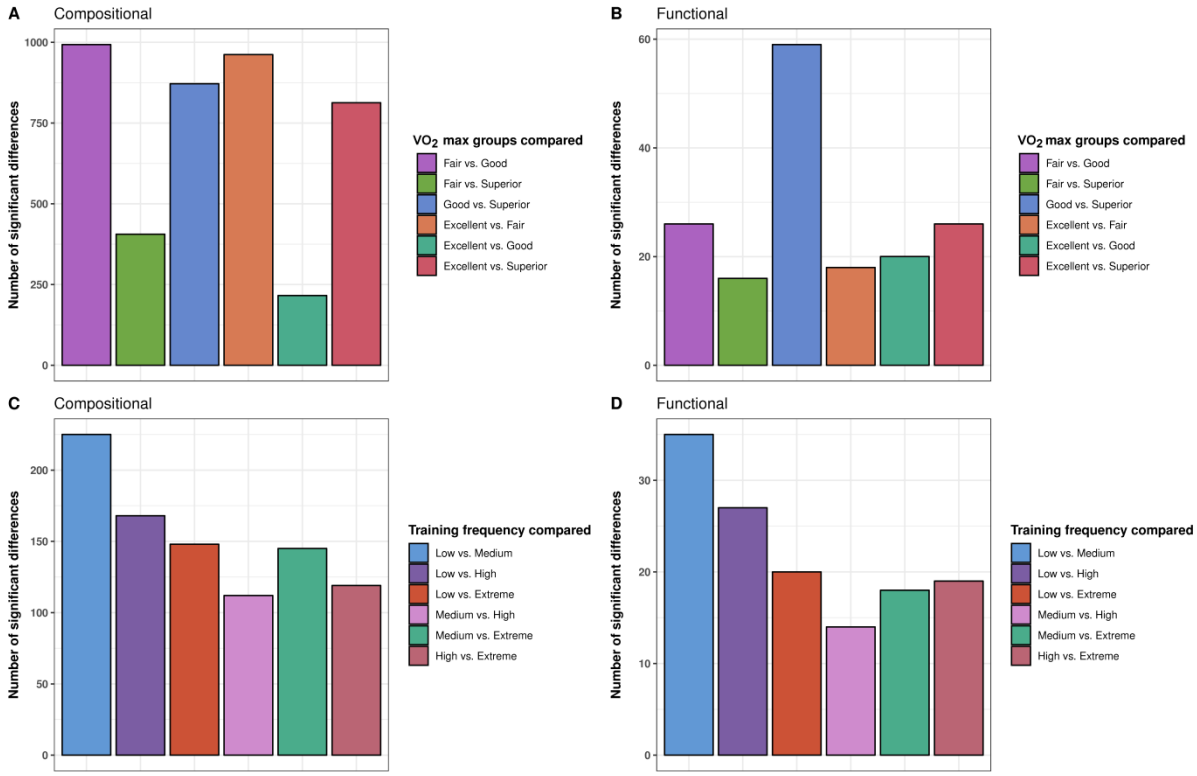


Supplementary figure 6.4| Alpha diversity measures by VO_2 max classification and training frequency

Diversity measures of gut microbiome (A + C) composition and (B + D) functional potential as determined using vegan in R coloured by (A + B) VO_2 max classification and (C + D) training frequency. Species level assignments from Bracken following taxonomic classification with Kraken 2 were analysed with vegan to determine Shannon and Simpson diversities, for compositional data. Unstratified pathway abundance (reads per million) assignments as determined using HUMAnN2 were analysed with vegan to determine Shannon and Simpson diversities, for functional data. VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Samples were classified into four groupings: fair (N=15), good (N=44), excellent (N=60), and superior (N=118). Samples were classified by training frequency into four groupings: low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23) based on self

Supplementary figure 6.4 continued

reported, average hours of different activities per week. No significant differences were established in alpha diversity by VO_2 max classification grouping or training frequency.



Supplementary figure 6.5| Number of significant differences overall

(A + C) Number of significant differences between groups in species, from Bracken following taxonomic classification with Kraken 2, and (B + D) pathways, from HUMAnN2, based on (A + B) VO_2 max classification and (C + D) training frequency (hours/week). VO_2 max classification groupings were established based on those determined by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Samples were classified into four groupings: fair (N=15), good (N=44), excellent (N=60), and superior (N=118). Samples were classified by training frequency into four groupings: low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+; N=23) based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method.

Supplementary table 6.1| Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
• Written informed consent • Healthy (gastrointestinal symptoms are allowed, but not needed) • Age ≥ 18 and ≤ 50 years at the day of screening • Weekly training load ≥ 4 hours within endurance sports as judged by the subject (minimum 1.5 hours of the training must be running) • Able to complete a 10 km run on a treadmill within 60 min, as judged by the subject • Willing to abstain from any probiotics products or medication known to alter gastrointestinal function throughout the participation of the trial	• Abdominal surgery which, as judged by the investigator, might affect the gastrointestinal function (except appendectomy and cholecystectomy) • Resting diastolic blood pressure ≥ 90 mmHg • Resting systolic blood pressure ≥ 140 mmHg • A current diagnosis of psychiatric disease • Systemic use of antibiotics, steroids (except contraceptives), or antimicrobial medication in the last 2 months • Daily usage of NSAIDs in the last 2 months or incidental use in the last 2 weeks prior to screening • Usage of medications, except contraceptives, in the last 2 weeks prior to screening • Diagnosed inflammatory gastrointestinal disease • Lactose intolerance • Any other disease that, by the Investigators discretion, could interfere with the intestinal barrier function of the subject • Participation in other clinical trials in the past 2 months prior to screening • Regular use of probiotics in the last 2 months • Smoking and/or frequent use of other nicotine products • Desire and/or plans on changing current diet and/or exercise regime during the participation of this trial • Use of laxatives, anti-diarrheals or anti-cholinergics within last 2 months prior to

	screening • Use of immunosuppressant drugs within last 4 weeks prior to screening • Women: Pregnancy or lactation

NSAIDs: non-steroidal anti-inflammatory drugs.

Supplementary table 6.2| BMI classifications

Classification	BMI (kg/m^2)	Number of individuals in classification
Normal weight	18.5 - 24.9	37 participants
Overweight	25 - 29.9	20 participants
Obese	>30	5 participants

BMI classifications adapted from Dwyer et al. [39] to show only those classifications which the participants of this study fell into. BMI: Body Mass Index.

Supplementary table 6.3| Taxa of interest

Taxa	Reason for interest	Reference(s)
Acholeplasma	Increased after marathon	Zhao et al., 2018
Akkermansia	Increased abundance with exercise and in athletes	Munukka et al., 2018; Clarke et al., 2014; Barton et al., 2017; Petersen et al., 2017; Bressa at al., 2017
Anaerofilum	Increased with endurance exercise	Munukka et al., 2018
Anaerostipes caccae	SCFA producer	Lordan et al., 2019
Anaerostipes hadrus	SCFA producer	Lordan et al., 2019; Baxter et al., 2019
Bacteroides caccae	Higher in body builders and distance runners	Jang et al., 2019
Bifidobacteriaceae	Increased with endurance exercise	Munukka et al., 2018
Bifidobacterium	Higher in active women	Bressa at al., 2017
Bulleidia	Increased after marathon	Zhao et al., 2018
Christensenella minuta	SCFA producer	Lordan et al., 2019
Clostridiales	Increased with exercise training	Allen et al., 2017
Clostridium	Higher in bodybuilders	Jang et al., 2019
Coprococcus	Increased abundance with exercise	Bressa at al., 2017; Zhao et al., 2018
Coriobacteriaceae	Increased after marathon	Zhao et al., 2018
Desulfovibrionaceae	Increased after marathon	Zhao et al., 2018
Dethiobacter	Increased after marathon	Zhao et al., 2018
Dorea	Increased with	Munukka et al., 2018

	endurance exercise	
Enterobacter cloacae	Higher in runners	Jang et al., 2019
Eubacterium eligens	SCFA producer	Lordan et al., 2019
Eubacterium rectale	SCFA producer	Lordan et al., 2019; Baxter et al., 2019
Faecalibacterium prausnitzii	Increased abundance with exercise	Bressa at al., 2017; Allen et al., 2017; Jang et al., 2019
Firmicutes	Higher within athletes	Clarke et al., 2014
Gardnerella	Increased after marathon	Zhao et al., 2018
Methanobrevibacter	Higher abundance of transcripts in professional cyclists	Petersen et al., 2017
Oscillospira	Increased with exercise	Taniguchi et al 2018
Prevotella	Correlated with time reported exercising	Petersen et al., 2017
Prevotellaceae	Higher in athletes and active women	Bressa at al., 2017; Clarke et al., 2014
RC9 gut group	Higher within athletes	Clarke et al., 2014
Roseburia hominis	Increased abundance with exercise	Bressa at al., 2017; Allen et al., 2017
Ruminococcaceae	Higher within athletes	Clarke et al., 2014
Ruminococcus	Increased abundance with exercise	Bressa at al., 2017; Zhao et al., 2018; Petersen et al., 2017; Jang et al., 2019
S24-7	Higher within athletes	Clarke et al., 2014
Succinivibrio	Increased abundance with exercise and in athletes	Zhao et al., 2018; Clarke et al., 2014
Sutterella	Higher in distance runners and body builders	Jang et al., 2019
Veillonella	Increased after marathon	Scheiman et al., 2019
Veillonella atypica	Improves treadmill	Scheiman et al., 2019

	runtime in mice	
Veillonellaceae	Increased after marathon	Zhao et al., 2018
Verrucomicrobia	Increased with endurance exercise	Munukka et al., 2018
Verrucomicrobiaceae	Increased with endurance exercise	Munukka et al., 2018
Victivallis	Increased after marathon	Zhao et al., 2018

Species of interest are those which have been previously identified as being increased in athlete cohorts or as a consequence of exercise, as well as those species which have been associated with short chain fatty acid (SCFA) production, as SCFA production has been found to be increased in athletes [4-9, 27-32].

Supplementary table 6.4| Species significantly different between VO_2 max classification groupings

Species	Excellent vs. Fair	Excellent vs. Good	Excellent vs. Superior	Fair vs. Good	Fair vs. Superior	Good vs. Superior
Acaryochloris marina	**0.01**	0.166	0.233	0.12	**0.019**	0.537
Acetobacter tropicalis	**0.006**	0.154	0.063	0.063	0.063	0.922
Acetobacterium woodii	0.531	0.123	0.235	0.123	0.909	**0.006**
Achromobacter denitrificans	**0.003**	0.737	**0.003**	**0.003**	0.096	**0.003**
Achromobacter insolitus	**0.008**	0.585	**0.002**	**0.02**	0.375	**0.006**
***Achromobacter* sp. AONIH1**	**0.02**	0.053	**0.034**	**0.003**	0.165	**0.001**
Achromobacter spanius	**0.017**	0.833	**0.011**	**0.017**	0.272	**0.011**
Achromobacter xylosoxidans	**0.041**	0.122	0.079	**0.002**	0.155	**<0.001**
Acidaminococcus fermentans	**0.001**	0.523	**0.016**	**<0.001**	**0.002**	**0.001**
Acidaminococcus intestini	0.155	0.461	0.443	**0.01**	0.091	0.091
***Acidiferrobacter* sp. SPIII_3**	**0.001**	0.396	**0.005**	**<0.001**	**0.028**	**0.001**
Acidihalobacter prosperus	**0.036**	0.646	**0.011**	**0.011**	0.506	**0.011**
Acidimicrobium ferrooxidans	**0.005**	0.227	**0.005**	0.056	0.12	0.264
Acidiphilium cryptum	**0.003**	**0.029**	0.368	0.29	**0.004**	0.054
Acidipropionibacterium acidipropionici	**0.016**	0.732	**0.021**	**0.013**	0.265	**0.016**
Acidithiobacillus ferrooxidans	**0.002**	**0.045**	**<0.001**	0.105	0.15	0.15

Acidobacterium capsulatum	**0.004**	0.454	**<0.001**	**0.003**	0.486	**0.003**
Acidothermus cellulolyticus	0.153	0.067	0.906	**0.015**	0.171	**0.047**
Acidovorax avenae	**0.003**	0.128	**0.03**	**0.001**	0.11	**0.001**
Acidovorax citrulli	0.075	0.259	0.078	**0.003**	0.351	**0.003**
***Acidovorax* sp. JS42**	**0.007**	0.517	**0.001**	**0.02**	0.419	**0.014**
***Acidovorax* sp. KKS102**	**0.001**	0.869	**0.001**	**0.002**	0.07	**0.003**
***Acidovorax* sp. RAC01**	**0.037**	0.492	0.072	0.094	0.167	0.407
***Acidovorax* sp. T1**	**0.005**	0.885	0.096	**0.019**	**0.026**	0.22
Acinetobacter haemolyticus	0.579	0.087	0.671	0.579	0.572	**0.019**
Acinetobacter johnsonii	**0.016**	**0.024**	**0.003**	0.426	0.426	0.677
Acinetobacter radioresistens	0.589	**0.038**	0.589	0.433	0.613	**0.038**
***Acinetobacter* sp. ADP1**	**0.034**	**0.034**	0.297	0.643	0.07	0.07
Actinoalloteichus hoggarensis	0.071	0.732	**0.043**	0.089	0.572	0.089
Actinoalloteichus hymeniacidonis	**0.045**	0.688	0.129	0.136	0.136	0.338
***Actinoalloteichus* sp. AHMU CJ021**	**0.001**	0.741	**0.001**	**0.001**	0.063	**0.002**
***Actinobacteria* bacterium IMCC19121**	0.415	**0.028**	0.415	0.415	0.758	0.12
***Actinobacteria* bacterium IMCC26077**	0.133	0.647	0.268	0.133	**0.012**	0.133
Actinomyces gaoshouyii	**0.006**	0.759	**0.017**	**0.001**	0.083	**0.006**
Actinomyces radicidentis	0.07	0.147	0.206	**0.005**	0.147	**0.008**
***Actinomyces* sp. Chiba101**	**0.021**	0.312	0.321	**0.006**	0.05	0.05

***Actinomyces* sp. oral taxon 897**	**0.018**	0.357	0.073	**0.007**	0.098	**0.016**
Actinoplanes friuliensis	**0.033**	0.869	**0.003**	**0.044**	0.779	**0.013**
Actinoplanes missouriensis	**0.022**	0.995	0.163	**0.035**	0.082	0.237
***Actinoplanes* sp. N902 109**	**0.006**	0.615	**0.001**	**0.006**	0.321	**0.001**
Actinopolyspora erythraea	**0.002**	0.573	**0.001**	**0.001**	0.247	**0.001**
Actinosynnema mirum	**0.037**	0.495	0.495	0.151	**0.038**	0.907
Actinosynnema pretiosum	**0.001**	0.788	**<0.001**	**0.005**	0.324	**0.005**
Advenella kashmirensis	0.079	0.72	**0.007**	0.273	0.768	0.123
Aerococcus christensenii	**0.016**	0.286	0.187	**0.006**	0.068	**0.016**
Aerococcus urinaehominis	0.32	0.137	0.689	0.689	0.137	**0.046**
Aeromonas hydrophila	**0.028**	0.366	0.071	0.087	0.169	0.38
Aeromonas salmonicida	**0.002**	0.599	0.126	**0.001**	**0.003**	**0.049**
Aeromonas schubertii	0.116	0.229	**0.01**	0.472	0.977	0.374
***Aeromonas* sp. CU5**	**<0.001**	0.438	**<0.001**	**<0.001**	0.228	**<0.001**
Aeromonas veronii	**0.005**	0.932	**0.004**	**0.006**	0.22	**0.005**
Aggregatibacter actinomycetemcomitans	0.247	0.095	**<0.001**	0.958	0.247	0.121
Aggregatibacter aphrophilus	0.236	**0.016**	0.471	0.589	0.471	0.086
Aggregatibacter segnis	0.155	**0.02**	0.28	0.726	0.448	0.155
Agrobacterium fabrum	**0.006**	0.502	**0.006**	0.073	0.417	0.122
***Agrobacterium* sp. RAC06**	**0.008**	0.65	**0.034**	**0.008**	0.056	0.082

Agrobacterium tumefaciens	**0.01**	0.792	0.311	**0.01**	**0.01**	0.326
Agromyces aureus	**0.002**	0.364	**0.043**	**0.001**	**0.01**	**0.003**
***Agromyces* sp. 30A**	**0.006**	0.36	0.179	**<0.001**	**0.01**	**0.008**
Ahniella affigens	0.06	0.366	0.099	**0.009**	0.265	**0.009**
Akkermansia muciniphila	0.337	0.879	**0.017**	0.432	0.989	0.178
Alcaligenes faecalis	**0.001**	0.371	**0.005**	**0.005**	**0.018**	0.152
Alcanivorax dieselolei	**0.005**	0.253	0.178	**0.001**	**0.009**	**0.005**
Alcanivorax pacificus	**0.041**	0.867	0.096	0.062	0.145	0.145
Algibacter alginicilyticus	0.629	0.449	**0.027**	0.876	0.629	0.511
Alicyclobacillus acidocaldarius	**0.004**	0.343	**0.033**	**0.001**	**0.045**	**0.004**
Aliivibrio wodanis	0.442	0.442	**0.008**	0.793	0.442	0.312
Alkalilimnicola ehrlichii	**0.003**	0.554	**0.005**	**0.006**	0.11	**0.02**
Alkaliphilus oremlandii	**0.03**	0.174	0.092	0.223	0.223	0.991
Alkalitalea saponilacus	0.45	**0.013**	0.233	**0.013**	0.814	**<0.001**
Allochromatium vinosum	**0.011**	0.1	0.076	**0.002**	0.08	**0.002**
Allofrancisella guangzhouensis	**0.039**	0.193	**0.001**	0.193	0.983	0.064
Altererythrobacter atlanticus	**0.002**	0.818	0.285	**0.002**	**0.007**	0.285
Altererythrobacter dongtanensis	**0.029**	0.159	**0.006**	0.291	0.508	0.291
Altererythrobacter epoxidivorans	**0.007**	0.63	**0.007**	**0.001**	0.092	**0.008**
Altererythrobacter ishigakiensis	**0.023**	0.243	0.33	0.252	**0.046**	0.33

Altererythrobacter mangrovi	**0.004**	0.308	**0.017**	**0.002**	0.055	**0.002**
Altererythrobacter marensis	**0.005**	**0.034**	0.325	0.156	**0.034**	0.108
Altererythrobacter namhicola	**0.015**	0.712	**0.01**	**0.01**	0.396	**0.01**
Alteromonas australica	**0.027**	0.883	0.114	**0.017**	0.114	0.114
Aminobacter aminovorans	**0.023**	0.407	**0.023**	0.065	0.265	0.202
***Aminobacter* sp. MSH1**	**0.016**	0.893	**0.002**	**0.017**	0.417	**0.005**
Ammonifex degensii	**0.048**	0.531	**0.007**	0.188	0.531	**0.048**
Amycolatopsis japonica	**0.011**	0.383	**0.011**	**0.037**	0.302	0.103
Amycolatopsis mediterranei	**0.014**	0.43	**0.021**	**<0.001**	0.105	**<0.001**
***Amycolatopsis* sp. AA4**	0.151	0.4	0.081	0.081	0.579	**0.037**
***Anabaena* sp. 90**	**0.004**	0.095	**0.04**	**0.019**	**0.04**	0.632
Anaeromyxobacter dehalogenans	0.056	0.076	0.343	**0.004**	0.093	**0.004**
***Anaeromyxobacter* sp. Fw109 5**	**0.005**	0.49	**0.001**	**0.001**	0.463	**0.001**
***Anaeromyxobacter* sp. K**	**<0.001**	0.527	**<0.001**	**0.001**	**0.022**	**0.005**
Anaplasma centrale	**0.022**	NA	0.386	**0.025**	**0.025**	0.386
Anaplasma phagocytophilum	**0.002**	0.295	0.453	**<0.001**	**<0.001**	0.453
Anomala cuprea entomopoxvirus	0.202	**0.039**	0.798	0.798	0.202	**0.019**
Anoxybacillus amylolyticus	0.342	**0.019**	0.552	0.484	0.153	**0.006**
Anoxybacillus flavithermus	**0.007**	0.974	0.974	**0.018**	**0.011**	0.974
Anoxybacillus gonensis	0.212	**0.033**	0.108	**0.009**	0.675	**<0.001**

Antarctobacter heliothermus	**0.018**	0.344	0.138	0.054	**0.041**	0.833
Aquabacterium olei	0.683	0.168	0.73	**0.027**	0.683	**0.027**
***Aquaspirillum* sp. LM1**	**0.003**	0.955	0.507	**0.003**	**0.003**	0.434
Aquifex aeolicus	0.087	0.432	0.11	**0.038**	0.168	**0.038**
Archangium gephyra	**0.004**	0.173	**0.014**	**0.001**	0.099	**0.001**
Aromatoleum aromaticum	**0.005**	0.078	**0.032**	**<0.001**	**0.035**	**<0.001**
***Arsenicicoccus* sp. oral taxon 190**	**0.012**	0.572	**0.037**	0.102	0.265	0.238
Arthrobacter crystallopoietes	**0.007**	0.457	**0.013**	**0.002**	**0.027**	**0.002**
***Arthrobacter* sp. DCT5**	**<0.001**	0.805	**0.001**	**<0.001**	**<0.001**	**0.007**
***Arthrobacter* sp. ERGS1 01**	**0.046**	0.576	0.181	**0.03**	0.118	0.118
***Arthrobacter* sp. FB24**	**<0.001**	0.921	**0.031**	**<0.001**	**0.015**	**0.031**
***Arthrobacter* sp. QXT 31**	**0.004**	0.236	**0.033**	**0.003**	0.114	**0.004**
***Arthrobacter* sp. U41**	**0.009**	0.158	**<0.001**	0.087	0.784	**0.009**
***Arthrobacter* sp. YN**	**0.007**	0.05	0.993	**0.001**	**0.018**	0.05
Arthrobacter virus ArV1	0.103	**0.009**	0.161	NA	0.161	**0.043**
***Arthrospira* sp. TJSD092**	**0.005**	0.556	**0.008**	**0.003**	0.131	**0.005**
Artogeia rapae granulovirus	0.308	0.123	**0.014**	NA	NA	NA
Asticcacaulis excentricus	0.073	0.698	0.183	**0.049**	0.183	0.14
***Aureimonas* sp. AU20**	0.288	0.288	**0.04**	0.091	0.76	**0.002**
Aureococcus anophagefferens virus	**0.018**	0.782	**0.018**	**0.023**	0.65	**0.036**

Azoarcus communis	**0.041**	0.563	**0.041**	**0.041**	0.404	**0.041**
***Azoarcus* sp. CIB**	**0.04**	0.744	**0.01**	**0.031**	0.744	**0.01**
***Azoarcus* sp. KH32C**	**0.021**	0.225	0.07	**0.011**	0.085	**0.011**
***Azoarcus* sp. SY39**	**<0.001**	0.451	**0.003**	**0.003**	**0.006**	0.075
Azorhizobium caulinodans	**0.008**	0.599	**0.008**	**0.008**	0.387	**0.008**
Azospira oryzae	**0.01**	0.885	**0.01**	**0.01**	0.253	**0.01**
Azospirillum brasilense	**0.034**	0.115	**0.04**	**0.004**	0.252	**0.002**
Azospirillum lipoferum	**0.001**	0.89	**0.001**	**0.001**	**0.038**	**0.002**
***Azospirillum* sp. TSH58**	**<0.001**	0.955	**0.004**	**0.002**	**0.037**	**0.009**
Azotobacter chroococcum	**0.001**	0.61	**0.001**	**0.001**	**0.039**	**0.001**
Azotobacter vinelandii	**0.013**	0.669	**0.008**	**0.008**	0.417	**0.007**
Bacillus altitudinis	0.301	0.516	**0.031**	0.265	0.097	0.127
Bacillus krulwichiae	**0.019**	0.176	0.242	0.45	0.176	0.5
Bacillus licheniformis	0.487	0.279	0.418	0.867	0.279	**0.035**
Bacillus methanolicus	0.068	**0.003**	0.068	0.794	0.282	0.068
Bacillus muralis	0.701	**0.011**	0.428	0.077	0.689	**0.033**
Bacillus pseudofirmus	**0.028**	0.337	0.337	**0.012**	**0.012**	0.931
Bacillus smithii	0.773	**0.007**	0.878	0.189	0.773	**0.003**
***Bacillus* sp. HBCD sjtu**	0.091	**0.004**	0.16	0.319	0.635	**0.008**
***Bacillus* sp. JS**	0.222	**0.029**	0.222	0.603	0.671	0.222

***Bacillus* sp. Y 01**	0.268	0.268	0.268	**0.044**	**0.012**	0.541
Bacillus thuringiensis	**0.033**	0.841	0.841	**0.014**	**0.014**	0.889
Bacillus virus TsarBomba	0.456	0.456	0.231	0.456	0.721	**0.04**
***Bacteroidales* bacterium CF**	0.916	0.216	0.216	0.216	0.409	**0.002**
Bacteroides caecimuris	0.111	**0.035**	0.882	0.889	0.111	**0.03**
Bacteroides cellulosilyticus	0.298	**0.011**	0.397	0.397	0.148	**0.001**
Bacteroides dorei	**0.012**	**0.002**	0.112	0.454	**0.001**	**<0.001**
Bacteroides helcogenes	0.374	**0.012**	0.808	0.472	0.374	**0.009**
Bacteroides heparinolyticus	0.392	**0.006**	0.707	0.392	0.392	**0.006**
Bacteroides ovatus	**0.029**	0.823	0.823	**0.029**	**0.021**	0.823
Bacteroides salanitronis	0.542	**0.026**	**0.046**	**0.048**	0.542	**<0.001**
Bacteroides thetaiotaomicron	0.101	**0.002**	0.676	0.496	**0.038**	**<0.001**
Bacteroides vulgatus	**0.045**	0.155	0.05	0.161	0.155	0.862
Bacteroides zoogleoformans	0.694	**0.007**	0.851	0.234	0.694	**0.004**
Barnesiella viscericola	0.21	0.21	0.06	0.073	0.955	**0.004**
Bartonella apis	**0.007**	**0.048**	**0.01**	0.084	**0.048**	0.94
Bartonella tribocorum	**0.044**	0.18	0.449	0.18	0.065	0.413
Bdellovibrio bacteriovorus	**0.037**	0.838	**0.041**	**0.041**	0.467	0.068
Beihai picorna like virus 116	0.38	0.201	**0.044**	NA	NA	NA
Beihai sobemo like virus 19	0.068	NA	NA	0.087	**0.015**	NA

Belliella baltica	0.362	0.314	0.327	0.062	0.994	**0.024**
Bernardetia litoralis	0.321	0.2	0.321	0.321	0.146	**0.005**
Beta *proteobacterium* CB	0.743	**0.048**	0.971	0.561	0.743	**0.028**
***Betaproteobacteria* bacterium GR16 43**	**0.013**	0.895	0.081	**0.021**	**0.027**	0.114
Beutenbergia cavernae	**0.008**	0.757	**0.012**	**0.008**	**0.036**	**0.016**
Bifidobacterium adolescentis	**0.043**	0.333	0.993	0.317	**0.04**	0.333
Bifidobacterium animalis	0.494	0.079	0.255	0.056	0.825	**0.026**
Bifidobacterium asteroides	**0.042**	**0.042**	**0.042**	0.585	0.484	0.585
Bifidobacterium thermophilum	**0.047**	0.416	0.491	**0.004**	**0.047**	0.078
Blastochloris viridis	**0.002**	**0.012**	**<0.001**	0.114	0.229	0.229
Blastococcus saxobsidens	**0.009**	0.838	0.057	**0.009**	0.059	0.057
***Blastomonas* sp. RAC04**	**0.032**	0.901	0.378	**0.038**	0.056	0.427
***Blattabacterium* sp. Blaberus giganteus**	0.24	0.053	**0.019**	0.896	0.896	0.896
***Blattabacterium* sp. Blatta orientalis**	0.13	0.062	0.724	**0.003**	0.066	0.062
***Blattabacterium* sp. Nauphoeta cinerea**	**0.027**	0.757	0.157	0.053	0.157	0.26
Blautia hansenii	0.246	**0.042**	0.375	0.497	0.085	**0.002**
***Blautia* sp. N6H1 15**	0.589	0.19	0.155	0.19	0.69	**0.004**
***Blautia* sp. YL58**	0.494	0.378	0.258	0.327	0.809	**0.029**
Bordetella bronchiseptica	0.111	0.111	0.828	**0.011**	0.112	0.05
Bordetella flabilis	**0.001**	0.8	**0.002**	**0.001**	**0.005**	**0.016**

***Bordetella* genomosp. 13**	**0.003**	0.833	**0.003**	**0.003**	0.204	**0.003**
***Bordetella* genomosp. 6**	**0.028**	**0.042**	0.096	0.378	0.096	0.306
***Bordetella* genomosp. 9**	**0.005**	0.541	**0.005**	**0.002**	**0.042**	**0.002**
Bordetella hinzii	**0.003**	0.767	0.058	**0.003**	**0.015**	**0.034**
Bordetella holmesii	0.76	0.207	**0.018**	0.635	0.437	0.76
Bordetella parapertussis	**0.003**	0.185	0.066	**0.001**	0.066	**0.013**
Bordetella petrii	**0.023**	**0.04**	**0.011**	**0.003**	0.374	**<0.001**
Bordetella pseudohinzii	0.066	0.626	**0.034**	0.066	0.496	**0.034**
***Bordetella* sp. H567**	0.053	0.885	**0.023**	**0.023**	0.392	**0.023**
Bordetella trematum	**0.017**	0.916	**0.001**	**0.017**	0.879	**0.001**
***Bosea* sp. AS 1**	**<0.001**	0.523	**0.004**	**0.001**	**0.009**	**0.018**
***Bosea* sp. RAC05**	0.057	0.901	**0.008**	0.057	0.892	**0.025**
Bosea vaviloviae	**<0.001**	0.324	**0.026**	**<0.001**	**0.016**	**0.004**
Brachybacterium ginsengisoli	0.152	0.412	0.154	**0.033**	0.422	**0.033**
***Brachybacterium* sp. P6 10 X1**	0.087	0.797	**0.028**	**0.049**	0.672	**0.028**
***Brachybacterium* sp. VM2412**	**0.012**	0.717	**0.003**	0.074	0.677	**0.047**
***Brachybacterium* sp. VR2415**	**0.016**	0.205	0.097	0.065	0.097	0.735
Brachyspira pilosicoli	0.989	0.558	**0.032**	0.558	0.103	0.103
Bradyrhizobium diazoefficiens	**0.001**	0.823	**0.001**	**0.001**	0.165	**0.001**
Bradyrhizobium japonicum	**0.011**	0.195	**0.029**	**0.002**	0.08	**0.002**

Bradyrhizobium oligotrophicum		0.117		0.921		**0.015**	0.094	0.921	**0.015**
Bradyrhizobium ottawaense		0.066		0.678		**0.026**	**0.039**	0.4	**0.026**
***Bradyrhizobium* sp.**		**<0.001**		0.658		0.148	**0.004**	**0.004**	0.421
***Bradyrhizobium* sp. 2 39S1MB**		0.17		**0.027**		0.115	**0.008**	0.955	**0.001**
***Bradyrhizobium* sp. BTAi1**		**0.014**		0.742		**0.014**	**0.014**	0.074	**0.041**
***Bradyrhizobium* sp. CCGE LA001**		0.121		0.942		**0.036**	0.121	0.931	**0.036**
***Bradyrhizobium* sp. ORS 278**		**0.006**		0.828		0.065	**0.006**	**0.029**	0.065
***Bradyrhizobium* sp. ORS 285**		0.467		0.216		0.216	0.12	0.904	**0.01**
***Bradyrhizobium* sp. SK17**		**0.02**		0.916		0.431	**0.02**	**0.02**	0.431
Brevefilum fermentans		**0.027**		0.995		0.302	**0.027**	0.07	0.302
Brevibacillus formosus		0.701		0.3		0.557	0.583	0.557	**0.037**
Brevibacterium aurantiacum		0.068		0.601		0.124	**0.03**	0.404	0.068
Brevibacterium linens		**0.016**		0.742		0.105	**0.013**	**0.041**	0.07
***Brevundimonas* sp. DS20**		**0.008**		0.979		**0.008**	**0.015**	0.217	0.05
***Brevundimonas* sp. LM2**		**0.026**		0.493		0.974	**0.007**	**0.009**	0.427
Brevundimonas subvibrioides		**0.008**		0.67		0.302	**0.025**	0.061	0.67
Brevundimonas vesicularis		0.428		0.545		**0.002**	0.527	0.527	**0.033**
Brochothrix phage A9	NA			**0.027**	NA		0.231	NA	**0.003**
Brucella pinnipedialis		**0.007**	NA		NA		**0.015**	**<0.001**	NA
Brucella vulpis		0.969		**0.015**		0.969	0.195	0.969	**0.015**

Burkholderia ambifaria	**0.022**	0.858	**0.022**	**0.022**	0.561	**0.022**
Burkholderia cenocepacia	0.159	0.067	0.339	**0.024**	0.289	**0.024**
Burkholderia cepacia	0.071	0.333	0.307	**0.02**	0.167	0.071
Burkholderia gladioli	**0.004**	0.486	**0.004**	**0.004**	0.215	**0.004**
Burkholderia glumae	**0.04**	0.739	0.119	**0.04**	0.101	0.226
Burkholderia insecticola	**0.002**	0.33	**0.011**	**0.002**	**0.018**	**0.002**
Burkholderia lata	**0.04**	0.479	0.054	**0.04**	0.057	0.164
Burkholderia latens	**0.043**	0.268	0.969	**0.015**	**0.024**	0.268
Burkholderia mallei	**0.001**	0.097	.	0.07	**<0.001**	**0.034**
Burkholderia metallica	**0.006**	0.759	0.071	**0.004**	**0.046**	0.055
Burkholderia multivorans	0.051	0.628	**0.045**	**0.045**	0.853	**0.045**
Burkholderia pseudomallei	0.408	0.237	0.237	**0.02**	0.787	**0.002**
Burkholderia pseudomultivorans	0.8	0.8	**0.008**	0.8	0.086	**0.019**
Burkholderia pyrrocinia	**0.01**	0.26	**0.004**	0.18	0.316	0.206
Burkholderia seminalis	0.092	0.592	**<0.001**	0.223	0.736	**0.013**
***Burkholderia* sp. BDU6**	**0.042**	0.392	**0.042**	**0.022**	0.258	**0.022**
***Burkholderia* sp. Bp7605**	0.086	0.984	**0.025**	0.13	0.766	0.088
***Burkholderia* sp. CCGE1002**	**0.026**	0.36	**0.036**	0.054	0.177	0.177
***Burkholderia* sp. MSMB617WGS**	0.308	0.534	0.086	0.207	0.534	**0.01**
***Burkholderia* sp. OLGA172**	0.512	0.106	0.961	**0.017**	0.424	**0.024**

***Burkholderia* sp. PAMC 26561**	0.325	0.492	**0.011**	0.672	0.672	0.325
***Burkholderia* sp. PAMC 28687**	**<0.001**	0.226	**0.008**	**0.005**	**0.025**	0.205
Burkholderia stabilis	0.256	0.256	0.353	**0.028**	0.299	**0.028**
Burkholderia stagnalis	**0.014**	0.336	0.323	**0.014**	0.067	0.078
Burkholderia territorii	**0.002**	0.162	**0.008**	**0.002**	**0.014**	**0.001**
Burkholderia ubonensis	**0.011**	0.139	**0.01**	**0.002**	0.139	**<0.001**
Burkholderia vietnamiensis	**<0.001**	0.884	0.406	**<0.001**	**<0.001**	0.406
***Burkholderiales* bacterium GJ E10**	0.512	**0.019**	0.285	**0.019**	0.983	**0.001**
***Burkholderiales* bacterium YL45**	0.832	0.495	**0.01**	0.832	**0.039**	**0.004**
Caldicellulosiruptor bescii	0.745	**0.044**	**0.015**	0.054	**0.044**	0.836
Caldicellulosiruptor kristjanssonii	**0.001**	0.657	0.906	**0.001**	**0.001**	0.657
Caldicellulosiruptor obsidiansis	0.555	0.164	**0.048**	0.862	0.749	0.749
***Calothrix* sp. 336 3**	0.516	**0.029**	0.272	**0.029**	0.147	0.057
***Calothrix* sp. NIES 4071**	**0.048**	0.385	0.384	0.153	**0.005**	0.153
***Calothrix* sp. NIES 4101**	0.284	0.245	0.245	**0.047**	0.623	**0.01**
***Calothrix* sp. PCC 7507**	0.167	0.324	**0.03**	0.324	0.434	0.364
Campylobacter cuniculorum	0.329	0.199	**0.023**	0.972	0.689	0.689
Campylobacter curvus	**0.001**	0.667	**0.002**	**0.002**	**0.021**	**0.015**
Campylobacter gracilis	**0.009**	0.766	**0.039**	**0.027**	0.06	0.11
Campylobacter hominis	**0.045**	0.543	0.084	0.084	0.205	0.286

Campylobacter hyointestinalis	**0.043**	0.732	0.658	0.091	0.091	0.732
Campylobacter iguaniorum	0.904	**0.002**	**0.002**	**0.019**	**0.026**	0.827
Campylobacter insulaenigrae	**0.009**	0.558	**0.018**	**0.018**	0.099	0.099
Campylobacter jejuni	**<0.001**	**0.049**	0.275	**<0.001**	**<0.001**	**0.001**
Campylobacter lari	**0.027**	0.742	0.155	**0.027**	0.065	0.255
Campylobacter sputorum	**0.024**	0.275	0.116	**0.005**	0.244	**0.018**
Candidatus Accumulibacter phosphatis	0.066	0.359	0.108	**0.017**	0.359	**0.017**
Candidatus Amoebophilus asiaticus	0.088	0.875	**0.026**	0.061	0.875	**0.026**
Candidatus Azobacteroides pseudotrichonymphae	0.177	0.245	0.11	0.054	0.589	**0.006**
Candidatus Baumannia cicadellinicola	**0.002**	0.666	0.666	**0.002**	**0.003**	0.57
Candidatus Carsonella ruddii	**0.013**	0.197	0.06	0.253	0.341	0.554
Candidatus Cloacimonas acidaminovorans	0.051	**0.023**	0.947	0.947	**0.023**	**0.006**
Candidatus Desulforudis audaxviator	0.106	0.371	0.309	**0.049**	0.16	0.106
Candidatus Endolissoclinum faulkneri	**0.016**	0.32	0.121	**0.016**	0.084	**0.048**
Candidatus Filomicrobium marinum	0.286	**0.029**	0.603	**0.021**	0.125	**0.029**
Candidatus Fokinia solitaria	0.064	0.348	0.172	**0.015**	**0.003**	0.86
Candidatus Kinetoplastibacterium oncopeltii	0.481	0.133	0.411	0.481	0.411	**0.034**
Candidatus Koribacter versatilis	0.091	0.195	0.091	**0.001**	0.195	**<0.001**
Candidatus Liberibacter solanacearum	**0.028**	**0.048**	0.272	**<0.001**	**0.048**	**0.002**

Candidatus Moranella endobia	0.76	**<0.001**	0.056	0.061	0.344	0.061
Candidatus Nanopelagicus hibericus	0.503	**0.005**	0.111	0.331	0.765	0.149
Candidatus Nitrospira inopinata	**0.004**	0.702	0.093	**0.001**	**0.007**	**0.036**
Candidatus Paracaedibacter acanthamoebae	**0.036**	0.463	**0.041**	0.063	0.233	0.223
***Candidatus Pelagibacter* sp. IMCC9063**	**0.014**	0.994	0.109	**0.014**	0.065	0.176
Candidatus Pelagibacter ubique	**0.001**	0.252	**0.031**	**0.017**	**0.04**	0.307
Candidatus Portiera aleyrodidarum	**0.049**	0.617	0.584	**0.049**	0.091	0.409
Candidatus Profftella armatura	0.123	0.264	**0.002**	0.434	0.872	0.123
Candidatus Promineofilum breve	**0.002**	0.906	**0.017**	**0.004**	**0.017**	**0.024**
Candidatus Protochlamydia naegleriophila	**0.008**	0.337	0.281	**0.033**	**0.033**	0.941
Candidatus Puniceispirillum marinum	**0.027**	0.865	0.865	**0.027**	**0.027**	0.865
Candidatus Rhodoluna planktonica	**0.001**	0.921	0.164	**0.001**	**0.008**	0.2
Candidatus Riesia pediculicola	0.119	0.463	0.416	0.345	**0.029**	0.227
Candidatus Ruthia magnifica	0.136	0.091	0.729	0.729	0.091	**0.033**
***Candidatus Saccharibacteria* oral taxon TM7x**	**0.033**	0.976	0.976	**0.033**	**0.033**	0.976
Candidatus Sodalis pierantonius	0.314	0.427	**0.001**	0.557	0.427	0.059
Candidatus Solibacter usitatus	**0.021**	0.757	0.137	**0.021**	0.058	0.137
Candidatus Thiodictyon syntrophicum	**0.013**	0.462	**0.001**	**0.013**	0.541	**0.009**
Candidatus Thioglobus autotrophicus	**0.003**	0.807	0.807	**0.003**	**0.003**	0.807
***Candidatus Xiphinematobacter* sp. Idaho**	0.205	0.265	0.164	0.071	0.759	**0.034**

Grape						
Capnocytophaga gingivalis	0.075	0.699	**0.023**	**0.026**	0.76	**0.005**
Capnocytophaga haemolytica	0.08	0.289	0.186	**0.038**	0.279	**0.038**
Capnocytophaga ochracea	0.257	0.257	0.257	0.784	0.125	**0.032**
***Capnocytophaga* sp. oral taxon 864**	0.894	0.118	0.266	0.269	0.34	**0.001**
***Capnocytophaga* sp. oral taxon 878**	0.863	0.097	0.547	0.13	0.797	**0.004**
Capnocytophaga sputigena	0.521	**0.024**	0.856	**0.022**	0.424	**0.022**
Capnocytophaga stomatis	0.589	0.589	0.152	0.152	0.589	**0.005**
Castellaniella defragrans	0.066	0.223	**0.004**	0.223	0.754	0.223
***Catenovulum* sp. CCB QB4**	0.457	0.293	0.431	0.14	0.814	**0.049**
Catenulispora acidiphila	**0.002**	0.411	**0.004**	**0.002**	0.11	**0.002**
Caulobacter henricii	**0.002**	0.203	**0.019**	**0.024**	**0.016**	0.651
Caulobacter mirabilis	**0.002**	0.355	**0.04**	**0.049**	0.052	0.433
Caulobacter segnis	**0.026**	0.554	**0.026**	0.077	0.256	0.12
***Caulobacter* sp. K31**	**0.022**	0.467	**0.024**	**0.001**	0.206	**0.001**
Caulobacter vibrioides	**0.001**	0.885	**0.018**	**0.001**	**0.018**	**0.026**
Cedecea neteri	**0.007**	0.803	0.225	**0.002**	**0.013**	0.332
Celeribacter manganoxidans	**0.026**	0.509	**0.026**	0.209	0.21	0.21
Celeribacter marinus	**0.001**	0.788	0.077	**0.002**	**0.002**	0.255
Cellulomonas fimi	**0.01**	0.291	0.291	**0.006**	**0.036**	0.05

Cellulomonas flavigena	**0.028**	0.664	**0.038**	**0.038**	0.137	**0.046**
Cellulomonas gilvus	**0.012**	0.617	**0.048**	**0.012**	0.099	**0.046**
Cellulophaga lytica	0.488	0.614	**0.007**	0.614	0.614	0.081
Cellulosimicrobium cellulans	0.062	0.239	**0.017**	**0.017**	0.589	**0.001**
***Cellulosimicrobium* sp. TH 20**	0.192	0.187	0.187	**0.02**	0.536	**0.02**
Cellvibrio japonicus	**0.025**	0.251	0.09	**0.006**	0.215	**0.009**
***Cellvibrio* sp. PSBB023**	0.054	0.343	0.149	**0.014**	0.186	**0.018**
***Chelativorans* sp. BNC1**	**0.01**	0.712	0.187	**0.01**	**0.03**	0.053
Chelatococcus daeguensis	**0.003**	0.916	**0.01**	**0.005**	0.064	**0.012**
***Chelatococcus* sp. CO 6**	**0.001**	0.281	**0.001**	**0.001**	0.167	**0.001**
Chitinophaga caeni	0.06	**0.047**	0.058	**0.005**	0.574	**<0.001**
Chloracidobacterium thermophilum	**0.003**	0.395	**0.031**	**0.031**	**0.031**	0.395
Chlorobaculum limnaeum	**0.003**	0.599	**0.009**	**0.003**	0.102	**0.009**
Chlorobaculum parvum	**0.038**	0.942	0.171	**0.04**	0.155	0.191
Chlorobaculum tepidum	**0.034**	0.608	**0.004**	**0.046**	0.619	**0.035**
Chlorobium chlorochromatii	**0.02**	0.093	0.134	**0.005**	0.088	**0.005**
Chlorobium limicola	**<0.001**	0.473	0.07	**<0.001**	**0.004**	**0.006**
Chlorobium phaeovibrioides	0.829	**0.016**	0.736	**0.038**	0.955	**0.001**
Chloroflexus aggregans	**0.004**	0.895	0.373	**0.004**	**0.021**	0.373
Chloroherpeton thalassium	**0.013**	0.063	0.326	**<0.001**	**0.007**	**0.001**

Chondrocystis sp. NIES 4102	0.947	**0.001**	0.06	0.052	0.526	**0.011**
Chondromyces crocatus	0.09	**0.035**	0.152	**0.001**	0.283	**<0.001**
Christensenella massiliensis	**0.005**	0.167	**0.008**	**0.02**	0.081	0.167
***Chromatiaceae* bacterium 2141T STBD 0c 01a**	0.093	0.67	**0.011**	**0.011**	0.749	**<0.001**
***Chromobacterium* sp. ATCC 53434**	**<0.001**	0.919	**0.039**	**<0.001**	**0.004**	**0.039**
***Chromobacterium* sp. IIBBL 274 1**	0.055	0.689	**0.003**	**0.031**	0.845	**0.003**
Chromobacterium vaccinii	**0.002**	0.426	**0.001**	**0.004**	0.165	**0.008**
Chroococcidiopsis thermalis	**<0.001**	0.974	**0.01**	**0.001**	**0.015**	**0.015**
Chryseobacterium gallinarum	0.184	0.184	0.394	**0.04**	0.184	**0.04**
Chryseobacterium taklimakanense	0.253	0.306	0.08	0.098	0.848	**0.007**
Citrobacter farmeri	0.097	0.328	0.5	**0.028**	**0.028**	0.5
Citrobacter freundii	**0.01**	**0.001**	**0.01**	0.741	0.11	**0.013**
***Citrobacter freundii* complex sp. CFNIH9**	0.76	0.319	0.171	0.715	0.171	**0.011**
***Citrobacter* sp. 86**	NA	0.062	NA	0.304	NA	**0.013**
***Citromicrobium* sp. JL477**	**0.004**	0.26	**0.045**	**0.003**	**0.029**	**0.004**
Clavibacter capsici	**0.004**	0.751	**0.022**	**0.008**	**0.022**	0.1
Clavibacter michiganensis	**0.01**	0.281	**0.018**	0.07	0.093	0.318
Cloacibacillus porcorum	**0.002**	0.35	**0.012**	**0.009**	**0.04**	0.14
Clostridioides difficile	**0.038**	**0.038**	0.314	0.52	0.058	0.081
Clostridium cellulosi	**0.001**	0.594	**0.014**	**0.001**	**0.014**	0.073

Cnuibacter physcomitrellae	**0.01**	0.494	0.056	**0.01**	**0.048**	**0.02**
Collimonas arenae	**0.006**	0.219	**0.038**	**0.004**	0.082	**0.006**
Collimonas pratensis	**0.006**	0.519	**0.035**	**0.035**	**0.035**	0.391
***Comamonadaceae* bacterium A1**	**0.006**	0.83	0.085	**0.006**	**0.029**	0.066
***Comamonadaceae* bacterium B1**	**0.005**	0.475	**<0.001**	**0.001**	0.696	**<0.001**
Comamonas kerstersii	**0.045**	0.571	**0.047**	0.091	0.188	0.188
Comamonas serinivorans	**0.016**	0.189	**0.005**	**0.001**	0.359	**<0.001**
Comamonas testosteroni	0.07	0.242	0.122	**0.012**	0.122	**0.012**
Conexibacter woesei	**0.017**	0.081	0.062	**0.002**	0.193	**0.001**
***Confluentimicrobium* sp. EMB200 NS6**	**0.005**	0.126	**0.003**	0.111	0.12	0.364
Coraliomargarita akajimensis	**0.005**	0.963	0.316	**0.005**	0.068	0.321
Corallococcus coralloides	**0.004**	0.554	**0.018**	**0.009**	**0.018**	0.08
***Coriobacteriaceae* bacterium 68 1 3**	**0.032**	0.4	0.268	**0.002**	0.064	**0.032**
Corynebacterium aquilae	0.207	0.703	**0.005**	0.057	0.541	**<0.001**
Corynebacterium argentoratense	**0.001**	0.15	0.116	**0.014**	**0.002**	0.997
Corynebacterium atypicum	**0.003**	0.961	0.052	**0.003**	**0.01**	0.055
Corynebacterium aurimucosum	**0.032**	0.927	**0.021**	**0.032**	0.506	**0.032**
Corynebacterium crudilactis	**0.008**	0.419	0.101	**0.008**	0.101	0.056
Corynebacterium diphtheriae	**0.021**	0.3	0.111	**0.017**	0.212	**0.021**
Corynebacterium doosanense	**0.001**	0.567	**0.003**	**<0.001**	0.107	**<0.001**

Corynebacterium epidermidicanis	**0.032**	**0.028**	0.854	**<0.001**	**0.033**	**0.028**
Corynebacterium falsenii	0.357	0.797	**0.013**	0.357	0.748	**0.013**
Corynebacterium frankenforstense	**0.001**	0.26	**0.003**	**0.001**	**0.007**	**0.045**
Corynebacterium glaucum	**0.026**	0.617	**0.019**	**0.019**	0.444	**0.019**
Corynebacterium glutamicum	**0.002**	0.803	**0.029**	**0.002**	**0.034**	**0.029**
Corynebacterium humireducens	**0.006**	0.232	**0.021**	0.062	0.117	0.336
Corynebacterium imitans	**0.046**	0.968	0.187	0.107	0.107	0.5
Corynebacterium jeikeium	**0.049**	0.882	**0.003**	**0.003**	0.882	**0.002**
Corynebacterium lactis	**0.001**	**0.043**	**0.005**	**0.012**	**0.043**	0.498
Corynebacterium marinum	**0.039**	0.654	**0.044**	**0.034**	0.164	0.056
Corynebacterium maris	**0.01**	0.343	**0.024**	**0.01**	**0.048**	**0.011**
Corynebacterium renale	0.134	0.095	0.112	**0.021**	**0.016**	0.418
***Corynebacterium* sp. 2183**	**0.007**	0.631	**0.007**	0.054	0.265	0.077
***Corynebacterium* sp. 2184**	**0.025**	0.06	0.144	**0.003**	0.144	**0.006**
***Corynebacterium* sp. NML98 0116**	0.429	0.259	0.101	0.101	0.727	**0.002**
Corynebacterium sphenisci	**0.003**	0.937	**0.006**	**0.003**	0.061	**0.006**
Corynebacterium striatum	0.182	0.4	0.094	0.094	0.744	**0.028**
Corynebacterium terpenotabidum	**0.015**	0.978	0.1	**0.015**	0.057	0.1
Corynebacterium testudinoris	**0.043**	0.777	**0.041**	**0.043**	0.736	**0.041**
Corynebacterium uterequi	**0.009**	0.831	**0.002**	**0.002**	0.409	**<0.001**

Corynebacterium variabile	0.179	0.051	0.179	**0.01**	0.639	**0.003**
Corynebacterium vitaeruminis	**0.001**	0.59	**0.012**	**0.003**	**0.003**	0.184
Croceibacter atlanticus	0.326	**0.016**	0.326	**0.016**	0.495	**<0.001**
Croceicoccus marinus	0.054	0.963	**0.001**	0.054	0.963	**0.002**
Croceicoccus naphthovorans	**0.002**	0.688	**0.011**	**0.002**	**0.007**	0.059
Cronobacter dublinensis	0.675	**0.034**	0.675	0.546	0.675	**0.034**
Cronobacter malonaticus	**0.049**	0.423	0.707	**0.049**	**0.049**	0.289
Cronobacter muytjensii	**0.006**	0.589	0.069	**0.006**	0.069	0.069
Cryobacterium arcticum	**0.004**	0.617	**0.014**	**0.001**	0.112	**0.004**
***Cryobacterium* sp. LW097**	**0.022**	0.708	**0.022**	**0.003**	0.167	**0.003**
Cryptobacterium curtum	**0.006**	0.083	0.14	**0.001**	**0.027**	**0.002**
Cupriavidus basilensis	**0.002**	0.89	**0.007**	**0.002**	**0.014**	**0.011**
Cupriavidus gilardii	**0.019**	0.655	**0.019**	**0.019**	0.099	0.075
Cupriavidus metallidurans	0.326	0.051	**0.036**	**0.017**	0.837	**<0.001**
Cupriavidus pinatubonensis	0.066	0.458	**0.015**	0.132	0.468	0.132
***Cupriavidus* sp. NH9**	**0.038**	0.292	0.125	**0.016**	0.174	**0.023**
Curtobacterium pusillum	**0.014**	0.303	0.241	**0.004**	0.056	0.056
***Curtobacterium* sp. BH 2 1 1**	0.07	0.152	0.193	**0.019**	0.152	**0.019**
Cutibacterium avidum	**0.003**	0.231	0.074	**0.033**	**0.02**	0.877
Cutibacterium granulosum	0.073	0.376	0.368	**0.023**	0.108	0.073

Cyanobium gracile	**0.001**	0.385	**<0.001**	**0.003**	0.174	**0.001**
***Cyanobium* sp. NIES 981**	**0.038**	0.471	**0.026**	**0.01**	0.471	**0.009**
Cyanophage Syn30	0.068	NA	NA	0.087	**0.015**	NA
***Cyanothece* sp. PCC 7424**	**<0.001**	**<0.001**	**0.004**	0.347	**0.011**	**0.029**
***Cyanothece* sp. PCC 8801**	**0.006**	**0.002**	**0.008**	0.952	0.336	0.167
***Cyanothece* sp. PCC 8802**	0.15	0.063	0.15	**0.003**	**0.003**	0.167
Cyclobacterium marinum	0.098	0.67	**0.013**	**0.026**	0.771	**0.002**
Cyprinid herpesvirus 2	0.068	NA	NA	0.087	**0.015**	NA
Cyprinid herpesvirus 3	0.068	NA	NA	0.087	**0.015**	NA
Cystobacter fuscus	**0.003**	0.532	**0.017**	**0.002**	**0.017**	**0.005**
Dechloromonas aromatica	0.367	0.367	**0.036**	0.52	0.966	0.367
Defluviimonas alba	**0.002**	0.342	**0.004**	**0.012**	**0.012**	0.156
Defluviitoga tunisiensis	0.38	**0.026**	0.38	0.38	0.411	**0.039**
Deinococcus actinosclerus	**0.003**	0.536	**0.004**	**0.003**	0.13	**0.003**
Deinococcus geothermalis	**0.009**	0.494	**0.007**	**0.018**	0.158	**0.03**
Deinococcus gobiensis	**0.007**	0.594	**0.008**	**0.007**	0.128	**0.007**
Deinococcus peraridilitoris	**0.023**	0.599	0.172	**0.021**	0.095	0.095
Deinococcus proteolyticus	0.132	0.316	**0.016**	0.316	0.685	0.316
Deinococcus puniceus	0.328	0.459	**0.002**	0.459	0.459	**0.029**
Deinococcus radiodurans	**0.006**	0.103	**0.038**	**0.001**	0.157	**0.001**

Deinococcus soli	**0.028**	0.864	**0.028**	**0.028**	0.167	**0.046**
***Deinococcus* sp. 17bor 2**	**0.034**	0.655	0.05	**0.034**	0.272	**0.034**
***Deinococcus* sp. NW 56**	**0.027**	0.619	**0.027**	0.057	0.247	0.161
Deinococcus swuensis	**0.025**	0.18	0.18	**0.02**	0.171	**0.025**
***Delftia* sp. HK171**	0.879	**0.048**	0.303	0.239	0.633	0.155
Delftia tsuruhatensis	0.209	0.209	0.209	0.098	0.536	**0.041**
Denitrobacterium detoxificans	**0.032**	0.191	0.861	**0.003**	**0.022**	0.05
Dermacoccus nishinomiyaensis	**0.013**	0.717	**0.02**	**0.013**	0.158	**0.013**
Dermatophilus congolensis	**0.001**	**0.002**	**0.002**	0.356	**0.028**	0.399
Desulfarculus baarsii	**0.008**	0.958	**0.008**	**0.009**	0.098	**0.016**
Desulfatibacillum alkenivorans	**0.021**	0.28	**0.014**	**0.002**	0.319	**0.001**
Desulfobacca acetoxidans	**0.011**	0.771	**0.016**	**0.013**	**0.048**	**0.016**
***Desulfobulbus* sp. ORNL**	**0.012**	0.34	**0.037**	**0.012**	0.059	0.215
Desulfococcus multivorans	0.185	0.346	**0.003**	0.346	0.859	0.112
Desulfococcus oleovorans	**0.003**	0.457	**0.006**	**0.023**	0.131	0.05
Desulfofarcimen acetoxidans	**0.017**	0.34	**0.033**	**0.033**	0.136	0.218
Desulfomicrobium baculatum	0.066	0.554	**0.043**	**0.017**	0.554	**0.017**
Desulfomicrobium orale	0.053	0.937	0.053	**0.007**	0.404	**0.007**
Desulfosporosinus meridiei	**0.047**	0.191	0.119	0.174	0.174	0.679
Desulfotomaculum ferrireducens	**0.037**	0.901	**0.039**	**0.037**	0.238	**0.037**

Desulfovibrio africanus	**0.001**	0.974	**0.005**	**0.001**	**0.032**	**0.009**
Desulfovibrio alaskensis	**0.033**	0.411	0.095	**0.003**	0.095	**0.007**
Desulfovibrio desulfuricans	0.152	0.152	0.177	**0.008**	0.177	**0.001**
Desulfovibrio fairfieldensis	**0.006**	0.163	**0.006**	**0.01**	0.088	0.088
Desulfovibrio magneticus	**<0.001**	0.984	**0.002**	**<0.001**	**0.002**	**0.002**
Desulfovibrio piger	**<0.001**	0.129	**0.001**	**0.021**	**0.045**	0.076
***Desulfovibrio* sp. G11**	**0.001**	0.139	**0.001**	0.075	0.2	0.139
Desulfovibrio vulgaris	0.058	0.396	**0.041**	**0.007**	0.339	**0.002**
Desulfurispirillum indicum	**0.008**	0.054	**0.018**	0.054	0.054	0.892
Desulfurobacterium thermolithotrophum	**0.001**	**0.016**	**0.022**	**0.037**	**0.016**	0.458
Desulfuromonas soudanensis	**0.033**	0.262	**0.002**	**0.013**	0.842	**<0.001**
***Desulfuromonas* sp. DDH964**	**0.001**	0.314	**0.011**	**0.001**	**0.04**	**0.001**
***Devosia* sp. A16**	**0.021**	0.971	**0.021**	**0.034**	0.238	**0.034**
***Devosia* sp. H5989**	**0.004**	0.945	**0.002**	**0.004**	0.183	**0.004**
***Devosia* sp. I507**	**0.005**	0.088	**0.013**	0.052	0.052	0.606
Dialister pneumosintes	0.187	**<0.001**	0.33	0.183	0.351	**<0.001**
Dichelobacter nodosus	**0.02**	0.841	0.892	**0.023**	**0.013**	0.841
Dickeya dianthicola	**0.021**	0.558	0.382	**0.001**	0.054	0.266
Dickeya solani	0.419	0.326	0.419	**0.028**	0.468	**0.028**
Dictyoglomus turgidum	**0.011**	**0.011**	**0.005**	0.254	0.337	0.83

Dietzia lutea	**0.009**		0.817		**0.009**		**0.009**		**0.048**		**0.009**	
Dietzia psychralcaliphila	0.39		0.47		**0.006**		0.304		0.387		**0.003**	
***Dietzia* sp. JS16 p6b**	**0.002**		0.206		**0.035**		**0.035**		**0.035**		0.608	
Dinoroseobacter shibae	**0.011**		0.489		**0.005**		**0.002**		0.299		**0.001**	
Dokdonella koreensis	**0.035**		0.767		**0.008**		**0.035**		0.472		**0.008**	
Draconibacterium orientale	0.15		0.15		0.15		**0.007**		0.238		**0.005**	
Dyadobacter fermentans	0.128		0.213		**0.026**		**0.022**		0.654		**0.001**	
Dyella japonica	**0.034**		0.354		0.397		**0.03**		**0.034**		0.064	
Dyella thiooxydans	**0.02**		0.236		0.067		**0.005**		0.07		**0.005**	
Echinicola strongylocentroti	0.114		0.418		0.114		**0.02**		0.418		**0.02**	
Echinicola vietnamensis	0.39		0.072		0.808		**0.045**		0.417		**0.045**	
***Ectothiorhodospira* sp. BSL 9**	**0.01**		0.727		**0.01**		**0.005**		0.082		**0.01**	
Ectromelia virus	0.068	NA		NA			0.087		**0.015**	NA		
Edwardsiella piscicida	**<0.001**	NA			0.268		**0.001**		**<0.001**		0.287	
Edwardsiella tarda	**0.005**		**0.021**		**0.012**		0.102		**0.045**		0.69	
Eggerthella lenta	0.605		**0.032**		0.377		0.081		0.312		0.081	
***Eggerthella* sp. YY7918**	**0.016**		0.747		0.076		**0.016**		**0.048**		0.056	
Ehrlichia canis	0.976		**0.008**		0.088		0.085		0.282		0.165	
Ehrlichia muris	0.57		0.061		0.576		0.57		0.57		**0.02**	
Eikenella corrodens	**<0.001**		0.575		**<0.001**		**<0.001**		**0.018**		**<0.001**	

Elizabethkingia anophelis	0.652	0.19	**0.02**	0.19	0.19	**0.001**
Emiliania huxleyi virus 86	0.122	0.161	0.648	**0.015**	0.122	0.122
Ensifer adhaerens	**0.003**	0.367	**0.032**	**0.005**	**0.013**	**0.022**
Ensifer sojae	**0.022**	0.29	**0.006**	**0.015**	0.522	**0.003**
Enterobacter hormaechei	0.068	0.718	**0.005**	0.086	0.781	0.052
Enterobacter kobei	NA	**0.039**	0.178	0.221	0.418	0.178
***Enterobacter* sp. 638**	**0.009**	**0.013**	0.18	0.835	0.117	0.117
***Enterobacter* sp. R4 368**	0.19	**0.012**	0.19	0.502	0.503	0.059
***Enterobacteriaceae* bacterium w17**	0.717	**0.028**	0.761	0.07	0.703	**0.026**
Enterococcus faecalis	**0.005**	0.797	**0.041**	**0.003**	**0.025**	**0.019**
Enterococcus gallinarum	**0.006**	0.226	**0.015**	0.086	0.188	0.262
Enterococcus phage EFDG1	NA	**0.039**	0.221	0.221	0.471	0.106
***Enterococcus* sp. FDAARGOS_375**	**0.006**	**0.039**	0.92	0.203	**0.006**	**0.017**
Enterococcus thailandicus	0.1	0.251	0.248	**0.02**	0.27	**0.02**
Erwinia billingiae	**0.016**	0.16	**<0.001**	0.203	0.949	0.112
Erwinia gerundensis	**0.004**	0.984	0.089	**0.004**	**0.037**	0.104
Erwinia phage PEp14	**0.026**	0.201	0.201	0.259	0.201	0.94
Erwinia phage vB_EamM_Caitlin	0.308	0.123	**0.014**	NA	NA	NA
Erwinia pyrifoliae	0.982	**0.044**	0.157	**0.044**	0.359	0.157
Erysipelothrix rhusiopathiae	0.127	0.148	0.413	**0.002**	**0.017**	0.247

Erythrobacter flavus	0.236	**0.016**	0.236	**0.003**	0.545	**<0.001**
Erythrobacter gangjinensis	**0.011**	0.267	0.123	**0.003**	0.121	**0.011**
Erythrobacter litoralis	**0.016**	0.215	0.215	**0.006**	**0.03**	**0.027**
Erythrobacter seohaensis	0.059	0.33	**0.016**	**0.002**	0.836	**0.001**
Escherichia fergusonii	**0.03**	0.647	0.604	**0.011**	**0.03**	0.273
Escherichia marmotae	**0.015**	0.575	**0.009**	**0.049**	0.174	0.071
Escherichia virus Cajan	**<0.001**	NA	**0.036**	**<0.001**	**<0.001**	0.06
Ethanoligenens harbinense	**0.001**	0.549	**0.038**	**0.001**	**0.011**	**0.011**
***Euzebya* sp. DY32 46**	**0.003**	0.389	0.064	**0.001**	**0.022**	**0.013**
Fermentimonas caenicola	0.442	**0.001**	0.429	0.102	0.268	**<0.001**
Fervidobacterium islandicum	**0.005**	0.974	0.477	**0.003**	**0.003**	0.477
***Fibrella* sp. ES10 3 2 2**	0.117	0.369	0.101	**0.019**	0.405	**0.006**
Fibrobacter succinogenes	0.229	0.229	0.294	**0.03**	0.442	**0.021**
Fimbriimonas ginsengisoli	**0.001**	0.262	0.262	**0.001**	**0.001**	0.845
***Fischerella* sp. NIES 3754**	0.958	0.22	**0.025**	0.398	0.22	0.497
Flavivirga eckloniae	0.085	0.301	0.085	**0.009**	0.309	**0.009**
Flavobacterium branchiophilum	0.319	0.184	0.069	0.067	0.69	**0.004**
Flavobacterium columnare	0.353	0.387	0.442	**0.029**	0.442	**0.046**
Flavobacterium crassostreae	0.534	**0.029**	**0.017**	0.057	**0.031**	0.534
Flavobacterium gilvum	**0.017**	0.063	0.649	0.26	**0.017**	0.063

***Flavobacterium* sp. HYN0048**	0.392	0.211	0.392	**0.021**	0.392	**0.021**
***Flavobacterium* sp. HYN0049**	0.937	0.395	0.245	0.317	0.453	**0.022**
***Flavobacterium* sp. HYN0056**	0.392	0.229	0.065	0.13	0.589	**0.002**
***Flavobacterium* sp. HYN0059**	0.54	0.111	0.54	0.111	0.82	**0.045**
Flavonifractor plautii	**0.006**	0.173	**0.001**	0.083	0.276	0.132
Fluviicola taffensis	0.116	0.262	0.135	0.384	**0.019**	**0.023**
***Formosa* sp. Hel3_A1_48**	0.931	0.188	0.282	0.282	0.704	**0.009**
Frankia alni	**0.019**	0.291	0.121	**0.012**	0.062	**0.022**
Frankia casuarinae	**0.035**	0.204	**0.035**	0.109	0.204	0.387
Frankia inefficax	**0.004**	0.502	**0.019**	**0.004**	0.066	**0.004**
***Frankia* sp. EAN1pec**	**0.018**	0.263	0.123	**0.011**	0.183	**0.018**
***Frondihabitans* sp. PAMC 28766**	**0.025**	**0.025**	0.295	**0.002**	**0.045**	**0.001**
***gamma proteobacterium* HdN1**	0.135	0.735	**0.006**	0.137	0.853	**0.041**
***Gammaproteobacteria* bacterium DM2**	**0.005**	0.549	0.515	**0.049**	**0.002**	0.336
***Gammaproteobacteria* bacterium ESL0073**	0.863	0.111	**0.005**	0.279	0.163	0.814
***Geitlerinema* sp. PCC 7407**	0.117	0.066	0.404	**0.002**	0.117	**0.002**
***Gemella* sp. oral taxon 928**	**0.024**	**0.047**	**0.047**	0.26	0.128	0.57
***Geminocystis* sp. NIES 3709**	**0.011**	0.757	0.541	**0.003**	**0.009**	0.321
Gemmata obscuriglobus	**<0.001**	0.469	**<0.001**	**0.002**	0.076	**0.002**
***Gemmata* sp. SH PL17**	**0.039**	0.162	0.073	**0.005**	0.323	**0.005**

Gemmatimonas aurantiaca	0.705	0.152	0.152	0.152	0.809	**0.002**
Gemmatimonas phototrophica	0.122	0.122	0.653	**0.003**	0.122	**0.031**
Gemmatirosa kalamazoonesis	**0.009**	0.59	**0.002**	**0.007**	0.506	**0.002**
Geoalkalibacter subterraneus	0.055	0.511	**0.008**	0.055	0.825	**0.016**
Geobacillus thermodenitrificans	**0.002**	**0.004**	**0.002**	0.173	0.209	0.885
Geobacillus thermoleovorans	0.613	0.243	**0.001**	0.83	0.315	0.2
Geobacter anodireducens	**0.005**	0.757	**0.016**	**0.005**	0.053	**0.013**
Geobacter bemidjiensis	**0.006**	0.927	**0.001**	**0.016**	0.259	**0.011**
Geobacter daltonii	**0.002**	**0.021**	**0.002**	**0.021**	**0.049**	0.493
Geobacter metallireducens	**0.02**	0.19	**0.021**	**0.005**	0.221	**0.002**
Geobacter pickeringii	**0.008**	0.396	0.091	**0.002**	**0.02**	**0.011**
***Geobacter* sp. M18**	**0.003**	0.844	0.055	**0.003**	**0.036**	0.062
***Geobacter* sp. M21**	**0.006**	0.131	**0.03**	**0.004**	0.131	**0.002**
Geobacter sulfurreducens	**0.012**	0.309	**0.007**	0.062	0.326	0.118
Geobacter uraniireducens	**0.046**	0.732	**0.046**	**0.046**	0.526	**0.046**
Geodermatophilus obscurus	**0.003**	0.437	**0.002**	**0.043**	0.114	0.093
Gibbsiella quercinecans	**0.025**	**0.018**	**<0.001**	0.484	0.972	0.422
Gloeobacter kilaueensis	**0.034**	**0.034**	**0.034**	0.59	0.405	0.59
Gloeobacter violaceus	**0.002**	0.378	**0.045**	**0.002**	**0.033**	**0.025**
***Gloeocapsa* sp. PCC 7428**	**0.003**	0.13	0.109	0.082	**0.041**	0.932

Gluconacetobacter diazotrophicus	**0.001**	0.605	**0.021**	**0.003**	**0.022**	**0.021**
Gluconobacter albidus	0.064	0.366	0.697	**0.042**	0.064	0.221
Gluconobacter oxydans	**0.021**	0.311	0.062	**0.002**	0.124	**0.005**
Gordonia bronchialis	0.259	0.217	0.294	**0.035**	0.696	**0.035**
Gordonia iterans	**0.014**	0.542	**0.014**	0.067	0.151	0.15
Gordonia polyisoprenivorans	**0.002**	0.747	**0.018**	**0.001**	**0.041**	**0.033**
***Gordonia* sp. KTR9**	**0.001**	0.412	0.053	**0.004**	**0.004**	0.46
***Gordonia* sp. YC JH1**	**0.006**	0.076	0.071	**<0.001**	**0.042**	**<0.001**
Gordonia terrae	**0.011**	0.219	0.109	**0.002**	0.071	**0.01**
Gordonibacter massiliensis	0.054	0.937	0.591	**0.038**	0.054	0.472
Granulicella mallensis	**0.04**	0.475	0.133	**0.014**	0.115	0.051
Granulicella tundricola	**0.021**	0.576	**0.012**	**0.012**	0.449	**0.011**
Gryllus bimaculatus nudivirus	0.068	NA	NA	0.087	**0.015**	NA
Hafnia paralvei	0.128	**0.018**	**0.024**	0.693	0.943	0.461
Haliangium ochraceum	**0.008**	0.299	**0.045**	**0.008**	0.133	**0.008**
Halioglobus pacificus	0.07	0.327	0.327	**0.009**	0.133	0.054
Halobacteriovorax marinus	0.485	0.408	0.408	0.274	0.696	**0.025**
Halomicronema hongdechloris	0.104	0.319	0.779	**0.012**	0.072	0.104
Halomonas aestuarii	**0.023**	0.963	**0.023**	**0.023**	0.339	**0.028**
Halomonas beimenensis	**0.001**	0.269	**0.009**	**<0.001**	**0.016**	**0.001**

Halomonas chromatireducens	0.066	0.245	0.074	**0.003**	0.292	**0.003**
Halomonas elongata	**0.029**	0.165	0.11	**0.014**	0.11	**0.014**
Halomonas hydrothermalis	0.064	**0.008**	**0.018**	0.532	0.983	0.214
***Halomonas* sp. 1513**	**0.006**	0.622	**0.033**	**0.019**	**0.045**	0.17
***Halomonas* sp. GFAJ 1**	0.77	**0.008**	0.77	0.211	0.77	**0.008**
***Halomonas* sp. SF2003**	0.464	0.084	0.464	0.073	0.464	**0.007**
Halorhodospira halochloris	**0.019**	**0.009**	0.243	0.807	**0.041**	**0.015**
Halotalea alkalilenta	**0.049**	0.89	**0.049**	**0.049**	0.572	**0.049**
***Halothece* sp. PCC 7418**	**0.047**	0.63	0.688	0.216	**0.018**	0.383
Halothermothrix orenii	0.741	**0.001**	0.051	0.051	0.463	0.051
***Halothiobacillus* sp. LS2**	**0.001**	0.833	**0.002**	**0.001**	**0.012**	**0.001**
Halovirus HCTV 2	0.269	0.269	0.269	**0.044**	**0.013**	0.541
Hartmannibacter diazotrophicus	**0.023**	0.465	**0.023**	0.107	0.306	0.139
Helicobacter mustelae	0.925	0.449	**0.01**	0.778	0.449	0.267
Helicobacter pylori	**0.04**	0.828	0.348	0.348	0.348	0.724
Heliobacterium modesticaldum	**0.034**	0.921	0.155	**0.008**	0.067	0.12
Herbaspirillum seropedicae	**0.004**	0.585	**<0.001**	**0.006**	0.128	**0.006**
***Herbaspirillum* sp. meg3**	**0.004**	0.859	**0.004**	**0.034**	0.215	0.1
Herminiimonas arsenicoxydans	0.688	0.179	0.236	0.179	0.762	**0.011**
Hippea maritima	**0.006**	0.437	0.054	**0.03**	0.066	0.324

***Hoeflea* sp. IMCC20628**	**0.003**	0.906	**0.003**	**0.003**	0.112	**0.003**
Hydrogenobacter thermophilus	0.451	0.536	**0.005**	0.576	0.451	0.06
***Hydrogenobaculum* sp. Y04AAS1**	0.132	**0.01**	**0.016**	0.822	0.822	0.516
Hydrogenophaga crassostreae	**0.027**	0.813	**0.031**	**0.027**	0.226	**0.027**
***Hydrogenophaga* sp. PBC**	**0.021**	0.459	**0.003**	**0.003**	0.459	**<0.001**
Hymenobacter nivis	**<0.001**	0.103	**0.003**	**<0.001**	**0.018**	**<0.001**
Hymenobacter sedentarius	0.212	0.083	**0.043**	**0.006**	1	**<0.001**
***Hymenobacter* sp. APR13**	**0.03**	0.186	**0.01**	**0.001**	0.114	**<0.001**
***Hymenobacter* sp. DG25A**	**0.019**	**0.043**	0.062	**0.002**	0.092	**<0.001**
***Hymenobacter* sp. DG25B**	0.121	0.101	**0.004**	**0.006**	0.744	**<0.001**
***Hymenobacter* sp. PAMC 26554**	**0.025**	0.061	0.052	**0.002**	0.188	**0.001**
***Hymenobacter* sp. PAMC 26628**	0.135	0.53	**0.027**	**0.045**	0.749	**0.008**
Hymenobacter swuensis	0.077	0.196	**0.03**	**0.014**	0.569	**0.001**
Hyphomicrobium denitrificans	**0.018**	0.89	**0.044**	**0.018**	0.139	0.081
Hyphomicrobium nitrativorans	**0.007**	0.085	**0.015**	**0.001**	0.085	**<0.001**
Hyphomonas neptunium	**0.025**	0.229	0.813	**0.002**	**0.025**	0.125
Ichthyobacterium seriolicida	0.484	0.28	**0.005**	0.839	0.37	0.28
Idiomarina loihiensis	**0.001**	**0.031**	**0.031**	**0.04**	**0.031**	0.719
Ignavibacterium album	**0.018**	0.49	0.118	**0.01**	0.091	**0.028**
Ilyobacter polytropus	0.076	**0.024**	0.737	**0.006**	0.073	**0.024**

Intestinimonas butyriciproducens	**0.02**	0.624	**0.003**	**0.004**	0.624	**0.001**
Intrasporangium calvum	**0.002**	0.961	**<0.001**	**0.001**	0.229	**<0.001**
Isoptericola variabilis	**0.017**	0.833	**0.017**	**0.023**	0.253	0.06
Isosphaera pallida	**0.023**	0.755	0.279	**0.023**	0.074	0.405
Janibacter indicus	**0.006**	0.086	**0.027**	**0.001**	0.086	**0.001**
Janthinobacterium agaricidamnosum	**0.002**	0.854	0.141	**0.004**	**0.011**	0.141
***Janthinobacterium* sp. 1_2014MBL_MicDiv**	**0.016**	0.473	**0.008**	0.077	0.379	0.077
***Janthinobacterium* sp. B9 8**	0.697	**0.041**	**0.041**	0.697	0.779	0.821
***Janthinobacterium* sp. LM6**	0.151	0.151	0.774	**0.022**	0.151	0.151
Janthinobacterium svalbardensis	0.11	0.511	**0.019**	**0.014**	0.825	**0.004**
Jonesia denitrificans	**0.003**	0.067	**0.003**	0.107	0.136	0.413
Kallithea virus	0.068	NA	NA	0.087	**0.015**	NA
Kangiella profundi	**0.037**	0.512	**0.046**	**0.046**	0.231	0.158
Kangiella sediminilitoris	**0.049**	0.526	**0.049**	**0.049**	0.526	0.068
Ketogulonicigenium robustum	**<0.001**	0.221	**0.022**	**<0.001**	**0.003**	0.221
Ketogulonicigenium vulgare	**0.013**	0.374	**0.009**	**0.036**	0.169	0.066
Kibdelosporangium phytohabitans	**<0.001**	0.126	**<0.001**	**0.026**	0.126	0.054
Kineococcus radiotolerans	**0.015**	0.675	**0.005**	**0.017**	0.675	**0.015**
Kiritimatiella glycovorans	**0.01**	0.118	0.13	**0.001**	**0.015**	**0.004**
Kitasatospora albolonga	0.111	0.328	**0.008**	0.328	0.87	0.241

Kitasatospora aureofaciens	**0.026**	0.979	**0.026**	0.067	0.22	0.115
Kitasatospora setae	**0.028**	0.992	**0.039**	**0.028**	0.149	**0.039**
Klebsiella phage vB_KpnM_KB57	0.068 NA	NA		0.087	**0.015** NA	
Kocuria palustris	0.259	0.497	**0.046**	0.115	0.798	**0.026**
***Kocuria* sp. BT304**	**<0.001**	0.739	**0.042**	**<0.001**	**<0.001**	**0.042**
Komagataeibacter europaeus	**0.04**	0.722	0.129	**0.04**	0.129	0.307
Komagataeibacter medellinensis	0.49	**0.002**	0.237	0.61	0.749	0.055
Komagataeibacter xylinus	**0.001**	0.247	**0.016**	**0.001**	**0.039**	**0.001**
Kosakonia cowanii	**0.001**	0.554	0.445	**0.001**	**0.001**	0.244
Kosakonia oryzae	**0.049**	0.759	**0.004**	**0.013**	0.759	**0.001**
Kribbella flavida	**0.031**	0.958	**0.025**	**0.025**	0.355	**0.025**
Kutzneria albida	**0.003**	0.415	**0.001**	**0.006**	0.238	**0.001**
***Kyrpidia* sp. EA 1**	**0.029**	0.595	**0.029**	0.097	0.588	0.097
Kytococcus sedentarius	**0.002**	0.171	**0.001**	**0.006**	0.074	0.06
Labrenzia aggregata	**0.03**	0.681	**0.038**	**0.03**	0.152	**0.03**
***Laceyella* sp. FBKL4 010**	**0.021**	0.066	**0.025**	0.252	0.066	0.735
Lachnoclostridium phocaeense	0.711	0.12	0.376	0.376	0.404	**0.007**
***Lacinutrix* sp. Bg11 31**	0.392	**0.006**	**0.019**	0.173	0.392	0.268
Lactobacillus agilis	**0.041**	0.433	0.131	0.297	0.153	0.66
Lactobacillus amylolyticus	**0.042**	0.846	0.234	**0.042**	0.137	0.215

Lactobacillus buchneri	0.398	**0.017**	**0.041**	0.398	0.458	0.458
Lactobacillus delbrueckii	**0.001**	**0.012**	0.176	**0.008**	**0.001**	0.112
Lactobacillus farciminis	0.07	0.07	0.453	**0.007**	0.07	**0.019**
Lactobacillus gallinarum	0.775	**0.018**	0.311	0.31	0.826	0.084
Lactobacillus heilongjiangensis	0.822	**0.015**	0.822	0.094	0.822	**0.005**
Lactobacillus kefiranofaciens	0.097	0.407	0.407	0.208	**0.048**	0.097
Lactobacillus koreensis	0.218	**0.036**	0.218	0.558	0.675	0.218
Lactobacillus phage JCL1032	**0.022**	NA	0.268	**0.036**	0.069	0.287
Lactobacillus plantarum	0.07	0.173	0.975	**0.012**	0.07	0.158
Lactobacillus sakei	0.073	0.48	0.277	**0.04**	**0.008**	0.949
Lactobacillus sanfranciscensis	0.111	0.111	**0.018**	0.945	0.989	0.945
Lactobacillus zymae	**0.007**	0.708	0.089	**0.007**	0.089	0.089
Lactococcus lactis	**0.035**	0.539	0.52	0.183	**0.035**	0.868
Lactococcus phage P335 sensu lato	**0.021**	0.873	0.866	**0.021**	**0.002**	0.873
Lactococcus phage ul36	0.068	NA	NA	0.087	**0.015**	NA
Lacunisphaera limnophila	**0.009**	0.629	**0.017**	**0.009**	0.179	**0.017**
Laribacter hongkongensis	**0.043**	0.688	**0.007**	0.074	0.516	0.056
Lawsonella clevelandensis	**0.029**	0.242	**0.011**	**0.007**	0.385	**0.001**
Lawsonia intracellularis	**0.011**	**0.048**	**0.011**	0.234	0.234	0.846
***Leclercia* sp. LSNIH3**	**0.041**	0.087	**0.041**	0.257	0.257	0.827

Legionella clemsonensis	0.633	**0.047**	0.633	0.054	0.678	**0.012**
Legionella hackeliae	**0.04**	0.434	**0.04**	0.082	0.332	0.18
Leifsonia xyli	**0.013**	0.469	**0.013**	**0.003**	0.191	**0.003**
***Lelliottia* sp. WB101**	0.058	0.061	0.52	0.52	**0.02**	**0.016**
Leminorella richardii	0.124	0.124	**0.033**	0.727	0.727	0.727
Lentzea guizhouensis	**0.007**	0.193	**0.007**	**0.001**	0.128	**<0.001**
***Leptolyngbya* sp. NIES 3755**	**0.026**	0.469	0.074	**0.049**	0.167	0.345
***Leptolyngbya* sp. O 77**	0.062	0.792	0.124	**0.046**	0.176	0.062
Leptospira borgpetersenii	**0.032**	0.699	0.747	0.163	**0.032**	0.699
Leptospira santarosai	0.603	**0.047**	0.562	0.562	1	0.113
Leptospirillum ferriphilum	0.055	**<0.001**	**0.027**	0.227	0.976	**0.027**
***Leptotrichia* sp. oral taxon 498**	**0.031**	0.747	0.609	**0.047**	**0.047**	0.747
Limnochorda pilosa	**0.022**	0.622	**0.043**	**0.005**	0.22	**0.011**
***Limnohabitans* sp. 63ED37 2**	**0.021**	0.481	0.131	**0.034**	**0.034**	0.461
Listeria phage LP 030 3	**0.002**	NA	0.134	**0.006**	**0.032**	0.167
Luteibacter rhizovicinus	**0.009**	0.921	**0.043**	**0.009**	0.093	**0.044**
***Luteimonas* sp. 100111**	0.186	0.358	**0.033**	0.066	0.743	**0.004**
***Luteimonas* sp. 83 4**	**0.039**	0.379	**0.005**	**0.016**	0.579	**0.002**
Luteitalea pratensis	**0.001**	1	**0.006**	**0.001**	**0.007**	**0.016**
Lutibacter profundi	0.076	0.075	0.983	**0.002**	0.076	0.075

***Lysinibacillus* sp. 2017**	**0.031**	0.466	0.678	0.107	**0.013**	0.325
Lysobacter antibioticus	**0.002**	0.49	**0.012**	**0.002**	0.057	**0.008**
Lysobacter capsici	**0.003**	0.99	**0.003**	**0.018**	0.171	**0.025**
Lysobacter enzymogenes	**0.014**	0.587	**0.018**	**0.014**	0.076	0.069
Lysobacter gummosus	**0.016**	0.306	**0.006**	**0.006**	0.522	**0.001**
Lysobacter maris	**0.003**	0.713	**0.007**	**0.003**	0.057	**0.007**
***Magnetospira* sp. QH 2**	**0.001**	0.445	0.088	**0.001**	**0.007**	**0.032**
Magnetospirillum gryphiswaldense	**0.022**	0.563	0.058	**0.022**	0.102	**0.033**
Magnetospirillum magneticum	**<0.001**	0.849	**<0.001**	**<0.001**	0.109	**<0.001**
***Magnetospirillum* sp. ME 1**	**0.003**	0.062	**0.002**	0.09	0.296	0.379
***Magnetospirillum* sp. XM 1**	**0.003**	0.732	**0.01**	**0.003**	**0.031**	**0.031**
Maricaulis maris	**0.001**	0.494	**0.039**	**0.001**	0.131	**0.039**
Marichromatium purpuratum	**0.023**	0.347	**0.023**	**0.023**	0.269	**0.014**
Marinithermus hydrothermalis	**0.048**	0.6	0.101	**0.04**	0.181	**0.04**
***Marinitoga* sp. 1137**	0.752	**0.042**	0.997	**0.043**	0.752	**0.029**
Marinobacter salinus	**0.009**	0.167	0.42	**0.001**	**0.015**	**0.022**
Marinobacter similis	0.336	0.558	**0.03**	0.558	0.749	0.52
***Marinobacter* sp. CP1**	0.656	0.229	**0.016**	0.834	0.356	0.229
***Marinobacter* sp. LQ44**	0.246	0.28	0.246	**0.038**	0.486	**0.034**
Marinomonas mediterranea	**0.014**	0.106	**0.007**	0.07	0.264	0.266

***Marinomonas* sp. MWYL1**	0.31	0.406	0.108	0.131	0.892	**0.033**
Mariprofundus aestuarium	**<0.001**	0.313	**0.034**	**<0.001**	**<0.001**	0.283
Mariprofundus ferrinatatus	**0.014**	0.932	**0.018**	**0.014**	0.141	**0.023**
Marivirga tractuosa	0.253	0.573	**0.017**	0.125	0.749	**0.007**
Martelella mediterranea	**0.005**	0.836	**0.01**	**0.005**	0.066	**0.01**
***Martelella* sp. AD 3**	**0.044**	0.742	0.1	**0.044**	0.143	0.111
Massilia putida	**0.015**	0.089	0.372	**0.007**	**0.01**	**0.01**
***Massilia* sp. NR 4 1**	0.117	0.667	**0.024**	0.111	0.926	0.072
***Massilia* sp. WG5**	**0.013**	0.385	**0.013**	**0.006**	0.241	**0.006**
***Massilia* sp. YMA4**	0.064	0.164	**0.001**	**0.006**	0.977	**<0.001**
Massilia timonae	**0.008**	0.89	0.106	**0.008**	**0.03**	0.242
Massilia violaceinigra	**0.013**	0.306	0.598	**0.003**	**0.011**	0.11
Megasphaera elsdenii	**<0.001**	0.221	**0.041**	**<0.001**	**<0.001**	0.438
Megasphaera hexanoica	**0.025**	0.101	0.408	0.19	**0.012**	0.101
Meiothermus ruber	**0.029**	0.381	0.105	0.105	0.105	0.626
Meiothermus silvanus	**0.017**	0.732	0.443	**0.015**	**0.036**	0.301
***Melaminivora* sp. SC2 7**	0.12	0.348	**0.001**	0.059	0.468	**0.001**
***Melaminivora* sp. SC2 9**	**<0.001**	0.767	**0.004**	**<0.001**	**0.008**	**0.002**
Melioribacter roseus	**0.003**	0.298	**0.005**	**<0.001**	0.099	**<0.001**
Melittangium boletus	**0.008**	0.974	**0.009**	**0.008**	0.112	**0.013**

Mesoplasma coleopterae	**0.002**	0.666	0.067	**0.002**	**0.031**	0.053
Mesoplasma syrphidae	**0.001**	0.947	0.095	**0.001**	**0.01**	0.095
Mesorhizobium amorphae	**0.005**	0.165	**0.002**	0.054	0.165	0.165
Mesorhizobium australicum	**0.006**	0.735	0.622	**0.003**	**0.002**	0.772
Mesorhizobium ciceri	0.071	0.547	**0.047**	**0.047**	0.547	**0.024**
Mesorhizobium japonicum	**0.01**	0.396	0.687	**0.01**	**0.01**	0.235
Mesorhizobium oceanicum	**0.006**	0.787	0.055	**0.004**	**0.012**	**0.037**
Mesorhizobium opportunistum	**0.003**	**0.04**	0.064	0.122	**0.04**	0.668
Mesotoga prima	**0.032**	**0.032**	0.194	0.444	0.075	0.075
Methylibium petroleiphilum	**0.015**	0.434	**0.015**	**0.015**	0.093	0.05
Methylobacillus flagellatus	0.265	0.344	0.192	0.081	0.634	**0.018**
Methylobacterium nodulans	**0.001**	0.603	**0.001**	**0.015**	0.169	**0.023**
Methylobacterium oryzae	NA	0.304	**0.02**	0.559	0.242	0.115
Methylobacterium radiotolerans	**<0.001**	0.784	**0.001**	**<0.001**	**<0.001**	**0.004**
***Methylobacterium* sp. 17SD2 17**	0.05	0.971	**0.025**	**0.025**	0.458	**0.025**
***Methylobacterium* sp. 17Sr1 1**	0.176	0.453	0.532	**0.037**	0.176	0.176
***Methylobacterium* sp. 17Sr1 28**	0.916	**0.001**	0.916	**0.019**	0.916	**0.003**
***Methylobacterium* sp. 17Sr1 43**	**0.013**	0.747	**0.017**	**0.013**	0.16	**0.017**
***Methylobacterium* sp. 4 46**	0.223	0.879	**0.026**	0.223	0.892	0.067
***Methylobacterium* sp. AMS5**	0.52	0.406	0.335	0.106	0.83	**0.007**

Methyloceanibacter caenitepidi	**0.012**	0.921	**0.021**	**0.024**	0.316	**0.043**
***Methyloceanibacter* sp. wino2**	**0.001**	0.078	**<0.001**	**0.043**	0.216	0.09
Methylococcus capsulatus	0.149	0.055	0.064	**0.011**	0.983	**0.001**
Methylocystis bryophila	**0.003**	0.25	**0.003**	**0.003**	**0.046**	**0.003**
***Methylocystis* sp. SC2**	**0.011**	0.48	**0.012**	**0.002**	0.128	**0.002**
Methylomonas methanica	0.168	0.693	0.08	0.098	0.693	**0.033**
Methylophaga nitratireducenticrescens	**0.021**	**0.021**	**0.021**	0.465	0.217	0.428
***Methylophilus* sp. TWE2**	0.472	**0.026**	**0.038**	0.472	0.727	0.472
Methylorubrum extorquens	**0.003**	0.683	**0.026**	**0.003**	**0.022**	**0.011**
Methylorubrum populi	**0.012**	0.473	**0.004**	**0.012**	0.292	**0.004**
Methylosinus trichosporium	**0.011**	0.226	**0.048**	**0.048**	**0.048**	0.465
Methylotenera versatilis	**0.001**	0.185	0.057	**0.012**	**0.005**	0.892
***Methyloversatilis* sp. RAC08**	0.131	0.282	0.149	**0.023**	0.477	**0.023**
Methylovorus glucosotrophus	0.127	0.551	**0.016**	0.31	0.933	0.172
***Methylovorus* sp. MP688**	0.151	0.151	0.707	**0.005**	**0.047**	0.165
Methylovulum psychrotolerans	0.103	**0.007**	**0.004**	0.807	0.766	0.766
Microbacterium chocolatum	0.06	**0.03**	**0.03**	0.931	0.911	0.862
Microbacterium hominis	**0.014**	0.69	**0.008**	**0.008**	0.4	**0.007**
Microbacterium paludicola	**0.001**	0.519	**0.001**	**0.01**	0.128	**0.047**
***Microbacterium* sp. BH 3 3 3**	0.159	0.118	0.261	**0.008**	0.261	**0.006**

***Microbacterium* sp. No 7**	**0.005**	0.698	**0.006**	**0.001**	0.112	**0.002**
***Microbacterium* sp. PAMC 28756**	**0.014**	0.203	**0.014**	**0.017**	0.134	0.228
***Microbacterium* sp. TPU 3598**	**0.034**	**0.048**	0.292	**0.005**	0.089	**0.009**
***Microbacterium* sp. XT11**	**0.027**	0.469	0.243	**0.027**	0.192	0.192
Microbulbifer aggregans	**0.047**	0.963	0.301	0.093	0.153	0.373
Microbulbifer thermotolerans	0.207	0.664	0.247	**0.034**	0.247	**0.048**
Micrococcus luteus	**0.003**	0.558	**0.013**	**0.001**	**0.049**	**0.004**
***Microcystis* sp. MC19**	0.308	**0.045**	**0.001**	0.528	0.308	0.367
Microlunatus phosphovorus	**0.014**	0.191	**0.014**	0.087	0.191	0.191
***Micromonospora* sp. L5**	**0.031**	0.453	0.162	**0.031**	0.077	0.453
Micropruina glycogenica	0.089	0.078	0.078	**0.002**	0.504	**<0.001**
Microterricola viridarii	**0.004**	0.422	**<0.001**	**0.002**	0.355	**<0.001**
Microvirga ossetica	0.054	0.416	**0.003**	0.188	0.983	0.054
***Microvirga* sp. 17 mud 1 3**	0.075	0.097	0.097	**0.008**	0.255	**0.004**
Mimivirus terra2	**0.047**	**0.004**	**0.046**	0.684	0.684	0.166
***Miniimonas* sp. S16**	**0.017**	0.698	0.063	**0.017**	0.147	0.063
***Mitsuaria* sp. 7**	**0.001**	0.374	**0.001**	**0.001**	0.086	**<0.001**
Modestobacter marinus	**0.007**	0.875	**0.009**	**0.016**	0.117	**0.024**
Moorea producens	0.62	0.187	0.603	0.148	0.814	**0.012**
Moorella thermoacetica	**0.011**	**0.02**	**0.012**	0.209	0.16	0.781

Moraxella catarrhalis	**0.022**	0.58	**0.005**	**0.034**	0.521	**0.022**
***Mordavella* sp. Marseille P3756**	0.35	0.063	0.77	0.623	0.21	**0.007**
Moritella viscosa	0.671	0.322	**0.026**	0.558	0.322	0.448
Mucilaginibacter gotjawali	**0.021**	0.175	**0.038**	**0.005**	0.2	**0.003**
***Mucilaginibacter* sp. BJC16 A31**	0.102	**0.033**	0.102	**0.004**	0.634	**<0.001**
Mucinivorans hirudinis	0.134	**0.046**	0.09	**0.004**	0.629	**<0.001**
Muricauda lutaonensis	**0.025**	0.496	**0.047**	**0.047**	0.113	0.294
***Mycetocola* sp. CGMCC 1 16372**	**0.038**	**0.038**	0.54	**0.001**	**0.045**	**0.012**
Mycobacterium haemophilum	**0.008**	0.929	0.929	**0.001**	**0.001**	0.929
Mycobacterium intracellulare	**0.002**	0.462	0.269	**0.002**	**0.002**	0.104
Mycobacterium marseillense	**0.001**	0.615	0.074	**0.007**	**0.007**	0.284
Mycobacterium phage Jobu08	**0.007**	NA	NA	**0.015**	**<0.001**	NA
Mycobacterium shigaense	**0.007**	0.59	**0.01**	**0.01**	0.06	**0.033**
***Mycobacterium* sp. EPa45**	**0.006**	0.401	**0.006**	**0.006**	0.292	**0.035**
***Mycobacterium* sp. JS623**	**0.048**	**0.038**	**0.048**	**<0.001**	0.292	**<0.001**
***Mycobacterium* sp. VKM Ac 1817D**	**0.002**	0.229	**0.001**	**0.036**	0.331	**0.036**
***Mycobacterium* sp. WY10**	0.357	0.328	**0.021**	0.93	0.545	0.328
Mycobacterium virus Bron	**0.027**	**0.028**	**0.011**	0.437	0.437	0.932
Mycobacterium virus Nigel	0.068	NA	NA	0.087	**0.015**	NA
Mycobacterium virus Rosebush	**0.013**	0.392	0.313	**0.013**	0.06	0.193

Mycobacteroides abscessus	0.223	0.718	**0.025**	0.125	0.798	**0.004**
Mycobacteroides chelonae	**0.008**	0.574	**0.019**	**0.019**	0.183	0.088
Mycobacteroides saopaulense	**0.021**	0.471	0.058	**0.019**	0.06	**0.019**
Mycolicibacter terrae	0.168	0.246	0.224	0.08	0.317	**0.017**
Mycolicibacterium aurum	**0.043**	0.248	0.348	**0.013**	0.063	**0.043**
Mycolicibacterium chubuense	**0.007**	0.451	**0.029**	**0.002**	0.115	**0.007**
Mycolicibacterium gilvum	0.175	0.175	0.175	**0.006**	0.504	**0.005**
Mycolicibacterium litorale	**0.01**	0.903	**0.008**	**0.01**	0.666	**0.01**
Mycolicibacterium phlei	**0.014**	1	**0.001**	**0.034**	0.521	**0.014**
Mycolicibacterium smegmatis	0.097	0.977	**0.016**	0.127	0.977	**0.024**
Mycolicibacterium thermoresistibile	**0.002**	0.742	0.117	**0.002**	**0.016**	0.068
Mycolicibacterium vaccae	**<0.001**	0.735	**<0.001**	**0.001**	0.262	**<0.001**
Mycolicibacterium vanbaalenii	**0.001**	0.286	**<0.001**	**0.01**	0.262	**0.001**
Mycoplasma anseris	0.067	**0.003**	**0.001**	1	1	1
Mycoplasma bovigenitalium	**<0.001**	0.203	**0.004**	**<0.001**	**<0.001**	0.135
Mycoplasma bovis	**0.049**	0.748	0.38	**0.049**	**0.049**	0.461
Mycoplasma canadense	**0.049**	0.953	0.953	**0.049**	**0.049**	0.953
Mycoplasma cynos	0.06	**0.033**	**0.033**	0.613	0.515	0.613
Mycoplasma dispar	**0.044**	0.299	0.262	**0.025**	0.081	**0.044**
Mycoplasma flocculare	**0.009**	0.486	0.156	**0.004**	**0.047**	**0.047**

Mycoplasma hominis	**0.001**	0.455	0.251	**0.001**	**0.002**	0.168
Mycoplasma mobile	**0.047**	0.373	**0.047**	**0.047**	0.187	0.216
Mycoplasma parvum	0.171	0.152	0.807	**0.018**	0.171	0.133
Mycoplasma phocidae	0.062	0.448	**0.022**	**0.014**	0.448	**0.004**
Mycoplasma pneumoniae	0.137	0.067	0.671	**0.015**	0.067	0.067
***Mycoplasma* sp. ex Biomphalaria glabrata**	0.473	0.284	**0.006**	0.717	0.284	0.284
Mycoplasma suis	**0.002**	0.435	0.068	**0.018**	0.065	0.307
Mycoplasma synoviae	0.401	**0.038**	0.764	**0.038**	0.349	**0.038**
Mycoplasma wenyonii	**0.011**	0.061	0.211	**<0.001**	0.061	**0.016**
Myroides profundi	0.969	**0.04**	0.969	0.09	0.969	**0.013**
***Myroides* sp. A21**	0.06	0.051	0.936	**0.008**	0.051	**0.008**
Myxococcus fulvus	**0.022**	0.135	0.135	**0.001**	0.076	**0.002**
Myxococcus hansupus	**0.023**	0.88	0.119	**0.023**	0.119	0.119
Myxococcus macrosporus	0.395	0.143	**<0.001**	0.781	0.142	0.06
Myxococcus stipitatus	**0.009**	0.083	**0.032**	**0.003**	0.11	**0.001**
Myxococcus xanthus	**0.008**	0.236	**0.033**	**0.005**	0.147	**0.005**
Nakamurella multipartita	**0.007**	0.974	**0.007**	**0.015**	0.145	**0.015**
Nautilia profundicola	**0.02**	**0.001**	0.602	0.602	0.053	**0.001**
Negativicoccus massiliensis	**0.025**	**0.001**	0.113	0.566	0.085	**0.003**
Neisseria cinerea	**0.023**	0.491	**0.023**	**0.048**	0.174	0.174

Neisseria lactamica	**0.018**	0.058	0.179	0.206	**0.043**	0.206
Neisseria meningitidis	0.458	**0.024**	**0.023**	**0.023**	0.449	**<0.001**
***Neisseria* sp. KEM232**	**0.007**	0.095	**0.019**	**0.007**	**0.016**	0.601
Neisseria weaveri	**0.004**	0.294	**0.007**	**0.016**	**0.034**	0.083
Neoasaia chiangmaiensis	0.143	0.88	**0.049**	0.13	0.88	**0.049**
Neomicrococcus aestuarii	**<0.001**	0.37	**0.014**	**<0.001**	**0.005**	**0.001**
Neorickettsia risticii	0.137	**0.047**	0.16	0.635	0.635	0.16
Nitratifractor salsuginis	**0.01**	0.572	**0.01**	**0.004**	0.303	**0.004**
Nitratireductor basaltis	**0.001**	0.18	**0.003**	**0.041**	**0.041**	0.311
***Nitratireductor* sp. OM 1**	**0.003**	0.086	**0.032**	**<0.001**	**0.043**	**<0.001**
Nitrobacter hamburgensis	**0.001**	0.122	**0.005**	**0.009**	**0.028**	0.231
Nitrobacter winogradskyi	**0.02**	0.984	**0.007**	**0.029**	0.483	**0.017**
Nitrosospira lacus	0.069	0.291	**0.006**	0.223	0.842	0.139
Nitrosospira multiformis	0.499	0.88	**0.029**	0.486	0.766	**0.029**
Nitrospira defluvii	**0.048**	0.548	**0.034**	**0.048**	0.87	**0.011**
Nitrospira moscoviensis	**0.048**	0.792	0.051	**0.048**	0.151	0.051
Nitrospirillum amazonense	**0.004**	0.603	0.06	**0.004**	**0.019**	**0.038**
Niveispirillum cyanobacteriorum	**0.017**	0.106	**0.036**	0.055	0.055	0.567
Nocardia brasiliensis	**0.016**	0.502	**0.002**	**0.006**	0.502	**0.002**
Nocardia cyriacigeorgica	**0.002**	0.239	**0.032**	**0.002**	**0.032**	**0.002**

Nocardia farcinica	**0.034**	0.213	**0.001**	0.309	0.803	0.136
Nocardia seriolae	**0.002**	0.64	**0.004**	**0.002**	**0.037**	**0.019**
Nocardia soli	**0.005**	0.064	**0.042**	**0.003**	**0.042**	**0.003**
Nocardia terpenica	**0.026**	0.787	**0.016**	**0.026**	0.422	**0.026**
Nocardioides dokdonensis	0.15	0.376	**<0.001**	**0.032**	0.378	**<0.001**
***Nocardioides* sp. 78**	**0.02**	0.254	**0.006**	**0.001**	0.359	**<0.001**
***Nocardioides* sp. JS614**	**0.038**	0.231	0.051	**0.013**	0.231	**0.013**
Nocardiopsis alba	**0.014**	0.969	**<0.001**	**0.003**	0.879	**<0.001**
Nocardiopsis dassonvillei	**0.004**	0.742	**0.004**	**0.004**	0.13	**0.005**
Nocardiopsis gilva	**0.019**	0.309	**0.024**	**0.014**	0.312	**0.014**
Nonlabens dokdonensis	**0.045**	0.244	**0.022**	**0.008**	0.459	**0.002**
***Nonlabens* sp. MB 3u 79**	0.507	0.507	0.077	0.944	0.077	**0.049**
***Nonomuraea* sp. ATCC 55076**	**0.031**	0.18	**0.033**	**0.01**	0.312	**0.009**
***Nostoc* sp. Peltigera membranacea cyanobiont N6**	**0.004**	**0.024**	**<0.001**	0.058	0.405	0.073
***Nostoc* sp. NIES 3756**	**0.003**	**<0.001**	**0.003**	0.566	0.115	0.115
***Nostoc* sp. PCC 7120**	**0.012**	0.06	0.06	0.077	0.079	0.833
Novibacillus thermophilus	**0.016**	0.239	0.051	**0.033**	0.11	0.341
Novosphingobium aromaticivorans	**0.012**	0.109	0.109	**0.001**	**0.023**	**0.001**
Novosphingobium pentaromativorans	0.526	0.099	**0.007**	0.601	0.526	0.526

Novosphingobium resinovorum	0.175	0.654	**0.048**	0.409	0.798	0.222
***Novosphingobium* sp. PP1Y**	**0.004**	**0.019**	0.928	**<0.001**	**0.003**	**0.009**
Oblitimonas alkaliphila	0.599	0.311	**0.007**	0.599	0.311	0.311
***Oceanimonas* sp. GK1**	**0.001**	0.519	**0.001**	**0.001**	**0.023**	**0.002**
Oceanithermus profundus	**0.002**	**0.047**	**0.002**	**0.001**	0.227	**<0.001**
Octadecabacter antarcticus	0.443	0.942	**0.025**	0.443	0.477	**0.036**
Oleiphilus messinensis	**0.001**	0.094	**0.039**	**0.03**	**0.009**	0.956
Oleispira antarctica	0.457	0.216	0.285	0.779	0.155	**0.009**
Oligotropha carboxidovorans	**0.021**	0.4	**0.018**	**0.015**	0.322	**0.009**
***Olleya* sp. Bg11 27**	0.694	**0.011**	0.743	0.146	0.651	**0.005**
***Olsenella* sp. Marseille P2300**	**0.002**	0.317	0.061	**0.019**	0.061	0.317
***Olsenella* sp. oral taxon 807**	**0.034**	0.901	0.449	**0.034**	0.054	0.449
Olsenella uli	**0.006**	0.713	**0.037**	**0.006**	0.145	0.068
Onyong nyong virus	NA	**0.027**	NA	0.231	NA	**0.003**
***Opitutaceae* bacterium TAV5**	**0.025**	0.206	0.061	**0.005**	0.133	**0.005**
Opitutus terrae	**0.004**	0.128	0.129	**0.002**	**0.022**	**0.008**
Ornithobacterium rhinotracheale	0.256	0.16	0.382	**0.035**	0.256	**0.01**
Oscillatoria nigro viridis	1	0.49	0.411	0.544	0.411	**0.039**
Oscillibacter valericigenes	**0.016**	0.507	**0.002**	**0.01**	0.685	**<0.001**
***Ottowia* sp. KADR8 3**	**<0.001**	0.441	**<0.001**	**0.007**	0.055	**0.02**

***Ottowia* sp. oral taxon 894**	**0.004**	0.506	**0.001**	**0.004**	0.128	**0.001**
Paenibacillus phage PG1	0.116	**0.023**	0.097	NA	0.3	0.116
Paenibacillus sabinae	0.139	0.16	0.642	**0.038**	0.16	**0.043**
***Paenibacillus* sp. FSL P4 0081**	**0.006**	0.125	0.742	0.08	**0.006**	0.08
Paenibacillus stellifer	**0.018**	0.762	0.149	**0.011**	0.065	0.149
Paludibacter propionicigenes	0.425	0.08	0.08	0.08	0.76	**<0.001**
Pandoraea apista	0.362	**0.005**	0.885	**0.006**	0.362	**0.005**
Pandoraea faecigallinarum	0.147	0.088	0.197	**0.014**	0.34	**0.004**
Pandoraea pnomenusa	0.218	0.67	**0.001**	0.218	0.67	**0.005**
Pandoraea pulmonicola	0.233	0.173	0.065	0.086	0.639	**0.008**
Pandoraea sputorum	**0.001**	0.71	**0.013**	**0.004**	**0.048**	**0.048**
Pandoraea thiooxydans	**0.002**	0.283	**0.007**	**0.001**	**0.03**	**0.001**
Pandoraea vervacti	**0.001**	0.247	0.074	**0.001**	**0.001**	**0.002**
Pannonibacter phragmitetus	**0.037**	0.396	**0.037**	**0.037**	0.396	**0.037**
Pantoea gaviniae	0.066	**0.015**	**0.005**	0.797	0.797	0.871
Pantoea rwandensis	**0.008**	**0.039**	0.172	**0.004**	**0.004**	0.167
Pantoea vagans	0.425	**0.038**	0.784	**0.027**	0.29	**0.027**
Parabacteroides phage YZ 2015b	0.872	**0.044**	0.133	**0.044**	0.47	0.121
Paraburkholderia aromaticivorans	0.134	0.134	0.134	**0.036**	0.555	**0.008**
Paraburkholderia caledonica	**0.004**	0.549	0.102	**0.004**	**0.038**	0.059

Paraburkholderia hospita	**<0.001**		0.826		**0.001**	**<0.001**		**<0.001**				**0.021**
Paraburkholderia phymatum	**0.042**		0.209		0.101	**0.005**		0.201				**0.005**
Paraburkholderia phytofirmans	**<0.001**		0.614		**0.007**	**<0.001**		**0.006**				**0.027**
Paraburkholderia rhizoxinica	0.968		0.494		0.056	0.495		0.302				**0.014**
Paraburkholderia sprentiae	**0.007**		0.512		0.09	**0.007**		**0.027**				**0.04**
Paraburkholderia terricola	**0.001**		0.117		**0.001**	**0.001**		0.165		**<0.001**		
Paracoccus aminophilus	**0.007**		0.846		**0.007**	**0.007**		0.204				**0.017**
Paracoccus aminovorans	**0.031**		0.273		0.054	**0.003**		0.13				**0.003**
Paracoccus contaminans	**0.038**		0.827		0.827	**0.047**		**0.047**				0.827
Paracoccus denitrificans	**0.005**		0.308		**0.008**	**0.017**		0.134				0.134
***Paracoccus* sp. BM15**	**0.01**		0.473		0.473	**0.001**		**0.004**				0.232
***Paracoccus* sp. CBA4604**	**0.025**		0.187	**<0.001**		0.118		0.842				**0.015**
***Paracoccus* sp. SC2 6**	**0.036**		**0.021**		0.88	**0.002**		**0.028**				**0.007**
Paracoccus yeei	**0.002**		0.836		**0.002**	**0.002**		0.074				**0.002**
Paracoccus zhejiangensis	**0.008**		0.364		**0.039**	**0.001**		**0.034**				**0.005**
Parageobacillus thermoglucosidasius	**0.001**		0.426		0.115	**0.003**		**0.003**				0.585
***Paraliobacillus* sp. X 1125**	**0.037**		0.65		0.4	**0.049**		0.072				0.65
Paramecium bursaria Chlorella virus NY2A	0.068	NA		NA		0.087		**0.015**	NA			
Paraphotobacterium marinum	**0.007**		0.766		**0.018**	**0.007**		0.225				0.056
Pararhodospirillum photometricum	**0.017**		0.324		0.07	**0.011**		0.07				**0.011**

Parvibaculum lavamentivorans	**0.038**	0.172	0.061	**0.009**	0.261	**0.005**
Parvularcula bermudensis	**0.01**	0.073	**0.026**	**0.042**	**0.042**	0.812
Pasteurella dagmatis	0.494	0.05	0.941	**0.014**	0.429	0.05
Paucimonas lemoignei	**0.011**	0.276	0.055	0.055	0.055	0.589
Pectobacterium atrosepticum	**0.027**	0.432	0.516	**0.008**	**0.008**	0.655
Pediococcus pentosaceus	0.64	**0.039**	0.983	**0.039**	0.64	**0.02**
***Pedobacter* sp. PACM 27299**	0.233	0.169	0.233	**0.014**	0.619	**0.005**
Pedobacter steynii	0.937	**0.044**	0.487	0.08	0.828	**0.002**
Pelagibaca abyssi	**0.01**	0.338	**0.01**	**0.01**	0.244	**0.006**
Pelobacter acetylenicus	**0.003**	0.532	0.085	**0.003**	**0.034**	0.255
Pelobacter propionicus	**0.013**	0.407	**0.013**	**0.013**	0.241	0.105
***Pelobacter* sp. SFB93**	**0.037**	0.644	**0.007**	**0.03**	0.711	**0.007**
Peptoclostridium acidaminophilum	**0.047**	0.48	0.326	0.14	0.14	0.865
***Persicobacter* sp. JZB09**	0.229	**0.007**	0.32	**0.001**	0.344	**<0.001**
Petrimonas mucosa	0.233	**0.023**	0.078	**0.005**	0.932	**<0.001**
Phaeobacter gallaeciensis	0.107	0.41	**0.047**	**0.047**	0.55	**0.007**
Phaeobacter porticola	0.337	0.679	**0.048**	0.487	0.679	0.189
Phenylobacterium sp. HYN0004	**0.001**	0.688	**0.002**	**0.002**	0.066	**0.008**
Phenylobacterium zucineum	**0.014**	0.767	**0.014**	**0.012**	0.096	**0.014**
Photorhabdus thracensis	**0.007**	0.116	**0.001**	0.116	0.522	0.157

Phreatobacter cathodiphilus	**0.016**	0.576	**0.001**	**0.006**	0.576	**0.001**
Phycisphaera mikurensis	**0.009**	0.984	**0.033**	**0.004**	0.082	**0.033**
Pimelobacter simplex	**0.011**	0.828	**0.011**	**0.011**	0.131	**0.016**
Pirellula staleyi	0.155	0.314	**<0.001**	0.398	0.245	**0.033**
***Planctomyces* sp. SH PL14**	**<0.001**	0.284	**0.001**	**<0.001**	**0.002**	**0.036**
***Planctomyces* sp. SH PL62**	**0.002**	0.916	**0.001**	**0.001**	0.115	**0.001**
Planctopirus limnophila	0.286	0.411	**0.001**	0.411	0.411	0.055
Planococcus plakortidis	**0.007**	0.549	0.336	**0.004**	**0.023**	0.122
Planococcus rifietoensis	**<0.001**	0.07	0.484	**<0.001**	**<0.001**	0.215
***Plantactinospora* sp. BB1**	0.719	0.719	**0.004**	0.719	0.204	**0.004**
***Plantactinospora* sp. KBS50**	**0.003**	0.608	**0.045**	**0.016**	**0.034**	0.276
Pluralibacter gergoviae	**0.001**	0.803	**0.002**	**0.002**	**0.031**	**0.016**
***Polaribacter* sp. ALD11**	**0.026**	0.139	0.052	**0.016**	0.355	**0.006**
Polaromonas naphthalenivorans	**0.043**	0.265	**0.043**	**0.005**	0.629	**0.005**
***Polaromonas* sp. JS666**	**0.004**	0.747	**0.003**	**0.016**	0.256	**0.016**
Polyangium brachysporum	0.207	0.437	**0.045**	0.07	0.887	**0.012**
Polymorphum gilvum	**0.004**	0.787	**0.004**	**0.009**	0.165	**0.012**
Polynucleobacter necessarius	**0.039**	0.963	**0.038**	0.071	0.841	**0.039**
Polynucleobacter sp. MWH Creno 4B4	0.3	0.198	**0.004**	0.817	0.659	0.367
***Porphyrobacter* HT 58 2**	**0.007**	0.362	**0.001**	**0.046**	0.615	**0.03**

Porphyrobacter neustonensis	**0.001**	0.252	**<0.001**	**0.022**	0.252	**0.017**
***Porphyrobacter* sp. LM 6**	0.185	0.185	0.76	**0.002**	0.185	**0.038**
Porphyromonas asaccharolytica	0.217	0.082	0.082	**0.009**	0.837	**<0.001**
Porphyromonas crevioricanis	0.503	0.063	0.127	0.063	0.87	**<0.001**
Porphyromonas gingivalis	0.312	**0.016**	0.261	**0.01**	0.717	**<0.001**
Prevotella dentalis	0.051	0.295	0.268	**0.044**	0.1	0.055
Prevotella denticola	0.052	0.132	0.132	**0.027**	0.147	**0.014**
Prevotella intermedia	0.221	0.221	**0.019**	0.107	0.983	**0.004**
Prevotella melaninogenica	0.408	0.195	0.408	0.125	0.565	**0.041**
Propionibacterium freudenreichii	**<0.001**	0.599	0.087	**<0.001**	**<0.001**	0.267
Prosthecochloris aestuarii	**0.021**	0.536	0.058	**0.027**	0.086	0.186
***Prosthecochloris* sp. CIB 2401**	0.562	**0.007**	0.714	**0.007**	0.626	**0.005**
***Prosthecochloris* sp. GSB1**	**0.002**	0.184	**0.04**	**<0.001**	**0.041**	**0.002**
***Prosthecochloris* sp. HL 130 GSB**	**0.002**	0.252	0.144	**<0.001**	**0.01**	**0.01**
Proteiniphilum saccharofermentans	0.306	0.071	0.071	**0.047**	0.989	**<0.001**
Providencia rettgeri	**0.035**	0.948	0.595	**0.035**	**0.035**	0.595
Pseudarthrobacter chlorophenolicus	**<0.001**	0.059	**<0.001**	0.17	0.186	0.375
Pseudarthrobacter phenanthrenivorans	**0.004**	0.849	0.173	**0.003**	**0.009**	0.173
Pseudoalteromonas arctica	0.869	0.675	**0.032**	0.676	0.122	**0.03**
Pseudoalteromonas atlantica	0.368	**0.048**	0.368	**0.039**	0.176	0.134

Pseudoalteromonas luteoviolacea	**0.046**	0.94	**0.007**	**0.038**	0.472	**0.004**
***Pseudoalteromonas* sp. Bsw20308**	**0.004**	0.305	0.065	**0.001**	0.157	**0.019**
***Pseudoalteromonas* sp. SM9913**	**0.013**	0.364	0.476	0.19	**0.012**	0.476
Pseudoalteromonas tetraodonis	0.088	**0.023**	0.574	**<0.001**	**0.035**	**0.027**
Pseudodesulfovibrio aespoeensis	**0.005**	0.383	**0.012**	**0.002**	0.062	**0.002**
Pseudodesulfovibrio indicus	**0.004**	0.722	**0.009**	**0.004**	**0.028**	**0.028**
Pseudodesulfovibrio profundus	**0.02**	0.309	0.071	0.071	0.071	0.5
***Pseudogulbenkiania* sp. NH8B**	**0.03**	**0.03**	0.367	**<0.001**	0.055	**<0.001**
Pseudomonas aeruginosa	**0.016**	0.844	0.089	**0.016**	0.056	0.089
Pseudomonas azotoformans	**0.031**	0.085	**0.026**	0.195	0.285	0.522
Pseudomonas brassicacearum	**0.012**	0.072	0.295	**0.002**	**0.031**	**0.01**
Pseudomonas cerasi	0.572	**0.018**	0.607	0.121	0.56	**0.013**
Pseudomonas chlororaphis	0.058	0.583	**0.013**	**0.031**	0.696	**0.013**
Pseudomonas cichorii	0.169	0.833	**0.001**	0.38	0.833	**0.006**
Pseudomonas citronellolis	**0.003**	0.688	**0.011**	**0.003**	0.058	0.056
Pseudomonas corrugata	**0.028**	0.279	**0.017**	**0.034**	0.292	0.109
Pseudomonas cremoricolorata	**0.019**	0.476	**0.013**	**0.019**	0.292	**0.022**
Pseudomonas entomophila	**0.003**	0.223	0.093	**0.002**	**0.026**	**0.012**
Pseudomonas fulva	**0.002**	0.569	**<0.001**	**0.002**	0.243	**0.002**
Pseudomonas furukawaii	**<0.001**	0.854	**0.002**	**<0.001**	**0.005**	**0.007**

Pseudomonas knackmussii	**0.001**	0.995	**0.011**	**0.004**	**0.009**	**0.048**
Pseudomonas mandelii	**0.003**	0.608	**0.04**	**0.005**	**0.029**	0.157
Pseudomonas mendocina	**0.006**	0.99	**0.001**	**0.007**	0.491	**0.004**
Pseudomonas monteilii	**0.015**	0.89	**0.026**	**0.011**	0.104	**0.038**
Pseudomonas orientalis	**0.009**	0.633	0.917	0.062	**0.009**	0.606
Pseudomonas oryzihabitans	0.212	**0.04**	0.271	**0.006**	**0.04**	0.139
Pseudomonas parafulva	**<0.001**	0.979	**0.048**	**<0.001**	**<0.001**	0.054
Pseudomonas pseudoalcaligenes	0.239	0.453	**0.009**	0.111	0.82	**0.001**
Pseudomonas psychrotolerans	**0.024**	0.945	**0.014**	**0.024**	0.426	**0.024**
Pseudomonas putida	**0.016**	0.916	**0.035**	**0.016**	0.204	0.056
Pseudomonas resinovorans	0.066	**0.031**	**<0.001**	0.665	0.809	0.156
Pseudomonas rhizosphaerae	**0.035**	**0.007**	0.942	**<0.001**	**0.01**	**0.001**
Pseudomonas silesiensis	0.604	0.423	0.189	0.258	0.604	**0.011**
***Pseudomonas* sp. ATCC 13867**	**0.001**	0.742	**0.001**	**0.001**	0.06	**0.001**
***Pseudomonas* sp. CC6 YY 74**	0.214	0.235	0.576	**0.034**	0.214	0.075
***Pseudomonas* sp. FDAARGOS_380**	0.775	0.714	**0.036**	0.882	0.528	0.104
***Pseudomonas* sp. GR 6 02**	**0.002**	0.269	**0.002**	**0.012**	0.269	**0.039**
***Pseudomonas* sp. HLS 6**	0.091	0.532	**0.008**	0.237	0.842	0.091
***Pseudomonas* sp. JY Q**	NA	NA	**0.045**	NA	0.219	0.055
***Pseudomonas* sp. LG1E9**	**0.008**	**0.025**	0.909	**<0.001**	**0.008**	**0.017**

***Pseudomonas* sp. LPH1**	0.083	**0.005**	0.619	0.812	0.085	**0.003**
***Pseudomonas* sp. Lz4W**	0.078	0.447	**0.011**	**0.042**	0.447	**0.002**
***Pseudomonas* sp. M30 35**	0.658	**0.037**	0.658	**0.008**	0.658	**0.008**
***Pseudomonas* sp. MRSN12121**	**<0.001**	0.824	0.055	**0.001**	**0.024**	0.148
***Pseudomonas* sp. R2A2**	0.101	0.127	**0.042**	**0.009**	0.592	**<0.001**
***Pseudomonas* sp. SWI6**	0.068	NA NA		0.087	**0.015**	NA
***Pseudomonas* sp. XWY 1**	0.778	0.148	0.281	0.148	0.795	**0.027**
Pseudomonas stutzeri	0.053	0.4	0.053	**0.031**	0.299	**0.031**
Pseudomonas synxantha	**<0.001**	0.165	**0.005**	**0.03**	0.158	0.166
Pseudomonas syringae	0.075	0.412	0.075	**0.04**	0.413	**0.039**
***Pseudomonas syringae* group genomosp. 3**	0.356	**0.009**	0.963	0.356	0.356	**0.003**
Pseudomonas taetrolens	0.661	**0.036**	0.988	0.347	0.661	**0.017**
Pseudonocardia dioxanivorans	**0.001**	0.209	**0.025**	**<0.001**	**0.014**	**0.001**
***Pseudonocardia* sp. AL041005 10**	**0.036**	0.802	**0.042**	**0.036**	0.167	**0.036**
***Pseudonocardia* sp. HH130629 09**	**0.001**	0.098	**0.017**	**0.001**	**0.017**	**0.001**
***Pseudonocardia* sp. HH130630 07**	0.101	0.612	**0.024**	0.094	0.551	**0.028**
Pseudopedobacter saltans	0.304	**0.023**	0.806	0.766	0.304	**0.023**
Pseudopropionibacterium propionicum	**0.025**	0.108	**0.032**	**0.047**	0.064	0.532
Pseudothermotoga thermarum	**0.001**	0.22	0.813	**0.009**	**0.003**	0.27
Pseudoxanthomonas spadix	**0.011**	0.717	**0.011**	**0.011**	0.359	**0.011**

Pseudoxanthomonas suwonensis	**0.002**	0.585	**0.002**	**0.002**	0.065	**0.002**
Psychrobacter cryohalolentis	0.88	**0.004**	0.612	0.172	0.964	**0.004**
***Psychrobacter* sp. AntiMn 1**	**0.038**	0.94	0.061	**0.036**	0.173	0.061
Psychroflexus torquis	0.079	**0.012**	0.228	**<0.001**	**0.007**	**0.047**
***Psychromonas* sp. CNPT3**	0.861	0.15	0.476	0.334	0.78	**0.011**
Ralstonia insidiosa	**0.001**	0.161	**0.015**	**0.001**	**0.017**	**0.002**
Ralstonia pickettii	**0.011**	0.81	**0.037**	**0.012**	**0.037**	**0.037**
Ralstonia solanacearum	**0.008**	0.594	**0.011**	**0.008**	0.169	**0.008**
Ramlibacter tataouinensis	**0.001**	0.286	**0.019**	**0.001**	**0.019**	**0.003**
Raoultella ornithinolytica	0.131	**0.009**	**0.014**	0.496	0.915	0.327
***Raoultella* sp. X13**	0.401	**0.038**	0.401	0.401	0.783	**0.049**
Rathayibacter toxicus	**0.001**	0.516	**0.032**	**0.011**	**0.032**	0.348
Rathayibacter tritici	**<0.001**	0.09	**0.016**	**<0.001**	**0.002**	**<0.001**
Rhizobacter gummiphilus	**0.017**	0.231	**0.027**	**0.013**	0.171	**0.01**
Rhizobium etli	**0.027**	0.477	**0.027**	**0.027**	0.244	0.071
Rhizobium gallicum	**<0.001**	0.679	**0.001**	**0.001**	**0.032**	**0.015**
Rhizobium leguminosarum	**0.006**	0.732	**0.034**	**0.012**	0.053	0.097
Rhizobium phaseoli	**0.04**	0.118	**0.006**	0.073	0.426	0.091
***Rhizobium* sp. ACO 34A**	0.06	0.752	**0.008**	0.127	0.909	**0.035**
***Rhizobium* sp. CIAT894**	0.068	0.793	**0.001**	0.068	0.831	**0.001**

***Rhizobium* sp. Kim5**	0.139	0.335	0.139	**0.035**	0.719	**0.035**
***Rhizobium* sp. N324**	**0.038**	0.215	**0.038**	0.215	0.32	0.376
***Rhizobium* sp. NT 26**	0.053	0.293	**0.002**	0.17	0.938	0.092
***Rhizobium* sp. NXC14**	**0.016**	**0.004**	**<0.001**	0.62	0.73	0.763
***Rhizobium* sp. NXC24**	**0.021**	0.576	0.054	**0.04**	**0.042**	0.239
Rhizorhabdus dicambivorans	**0.01**	0.396	**0.021**	**0.004**	0.081	**0.003**
Rhodanobacter denitrificans	**0.026**	0.455	0.12	**0.009**	0.126	**0.009**
Rhodobacter capsulatus	**0.001**	0.927	**0.001**	**0.001**	0.08	**0.001**
***Rhodobacter* sp. CZR27**	**0.003**	0.785	0.104	**0.008**	0.104	0.285
Rhodobacter sphaeroides	**0.027**	0.091	**0.014**	**0.003**	0.422	**<0.001**
***Rhodobacteraceae* bacterium QY30**	**<0.001**	0.843	**0.001**	**<0.001**	**0.049**	**<0.001**
Rhodococcus coprophilus	**0.01**	**0.01**	0.7	**<0.001**	**0.008**	**0.024**
Rhodococcus hoagii	**0.002**	0.396	**0.002**	**0.002**	0.135	**0.005**
Rhodococcus jostii	**0.009**	0.553	**0.001**	**0.038**	0.518	**0.016**
Rhodococcus opacus	**0.001**	0.389	**0.011**	**0.002**	**0.005**	0.32
Rhodococcus ruber	**0.02**	0.995	**0.011**	**0.011**	0.294	**0.011**
***Rhodococcus* sp. B7740**	0.081	0.305	**0.046**	**0.017**	0.776	**0.007**
***Rhodococcus* sp. BH4**	0.124	0.572	0.231	0.066	**0.002**	0.713
***Rhodococcus* sp. MTM3W5 2**	**0.031**	0.591	**0.022**	**0.031**	0.347	**0.035**
***Rhodococcus* sp. PBTS 2**	0.055	0.103	0.453	**0.006**	0.107	**0.032**

***Rhodococcus* sp. S2 17**	**0.046**	0.808	**0.019**	0.152	0.68	**0.046**
***Rhodococcus* sp. WMMA185**	**<0.001**	0.792	0.079	**<0.001**	**0.003**	0.168
Rhodoferax koreense	**0.015**	0.681	**0.003**	**0.013**	0.516	**0.003**
Rhodomicrobium vannielii	0.938	0.138	0.716	0.14	0.938	**0.016**
Rhodopseudomonas palustris	**0.013**	0.747	**0.006**	**0.004**	0.409	**0.004**
Rhodospirillum centenum	**0.001**	0.813	**0.001**	**0.02**	0.151	**0.021**
Rhodospirillum rubrum	0.818	0.995	**0.021**	0.842	0.818	**0.021**
Rhodothermaceae bacterium	**0.007**	0.142	**0.018**	**0.001**	0.131	**0.001**
Rhodothermus marinus	**0.001**	**0.035**	**0.024**	**0.001**	**0.024**	**0.001**
***Rhodovulum* sp. P5**	**0.04**	0.463	**0.027**	0.07	0.409	0.097
Rhodovulum sulfidophilum	**<0.001**	0.506	**<0.001**	**0.002**	0.226	**0.004**
Robiginitalea biformata	**0.004**	0.155	**0.005**	**<0.001**	0.104	**<0.001**
Roseateles depolymerans	**0.005**	0.438	**0.002**	**0.002**	0.244	**0.001**
Roseiflexus castenholzii	**0.018**	0.377	**0.002**	**0.032**	0.434	**0.032**
***Roseiflexus* sp. RS 1**	0.155	0.433	0.063	0.063	0.639	**0.017**
Roseomonas gilardii	**0.01**	0.36	**0.01**	**0.001**	0.215	**0.001**
Rothia mucilaginosa	0.241	0.265	**0.041**	0.265	0.659	0.265
Rubrivivax gelatinosus	**0.025**	0.932	**0.006**	**0.041**	0.649	**0.025**
Rubrobacter radiotolerans	**0.004**	0.914	**0.007**	**0.005**	0.191	**0.014**
Rubrobacter xylanophilus	**0.009**	0.813	**0.007**	**0.014**	0.269	**0.014**

Ruegeria phage DSS3 P1	0.413	0.451	0.322	0.784	**0.03**	0.061
Ruegeria pomeroyi	**<0.001**	0.416	**<0.001**	**<0.001**	**0.001**	**0.002**
***Ruegeria* sp. NKC1 1**	0.052	0.806	0.806	0.052	**0.012**	0.543
***Ruegeria* sp. TM1040**	**0.027**	**0.027**	**0.028**	0.209	0.15	0.591
***Rufibacter* sp. DG31D**	**0.032**	0.528	**0.015**	**0.01**	0.417	**0.004**
Rufibacter tibetensis	0.366	0.108	0.366	**0.027**	0.366	**0.005**
***Ruminiclostridium* sp. KB18**	**0.014**	0.327	**0.014**	**0.014**	0.099	0.09
***Ruminococcaceae* bacterium CPB6**	**0.029**	0.772	0.126	**0.029**	0.126	0.129
Ruminococcus albus	0.374	0.388	**0.032**	0.135	0.442	**0.003**
Saccharomonospora viridis	0.076	0.277	**0.013**	0.277	0.397	0.277
Saccharopolyspora erythraea prophage pSE211	0.091	0.19	0.195	**0.005**	0.224	**0.006**
Saccharothrix espanaensis	**0.033**	0.622	**0.005**	**0.009**	0.622	**0.002**
***Salegentibacter* sp. T436**	0.968	0.597	0.143	0.597	0.344	**0.017**
Salinibacter ruber	0.088	0.226	0.183	**0.005**	0.1	**0.005**
***Salinibacterium* sp. CGMCC 1 16371**	**0.009**	**0.044**	**0.001**	0.099	0.196	0.358
Salinicola tamaricis	**0.036**	0.549	**0.003**	0.223	0.921	0.108
Salinispira pacifica	**0.036**	0.171	0.199	0.135	**0.036**	0.46
Salinispora arenicola	**0.003**	0.102	0.358	**<0.001**	**0.005**	**0.005**
Salinivirga cyanobacteriivorans	0.09	0.347	**0.003**	**0.01**	0.803	**<0.001**

Salipiger profundus	**0.002**	**0.027**	0.139	**<0.001**	**0.01**	**0.001**
Salmonella enterica	**0.028**	0.118	0.09	0.417	0.209	0.701
Sandaracinus amylolyticus	**0.007**	**0.043**	**0.043**	**<0.001**	**0.043**	**<0.001**
Sanguibacter keddieii	**0.002**	0.864	**0.003**	**0.002**	0.071	**0.008**
Saprospira grandis	0.091	0.825	**0.021**	0.294	0.825	0.102
Sedimenticola thiotaurini	0.62	0.378	0.378	0.245	0.96	**0.031**
Segniliparus rotundus	**0.041**	0.802	0.11	0.084	0.131	0.218
Selenomonas ruminantium	**<0.001**	0.175	**0.013**	**0.002**	**0.016**	0.175
***Selenomonas* sp. oral taxon 136**	**0.048**	0.328	0.328	**0.017**	0.086	0.086
***Selenomonas* sp. oral taxon 478**	**0.001**	**0.009**	**0.001**	**0.009**	**0.022**	0.562
***Selenomonas* sp. oral taxon 920**	**<0.001**	0.727	**0.032**	**<0.001**	**0.002**	**0.038**
Selenomonas sputigena	**0.005**	0.599	0.131	**0.001**	**0.019**	0.056
***Seonamhaeicola* sp. S2 3**	0.11	0.374	0.11	**0.024**	0.374	**0.024**
***Serinicoccus* sp. JLT9**	**0.011**	**0.038**	**<0.001**	0.182	0.486	0.192
Serratia ficaria	**0.003**	0.622	**0.006**	**0.003**	0.089	**0.004**
Serratia fonticola	**0.006**	0.06	**0.006**	0.087	0.302	0.302
Serratia marcescens	**0.003**	0.063	**0.001**	**0.038**	0.195	**0.039**
Serratia plymuthica	**0.001**	0.612	**0.015**	**0.003**	**0.004**	0.129
Serratia proteamaculans	0.863	**0.008**	0.637	**0.026**	0.637	**0.02**
***Serratia* sp. ATCC 39006**	0.937	0.265	0.162	0.315	0.379	**0.008**

***Serratia* sp. FGI94**	**<0.001**	0.477	0.09	**<0.001**	**0.001**	**0.04**
***Serratia* sp. YD25**	0.675	0.399	0.099	0.399	0.571	**0.016**
Shewanella baltica	0.218	0.832	**0.048**	0.365	0.949	0.21
Shewanella bicestrii	0.387	**0.014**	0.716	0.174	0.379	**0.014**
Shewanella psychrophila	0.098	0.066	0.1	**0.008**	**0.01**	0.368
Shewanella sediminis	0.683	**0.031**	0.546	**0.031**	0.749	**0.002**
Shewanella sp. phage 1 40	0.621	**0.043**	**0.004**	0.695	0.695	0.695
***Shewanella* sp. WE21**	0.611	0.057	0.546	0.25	0.435	**0.003**
Shewanella violacea	0.979	0.126	0.963	0.181	0.963	**0.046**
Shimwellia blattae	0.168	0.184	0.645	**0.021**	0.165	0.118
***Shinella* sp. HZN7**	**0.033**	0.337	0.106	**0.015**	0.196	**0.015**
Sideroxydans lithotrophicus	**0.011**	**0.005**	0.919	**<0.001**	**0.009**	**0.003**
Simplicispira suum	**0.015**	0.787	**0.014**	**0.015**	0.417	**0.015**
Singulisphaera acidiphila	**0.029**	0.21	0.689	**0.003**	**0.032**	0.081
Sinomonas atrocyanea	**0.001**	0.81	**0.001**	**0.003**	**0.036**	**0.006**
Sinorhizobium fredii	**0.011**	0.347	**0.033**	**0.006**	0.16	**0.007**
Sinorhizobium medicae	0.071	0.385	**0.002**	0.297	0.842	0.073
Sinorhizobium meliloti	**0.007**	**0.035**	**0.012**	**<0.001**	0.114	**<0.001**
***Sinorhizobium* sp. RAC02**	**0.004**	0.674	**0.046**	**0.004**	0.095	**0.046**
Slackia heliotrinireducens	0.075	0.636	0.636	**0.03**	0.075	0.313

Sneathia amnii	**0.023**	0.293	0.089	0.085	0.14	0.47
Sodalis glossinidius	**0.009**	0.837	**0.026**	**0.013**	0.061	0.061
Sodalis phage phiSG1	**0.026**	0.082	0.062	0.612	0.612	0.761
Sorangium cellulosum	**0.01**	0.803	**0.007**	**0.007**	0.285	**0.007**
Sphaerobacter thermophilus	**0.001**	0.211	**0.036**	**<0.001**	**0.014**	**0.001**
Sphingobacteriaceae **bacterium GW460 11 11 14 LB5**	0.172	0.189	0.172	**0.035**	0.624	**0.007**
Sphingobacterium **sp. 21**	0.259	0.056	0.366	**0.016**	0.366	**0.015**
Sphingobium chlorophenolicum	**0.042**	0.89	**0.026**	**0.026**	0.379	**0.026**
Sphingobium cloacae	**0.006**	**0.042**	0.192	**<0.001**	**0.016**	**0.001**
Sphingobium herbicidovorans	**0.001**	0.625	**0.001**	**0.001**	0.073	**0.001**
Sphingobium hydrophobicum	**0.002**	0.915	**0.017**	**0.002**	**0.013**	**0.013**
Sphingobium japonicum	**0.043**	0.147	0.083	**0.002**	0.147	**0.002**
Sphingobium **sp. EP60837**	0.05	0.529	0.05	**0.022**	0.256	**0.022**
Sphingobium **sp. MI1205**	0.177	0.177	**0.039**	0.572	0.572	0.572
Sphingobium **sp. SYK 6**	**0.007**	0.995	**0.007**	**0.007**	0.169	**0.019**
Sphingobium **sp. TKS**	**0.029**	0.828	0.165	**0.024**	0.08	0.134
Sphingobium **sp. YBL2**	0.122	0.869	**0.001**	0.264	0.851	**0.013**
Sphingobium yanoikuyae	0.056	**0.024**	0.056	**0.005**	0.258	**<0.001**
Sphingomonas koreensis	**0.005**	0.688	**0.005**	**0.004**	0.171	**0.005**

Sphingomonas panacis	0.208	0.208	0.351	**0.035**	0.351	**0.035**
Sphingomonas sanxanigenens	**0.008**	0.519	**0.008**	**0.008**	0.104	**0.025**
***Sphingomonas* sp. Cra20**	**0.001**	0.995	**0.002**	**0.004**	**0.037**	**0.01**
***Sphingomonas* sp. KC8**	0.584	0.12	0.192	0.096	0.831	**0.002**
***Sphingomonas* sp. LK11**	**0.044**	0.56	**0.02**	0.05	0.56	**0.02**
***Sphingomonas* sp. LM7**	**0.005**	0.808	**0.017**	**0.005**	0.076	0.059
***Sphingomonas* sp. MM 1**	**0.005**	0.838	**0.015**	**0.015**	**0.045**	**0.026**
***Sphingomonas* sp. NIC1**	**0.022**	0.409	**0.022**	**0.022**	0.17	0.112
Sphingomonas taxi	0.261	0.091	0.352	**0.018**	0.353	**0.005**
Sphingomonas wittichii	0.385	0.063	**0.006**	**0.031**	0.43	**<0.001**
Sphingopyxis alaskensis	0.489	0.589	0.068	0.203	0.589	**0.015**
Sphingopyxis fribergensis	**0.032**	0.578	0.089	**0.047**	0.101	0.204
Sphingopyxis granuli	**0.017**	0.254	**0.013**	0.168	0.254	0.254
***Sphingopyxis* sp. 113P3**	0.654	0.886	**0.012**	0.528	0.313	**0.012**
***Sphingopyxis* sp. MG**	**<0.001**	0.325	**<0.001**	**0.009**	0.214	**0.013**
***Sphingopyxis* sp. QXT 31**	0.212	0.212	0.212	**0.047**	0.438	**0.03**
Sphingopyxis terrae	**<0.001**	0.82	**<0.001**	**0.001**	**0.011**	**0.002**
Spiribacter curvatus	**0.019**	0.402	**<0.001**	**0.006**	0.887	**<0.001**
Spiribacter salinus	**0.004**	0.139	**0.007**	**0.024**	0.086	0.139
Spirochaeta thermophila	**0.018**	0.984	0.107	**0.002**	**0.035**	**0.035**

Spiroplasma cantharicola	**0.001**	**0.017**	**0.008**	**0.017**	**0.032**	0.833
Spiroplasma corruscae	**0.005**	0.586	0.613	**0.005**	**0.007**	0.432
Spiroplasma eriocheiris	**0.008**	0.077	**0.008**	0.056	0.082	0.399
Spiroplasma helicoides	**0.029**	0.414	0.253	**0.005**	0.083	0.062
Spiroplasma kunkelii	**<0.001**	0.075	0.111	0.053	**0.005**	0.53
Spiroplasma phoeniceum	**<0.001**	**0.004**	**0.004**	**0.017**	**0.004**	0.677
Spiroplasma syrphidicola	0.912	**0.023**	0.739	0.242	0.761	**0.003**
Spiroplasma turonicum	**0.006**	**0.04**	0.226	0.05	**0.018**	0.226
Spirosoma aerolatum	0.075	0.077	0.075	**0.017**	0.584	**<0.001**
Spirosoma montaniterrae	**0.002**	0.202	**0.002**	**0.001**	0.077	**<0.001**
Spirosoma rigui	0.229	0.408	0.176	0.068	0.513	**0.033**
***Spirosoma* sp. HA7**	0.503	0.204	0.263	0.154	0.966	**0.008**
***Spongiibacter* sp. IMCC21906**	**0.046**	0.904	0.34	**0.046**	0.054	0.34
***Sporosarcina* sp. P37**	**0.005**	0.426	**0.003**	**0.047**	0.587	**0.047**
Stackebrandtia nassauensis	**0.023**	0.545	**0.047**	**0.023**	0.16	**0.023**
Stanieria cyanosphaera	0.28	0.13	**0.018**	0.795	0.989	0.795
***Stanieria* sp. NIES 3757**	**0.011**	**0.011**	**0.044**	0.613	0.23	0.359
Staphylococcus cohnii	**0.001**	0.869	0.213	**0.003**	**0.007**	0.283
Staphylococcus condimenti	**0.04**	0.644	0.304	0.173	0.201	0.684
Staphylococcus lugdunensis	0.055	0.713	**0.028**	0.055	0.572	**0.028**

***Staphylococcus* sp. AntiMn 1**	**0.002**	NA		0.092	**0.006**	0.092	0.1
Staphylococcus stepanovicii	0.139		0.146	0.676	**0.032**	0.139	0.2
Starkeya novella	**0.01**		0.587	**0.01**	**0.004**	0.209	**0.004**
Stenotrophomonas acidaminiphila	0.229		0.68	**0.023**	0.211	0.989	**0.023**
Stenotrophomonas maltophilia	**0.005**		0.263	**0.031**	**0.001**	**0.031**	**0.001**
Stenotrophomonas rhizophila	**0.021**		0.64	0.098	**0.027**	0.09	0.101
***Stenotrophomonas* sp. LM091**	**0.009**		0.521	0.131	**0.009**	**0.028**	0.079
***Stenotrophomonas* sp. MYb57**	0.21		0.352	**0.004**	0.507	0.856	0.201
Stigmatella aurantiaca	**0.01**		0.844	**0.01**	**0.01**	0.199	**0.011**
Streptococcus agalactiae	**0.004**		0.278	**0.015**	**0.015**	0.077	0.227
Streptococcus australis	**0.037**		0.787	**0.001**	0.147	0.787	**0.035**
Streptococcus constellatus	0.794		0.567	0.094	0.876	0.148	**0.002**
Streptococcus infantarius	0.705		0.963	**0.008**	0.705	0.165	**0.009**
Streptococcus lutetiensis	0.612		0.284	**0.02**	0.931	0.612	0.612
Streptococcus marmotae	0.921		**0.002**	0.921	**0.039**	0.921	**<0.001**
Streptococcus mitis	0.624		0.239	0.116	0.52	0.239	**0.006**
Streptococcus oralis	0.139		0.838	**0.023**	0.073	0.838	**0.017**
Streptococcus pasteurianus	0.5		0.109	0.511	0.52	0.252	**0.015**
Streptococcus phage SpSL1	0.38		0.38	**0.036**	0.38	NA	**0.005**
Streptococcus phage YMC 2011	0.288		0.288	**0.001**	0.595	0.595	0.152

Streptococcus sobrinus	0.054	0.867	0.931	**0.031**	**0.031**	0.733
***Streptococcus* sp. A12**	0.075	0.937	**0.005**	0.137	0.937	**0.025**
***Streptococcus* sp. I G2**	0.119	0.807	**0.015**	0.162	0.807	0.07
***Streptococcus* sp. I P16**	0.165	0.89	**0.009**	0.195	0.915	**0.028**
***Streptococcus* sp. oral taxon 431**	**0.014**	0.64	**0.029**	**0.031**	**0.036**	0.09
Streptococcus uberis	**0.001**	0.516	0.719	**0.006**	**0.001**	0.347
Streptomyces actuosus	**0.046**	0.735	**<0.001**	**0.039**	0.792	**<0.001**
Streptomyces albireticuli	**0.01**	0.477	**0.026**	**0.01**	0.204	**0.01**
Streptomyces alboflavus	**<0.001**	0.286	**0.005**	**<0.001**	**0.005**	**0.001**
Streptomyces albulus	**0.003**	0.906	**0.011**	**0.004**	**0.014**	**0.037**
Streptomyces albus	0.212	0.242	0.11	0.061	0.543	**0.014**
Streptomyces ambofaciens	**0.004**	**0.004**	**<0.001**	0.376	0.486	0.515
Streptomyces atratus	**0.042**	0.585	**0.032**	**0.042**	0.435	**0.042**
Streptomyces autolyticus	0.246	0.283	**0.016**	0.583	0.9	0.283
Streptomyces avermitilis	**0.011**	0.344	**0.045**	0.095	0.095	0.405
Streptomyces bacillaris	0.182	0.212	0.551	**0.041**	0.212	0.154
Streptomyces bingchenggensis	**0.005**	0.087	**0.018**	**0.028**	**0.015**	0.848
Streptomyces cattleya	**0.006**	0.24	**0.007**	**0.004**	0.124	**0.004**
Streptomyces chartreusis	**0.016**	0.818	**0.044**	**0.016**	0.107	**0.044**
Streptomyces clavuligerus	**0.041**	0.09	0.093	**0.01**	0.145	**0.004**

Streptomyces collinus	0.084	0.654	**0.028**	**0.028**	0.654	**0.022**
Streptomyces cyaneogriseus	**0.024**	0.899	**0.024**	**0.038**	0.39	**0.039**
Streptomyces formicae	**0.015**	0.181	**0.016**	**0.004**	0.143	**0.001**
Streptomyces fulvissimus	**0.036**	0.052	**0.01**	0.532	0.532	0.861
Streptomyces gilvosporeus	**0.008**	0.502	**0.047**	**0.004**	0.096	**0.038**
Streptomyces glaucescens	**0.012**	0.802	**0.042**	**0.033**	0.074	0.074
Streptomyces globisporus	**<0.001**	0.812	**0.001**	**0.002**	**0.029**	**0.02**
Streptomyces globosus	**0.001**	0.523	**0.003**	**0.008**	**0.033**	0.069
Streptomyces griseochromogenes	**0.004**	0.397	**0.004**	**0.043**	0.199	0.099
Streptomyces griseus	**0.02**	0.727	0.463	**0.004**	**0.026**	0.176
Streptomyces lavendulae	**0.005**	0.515	**0.01**	**0.002**	0.269	**0.005**
Streptomyces lunaelactis	**0.015**	0.485	**0.011**	**0.011**	0.485	**0.011**
Streptomyces lydicus	**0.001**	0.091	**0.032**	**0.001**	**0.046**	**0.001**
Streptomyces malaysiensis	0.56	**0.037**	0.73	**0.017**	0.56	**0.017**
Streptomyces noursei	0.163	0.434	0.574	**0.033**	0.163	0.163
Streptomyces olivaceus	**<0.001**	0.916	**<0.001**	**<0.001**	**0.014**	**0.001**
Streptomyces pactum	**0.004**	0.123	**0.003**	**0.004**	**0.039**	**<0.001**
Streptomyces parvulus	**0.002**	0.958	**<0.001**	**0.003**	0.282	**0.002**
Streptomyces pluripotens	**0.01**	**0.038**	**<0.001**	0.336	0.634	0.336
Streptomyces pristinaespiralis	0.058	0.571	0.097	**0.027**	0.244	**0.035**

Streptomyces rubrolavendulae	**0.019**	0.131	0.93	**0.005**	**0.009**	0.131
Streptomyces scabiei	**0.016**	0.158	**0.016**	0.158	0.158	0.345
Streptomyces silaceus	**0.005**	0.093	0.103	**0.003**	0.061	**0.007**
***Streptomyces* sp. 3211**	**<0.001**	0.911	**<0.001**	**<0.001**	**0.007**	**0.001**
***Streptomyces* sp. 452**	**0.046**	0.895	**0.021**	**0.046**	0.816	**0.021**
***Streptomyces* sp. 4F**	0.09	0.051	0.314	**0.004**	0.33	**0.004**
***Streptomyces* sp. CCM_MD2014**	**0.028**	0.787	**0.028**	**0.028**	0.254	**0.034**
***Streptomyces* sp. CdTB01**	**0.009**	0.071	**0.003**	0.071	0.147	0.328
***Streptomyces* sp. CLI2509**	**0.001**	0.061	**0.001**	**<0.001**	0.103	**<0.001**
***Streptomyces* sp. CMB StM0423**	**0.039**	0.729	**0.039**	**0.039**	0.201	**0.039**
***Streptomyces* sp. CNQ 509**	**0.006**	0.774	**0.002**	**0.006**	0.229	**0.006**
***Streptomyces* sp. fd1 xmd**	0.111	**0.036**	0.783	**0.018**	0.1	**0.035**
***Streptomyces* sp. FR 008**	**<0.001**	**0.045**	0.079	**0.031**	**0.001**	0.483
***Streptomyces* sp. GSSD 12**	0.202	0.78	0.06	0.125	0.921	**0.041**
***Streptomyces* sp. HNM0039**	**0.013**	0.464	**0.041**	**0.012**	0.151	**0.013**
***Streptomyces* sp. Mg1**	0.06	0.664	**0.003**	**0.01**	0.887	**<0.001**
***Streptomyces* sp. MOE7**	0.127	0.135	0.169	**0.009**	0.261	**0.009**
***Streptomyces* sp. RTd22**	**0.044**	0.411	0.185	**0.003**	0.11	**0.007**
***Streptomyces* sp. S10 2016**	**0.001**	0.465	**0.037**	**0.001**	0.054	**0.035**
***Streptomyces* sp. S8**	**0.034**	0.365	**0.019**	0.143	0.365	0.189

***Streptomyces* sp. SAT1**	**0.005**	0.942	**0.005**	**0.007**	0.193	**0.011**
***Streptomyces* sp. SCSIO 03032**	**0.011**	0.521	**0.011**	**0.011**	0.209	**0.011**
***Streptomyces* sp. Sge12**	0.294	0.901	**0.005**	0.294	0.732	**0.005**
***Streptomyces* sp. SirexAA E**	0.315	0.574	**0.005**	0.384	0.733	**0.027**
***Streptomyces* sp. TN58**	**0.002**	0.262	0.061	**<0.001**	**0.013**	**0.006**
***Streptomyces* sp. Tue 6075**	0.781	0.781	**0.032**	0.781	0.357	**0.032**
***Streptomyces* sp. WAC00288**	**0.008**	0.493	**0.022**	**0.005**	0.086	**0.008**
***Streptomyces* sp. XZHG99**	**0.016**	0.987	**0.049**	**0.016**	0.067	0.067
***Streptomyces* sp. ZFG47**	**0.002**	0.515	**0.011**	**0.002**	**0.035**	**0.004**
Streptomyces spongiicola	**0.005**	0.074	**0.021**	0.074	0.074	0.595
Streptomyces venezuelae	**0.001**	0.782	**0.001**	**0.001**	**0.016**	**0.004**
Streptomyces vietnamensis	0.166	0.418	0.063	0.063	0.624	**0.009**
Streptomyces violaceoruber	**0.007**	0.619	**0.033**	**0.007**	**0.04**	**0.028**
Streptomyces violaceusniger	**0.023**	0.548	**0.013**	**0.01**	0.426	**0.01**
Streptomyces xiamenensis	0.08	0.162	0.249	**0.013**	0.162	**0.013**
Streptosporangium roseum	**0.038**	0.143	0.071	**0.01**	0.179	**0.01**
***Sulfitobacter* sp. AM1 D1**	**0.03**	0.288	0.145	**0.03**	0.145	**0.03**
Sulfuricaulis limicola	**0.001**	0.21	**0.004**	**0.001**	0.089	**0.001**
Sulfuricurvum kujiense	**0.039**	0.898	0.626	**0.028**	**0.015**	0.419
***Sulfuriferula* sp. AH1**	0.081	0.493	**0.019**	0.188	0.599	0.152

Sulfurifustis variabilis	**0.017**	0.453	0.099	**0.011**	**0.027**	**0.024**
Sulfurihydrogenibium azorense	0.057	**0.042**	**0.042**	0.676	0.496	0.496
Sulfurimonas autotrophica	**0.011**	0.241	0.834	0.057	**0.011**	0.241
Sulfurimonas denitrificans	**0.037**	0.591	0.375	**0.017**	**0.049**	0.195
Sulfuritalea hydrogenivorans	**0.002**	0.383	**0.02**	**0.002**	0.059	**0.002**
Sulfurospirillum barnesii	0.54	**<0.001**	**0.039**	**0.001**	**0.039**	**0.01**
***Sulfurospirillum* sp. JPD 1**	0.319	**0.003**	0.506	**0.004**	0.216	**0.004**
***Sulfurospirillum* sp. SL2 1**	0.162	0.858	**0.022**	0.162	0.858	**0.022**
Sulfurovum lithotrophicum	0.106	0.319	0.978	**0.009**	0.062	0.199
Symbiobacterium thermophilum	**0.01**	0.655	**0.01**	**0.009**	0.188	**0.009**
Synechococcus phage ACG 2014e	0.068	NA NA		0.087	**0.015**	NA
Synechococcus phage S MbCM100	0.124	0.168	0.277	0.058	0.168	**0.049**
Synechococcus phage S PM2	0.068	NA NA		0.087	**0.015**	NA
***Synechococcus* sp. CC9311**	0.976	0.862	**0.015**	0.862	0.197	0.091
***Synechococcus* sp. CC9902**	**<0.001**	0.812	0.728	**<0.001**	**<0.001**	0.642
***Synechococcus* sp. JA 2 3B a 2 13**	**0.022**	0.065	0.349	**0.001**	0.051	**0.004**
***Synechococcus* sp. JA 3 3Ab**	**0.023**	0.676	**0.023**	**0.004**	0.313	**0.004**
***Synechococcus* sp. KORDI 100**	**0.045**	0.663	0.592	**0.023**	**0.045**	0.251
***Synechococcus* sp. NIES 970**	0.135	0.208	**0.047**	**0.047**	0.814	**0.015**
***Synechococcus* sp. PCC 7003**	**0.002**	**0.007**	0.179	0.137	**0.006**	0.056

***Synechococcus* sp. PCC 7117**	**0.005**	0.061	0.156	0.132	0.061	0.4
***Synechococcus* sp. RCC307**	0.239	0.298	**0.041**	0.637	0.876	0.637
***Synechococcus* sp. WH 8103**	**0.003**	0.291	0.226	**0.046**	**0.018**	0.734
Syntrophobacter fumaroxidans	**0.04**	0.317	**0.039**	**0.012**	0.317	**0.007**
Syntrophomonas wolfei	**0.002**	0.659	0.187	**0.002**	**0.003**	0.494
Syntrophothermus lipocalidus	0.073	**0.03**	**0.03**	0.873	0.873	0.901
Syntrophus aciditrophicus	**0.015**	0.525	**0.017**	**0.015**	0.139	0.085
Tannerella forsythia	0.661	**0.043**	0.329	0.073	0.943	**0.001**
***Tannerella* sp. oral taxon HOT 286**	**0.049**	0.079	**0.017**	**0.017**	0.508	**<0.001**
Tateyamaria omphalii	**0.023**	0.394	**0.001**	0.071	0.881	**0.029**
Taylorella asinigenitalis	0.548	0.457	**0.003**	0.847	0.76	0.457
***Tenericutes* bacterium MO XQ**	0.075	0.396	0.386	**0.022**	**0.022**	0.973
Teredinibacter turnerae	0.098	**0.029**	0.924	0.889	0.098	**0.022**
Terriglobus roseus	**0.001**	0.885	**0.003**	**0.001**	**0.015**	**0.003**
Terriglobus saanensis	**0.001**	0.239	**0.006**	**0.02**	**0.012**	0.293
Tessaracoccus aquimaris	**0.001**	0.657	**0.001**	**0.001**	0.07	**0.005**
Tessaracoccus flavescens	**0.003**	0.969	**0.003**	**0.004**	0.124	**0.005**
Tessaracoccus flavus	**0.041**	0.808	**0.041**	**0.041**	0.351	0.061
***Tessaracoccus* sp. Marseille P5995**	**0.004**	0.782	0.063	**0.004**	0.063	0.063

***Tessaracoccus* sp. T2 5 30**	**0.028**	0.629	0.066	**0.023**	0.196	**0.039**
***Thalassococcus* sp. SH 1**	**0.041**	0.503	0.125	0.125	0.14	0.503
Thauera aromatica	0.526	0.141	**0.034**	0.695	**0.031**	**<0.001**
Thauera chlorobenzoica	0.054	0.487	**0.046**	**0.032**	0.569	**0.024**
Thauera humireducens	0.286	0.286	**0.022**	0.848	0.286	0.286
***Thauera* sp. K11**	**0.002**	0.225	**0.002**	**0.001**	0.088	**<0.001**
***Thauera* sp. MZ1T**	**0.022**	0.073	0.073	**0.001**	0.095	**<0.001**
Thermaerobacter marianensis	**0.002**	0.777	**0.002**	**0.002**	0.066	**0.002**
Thermanaerovibrio acidaminovorans	**0.001**	0.069	**0.001**	**0.005**	**0.049**	0.11
Thermoanaerobacter mathranii	**0.039**	**0.039**	**0.004**	0.292	0.495	0.495
Thermobacillus composti	0.054	0.441	0.225	**0.02**	0.151	0.054
Thermobispora bispora	**0.005**	0.36	**0.021**	**0.005**	**0.021**	0.188
Thermodesulfobacterium commune	1	**0.008**	1	0.142	1	**0.008**
Thermogutta terrifontis	**0.013**	0.164	**0.013**	**0.013**	**0.032**	0.31
Thermomonospora curvata	**0.042**	0.911	**0.036**	**0.042**	0.615	**0.036**
Thermosediminibacter oceani	0.259	**0.032**	0.676	**0.006**	0.262	**0.014**
Thermosipho melanesiensis	**0.002**	0.196	**0.006**	0.116	0.196	0.314
Thermovibrio ammonificans	**<0.001**	0.315	0.05	**<0.001**	**0.001**	0.315
Thermus aquaticus	**0.001**	0.729	**0.001**	**0.004**	**0.048**	**0.006**
Thermus brockianus	**0.002**	0.196	**0.007**	0.111	0.111	0.33

Thermus scotoductus	0.135	0.215	**0.006**	0.49	0.994	0.241
***Thermus* sp. CCB_US3_UF1**	**<0.001**	0.283	**0.001**	**0.003**	**0.022**	**0.044**
Thermus thermophilus	**<0.001**	0.597	**0.001**	**<0.001**	**0.001**	**0.003**
Thioalkalivibrio paradoxus	**0.013**	0.169	**0.004**	**0.031**	0.261	**0.031**
Thioalkalivibrio sulfidiphilus	**0.002**	0.629	**0.015**	**0.006**	**0.011**	**0.011**
Thiobacillus denitrificans	**0.017**	0.498	**0.024**	**0.017**	0.244	**0.017**
Thiocystis violascens	0.118	0.73	0.08	0.118	0.626	**0.029**
Thiohalobacter thiocyanaticus	**0.006**	0.844	**0.006**	**0.026**	0.262	**0.033**
Thiolapillus brandeum	0.151	0.938	**0.025**	0.151	0.938	**0.025**
Thiomonas intermedia	**0.008**	0.561	**<0.001**	**0.008**	0.619	**<0.001**
Thioploca ingrica	**0.027**	0.166	0.091	0.091	0.091	0.895
Tistrella mobilis	**0.007**	0.813	**0.016**	**0.007**	0.093	**0.016**
Tokyovirus A1	0.951	0.108	**0.013**	0.108	**0.013**	NA
Treponema azotonutricium	**0.007**	0.169	**0.007**	**0.021**	0.079	0.079
Treponema primitia	**0.023**	0.112	0.102	0.118	0.102	0.638
Trichormus azollae	0.122	0.114	**0.025**	**0.025**	0.972	**0.001**
Tropheryma whipplei	**0.003**	0.973	**0.003**	**0.003**	0.279	**0.006**
Truepera radiovictrix	**0.034**	0.211	0.173	**0.02**	0.173	**0.032**
Trueperella pyogenes	**0.024**	0.371	0.286	**0.001**	**0.044**	**0.044**
Tsukamurella paurometabola	**0.002**	0.432	**0.002**	**0.009**	0.171	**0.034**

Tsukamurella tyrosinosolvens	**0.006**	0.932	**0.005**	**0.006**	0.178	**0.005**
Tsukamurella virus TIN2	**<0.001**	0.392	0.124	**<0.001**	**<0.001**	0.091
Tumebacillus avium	**0.027**	0.823	0.217	**0.014**	0.085	0.167
Turneriella parva	0.066	0.247	**0.013**	**0.013**	0.749	**0.001**
Uncultured crAssphage	**0.001**	**0.006**	0.12	**<0.001**	**<0.001**	**0.022**
Uncultured phage WW nAnB	0.376	0.903	**0.025**	0.376	0.866	**0.025**
Uncultured phage WW nAnB strain 2	0.925	0.925	**0.025**	0.925	0.171	**0.025**
Uncultured phage WW nAnB strain 3	0.376	0.861	**0.026**	0.376	0.861	**0.026**
Ureaplasma parvum	**0.01**	0.929	**0.01**	**0.01**	0.289	**0.021**
Ureaplasma urealyticum	0.103	0.103	0.429	**0.024**	**0.047**	0.198
Variibacter gotjawalensis	**0.001**	0.828	0.132	**0.001**	**0.002**	0.082
Variovorax paradoxus	**0.031**	0.969	0.052	**0.031**	0.141	0.055
***Variovorax* sp. PMC12**	**0.012**	0.465	**0.007**	**0.047**	0.256	0.103
Veillonella rodentium	0.898	**0.004**	0.736	**0.045**	0.898	**0.012**
Verminephrobacter eiseniae	**0.007**	0.122	**0.017**	0.087	0.065	0.738
***Verrucomicrobia* bacterium IMCC26134**	**0.014**	0.162	**0.012**	**0.001**	0.312	**<0.001**
Verrucosispora maris	**0.005**	0.849	**0.024**	**0.005**	**0.042**	**0.035**
Vibrio diabolicus	0.773	0.449	**0.049**	0.449	0.233	0.317
Vibrio phage henriette 12B8	NA	0.062	NA	0.304	NA	**0.013**
Vibrio rotiferianus	0.408	**0.007**	0.811	**0.002**	0.187	**0.002**

Vibrio vulnificus	**0.002**	0.131	0.775	**0.001**	**0.002**	0.156
***Victivallales* bacterium CCUG 44730**	**0.014**	0.99	**0.003**	**0.009**	0.506	**0.003**
Virgibacillus pantothenticus	**0.027**	0.921	0.921	**0.027**	**0.027**	0.921
***Vitreoscilla* sp. C1**	0.691	0.195	0.195	0.484	0.195	**0.003**
Vulgatibacter incomptus	0.07	0.223	0.08	**0.001**	0.227	**0.001**
Weissella ceti	0.196	0.874	0.056	0.196	**0.015**	0.173
Weissella cibaria	**0.045**	**0.019**	0.629	0.917	**0.045**	**0.019**
Weissella paramesenteroides	**0.043**	**0.048**	0.709	**0.001**	0.061	**0.019**
Wenzhouxiangella marina	**0.004**	0.817	0.817	**0.001**	**0.002**	0.817
***Winogradskyella* sp. PG 2**	0.545	**0.025**	0.306	0.125	0.932	**0.025**
Xanthobacter autotrophicus	**0.001**	0.415	**0.001**	**0.004**	0.104	**0.018**
Xanthomonas citri	0.711	0.741	**0.024**	0.741	0.401	0.184
Xanthomonas sacchari	**0.001**	0.979	**0.041**	**0.001**	**0.005**	0.063
Xanthomonas translucens	**0.003**	0.465	**0.037**	**0.002**	**0.015**	**0.009**
Xanthomonas vasicola	**0.017**	0.334	**0.001**	0.378	0.864	0.32
Xanthomonas vesicatoria	**0.002**	0.932	0.057	**0.002**	**0.01**	0.057
Xanthomonas virus Xop411	0.38	0.201	**0.044**	NA	NA	NA
Xenorhabdus doucetiae	0.979	**0.037**	0.475	0.307	0.809	**0.001**
Xenorhabdus hominickii	0.054	0.054	**0.011**	0.365	0.579	0.579
Xylella fastidiosa	**0.02**	0.598	0.159	**0.018**	0.065	**0.02**

Yangia pacifica	0.135	0.837	0.076	0.125	0.949	**0.029**
***Yangia* sp. CCB MM3**	**0.004**	0.898	**0.004**	**0.001**	0.413	**0.001**
Yersinia aleksiciae	0.267	0.567	0.08	0.183	**0.047**	0.267
Zhongshania aliphaticivorans	0.171	0.814	**0.026**	0.204	0.814	0.072
Zobellella denitrificans	**0.001**	0.948	**0.015**	**0.001**	**0.014**	**0.015**
Zobellia galactanivorans	0.579	0.311	0.379	0.173	0.983	**0.036**
***Zoogloeaceae* bacteirum Par f 2**	0.425	0.451	**0.01**	0.642	0.89	0.33

Significantly different species, as established from Bracken following taxonomic classification with Kraken 2, between VO_2 max classification groupings (fair (N=15), good (N=44), excellent (N=60), and superior (N=118)). VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Significant differences are highlighted in bold. NA indicates it was not possible in this instance to determine a p value.

Supplementary table 6.5| Species significantly different between training frequency groupings

Species	Low vs. Medium	Low vs. High	Low vs. Extreme	Medium vs. High	Medium vs. Extreme	High vs. Extreme
Acetobacter aceti	0.273	**0.038**	0.642	0.273	0.642	0.273
Acetomicrobium mobile	**0.036**	0.193	**0.036**	0.639	0.576	0.441
Achromobacter insolitus	**0.008**	0.647	0.521	0.103	0.452	0.647
Acidaminococcus fermentans	0.823	0.823	**0.001**	0.863	**<0.001**	**<0.001**
Acidaminococcus intestini	0.052	**0.039**	**0.037**	0.3	**0.001**	**0.004**
Acidihalobacter prosperus	0.781	0.982	**0.002**	0.781	**<0.001**	**0.002**
Acidiphilium cryptum	**0.008**	0.208	0.625	0.551	0.121	0.549
Acidithiobacillus ferrivorans	**0.025**	**0.014**	0.403	0.405	0.523	0.175
Acinetobacter equi	0.167	0.062	0.721	**0.014**	0.357	0.343
Acinetobacter johnsonii	0.135	**0.003**	0.244	0.067	0.935	0.166
Acinetobacter nosocomialis	**0.017**	**0.015**	**0.015**	0.625	0.343	0.385
Acinetobacter phage ZZ1	0.611	**0.003**	0.611	**0.005**	0.611	0.07
Acinetobacter schindleri	**0.028**	0.136	0.608	0.901	0.608	0.737
Actinoalloteichus hoggarensis	0.964	0.964	**0.009**	0.964	**0.009**	**0.012**
Actinobacillus porcitonsillarum	0.177	**0.015**	**0.015**	0.16	0.055	0.25
Actinobacillus suis	0.235	0.388	0.066	0.127	**0.009**	0.202

Actinomyces meyeri	0.292	**0.01**	0.085	0.087	0.165	0.858
Actinomyces pacaensis	**0.001**	**0.001**	**0.018**	0.661	0.661	0.909
Actinosynnema mirum	0.321	0.651	0.052	0.651	**0.005**	**0.026**
Aequorivita sublithincola	0.515	0.205	0.052	0.052	**0.018**	0.296
Aerococcus christensenii	0.363	0.155	0.055	0.363	**0.012**	**0.012**
Aeromonas rivipollensis	0.277	**0.002**	0.269	**0.01**	0.371	0.457
***Aeromonas* sp. ASNIH3**	**0.006**	0.883	**<0.001**	0.057	0.07	**0.006**
***Aeromonas* sp. ASNIH4**	0.861	0.852	**0.042**	0.787	**0.037**	0.062
Alcanivorax dieselolei	0.218	0.344	**0.016**	0.087	0.109	**0.002**
Alcanivorax xenomutans	**0.002**	0.342	0.105	0.156	0.994	0.269
Aliivibrio salmonicida	0.69	0.094	0.451	**0.023**	0.451	**0.006**
Alkalilimnicola ehrlichii	0.383	**0.043**	0.383	0.266	0.609	0.609
Allofrancisella guangzhouensis	**0.033**	0.364	**0.009**	0.33	0.072	**0.033**
Altererythrobacter epoxidivorans	0.291	0.65	0.257	0.26	**0.034**	0.29
Altererythrobacter marensis	**0.001**	0.075	**<0.001**	0.481	**0.046**	**0.04**
***Alteromonas* sp. RW2A1**	0.972	**0.013**	0.461	**0.013**	0.461	**0.02**
Amsacta moorei entomopoxvirus	0.974	0.662	**0.008**	0.662	**0.008**	0.119
Amycolatopsis mediterranei	**0.044**	0.494	0.934	0.494	0.437	0.494
Anaerostipes hadrus	0.91	0.91	**0.009**	0.91	**0.002**	**0.01**
Anaplasma marginale	0.772	**0.006**	0.119	**0.005**	0.101	0.614

Anoxybacillus flavithermus		0.104	0.901	**0.034**	0.282	**0.008**	0.091
***Aureitalea* sp. RR4 38**		0.474	0.491	0.057	0.934	**0.04**	**0.04**
***Azospirillum* sp. M2T2B2**		**0.04**	0.508	0.519	0.613	0.813	0.813
Bacillus coagulans		**0.011**	**0.011**	0.315	0.579	**0.011**	**0.011**
Bacillus licheniformis		0.686	0.282	**0.006**	0.165	**0.003**	**0.041**
Bacillus smithii		**0.026**	**0.026**	0.695	0.607	0.327	0.269
***Bacillus* sp. IHB B 7164**	NA		NA	**0.008**	NA	**0.008**	0.083
Bacillus virus G		0.989	**0.001**	0.327	**0.001**	0.327	0.282
Bacillus virus TsarBomba		0.442	0.182	0.449	**0.044**	0.615	0.182
Bacterioplanes sanyensis		**0.003**	0.408	0.468	0.298	0.488	0.745
Bacteriovorax stolpii		**0.047**	0.323	0.323	**0.047**	0.791	0.323
Bacteroides heparinolyticus		0.876	0.147	0.147	0.091	0.091	**0.046**
Bacteroides thetaiotaomicron		**0.005**	**0.027**	**0.014**	0.732	0.732	0.29
Bartonella ancashensis		0.385	0.551	**0.016**	0.262	0.079	**0.016**
Bartonella bacilliformis		0.519	**0.046**	**0.046**	**0.024**	**0.024**	0.758
Bartonella henselae		**0.048**	**<0.001**	**0.001**	**0.004**	**0.03**	0.961
***Bartonella* sp. Raccoon60**		**0.042**	0.211	**0.031**	0.489	0.738	0.485
***Bartonella* sp. WD16 2**		0.702	**0.044**	0.706	0.098	0.972	0.26
Bartonella tribocorum		0.725	**0.005**	0.659	**0.015**	0.666	0.216
Beihai picorna like virus 72		0.22	0.28	**0.002**	0.146	**0.037**	**0.006**

Bibersteinia trehalosi	0.144	0.097	0.238	**0.025**	0.963	0.097
Bifidobacterium longum	**0.012**	**0.033**	0.241	0.987	0.987	0.987
Blattabacterium clevelandi	0.055	0.953	**0.002**	0.078	**0.021**	**0.002**
***Blattabacterium* sp. Blattella germanica**	0.081	**0.002**	0.881	0.076	0.405	0.055
Bordetella avium	0.059	**0.002**	0.866	0.053	0.268	**0.01**
Bordetella pertussis	0.16	0.103	**0.006**	0.539	0.112	0.326
Bordetella pseudohinzii	**0.014**	0.995	0.995	**0.036**	0.05	0.995
***Bordetella* sp. HZ20**	**0.001**	0.73	0.316	**0.031**	0.34	0.34
***Bradyrhizobium* sp. ORS 285**	0.272	0.339	**0.001**	0.938	**0.013**	**0.013**
***Bradyrhizobium* sp. S23321**	**0.008**	**0.008**	0.463	0.564	**0.008**	**0.008**
***Brucella* sp. 09RB8471**	0.755	0.755	**0.027**	0.742	**0.027**	**0.027**
Burkholderia insecticola	0.888	**0.007**	0.888	**0.001**	0.888	**0.007**
Burkholderia latens	0.176	0.214	0.214	0.541	**0.03**	0.096
Burkholderia oklahomensis	**0.046**	0.249	0.235	0.816	**0.046**	**0.046**
Burkholderia seminalis	0.305	**0.03**	0.154	**0.008**	**0.03**	0.726
Burkholderia virus BcepNY3	NA	**0.032**	NA	**0.032**	NA	0.241
Burkholderiales bacterium YL45	**0.017**	0.094	0.27	0.903	0.879	0.879
Caldisericum exile	**0.013**	**0.003**	0.271	0.271	0.734	0.271
***Calothrix* sp. NIES 3974**	**0.002**	0.905	0.414	**0.031**	0.791	0.414
Campylobacter fetus	0.144	0.256	**0.028**	0.051	0.051	**0.028**

***Campylobacter* sp. NCTC 13003**	**0.005**	**0.023**	0.674	0.938	0.376	0.376
Campylobacter volucris	0.702	**0.017**	0.702	**0.024**	0.702	**0.036**
Candidatus Atelocyanobacterium thalassa	**0.045**	0.274	**0.012**	0.491	0.128	0.064
Candidatus Blochmannia vafer	0.341	0.228	**0.042**	0.409	0.131	0.34
Candidatus Cardinium hertigii	**0.013**	0.508	0.613	0.508	0.508	0.845
Candidatus Cloacimonas acidaminovorans	**0.028**	0.105	**0.028**	0.943	0.85	0.85
Candidatus Hoaglandella endobia	0.052	0.972	**0.01**	0.13	0.13	**0.022**
Candidatus Kinetoplastibacterium sorsogonicusi	**0.026**	**0.032**	0.937	0.937	0.116	0.11
Candidatus Liberibacter asiaticus	0.383	0.705	**0.034**	0.705	0.075	0.075
Candidatus Moranella endobia	0.114	**0.025**	0.06	0.112	0.169	0.799
Candidatus Nitrosoglobus terrae	0.082	**0.002**	0.78	0.067	0.178	**0.018**
Candidatus Planktophila limnetica	**0.024**	0.46	0.46	0.46	0.596	0.596
Candidatus Planktophila versatilis	0.539	**0.016**	0.053	**0.038**	0.105	0.948
Candidatus Puniceispirillum marinum	**0.008**	0.131	0.679	0.477	0.344	0.477
Candidatus Walczuchella monophlebidarum	0.441	0.186	0.675	**0.037**	0.441	0.485
Capnocytophaga cynodegmi	**0.048**	0.79	0.82	0.267	0.267	0.82
***Catenovulum* sp. CCB QB4**	**0.04**	**0.04**	**0.04**	0.881	0.552	0.552
Caulobacter henricii	0.36	**0.032**	0.467	**0.004**	0.808	**0.034**
***Caulobacter* sp. K31**	**0.028**	0.76	0.488	0.079	0.488	0.488
Caulobacter virus phiCbK	0.615	0.087	0.615	**0.044**	0.615	0.182

Cellulomonas gilvus	**0.046**	**0.046**	0.528	0.681	0.638	0.531
***Cellvibrio* sp. PSBB023**	**0.025**	0.87	0.687	0.132	0.687	0.751
Chelatococcus daeguensis	0.381	0.888	**0.007**	0.517	**0.021**	**0.01**
Chitinophaga caeni	0.083	0.113	**0.023**	0.773	0.2	0.331
Chlamydia muridarum	0.165	**0.01**	0.369	0.275	0.653	0.284
Chlamydia trachomatis	**0.024**	0.342	**0.044**	**0.024**	**0.008**	0.109
Chlorobaculum parvum	**0.039**	0.112	0.263	0.89	0.84	0.84
Chlorobium phaeovibrioides	0.951	**0.043**	0.226	**0.043**	0.226	0.865
Chloroflexus aurantiacus	**0.007**	0.393	0.137	0.233	0.935	0.393
Chloroherpeton thalassium	0.829	0.842	**0.015**	0.842	**0.01**	**0.019**
Chroococcidiopsis thermalis	0.281	0.345	**0.041**	**0.049**	**0.049**	**0.041**
Chryseobacterium gallinarum	0.444	**0.026**	0.809	**0.007**	0.633	**0.026**
Chryseobacterium indologenes	**0.022**	0.077	**0.01**	0.885	0.361	0.361
***Chryseobacterium* sp. StRB126**	**0.038**	0.326	0.986	**0.005**	0.213	0.488
Chryseobacterium taklimakanense	**0.03**	0.499	0.499	0.481	0.685	0.685
***Citrobacter* sp. FDAARGOS_156**	0.418	**0.03**	0.787	0.113	0.787	0.113
Clavibacter insidiosus	0.188	0.624	0.071	0.623	**0.01**	0.071
Collimonas arenae	**0.022**	0.627	0.627	0.137	0.627	0.627
Colwellia beringensis	0.431	0.951	0.079	0.484	**0.03**	0.079
Colwellia psychrerythraea	0.104	0.312	**0.022**	0.058	**0.003**	0.129

***Confluentimicrobium* sp. EMB200 NS6**	**0.032**	**0.032**	0.228	0.228	0.614	0.228
Corynebacterium atypicum	0.22	**0.038**	0.315	0.22	0.958	0.281
Corynebacterium callunae	0.074	0.08	**0.001**	0.934	0.05	0.052
Corynebacterium casei	0.156	**0.044**	0.344	0.344	0.128	**0.037**
Corynebacterium deserti	**0.015**	0.088	0.14	0.868	0.865	0.865
Corynebacterium halotolerans	0.286	0.334	0.096	0.111	**0.021**	0.298
Corynebacterium humireducens	0.312	0.126	**0.028**	0.312	0.126	0.126
Corynebacterium minutissimum	**0.013**	0.141	0.102	0.445	**0.004**	**0.02**
Corynebacterium phocae	**0.01**	0.511	0.291	0.291	0.628	0.628
Corynebacterium pseudotuberculosis	0.562	**0.037**	0.302	**0.042**	0.142	**0.037**
Corynebacterium singulare	0.221	0.221	0.149	**0.048**	0.362	**0.048**
***Corynebacterium* sp. ATCC 6931**	**<0.001**	0.327	0.074	0.06	0.73	0.466
Cotia virus	0.074	0.626	0.438	**0.042**	0.782	0.324
Coxiella burnetii	0.583	**0.016**	**0.001**	**0.016**	**0.001**	0.097
***Cyanobium* sp. NIES 981**	0.281	0.696	0.053	0.428	**0.005**	**0.046**
Cyanophage S RIM32	NA	**0.032**	NA	**0.032**	NA	0.241
Dehalobacter restrictus	**0.009**	0.094	0.789	0.903	0.134	0.136
***Dehalobacter* sp. CF**	**0.016**	0.269	0.269	0.343	**0.016**	0.063
Deinococcus puniceus	**0.007**	0.053	0.986	0.986	0.271	0.424
Dermatophilus congolensis	**0.002**	**<0.001**	0.187	0.154	0.435	0.075

Desulfatibacillum alkenivorans	0.057	0.663	0.27	0.27	**0.033**	0.27
Desulfosporosinus acidiphilus	0.769	**0.049**	0.15	**0.049**	0.15	0.769
Desulfotalea psychrophila	**0.036**	0.218	**0.028**	**0.047**	0.452	**0.028**
Desulfotomaculum ferrireducens	**0.045**	0.487	0.487	**0.045**	0.487	0.445
Desulfurobacterium thermolithotrophum	**0.004**	0.491	0.986	0.491	0.102	0.545
***Devosia* sp. I507**	**0.04**	0.234	0.911	0.911	0.22	0.286
Dialister pneumosintes	**0.007**	**0.007**	0.465	0.465	0.223	0.084
Dickeya dadantii	**0.03**	0.309	0.302	0.309	**0.03**	0.094
***Dietzia* sp. oral taxon 368**	**0.029**	0.426	0.105	0.294	0.484	0.294
Dokdonia donghaensis	**0.018**	0.564	0.564	0.564	0.564	0.934
Eggerthella lenta	0.065	**0.011**	0.782	0.137	0.117	**0.031**
Eikenella corrodens	0.102	0.304	0.117	0.842	**0.005**	**0.04**
Elizabethkingia bruuniana	0.122	**0.006**	0.065	0.065	0.39	0.394
Elizabethkingia miricola	0.69	**0.038**	0.849	**0.011**	0.847	0.109
Elizabethkingia ursingii	0.141	**0.017**	0.689	0.141	0.196	**0.048**
Ensifer adhaerens	**0.002**	0.106	0.106	0.497	0.748	0.757
Enterobacter kobei	**0.044**	0.989	0.335	**0.044**	**0.003**	0.384
Enterobacter roggenkampii	0.123	0.05	0.296	0.233	0.05	**0.046**
***Enterobacter* sp. E20**	0.45	0.883	0.199	0.57	**0.035**	0.264
***Enterobacter* sp. FY 07**	0.881	0.439	**0.048**	0.439	0.069	**0.018**

***Enterobacter* sp. HK169**	0.887	0.887	**0.032**	0.887		**0.032**	0.119
***Enterobacter* sp. R4 368**	**0.002**	0.199	**0.035**	0.301		0.885	0.414
Enterococcus faecalis	**0.037**	**0.037**	0.501	0.706		0.482	0.324
Erwinia tasmaniensis	0.086	0.254	**<0.001**	0.85	**<0.001**		**0.001**
***Erysipelotrichaceae* bacterium I46**	**0.031**	0.217	0.888	0.541		0.071	0.243
Erythrobacter atlanticus	**0.015**	0.179	0.846	0.716		0.179	0.424
Escherichia albertii	0.63	**0.022**	0.579	**0.04**		0.503	**0.031**
Escherichia marmotae	**0.009**	**0.004**	0.273	0.323		0.734	0.273
Escherichia virus Cajan	0.117	0.417	**0.03**	0.794	**<0.001**		**0.03**
Escherichia virus JenP2	NA	0.132	**0.013**	0.132		**0.013**	0.366
Eubacterium eligens	**0.003**	0.413	0.314	0.314		0.455	0.733
***Exiguobacterium* sp. U13 1**	**<0.001**	**0.005**	0.232	0.475		0.327	0.493
Fermentimonas caenicola	**0.023**	0.224	0.188	0.622		0.89	0.765
***Flammeovirga* sp. MY04**	0.223	**0.041**	0.444	0.331		0.802	0.425
***Flavobacteriaceae* bacterium 3519 10**	0.552	**0.011**	0.827	0.083		0.885	0.11
Flavobacterium arcticum	**0.046**	0.287	0.863	0.863		0.38	0.575
Flavobacterium branchiophilum	0.223	0.95	0.075	0.24		**0.026**	0.075
Flavobacterium commune	**0.021**	**0.021**	0.455	0.89		0.455	0.455
Flavobacterium gilvum	**0.007**	**0.005**	0.325	0.441		0.377	0.222
***Flavobacterium* sp. HYN0056**	**0.012**	0.551	0.322	0.322		0.551	0.551

***Flavobacterium* sp. MEBiC07310**	**0.038**	0.352	0.484	0.65	0.65	0.909
Flexistipes sinusarabici	0.304	**0.041**	0.226	0.176	0.164	**0.022**
Francisella noatunensis	0.105	**0.007**	0.44	0.165	0.571	0.152
***Francisella* sp. FDC440**	0.061	**0.029**	0.517	0.294	0.517	0.294
***Francisella* sp. FSC1006**	**0.001**	0.126	0.61	0.61	0.587	0.671
***Francisella* sp. TX077308**	**0.019**	0.277	0.42	0.446	**0.019**	0.12
***gamma proteobacterium* HdN1**	0.12	**0.032**	0.405	0.405	0.405	0.12
Gardnerella vaginalis	**0.03**	0.829	0.114	0.054	0.829	0.114
***Geobacillus* sp. GHH01**	0.9	0.917	**0.021**	0.9	**0.021**	0.117
Geobacillus thermocatenulatus	**<0.001**	0.239	0.37	0.081	**0.003**	0.09
Geobacillus thermoleovorans	**0.009**	0.065	**<0.001**	0.791	**0.005**	**0.005**
Geobacter daltonii	0.087	**0.03**	0.663	0.156	0.104	**0.031**
Gordonibacter urolithinfaciens	**0.01**	0.312	0.389	0.312	**0.03**	0.295
Gramella flava	**0.036**	0.279	0.29	**0.036**	0.669	0.201
Granulosicoccus antarcticus	0.95	0.401	**0.02**	0.401	**0.023**	**0.004**
Haematospirillum jordaniae	**0.01**	0.085	0.085	0.747	**0.002**	**0.012**
Haemophilus haemolyticus	**0.001**	0.158	0.48	0.506	0.34	0.518
Haemophilus influenzae	0.193	**0.017**	0.542	0.193	0.659	0.265
Haemophilus parainfluenzae	0.097	**0.028**	0.238	0.238	0.874	0.415
Haemophilus pittmaniae	**0.022**	**0.01**	0.357	0.246	0.846	0.357

***Haemophilus* sp. oral taxon 036**	**0.002**	0.082	0.44	0.833	0.44	0.44
Halocynthiibacter arcticus	0.625	0.072	0.625	**0.03**	0.459	0.375
***Halomonas* sp. GFAJ 1**	**0.037**	0.165	0.22	0.929	0.929	0.929
Halorhodospira halochloris	0.117	**0.018**	**0.02**	0.176	0.117	0.334
Helicobacter acinonychis	0.089	**0.001**	0.258	0.059	0.977	0.198
Helicobacter cinaedi	0.865	0.528	**0.001**	0.602	**0.001**	**0.01**
Helicobacter himalayensis	**0.013**	0.364	0.227	0.699	0.747	0.608
Herbaspirillum rubrisubalbicans	0.152	0.565	**0.032**	0.585	0.192	0.152
Human fecal virus Jorvi2	0.115	0.537	**0.036**	0.19	0.344	0.096
Human gammaherpesvirus 8	**0.016**	0.202	**0.002**	**0.014**	0.181	**0.002**
Hydrogenovibrio crunogenus	0.201	0.201	0.764	**0.048**	0.764	0.356
Hyphomicrobium nitrativorans	0.336	0.336	0.754	**0.019**	0.336	0.612
***Hyphomonas* sp. Mor2**	0.12	**0.021**	0.888	0.341	0.316	0.12
Idiomarina loihiensis	**0.03**	0.74	**0.013**	0.188	0.194	**0.03**
Idiomarina piscisalsi	**0.001**	0.216	**0.021**	0.241	0.91	0.384
Janthinobacterium agaricidamnosum	**0.011**	**0.011**	0.86	0.357	0.357	0.357
Klebsiella michiganensis	**0.003**	0.55	0.097	0.053	0.641	0.166
***Klebsiella* sp. M5al**	0.363	0.159	**0.015**	0.363	**0.015**	**<0.001**
Komagataeibacter medellinensis	**0.019**	**0.002**	**0.025**	0.223	0.91	0.223
Kosakonia cowanii	**0.001**	**0.009**	0.366	0.898	0.116	0.131

Kosakonia radicincitans	**0.003**	0.228	0.228	0.291	0.558	0.909
***Laceyella* sp. FBKL4 010**	**0.013**	0.231	0.444	0.538	**0.013**	0.114
***Lacinutrix* sp. 5H 3 7 4**	**0.037**	0.517	0.928	0.517	0.509	0.71
Lactobacillus acetotolerans	0.179	0.246	0.129	0.833	**0.04**	**0.04**
Lactobacillus apis	**0.005**	0.483	0.477	**0.006**	0.451	0.451
Lactobacillus gallinarum	0.361	0.095	**0.047**	**0.047**	**0.042**	0.345
Lactobacillus helsingborgensis	**<0.001**	**0.005**	**0.033**	0.496	0.718	0.854
Lactobacillus helveticus	**0.02**	**0.02**	0.492	0.238	0.271	0.123
Lactobacillus parabuchneri	**0.049**	0.421	0.365	**0.049**	**0.049**	0.782
***Lactobacillus* sp. wkB8**	0.691	**0.011**	0.429	**0.003**	0.442	**0.003**
Lactococcus phage jm2	0.458	0.509	0.199	0.243	**0.035**	0.467
Lactococcus phage P680	0.081	0.916	**0.022**	0.146	**<0.001**	0.081
Lactococcus virus ASCC191	0.366	0.519	**0.008**	0.182	**0.001**	0.182
Lactococcus virus Bibb29	0.994	0.184	**0.018**	0.184	**0.018**	0.467
Lactococcus virus CB13	0.145	0.91	**0.011**	0.146	**<0.001**	0.093
Lactococcus virus CB14	0.557	0.14	**0.013**	0.061	**0.005**	0.461
Lactococcus virus Sl4	0.227	0.705	0.108	0.227	**0.004**	0.227
Leadbetterella byssophila	0.054	**0.022**	**0.022**	0.309	0.136	0.672
Legionella sainthelensi	**0.023**	0.406	0.921	**0.023**	0.138	0.633
Legionella spiritensis	0.141	**0.045**	0.141	0.381	0.459	0.987

Leptolyngbya boryana	**0.001**	0.838	0.488	**0.03**	0.547	0.547
Listeria innocua	0.253	0.084	0.091	0.253	**0.035**	**0.018**
Luteibacter rhizovicinus	**0.012**	0.488	0.934	0.199	0.133	0.488
Lutibacter profundi	**0.011**	**0.005**	0.727	0.405	0.282	0.175
Magnetococcus marinus	0.645	0.066	0.087	**0.044**	0.141	**0.01**
***Magnetospirillum* sp. XM 1**	**0.004**	0.091	0.743	0.743	0.091	0.222
Maribacter cobaltidurans	0.112	**0.008**	0.112	0.195	0.464	0.916
***Maribacter* sp. 1_2014MBL_MicDiv**	**0.003**	**0.017**	0.351	0.777	0.757	0.765
***Maribacter* sp. T28**	0.277	**0.029**	**0.029**	0.285	0.105	0.371
Marinobacter adhaerens	0.983	**0.019**	0.176	**0.016**	0.176	**0.016**
Marinovum algicola	**0.042**	0.581	0.495	0.495	0.732	0.581
Mariprofundus aestuarium	**0.022**	**0.01**	0.2	0.411	0.69	0.382
Megasphaera hexanoica	**0.006**	**0.039**	0.888	0.877	**0.022**	**0.039**
Methylacidiphilum fumariolicum	0.385	0.119	0.997	**0.012**	0.644	0.369
Methylobacillus flagellatus	**0.016**	0.532	0.589	0.514	0.589	0.845
Methylobacterium aquaticum	**0.016**	0.59	0.124	**0.016**	0.989	0.076
Methyloceanibacter caenitepidi	0.214	0.348	**0.019**	0.756	0.073	0.073
Methylomonas methanica	**0.022**	0.415	0.238	0.445	0.992	0.445
Methylotenera mobilis	0.395	**0.026**	0.98	0.061	0.575	0.075
***Microbacterium* sp. BH 3 3 3**	0.089	0.508	0.891	**0.033**	0.508	0.587

Microcystis aeruginosa	0.199	0.11	0.334	**0.018**	0.11	0.558
***Mitsuaria* sp. 7**	0.286	0.795	**0.037**	0.49	**0.005**	**0.037**
Mobiluncus curtisii	0.12	0.052	0.052	0.279	**0.009**	**0.009**
Moorea producens	**0.015**	0.56	**0.004**	**0.013**	0.136	**0.002**
Moorella thermoacetica	**0.019**	**0.019**	0.595	0.645	**0.019**	**0.019**
Moraxella catarrhalis	**0.001**	**0.002**	**0.015**	0.45	0.835	0.307
Muricauda ruestringensis	**<0.001**	0.08	0.087	0.192	0.283	0.981
Mycobacterium bovis	NA	NA	**0.008**	NA	**0.008**	0.083
Mycobacterium chimaera	0.859	0.859	**0.039**	0.859	**0.039**	**0.039**
Mycobacterium haemophilum	0.898	0.871	**0.031**	0.871	**0.031**	0.077
Mycobacterium leprae	**0.034**	0.075	**0.001**	0.996	**0.005**	**0.005**
Mycobacterium marinum	**0.003**	**0.024**	0.541	0.895	**0.012**	**0.025**
***Mycobacterium* sp. EPa45**	**<0.001**	**0.035**	0.362	0.362	0.099	0.362
***Mycobacterium* sp. JS623**	0.85	**0.009**	0.85	**0.009**	0.85	0.147
***Mycobacterium* sp. QIA 37**	**0.046**	0.154	0.2	0.934	0.934	0.934
Mycobacterium tuberculosis	**0.024**	0.602	0.266	0.211	0.833	0.456
Mycobacterium virus Ff47	NA	NA	**0.008**	NA	**0.008**	0.083
Mycobacteroides chelonae	**0.03**	0.123	0.439	1	0.439	0.439
Mycolicibacterium fortuitum	0.474	0.755	**0.044**	0.755	0.05	**0.044**
Mycolicibacterium goodii	0.076	**0.014**	0.408	0.408	0.575	0.365

Mycoplasma leachii	**0.001**	0.411	0.411	0.062	**0.017**	0.243
Mycoplasma pneumoniae	0.222	**0.028**	0.783	0.204	0.671	0.204
Myroides odoratimimus	0.124	0.136	0.541	**0.023**	0.136	0.424
***Myroides* sp. A21**	**0.047**	0.944	0.944	0.117	0.342	0.944
Myxococcus macrosporus	0.975	0.133	**0.014**	0.109	**0.014**	**0.002**
Nakamurella multipartita	**0.013**	0.737	0.911	0.155	0.153	0.737
Nautilia profundicola	0.072	**0.008**	0.334	0.194	0.721	0.202
Negativicoccus massiliensis	**0.016**	**0.007**	0.529	0.398	**0.016**	**0.007**
Neisseria elongata	0.236	0.657	**0.028**	0.657	**0.003**	**0.013**
Neisseria gonorrhoeae	**0.018**	0.356	0.857	0.356	0.125	0.356
***Neisseria* sp. 10022**	0.496	0.897	**0.017**	0.496	**0.031**	**0.02**
Neisseria zoodegmatis	0.672	**0.017**	0.167	**0.004**	0.103	0.672
Neomicrococcus aestuarii	0.372	0.611	**0.012**	0.302	**0.003**	**0.04**
Nitratireductor basaltis	**0.011**	**0.011**	0.504	0.714	**0.031**	**0.021**
Nitrosococcus halophilus	0.698	**<0.001**	0.698	**0.002**	0.767	0.09
Nitrosomonas europaea	0.677	**0.018**	0.456	**0.018**	0.456	0.456
***Nitrosomonas* sp. Is79A3**	0.089	**0.008**	0.994	0.242	0.311	0.142
Nitrosospira briensis	**0.045**	**0.049**	0.053	**0.001**	0.577	**0.005**
Nitrospira japonica	0.658	0.201	0.061	0.321	0.061	**0.042**
Niveispirillum cyanobacteriorum	**0.036**	0.922	0.922	0.23	0.232	0.922

Nocardia phage NBR1	NA	NA	**0.008**	NA	**0.008**	0.083
Nocardiopsis alba	0.767	0.471	**0.004**	0.471	**0.003**	**0.003**
Nonlabens dokdonensis	0.771	**0.014**	0.16	**0.007**	0.16	0.557
Nonlabens sediminis	**0.024**	0.69	0.535	0.356	0.535	0.706
***Nonlabens* sp. MB 3u 79**	0.1	0.164	0.1	0.991	**0.024**	**0.026**
Novosphingobium resinovorum	0.363	0.121	0.18	0.263	0.123	**0.039**
Oceanisphaera avium	**0.035**	0.123	**0.043**	0.764	0.538	0.495
Ochrobactrum anthropi	**0.016**	0.183	0.911	0.375	0.141	0.183
***Ochrobactrum* sp. A44**	0.629	**0.049**	0.679	**0.049**	0.629	0.398
Octadecabacter antarcticus	0.786	0.072	0.079	0.079	0.079	**0.026**
Oleispira antarctica	**<0.001**	**0.046**	0.256	0.256	0.256	0.89
Oscillatoria acuminata	0.325	0.782	**0.025**	0.325	**0.002**	0.137
Owenweeksia hongkongensis	**0.026**	0.619	0.115	0.161	0.619	0.161
***Paenibacillus* sp. CAA11**	0.165	**0.02**	**0.02**	0.12	0.165	0.896
***Paenibacillus* sp. FSL R5 0345**	**0.038**	0.928	0.985	0.349	0.479	0.987
Paludibacter propionicigenes	**0.042**	0.269	0.461	0.755	0.755	0.974
Pandoraea oxalativorans	0.876	0.876	**0.047**	0.916	**0.047**	**0.047**
Pandoraea vervacti	0.918	0.08	0.918	**0.024**	0.918	0.364
Pantoea gaviniae	**0.012**	**0.026**	0.135	0.661	**0.006**	**0.013**
Paraburkholderia terricola	0.487	0.146	0.146	0.146	0.088	**0.04**

Paracoccus mutanolyticus	**0.002**	0.333	0.872	0.197	**0.01**	0.352
***Paracoccus* sp. CBA4604**	0.136	0.201	0.15	0.987	**0.005**	**0.007**
Parageobacillus thermoglucosidasius	0.234	**0.018**	**0.019**	0.065	0.082	0.964
Pectobacterium atrosepticum	0.058	0.094	**0.005**	0.575	**0.015**	0.147
Pediococcus acidilactici	0.491	0.695	**<0.001**	0.874	**0.001**	**0.01**
Pediococcus inopinatus	0.297	0.293	0.081	0.649	**0.014**	**0.017**
Pedobacter cryoconitis	**0.011**	0.339	0.597	**0.022**	0.154	0.339
Pedobacter steynii	**0.011**	0.631	0.405	0.086	0.405	0.405
Pelagibacterium halotolerans	**0.038**	0.395	0.849	**0.029**	0.395	0.395
Pelodictyon phaeoclathratiforme	0.438	**0.02**	0.612	0.23	0.885	0.438
***Peptoniphilus* sp. ING2 D1G**	**0.037**	0.147	0.934	0.934	0.228	0.228
Perinet vesiculovirus	**0.004**	0.334	0.107	0.289	0.764	0.437
Photorhabdus asymbiotica	**0.003**	**0.021**	0.907	0.907	0.06	0.079
***Phycisphaerae* bacterium L21 RPul D3**	**0.02**	**0.019**	0.917	0.632	0.22	0.087
***Planctomyces* sp. SH PL14**	**0.008**	0.991	0.814	**0.039**	0.236	0.991
Planococcus faecalis	0.156	**0.003**	0.854	0.064	0.347	**0.019**
Planococcus plakortidis	**0.036**	0.333	0.632	0.375	0.158	0.333
Planococcus rifietoensis	0.413	0.733	**0.023**	0.816	**0.004**	**0.034**
***Plantactinospora* sp. BC1**	**0.021**	0.557	**0.04**	0.137	0.357	0.097
***Polaribacter* sp. ALD11**	0.987	**0.039**	**0.039**	**0.039**	**0.039**	0.987

***Polynucleobacter* sp. MWH Creno 4B4**		0.211	0.211	0.714	**0.028**	0.658	0.309
Porphyromonas crevioricanis		0.127	0.127	**0.047**	0.56	0.264	0.56
Prevotella enoeca		**0.004**	0.676	**0.03**	0.066	0.676	0.068
Prevotella fusca		**0.008**	0.534	**0.008**	0.146	0.534	0.132
Prevotella jejuni		**0.007**	0.313	**0.008**	0.215	0.423	0.215
Prevotella melaninogenica		**0.001**	0.339	**0.01**	0.188	0.78	0.273
Prevotella ruminicola		**0.002**	0.384	**0.036**	0.123	0.824	0.281
***Prevotella* sp. oral taxon 299**		**0.001**	0.286	**0.028**	0.327	0.935	0.385
Proteus phage PM 85	NA		NA	**0.008**	NA	**0.008**	0.083
Proteus vulgaris		**0.02**	0.491	0.634	**0.02**	0.064	0.896
Providencia alcalifaciens		**0.002**	0.104	**0.001**	**0.001**	0.169	**0.001**
Providencia heimbachae		**0.005**	0.341	0.341	0.341	0.644	0.708
Providencia rettgeri		0.401	0.102	0.705	**0.043**	0.705	0.164
Providencia stuartii		0.551	**0.028**	**0.02**	**0.028**	**0.02**	0.82
***Pseudanabaena* sp. PCC 7367**		0.068	**0.04**	0.368	0.368	0.837	0.769
Pseudoalteromonas arctica		**0.044**	**0.044**	0.368	0.783	0.783	0.783
Pseudoalteromonas luteoviolacea		0.418	0.108	0.703	**0.031**	0.418	0.365
Pseudoalteromonas marina		**0.019**	0.303	0.829	0.469	0.142	0.367
Pseudoalteromonas phage H101		0.241	**0.028**	0.148	0.182	0.584	0.582
Pseudoalteromonas translucida		0.439	0.078	0.438	**0.026**	0.55	0.078

Pseudoalteromonas tunicata	0.596	**0.001**	0.135	**<0.001**	0.13	0.49
Pseudodesulfovibrio profundus	**0.017**	**0.032**	0.491	0.934	0.437	0.344
Pseudomonas alcaliphila	0.709	**0.02**	0.391	**0.02**	0.368	**0.02**
Pseudomonas amygdali	**0.019**	0.378	0.378	0.219	0.051	0.219
Pseudomonas antarctica	**0.001**	0.625	0.759	**0.003**	0.079	0.625
Pseudomonas frederiksbergensis	0.953	**0.043**	0.606	**0.043**	0.606	0.069
Pseudomonas mesoacidophila	**0.044**	0.676	0.927	0.329	0.093	0.676
Pseudomonas orientalis	**0.008**	0.249	**0.034**	0.309	0.7	0.309
Pseudomonas phage PaBG	NA	**0.032**	NA	**0.032**	NA	0.241
Pseudomonas resinovorans	0.052	0.709	0.263	0.236	**0.016**	0.236
***Pseudomonas* sp. CC6 YY 74**	**0.007**	0.401	0.401	0.401	0.401	0.726
***Pseudomonas* sp. GR 6 02**	0.345	0.345	0.145	0.145	**0.021**	0.345
***Pseudomonas* sp. JY Q**	**0.014**	0.347	0.965	0.204	**0.014**	0.388
***Pseudomonas* sp. MRSN12121**	**0.015**	0.203	0.941	0.604	0.082	0.273
***Pseudomonas* sp. R2A2**	0.074	0.344	0.344	0.344	**0.025**	0.074
***Pseudomonas* sp. S 6 2**	**0.042**	**0.042**	0.114	0.833	0.833	0.833
***Pseudomonas* sp. SWI44**	0.201	**0.024**	.	0.201	0.475	0.201
Pseudomonas synxantha	**0.017**	**0.017**	0.592	0.581	0.483	0.363
***Pseudomonas syringae* group genomosp. 3**	**0.037**	0.108	0.228	0.934	0.934	0.934
Pseudomonas viridiflava	0.629	0.186	**0.039**	0.186	**0.039**	0.186

Pseudomonas virus PaMx11	0.104	0.263	0.472	NA		**0.024**	0.138
Pseudomonas virus PaMx28	NA	**0.032**	NA		**0.032**	NA	0.241
Pseudomonas yamanorum	0.344	0.121	0.052		0.344	**0.006**	**0.001**
Pseudopropionibacterium propionicum	**0.043**	**0.043**	0.721		0.369	0.309	0.237
Psychrobacter alimentarius	**0.04**	0.993	0.993		0.083	0.264	0.993
Psychrobacter arcticus	**0.037**	0.853	0.565		0.091	0.585	0.565
***Psychrobacter* sp. PRwf 1**	**0.036**	0.183	0.203	.	.	.	
Psychroflexus torquis	0.892	**0.012**	0.84		**0.012**	0.84	**0.023**
***Pusillimonas* sp. T7 7**	0.348	0.393	**0.008**		0.792	0.079	0.055
***Rhizobium* sp. 11515TR**	**0.003**	0.106	0.1		0.581	0.913	0.581
***Rhizobium* sp. CIAT894**	0.17	0.486	0.17		0.859	**0.027**	0.06
***Rhizobium* sp. Kim5**	0.446	0.295	**0.032**		0.531	0.127	0.295
***Rhizobium* sp. NT 26**	0.138	0.79	0.138		0.138	**0.005**	0.138
***Rhodobacteraceae* bacterium QY30**	**0.01**	0.445	0.418		0.418	0.724	0.757
Rhodococcus erythropolis	0.146	**0.002**	0.562		0.122	0.909	0.244
Rhodococcus qingshengii	0.241	**0.028**	0.148		0.194	0.567	0.567
Rhodococcus rhodochrous	0.658	**0.002**	**0.002**		**0.021**	**0.012**	0.573
Rhodococcus ruber	**0.031**	0.336	0.672		0.672	0.604	0.672
***Rhodococcus* sp. PBTS 2**	0.659	0.659	**0.004**		0.924	**0.011**	**0.004**
***Rhodococcus* sp. WB1**	**0.048**	0.646	0.646		**0.036**	0.499	0.533

Rhodococcus virus Pepy6	0.201	NA	**0.026**	0.385	0.201	0.201
Rickettsia felis	0.228	0.1	**0.001**	0.435	**0.01**	0.1
Rubrobacter xylanophilus	0.152	0.508	0.235	0.494	**0.048**	0.152
***Ruegeria* sp. NKC1 1**	**<0.001**	**0.004**	0.26	0.965	0.126	0.146
***Rufibacter* sp. DG15C**	**0.025**	0.174	0.94	0.569	0.26	0.569
Saccharophagus degradans	0.385	0.719	**0.008**	0.385	**0.002**	**0.027**
***Salinibacterium* sp. CGMCC 1 16371**	0.068	0.146	0.386	**0.014**	0.068	0.678
Salinivirga cyanobacteriivorans	**0.039**	0.357	0.104	0.357	0.963	0.357
***Selenomonas* sp. oral taxon 478**	0.155	**0.049**	0.923	0.178	0.327	0.155
***Serratia* sp. MYb239**	0.671	0.673	**0.042**	0.611	**0.043**	**0.043**
Shewanella benthica	0.142	0.073	0.165	0.438	0.05	**0.032**
Shewanella halifaxensis	**0.034**	0.678	0.779	0.361	0.361	0.974
Shewanella loihica	0.345	0.106	0.435	**0.033**	0.339	0.435
Shewanella marisflavi	**0.001**	**<0.001**	**0.014**	0.325	0.644	0.683
Shewanella putrefaciens	**0.022**	**0.026**	**0.001**	0.969	0.181	0.181
***Shewanella* sp. ANA 3**	0.69	0.881	**0.026**	0.881	**0.02**	**0.033**
***Shewanella* sp. MR 4**	**0.001**	0.1	0.1	0.354	0.638	0.651
***Shewanella* sp. MR 7**	**0.027**	0.314	0.944	0.384	0.285	0.384
***Shewanella* sp. W3 18 1**	0.201	**0.005**	.	0.062	0.475	0.152
***Shewanella* sp. WE21**	0.112	**0.006**	**0.007**	0.112	0.112	0.577

Siansivirga zeaxanthinifaciens	**<0.001**	0.086	0.504	0.514	0.278	0.514
Solitalea canadensis	**0.007**	0.36	0.579	0.36	0.482	0.948
Sphingobacteriaceae **bacterium GW460 11 11 14 LB5**	**0.005**	0.538	**0.033**	0.136	0.669	0.136
Sphingobacterium mizutaii	**0.044**	0.116	**0.044**	0.714	0.398	0.343
Sphingobacterium **sp. B29**	0.087	**0.023**	0.477	0.477	0.969	0.778
Sphingobacterium **sp. ML3W**	**0.017**	0.517	0.891	0.314	0.314	0.593
Sphingobium **sp. MI1205**	**0.034**	0.128	**0.017**	0.462	0.462	0.17
Sphingomonas hengshuiensis	**0.023**	0.542	0.9	0.352	0.096	0.542
Sphingomonas koreensis	**0.035**	0.303	0.621	0.621	0.621	0.621
Sphingopyxis terrae	0.387	**0.026**	0.515	0.164	0.896	0.387
Sphingorhabdus flavimaris	**0.008**	0.511	0.516	0.511	0.511	0.795
Sphingorhabdus **sp. YGSMI21**	0.64	0.623	**0.008**	0.426	**0.012**	**0.006**
Spiroplasma helicoides	0.304	0.304	**0.026**	0.12	**0.025**	0.12
Spirosoma aerolatum	**0.021**	0.537	**0.047**	0.197	0.537	0.164
Staphylococcus virus IPLAC1C	0.39	0.075	0.611	**0.021**	NA	0.245
Streptococcus mitis	0.524	**0.022**	0.341	0.141	0.524	0.341
Streptococcus pantholopis	0.399	**0.017**	0.245	**0.004**	0.465	**0.017**
Streptococcus pasteurianus	0.387	0.756	**0.046**	0.486	**0.007**	**0.046**
Streptococcus phage YMC 2011	**0.009**	**0.003**	0.47	0.391	0.084	**0.043**

Streptococcus pyogenes	0.606	0.088	0.362	**0.042**	0.314	0.362
***Streptococcus* sp. I G2**	0.795	0.795	0.05	0.982	**0.038**	0.058
Streptococcus virus 7201	0.156	0.338	0.831	**0.026**	0.14	0.668
Streptococcus virus DT1	0.195	0.824	0.083	0.195	**0.003**	0.195
Streptococcus virus Sfi21	0.39	0.078	0.611	**0.021**	NA	0.245
Streptomyces actuosus	0.83	0.524	**0.011**	0.524	**0.004**	**0.011**
Streptomyces ambofaciens	**0.015**	**0.011**	0.771	0.262	0.051	**0.015**
Streptomyces bacillaris	0.854	0.59	**0.049**	0.586	**0.049**	0.154
Streptomyces globosus	**0.014**	**0.002**	0.439	0.243	0.439	0.194
Streptomyces griseochromogenes	**0.009**	0.31	0.997	0.569	0.317	0.569
Streptomyces hygroscopicus	**0.024**	0.351	0.705	**0.014**	0.339	0.351
***Streptomyces* sp. SirexAA E**	0.468	0.887	0.055	0.551	**0.048**	0.055
***Streptomyces* sp. SM17**	0.439	0.24	0.107	0.343	**0.038**	**0.038**
Sulfitobacter pseudonitzschiae	**0.03**	0.088	0.617	0.96	**0.03**	0.066
Sulfuricurvum kujiense	**0.04**	0.513	**0.015**	0.307	0.513	**0.049**
Sulfurospirillum cavolei	0.201	**0.023**	0.681	0.309	0.5	0.222
Sulfurospirillum halorespirans	**0.024**	0.362	0.591	0.591	0.591	0.753
***Synechococcus* sp. CC9605**	**0.034**	0.358	0.348	0.358	**0.026**	0.131
***Synechococcus* sp. JA 3 3Ab**	0.117	0.68	**0.005**	0.069	**0.033**	**0.005**
***Synechococcus* sp. PCC 73109**	0.938	0.054	0.992	**0.033**	0.938	0.324

***Synechococcus* sp. PCC 7502**		0.379	**0.02**	0.086	0.065	**0.02**	**<0.001**
Tatumella citrea		0.894	**0.043**	0.894	0.069	0.894	0.072
Thalassolituus oleivorans		**0.006**	0.627	0.62	0.196	0.538	0.627
Thalassotalea crassostreae		0.091	**0.028**	0.453	0.453	0.743	0.453
Thauera chlorobenzoica		0.295	0.298	**0.01**	0.697	**0.042**	0.1
Thermoanaerobacter kivui		0.864	**0.019**	0.925	**0.043**	0.958	0.344
Thermodesulfobacterium commune		0.393	0.202	**0.048**	0.081	**0.024**	0.202
Thermodesulfobium acidiphilum		0.076	**0.001**	**0.024**	**0.04**	0.225	0.569
Thermosulfidibacter takaii		**0.007**	0.449	0.968	0.449	0.373	0.618
Thermus brockianus		0.065	**0.048**	0.384	0.537	**0.048**	**0.043**
Torque teno virus 4		**0.016**	0.322	0.125	0.322	0.653	0.322
Treponema azotonutricium		**0.007**	**0.007**	0.326	0.79	0.271	0.269
Treponema pedis		0.126	**0.024**	0.854	0.177	0.46	0.099
Treponema primitia		**0.004**	0.366	0.657	0.603	0.366	0.935
Treponema succinifaciens		0.857	**0.04**	0.857	**0.031**	0.857	0.106
Trichodesmium erythraeum		**0.001**	**0.033**	**0.034**	0.692	0.769	0.692
Tsukamurella virus TIN2		0.571	0.571	0.072	0.738	**0.032**	0.072
Turnip mosaic virus	NA		NA	**0.001**	NA	**0.001**	**0.032**
Uncultured crAssphage		0.717	0.347	**0.016**	**0.042**	**0.006**	**0.003**
Uncultured phage WW nAnB		0.626	0.074	0.561	**0.041**	0.57	0.113

Uncultured phage WW nAnB strain 2	0.475	0.113	0.454	**0.042**	0.475	0.113
Uncultured phage WW nAnB strain 3	0.612	0.081	0.561	**0.041**	0.57	0.113
Vibrio alginolyticus	0.782	0.585	**0.018**	0.646	**0.017**	**0.017**
Vibrio casei	**<0.001**	**0.036**	0.307	0.58	0.307	0.58
Vibrio gazogenes	0.838	0.533	**0.014**	0.533	**0.014**	0.164
Vibrio harveyi	0.85	0.214	**0.039**	0.214	**0.033**	**0.033**
Vibrio nigripulchritudo	0.752	**0.034**	**0.034**	**0.034**	**0.034**	0.618
Vibrio rotiferianus	0.069	0.182	**0.029**	0.89	0.267	0.267
Vibrio scophthalmi	**0.012**	0.793	0.866	0.201	0.201	0.793
Vibrio tubiashii	0.939	0.776	**<0.001**	0.776	**0.001**	**0.001**
Vibrio vulnificus	0.098	**0.015**	0.152	0.189	0.629	0.569
***Winogradskyella* sp. J14 2**	0.096	0.856	0.057	0.098	**0.015**	0.057
***Winogradskyella* sp. PG 2**	0.674	**0.01**	0.674	**0.047**	0.963	0.15
Woeseia oceani	0.149	0.971	**0.046**	0.432	0.149	0.149
Xanthomonas albilineans	**0.036**	0.506	0.586	0.506	0.506	0.85
Xanthomonas euvesicatoria	0.528	0.069	0.429	**0.036**	0.528	**0.036**
Xanthomonas virus Xop411	NA	**0.007**	NA	**0.007**	NA	0.147
Xenorhabdus nematophila	0.179	**0.012**	0.991	0.134	0.419	0.074
Yersinia aleksiciae	0.831	**0.014**	0.067	**0.014**	0.052	**0.008**

Significantly different species, as established from Bracken following taxonomic classification with Kraken 2, between training frequency groupings (low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week; N=23)). Training groups were based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. The significance threshold was set at 0.05. Significant differences are highlighted in bold. NA indicates it was not possible in this instance to determine a p value.

Supplementary table 6.6| Species trending towards increased or decreased with increasing VO_2 max grouping

Species	Direction of relative abundance with increased VO_2 max	Excellent	Fair	Good	Superior
Taylorella asinigenitalis	Increased	0.00029	0.00047	0.00067	0.00093
Oblitimonas alkaliphila	Increased	0.00079	0.00088	0.00098	0.00105
Helicobacter mustelae	Increased	0.00086	0.00095	0.00099	0.00117
Campylobacter iguaniorum	Increased	0.00095	0.00114	0.00138	0.00141
Moritella viscosa	Increased	0.00146	0.00156	0.00166	0.00193
***Janthinobacterium* sp. B9_8**	Increased	0.00148	0.0017	0.00193	0.00195
Sulfurihydrogenibium azorense	Increased	0.00158	0.00212	0.00213	0.00217
Thauera humireducens	Increased	0.00185	0.00207	0.00216	0.00249
Bordetella holmesii	Increased	0.00242	0.00283	0.00291	0.00298
Algibacter alginicilyticus	Increased	0.00243	0.00268	0.00306	0.00316
Streptococcus constellatus	Decreased	0.00268	0.00266	0.00257	0.00208

Species found to be increasing or decreasing over increasing VO_2 max groupings (fair (N=15), good (N=44), excellent (N=60), and superior (N=118)). VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2).Values are given as median relative abundances of each species per grouping.

Supplementary table 6.7| Species trending towards increased or decreased with increasing training frequency

Species	Direction of relative abundance with increased training frequency	Low	Medium	High	Extreme
Oceanisphaera avium	Increased	0	0.00035	0.00049	0.0007
Pseudomonas viridiflava	Increased	0.00112	0.00125	0.00151	0.00177
Erwinia tasmaniensis	Increased	0.00127	0.00143	0.0015	0.00298
Pectobacterium atrosepticum	Increased	0.00131	0.00154	0.00164	0.0024
Acinetobacter nosocomialis	Increased	0.00153	0.0019	0.00204	0.00226
***Polaribacter* sp. ALD11**	Increased	0.00183	0.00195	0.00265	0.00282
Maribacter cobaltidurans	Increased	0.00201	0.00262	0.00323	0.00336
***Sphingobacterium* sp. B29**	Increased	0.00253	0.0031	0.00321	0.00326
Chitinophaga caeni	Increased	0.0026	0.00321	0.00386	0.00403
Sphingobacterium mizutaii	Increased	0.00367	0.00441	0.0045	0.00483
Actinoalloteichus hoggarensis	Increased	0.00476	0.00489	0.00502	0.00687
Porphyromonas crevioricanis	Increased	0.00538	0.00594	0.00618	0.00714
Helicobacter cinaedi	Decreased	0.00103	0.0009	0.00078	0
Legionella spiritensis	Decreased	0.00111	0.00087	0.00076	0.00067
Mycobacterium leprae	Decreased	0.00139	0.00119	0.00114	0.00065
***Halomonas* sp. GFAJ 1**	Decreased	0.00141	0.00106	0.00102	0.00096

***Pseudanabaena* sp. PCC 7367**	Decreased	0.00151	0.00117	0.00112	0.0011
Aeromonas rivipollensis	Decreased	0.00152	0.00113	0.0006	0.00049
Actinobacillus porcitonsillarum	Decreased	0.00154	0.00132	0.00106	0.00084
Halorhodospira halochloris	Decreased	0.00158	0.00125	0.00101	0.00075
Methyloceanibacter caenitepidi	Decreased	0.00182	0.00143	0.0014	0.00089
Rhodococcus rhodochrous	Decreased	0.00183	0.00168	0.00099	0
Thalassotalea crassostreae	Decreased	0.00201	0.00159	0.00151	0.00144
Thauera chlorobenzoica	Decreased	0.00226	0.00205	0.00193	0.00106
Acetomicrobium mobile	Decreased	0.00248	0.00209	0.00208	0.00179
***Maribacter* sp. T28**	Decreased	0.00265	0.00214	0.00212	0.00177
Actinomyces meyeri	Decreased	0.01192	0.01073	0.00887	0.00876

Species found to be increasing or decreasing over increasing training frequency groupings (low (4-6 hours/week; N=89), medium (7-9 hours/week; N=91), high (10-12 hours/week; N=34), and extreme (13+ hours/week, N=23)). Training groups were based on self-reported, average hours of different activities per week. Values are given as median relative abundances of each species per grouping.

Supplementary table 6.8| Pathways significantly different between VO_2 max classification groupings

Pathway	Excellent vs. Fair	Excellent vs. Good	Excellent vs. Superior	Fair vs. Good	Fair vs. Superior	Good vs. Superior
ANAGLYCOLYSIS-PWY: glycolysis III from glucose	**0.013**	0.438	**0.017**	**0.026**	0.134	0.182
ARG POLYAMINE SYN: superpathway of arginine and polyamine biosynthesis	0.319	**0.034**	0.798	**0.016**	0.235	**0.034**
ARGORNPROST-PWY: arginine ornithine and proline interconversion	**0.017**	0.315	0.315	0.203	0.06	0.774
ARGSYN-PWY: L-arginine biosynthesis I via L-ornithine	0.63	0.052	0.362	**0.028**	0.831	**0.001**
ARGSYNBSUB-PWY: L-arginine biosynthesis II (acetyl cycle)	0.64	0.066	0.352	**0.03**	0.825	**0.001**
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	**0.031**	0.094	0.094	0.161	0.077	0.542
CALVIN-PWY: Calvin Benson Bassham cycle	0.289	0.098	0.922	**0.038**	0.289	**0.038**
FOLSYN-PWY: superpathway of tetrahydrofolate biosynthesis and salvage	0.087	0.479	0.479	0.094	**0.022**	0.087
GLUCONEO-PWY: gluconeogenesis I	0.595	0.595	**0.017**	0.827	0.595	0.225
GLUTORN-PWY: L-ornithine biosynthesis	0.569	0.05	0.382	**0.019**	0.972	**0.001**
GOLPDLCAT-PWY: superpathway of glycerol degradation to 1-3-propanediol	0.368	**0.019**	0.058	0.252	**0.049**	**<0.001**

HISDEG-PWY: L-histidine degradation I	0.177	0.41	0.052	0.052	0.881	**0.008**
HOMOSER METSYN PWY: L-methionine biosynthesis I	0.5	0.164	0.561	0.689	0.318	**0.019**
MET SAM PWY: superpathway of S-adenosyl-L-methionine biosynthesis	0.124	0.091	0.972	0.972	0.124	**0.046**
METHANOGENESIS-PWY: methanogenesis from H_2 and C_{O2}	**0.003**	0.76	0.516	**0.014**	**0.015**	0.666
METSYN-PWY: L-homoserine and L-methionine biosynthesis	0.366	0.15	0.629	0.807	0.239	**0.029**
NONOXIPENT-PWY: pentose phosphate pathway non-oxidative branch	0.183	**0.005**	0.988	**0.001**	0.146	**0.001**
P108-PWY: pyruvate fermentation to propanoate I	0.231	0.491	**0.005**	0.552	0.915	0.231
P161-PWY: acetylene degradation	0.661	**0.004**	0.41	0.116	1	**0.008**
P23-PWY: reductive TCA cycle I	0.578	0.065	0.228	0.124	0.987	**<0.001**
P261-PWY: coenzyme M biosynthesis I	0.617	0.294	0.294	0.379	NA	**0.022**
P441-PWY: superpathway of N acetylneuraminate degradation	0.362	**0.008**	0.362	0.45	0.517	**0.045**
PENTOSE P PWY: pentose phosphate pathway	0.521	0.521	**0.021**	0.944	0.464	0.344
POLYAMSYN-PWY: superpathway of polyamine biosynthesis I	0.103	0.219	0.528	**0.038**	**0.038**	0.42
PPGPPMET-PWY: ppGpp biosynthesis	0.916	0.123	0.766	0.372	0.828	**0.038**
PRPP-PWY: superpathway of histidine, purine, and pyrimidine biosynthesis	0.06	0.536	**0.018**	0.083	0.531	0.06
PWY-1042: glycolysis IV plant cytosol	0.207	0.068	0.546	**0.023**	0.285	**0.023**
PWY-3841: folate transformations II	0.369	**0.039**	0.813	0.736	0.369	**0.039**

PWY-4242: pantothenate and coenzyme A biosynthesis III	0.52	**0.046**	**0.046**	0.931	0.931	0.931
PWY-4702: phytate degradation I	0.507	0.507	**0.048**	0.64	0.899	0.507
PWY-5005: biotin biosynthesis II	0.686	**0.008**	0.686	0.36	0.564	**0.002**
PWY-5097: L-lysine biosynthesis VI	0.863	0.583	0.337	0.583	0.583	**0.038**
PWY-5121: superpathway of geranylgeranyl diphosphate biosynthesis II via MEP	0.542	0.147	0.513	0.398	0.398	**0.012**
PWY-5138: unsaturated even numbered fatty acid beta oxidation	0.456	0.105	1	0.159	0.456	**0.049**
PWY-5154: L-arginine biosynthesis III via N-acetyl-L-citrulline	0.595	0.054	0.747	**0.037**	0.649	**0.012**
PWY-5188: tetrapyrrole biosynthesis I from glutamate	0.256	**0.004**	0.601	0.256	0.256	**0.004**
PWY-5189: tetrapyrrole biosynthesis II from glycine	**0.04**	0.203	**0.004**	**0.011**	**0.003**	0.25
PWY-5198: factor 420 biosynthesis	**0.001**	0.533	0.444	**0.008**	**0.008**	0.722
PWY-5347: superpathway of L-methionine biosynthesis transsulfuration	0.342	**0.042**	0.599	0.599	0.27	**0.014**
PWY-5384: sucrose degradation IV sucrose phosphorylase	0.167	0.058	**<0.001**	0.944	0.317	0.058
PWY-5509: adenosylcobalamin biosynthesis from cobyrinate a,c-diamide I	**0.007**	NA	NA	**0.015**	**<0.001**	NA
PWY-5656: mannosylglycerate biosynthesis I	0.548	**0.044**	0.697	0.548	0.548	**0.041**
PWY-5676: acetyl CoA fermentation to butanoate II	0.067	0.253	**0.039**	0.253	0.356	0.345
PWY-5690: TCA cycle II (plants and fungi)	0.087	0.398	**0.008**	**0.03**	0.814	**0.003**

PWY-6167: flavin biosynthesis II (archaea)	**0.02**	0.922	0.473	**0.02**	0.067	0.473
PWY-6168: flavin biosynthesis III (fungi)	**0.014**	0.561	0.66	0.073	**0.014**	0.561
PWY-621: sucrose degradation III (sucrose invertase)	0.701	0.409	0.071	0.409	0.409	**0.009**
PWY-6270: isoprene biosynthesis I	0.5	0.169	0.5	0.5	0.426	**0.02**
PWY-6317: galactose degradation I (Leloir pathway)	0.298	0.617	**0.004**	0.469	0.617	**0.029**
PWY-6318: L-phenylalanine degradation IV mammalian via side chain	0.114	0.304	NA	0.405	**0.025**	0.169
PWY-6353: purine nucleotides degradation II (aerobic)	0.604	0.087	0.27	0.12	0.604	**0.003**
PWY-6527: stachyose degradation	0.741	0.226	**0.009**	0.385	0.149	0.385
PWY-6595: superpathway of guanosine nucleotides degradation plants	0.182	0.316	**0.046**	**0.046**	**0.01**	0.451
PWY-6609: adenine and adenosine salvage III	**0.033**	0.272	0.12	**0.033**	0.247	**0.033**
PWY-6612: superpathway of tetrahydrofolate biosynthesis	0.072	0.404	0.502	0.098	**0.023**	0.072
PWY-6628: superpathway of L-phenylalanine biosynthesis	0.636	0.966	**0.03**	0.636	0.636	**0.03**
PWY-6700: queuosine biosynthesis	0.427	0.071	0.418	0.427	0.235	**0.002**
PWY-6892: thiazole biosynthesis I (*E. coli*)	**0.008**	0.07	**0.008**	0.338	0.338	0.735
PWY-6936: seleno amino acid biosynthesis	0.168	0.869	0.168	0.168	**0.015**	0.222
PWY-6969: TCA cycle V 2-oxoglutarate:ferredoxin oxidoreductase	0.066	0.355	**0.031**	**0.031**	0.754	**0.008**
PWY-7039: phosphatidate metabolism as a signaling molecule	**0.013**	0.364	0.476	0.19	**0.012**	0.476

PWY-7115: C4 photosynthetic carbon assimilation cycle (NAD-ME type)		0.525	0.525	0.074	0.401	0.827	**0.022**
PWY-7211: superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis		0.155	0.892	**0.002**	0.12	0.892	**0.001**
PWY-7219: adenosine ribonucleotides de novo biosynthesis		0.262	0.076	0.145	**0.022**	0.323	**<0.001**
PWY-7221: guanosine ribonucleotides de novo biosynthesis		0.476	0.342	0.342	0.295	0.82	**0.041**
PWY-7286: 7-(3-amino-3-carboxypropyl)-wyosine biosynthesis		**0.004**	0.263	0.131	0.08	0.08	0.733
PWY-7357: thiamin formation from pyrithiamine and oxythiamine yeast		0.569	**0.043**	0.556	0.091	0.771	**0.011**
PWY-7383: anaerobic energy metabolism (invertebrates, cytosol)		**0.023**	0.325	**0.002**	**0.003**	0.325	**<0.001**
PWY-7392: taxadiene biosynthesis (engineered)		**0.022**	0.831	0.692	0.058	0.058	0.952
PWY-7399: methylphosphonate degradation II	NA		**0.039**	0.268	0.268	0.534	0.056
PWY-7400: L-arginine biosynthesis IV (Archaebacteria)		0.64	0.055	0.363	**0.037**	0.809	**0.001**
PWY-7560: methylerythritol phosphate pathway II		0.5	0.169	0.5	0.5	0.432	**0.02**
PWY-7663: gondoate biosynthesis (anaerobic)		0.761	**0.018**	0.494	0.163	0.494	**0.001**
PWY-821: superpathway of sulfur amino acid biosynthesis (*Saccharomyces cerevisiae*)		0.691	**0.043**	0.682	0.166	0.896	0.078
PWY-841: superpathway of purine nucleotides *de novo* biosynthesis I		0.676	0.676	**0.021**	0.676	0.676	0.234

PWY-922: mevalonate pathway I	0.638	0.638	**0.034**	0.689	0.689	0.341
PWY0-1061: superpathway of L-alanine biosynthesis	0.314	0.201	0.121	0.056	0.919	**0.005**
PWY0-1297: superpathway of purine deoxyribonucleosides degradation	0.24	**0.012**	0.276	**0.009**	0.276	**<0.001**
PWY0-1298: superpathway of pyrimidine deoxyribonucleosides degradation	0.438	0.395	0.093	0.093	0.438	**0.013**
PWY0-1479: tRNA processing	0.059	**0.026**	0.717	0.717	**0.047**	**0.01**
PWY0-42: 2-methylcitrate cycle I	0.617	0.294	0.294	0.379	NA	**0.022**
PWY0-845: superpathway of pyridoxal 5'-phosphate biosynthesis and salvage	0.771	0.422	0.422	0.771	0.422	**0.049**
PWY4LZ-257: superpathway of fermentation (*Chlamydomonas reinhardtii*)	0.662	**0.008**	0.662	0.219	0.744	**0.008**
PWY66-422: D-galactose degradation V (Leloir pathway)	0.269	0.706	**0.008**	0.398	0.706	**0.029**
PYRIDNUCSAL PWY NAD: salvage pathway I	0.132	0.889	0.367	0.132	**0.022**	0.367
PYRIDOXSYN-PWY: pyridoxal 5'-phosphate biosynthesis I	0.738	0.356	0.356	0.754	0.356	**0.039**
RHAMCAT-PWY: L- rhamnose degradation I	**0.018**	0.727	**0.032**	**0.018**	0.244	**0.025**
RIBOSYN2-PWY: flavin biosynthesis I (bacteria and plants)	0.943	**0.008**	0.943	0.13	0.943	**0.008**
TCA: TCA cycle I prokaryotic	0.055	0.416	**0.005**	**0.016**	0.776	**0.002**
TEICHOICACID-PWY: teichoic acid poly glycerol biosynthesis	0.068	NA	NA	0.087	**0.015**	NA
THISYN-PWY: superpathway of thiamin diphosphate	**0.009**	0.1	**0.007**	0.179	0.299	0.348

biosynthesis I						
TRNA CHARGING PWY: tRNA charging	**0.005**	0.603	**0.004**	**0.007**	0.303	**0.01**
X1CMET2-PWY: N10-formyl tetrahydrofolate biosynthesis	0.313	**0.049**	0.61	0.728	0.374	**0.049**
X3 HYDROXYPHENYLACETATE DEGRADATION PWY: 4-hydroxyphenylacetate degradation	0.477	0.289	0.117	0.289	NA	**0.005**

Significantly different pathways, as established from HUMAnN2, between VO_2 max classification groupings (fair (N=15), good (N=44), excellent (N=60), and superior (N=118)). VO_2 max classification groupings were established based on those established by Heyward and The Cooper Institute for Aerobics Research [17, 18] (table 6.2). Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. Significant differences are highlighted in bold. NA indicates it was not possible in this instance to determine a p value.

Supplementary table 6.9| Pathways significantly different between training frequency groupings

Pathway	Low vs. Medium	Low vs. High	Low vs. Extreme	Medium vs. High	Medium vs. Extreme	High vs. Extreme
7 ALPHADEHYDROX PWY: cholate degradation(bacteria;anaerobic)	0.098	0.28	**0.021**	NA	**<0.001**	**0.021**
AEROBACTINSYN-PWY: aerobactin biosynthesis	**0.021**	0.85	0.554	**0.021**	0.489	0.554
ANAEROFRUCAT-PWY: homolactic fermentation	0.318	0.318	0.108	0.807	**0.042**	**0.042**
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	0.067	**0.002**	0.968	0.09	0.196	**0.019**
AST-PWY: L-arginine degradation II (AST pathway)	**0.02**	**0.045**	0.453	0.529	0.325	0.264
BRANCHED CHAIN AA SYN PWY: superpathway of branched amino acid biosynthesis	0.721	**0.005**	0.319	**0.005**	0.229	**0.001**
CENTFERM-PWY: pyruvate fermentation to butanoate	0.471	**0.042**	0.839	0.068	0.471	**0.042**
CITRULBIO-PWY: L-citrulline biosynthesis	**0.042**	0.131	0.131	**0.008**	**0.014**	0.696
COA-PWY: coenzyme A biosynthesis I	0.551	**0.023**	0.458	0.104	0.36	0.104
CRNFORCAT-PWY: creatinine degradation I	**0.007**	**0.003**	**0.002**	0.346	0.077	0.311
ENTBACSYN-PWY: enterobactin biosynthesis	**0.005**	0.08	0.513	0.928	0.139	0.228
FOLSYN-PWY: superpathway of tetrahydrofolate biosynthesis and salvage	**0.003**	**0.002**	0.341	0.341	0.341	0.194
GLCMANNANAUT-PWY: superpathway of N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminate degradation	**0.048**	0.495	0.495	0.531	0.627	0.922
GLUCONEO-PWY: gluconeogenesis I	0.299	0.198	**0.026**	0.052	**0.003**	0.195

GLUCUROCAT-PWY: superpathway of beta D-glucuronide and D-glucuronate degradation	**0.033**	0.31	0.547	0.547	0.547	0.547
GLUDEG I PWY: GABA shunt	**0.038**	0.254	0.254	0.477	0.703	0.703
GLYCOL GLYOXDEG PWY: superpathway of glycol metabolism and degradation	**0.048**	0.484	0.336	0.539	0.977	0.603
GLYCOLYSIS E-D: superpathway of glycolysis and Entner Doudoroff	0.255	0.706	0.06	0.255	**0.027**	0.255
GLYCOLYSIS: glycolysis I from glucose 6 phosphate	0.387	0.387	**0.046**	0.71	**0.022**	**0.022**
GOLPDLCAT-PWY: superpathway of glycerol degradation to 1-3-propanediol	0.15	0.237	**0.01**	0.863	0.055	0.091
HEMESYN2-PWY: heme biosynthesis II anaerobic	**0.005**	**0.015**	0.402	0.402	**0.004**	**0.012**
KETOGLUCONMET-PWY: ketogluconate metabolism	**0.003**	0.988	0.198	**0.003**	0.8	0.198
METHGLYUT-PWY: superpathway of methylglyoxal degradation	**0.023**	0.079	0.196	0.911	0.976	0.911
NONOXIPENT-PWY: pentose phosphate pathway non-oxidative branch	0.901	**0.019**	0.494	**0.019**	0.644	0.457
P162-PWY: L-glutamate degradation V via hydroxyglutarate	0.631	0.077	0.362	**0.049**	0.453	**0.049**
P4-PWY: superpathway of L -lysine L-threonine and L-methionine biosynthesis I	0.167	0.167	**0.048**	0.629	0.245	0.445
P461-PWY: hexitol fermentation to lactate, formate, ethanol, and acetate	0.59	0.677	**0.03**	0.881	**0.008**	**0.026**
PANTO-PWY: phosphopantothenate biosynthesis I	**0.04**	**0.044**	0.136	0.765	0.765	0.795
PANTOSYN-PWY: pantothenate and coenzyme A biosynthesis I	0.177	**0.008**	0.582	0.143	0.284	0.099
PWY-1042: glycolysis IV plant cytosol	0.172	0.172	**0.003**	0.625	**0.028**	0.102
PWY-2723: trehalose degradation V	**0.009**	0.33	0.33	0.33	0.598	0.598

PWY-3001: superpathway of L- isoleucine biosynthesis I	0.557	**0.049**	0.158	0.074	0.074	**0.016**
PWY-4984: urea cycle	0.07	0.126	0.123	**0.011**	**0.011**	0.649
PWY-5030: L -histidine degradation III	**0.046**	0.557	0.325	0.549	0.938	0.557
PWY-5097: L- lysine biosynthesis VI	0.181	0.639	**0.007**	0.181	0.083	**0.007**
PWY-5101: L- isoleucine biosynthesis II	0.074	0.308	0.074	**0.031**	0.308	**0.031**
PWY-5103: L -isoleucine biosynthesis III	0.767	**0.002**	0.244	**0.002**	0.184	**<0.001**
PWY-5189: tetrapyrrole biosynthesis II from glycine	**0.001**	**0.002**	**<0.001**	0.451	**0.003**	**0.009**
PWY-5198: factor 420 biosynthesis	0.167	0.224	**0.032**	0.852	0.154	0.167
PWY-5415: catechol degradation I (meta-cleavage pathway)	0.403	NA	**0.026**	0.541	0.107	0.138
PWY-5464: superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase, and TCA cycle	0.377	0.121	0.304	**0.046**	0.454	0.068
PWY-5484: glycolysis II from fructose 6 phosphate	0.39	0.39	0.068	0.748	**0.03**	**0.03**
PWY-5920: superpathway of heme biosynthesis from glycine	**0.014**	0.559	0.426	0.087	0.426	0.559
PWY-5973: *cis* vaccenate biosynthesis	**0.009**	0.085	**0.025**	0.532	0.532	0.47
PWY-6147: 6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	0.103	**0.034**	**0.034**	0.173	0.103	0.733
PWY-6151: S-adenosyl-L-methionine cycle I	**0.042**	0.497	0.527	0.497	0.497	0.696
PWY-6507: 4-deoxy-L-threo-hex-4-enopyranuronate degradation	**0.035**	0.371	0.837	0.727	0.264	0.494
PWY-6590: superpathway of *Clostridium acetobutylicum* acidogenic fermentation	0.471	**0.042**	0.839	0.065	0.471	**0.042**
PWY-6595: superpathway of guanosine nucleotides degradation (plants)	0.175	**0.029**	0.052	0.175	0.237	0.948

PWY-6606: guanosine nucleotides degradation II	0.092	**0.017**	0.346	0.34	0.589	0.34
PWY-6612: superpathway of tetrahydrofolate biosynthesis	**0.004**	**0.001**	0.374	0.374	0.399	0.207
PWY-6630: superpathway of L-tyrosine biosynthesis	**0.001**	0.309	0.309	0.117	0.309	0.73
PWY-6700: queuosine biosynthesis	**0.019**	0.65	0.171	0.171	0.952	0.494
PWY-6859: all-*trans*-farnesol biosynthesis	**0.023**	**0.011**	**0.011**	0.329	0.315	0.924
PWY-6891: thiazole biosynthesis II (*Bacillus*)	**0.049**	0.425	0.618	0.618	0.425	0.618
PWY-6895: superpathway of thiamin diphosphate biosynthesis II	**0.048**	0.416	0.61	0.61	0.416	0.61
PWY-6901: superpathway of glucose and xylose degradation	0.261	0.246	0.14	**0.041**	**0.041**	0.708
PWY-6936: seleno amino acid biosynthesis	0.379	0.144	0.302	**0.048**	0.604	**0.048**
PWY-7013: L-1,2-propanediol degradation	**0.034**	0.668	0.14	**0.023**	**0.001**	0.381
PWY-7204: pyridoxal 5-phosphate salvage II (plants)	**0.043**	0.451	0.451	0.451	0.451	0.961
PWY-7234: inosine 5-phosphate biosynthesis III	0.187	0.187	**0.033**	0.506	0.187	0.298
PWY-7237: myo, chiro, and scillo inositol degradation	0.716	**0.04**	0.885	0.082	0.885	0.114
PWY-7392: taxadiene biosynthesis	0.069	**0.001**	**0.021**	0.096	0.22	0.661
PWY-7539: 6-hydroxymethyl-dihydropterin diphosphate biosynthesis III	0.275	0.159	**0.006**	0.387	**0.016**	0.11
PWY-7663: gondoate biosynthesis (anaerobic)	**0.002**	0.106	0.072	0.43	0.913	0.533
PWY0-1298: superpathway of pyrimidine deoxyribonucleosides degradation	0.734	0.85	**0.011**	0.85	**0.017**	**0.017**
PWY0-1338: polymyxin resistance	**0.032**	**0.034**	0.409	0.409	0.409	0.264
PWY0-1415: superpathway of heme biosynthesis from uroporphyrinogen III	**0.018**	0.239	0.237	0.639	0.834	0.639
PWY0-1479: tRNA processing	**0.001**	0.07	0.287	0.398	0.234	0.51

PWY0-162: superpathway of pyrimidine ribonucleotides de novo biosynthesis	0.488	0.508	0.069	0.288	0.138	**0.04**
PWY66-389: phytol degradation	0.114	**0.005**	NA	0.114	0.38	0.114
PWY66-399: gluconeogenesis III	0.122	0.519	0.065	0.065	**0.005**	0.147
PWY66-400: glycolysis VI (metazoan)	0.307	0.392	0.097	0.097	**0.027**	0.376
PYRIDNUCSAL-PWY: NAD salvage pathway I	0.584	**0.021**	0.718	**0.012**	0.584	0.168
RIBOSYN2-PWY: flavin biosynthesis I (bacteria and plants)	**0.003**	**0.003**	**0.044**	1	1	1
UBISYN-PWY: superpathway of ubiquinol 8 biosynthesis (prokaryotic)	**0.034**	**0.034**	0.245	0.269	0.976	0.269

Significantly different pathways, as established from HUMAnN2, between training frequency groupings (low (4-6 hours/week, N=89); medium (7-9 hours/week, N=91); high (10-12 hours/week, N=34), and extreme (13+ hours/week, N=23)). Training groups were based on self-reported, average hours of different activities per week. Statistical analysis was completed using the Kruskal-Wallis test, with a Mann Whitney U test performed for significant results to determine the groups between which this difference applied. All reported p values were adjusted using the Benjamini-Hochberg method. Significant differences are highlighted in bold.

Chapter 7

General Discussion

As highlighted in chapter 1, the human gastrointestinal tract (GIT) is home to a complex community of microorganisms, which interact with the host in a symbiotic relationship [1]. These individually unique populations participate in a wide variety of functions and produce several metabolites that interact with the human host. Such contributions include involvement with immune system functionality, food digestion, and short chain fatty acid (SCFA) and vitamin production [2, 3]. In parallel, the diversity and population composition, as well as the functional capacity, of the GIT has been associated with several diseases and disorders, with evidence supporting the influence of the gut microbiome on a wide variety of health outcomes [3, 4]. Furthermore, as discussed in chapter 2, the gut microbiome of athletes has a potential role in general athlete health, musculoskeletal health, and performance [5-7]. The gut microbiome, which begins colonization at birth, is further influenced by factors such as first foods, antibiotic or other drug use, age, and is considerably influenced by dietary intake [8-10]. Of particular relevance to this thesis, exercise has also been found to influence the gut microbiome, although the impact of exercise regimens has been variable in human studies [11-14].

While to date, evidence suggests that the athlete gut microbiome has been identified as differing from that of sedentary individuals, the range of factors driving the athlete gut microbiome have not been fully explored [15, 16]. In this thesis, we begin to unravel some of these factors including travel, type of sport, fitness, and training frequency by exploring the impact of these variables on the gut microbiome composition and functionality, as well as studying metabolites produced by both microbes and host among athlete cohorts.

Chapters 3, 4, and 6 explore some of the factors that may shape the gut microbiome composition and functionality in athlete cohorts. In chapter 3, we investigate the impact of travel on the gut microbiome of the Irish cricket teams using a combination of 16S rRNA and shotgun metagenomic sequencing. Previously, these individuals had self-reported an

increased incidence of gastrointestinal distress during travel, a phenomenon that can potentially increase injury potential and reduce performance [17-20]. In this chapter, we found gut microbiome instability was individual specific and was more commonly associated with gastrointestinal distress prior to travel to India (albeit, due to small cohort size, not in a significant way). Additionally, we identified the presence of several strains of *Escherichia coli* that were only present in instances where gastrointestinal distress was experienced before travel to India (i.e. during travel to UAE); with these individuals also harbouring an increased antibiotic resistance and virulence profile post-India. These findings highlight potential risks that athletes are exposed to during travel. Some of those taxa acquired during travel potentially placed these athletes at risk of developing post-infectious IBS [21, 22]. Additionally, the acquisition of antibiotic resistance during travel to India may persist within these individuals over time, with the potential to acquire and harbour an antibiotic resistant infection increased [23]. Although beyond the scope of this thesis, future investigations may benefit from the inclusion of a larger number of participants and the inclusion of an intervention, such as a specific probiotic, known to have potential with respect to preventing travellers' diarrhoea [24], to counteract the undesirable microbiome changes observed in this study.

In chapter 4, we characterise the gut microbiome and metabolome across several sports groups to determine the variance between sports classification groupings (SCGs) using shotgun metagenomic sequencing, GC-MS, UPLC-MS, and ^{1}H-NMR. SCGs were used in this instance to classify sports by their static and dynamic components [25], which was found to separate the gut microbiome (both compositionally and functionally) and metabolome (both urine and faecal). No significant dietary variation was established between different types of sports. The absence of this potential confounding factor was important given that diet has been suggested to be a major driver of microbiome differences between athletes and sedentary controls [15, 16]. Variation in species present and functionality, as identified

here, may be as a consequence of the variable demands of the different sports, including duration of the activity and training involved. Additionally, those functions executed by the gut microbiome and metabolites created, may confer a beneficial impact for those athletes within a specific SCG. Further expansion of the numbers per sport and sports included in this type of analysis should be the focus of future studies to determine if these patterns are observed in other sports from within the same groupings and across a wide range of individuals, as well as in those sports classifications not studied in this thesis. It will also be important to expand this work to athletes from other countries.

In chapter 6 we revealed an association between both VO_2 max and training frequency with gut microbiome composition and functionality in an active cohort, using shotgun metagenomic sequencing. In this study, many taxa were found to be associated with either VO_2 max or training frequency, and so only those taxa which had previously been identified as associated with exercise or athlete cohorts, as well as several SCFA producers, were investigated. Several intriguing patterns were observed when identifying the significantly different species and pathways based on VO_2 max or training frequency. In particular, an observed increase or decrease in particular taxa in samples collected from individuals within the lowest VO_2 max or training frequency group, followed by a subsequent trend in the opposite direction towards the highest VO_2 max or training frequency group. This is of interest as it suggests the variation of the impact of partially increased fitness or training frequency from the more extreme end of the scale. Furthermore, those pathways identified as variable between groups are generally related to energy metabolism, highlighting potential exploitation of these functions which may be utilised by certain VO_2 max or training frequency groupings. These patterns potentially warrant further investigation to determine potential intestinal changes occurring within individuals of certain VO_2 max or training frequency groups, which may drive the alterations observed in the gut microbiome here.

Chapter 5 builds on previous work that established significant differences in the gut microbiome of athletes from low and high BMI controls [15, 16]. Here, we focused on metabolites of interest, while also investigating the bile salt hydrolysing capacity of a functional metagenomic bank generated from athlete faecal metagenomic DNA. We decided to focus primarily on the host produced and microbially modified bile acids, as well as fatty acids which are predominantly derived from dietary sources. We identified the bile and fatty acid profiles of the athletes as being significantly different from that of high and low BMI controls. In particular, lower concentrations of tauro-conjugated bile acids were identified in the athlete samples, potentially indicating microbial modifications to reduce bile acid toxicity in the GIT. Bile acids that are agonists for liver X receptor (LXR) and G-protein-coupled bile acid receptor (TGR5), which are both associated with inflammatory responses, were found to be lower in the athletes with respect to the controls, while the inflammatory cytokines of these athletes were previously identified as lower with respect to the controls [16]. As noted, we also investigated the bile acid metabolizing capacity of the athlete gut microbiome using a metagenomic bank (functional fosmid library containing metagenomic DNA from the faecal samples of 40 athletes). The benefit of utilising this type of screen is that it highlights the functional capabilities of these athlete samples without the need for culturing individual species. Through this analysis we identified the presence of 5 distinct clusters of bile acid hydrolysing capabilities from among the samples, which is similar to recent findings of 7 phylotypes worldwide based on deconjugation capacity [26]. The analysis of these clusters from new cohorts, to assess how their presence corresponds with health and disease, would be of interest for future studies.

Overall, in this thesis I have added to the wealth of knowledge surrounding the factors driving the gut microbiome composition and functionality in athlete cohorts. In particular, we have found travel, sports grouping, VO_2 max and training frequency are all associated with differences in the gut microbiome of athletes. This has been found both from a

compositional and functional perspective, supported by the analysis of the faecal and urine metabolomes of these athletes. Overall, those differences identified may impact on the athlete cohorts investigated in a variety of mechanisms, although investigation of this was beyond the scope of this thesis. Many of those variations identified, particularly in functionality, may confer specific benefits within the athlete cohorts investigated, or be exploited by these cohorts. There are many potential avenues for the future of this research to further appreciate and understand the related changes driving these gut microbiome differences, which will be of benefit to athletes, aspiring athletes, and the general public.

7.1 References

1. 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., and Gordon, J.I., *Evolution of mammals and their gut microbes.* Science, 2008. **320**(5883): p. 1647-1651.

2. Wu, H.J. and Wu, E., *The role of gut microbiota in immune homeostasis and autoimmunity.* Gut Microbes, 2012. **3**(1): p. 4-14.

3. D'Argenio, V. and Salvatore, F., *The role of the gut microbiome in the healthy adult status.* Clinica Chimica Acta, 2015. **451**(Pt A): p. 97-102.

4. Laursen, M.F., Bahl, M.I., Michaelsen, K.F., and Licht, T.R., *First foods and gut microbes.* Frontiers in Microbiology, 2017. **8**: p. 356.

5. Bermon, S., Petriz, B., Kajeniene, A., Prestes, J., Castell, L., and Franco, O.L., *The microbiota: An exercise immunology perspective.* Exercise Immunology Review, 2015. **21**: p. 70-79.

6. Ohlsson, C. and Sjogren, K., *Effects of the gut microbiota on bone mass.* Trends in Endocrinology & Metabolism, 2015. **26**(2): p. 69-74.

7. Scheiman, J., Luber, J.M., Chavkin, T.A., MacDonald, T., Tung, A., Pham, L.D., Wibowo, M.C., Wurth, R.C., Punthambaker, S., Tierney, B.T., Yang, Z., Hattab, M.W., Avila-Pacheco, J., Clish, C.B., Lessard, S., Church, G.M., and Kostic, A.D., *Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism.* Nature Medicine, 2019. **25**(7): p. 1104-1109.

8. Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., Marchesi, J.R., and Collado, M.C., *The composition of the gut microbiota throughout life, with an emphasis on early life.* Microbial Ecology in Health and Disease, 2015. **26**: p. 26050.

9. Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M.T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y.S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Masen, L., Kristiansen, K., Dahlgren, J., and Wang, J., *Dynamics and stabilization of the human gut microbiome during the first year of life.* Cell Host & Microbe, 2015. **17**: p. 690-703.

10. Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.s.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., and O'Toole, P.W., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**: p. 178-185.

11. Bressa, C., Bailén-Andrino, M., Pérez-Santiago, J., Gonzalez-Soltero, R., Pérez, M., Montalvo-Lomichar, M.G., Maté-Muñoz, J.L., Domínguez, R., Moreno, D., and Larrosa, M., *Differences in gut microbiota profiles between women with active lifestyle and sedentary women.* PLoS ONE, 2017. **12**(2).

12. Cronin, O., Barton, W., Skuse, P., Penney, N.C., Garcia-Perez, I., Murphy, E.F., Woods, T., Nugent, H., Fanning, A., Melgar, S., Falvey, E.C., Holmes, E., Cotter, P.D., O'Sullivan, O., Molloy, M.G., and Shanahan, F., *A prospective metagenomic and metabolomic analysis of the impact of exercise and/or whey protein supplementation on the gut microbiome of sedentary adults.* mSystems, 2018. **3**(3).

13. Allen, J.M., Mailing, L.J., Niemiro, G.M., Moore, R., Cook, M.D., White, B.A., Holscher, H.D., and Woods, J.A., *Exercise alters gut microbiota composition and function in lean and obese humans.* Medicine & Science in Sports & Exercise, 2018. **50**(4): p. 747-757.

14. Zhao, X., Zhang, Z., Hu, B., Huang, W., Yuan, C., and Zou, L., *Response of gut microbiota to metabolite changes induced by endurance exercise.* Frontiers in Microbiology, 2018. **9**: p. 765.

15. Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., Shanahan, F., Cotter, P.D., and O' Sullivan, O., *The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level.* Gut, 2017. **67**(4): p. 625-633.

16. 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., and Cotter, P.D., *Exercise and associated dietary extremes impact on gut microbial diversity.* Gut, 2014. **63**: p. 1913-1920.

17. Chapman, D.W., Bullock, N., Ross, A., Rosemond, D., and Martin, D.T., *Detrimental effects of west to east transmeridian flight on jump performance.* European Journal of Applied Physiology, 2012. **112**(5): p. 1663-1669.

18. McGuckin, T.A., Sinclair, W.H., Sealey, R.M., and Bowman, P., *The effects of air travel on performance measures of elite Australian rugby league players.* European Journal of Sport Science, 2014. **14 Suppl 1**: p. S116-122.

19. Bullock, N., Martin, D.T., Ross, A., Rosemond, D., and Marino, F.E., *Effect of long haul travel on maximal sprint performance and diurnal variations in elite skeleton athletes.* British Journal of Sports Medicine, 2007. **41**(9): p. 569-573.

20. Huyghe, T., Scanlan, A.T., Dalbo, V.J., and Calleja-Gonzalez, J., *The negative influence of air travel on health and performance in the National Basketball Association: A narrative review.* Sports, 2018. **6**(3).

21. Tap, J., Derrien, M., Tornblom, H., Brazeilles, R., Cools-Portier, S., Dore, J., Storsrud, S., Le Neve, B., Ohman, L., and Simren, M., *Identification of an intestinal microbiota signature associated with severity of irritable bowel syndrome.* Gastroenterology, 2017. **152**(1): p. 111-123 e118.

22. Thabane, M. and Marshall, J.K., *Post-infectious irritable bowel syndrome.* World Journal of Gastroenterology, 2009. **15**(29): p. 3591-3596.

23. Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K., *Long-term impacts of antibiotic exposure on the human intestinal microbiota.* Microbiology, 2010. **156**(Pt 11): p. 3216-3223.

24. Reilly, T., Waterhouse, J., Burke, L.M., Alonso, J.M., and International Association of Athletics, F., *Nutrition for travel.* Journal of Sports Sciences, 2007. **25 Suppl 1**: p. S125-134.

25. Mitchell, J.H., Haskell, W., Snell, P., and Van Camp, S.P., *Task force 8: Classification of sports.* Journal of the American College of Cardiology, 2005. **45**(8): p. 1364-1367.

26. Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., Kalavagunta, P.K., Liao, J., Jin, L., Shang, J., and Li, J., *Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome.* Microbiome, 2019. **7**(1): p. 9.

Acknowledgements

The completion of this thesis would not have been possible without the support, guidance, friendship, and help of all the wonderful individuals I have had the pleasure of meeting along the way. I will probably forget some individual names here but every person I have crossed paths with during this experience have been helpful in ways they do not understand.

To my supervisory team, Prof. Paul Cotter, Dr. Orla O' Sullivan, Dr. Susan Joyce, and Dr. Cormac Gahan, thank you for accepting me as a PhD student and your constant encouragement throughout, it has been instrumental in me reaching this point. Paul- I have really appreciated your open-door policy and willingness to help in all situations. Your positive attitude when things don't exactly work out as hoped has been incredibly grounding during this experience and your attitude towards, and interest in, research is truly infectious. Orla- I have appreciated your guidance, patience, and assistance with all things bioinformatics, this PhD is truly founded on all of the skills I was lucky enough to learn from you. I also wish to thank you for exposing me to the world of education and outreach, a pursuit which has brought me great joy during my PhD, it is a passion that I hope to be able to continue working on. Susan – I will forever be grateful for the kindness and openness you have shown me whenever I have visited UCC. Being able to work on the metabolomics analysis for this thesis under your guidance was a wonderful experience. Cormac- I thank you for your advice, support, and continual encouragement throughout my PhD.

A special mention to Dr. Caitriona Guinane and Dr. Fiona Fouhy. Caitriona- you inspired my love of research when I had the fortune of completing a placement in Teagasc during my undergraduate degree. I am thankful to you for giving me that opportunity and letting me have the reigns with that project, which started my love for research, without which I know

this PhD would not have come to fruition. Fiona - you were the first person to suggest to me that I pursue a PhD, for that, as well as all the tips, chats, and laughs along the way, I am incredibly grateful.

Thank you to all of those who were involved with sample collection across the studies which formed this thesis, in particular to those in the Sport Ireland Institute and Cricket Ireland. A special thank you to Dr. Sharon Madigan for your enthusiasm on these projects and hard graft to get samples collected. Additionally, a massive thanks to all those who donated samples to any of these studies, without whom this work would not have been possible.

To all of the members of Vision 1, both past and present, this PhD would not have been possible without each of you, from your guidance in the lab and with practical bioinformatic advice and assistance, to the chats about nonsense to offer a distraction, and the copious amount of coffee (and occasional gin), I will forever be grateful for all of it, and thankful I have had this opportunity not only to grow as a researcher but also gain some true friends for life. In particular to Cathy, Chloe, and Laura, thanks for the all laughs and for just being all around good eggs! To Aaron, Calum, and Wiley, I have really appreciated your bioinformatics assistance and your patience with me, even when I have had to ask a question for the millionth time. To anyone who graced the student office for a long or short time, it has been a great pleasure to work on Master's and PhD's alongside such wonderful individuals.

To Aoife, having someone to go through the PhD process with has been invaluable. Thank you for answering all my questions (even when they were repetitious), for being a soundboard to bounce ideas and thoughts off, especially in these past few months, and not rolling your eyes too much at me, it's been a pleasure. I can't quite believe we have got

here in what seems like a blink of an eye. Onwards and upwards to the next chapter (or down the hall at least!).

To all those in lab 4.28 (and also 2.04) in the BioSciences Institute, UCC, thank you all for opening your lab to me on the occasions when I would show up. For your help with mass spec questions and always making sure I had a cup of coffee. In particular to Alvaro and Sarah, for your help with methods and running the mass spec I will forever be grateful.

To all my other friends, I thank you for your patience over the last few years where I maybe haven't been the easiest to meet up with, your well wishes, and catch up dinners when life allowed.

To my family, thanks for always believing in me. Mam- thank you for always taking an interest in everything I have had to update you on, for your kind words of encouragement in times when I struggled, and sorry I haven't been the easiest to be around over these last few months. Dad- thank you for always believing in me and your continued interest in my work. To my siblings, Barry and Niamh, and my sister- and brother-in-law, Paula and Alan, your continuous support and kindness have been instrumental in my PhD process, thank you for everything you have done for me as the youngest sibling, I owe where I am now in no small part to you. To my nephews (Ethan and Luca) and my niece (Chloé), thanks for always making me smile, maybe we can celebrate with some pannacotta. To my nana Ann, I am incredibly grateful for your unwavering belief that I could get this far, candle lighting, and constant encouragement in my education. To Aaron's family, thank you for all your support, encouragement, and kindness always, but especially as I have been working on this thesis, and also pushing me to ensure I actually submitted!

Finally, to Aaron, thank you for standing by my side throughout this whole experience; for listening to my PhD troubles, being there when I was excited when something worked,

celebrating the big and the small things, and supporting me in every way you could. I really wouldn't have gotten through this without you my love.

To all those I have lost along the way, I miss you all dearly and dedicate this thesis to you.