Investigating gut-microbiota mediated polymerisation of deoxycholic acid in humans

A thesis submitted for the degree of Doctor of Philosophy from

Imperial College London

Adesola Temitope Bello

2020

Section of Nutrition Research
Division of Digestive Diseases
Department of Metabolism, Digestion and Reproduction
Imperial College London

ABSTRACT

Cardiovascular diseases (CVD) continue to be the leading cause of death worldwide. Hypercholesterolemia (HC) is a modifiable risk factor in the development of CVD. Diet-based therapies, such as dietary fibres (DF) are first-line treatment options to be used solely or in conjunction with pharmacological options for cholesterol management. DF have been shown to reduce HC, but underlying mechanisms are unknown. Existing literature suggest that the hypocholesterolemic effect may be as a result of an interplay of diet, host metabolism, and gut microbiota (GM) which leads to the increased faecal excretion of bile acids (BA). Faecal BA excretion (FBAE) leads to a reduction of circulating cholesterol via a cascade of reactions, therefore increasing FBAE may be a cardioprotective mechanism.

A novel mechanism was thus hypothesised. DF may modulate GM to promote the polymerisation of secondary BA deoxycholic acid (DCA) to polydeoxycholic acid (PDC). PDC cannot be reabsorbed; it is subsequently excreted, triggering compensatory reactions that lead to reduced plasma cholesterol and its associated CVD risk.

This PhD project explored the mechanisms involved in and the effects of diet on PDC production in humans using a combination of cutting edge metabolomic and metagenomic tools.

An ultra-performance liquid chromatography-based method was developed for high throughput and sensitive detection and quantification of PDC. PDC was reliably detected in human faeces (up to 90% of total faecal DCA) and was found to correlate with anti-inflammatory butyrate producing GM and established beneficial GM like *Bifidobacteria*, and also with a high fibre high protein diet. More investigations are warranted to further elucidate the role of PDC in host physiology. Current indications suggest a beneficial role for PDC.

Abbreviations: CVD cardiovascular diseases; HC Hypercholesterolemia; DF dietary fibres; BA bile acids; GM gut microbiota, FBAE faecal bile acid excretion, DCA deoxycholic acid; PDC polydeoxycholic acid.

COPYRIGHT STATEMENT

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

DECLARATION OF CONTRIBUTIONS

The majority of the work presented in this thesis, and the research to which it refers, was performed by the author. All ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

The author would like to mention the following contributions:

Chapter 4: Dean and Stark distillation experiments were carried out in the Department of Chemistry General Labs, Imperial College London under the supervision of the lab managers and Ms Olga Kuzmina. MALDI-MS (including lab processes and chromatogram generation) was carried out as a paid-for service by the Chemistry Mass Spectrometry facility by Dr Lisa Haigh. Dr Alex Pechlivanis developed the “OLIG” LC-MS method for detecting PDC oligomers. NMR experiments were carried out under the supervision and with the assistance of Dr Fahmina Fardus-Reid.

Chapter 3-5: Samples were obtained from a cross-sectional cohort study run and managed by collaborators at the University of Copenhagen, Denmark led by Prof Tue Hansen. Metagenomics analysis was carried out by a team of collaborators from MetaGenoPolis (INRA), France namely Florence Thirion, Emmanuelle Le Chatelier, Nicolas Pons, Florian Plaza-Oñate, Florence Levenez, Nathalie Galleron, Benoît Quinquis, and Dusko S. Ehrlich (list as provided by Ms Florence Thirion). Urine and serum samples were analysed for bile acids by Ms Jerusa Brignardello, Department of Metabolism, Digestion and Reproduction, Imperial College London. All the data from this study generated by persons other than author and used in these chapters were obtained with permission.

Chapter 6: Samples were obtained from a randomised controlled trial run and managed by a research team led by Dr Aikaterini Petropoulou, Department of Metabolism, Digestion and Reproduction, Imperial College London. Data from this study used in this thesis were obtained with permission.

Chapter 7: Oat β-glucan substrate was provided on request as a free sample by Garuda International Inc. (Exeter, California, USA). Barley grains were obtained from Hodmedod’s (Suffolk, UK). Inulin and cellulose substrates were provided by Dr Claire Byrne. *In vitro* batch cultures were performed at the University of Reading under the supervision and assistance of Dr Gemma Walton.

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Prof Gary Frost and Dr Jonathan Swann, for giving me the opportunity to undertake scientific research under their guidance. Thank you for having afforded me your invaluable wisdom, patience and support over the course of my PhD, as well as for your time and efforts invested in reviewing this thesis.

A big thank you to Profs Elaine Holmes and Julian Griffin for giving me the opportunity to work with your team at the UK Dementia Research Institute at Imperial College. Thank you for being very understanding especially during the last months of writing up my thesis. Also, thank you to Parisa, Manuja and Luke for your help and support.

I would like to appreciate the Mass Spectrometry team of the (then) Computational and Systems Medicine (CSM), Faculty of Medicine for their support. I am particularly indebted to Drs Leanne Nye, Nyasha Munjoma and Verena Horneffer-van der Sluis for their help with day-to-day challenges with the mass spectrometer in the lab. Special thanks to Prof Ian Wilson for his advice on liquid chromatography techniques.

I would also like to thank the members of the Gary Frost and Jon Swann labs, and really all the PhD students in the “old CSM”. I am most indebted to Jerusa, my other bile acids comrade 😊. Thank you for your friendship and for making this PhD journey brighter and feel smoother, especially when we had to work late into the night, which was a lot of times.

To my London Church (West London) family, although God brought us together in the last lap of this PhD marathon, it was exactly what I needed to finish the race. Thank you for your prayers, support and love.

I cannot but express my gratitude to my parents and siblings, who despite not understanding an iota of what my research was about, continually gave me morale support and the will to continue when the times got tough. A worthy mention is my sister, Ibiyemi, who can maybe now stop asking me about the autoclave… Thank you for being in my corner always!

To everyone I may have missed due to the pressure of time, I will personally sing you a thank you song (brace yourself)!

Finally, it sure has been an ardous journey that at some point was overwhelming, but all thanks to God Almighty Who makes the impossible possible and gave me the strength to get to the finishing line.

TABLE OF CONTENTS

ABSTRACT 2
COPYRIGHT STATEMENT 3
DECLARATION OF CONTRIBUTIONS 4
ACKNOWLEDGEMENTS 5
TABLE OF CONTENTS 6
LIST OF FIGURES 13
LIST OF TABLES 16
ABBREVIATIONS 18
1. GENERAL INTRODUCTION 23
1.1 The Burden of Hypercholesterolemia and Cardiovascular Diseases 23
1.2 Cholesterol metabolism 24
1.3 The Role of Diet in Cholesterol Management 26
1.4 Bile acid metabolism 28
1.4.1 Formation of BA 28
1.4.2 Physiological Functions of BA 30
1.4.3 Enterohepatic circulation of BA 32
1.4.4 Gut microbial modulation of bile acids 33
1.5 Bile Acid Excretion by Dietary Soluble Fibres 37
1.5.1 Hypothesis 40
1.6 Thesis aims and objectives 40
1.6.1 Main Aim 40
1.6.2 Objectives of this PhD 40
1.6.3 Justification for this PhD project 40
1.7 Conclusion 41
2. ANALYTICAL STRATEGIES 43
2.1 Introduction 43
2.2 Aim and Objectives 44

2.2.1 Aim 44
2.2.2 Objectives 44
2.3 Sample Collection Techniques 45
2.3.1 Sample selection, collection and handling 45
2.3.2 Study size estimation 47
2.3.3 Cross-sectional cohort study 48
2.3.4 Randomised controlled trial 50
2.3.5 *In vitro* sample collection strategies 51
2.4 Sample analysis 52
2.4.1 Metabolite analysis 52
2.4.2 Microbiome Techniques 59
2.5 Data and statistical analysis 63
2.5.1 Analysis of metabolite data 63
2.5.2 Analysis of microbiome data 64
2.5.3 Integration of metabolite and microbiome data 65
2.6 Conclusion 65
3. QUANTIFYING POLYDEOXYCHOLIC ACID: VALIDATION OF AN ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY MASS SPECTROMETRY METHOD FOR BILE ACID ANALYSIS IN FAECAL SAMPLES 66
3.1 Introduction 66
3.2 Aim and objectives 68
3.2.1 Aim 68
3.2.2 Objectives 68
3.3 Methods 68
3.3.1 Chemicals and reagents 68
3.3.2 Preparation of calibration curve and quality control concentrations 69
3.3.3 Preparation of stock and working solutions for validation experiments 70
3.3.4 Linearity 70
3.3.5 Intra-day accuracy and precision 70
3.3.6 Analyte Recovery 70

3.3.7 Matrix Factor Analysis 71
3.3.8 Method Selectivity 71
3.3.9 Collection of clinical faecal samples from Vegans vs Omnivores 71
3.3.10 Preparation of faecal samples for analysis 72
3.3.11 UPLC-MS conditions 72
3.3.12 MS Data Analyses 73
3.4 Results 78
3.4.1 Optimisation of injection volume 78
3.4.2 Analytical range and linearity: 78
3.4.3 Intra-day accuracy and precision 78
3.4.4 Analyte Selectivity of method 78
3.4.5 DIS recovery and matrix factor 79
3.4.6 Application to clinical samples 83
3.5 Discussion 91
3.5.1 Summary of findings 91
3.5.2 General discussion 91
3.5.3 Strengths and Limitations of the study 95
3.6 Conclusion 97
4. QUANTIFYING POLYDEOXYCHOLIC ACID: DEVELOPMENT OF A HIGH THROUGPUT WORKFLOW FOR QUANTIFYING POLYDEOXYCHOLIC ACID IN HUMAN FAECAL SAMPLES 98
4.1 Introduction 98
4.2 Aim and Objectives 101
4.2.1 Aim 101
4.2.2 Objectives 101
4.3 Methods 102
4.3.1 Chemicals and reagents 102
4.3.2 Chemical synthesis of PDC standards for method development 102
4.3.3 UPLC-MS Experimental Conditions 104
4.3.4 Proton-NMR Experimental Conditions 106

4.3.5 MALDI-MS Experimental Conditions 106
4.3.6 Preparation of working solutions 106
4.3.7 Collection of clinical faecal samples from Vegans vs Omnivores Study 106
4.3.8 Preparation of stool samples 107
4.3.9 MS Data Analyses 108
4.4 Results and Discussion 109
4.4.1 Synthesis of PDC oligomers 109
4.4.2 LC-MS detection of PDC oligomers: 109
4.4.3 NMR detection of PDC oligomers: 115
4.4.4 MALDI-MS detection of PDC oligomers 117
4.4.5 Optimisation of PDC extraction 119
4.4.6 DCA and PDC Recovery in Stool Samples 124
4.4.7 Validation of Hydrolysis of PDC Oligomers into DCA 124
4.4.8 Final Sample Preparation Workflow for Analysing PDC in Stool Samples 126
4.4.9 Application to clinical samples: PDC in vegans and omnivores 127
4.5 Discussion 129
4.5.1 Summary of results 129
4.5.2 General discussion 129
4.5.3 Strengths and Limitations of this work 135
4.6 Conclusion 140
5. EXPLORING POLYDEOXYCHOLIC ACID *IN VIVO* IN HEALTHY HUMAN ADULT COHORT OF VEGANS AND OMNIVORES 141
5.1. Introduction 141
5.2. Aim and Objectives 142
5.2.1 Aim 142
5.2.2 Objectives 142
5.3. Methods 143
5.3.1 Sample collection and analysis 143
5.3.2 Data and Statistical analyses 143

5.4. Results 144
5.4.1 Summary of Patient Characteristics 144
5.4.2 Investigating lower and upper quartiles of PDC 147
5.4.3 Investigating lower and upper quartiles of PDC% 150
5.4.4 Investigating relationship of PDC with cholesterol, diet and other biofluid-specific BAs 156
5.5. Discussion 167
5.5.1 Summary of findings 167
5.5.2 General discussion 168
5.5.3 Strengths and limitations 174
5.6. Conclusion 175
6. EXPLORING THE EFFECT OF AN INCREASED INTAKE OF LEGUMES ON POLYDEOXYCHOLIC ACID IN HEALTHY ADULT HUMANS 176
6.1 Introduction 176
6.2 Aim and Objectives 178
6.2.1 Aim 178
6.2.2 Objectives 178
6.3 Methods 179
6.3.1 Sample collection 179
6.3.2 Sample analysis 179
6.3.3 Data and Statistical analyses 181
6.4 Results 182
6.4.1 Participant Characteristics 182
6.4.2 DCA and PDC on both *rr* and *RR* interventions 184
6.4.3 Correlations between blood lipoproteins and PDC/PDC% 187
6.4.4 Correlations between gut microbiota and PDC/PDC% 187
6.5 Discussion 191
6.5.1 Summary of findings 191
6.5.2 General discussion 191
6.5.3 Strengths and Limitations of this study 196

6.6 Conclusion 198
7. *IN VITRO* INVESTIGATIONS OF THE COLONIC PRODUCTION OF POLYDEOXYCHOLIC ACID 199
7.1 Introduction 199
7.2 Aim and Objectives 200
7.2.1 Aim 200
7.2.2 Objectives 200
7.3 Materials and Methods 201
7.3.1 Raw materials for buffers, reagents and culture media 201
7.3.2 Substrates 201
7.3.3 Preparation of basal nutrient medium 201
7.3.4 Collection and preparation of faecal sample 203
7.3.5 Simulated *in vitro* upper gut digestion 203
7.3.6 Batch culture fermentations 204
7.3.7 PDC analysis 206
7.3.8 Data and Statistical analyses 206
7.4 Results 207
7.4.1 Investigating preferred precursors for PDC production *in vitro* (Experiment I) 207
7.4.2 Effect of fibres on PDC production *in vitro* (Experiment II) 210
7.5 Discussion 213
7.5.1 Summary of key findings 213
7.5.2 General discussion 213
7.5.3 Strengths and Limitations of this study 216
7.6 Conclusion 217
8. GENERAL DISCUSSION 218
8.1 Summary of work and future recommendations 218
9. REFERENCES 222
10. APPENDICES 266
Appendix 1 266

Appendix 2 266
Appendix 3 267
Appendix 4 271
Appendix 5 271
Vegan Study Methods For Sample Collection And Data Processing Obtained From Collaborators 271
Supplementary Figures and Results for Chapter 5 275
Appendix 6 286
Appendix 7 288
Appendix 8 288

LIST OF FIGURES

Figure 1-1: A simple outline of cholesterol metabolism in the body. 25
Figure 1-2: Structure of three most common soluble fibres 27
Figure 1-3: Simplified pathway of biosynthesis of primary bile acids from cholesterol in the liver. 29
Figure 1-4: Diagrammatic representation of the enterohepatic circulation of bile acids..... 33
Figure 1-5: Proposed structure for polydeoxycholic acid 39
Figure 1-6: The “Uninhabited” Niche this PhD aims to contribute to. 42
Figure 2-1: Hierarchy of evidence 49
Figure 2-2: A simplified overview of a mass spectrometry analysis pipeline. 53
Figure 2-3: A diagrammatic description of electrospray ionisation process..... 55
Figure 2-4: Schematic representation of MALDI-ToF system 56
Figure 2-5: A simplified illustration of the 16S rRNA gene showing conserved (orange) and hypervariable (grey) regions..... 61
Figure 2-6: Secondary structure of the 16S rRNA of Escherichia coli..... 61
Figure 3-1: A schematic diagram showing the principle of action for alkaline hydrolysis of polydeoxycholic acid 67
Figure 3-2: Average percentage recoveries of deuterated internal standards in test stool samples 81
Figure 4-1: Chromatogram of deoxycholic acid peak; m/z 391.3, RT 10.36mins on the EL method..... 100
Figure 4-2: A Dean-Stark set-up for the synthesis of polydeoxycholic acid standards. 103
Figure 4-3: Peak areas of deoxycholic acid (DCA) and polydeoxycholic acid dimer (PDC-dimer) over the course of the synthesis experiment. 110
Figure 4-4: Extracted mass chromatograms of DCA and PDC-dimer on the EL method. .. 111
Figure 4-5: Extracted mass chromatograms showing peaks (after smoothing using mean 3x2) of DCA and PDC oligomers from synthesised standard samples (t = 360mins) on the OLIG method..... 113
Figure 4-6: Peak areas of deoxycholic acid (DCA) and polydeoxycholic acid oligomers [-dimer (PDC2), -trimer(PDC3), -tetramer(PDC4) and -pentamer(PDC5)] detected on OLIG method. 114
Figure 4-7: 1D Proton NMR spectra of synthesized PDC standards..... 116
Figure 4-8: MALDI-MS chromatogram of synthesized PDC standards using 2,5-Dihydroxybenzoic acid as matrix. 117
Figure 4-9: MALDI-MS chromatogram of synthesized PDC standards using nitroaniline as matrix. 118

Figure 4-10: Mass chromatograms showing MRM transition peaks (after smoothing using mean 3x2) of the PDC-pentamer in faecal extracts. 121
Figure 4-11: Mass chromatograms showing MRM transition peaks (after peak smoothing using mean 3x2) of the PDC-tetramer in faecal extracts. 122
Figure 4-12: Mass chromatograms showing MRM transition peaks (after peak smoothing using mean 3x2) of the PDC-trimer in faecal extracts..... 123
Figure 4-13: Bar graph of mean DCA and PDC peak areas before (pre-) and after (post-) PDC extraction including alkaline hydrolysis step. 125
Figure 4-14: Boxplots showing percentage of total faecal deoxycholic acid (fDCA) in monomeric (DCA) and polymeric forms (PDC) in vegans vs omnivores. 128
Figure 4-15: Boxplots showing DCA concentrations obtained from EL validation experiments described in the Chapter 3 (“fBA23”) and modified extraction protocol from optimisation experiments described in the chapter 4 (“DCA”)..... 133
Figure 4-16: A dot plot showing the differences in DCA concentrations obtained from old extraction and modified protocols in vegans and omnivores..... 134
Figure 4-17: Correlation heatmap showing Spearman correlations between PDC, PDC%, DCA, DCA%, total faecal DCA (DCA + PDC) and total faecal BA in whole cohort..... 136
Figure 4-18: Scatterplot showing a strong correlation (Spearman rho, $\rho = 0.792$, $p_{raw} < 0.001$) between DCA and total faecal bile acid concentrations in vegans and omnivores..... 137
Figure 4-19: Scatterplot showing no correlation (Spearman rho, $\rho = 0.035$, $p_{raw} = 0.718$) between PDC and DCA concentrations in vegans and omnivores. 138
Figure 5-1: Boxplot showing the spread of concentrations of PDC concentrations (ng/mg dry stool) in vegans (n = 57) and omnivores(n = 49). 148
Figure 5-2: Boxplot showing PDC percentage of total faecal DCA output in vegans and omnivores. 151
Figure 5-3: Boxplots showing A)PDC B)DCA and C)total faecal DCA concentrations in low-%PDC and high-%PDC groups. 153
Figure 5-4: Boxplots showing total faecal bile acid concentrations between low %PDC and high %PDC groups. 154
Figure 5-5: Boxplots showing average nut and legume intakes A)with and B)without outliers in low-%PDC and high-%PDC groups..... 154
Figure 5-6: Boxplots showing fibre intakes per 1000kJ energy consumed A)with and B)without outliers in low-%PDC and high-%PDC groups. 155
Figure 5-7: Diagrammatic summary of Spearman correlations between PDC, DCA, total faecal DCA, PDC% and DCA% and dietary variables..... 157

Figure 5-8: Correlation heatmap showing only statistically significant Spearman correlations of an absolute value ≥0.1 between PDC, DCA, total faecal DCA, PDC% and DCA% and intakes of “minor” nutrients in entire cohort. 159
Figure 5-9: Correlation heatmaps and corresponding dendrograms showing Spearman correlations of an absolute value ≥0.1 between PDC, DCA, total faecal DCA, PDC% and DCA% and intakes of fibres in A) the whole cohort, B) vegans only and C) omnivores only. 161
Figure 5-10: Heatmap showing correlations of PDC, DCA and total faecal DCA concentrations and percentages with faecal bile acids in whole cohort (n=106). 163
Figure 5-11: Heatmap showing correlations between PDC, DCA and total faecal DCA concentrations and percentages with serum bile acids. 164
Figure 5-12: Heatmap showing correlations between PDC, DCA and total faecal DCA concentrations and percentages with urine bile acids. 165
Figure 5-13: Heatmap showing top 20 significant correlations between PDC, DCA and total faecal DCA concentrations and percentages with metagenomic species (MSP). 166
Figure 6-1: Study design of crossover randomised controlled trial. 180
Figure 6-2: PDC and PDC% at baseline and follow-up visits on both rr and RR interventions. 185
Figure 6-3: Line graphs of participants showing compliant and non-compliant participants on both interventions. 186
Figure 6-4: Spearman correlations between metagenomic species and each of PDC, PDC%, DCA, DCA% and total faecal DCA. 189
Figure 6-5: Spearman correlations of delta changes in A) rr and B) RR phases. 190
Figure 7-1: Bar plots showing average taurocholic acid (TCA), glycocholic acid (GCA), cholic acid (CA) and deoxycholic acid (DCA) percentages in different batch culture vessels throughout Experiment I. 208
Figure 7-2: Line graphs showing DCA concentrations of batch culture samples from Experiment I. No difference in DCA concentrations before (grey) and after (blue) alkaline hydrolysis indicates that no PDC was produced. 209
Figure 7-3: Summary charts of bile acid concentrations and proportions from Experiment II batch cultures. 211

LIST OF TABLES

Table 1-1: Table showing NHS recommended plasma cholesterol levels..... 24
Table 1-2: Gut bacterial genera that are involved in biotransformation of bile acids. 36
Table 3-1: Gradient for chromatographic separation for analysing bile acids using the existing lab "EL" method. 74
Table 3-2: Full experiment details of selected ion/reaction monitoring experiments for bile acids standards. 75
Table 3-3: Full experiment details of selected ion/reaction monitoring experiments for internal standards for bile acid analysis. 77
Table 3-4: List of bile acids with interference less than lower limit of quantification (LLOQ) response. 80
Table 3-5: Faecal BA concentrations in vegans vs omnivores. 84
Table 3-6: Percentage compositions of BA in vegans and omnivores. 85
Table 4-1: Relative accuracy and precision for detecting and quantifying deoxycholic acid (BA23) on the EL method 99
Table 4-2:DCA-d4 (deuterated internal standard for deoxycholic acid) recovery and matrix factor results from the validation of EL method. 99
Table 4-3: Adjusted gradient for chromatographic separation for analysing bile acids using the "EL" UPLC-MS based method. 105
Table 4-4: Predicted molecular weights for PDC oligomers up to the decamer (10 DCA monomer units). 109
Table 4-5:Gradient for chromatographic separation for analysing synthesised PDC oligomers using the "OLIG" UPLC-MS based method 112
Table 4-6: MRM transitions created for polydeoxycholic acid oligomers up to the pentamer for analysis on the OLIG method 120
Table 4-7: Mean % Recoveries (± standard deviation) of DCA-d4 from test faecal samples. 124
Table 4-8: DCA and PDC measures in vegans and omnivores. 127
Table 5-1: Demographic and clinical characteristics of study participants in the subset data (complete cases “CC” dataset) compared to the dataset including all participants (master “M” dataset) and excluded values (“NA” dataset). 145
Table 5-2: Demographic and clinical characteristics of vegans and omnivores in the subset data (n=106). 146
Table 5-3: Demographic, clinical and lab data for lower ("low PDC") and upper ("high PDC") quartile samples for PDC concentrations (ng per mg dry stool). 149

Table 5-4: Demographic, clinical and lab data for lower ("low PDC%") and upper ("high PDC%") quartile samples for percentage PDC proportions. 152
Table 6-1: Patient characteristics from before and after both RR and rr interventions 183
Table 7-1: Components of basal nutrient medium for batch culture experiments..... 202
Table 7-2: Schematic summary of the different sets of batch culture fermentations carried out in Experiment I investigations..... 205
Table 7-3: Schematic summary of the different sets of batch culture fermentations carried out in Experiment II investigations..... 205
Table 7-4: P-values of analysis of variance conducted on DCA concentrations (unsaponified samples) in Experiment I..... 210
Table 7-5: P-values of analysis of variance conducted on DCA concentrations (unsaponified samples) in Experiment II..... 212

ABBREVIATIONS

1D ^{1}H NMR	One-Dimensional Proton NMR
7KLCA	3α-Hydroxy-7-Ketolithocholic Acid
ACN	Acetonitrile
allo-LCA	Allolithocholic Acid
ANOVA	Analysis of Variance
APCI	Atmospheric Pressure Chemical Ionisation
ASBT	Apical Sodium co-dependent BA Transporter
ASV	Amplicon Sequence Variance
BA	Bile Acids
BBG	Barley B-Glucan
βMCA	β-Muricholic Acid
BEH	Ethylene Bridged Hybrid
BHF	British Heart Foundation
BSH	Bile Salt Hydrolase
CA	Cholic Acid
$CaCl_2$	Calcium Chloride
$CaCl_2.6H_2O$	Calcium Chloride Hexahydrate
CAR	Constitutive Androstane Receptor
CC	Calibration Curve
CC^1	Complete Cases
CDCA	Chenodeoxycholic Acid
$CDCl_3$	Deuterated Chloroform
CE	Collision Energy
$CHCl_3$	Chloroform
CHD	Coronary Heart Disease

CI	Chemical Ionisation
CVD	Cardiovascular Diseases
DADA2	Divisive Amplicon Denoising Algorithm 2
DC	Direct Current
DCA	Deoxycholic Acid
DIS	Deuterated Internal Standard
EDTA	Ethylenediaminetetraacetic Acid
EGA	Extreme Group Analysis
EHC	Enterohepatic Circulation
EI	Electron Ionisation
EL method	Existing Lab Method
ESI	Electrospray Ionisation
FDA	Food and Drug Administration
FDR	False Discovery Rate
FH	Familial Hypercholesterolemia
FXR	Farnesoid X Receptor
GC	Gas Chromatography
GCA	Glycocholic acid
GCDCA	Glycochenodeoxycholic acid
GM	Gut Microbiota/Microbiome
GUDCA	Glycoursodeoxycholic acid
HC	Hypercholesterolemia
HCl	Hydrochloric acid
HDL	High Density Lipoprotein
HMG-CoA reductase	3-Hydroxyl-3-Methylglutaryl CoA Reductase

HQC	High Quality Control
IBABP	Ileal Bile Acid Binding Protein
IPA	2-Propanol
iso-LCA	Isolithocholic acid
K_2HPO_4	Potassium phosphate dibasic
KH_2PO_4	Potassium phosphate monobasic
LC	Liquid Chromatography
LCA	Lithocholic acid
LDL	Low Density Lipoprotein
LLOQ	Lower Limit of Quantification
LQC	Low Quality Control
MALDI	Matrix Assisted Liquid Desorption/Ionisation
MeOH	Methanol
$MgSO_4.7H_2O$	Magnesium Sulphate Heptahydrate
MQC	Medium Quality Control
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
NaCl	Sodium Chloride
$NaHCO_3$	Sodium Bicarbonate
NaOH	Sodium Hydroxide
NCEP	National Cholesterol Education Program
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance
OBG	Oat β-Glucan
OLIG method	Oligomer Method

OSTα/β	Organic Solute Transporter
OTU	Operational Taxonomic Unit
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDC	Polydeoxycholic Acid
PTSA	p-Toluenesulfonic Acid
PXR	Pregnane X Receptor
QC	Quality Control
RCT	Randomised Controlled Trial
RFAC	Radio Frequency Alternating Current
RP-LC	Reversed Phase Liquid Chromatography
rRNA	Ribosomal Ribonucleic Acid
S1P2	Sphingosine-1-Phosphate Receptor 2
SD	Standard Deviation
SEM	Standard Error of Mean
SF	Soluble Fibres
SGS	Shotgun Sequencing
SIM	Selected Ion Monitoring
SPE	Solid Phase Extraction
SRM	Selected Reaction Monitoring
TCA	Taurocholic acid
TCDCA	Taurochenodeoxycholic acid
TGR5	G-Protein-Coupled Bile Acid Receptor
ToF	Time of Flight

TQ	Tandem Triple Quadrupole Analyser
TUDCA	Tauroursodeoxycholic Acid
Tukey HSD	Tukey Honest Significant Differences
UDCA	Ursodeoxycholic Acid
U(H)PLC	Ultra (High) Performance Liquid Chromatography
ULOQ	Upper Limit of Quantification
VDR	Vitamin D Receptor
VLDL	Very Low-Density Lipoprotein
WHO	World Health Organisation

1. GENERAL INTRODUCTION

1.1 The Burden of Hypercholesterolemia and Cardiovascular Diseases

Hypercholesterolemia (HC), defined as elevated fasting blood lipoproteins, is a well-documented modifiable and independent risk factor for the development of atherosclerotic cardiovascular diseases (CVD) (Furtado & Giugliano, 2020; Kreisberg & Oberman, 2002; Rose & Shipley, 1986). Atherosclerotic CVD, including ischemic/coronary heart disease (CHD), cerebrovascular disease and hypertension, remain the leading causes of death worldwide, despite the reduced prevalence (Benjamin et al., 2017; WHO, 2018). In the UK, CHD is the single leading cause of death being responsible for 170, 000 deaths per year. Furthermore, current estimates show that CVD cost the UK economy up to £19 billion each year when premature death, disability and informal costs are accounted for (BHF, 2020). According to WHO estimates, HC can be attributed to a third of global coronary heart disease (CHD) cases. Thus, maintaining normal blood cholesterol levels is unequivocally important to living a healthy life.

Atherosclerotic CVD make up 85% of all CVD cases (Mendis, Puska, & Norrving, 2011). Atherosclerosis is a complex pathological process characterised by thickening of blood vessel walls due to progressive plaque deposits resulting in reduced blood flow. When a plaque ruptures, it triggers the formation of a blood clot, blocking off blood flow completely and may lead to a heart attack if formed in a coronary artery, or a stroke if formed in the brain. Atherosclerosis develops as an overall effect of a number of risk factors including, but not limited to, elevated blood lipids, unhealthy diet, tobacco use, raised blood pressure and genetic disposition.

Evidence for an atherogenic role for cholesterol, particularly LDL, is obtained from Familial HC (FH). FH is an autosomal dominant inherited disorder associated with elevated levels of LDL-C and premature atherosclerotic CVD (M. S. Brown & Goldstein, 1986). It is caused by mutations in one of LDL-receptor, apolipoprotein-B or the PCSK9 genes (Defesche et al., 2017). People who are heterozygous for the mutations have cholesterol levels that are about twice the normal levels and are at a greater risk of developing CVD, while those who are homozygous develop atherosclerosis, CVD and may die very young (Linton, Yancey, Davies, & Al., 2000). Furthermore, recent large human clinical trials have shown that a reduction in LDL is associated with improved clinical outcomes (Baigent et al., 2010; Sabatine et al., 2017; Schwartz et al., 2018; M. G. Silverman et al., 2016).

1.2 Cholesterol metabolism

Cholesterol is a fundamental component of animal cell membranes; it modulates the fluidity and structure of the membrane, and serves as a precursor for all steroid hormones, bile acids (BA) and vitamin D (García, Uhía, & Galán, 2012). Cholesterol is synthesised *de novo* and can also be obtained from the diet as dietary cholesterol from foods of animal origin such as eggs, milk and meat. In a healthy individual, cholesterol synthesis, absorption and excretion are tightly regulated in response to the diet ensuring optimal amounts are present in the body at all times (Berg, Tymoczko, & Stryer, 2002). However, due to genetic factors and lifestyle choices, cholesterol homeostasis may be compromised.

The principal site of cholesterol synthesis is the liver, although it can be synthesised in the intestines, endocrine organs, skin and muscles (Berg et al., 2002). Cholesterol is synthesised in a series of enzyme-catalysed reactions from acetyl-CoA (Figure 1-1). The rate-limiting step of this series is catalysed by the enzyme 3-hydroxyl-3-methylglutaryl CoA reductase (HMG-CoA reductase).

As cholesterol is a lipid molecule, it is insoluble in plasma and is transported within four major lipoprotein molecules namely chylomicrons, low-density lipoproteins (LDL), very low-density lipoproteins (VLDL) and high-density lipoproteins (HDL). Lipoprotein particles are composed of lipids and proteins in varying proportions; they also differ in size, shape, function and roles they play in the pathogenesis of vascular diseases (Daniels, Killinger, Michal, Wright, & Jiang, 2009). LDL transports cholesterol from the liver to peripheral cells and organs in the body while HDL shuttles cholesterol from other cells back to the liver. LDL is the main carrier of cholesterol in the blood and high plasma levels of LDL are a major contributor to the initiation and progression of atherosclerotic plaques (Furtado & Giugliano, 2020). Hence, maintaining lower plasma LDL cholesterol is strongly advocated to prevent and manage CVD-related deaths and morbidity. NHS guidelines recommend that healthy adults should have no more than 5mmol/l total plasma cholesterol and 3 mmol/L of LDL cholesterol, and at least 1 mmol/L of HDL cholesterol (Table 1-1).

Table 1-1: Table showing NHS recommended plasma cholesterol levels

NHS Guidelines for blood cholesterol levels		
Lipoprotein	**Healthy Individuals**	**Individuals at risk***
Total cholesterol	<5 mmol/L	<4mmol/L
LDL cholesterol	<3 mmol/L	<2mmol/L
HDL cholesterol	>1 mmol/L	>1 mmol/L
***Presence of other risk factors**		

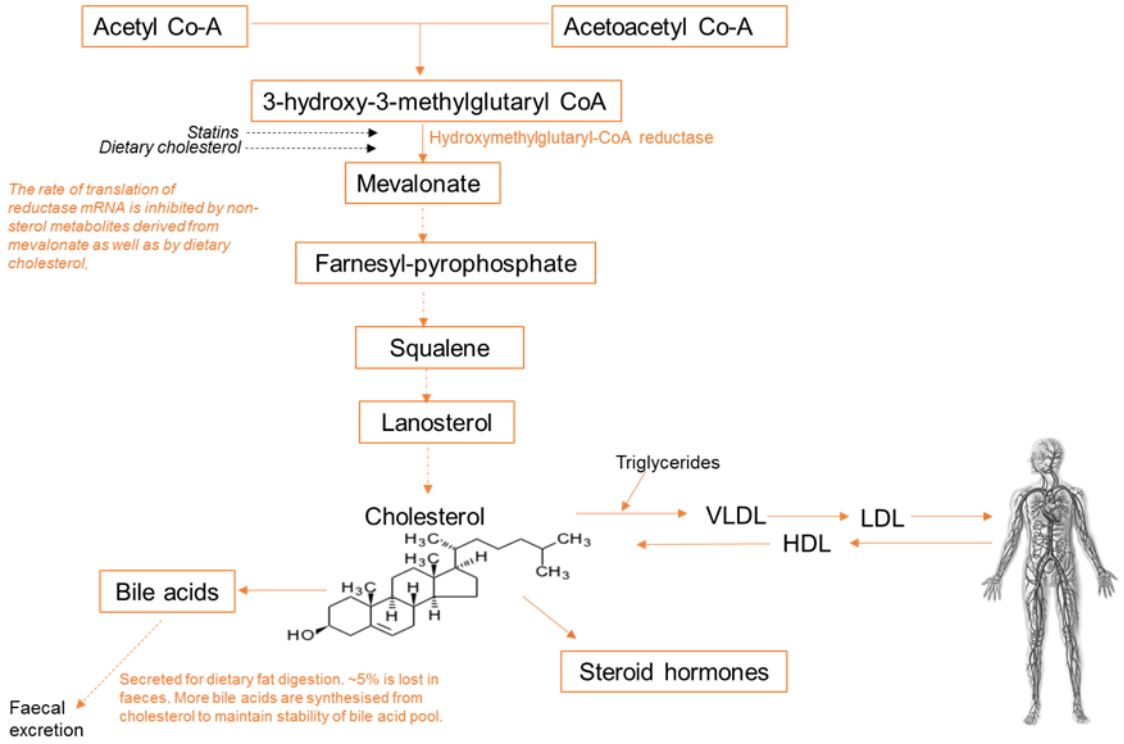


Figure 1-1: A simple outline of cholesterol metabolism in the body.

VLDL: very low-density lipoprotein, LDL: low density lipoprotein, HDL: high density lipoprotein

1.3 The Role of Diet in Cholesterol Management

HC is managed by both pharmacological and lifestyle intervention strategies. Statins are the most common pharmacological therapies used to manage blood cholesterol. They are a class of cholesterol-lowering medications that inhibit HMG-CoA reductase to block cholesterol synthesis. However, their use is hampered by adverse side effects on muscles, adverse drug-drug interactions, increase in blood glucose and liver damage (Golomb & Evans, 2008; M. Jones, Tett, Peeters, Mishra, & Dobson, 2017; Jose, 2016). Furthermore, some patients still develop CVD even after using statins (Sampson, Fazio, & Linton, 2012). Other medications include Ezetimibe which inhibits the absorption of dietary cholesterol by blocking the action of the NPC1L1 protein responsible for the transportation of dietary cholesterol within the intestine (Sudhop et al., 2002), which are not always as effective or without side effects (Hammersley & Signy, 2017).

However, lifestyle changes, particularly in the diet, remain the most important fundamental steps in the primary and secondary prevention of CVD, without or alongside medication (Yu, Malik, & Hu, 2018). The National Cholesterol Education Program recommends dietary modifications as a first-line option for preventing and treating hypercholesterolaemia (Expert Panel on Detection, Evaluation, 2001). These changes include reducing intake of saturated fat and increasing intake of soluble dietary fibres.

Dietary fibres are defined as edible carbohydrate polymers that are not hydrolysed by the endogenous enzymes in the small intestine of humans (J. M. Jones, 2014). They are commonly categorised based on their ability to disperse in water; namely insoluble dietary fibres and soluble dietary fibres (Figuerola, Hurtado, Estévez, Chiffelle, & Asenjo, 2005). While the former passes through the gastrointestinal tract almost intact and simply adds bulk to stool promoting regular bowel movements; soluble fibres (SF) dissolve, disperse or swell in water during digestion to form viscous gels and are mostly fermentable by gut microbiota (GM). Examples of SF include β-glucans, hemicelluloses and pectin (Figure 1-2).

The hypocholesterolaemic and CVD risk-reducing effects of consumption of SF are well-documented in literature (Anderson, Allgood, et al., 2000; Anderson, Davidson, et al., 2000; L. Brown, Rosner, Willett, & Sacks, 1999; Talati, Baker, Pabilonia, White, & Coleman, 2009; Theuwissen & Mensink, 2008; A. Whitehead, Beck, Tosh, & Wolever, 2014). A recent systematic review and meta-analysis of cohort studies found an inverse association between CVD and intakes of total dietary fibre, soluble fibre, cereal fibre, fruit fibre, vegetables and insoluble (Threapleton et al., 2013). Jovanovski *et al* reported a modest but significant decrease of 0.33 mmol/L (95% CI − 0.38 to − 0.27 mmol/L; $p < 0.00001$) in LDL with a daily median intake of 10.2 g (range 2.4–20.4g) of psyllium (Jovanovski et al., 2018). Adding

psyllium to statin therapy increased the hypocholesterolemic effect of the statins (Brum, Ramsey, McRorie, Bauer, & Kopecky, 2018). McRae discovered from an umbrella review of previously published meta-analyses that dietary fibre intake, particularly β-glucan and psyllium, reduced the risk of developing CHD and stroke by about 7% to 24%, as well as reduced their overall CVD related morbidity and mortality by 17% to 28% (McRae, 2017).

However, the mechanism of action behind SF-induced cholesterol lowering is poorly understood, especially in humans. Several hypotheses have been proposed, however literature strongly suggests that the major mechanism may involve a disruption of the BA metabolism leading to enhanced BA excretion (Fernandez, 2001; Gunness & Gidley, 2010).

To better understand the significance of an enhanced BA excretion, the next few sections will provide an overview of BA metabolism.

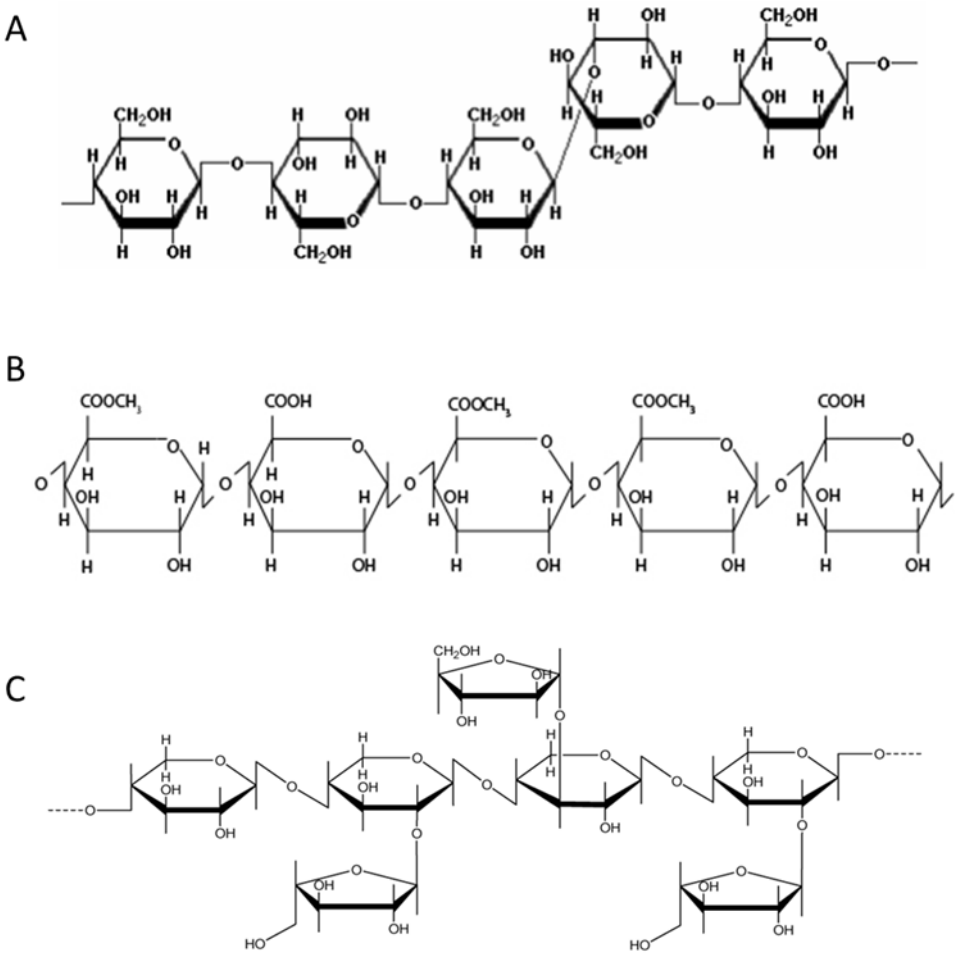


Figure 1-2: Structure of three most common soluble fibres

A) Cereal β-glucan with mixed link 1-3 and 1-4 glycosidic bonds. B) Pectin is a complex structure fundamentally made up of galacturonic acid moieties (shown here) with side chain sugars like rhamnose, xylose, fucose and arabinose. C) Psyllium - is made up of arabinoxylan, a hemicellulose (shown here).

1.4 Bile acid metabolism

BA are one of the most abundant and important classes of metabolites that are produced from a collaborative host-gut microbiota effort, the others being short-chain fatty acids and tryptophan pathway metabolites (E.-T. Jia, Liu, Pan, Lu, & Ge, 2019). They have emerged as physiologically important molecules playing crucial roles in digestion and absorption of nutrients, host metabolic homeostasis and immune processes (Brestoff & Artis, 2013; Fiorucci, Mencarelli, Palladino, & Cipriani, 2009; Russell, 2003). BAs are first synthesised by host hepatic cells, secreted in response to diet and are subsequently modified by certain members of the gut microbiota (GM) in a variety of biotransformations leading to the production of metabolites which have been implicated in cardiovascular diseases, colorectal cancer, inflammatory bowel disease, liver diseases and gastrointestinal diseases to name a few (Heinken et al., 2019; Molinero, Ruiz, Sánchez, Margolles, & Delgado, 2019; Ridlon, Harris, Bhowmik, Kang, & Hylemon, 2016; Ridlon, Kang, & Hylemon, 2006).

1.4.1 Formation of BA

BA are amphipathic metabolites formed from the catabolism of cholesterol in the liver in a series of oxidative transformations. About 500mg of cholesterol is broken down into BA daily in the adult human liver (Russell, 2003). This breakdown is initiated by the 7α-hydroxylation of the cholesterol molecule, which is the rate-limiting step in the reaction, catalysed by cholesterol 7α-hydroxylase enzyme. The 7α-hydroxylated intermediate undergoes further modifications in a series of steps involving 16 other enzymes to produce the primary BA (Javitt, 2002). In humans, the primary BA are cholic acid (CA) and chenodeoxycholic acid (CDCA). This is known as the classical or neutral pathway which accounts for most of the BA produced in humans under normal condition (Figure 1-3 summarises the main classical pathway in man). The acidic pathway, initiated by sterol-27-hydroxylase (CYP27A1), is the other major pathway in man, contributing about 10% (mostly CDCA) of total BA produced (Pandak & Kakiyama, 2019). Other minor pathways have also been identified in man and other animals.

Following synthesis in the liver these primary BAs are N-acyl amidated at the C-24 carboxyl group with the amino acids glycine, to yield glycocholic (GCA) and glycochenodeoxycholic (GCDCA) acids, or taurine, to yield taurocholic (TCA) and taurochenodeoxycholic (TCDCA) acids. This increases their solubility, reduces their ability to cause cell membrane damage and enables their secretion into bile and storage in the gallbladder (Russell, 2003). Secondary BAs are produced by GM-mediated amido-deconjugation and 7α-dehydroxylation of primary BA,

where G/TCA is converted into deoxycholic acid (DCA) and G/TCDCA into lithocholic acid (LCA).

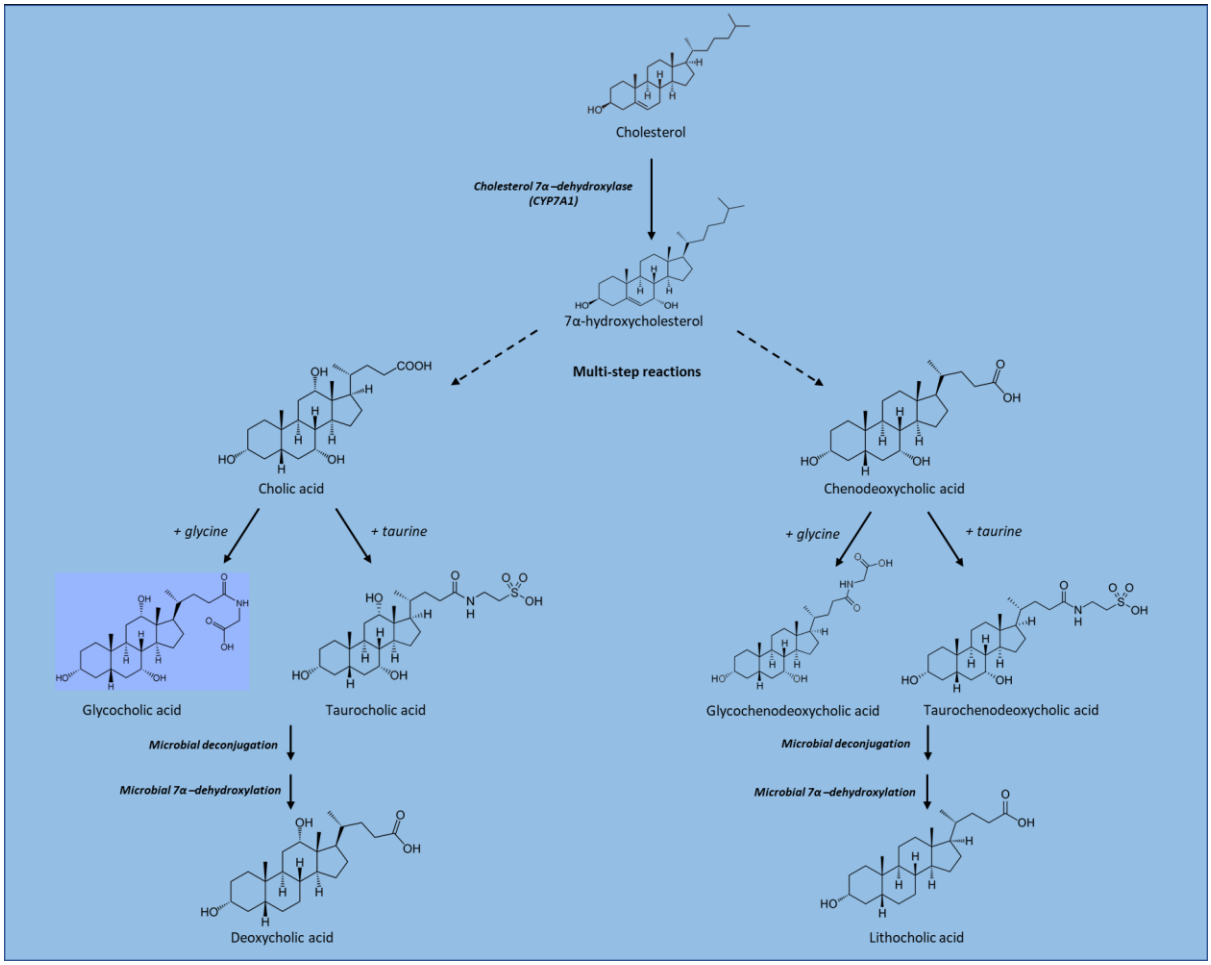


Figure 1-3: Simplified pathway of biosynthesis of primary bile acids from cholesterol in the liver.

Hepatic biosynthesis of BAs occurs principally via the classical pathway accounting for ≥75% of BA production under normal physiological conditions. The multi-step cascade of enzymatically catalysed reactions is initiated by the 7α-hydroxylation of cholesterol. This first reaction is catalysed by cholesterol 7α-hydroxylase (CYP7A1) which is the rate-limiting enzyme and determines the amount of BA synthesised. In man, primary BAs cholic and chenodeoxycholic acids are produced and conjugated with glycine or taurine. Microbiota-mediated deconjugation and subsequent 7α-dehydroxylation converts cholic to deoxycholic and chenodeoxycholic to lithocholic acids in the colon.

1.4.2 Physiological Functions of BA

1.4.2.1 Elimination of cholesterol

Cholesterol is eliminated from the body via two main routes namely the faecal excretion of cholesterol and of BA. Quantitatively, the latter is the more important route for getting rid of cholesterol. Cholesterol is catabolised into BA which are then excreted from the body. Apart from the breakdown of cholesterol into BA, the BAs also contribute to cholesterol excretion by micellar solubilization of cholesterol in bile, moving cholesterol from the liver environment to the intestines, from where it is ultimately eliminated via the fecal route (Alan F Hofmann, 1999).

1.4.2.2 Digestion of dietary lipids

BA are responsible for the emulsification and absorption of dietary lipids and fat-soluble vitamins and minerals in the small intestine. Upon entry of fats into the jejunum, a cholecystokinin-induced contraction of the gall bladder leads to the secretion of BA. Via the formation of mixed micelles, BAs are able to facilitate the solubilization, digestion, and absorption of lipid contents of the digesta in the duodenum (de Aguiar Vallim, Tarling, & Edwards, 2013). The amphipathic nature of conjugated BA confers detergent properties to these molecules that facilitate the digestion of dietary fats (Monte, Marin, Antelo, & Vazquez-Tato, 2009).

1.4.2.3 Hormones/signalling molecules

BAs were recently found to both be ligands for the nuclear receptors Farnesoid X receptor (FXR), Pregnane X receptor (PXR), Vitamin D receptor (VDR); G protein-coupled receptors G-protein-coupled bile acid receptor (GPBAR1/TGR5) and sphingosine-1-phosphate receptor 2 (S1P2) and several cell-signalling pathways (Hylemon et al., 2009; Studer et al., 2012). The constitutive androstane receptor (CAR) is another nuclear receptor that is activated indirectly by BA exposure (M. L. Chen, Takeda, & Sundrud, 2019).Nuclear receptors are a family of ligand-activated transcription factors that are activated by hormones and are important for embryogenesis, development, and metabolism (T. Li & Chiang, 2014; Mangelsdorf et al., 1995). The different BAs vary in their affinity for the different receptors in the following order; CDCA > LCA = DCA > CA for FXR; LCA > DCA > CA for PXR, 3-oxo-LCA > LCA > DCA > CA for VDR, LCA > DCA > CDCA > CA for TGR5 and conjugated BAs for SIP2 (Ridlon, Kang, Hylemon, & Bajaj, 2014; Zhou & Hylemon, 2014). While activation of PXR, VDR and CAR is thought to lead only to the hepatoprotective clearance of hydrophobic BAs, FXR and TGR5 are more involved in the regulation of lipid, lipoprotein and glucose homeostasis as well as

general BA metabolism (synthesis, secretion, absorption, transport and circulation) (Joyce & Gahan, 2016).

1.4.2.4 Modulation of the gut microbiota

Numerous reports have shown that BAs alter GM structure in clinically significant ways (Ridlon et al., 2014). BAs, particularly DCA, exert bactericidal effects by disintegrating membrane structure and inducing DNA damage (Begley, Gahan, & Hill, 2005). Their effect on membrane integrity may be due to their high detergent capacity conferred by their amphipathic property. Also, BA-mediated dysbiosis has been reported in cirrhosis (Kakiyama et al., 2013). The oral administration of BA has been shown to selectively alter GM populations, to counteract bacterial proliferation and translocation caused by obstruction of bile flow, and to block the effects of endotoxins and enterotoxins via direct and indirect mechanisms (Alan F Hofmann, 1999; Inagaki et al., 2006; Islam et al., 2011; Lorenzo-Zúñiga et al., 2003).

1.4.2.5 Intestinal immunity and inflammation

This is a promising new role for BAs that has only recently gained attention. So far, BAs have been demonstrated to have divergent roles in inflammation and mucosal immune responses. DCA and LCA are less soluble than CA and CDCA and as such are not easily reabsorbed from the intestines and are generally considered to be toxic. They have been previously linked with colorectal cancer and tumour proliferation, although DCA was shown to only promote tumorigenesis only in the presence of a carcinogen (Yoshimoto et al., 2013). Reduced faecal BA was associated with an increase in pro-inflammatory pathobiont *Enterobacteriaceae*, while faecal TCA was found to be positively associated with *Bilophila wadsworthia* which is a potentially pathogenic gut microbe linked with ulcerative colitis (Devkota et al., 2012; Kakiyama et al., 2013). BAs have been demonstrated to induce or suppress immune response factors via the activation of aforementioned receptors (M. L. Chen et al., 2019). However, there is a lot yet to be explored in this area.

1.4.2.6 Biliary functions

BA stimulate bile flow via their osmotic properties from liver cells into the bile canaliculi and then gall bladder, and induce the secretion of phospholipids into bile from the canalicular membrane (Alan F Hofmann, 1999). The presence of the phospholipids in bile enhances the stability of micelles which keeps BAs from damaging biliary epithelium and prevents

cholesterol crystallization and gallstone formation (de Aguiar Vallim et al., 2013; Alan F Hofmann, 1999).

1.4.2.7 Colonic functions

BAs modulate the functioning of colonic muscles and the absorption of water and electrolytes in the colon (Alan F Hofmann, 1999). The main human BAs, but not LCA, have been shown to decrease fluid and electrolyte absorption and promote secretion across the gut epithelium, although this effect on absorption of water and electrolytes is not always seen except at unusually high concentrations of 3–10 mM (Bajor, Gillberg, & Abrahamsson, 2010). Also, *in vitro*, animal and human studies have demonstrated that BAs may significantly impact gut motility and transit time (Bajor et al., 2010).

1.4.3 Enterohepatic circulation of BA

Primary BAs are secreted into the small intestine during digestion for the emulsification of fats. Once that task is completed, the BAs are recycled and transferred back to the liver via the portal blood by a combination of passive absorption in the proximal ileum, active transport in the terminal ileum, and passive absorption in the colon, with ileal active transport accounting for most of the reabsorption (Boyer, 2013). The active reabsorption and transcellular transport across the intestinal epithelium is facilitated in three stages by the apical sodium co-dependent BA transporter (ASBT) mediated uptake from ileum, ileal bile acid binding protein (IBABP) mediated transcellular transport on the apical side, and the heteromeric transporter, organic solute transporter OSTα-OSTβ mediated efflux on the basolateral membrane (Alrefai & Gill, 2007; Joyce & Gahan, 2016).

The (ileal) reabsorption is very efficient as 95-98% of BA are returned to the liver. However, 2-5% (~200 – 800 mg daily) “escape” into the colon where resident GM biotransform the primary BA into secondary BA (Gérard, 2013). Some BA may also spill into systemic circulation, in urinary system and in other tissue compartments (Dawson, Lan, & Rao, 2009; Swann et al., 2011). It is important to note here that the GM strongly influences BA metabolism by altering the composition and size of the BA pool. To fully appreciate the role of the GM in the BA metabolism, the GM will be discussed more extensively in a following sub-section.

The two main processes that occur in the colon are first, the deconjugation of taurine- and glycine-conjugated primary BA and then the 7α-dehydroxylation of the unconjugated primary BA into secondary BA. CA is converted to DCA while CDCA is converted into LCA. About 30% of DCA and very little of the LCA may be reabsorbed by passive reabsorption in the colon and

returned to the liver for further processing while the rest are excreted in faeces. DCA and LCA that are transported to the liver cannot be 7α-hydroxylated because humans lack the enzymes. LCA is very toxic and is readily sulphated by liver enzymes to reduce its toxicity and excreted; while DCA may be conjugated and enter into systemic circulation.

Most of the secondary BA is eventually excreted in faeces, with DCA being the most abundant. The BA pool is tightly regulated such that any BA loss would result in a *de novo* synthesis of BA in the liver to replenish the lost BA. This leads to a depletion of hepatic cholesterol stores. There is a consequent increased expression of LDL receptors and *de novo* cholesterol synthesis. This completes an enterohepatic circulation (EHC). The BA pool is tightly regulated such that very minimal perturbations are observed under normal conditions.

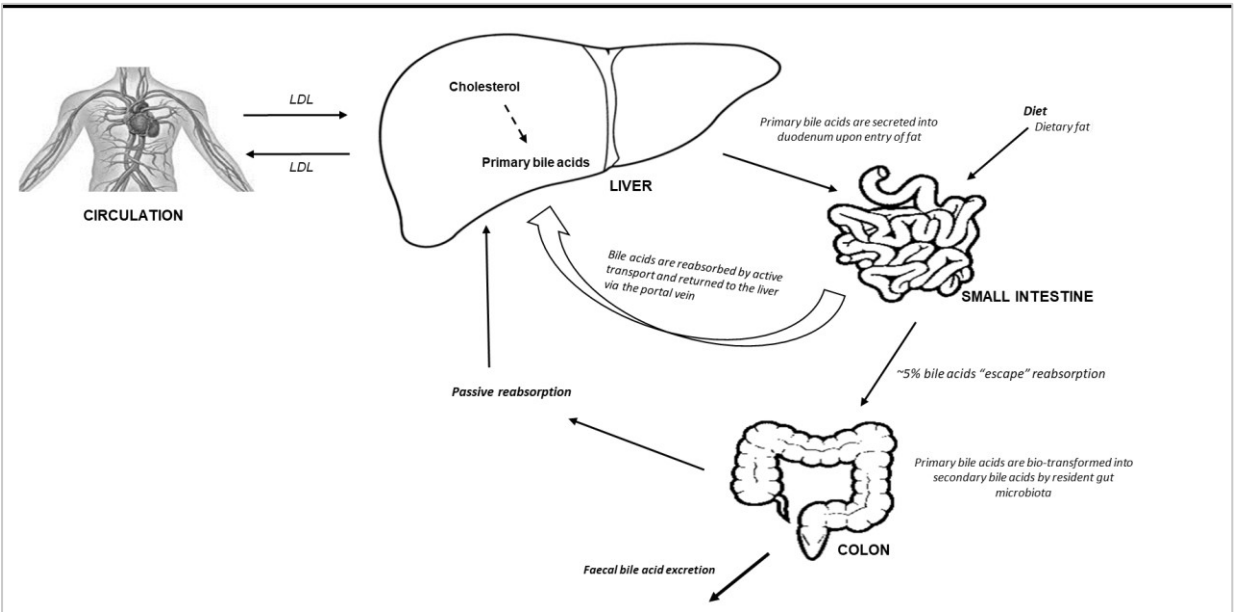


Figure 1-4: Diagrammatic representation of the enterohepatic circulation of bile acids.

1.4.4 Gut microbial modulation of bile acids

1.4.4.1 The gut microbiota

The gut microbiota (GM) is the sum total of microorganisms, such as bacteria, archaea, fungi and virus, living in the gastrointestinal tract, particularly the colon which is the most microbially dense site in the body (Sender, Fuchs, & Milo, 2016). Their genomes are collectively referred to as the "gut microbiome". The GM is heavily influenced by the host physiological state as

well as the environment (such as diet, exercise and xenobiotics) but also can in turn impact host metabolic processes. GM-associated diseases are on the rise as reports increasingly link dysbiosis (an altered microbial composition) with many diseases and infections (Kho & Lal, 2018; Thursby & Juge, 2017). This has warranted reinforced exploration of the interface between host's metabolism and the microbial genome it harbours.

GM derive energy from dietary substrates (e.g. dietary fibres), endogenous compounds (e.g. BA) and xenobiotics that enter the gut, and in the process produce metabolites that may benefit or harm the host (Patterson et al., 2014). Conversely, the immune system of the host and gut physiology (e.g. pH which could be influenced by the presence of BA) may alter the gut environment and modulate the microbial community (Zhao, 2013).

The diversity and abundance of microbial populations in the gut at any time is thus determined by a balance between the abilities of a microbial population to make use of available nutrients for proliferation and to adopt strategies that allow them survive unfavourable conditions (Zhao, 2013). Although not fully characterised, the gut microbial community is known to be dominated by the bacterial phyla *Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria* and *Verrucomicrobia* (Eckburg et al., 2005; Tremaroli & Bäckhed, 2012). Disturbances in the "normal structure" (e.g. increase/decrease in a particular phyla) of GM have been linked to the development of various diseases, such as CVD, metabolic syndrome and inflammatory bowel disease (Ramezani & Raj, 2014; Tremaroli & Bäckhed, 2012). However, it is not yet clear whether associated pathogenesis is caused by an altered GM or whether the pathogenesis drives the changes seen in this microbial community (Patterson et al., 2014).

1.4.4.2 A few words about technical terms…

Having been disproportionately explored more compared to the other communities, the bacteria and archaea communities are almost exclusively referred to as "microbiome" in literature, while the virus and fungi communities are specifically called virome and mycobiome respectively. Thus, "gut microbiota", unless otherwise indicated, will refer only to the gut bacteria and archaea components of the total gut microbiota in this thesis.

Also, despite the technical differences, both "gut microbiota" and "gut microbiome" may be used interchangeably in literature (and in this thesis). Lastly, the gut microbiome/microbiota will be referring to the colonic microbiota community, and not the whole of the gastrointestinal tract, unless otherwise stated.

1.4.4.3 Gut-microbiota mediated transformations of bile acids

As previously stated, approximately 5% of conjugated primary BA escape into the colon where they undergo a variety of biotransformations mediated by resident bacteria. The GM strongly influences BA metabolism by altering the composition and size of the BA pool.

GM-mediated biotransformation modifies BAs both at the side chain and on the steroid nucleus of the BA structure. At the side chain, a considerable number of GM possess bile salt hydrolases (BSH) which catalyse the deamidation (deconjugation) of the attached taurine and glycine. BSH genes are widespread within the three phyla *Firmicutes*, *Bacteroidetes* and *Actinobacteria*, and more specifically within the genera *Bacteroides, Bifidobacterium, Clostridium, Enterococcus* and *Lactobacillus* (B. V Jones, Begley, Hill, Gahan, & Marchesi, 2008), although a more recent analysis further identified that *Blautia, Eubacterium*, and *Roseburia* in human samples alongside other genera possessed BSH genes as well (Song et al., 2019).

A wider range of transformations occurs at the steroid nucleus of the BA. In addition to deconjugation and 7α-dehydroxylation already discussed, bacteria in the gut are also able to transform the BAs in many other reactions such as dehydroxylation, oxidation and epimerization of the hydroxyl groups at the C3, C7 and C12 positions, in addition to esterification and desulfatation, thereby diversifying the BA pool (Ridlon et al., 2006). Over 50 chemically distinct BAs have been discovered in blood, urine and faeces (Sarafian et al., 2015; K. D. Setchell, Lawson, Tanida, & Sjövall, 1983). Table 1-2 summarizes the identified transformations that have been characterised as well as GM members that mediate these effects.

Of all the GM-mediated transformations, 7α-dehydroxylation is the most quantitatively important, as faeces are dominated by the secondary BA, DCA and LCA (Ridlon et al., 2006). The microbial actions on BAs have been hypothesised to provide energy sources (e.g. taurine is a source of nitrogen and energy) and reduce toxicity of conjugated BAs (e.g. the 7α-hydroxyl group on CDCA is epimerised to the 7β-conformation, yielding ursodeoxycholic acid (UDCA, also 7β-CDCA) and thus decreasing its toxicity to bacteria (Winston & Theriot, 2020).

Table 1-2: Gut bacterial genera that are involved in biotransformation of bile acids.

Reactions	Enzymes involved	Examples of GM genera that express these enzymes
Deconjugation of conjugated bile acids to yield free BA	Bile salt hydrolases	*Bacteroides*, *Bifidobacterium*, *Clostridium*, *Lactobacillus*, *Listeria,* certain strains of *enterococci, Veillonella*
Oxidation and epimerization	Hydroxysteroid dehydrogenase	*Bacteroides*, *Clostridium*, *Escherichia*, *Egghertella*, *Eubacterium*, *Peptostreptococcus*, *Ruminococcus*
7-dehydroxylation at C-7 hydroxyl group of the steroid nucleus	Series of steps catalysed by a number of enzymes	*Clostridium*, *Eubacterium,* Certain strains of anaerobic *Lactobacilli, Bacteroides,*
Esterification		*Bacteroides*, *Eubacterium*, *Lactobacillus*
Desulfatation (Hydrolytic cleavage of sulphate esters)		*Clostridium*, *Fusobacterium*, *Peptococcus*, *Pseudomonas*

Table summarising gut microbial-mediated biotransformation of bile acids. Based on information retrieved from Gérard (2013) and Prabha & Ohri (2006). GM: gut microbiota

1.5 Bile Acid Excretion by Dietary Soluble Fibres

Increased BA excretion has been explored in a number of interventions for cholesterol-reduction. This has been achieved through a variety of ways including but not limited to blockage of important enzymes in the EHC, physical disruption of the EHC by surgical interventions or direct binding of BA. All have a main goal which is to increase the faecal excretion of BA in order to trigger the cascade of reactions involving the increased catabolism of hepatic cholesterol to replenish lost BA and maintain the stability of the BA pool and the eventual reduction in blood cholesterol.

The most common point at which the EHC is disturbed is the ileum because that is where to obtain the most effect. Blocking ASBT led to an increase in hepatic BA synthesis and decrease in LDL cholesterol in rabbits (Bhat et al., 2003). Partial ileal bypass surgery on the Program on the Surgical Control of Hyperlipidaemias (POSCH) trial resulted in significant reductions in both plasma LDL levels and CAD (Buchwald et al., 1990).

The promotion of hepatic cholesterol metabolism following SF-induced BA excretion has been evidenced by an up-regulation of 7α-hydroxylase activity reported in guinea pigs (Fernandez, 1995), and increased serum concentrations of 7α-hydroxy-4-cholesten-3-one in humans (Bertolotti et al., 2008; Ellegård & Andersson, 2007) when following high SF-diets. 7α-hydroxy-4-cholesten-3-one has been shown to be a reliable marker of hepatic cholesterol 7α-hydroxylase activity in humans (Axelson, Björkhem, Reihnér, & Einarsson, 1991; Bertolotti et al., 2008; Sauter, Berr, Beuers, Fischer, & Paumgartner, 1996). Increased conversion to BA depletes cholesterol stores in the liver (Fernandez, 1995; Rideout et al., 2007), creating a hepatic demand for cholesterol, which may be met by enhanced uptake of lipoprotein cholesterol from circulation (Rideout, Harding, Jones, & Fan, 2008). Indeed, SF consumption has been demonstrated to increase hepatic LDL-receptor expression and LDL turnover (Fernandez, 1995; Fernandez et al., 1995; J. D. Horton, Cuthbert, & Spady, 1994; Rideout et al., 2007).

Elevated demand for cholesterol in the liver may also lead to an increase in *de novo* synthesis, indicated by an up-regulation of HMG-CoA (Garcia-Diez, Garcia-Mediavilla, Bayon, & Gonzalez-Gallego, 1996), and sterol synthesis in liver of hamsters (J. D. Horton et al., 1994; Turley, Daggy, & Dietschy, 1991). However, it should be noted that these same groups reported a net decrease in plasma and hepatic cholesterol showing that enhanced HMG-CoA activity and *de novo* cholesterol synthesis may not be sufficient to compensate for SF-induced faecal BA loss (Fernandez et al., 1995; Garcia-Diez et al., 1996; J. D. Horton et al., 1994; Moundras, Behr, Rémésy, & Demigné, 1997; Turley et al., 1991).

While the aforementioned are consequences of SF-induced altered BA metabolism, the underlying mechanism(s) by which SF successfully disrupt BA circulation is (are) yet to be elucidated. There are suggestions that SF may bind BA in a similar fashion to BA sequesterants (Kritchevsky & Story, 1974). BA sequesterants are anion-exchange polymers that bind BA in exchange for their chloride ions (Casdorph, 1967; Tennent et al., 1960). However, ionic binding as a mechanism for SF is highly unlikely because SF are mostly neutral or even acidic (as in the case of pectins) and as such have a weak binding ability. For instance, Daggy et al. found that psyllium increased the excretion of unconjugated BA while cholestyramine enhanced excretion of conjugated BA, and thus concluded that psyllium does not bind BA in vivo, as the forms of BA excreted would have been similar in both cases (Daggy, O'Connell, Jerdack, Stinson, & Setchell, 1997).

In vitro experiments investigating the BA binding activity of dietary fibres have suggested that binding in the ileum may be hydrophobic or a combination of different types of interactions (Story & Kritchevsky, 1978). The experiments of Eastwood and Mowbray demonstrated that adsorption and binding of BA by SF may be diminished and reversible in the presence of fatty acids and other possible components present in the digesta and the colon (Eastwood & Mowbray, 1976). This indicates that even if SF bind BA, it is likely they dissociate in the colon, suggesting that the mechanism of action may not solely involve the binding of BA. They surmised that SF-induced BA binding in the small intestine might only be sufficient to obstruct reabsorption and enhance delivery of BA to the colon where gut bacteria may metabolise them into less soluble products that cannot be reabsorbed by the body and are consequently excreted (Eastwood & Mowbray, 1976).

In a relatively recent investigation, Benson and colleagues (Benson et al., 1993) discovered that a considerable amount of labelled cholic acid metabolites in hamster and human faeces were trapped in a water/methanol-insoluble fraction and were only analysable when subjected to hydrolysis (saponification). Further analyses revealed that these "saponifiable" metabolites contained 3α-hydroxy polyesters of deoxycholic acid, hence called polydeoxycholic acid (PDC). They reported that PDC constituted a high proportion of total faecal BA and correlated negatively with serum cholesterol levels.

Few other investigations of faecal samples from animals and humans have identified the presence of variable proportions of "saponifiable BA metabolites" (Korpela, Adlercreutz, & Turunen, 1988; Korpela, Fotsis, & Adlercreutz, 1986; Korpela, Korpela, & Adlercreutz, 1992; Norman, 1964; J.-X. Zhang et al., 1993). It is likely that these polymeric BA are lost during extraction or purification procedures where their presence was not being monitored (Benson et al., 1993), are extracted alongside neutral sterols (Korpela et al., 1986) or hydrolysed by

extraction conditions, as most extraction procedures include alkali treatment. Indeed, Kakiyama and colleagues reported an increase in deoxycholic acid when an alkaline hydrolysis step was added to their BA extraction process (Kakiyama et al., 2014).

A proposed structure is presented in Figure 1-5. PDC is most likely synthesised by microbiota in the gut because of its absence in bile (Gérard, 2013) and decreased synthesis during antibiotic treatment (Korpela et al., 1986). In addition, microorganisms and not humans are known to make polyesters from substrates (Gumel, Annuar, & Chisti, 2013). In their conjugated form, BA are highly toxic to microorganisms; it is likely that converting BA to their less toxic metabolites through deconjugation, 7α-hydroxylation and esterification are survival mechanisms that have been adopted by some gut bacteria. Interestingly, some studies have reported that these saponifiable BA metabolites may be modifiable by dietary modulation. Korpela et al. and Zhang et al. reported increases in faecal saponifiable BA in humans and hamsters respectively following administration of diets rich in oat and rye (Korpela et al., 1992; J.-X. Zhang et al., 1993). In another study, Korpela and colleagues reported that vegetarians had the highest relative proportions of saponifiable BA in faeces, with some up to 80% of total faecal BA, when compared to omnivores and patients with colorectal cancer (Korpela et al., 1988). Vegetarians are known to have a higher intake of fermentable carbohydrates (SFs) and lower plasma cholesterol levels compared to omnivores (F. Wang et al., 2015), suggesting a possible link between higher intake of SF, production of saponifiable BA, and reduced blood cholesterol.

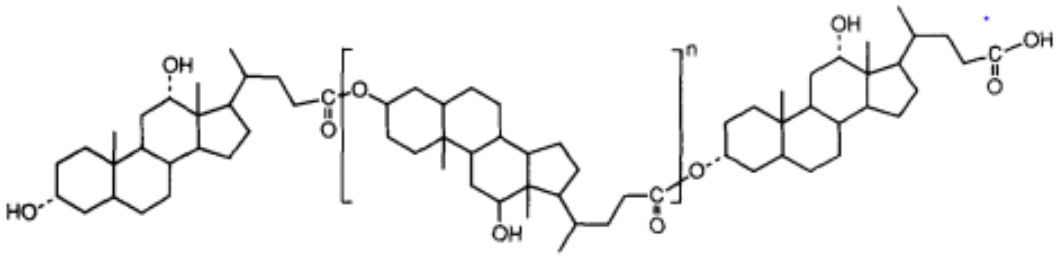


Figure 1-5: Proposed structure for polydeoxycholic acid

The C3-hydroxyl group in the preceding deoxycholic acid is esterified to the C24-carboxyl group in the consecutive deoxycholic acid monomer. Reprinted with permission from Benson *et al.,* 1993 (Benson et al., 1993).

1.5.1 Hypothesis

Based on this evidence, the following mechanism was hypothesised as a mechanism for the observed hypocholesterolaemic effect of SFs: SFs disrupt the enterohepatic cycle of BA by binding primary BA in the ileum via hydrophobic interactions, preventing reabsorption and enhancing their flow into the colon. In the colon, SFs, serving as a nutrient source, alter gut microbial metabolism to selectively increase the formation and subsequent excretion of PDC resulting in a consequent reduction in blood cholesterol and its associated CVD risk.

1.6 Thesis aims and objectives

1.6.1 Main Aim

To test the new hypothesis, the mechanisms involved in PDC synthesis and the effects of high SF diets on its production will be investigated. The relationship between PDC production, blood lipid biomarkers and colonic bacterial metabolism will also be explored.

1.6.2 Objectives of this PhD

The objectives of this PhD project are:

- To identify, isolate and characterise bacteria responsible for and mechanisms involved in intestinal formation of PDC.
- To investigate the effect of diet on faecal PDC production in humans.

1.6.3 Justification for this PhD project

The main aim of this PhD project is to elucidate the *in vivo* mechanism(s) behind the hypocholesterolaemic effect of SF in humans. Although decreases in cholesterol because of dietary interventions may be small, improvements as little as 0.3-0.6 mmol/L may considerably reduce disease burden due to CVD on an epidemiological scale (L. Brown et al., 1999). A thorough understanding of how SF reduce cholesterol would allow for the development of effective dietary strategies to reduce the prevalence of hypercholesterolaemia and associated CVD risks and may also serve as the basis for public health advice on the cardioprotective role of dietary fibre.

Evidence from this study may also be of great benefit to the food industry. The flexible nature of the GM, their influence on host health and their accessibility for modulation has resulted in this indigenous community becoming a therapeutic target in recent years. Various approaches have been, and are being, developed to modulate the gut microbiota in order to prevent or treat various disease conditions (W. Jia, Li, Zhao, & Nicholson, 2008).

Already, various diet-based therapeutic strategies taking advantage of the propensity of the diet to modulate the composition of the human gut microbiota have been extensively researched. These include probiotics, live microorganisms introduced through the diet to confer beneficial effects on the host, such as various *Bifidobacteria* and *Lactobacilli* species (Hill et al., 2014; M. L. Jones, Martoni, & Prakash, 2012), and prebiotics, defined as non-digestible foods that selectively stimulate the growth and/or activity of particular microbial populations in the colon to benefit host metabolism (Gibson et al., 2017). Established prebiotics include fermentable carbohydrates such as psyllium and acacia gum with demonstrated cholesterol-lowering activity as well (Slavin, 2013), while others like β-glucans have been shown to possess prebiotic potential (Arena et al., 2014; Carlson, Erickson, Hess, Gould, & Slavin, 2017). Thus, the results from this study may lead to the development of strategies such as a prebiotic approach using SF to selectively modulate the synthesis of PDC-producing bacteria or/and a probiotic approach by directly administering the candidate bacteria.

1.7 Conclusion

In summary, this thesis aims to investigate the effect of an environmental factor (diet) on the interplay of host metabolism (BA) and GM. BA are one of the most abundant and important classes of metabolites that are produced from host-GM co-metabolism. BA and GM are heavily influenced by the other and in turn both can be modulated by the diet through a variety of pathways, creating a complex 3-way dynamic that has wide-ranging implications for host health (Joyce & Gahan, 2016; Ridlon et al., 2014).

The main aim of this chapter was to introduce the reader to the concepts explored within the PhD project to assist the reader better appreciate the scientific background and justification for the work reported in this thesis.

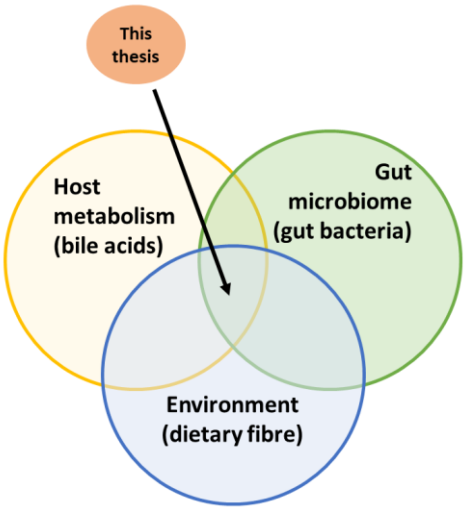


Figure 1-6: The "Uninhabited" Niche this PhD aims to contribute to.

This thesis is a multicomponent and multidisciplinary project that draws from the fields of analytical chemistry techniques and microbiology to solve a biological question in humans (clinical research). Overall, this thesis aims to make a modest contribution to the wealth of scientific knowledge being generated in this budding research area by exploring the interplay between the environment (diet), gut microbiota GM and host BA metabolism and implications on host physiology.

2. ANALYTICAL STRATEGIES

2.1 Introduction

Any research undertaking is only as good as the data collected, and this in turn is largely dependent on the methodology and strategies adopted in acquiring the data. Strategies for sample collection and analysis as well as methods for data collection, analysis and interpretation must be carefully considered to ensure robust, unbiased and reliable results. Other important and pertinent issues to consider when designing a methodology may include determining the current best analytical technique available to acquire the desired data by reviewing the literature, costs of analyses and infrastructural constraints to ensure practicality/feasibility of the research endeavour.

As identified in Chapter 1, the overall aim of this thesis is to investigate the mechanisms and gut microbes involved in the polymerisation of deoxycholic acid (DCA) to produce polydeoxycholic acid (PDC). To achieve this aim, the two main types of data to be collected were metabolite (*i.e.* DCA and PDC) and microbiome (candidate bacterial producers) data. These two data types respectively required metabolomic and metagenomic tools for their acquisition.

Metabolomics, or metabonomics, refers to the comprehensive characterisation of the metabolites in a biological sample using cutting-edge analytical techniques (Nicholson & Lindon, 2008). This can be untargeted, where global profiles of biological specimen are assessed, or targeted, where an array of known compounds are quantified usually against chemical standards. Metabonomics may also be qualitative, as applied in the identification and structural elucidation of metabolites, or quantitative, as applied to the quantification of metabolites of interest.

Metagenomics, on the other hand, refers to the detailed study of the genomes of microorganisms in a sample using modern genomic techniques (Handelsman, 2004). Metagenomic analytical techniques may target the whole genome, such as in whole-genome sequencing, or specific marker genes, such as 16S and 23S ribosomal RNA (rRNA) gene sequencing methods.

In summary, the two main types of data required to address the research questions posed in this thesis were biological analytes (metabonomics/metabolomics) and microbiome/microbial data (metagenomics). To ensure these data are collected with scientific rigour, various cutting-edge platforms/techniques from both fields were deployed for the acquisition of robust good quality data.

2.2 Aim and Objectives

2.2.1 Aim

The main aim of this chapter is to discuss the research strategy for all of the work presented in this thesis.

2.2.2 Objectives

To help the reader better appreciate and interpret the results obtained from the research work carried out in this thesis, this chapter will

- Introduce the reader to the principles underlying the analytical techniques employed in this thesis;
- Educate them on the strengths and limitations of these methods;
- In addition, the specific applications of these techniques in the thesis are briefly described but full details of procedures are reserved for the respective results chapters.

2.3 Sample Collection Techniques

Sample collection is very important because this is the first step in the research process that ensures the quality of outcomes and results. This starts from deciding what samples to collect through to designing experiments to ensure the selected sample types are collected and stored to high standards. For this project, samples were obtained from a mixture of *in vivo* and *in vitro* experiments. Also, determining the sample size is important, more so in *in vivo* than in *in vitro* experiments, in order to facilitate statistical analysis and obtain robust meaningful results.

2.3.1 Sample selection, collection and handling

The human faecal material is the principal sample that has been used in this project as the primary analyte of interest, PDC, is found in faeces. Another aim of the thesis was to identify potential gut microbial producers of PDC. However, the colon is inaccessible and impractical for frequent sampling. The faecal microbiota is generally accepted as a proxy for colonic microbiota (Falony et al., 2016). However, previous reports have shown that faecal microbiota differ significantly from colonic microbiota sampled from the ascending, transverse, descending or sigmoid colon (Rangel et al., 2015; Zoetendal et al., 2002), rectum (Momozawa, Deffontaine, Louis, & Medrano, 2011) and caecum (Marteau et al., 2001). The Eckburg and Ringel teams have suggested that faecal microbiota may be more indicative of luminal microbial communities than mucosal microbiota (Eckburg et al., 2005; Ringel et al., 2015). While mucosa-associated microorganisms interact more closely with the host epithelium and may therefore be predominantly involved in host immune processes, gut microbes in the lumen are more involved with nutrient digestion and may be less involved with host immune system via production of physiologically important metabolites (Rangel et al., 2015; Van den Abbeele, Van de Wiele, Verstraete, & Possemiers, 2011).

Although versatile, non-invasive and easy to collect, stool is a complex and sensitive material that demands proper handling to avoid confounding or biased results. Major issues to consider include but are not limited to the following:

- Homogeneity of sample: Stool is very heterogenous and different sampling sites can be very dissimilar in both microbiome (Swidsinski, Loening–Baucke, Verstraelen, Osowska, & Doerffel, 2008), and metabolite (Y. Liang et al., 2020) investigations;
- Day to day fluctuations in faecal BA metabolome: Due to wide day to day variations previously observed in faecal BA, collection of faecal samples over 3-5 days has been advised (Imray et al., 1992; K. D. R. Setchell, Ives, Cashmore, & Lawson, 1987);

- Burden on the study participant or donor: Donor's concerns about sanitation, sensory and personal feelings of awkwardness constitute a substantial challenge that may arise and may impact the readiness and compliance of the donor which will unarguably determine the quality of sampling (Wu et al., 2019);
- Presence of microbiome-altering xenobiotics: As the primary analytes of interest are by themselves microbial species or are modified by these organisms, it is important to ensure that study participants and sample donors have not recently used oral or topical medications that may disturb their "normal" gut microbiome;
- Storage: In many studies, stool samples are collected and immediately frozen, immersed in Tris-EDTA, RNAlater or ethanol or collected in commercial stool preservation tubes (Watson, Giles, Scherer, & Blatchford, 2019). The criterion standard for proper storage is immediate freezing at -80 ºC but this is most times impractical in cases where samples are collected in free-living conditions. The next best option is storing at -20 ºC which is the temperature for most home freezers. Samples stored at both temperatures were found to be similar in microbial profiles (Bundgaard-Nielsen, Hagstrøm, & Sørensen, 2018), while metabolic profiles were minimally affected by storage conditions (Y. Liang et al., 2020). Duration of storage may also affect downstream microbiome analysis (Carruthers et al., 2019); and
- Maintenance of anaerobic conditions: It is essential to keep the sample as intact as possible after voiding by ensuring anaerobic conditions are maintained. This is particularly important because most of the intestinal microbiota are anaerobic and an exposure to aerobic environments may significantly alter the faecal metabolome and microbiome (N. Martínez, Hidalgo-Cantabrana, Delgado, Margolles, & Sánchez, 2019). For analyses requiring fresh stool samples, samples have been collected and rapidly transferred to anaerobic containers with anaerobic atmosphere generation sachets (Cueva et al., 2013; Grimaldi et al., 2017; Martín-Peláez et al., 2008).

To circumvent most of these challenges, the go-to option is to provide donors with user-friendly DIY at-home kits with clear and simple instructions on how to collect samples. Participants are usually then required to freeze samples immediately after voiding or to keep in anaerobic conditions while the research team retrieves the samples as quickly as feasible. It is highly recommended to homogenize the stool sample before aliquoting for analysis (Gorzelak et al., 2015). Samples are stored at -80 ºC until day of analysis, which ideally should be as quickly as possible (Lauber, Zhou, Gordon, Knight, & Fierer, 2010).

For this thesis, stool samples were obtained from volunteers recruited onto clinical studies for direct analyses or from donors and used as inoculum for *in vitro* experiments from which

samples were obtained and analysed by targeted metabolomic (*i.e.* BA analysis) and metagenomic assays.

2.3.2 Study size estimation

Determining the sample size of a study is a crucial part of study design because it allows for judicious use of resources, ensures validity of the study and generalisability of results obtained (Lachin, 1981; Lenth, 2001). In a hypothesis-driven research experiment, there are always two hypotheses to consider (S. R. Jones, Carley, & Harrison, 2003):

- The null hypothesis: which states that there is no difference between the intervention or treatment arm and the control arm; and
- The alternative hypothesis: which states there is a difference between the two arms.

Sample size estimation is done statistically by carrying out a power calculation which require four pieces of information namely: the desired effect size (which is the minimum clinically relevant difference between the treatment and control groups), the usual amount of variance within that sample population (usually given by standard deviation determined previously), the significance level (denoted by α, which is the probability of rejecting the null hypothesis when it is true i.e. a false positive aka type I error.) and the power (denoted by β, which is the probability of detecting a difference when there is one to find to avoid a false negative aka type II error).

A sample size that is too small is said to be “underpowered” and may not be able to detect significant effects thus producing results that cannot be used, while one that is too large may not detect a small but clinically relevant effect (Lenth, 2001). Particularly in human clinical studies, this becomes an ethical issue as an undersized or oversized study will be exposing participants to potentially harmful interventions (or placing a burden on them in the case of non-interventional trials) unnecessarily and/or without advancing knowledge (Lenth, 2001).

However, power calculations for pilot studies may not be possible because the biological variation or clinically significant difference may not be known (Dell, Holleran, & Ramakrishnan, 2002). A rule of thumb of $n = 12$ per intervention arm is usually advised in this instance (A. L. Whitehead, Julious, Cooper, & Campbell, 2016).

2.3.3 Cross-sectional cohort study

One of the *in vivo* study designs adopted for sample and data collection in this thesis was a cross-sectional cohort study (Gallo et al., 2014; Hudson, Pope, & Glynn, 2005). This is an observational clinical study design that is a merger of a conventional cohort study and an analytic cross-sectional survey, therefore may rank similarly to case-control studies on the hierarchy of evidence (Figure 2-1). The cross-sectional cohort study is a hypothesis-driven study where exposure and outcomes are concurrently assessed in a defined human population at a particular point in time. Study participants are recruited based on their exposure status and appropriate controls may be selected by matching. This kind of study design is retrospective, relatively inexpensive (compared to prospective cohort studies and randomised human trials), simple to run, ethically safe and suitable for assessing multiple outcomes. However, due to its retrospective nature, data obtained from this study may be subject to recall bias and causality cannot be established due to a lack of randomisation or *a priori* allocation to exposure. Although, participant matching when done helps to reduce the effect of confounders and bias, this study design may not completely eradicate hidden and selection biases (Thiese, 2014). Also, temporality cannot be established because sample and data collection is a one-off exercise (Michael, N., & A., 2008).

Specifically, in this thesis, the cross-sectional cohort study was conducted in vegans and omnivores who were recruited from urban areas in Denmark through advertisements placed in print, social and online media resources. Comprehensive clinical metadata and dietary information were obtained, and blood, urine and stool samples were collected for various analyses. Further details are given in Chapter 5 where extensive results from this cohort are presented.



Figure 2-1: Hierarchy of evidence

Adapted with permission from Yetley et al (2017) (Yetley et al., 2017)

2.3.4 Randomised controlled trial

The randomised controlled trial (RCT) is regarded as the gold standard for human studies and ranks very high on the hierarchy of evidence, second only to systematic reviews of RCTs (Evans, 2003). Trial participants are recruited using prespecified inclusion and exclusion criteria and are prospectively assigned to a treatment arm or a control arm in a random manner. The randomisation results in comparable parallel groups with evenly distributed confounders and with the intervention/treatment as the only difference between groups, theoretically (Michael et al., 2008). Methodological additions like blinding and compliance measurement may help to ensure the internal validity of the study and results obtained. Blinding is the concealment of intervention allocation from at least one group of stakeholders in the trial, be it participants, assessors or both (Jadad et al., 1996). Single blinding is when allocation is concealed from one of these groups, usually the trial volunteers, while in double blinding both participant and assessor are blinded to the randomisation. Compliance is a measure of how well the participants adhered to the intervention or treatment they were allocated to. Thus, the internal validity of an RCT is threatened if there is differential compliance between treatment arms (Thiese, 2014). Furthermore, this conventional parallel RCT design may require a large sample size to overcome interindividual variation and thus may be very expensive to run.

A crossover RCT overcomes this disadvantage by recruiting participants onto all treatment arms in a sequence separated by adequate “washout” periods (Mills et al., 2009). By serving as their own control, the error variance due to interindividual variation is greatly reduced and so is the required sample size (T. A. Louis, Lavori, Bailar, & Polansky, 1984) which may make it less costly. The crossover RCT also overcomes the effect of poorly conducted randomisation since every recruited participant will get all treatments (intervention, placebo or control). However, washout period must be adequate to ensure a successful trial. This design is also only applicable for short-term investigations, as patient dropout rate increases as study duration increases (Alexander, 2013).

In this body of work, samples were obtained from a double blinded crossover RCT investigating the effect of a specific dietary intervention on the production of PDC. Details and results from this study are presented in Chapter 6.

2.3.5 *In vitro* sample collection strategies

In order to better understand processes going on in the colon with respect to gut-microbiome and bile acid metabolism interactions, colonic processes have to be studied. However, the colon is very inaccessible. Therefore, *in vitro* models may be used to explore these processes.

Not only do *in vitro* models provide practical ways to properly investigate hypotheses in otherwise inaccessible conditions (such as the human colon in this case), they also help to simplify processes which are normally physiologically complex. This gives the investigator control over the system and enables them focus on a small number of analytes, pathways, etc at a time, and thus delineate and untangle the complexities found in higher-level organisms like humans. They are also low cost, can be reproduced fairly quickly, and ensures all factors are controlled for.

However, care must be taken to ensure that *in vitro* experiments are carried out such that they closely resemble the physiological state being studied. In a way, their greatest advantage is their limitation. While *in vitro* models remove the complexities of a functional physiological system as found in multicellular organisms, to enable the detailed investigation of physiological processes, this lack also threatens the translatability of results from test tube to humans.

In this thesis, *in vitro* models of the digestive tract were employed. Upper digestive processes were simulated by a series of enzymatic treatments of foods. The resultant digesta was then supplemented into validated models of the colon (Macfarlane, Macfarlane, & Gibson, 1998). More details may be found in Chapter 7.

2.4 Sample analysis

BAs are very important for the digestion and absorption of dietary lipids, maintenance of cholesterol homeostasis, modulation of gut microbiota and regulation of glucose, fatty acid and energy metabolism (Chiang, 2013; Hylemon et al., 2009; Kakiyama et al., 2014). Thus, there is a vast amount of literature on the analysis of BA in different biological materials. This section attempts to shed more light on the targeted metabolomic techniques used in BA analysis in faecal samples.

2.4.1 Metabolite analysis

Generally, the two major platforms used for metabolite identification and quantification are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectrometry. Both platforms were used in this thesis for the analysis of the principal analytes which were monomeric and oligomeric BAs, DCA and PDC respectively. This section will focus specifically on the methods used which were ultra-high-performance liquid chromatography coupled to MS (UHPLC-MS), matrix assisted liquid desorption/ionisation MS (MALDI-MS) and one-dimensional proton NMR (1D ^{1}H NMR). For a comprehensive library on the history of BA analysis, the reader is directed to the thematic review by Griffiths & Sjövall (Griffiths & Sjövall, 2010). For a more concise review of current platforms deployed for BA analyses, the reader is referred to Dutta and colleagues (Dutta, Cai, Gui, & Patterson, 2019).

2.4.1.1 *UHPLC-MS*

Of the aforementioned top two analytical platforms for metabolomic analyses, MS is the most widely applied in the analysis of BA in biological matrices (Griffiths & Wang, 2017). In particular, UHPLC-MS is the dominant analytical technique of choice when analysing BA (Dutta et al., 2019).

MS is a high throughput and powerful technique which may be used for metabolite identification (qualitative) and for quantifying known analytes (quantitative). An MS analysis involves producing gaseous ions from a sample (this process is called ionisation) which are then separated and sorted by their mass-to-charge ratios (*m/z*) in a mass analyser and detected by an ion detector system. A mass spectrum is obtained containing information on the *m/z* ratios of the ions and their relative abundances or intensities. A simplified schematic is presented in Figure 2-2.

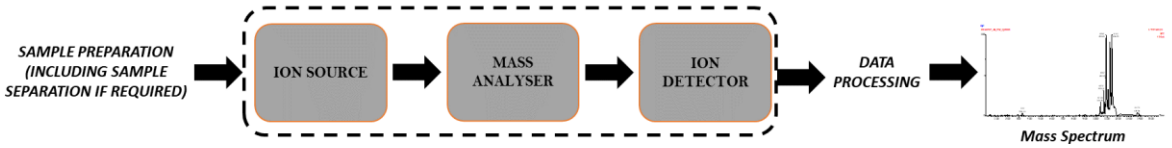


Figure 2-2: A simplified overview of a mass spectrometry analysis pipeline.

The ion source, mass analyser and ion detector are encased in a vacuum system (dashed line box).

Results obtained from MS analysis are highly dependent on the ionisation and detection efficiencies of the analyte ions. Indeed, ionisation efficiency may be reduced due to the simultaneous presence of more than one ion at the ionisation source. This leads to competition for ionisation and a consequent reduction in MS signal getting to the detector. Furthermore, isomeric bile acids cannot be delineated from each other using just MS. To overcome these problems, MS is usually coupled to chromatographic techniques such as gas chromatography (GC) or liquid chromatography (LC) which can separate compounds in the sample prior to introduction into MS.

A typical chromatographic system consists of the mobile phase and the stationary phase; and constituents of mixtures are separated based on their differential affinity for either of the phases. Thus, analytes of interests within complex biological materials can be separated from other unwanted compounds present in the matrix. Column reversed phase LC (RP-LC) is the most popular technique used in MS-based BA analysis (Cajka & Fiehn, 2014). In RP-LC, analytes are separated based on their hydrophobicity; with a non-polar column as the stationary phase and a polar solvent system as the mobile phase. In this set-up, the most polar analytes are the first to be eluted while the least polar are the most retained on the column (Lange, Ni, Criscuolo, & Fedorova, 2019). Evidently, the effectiveness of the process is largely determined by column chemistry and the mobile phase solvents used. These must be carefully selected considering the analytes of interest and the sample matrix to be analysed.

The coupling of LC to MS enhances sensitivity, as fewer compounds enter the MS per time, thus potentially reducing ion suppression. This also adds another dimension, the retention time, to the mass spectrum and may improve analyte identification. However, adding LC increases the length of analysis time and reduces throughput. Nevertheless, it remains the most popular technique used for the analysis of BA as well as other biological lipids. Advancements in the field have yielded in better quality instrumentation and shorter run times.

In MS, there are many ionisation methods such as electron ionisation (EI) (R. B. Silverman & Holladay, 2014), chemical ionisation (CI) (Poole, 2003), electrospray ionisation (ESI) (Fenn, Mann, Meng, Wong, & Whitehouse, 1989; Wong, Meng, & Fenn, 1988), atmospheric pressure chemical ionisation (APCI) (Byrdwell, 2001) and matrix assisted laser desorption/ionisation (MALDI) (Karas & Hillenkamp, 1988). ESI and MALDI are ionisation techniques that have been mostly applied to biological samples. Particularly, ESI is the most widely applied ionisation for BA analysis (Dutta et al., 2019).

ESI is a "soft" ionisation technique used for ionising non-volatile, thermally labile and high molecular weight analytes at atmospheric pressure. (Roberts, McCombie, Titman, & Griffin, 2008; R. B. Silverman & Holladay, 2014). It is "soft" because it uses a "gentler" electrical and not "harsh" heat energy to generate gaseous ions from the sample; the energy applied, although at relatively high voltages (e.g. 2 – 6 kV), is usually just enough to create adducted (e.g. deprotonated $[M-H]^-$, protonated $[M+H]^+$) molecular ions (Greaves & Roboz, 2014). For a detailed reading about ESI, the reader may refer to review papers by Andres Bruins (Bruins, 1998) and Banerjee *et al* (Banerjee & Mazumdar, 2012).

The ESI process is summarised in Figure 2-3. Briefly, in ESI, analyte ions are generated by dissolving sample in a suitable solvent; then the solution (or eluent, if interfaced with LC) is introduced into the MS by spraying; a slightly elevated temperature or a stream of drying gas (e.g. nitrogen) is applied to dry the droplets and break up clusters formed as the carrier solvent is evaporated, and, due to Coulomb repulsion, multiply charged ions are released in the gas phase (C. S. Ho et al., 2003). Thus, ESI allows for analysis in multiple modes (positive and negative) and also serves as a versatile interface that allows LC coupling to MS (Poole, 2003).

However, the efficiency of the process is affected by solution parameters such as the nature and suitability of solvent as well as the structure, conformation and concentration of analytes. For example, more polar and stronger ionising analytes in the solution are more favoured in ESI than less polar analytes and therefore can reduce the signals of the latter (Greaves & Roboz, 2014). The presence of non-volatile non-analyte ions in the sample matrix at the spraying stage may also result in ion enhancement or suppression (King, Bonfiglio, Fernandez-Metzler, Miller-Stein, & Olah, 2000) even when these interfering species cannot be observed on the mass spectrum (Roberts et al., 2008). Thus, extra care must go into adequate sample preparation to remove as much of these matrix effects as possible before the sample extract is introduced into the ionisation source. Sample preparation is discussed later in this chapter. Another strategy is to modify LC conditions in a way that the analytes of interest may elute at different times to the interfering ions (Annesley, 2003).

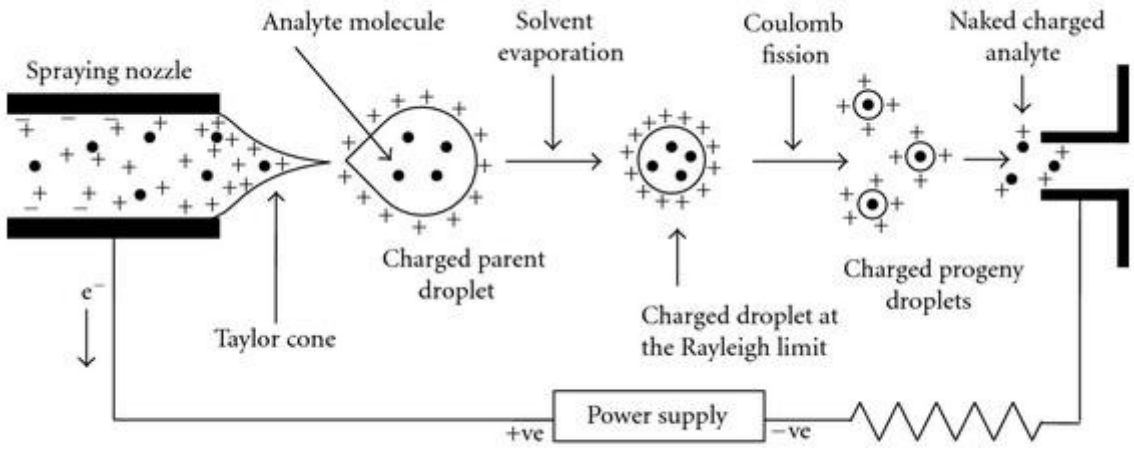


Figure 2-3: A diagrammatic description of electrospray ionisation process.

Adapted with permission from Banerjee & Mazumdar (2012) (Banerjee & Mazumdar, 2012).

After ionisation, analyte ions are then made to travel through a mass analyser where they are separated according to their *m/z* ratio. A range of mass analysers exist for different applications. For a targeted metabolomic assay such as BA quantification, the tandem triple quadrupole analyser (QqQ or TQ for short) is utilised. As the name suggests, a TQ system is made up of three quadrupoles which are arranged linearly. Each quadrupole is a collection of four parallel metal (molybdenum) rods arranged equidistantly from each other in a square formation, with each pair of rods connected electrically (Greaves & Roboz, 2014). A radio frequency alternating current (RFAC) voltage is superimposed with a direct current (DC) voltage and mass-dependent separation of ions occurs when the ions oscillate within the electrical field generated between the rods (C. S. Ho et al., 2003). The amplitude of oscillation of an ion is related to its *m/z* ratio and can be controlled by manipulating the RFAC and DC voltages. In a TQ, the first quadrupole (Q1) is used to select desired ions by setting the RFAC and DC voltages to allow selected ions to travel through while undesired ions are filtered out. The selected "parent" or precursor ion in Q1 travels to Q2 where a collision energy is applied to generate smaller "daughter" or product ions which are scanned and monitored in Q3. Parent and daughter ions can thus be pre-determined, allowing for highly specific and selective quantification of analytes of interest. This type of monitoring with pre-selected precursor and product *m/z* ratios is called selected reaction monitoring (SRM) when applied to a single product ion or multiple reaction monitoring (MRM) when applied to more than one product ion. When dealing with ions that may not easily fragment into daughter ions, the system can be set to select the parent ion at both Q1 and Q3 (practically, most MS systems do not allow for

same masses to be inputted. This is circumvented by setting the product ion at a mass minimally different from the parent ion, e.g. DCA with *m/z* 391.2 may be set as 391.2 > 391.19). This is termed selected ion monitoring (SIM). Both SIM and MRM were used in this thesis to monitor and quantify BAs.

Following ionisation, the separated ions arrive at an electron multiplier detector system where they create a cascade that results in the multiplication of the signal by up to a million-fold (Greaves & Roboz, 2014). The amplified signal is recorded and digitally processed to produce the three-dimensional mass spectrum containing retention time (from LC), *m/z* and relative abundance information for respective analyte ions. In quantification experiments, a series of standard solutions are prepared, analysed and used to estimate analyte concentrations in samples.

2.4.1.2 *MALDI-MS*

As with any MS system, MALDI-MS is made up of the three components namely the ionisation source (MALDI), mass analyser (almost always is a time-of-flight (ToF) mass analyser) and detector (Greaves & Roboz, 2014). MALDI is another "soft" ionisation technique that is used for higher molecular weight analytes, for example in the quantification of biopolymers (Hillenkamp, Karas, Beavis, & Chait, 1991). The sample is mixed with a special matrix, usually an aromatic acid, spotted and dried on a metal plate (Roberts et al., 2008). The mixture is irradiated with an ultraviolet laser pulse, the matrix is ionised and both matrix and analyte ions are evaporated (Greaves & Roboz, 2014), generating gas-phase ions with little to no fragmentation (R. B. Silverman & Holladay, 2014). A schematic diagram is shown in Figure 2-4.

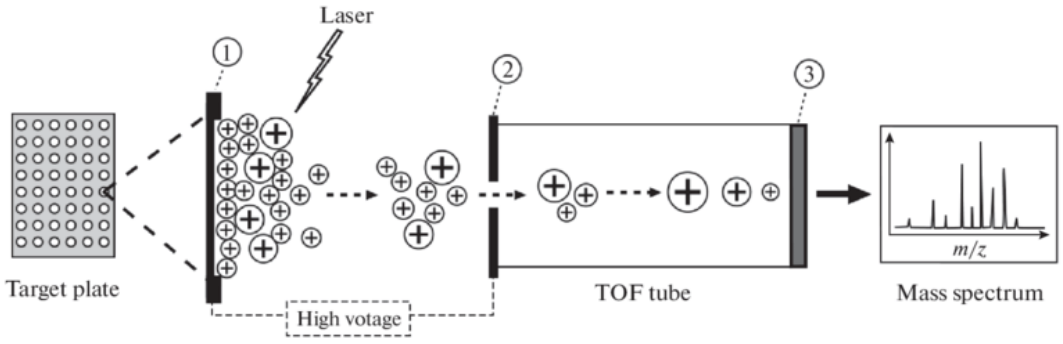


Figure 2-4: Schematic representation of MALDI-ToF system

Adapted with permission from Velichko & Pinevich (2019)(Velichko & Pinevich, 2019)

Unlike the quadrupole analyser which scans for selected ions by *m/z* , the ToF mass analyser is a non-scanning technique where the full mass spectrum of the sample is acquired (Williamson & Bartlett, 2007). The ToF consists of an accelerator, a field-free region called the flight tube and a detector. Ions are generated and introduced into the flight tube, accelerated by applying a high voltage electrical field, and the pulsed ions are then allowed to travel in the field-free tube to the detector (Greaves & Roboz, 2014). The velocity, and thus time taken to reach the detector, of the ions is proportional to their mass. Therefore, at a constant voltage and kinetic energy, the smaller ions will travel and arrive at the detector faster than the larger ions (Pitt, 2009). The *m/z* ratios can be calculated from the time of flight and used to produce a mass spectrum. The ToF analyser is therefore almost unrestricted in mass range and is used for accurate mass determination. The pulsed ionization also allows for easy coupling to MALDI (Rockwood, Kushnir, & Clarke, 2018). However, MALDI, like ESI, may also suffer from ion suppression (Petković et al., 2001) which may be ameliorated by modifying the matrix (Gharahdaghi, Kirchner, Fernandez, & Mische, 1996) or the pressure conditions (G. Li et al., 2019).

2.4.1.3 *^{1}H NMR*

NMR, although not used for BA quantification in this thesis, has been seldomly applied to BA analysis (Nagana Gowda, 2011; Waterhous, Barnes, & Muccio, 1985). Although highly reproducible, NMR has had limited application in this area as it is not as sensitive as MS-based methods (Dutta et al., 2019). NMR, however, is excellent for structural elucidation, and was used for PDC structure analysis. NMR is based on complex quantum physics principles. However, the author has attempted to present a simplified explanation in the following paragraphs. There is vast literature already on the topic. For more reading, the reader is suggested to refer to the NMR handbook by Andrew Lane (Lane, 2012).

The underlying principle of NMR is based on deducing chemical information from the magnetic properties of atomic nuclei in samples. Subatomic particles, including the atomic nucleus, possess a fundamental intrinsic property called “spin”, denoted by *I*. Nuclear spin is quantized, meaning it can only be measured in discrete integer or half-integer units (e.g. 1/2, 1, 3/2, 2, 5/2, 3…up to 8). Nuclei with odd number of protons or/and odd number of neutrons, such as ^{1}H, ^{2}H, ^{13}C, ^{14}N, ^{15}N, ^{23}Na, ^{25}Mg, ^{31}P, ^{33}S, ^{39}K, ^{77}Se have an overall/net spin while nuclei with even number of both protons and neutrons such as ^{14}C and ^{16}O have no net spin. These nuclei with no spins ($I = 0$) cannot absorb and emit electromagnetic radiation or undergo "resonance" when placed in a magnetic field, and therefore cannot be detected by NMR. ^{1}H is the most widely used nuclei of biological interest in proton NMR (^{1}H NMR) experiments (Lane, 2012).

In particular, 1D ^{1}H NMR provides a quick and simple method of analysing samples (Emwas et al., 2019).

Briefly, when a sample is placed in a magnetic field and electromagnetic radiation is applied at the appropriate frequency, the atomic nuclei in the sample can absorb the applied energy. This frequency, called Larmor frequency, is characteristic of each nucleus, therefore for instance ^{1}H and ^{13}C will have characteristic frequencies. Furthermore, the Larmor frequency, also called resonance frequency, is dependent on the chemical environment of the nucleus. This is because nuclei may be affected by the fields of surrounding atomic particles (nuclei and electrons), thus, chemically different nuclei will have different electronic environments and will produce different signals. For instance, in ^{1}H NMR, the methyl and hydroxyl protons in the methanol molecule will absorb energy at different frequency because their chemical environments are dissimilar.

On removal of the magnetic field, the excited nuclei return to original state (equilibrium) and in the process (called relaxation) emit energy which creates a signal that can also be measured. The time it takes and the manner in which the nucleus emits energy during this relaxation process can also provide chemical information about the nuclei. Signals generated by resonance and relaxation are recorded and processed into an NMR spectrum. The 1D ^{1}H NMR spectrum of a sample thus is composed of the superimposed ^{1}H NMR spectra of all analytes in that sample. NMR signals are measured as a “chemical shift” relative to a standard compound and their intensities are recorded. All of this information contained in the spectrum can then be processed to elucidate the structure or quantify the analytes (Dona et al., 2016).

NMR is non-destructive unlike LC-MS, highly reproducible, requires minimal sample preparation and is thus very high throughput (Nicholson & Lindon, 2008). However, it is not as sensitive as MS meaning metabolites have to be present in much higher amounts to be detected by the technique (Lane, 2012).

2.4.1.4 *Sample preparation (Analyte extraction)*

Although these platforms discussed above are very sophisticated and sensitive, their performance may be compromised by the quality of the sample extraction procedure. Sample preparation is thus a significant and important step in the analytical process.

A sample preparation procedure should be able to transform the sample (if required) from its initial state to a state compatible with the analytical platform being utilised for analysis (Gao, Ho, & Williams, 2013). For example, the sample to be introduced into the UPLC-(ESI)-MS system must be in liquid form, therefore solid samples have to be processed appropriately by

a homogenization method (e.g. bead-beating) to increase the surface area and ensure the extraction solvent is able to mix properly with samples. Furthermore, the sample must be completely dissolved in a carrier solvent that is dissolves the analytes of interest and is compatible with all components of the analytical system (column in LC, ionization source, etc).

Biological samples are usually very complex matrices with interfering analytes. An effective extraction procedure must not only successfully extract the analyte of interest without a loss of sample integrity, it must also easily remove interference (i.e. reduce matrix effect) (Sobotka, 1938 as quoted by Griffiths & Sjövall, 2010). Matrix effects in LC-MS may be mitigated by spiking stable isotope labelled internal standards into the sample (Tan & Awaiye, 2013). This standard must be similar to the endogenous analyte of interest and must behave similarly (elute at the same time, etc) for this to be effective (Annesley, 2003).

Lastly, the sample preparation procedure must also ensure the final concentration of the analyte of interest is within detectable limits of the instrument of analysis; which may mean concentrating or diluting the sample extract as required (G. Liu & Aubry, 2013).

Sample preparation methods were considered in the development of an analytical method for detecting and measuring PDC in faecal material. More details about this are given in Chapters 3 and 4.

2.4.2 Microbiome Techniques

Metagenomic techniques include a plethora of sequencing techniques encompassing whole-genome sequencing, metatranscriptomics or amplicon-based analyses. In this thesis, both targeted amplicon-based and whole-genome metagenomic approaches were utilised for studying the microbial ecology and diversity in human faecal samples.

2.4.2.1 *16S rRNA gene sequencing*

Amplicon-based metagenomic analyses usually involve targeting a marker gene and exploring the diversity and abundance of that gene in a sample, for example the 16S rRNA gene in prokaryotes and the 18S rRNA gene in eukaryotes. A typical pipeline will involve first the extraction of DNA of sample, amplification of marker gene of interest and sequencing of the amplicons. Downstream data analyses will be discussed in further sections.

The prokaryotic 16S rRNA gene is the most commonly targeted gene marker because it is highly conserved among *Bacteria* and *Archaea*. It is a component of the 30S ribosomal subunit which binds to the Shine-Dalgarno sequence located on messenger RNAs which initiates

polypeptide synthesis in the bacteria. As such, all bacterial genomes encode a 16S rRNA gene making it a reliable marker for bacterial presence in a sample. The 16S rRNA gene is approximately 1 500 base pairs (bp) long and consists of conserved regions interspersed by hypervariable regions (Figure 2-5 and 2-6). There are nine of these variable regions (V1-9), containing 30-100 bp each, which contain specific signature sequences that can be exploited to confidently distinguish between different bacterial genera but not species (Deurenberg et al., 2017).

In a typical analysis, one or more variable regions will be selected to be amplified by polymerase chain reaction (PCR) and sequenced. PCR is carried out using primers that were designed using the conserved regions bordering the selected hypervariable region(s). The selection of variable regions to amplify should be informed by consulting literature to find out which has been used for similar investigations previously and which is optimal for the group of bacteria being studied. This is because sequencing different regions have varying resolution power across the taxonomies (Yang, Wang, & Qian, 2016). For instance, the V1-V3 region has been shown to distinguish between members of the kingdom *Bacteria* but not *Archaea*, while V3-V5, although can distinguish both kingdoms with high resolution, cannot resolve between some key families at genus level (Davidson & Epperson, 2018).

The amplified reads are then sequenced using next-generation sequencing (NGS) techniques. NGS is a modern, high throughput sequencing technology that is faster, more sensitive and covers a broader range of genomic variation than the conventional Sanger sequencing (Behjati & Tarpey, 2013). However, it is very expensive to run and requires highly specialist expertise. Examples of NGS platforms available are the Illumina, ion torrent, Roche 454 and so on; the most commonly used being Illumina. Illumina sequencers are capable of producing short reads of up to 300 bp. These may be single-end or paired end, the latter providing more accurate sequences and consequently taxonomic identification. Finally, sequences are pre-processed and assigned to taxonomies by matching sequences to databases (discussed below).

An amplicon-based technique like 16S rRNA gene sequencing has the advantages of being culture-independent which allows for an unrestricted comprehensive study of the microbiome. However, the 16S rRNA gene sequencing technique is plagued with usual PCR limitations, a most crucial one being the differential amplification efficiency due to differential affinity for the primer used which may underestimate/overestimate the presence of certain bacterial groups at the expense of others. As already hinted above, the different variable regions differ in resolution between taxonomies which may also inflate the presence of certain bacterial groups.

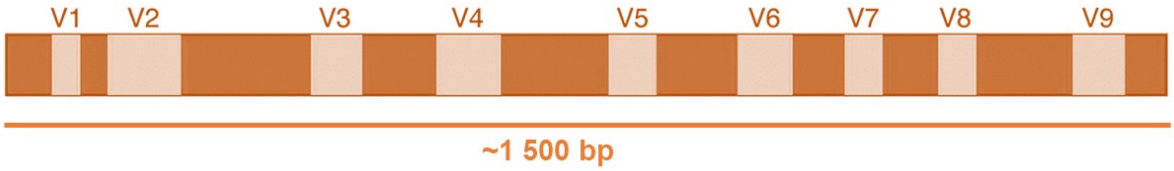


Figure 2-5: A simplified illustration of the 16S rRNA gene showing conserved (orange) and hypervariable (grey) regions.

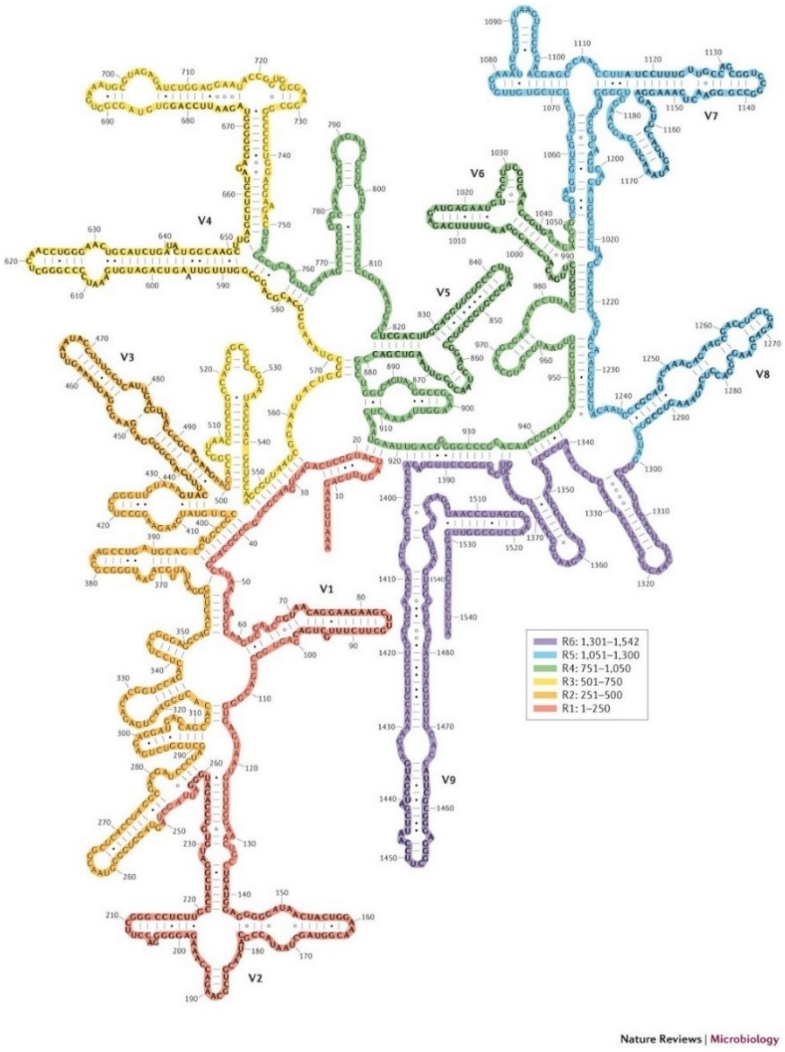


Figure 2-6: Secondary structure of the 16S rRNA of Escherichia coli

(Adapted with permission from Yarza et al (2014).(Yarza et al., 2014)

Fragment R1 (in red) includes regions V1 and V2; fragment R2 (in orange) depicts region V3; fragment R3 (in yellow) depicts region V4; fragment R4 (in green) includes regions V5 and V6; fragment R5 (in blue) includes regions V7 and V8; and fragment R6 (in purple) depicts region V9.

Furthermore, results obtained from amplifying different hypervariable regions are different, reducing the generalisability of results. Lastly, unless accompanied by quantitative PCR, only relative and not absolute abundance is obtained from the analyses, meaning if one group "increases" another group must "decrease", so results must be interpreted with caution as the actual numbers of bacteria are unknown (Jo, Kennedy, & Kong, 2016).

2.4.2.2 Shotgun sequencing

Unlike 16S rRNA gene sequencing, shotgun sequencing (SGS) allows for broad sequencing and analysis of the entire metagenomic microbial diversity present in a sample. DNA from both prokaryotic and eukaryotic organisms may be sequenced in parallel, as such, cross-genome interferences may also be explored, although it may miss out microbes at lower abundances. A typical SGS pipeline involves extraction of DNA, random fragmentation of purified DNA, adaptor ligation to fragments, DNA fragment sequencing, genome assembly and further bioinformatic processing.

Since it sequences the whole genome, it is more specific, sensitive and it can resolve confidently to the species and strain level, although accuracy may be limited (Dai, Rhoads, Edwards, & Pruden, 2018). This allows for functional profiling of the gut microbiota. Furthermore, results from different studies may be compared. However, it is much more costly than 16S rRNA and it also has the challenge of reporting compositional, not absolute, numbers for microbial abundance (Quince, Walker, Simpson, Loman, & Segata, 2017).

2.5 Data and statistical analysis

Data analysis was carried out with respect to the different types of data. The different data types were also combined and analysed.

2.5.1 Analysis of metabolite data

Data generated from UHPLC-MS BA analysis were pre-processed on relevant software and tables of concentrations were obtained. A combination of univariate and multivariate statistical methods were utilised in further data analysis.

Univariate statistics deal with one variable at a time, while multivariate statistics involve the simultaneous analysis of variables. Univariate methods employed include student's *t* tests and analysis of variance (ANOVA) or their non-parametric equivalents to test for statistical significance. Normality has to be assessed to determine if data transformation is required and whether parametric or non-parametric techniques should be used. Due to the large number of variables generated in metabolomic datasets, these univariate tests of significance are corrected for multiple comparisons to reduce the possibility of false positives (type I error), although this then increases the probability of obtaining false negatives (type II error). Methods for correction (such as Bonferroni, Hochberg and Benjamini, Bonferroni–Holm, Westfall–Young, etc) may be used to adjust *p* values (Saccenti, Hoefsloot, Smilde, Westerhuis, & Hendriks, 2014). It is important to note that multiple testing correction methods do not account for the collinearity or correlations between the metabolites as the methods are based on the assumption that the variables are independent of each other (Hendriks et al., 2011).

The main multivariate statistical technique used in this thesis was the principal component analysis (PCA). PCA is an unsupervised chemometric method that reduces the dimensionality of the data to produce "components" which explain for variation and emphasises patterns in the data. It does this without using any metadata information for grouping the data. The first, or "principal" component explains for the greatest amount of variation, the second explains for the remaining variation that is orthogonal (uncorrelated) to the first, and so on (Steuer, Morgenthal, Weckwerth, & Selbig, 2007). Graphically, this can be presented as scores plots, which display the projection of observations/cases, and the loading plots which display to the projection of the variables. PCA depends on scaling of data and can be heavily influenced by few data with significantly greater magnitude than the others in the dataset (Lever, Krzywinski, & Altman, 2017).

2.5.2 Analysis of microbiome data

16S rRNA gene sequencing reads are processed using bioinformatic pipelines and algorithms that abound in literature but can be mainly categorised as operational taxonomic unit (OTU) clustering-based methods or amplicon sequence variance (ASV)-based methods (Prodan et al., 2020). In OTU methods, the researcher determines the level of dissimilarity (usually set at 3%) between sequences that is a safe guess of the variation due to taxonomic differences and not due to random sequencer noise (Caruso, Song, Asquith, & Karstens, 2019). Sequences that pass this set threshold are clustered into a single OTU. However, this leads to grouping together many bacteria belonging to vastly different taxonomies that have up to or more than 97% similarity in their 16S rRNA genome but may have different clinical impact/relevance (Mysara et al., 2017). ASV methods on the other hand are able to distinguish between sequences differing by as few as one nucleotide based on an error model constructed from the sample set (Callahan, McMurdie, & Holmes, 2017). This error model is built on the assumption that sequences of true biological origin are more likely to be repeatedly observed in a sample set than spurious sequences are (Callahan et al., 2017; Rosen, Callahan, Fisher, & Holmes, 2012). Furthermore, ASV methods have been shown to be more sensitive and reliable (Caruso et al., 2019; Prodan et al., 2020) but may be limited by the resolution of the NGS platform used for sequencing.

For the 16S rRNA analyses reported in this thesis, an ASV method, Divisive Amplicon Denoising Algorithm 2 (DADA2) method was used to process metagenomic results discussed in Chapter 6 (Callahan et al., 2016).

Following pre-processing, sequences are assigned taxonomies by matching or aligning to databases, examples of which are SILVA (Yilmaz et al., 2013), Ribosomal Database Project (RDP) Classifier (Qiong Wang, Garrity, Tiedje, & Cole, 2007) and Green Genes (D. McDonald et al., 2012). However, discrepancies may exist between databases (Balvočiūtė & Huson, 2017), therefore it is advised to compare results obtained from these databases especially when using OTUs (Beiko, 2015).

The SGS data was processed by collaborators according to a pipeline explained in detail in Appendix 5.0.

2.5.3 Integration of metabolite and microbiome data

While either of metabolomic and microbiome data generated may respectively describe the host metabolic processes and the gut microbial eco-diversity, the goal is to integrate both datasets in order to provide deeper insights into the interaction between host metabolism and gut microbial genome, and the impact of this interaction on host overall health.

Strategies for integrating these different data types are scarce due to a lack of computational and statistical tools that can handle the peculiarities of the individual datasets. Limitations arise when coupling sparse and relative microbiome data to continuous metabolite data. Most conventional statistical methods are based assumptions suitable for independent continuous data. This is a burgeoning area in the field, and many approaches are being developed such as correlation-, regression- and network-based methods (Jiang et al., 2019; Qing Wang et al., 2019).

Univariate correlations between metabolite and microbiome variables, such as Spearman correlation which was used in this thesis, are a simple and commonly utilised approach (Le Chatelier et al., 2013; Mao, Huo, & Zhu, 2016; Theriot et al., 2014). However, results have to be interpreted in context to derive accurate mechanistic insights into the biological processes being explored (Chong & Xia, 2017).

2.6 Conclusion

In conclusion, this chapter provided a brief description of the various state-of-the-art technologies employed in this PhD project. These analytical platforms were selected based on their suitability and superiority for assessing the biological analytes of interest as well as the feasibility of adopting them. Their underlying principles, strengths and limitations were discussed in order to assist the reader in interpreting the results discussed in subsequent chapters of this thesis.

3. QUANTIFYING POLYDEOXYCHOLIC ACID: VALIDATION OF AN ULTRA-PERFORMANCE LIQUID CHROMATOGRAPHY MASS SPECTROMETRY METHOD FOR BILE ACID ANALYSIS IN FAECAL SAMPLES

3.1 Introduction

Given the potential role for the formation of polydeoxycholic acid (PDC) in human physiology and health, a high-throughput analytical method was required to accurately detect and quantify PDC in relevant biofluids.

Previously, PDC has been quantified using a multiple step extraction process and a subsequent enzyme-based assay. First, lyophilised faecal samples were extracted using an aqueous/polar solvent and centrifuged to yield a supernatant fraction containing the "monomeric" bile acids (BA) and a pellet containing "polymeric" BA. The supernatant fraction was freeze-dried while the pellet was dried and hydrolysed using a strong alkaline solution to "saponify" the polyester into its monomeric units, deoxycholic acid (DCA) (see Figure 3-1). "Monomeric" and "polymeric" extracts were purified on C18 SPE columns, dried down and resuspended in reaction buffer for total bile acid quantification using a 3α-hydroxysteroid dehydrogenase assay (Benson et al., 1993). However, this method had a number of drawbacks including time-consuming sample extraction, insensitivity of the enzyme assay below a relatively high limit of detection, over-estimation of BA abundance due to the reaction of 3α-hydroxysteroid dehydrogenase with other nonacidic sterols containing a hydroxyl group in the 3α position and the inaccuracy of the estimation of concentration based on assumed proportions (Griffiths & Sjövall, 2010; Yanan Liu, Rong, Xiang, Zhang, & Liu, 2018).

To improve the quantification of PDC in faecal samples and overcome the issues identified with the previous method, a new approach for PDC quantification was developed. To achieve this, an existing and validated high-throughput targeted reverse phase ultra-high performance liquid chromatography tandem mass spectrometry (UHPLC-MS)-based method was used (referred to as the EL method for the remainder of this chapter) (Sarafian et al., 2015). This method has greater specificity (distinguished BA species from other lipid species), selectivity (resolved individual BA peaks) and sensitivity (measured as low as 0.5 nM) compared to the 3α-hydroxysteroid dehydrogenase assay.

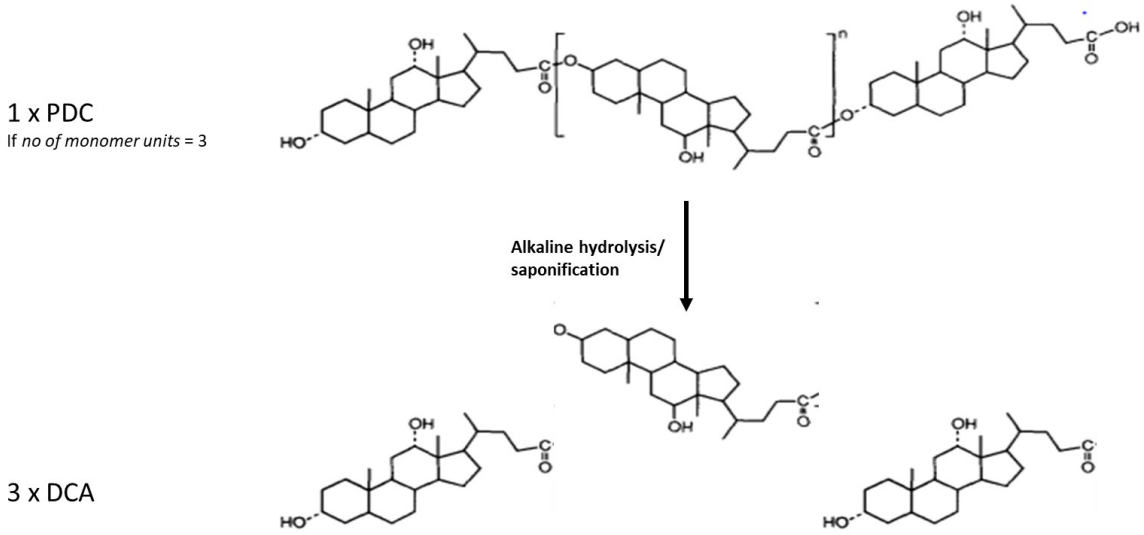


Figure 3-1: A schematic diagram showing the principle of action for alkaline hydrolysis of polydeoxycholic acid

Strong alkali, in the presence of heat, will break down the ester bonds in between the consecutive monomeric deoxycholic acid (DCA) moelcules in the polymeric polydeoxycholic acid (PDC) molecule. An example is given with a PDC trimer which will yield three DCA monomers when saponified.

Although the EL method was previously validated in urine, plasma and serum, it had not yet been validated in faecal samples. To apply the method to faecal samples, a partial validation of the method was therefore warranted. This was recommended by in-house protocols based on U.S. Food and Drug Administration (FDA) guidelines, which stated that a partial validation was to be carried out when applying a previously validated bioanalytical method in a new biofluid/matrix (Center for Drug Evaluation and Research, 2013). Following validation, the method was subsequently applied to the analysis of clinical faecal samples obtained from a cohort of vegans and omnivores. This sample set was chosen for the proof-of-concept application because the samples were to be further analysed for PDC, and as such, the data generated would contribute to the research questions explored in this thesis. Also, the sample set was considered appropriate at the time because to the best of the author's knowledge, fecal BA excretion had not been investigated in habitual vegans.

In summary, a novel approach for PDC quantification in the biological sample of interest (*i.e.* faeces) was developed using a UHPLC-MS method in combination with an enzymatic step to liberate the monomeric DCA units. The full optimisation process was carried out in two major steps; first, a validation of the targeted BA profiling method in faecal extracts, and then a further streamlined method development for PDC detection and quantification. This chapter describes the experiments carried out as part of the initial step, *i.e.* the partial validation of the EL method.

3.2 Aim and objectives

3.2.1 Aim

The aim of this chapter was to partially validate and optimise the EL method in faecal samples.

3.2.2 Objectives

To validate the EL method for faecal samples the following parameters were assessed:

1. linearity of calibration ranges;
2. intra-assay accuracy and precision;
3. selectivity;
4. analyte recovery; and
5. matrix factor

To optimise the EL method for faecal samples the following steps were taken:

6. To adapt sample preparation for targeted analysis;
7. To apply validated method to clinical samples

As far as practical, the validation was carried out in accordance to FDA guidelines.

3.3 Methods

3.3.1 Chemicals and reagents

Lithocholenic acid, 5-cholenic acid-3β-ol, 3-ketocholanic acid, isolithocholic acid, allolithocholic acid, 9(11), (5β)-cholenic acid-3α-ol-12-one, 3,7-diketocholanic acid, 3α-hydroxy-7-ketolithocholic acid, 3α-hydroxy-12-ketolithocholic acid, 3,7,12-dehydrocholic acid, 3α-hydroxy-7,12-diketocholanic acid, 3,7,12-glycodehydrocholic acid, glycocholic acid, taurolithocholic acid, tauroursodeoxycholic acid, 3,7,12-taurodehydrocholic acid, 5α-cholanic acid-3,6-dione, 5α-cholanic acid-3α-ol-6-one, glycoursocholanic acid, glycohyodeoxycholic acid, glycoursodeoxycholic acid, tauroursocholanic acid, 8(14),(5β)-cholenic acid-3α,12α-diol and taurohyodeoxycholic acid were purchased from Makaira Ltd (London, England). Hyodeoxycholic acid, chenodeoxycholic acid, glycochenodeoxycholic acid, taurochenodeoxycholic acid, ursodeoxycholic acid, deoxycholic acid, glycodeoxycholic acid, taurodeoxycholic acid, lithocholic acid, hyocholic acid, cholic acid, taurocholic acid, ursocholanic acid, murocholic acid, 5β-cholanic acid-3β,12α-diol, 3α,12α,23-nordeoxycholic

acid, 3,6-diketocholanic acid, 3,12-diketocholanic acid, 5β-cholenic acid-7α-ol-3-one, isodeoxycholic acid, 3α-hydroxy-6,7-diketocholanic acid, 5β-cholanic acid-3α, 6α-diol-7-one, 3-dehydrocholic acid, 12-dehydrocholic acid, ω-muricholic acid, β-muricholic acid, α-muricholic acid, glycolithocholic acid, glycohyocholic acid, and sodium salts of taurohyocholic acid, tauro-β-muricholic acid, tauro-α-muricholic acid and tauro-ω-muricholic acid were purchased from Steraloids Inc., Newport (Rhodes Island, USA).

Deuterated bile acid standards were used as internal standards (DIS). Ursodeoxycholic acid-d4, β-hyodeoxycholic acid-d4, chenodeoxycholic acid-d5, glycolithocholic acid-d4, glycodeoxycholic acid-d4, glycoursodeoxycholic acid-d4, taurolithocholic acid-d4, taurochenodeoxycholic acid-d4, tauroursodeoxycholic acid-d5, taurodeoxycholic acid-d4, and taurocholic acid-d5 were purchased from Medical Isotopes Inc., Pelham (New Hampshire, USA) while Lithocholic acid-d4, cholic acid-d4, glycocholic acid-d4, deoxycholic acid-d4 and glycochenodeoxycholic acid -d4 were purchased from Steraloids Inc., Newport (Rhode Island, USA).

UHPLC-MS grade Chromasolv^{TM} solvents acetonitrile (ACN) and 2-propanol (IPA) from Honeywell International Inc (USA) and UHPLC-UV grade water (H_2O) purchased from Fischer Scientific (Loughborough, UK) were used for preparation of mobile phases and sample extraction solutions. Eluent additives ammonium acetate and acetic acid (used for adjusting pH of mobile phase A) were purchased from Sigma-Aldrich (Dorset/Gillingham, UK).

3.3.2 Preparation of calibration curve and quality control concentrations

A calibration range of 0.5 nM – 5 μM as previously reported by Sarafian and colleagues (Sarafian et al., 2015) was adopted. The lower limit of quantification (LLOQ) of each BA standard was accepted as the lowest calibration curve (CC) concentration with a signal-to-noise ratio greater than 5 that could be quantified with a coefficient of variation (CV) below 20%. The upper limit of quantification (ULOQ) of each BA standard was accepted as the highest concentration of the CC that was detected at an intensity below the detector saturation level. Five quality control (QC) solutions were prepared as follows, LLOQQC (QC corresponding to the LLOQ), Low QC (LQC), Medium QC (MQC), High QC (HQC) and ULOQQC (QC corresponding to the ULOQ). LQC concentration was calculated as three times LLOQQC while MQC and HQC were 40% and 80% of ULOQQC respectively. Acceptance criteria for each CC and QC concentration were a signal-to-noise ratio greater than 5 and a deviation of less than 25% from the nominal concentration.

3.3.3 Preparation of stock and working solutions for validation experiments

Parent 1 mg/ml stock solutions of each of the 57 BA standards were prepared in H_2O: ACN: IPA 2:1:1 solution. A working "cocktail" solution of all the BA standards was then prepared from the parent stocks to yield a final concentration of 10 µM for each BA. Standard calibration curve (CC) solutions were prepared by serial dilutions of BA cocktail mix into a 13-point CC. A duplicate stock cocktail was also prepared and diluted as required with the 2:1:1 solution to produce suitable quality control (QC) solutions.

Similarly, 1 mg/ml parent stocks of the 16 labelled DIS were prepared in H_2O: ACN: IPA 2:1:1 solution. A working cocktail mix of all DIS was then prepared and spiked into CC, QC and sample extracts as appropriate. The DIS concentration was kept constant in all CC and QC solutions.

3.3.4 Linearity

Linearity was assessed for each BA over their respective CC ranges. CCs were generated by plotting the ratio of the peak area of the respective BA standard to that of its corresponding DIS against the nominal concentration. The line of best fit was determined by linear regression using a 1/x weighting factor. The acceptance criterion for linearity was a correlation coefficient R^2 value equal to or greater than 0.99.

3.3.5 Intra-day accuracy and precision

The intra-day accuracy and precision were determined for each BA by analysing six replicates of their respective LLOQQC, LQC, MQC, HQC and ULOQQC. At least four out of the six replicates were to be valid for the QC to be accepted. Accuracy was defined as the average of the calculated concentrations of the QC replicates expressed as a percentage of the nominal concentration. Cut-off for percentage error of accuracy was set as 15% (or 20% for LLOQQC). Precision was computed as the standard deviation expressed as a percentage of the average QC concentration. Intra-day precision was to be within +/- 15% (or 20% for LLOQQC) to be accepted.

3.3.6 Analyte Recovery

Because BAs are metabolites endogenous in the biofluid to be analysed (*i.e.* faecal samples), a true "blank" analyte-free matrix was thus unavailable. Instead, the recovery of an internal

standard (DIS) spiked into the analytical samples was investigated. DIS recoveries were determined in *n* = 3 differently sourced samples. Samples were spiked in with the DIS cocktail at the start of sample preparation (pre-spiked) and after sample preparation (post-spiked). Recovery was calculated as response of pre-spiked sample divided by the response of post-spiked samples at low (QC1) and high (QC2) final concentrations of 10 nM and 500 nM respectively.

3.3.7 Matrix Factor Analysis

Similarly, analyte matrix factor could not be calculated because there was no “blank” matrix free of endogenous BAs. Instead, internal standard matrix factor was assessed. Reference single blank solutions were prepared in 2:1:1 from the DIS stock mix at QC1 (10 nM) and QC2 (500 nM) final concentrations. Test samples (*n* = 3) for matrix factor analysis were extracted and spiked in with DIS after sample preparation (post -spiked) at both QC1 and QC2 concentrations. Matrix factor was calculated as the response of post-spiked samples (response in presence of matrix ions) divided by the response of corresponding reference single blank solution (response in absence of matrix ions). A ≤5% difference between reference solutions and test sample was acceptable.

3.3.8 Method Selectivity

To ascertain the selectivity of the method, interference was assessed. To evaluate analyte-analyte interference, solutions of each BA were prepared at its ULOQ concentration. Three aliquots of each solution were assessed for interference due to that BA at the retention time (RT) of other BAs. Any response at the RT of other BAs was compared to the LLOQ response of the BA being assessed. The interference was expressed as a percentage of the LLOQ and acceptable below 20%. To assess matrix-internal standard interference, *n* = 6 faecal samples were processed without spiking in DIS. Any response at the RT of each DIS was compared to the expected response of QC1 (10 nM) and of QC2 (500 nM).

3.3.9 Collection of clinical faecal samples from Vegans vs Omnivores

Clinical samples were obtained from a cross-sectional study of vegans and omnivores already described elsewhere (T. H. Hansen, Kern, et al., 2018; T. H. Hansen, Madsen, et al., 2018; Kristensen et al., 2015). Briefly, adults aged 18-65 years were recruited onto the study through print and online media. Eligible participants were weight stable for a minimum of two months

prior to the study, not pregnant or lactating, had no known gastrointestinal disease and had not taken antibiotics within 3 months prior to the study. Vegan participants were to have adhered to a vegan diet for at least 1 year. Participants were provided with Sarstedt sampling tubes to collect their faecal samples at home. Samples were kept at -18 C and transferred on dry ice within 48 hours for storage at -80ºC until analysis. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethical Committee for the Capital Region of Denmark (ethical approval ref: j no. H-3-2012-145). All participants gave their written informed consent. Participants were recruited from November 2013 to November 2014.

3.3.10 Preparation of faecal samples for analysis

The faecal sample preparation protocol used by Mullish and colleagues (Mullish et al., 2018) for monomeric BA profiling by untargeted LC-MS was adapted for this analysis. Approximately 1 g of stool was weighed into a 2 ml Eppendorf tube, which was sealed with parafilm wax. The parafilm was pierced slightly and the tubes were allowed to freeze overnight at -80 ºC before freeze-drying (VirTis BenchTop Pro, SP Scientific) until completely dry. 50 mg of freeze-dried stool sample was weighed into 2 ml tubes containing about 0.05 g of 1.0 mm Zirconia beads (Biospec Products, USA) and 1 ml of 2:1:1 solution. The sample was homogenised in a bead beater for 3 x 30 s (Precellys 24, Bertin Instruments, France) and centrifuged at 4 ºC for 20 minutes at 16 000 x *g* (Eppendorf Centrifuge 5430, Hamburg, Germany). 500 ul of the supernatant was filtered using 0.45 µm pore size nylon Corning® Costar® Spin-X® filters (Massachusetts, USA). The spin filters were centrifuged for 15 minutes using similar settings after which the filtrate was injected and transferred into vials (Waters, UK) for subsequent LC-MS analysis.

3.3.11 UPLC-MS conditions

UPLC-MS validation and analyses were carried out on a Waters ACQUITY UHPLC instrument (Waters Ltd., Elstree, UK) coupled to a Xevo triple quadrupole TQ-S mass spectrometer (Waters, Manchester, UK). The UPLC-MS system was as previously described (Sarafian et al., 2015) with minor changes. Chromatographic separation was carried out on an ACQUITY BEH C8 column (1.7 µm, 100 mm × 2.1 mm) at an operating temperature of 60ºC. Mobile phase A was made up of 1 mM ammonium acetate in an ACN:H_2O mixture in a 1:10 ratio. pH was adjusted to 4.16 using acetic acid. Mobile phase B was made up of ACN:IPA in a 1:1 ratio. The C8 column was first equilibrated with an initial mobile phase of 90% A at a flow rate

of 0.6 ml/min. The full gradient separation is given in Table 3-1. The strong and weak washes were 5:95 and 95:5 H_2O/IPA, respectively. Instrument seal wash was prepared as 10% IPA in H_2O. A needle wash cycle was set as first 500 µL of weak wash followed by 500 µL of strong wash.

The TQ-S was equipped with an electrospray ionization source operating in negative ion mode (ESI-). Capillary voltage was set at 1.5 kV, cone voltage at 60 V, source temperature maintained at 150 °C, desolvation temperature at 600 °C, desolvation gas flow at 1000 L/h, and cone gas flow at 150 L/h. Deprotonated species (M-H) of unconjugated BA were detected by selected ion monitoring (SIM) while conjugated BA were subjected to collision-induced dissociation to yield characteristic fragments detected by selected reaction monitoring (SRM). The SIM and SRM transitions for each of the BA standards and deuterated internal standards, as well as their optimum collision energy (CE) are provided in Tables 3-2 and 3-3.

Samples were kept at 4°C during the LC-MS analysis in a Waters ACQUITY autosampler (Waters, UK).

3.3.12 MS Data Analyses

MS data from the targeted LC-MS analysis was acquired on Waters MassLynx software 4.1 and pre-processed using Waters TargetLynx software 4.2. Processed MS data were further analysed, and graphs created using Microsoft Excel for Office 365. Multivariate analysis was carried out using SIMCA 15 (Sartorius Stedim, Umetrics® Data Analytics Suite, Sweden).

Table 3-1: Gradient for chromatographic separation for analysing bile acids using the existing lab "EL" method.

TIME	FLOW	%A	%B	Curve
(min)	(mL/min)			
0.00	0.6	90	10	6
0.10	0.6	90	10	6
9.25	0.6	65	35	6
11.50	0.65	15	85	6
11.80	0.8	0	100	6
12.00	0.95	0	100	6
12.10	1.0	0	100	6
12.40	1.0	0	100	6
12.45	0.85	45	55	6
12.50	0.85	90	10	6
12.60	0.8	90	10	6
12.70	0.7	90	10	6
12.80	0.6	90	10	6
15.00	0.6	90	10	6

Table 3-2: Full experiment details of selected ion/reaction monitoring experiments for bile acids standards.

SRM/ SIM No	BA ID	DIS for normalisation	NAME	Formula	M (g/mol)	*m/z* (M-H)	SRM/SIM transitions	CE (eV)
1	**BA1**	IS60	Ursocholanic Acid	$C_{24}H_{40}O_2$	360.3	359.28	359.29>359.28	10
2	**BA3**	IS60	5-Cholenic acid-3β-ol	$C_{24}H_{38}O_3$	374.28	373.274	373.27 > 373.26	10
	BA2	IS60	Lithocholenic acid		374.28			10
	BA4	IS60	3-Ketocholanic acid		374.28			10
3	**BA7**	IS60	Allolithocholic acid	$C_{24}H_{40}O_3$	376.3	375.29	375.29 > 375.28	10
	BA6	IS60	Isolithocholic acid		376.3			10
	BA5	IS60	Lithocholic acid		376.29			10
4	**BA9**	IS64	3α,12α, 23-Nordeoxycholic acid	$C_{23}H_{38}O_4$	378.28	377.27	377.27 > 377.26	10
6	**BA10**	IS64	9(11), (5β)-Cholenic acid-3α-ol-12-one	$C_{24}H_{36}O_4$	388.26	387.25	387.25 > 387.24	10
	BA11	IS63	5α-Cholanic acid-3, 6-dione		388.26			10
	BA12	IS65	3,7-Diketocholanic acid		388.26			10
	BA14	IS64	3,12-Diketocholanic acid		388.26			10
	BA13	IS63	3,6-Diketocholanic acid		388.26			10
7	**BA15**	IS64	8(14),(5β)-Cholenic acid-3α, 12α-diol	$C_{24}H_{38}O_4$	390.28	389.26	389.26 > 389.25	10
	BA16	IS65	5β-Cholenic acid-7α-ol-3-one		390.28			10
8	**BA17**	IS63	5α-Cholanic acid-3α-ol-6-one		390.28	389.27	389.27 > 389.26	10
	BA18	IS65	3α-Hydroxy-7 Ketolithocholic acid		390.28			10
	BA19	IS64	3α-Hydroxy-12 Ketolithocholic acid		390.28			10
9	**BA20**	IS63	Murocholic acid	$C_{24}H_{40}O_4$	392.29	391.28	391.28 > 391.27	10
10	**BA32**	IS62	Ursodeoxycholic acid		392.28	391.28	391.28 > 391.27	10
	BA25	IS63	Hyodeoxycholic acid		392.29			10
	BA22	IS64	5β-Cholanic acid-3β, 12α-diol		392.29			10
11	**BA24**	IS65	Chenodeoxycholic acid		392.29	391.28	391.28 > 391.27	10
	BA23	IS64	Deoxycholic acid		392.29			10
	BA21	IS64	Isodeoxycholic acid		392.29			10
15	**BA26**	IS66	3,7,12 Dehydrocholic acid	$C_{24}H_{34}O_5$	402.24	401.26	401.26 > 401.25	10
16	**BA27**	IS66	3α-Hydroxy-7,12-DiketoCholanic acid	$C_{24}H_{36}O_5$	404.26	403.25	403.25 > 403.24	10
17	**BA28**	IS66	3α-Hydroxy-6,7-DiketoCholanic acid		404.26	403.25	403.25 > 403.24	10
18	**BA31**	IS66	12 Dehydrocholic acid	$C_{24}H_{38}O_5$	406.27	405.26	405.26 > 40.5.25	10
	BA29	IS66	5β-Cholanic acid-3α, 6α-diol-7-one		406.27			10
	BA30	IS66	3 Dehydrocholic acid		406.27			10

SRM/ SIM No	BA ID	DIS for normalisation	NAME	Formula	M (g/mol)	*m/z* (M-H)	SRM/SIM transitions	CE (eV)
19	**BA33**	IS66	ω Muricholic acid	$C_{26}H_{40}O_5$	408.29	407.28	407.28 > 407.27	10
	BA34	IS66	β Muricholic acid		408.29			10
	BA37	IS66	α Muricholic acid		408.29			10
20	**BA36**	IS66	Hyocholic acid		408.29	407.28	407.28 > 407.27	10
	BA35	IS66	Cholic acid		408.29			10
22	**BA38**	IS67	Glycoursocholanic acid	$C_{26}H_{43}NO_3$	417.32	416.32	416.32 > 74	40
23	**BA39**	IS67	Glycolithocholic acid	$C_{26}H_{43}NO_4$	433.32	432.31	432.31 > 74	40
25	**BA41**	IS68	Glycochenodeoxycholic acid	$C_{26}H_{43}NO_5$	449.31	448.31	448.31 > 74	40
	BA42	IS69	Glycodeoxycholic acid		449.31			40
26	**BA43**	IS70	Glycoursodeoxycholic acid		449.31	448.31	448.31 > 74	40
	BA40	IS69	Glycohyodeoxycholic acid		449.31			40
29	**BA44**	IS71	3,7,12-Glycodehydrocholic acid	$C_{26}H_{37}NO_6$	459.26	458.25	458.25 > 74	40
30	**BA49**	IS71	Glycohyocholic acid	$C_{26}H_{43}NO_6$	465.31	464.28	464.28 > 74	40
31	**BA46**	IS71	Glycocholic acid		465.3	464.3	464.3 > 74	40
32	**BA47**	IS72	Tauroursocholanic acid	$C_{26}H_{45}NO_4S$	467.31	466.3	466.3 > 80	60
34	**BA48**	IS72	Taurolithocholic acid	$C_{26}H_{45}NO_5S$	483.3	482.29	482.29 > 80	60
36	**BA50**	IS74	Tauroursodeoxycholic acid	$C_{26}H_{45}NO_6S$	499.3	498.29	498.29 > 80	60
	BA51	IS75	Taurohyodeoxycholic acid		499.3			60
	BA53	IS75	Taurodeoxycholic acid		499.3			60
37	**BA52**	IS73	Taurochenodeoxycholic acid		499.3	498.29	498.29 > 80	60
40	**BA54**	IS77	3,7,12 Taurodehydrocholic acid	$C_{26}H_{39}NO_7S$	509.25	508.24	508.24 > 80	60
41	**BA58**	IS77	Tauro-ω-Muricholic acid	$C_{26}H_{45}NO_7S$	515.29	514.28	514.28 > 80	60
	BA56	IS77	Tauro-β-Muricholic acid		515.29			60
	BA57	IS77	Tauro-α-Muricholic acid		515.29			60
42	**BA55**	IS77	Taurohyocholic acid		515.29	514.28	514.28 > 80	60
43	**BA59**	IS77	Taurocholic acid		515.29	514.28	514.28 > 80	60

Identification (ID) short codes, matched deuterated internal standards for normalisation, full names, molecular weights, collision energies (CE), and SIM/SRM transitions for all 57 bile acid standards. Bile acids with similar collision energies, *m/z* values and SRM/SIM transitions were grouped into the same SRM window. For DIS short code IDs, the reader is referred to Table 3-3.

Table 3-3: Full experiment details of selected ion/reaction monitoring experiments for internal standards for bile acid analysis.

SRM	DIS ID	NAME	Formula	M (g/mol)	M/Z (M-H)	SRM/SIM transitions	CE (eV)
5	**IS60**	Lithocholic acid-d4	$C_{24}H_{36}O_3d_4$	380.32	379.09	379.09 > 378.08	10
12	**IS63**	β Hyodeoxycholic acid-d4	$C_{24}H_{40}O_4d4$	396.6	395.06	395.06 > 395.05	10
	IS62	Ursodeoxycholic acid-d4		396.6			10
13	**IS64**	Deoxycholic Acid-d4	$C_{24}H_{36}O_4d_4$	396.6	395.06	395.06 > 395.05	10
14	**IS65**	Chenodeoxycholic Acid -d5	$C_{24}H_{35}O_4d_5$	397.6	396.07	396.07 > 396.06	10
21	**IS66**	Cholic Acid-d4	$C_{24}H_{36}O_5d_4$	412.6	411.03	411.03 > 411.02	10
24	**IS67**	Glycolithocholic acid-d4	$C_{26}H_{39}NO_4d_4$	437.65	436.08	436.08 > 74	40
27	**IS70**	Glycoursodeoxycholic acid-d4	$C_{26}H_{39}NO_5d_4$	453.65	452.05	452.05 > 74	40
28	**IS68**	Glycochenodeoxycholic acid -d4	$C_{26}H_{29}NO_5d_4$	453.65	452.2	452.2 > 74	40
	IS69	Glycodeoxycholic acid-d4	$C_{26}H_{39}NO_5d_4$	453.65			40
33	**IS71**	Glycocholic Acid-d4	$C_{26}H_{39}NO_6d_4$	469.65	468.06	468.06 > 74	40
35	**IS72**	Taurolithocholic Acid-d4	$C_{26}H_{41}NO_5Sd_4$	487.75	486.02	486.02 > 80	60
38	**IS73**	Taurochenodeoxycholic Acid-d4	$C_{26}H_{41}NO_6Sd4$	503.73	502.03	502.03 > 80	60
	IS75	Taurodeoxycholic Acid-d4	$C_{26}H_{40}NO_6Sd_4$	503.73			60
39	**IS74**	Tauroursodeoxycholic acid-d5	$C_{26}H_{40}NO_6Sd5$	504.73	503	503 > 80	60
45	**IS77**	Taurocholic Acid-d_5	$C_{26}H_{39}NO_7Sd_5$	520.3	519.01	519.01 > 80	60

Identification (ID) short codes, full names, molecular weights, collision energies (CE), and SIM/SRM transitions for all 16 deuterated internal standards. Internal standards with similar collision energies, *m/z* values and SRM/SIM transitions were grouped into the same SRM window.

3.4 Results

All BAs and DIS will be identified by the ID short code throughout the Results section, unless otherwise stated. The ID short codes may be found in Tables 3-2 and 3-3.

3.4.1 Optimisation of injection volume

An injection volume of 2 µL was found to be optimal for both standard and sample extracts. Larger volumes resulted in broader and lagging peaks, especially of conjugated BAs which eluted before 8 mins RT. In the case of samples, faecal extracts were found to contain high concentrations of certain BAs (termed “major” BAs) that reached detector saturation when 5 µL of sample was injected. Thus, the sample injection volume was reduced from 5 µL to 2 µL for better peak shape and detection.

3.4.2 Analytical range and linearity:

Calibration curves were found to be linear over the selected analytical ranges. LLOQs ranged between 0.5 nM – 5 nM, while ULOQs ranged from 500 nM – 5000 nM (5µM). The more lipophilic unconjugated BA species were found to reach detector saturation at a lower concentration than the conjugated BAs. The LLOQ, ULOQ and r^2 of the different BAs are presented in Table S3-1.

3.4.3 Intra-day accuracy and precision

A single batch intra-day accuracy and precision was recommended for this partial validation process. Percentage accuracy for all BAs were within a range of 83.6-114.4% (120.8% for LLOQ) while precision was between 0.1-13.1% (16.9% for LLOQ). Percentage accuracy and precision values are presented in Table S3-1.

3.4.4 Analyte Selectivity of method

The SIM and SRM windows created were assessed for analyte transition specificity. Even though BAs of similar *m/z* values and close RTs were included in the same SIM/SRM windows, all BAs, except the diketocholanic acids and tauromuricholic acids, were resolved into distinct peaks. The diketocholanic acids, BA11-14, could not be distinguished when analysed together in the same BA standard cocktail because they elute at very close RTs (8.11, 8.16. 8.21 and

8.24 mins respectively). This resulted in a merging of the peaks into one wide “continuous” peak. However, their respective RTs were noted to allow for peak integration in biological samples, and their concentrations were quantified using the CC for BA10 which elutes as a fully resolved peak in the same window at 8.47 mins. Similarly, BA56 and BA57 co-elute at almost the same RT (3.30 and 3.31 mins respectively), therefore, the peaks could not be separated. Peaks detected at that RT in biological samples were recorded as “BA56/BA57”.

In addition, analyte-analyte interference was computed for all the BAs. Most of the interference was within the FDA recommended range of 20% of LLOQ response for the respective BA being investigated. Except for few BAs listed in Table 3-4, all interference responses were ≤LLOQ response.

Results from matrix-internal standard interference tests revealed that faecal extracts contained contaminants eluting at or very close to RTs of all unconjugated DIS (IS60, IS62, IS63, IS64, IS65 and IS66). Intensities of contaminants were reduced by diluting the extracts 10-fold with 2:1:1 solution.

3.4.5 DIS recovery and matrix factor

Samples were spiked with DIS at bead-beating (pre-spiked) and filtration (post-spiked) steps with QC1 (10nM) and QC2 (500nM). Average recoveries are given in Figure 3-2. Recoveries at both low and high QC concentrations ranged between 80% - 110%.

Test samples also revealed that concentrations of endogenous BAs were widely variant in different orders of magnitude. Several BAs (such as BA23 and BA05) were present in high concentrations exceeding the ULOQ of the analytical range, and in some instances reaching detector level of saturation. These were called the “major BAs”. Other BAs were present at much lower concentrations and these were referred to as the “minor BAs”. This observation suggested that depending on the BA(s) of interest, faecal extracts may have to be diluted, or sample preparation procedures adjusted to bring concentrations within analytical range.

For matrix factor experiments, compared to reference solutions, DIS signals were moderately affected by the presence of matrix ions (Figure 3-3) with matrix factors ranging between 80% - 115% at QC1 and 98% - 107% at QC2.

Table 3-4: List of bile acids with interference less than lower limit of quantification (LLOQ) response.

BA STANDARD		INTEFERING BA	
BA04	***3-Ketocholanic acid***	BA01	Ursocholanic acid
BA10	***9(11), (5β)-cholenic acid-3α-ol-12-one***	BA19	3α-hydroxy-12-ketolithocholic acid
BA17	***5α-cholanic acid-3α-ol-6-one***	BA32	Ursodeoxycholic acid
BA19	***3α-hydroxy-12-ketolithocholic acid***	BA22	5β-Cholanic Acid-3β, 12α-diol
BA25	***Hyodeoxycholic acid***	BA36	Hyocholic acid
BA26	***3,7,12-Dehydrocholic acid***	BA11-14	5α-cholanic acid-3,6-dione, Diketocholanic acids
BA27	***3α-hydroxy-7,12-diketocholanic acid***	BA19, BA31	3α-hydroxy-12-ketolithocholic acid, 12-Dehydrocholic acid
BA29	***5β-cholanic acid-3α, 6α-diol-7-one***	BA37	α-Muricholic acid
BA30	***3-dehydrocholic acid***	BA35	Cholic acid
BA32	***Ursodeoxycholic acid***	BA24	Chenodeoxycholic acid
BA35	***Cholic acid***	BA24, BA31	Chenodeoxycholic acid, 12-Dehydrocholic acid
BA37	***α-Muricholic acid***	BA29	5β-cholanic acid-3α, 6α-diol-7-one
BA39	***Glycolithocholic acid***	BA05, BA38	Lithocholic acid, Glycoursocholanic acid
BA42	***Glycodeoxycholic acid***	BA23	Deoxycholic acid
BA43	***Glycoursodeoxycholic acid***	BA32	Ursodeoxycholic acid
BA44	***3,7,12-Glycodehydrocholic acid***	BA26	3,7,12-Dehydrocholic acid
BA46	***Glycocholic acid***	BA35	Cholic acid
BA47	***Tauroursocholanic acid***	BA01	Ursocholanic acid
BA48	***Taurolithocholic acid***	BA47, BA53	Tauroursocholanic acid, Taurodeoxycholic acid
BA50	***Tauroursodeoxycholic acid***	BA32, BA52	Ursodeoxycholic acid, Taurochenodeoxycholic acid
BA51	***Taurohyodeoxycholic acid***	BA25	Hyodeoxycholic acid
BA54	***3,7,12-Taurodehydrocholic acid***	BA26	3,7,12-Dehydrocholic acid
BA56	***Tauro-β-muricholic acid***	BA34	β-Muricholic acid

Varying interference patterns are observed. Interfering bile acids which differ in the loss of a functional group was observed, such as BA04 and BA01 which differ only in the ketone group on C3 in the former; BA10 and BA19 which differ only in the C9=C11 double bond in the former; and BA25 and BA36 differ in the presence of -OH group on C7 in the latter. BA24 (CDCA) was observed interfering with BA32 (UDCA, a 7-beta isomer of CDCA). Interfering responses were observed for glycoconjugates BA42, BA43, BA44 and BA46 at the RT of their unconjugated moieties BA23, BA32,BA26 and BA35 respectively. Similarly, tauroconjugates BA47, BA51,BA54 and BA56 had interfering responses at the RT of their unconjugated parts BA01, BA25, BA26 and BA34 respectively. For bile acid structures, the reader may refer to Figure S3-1 in Appendix 3.

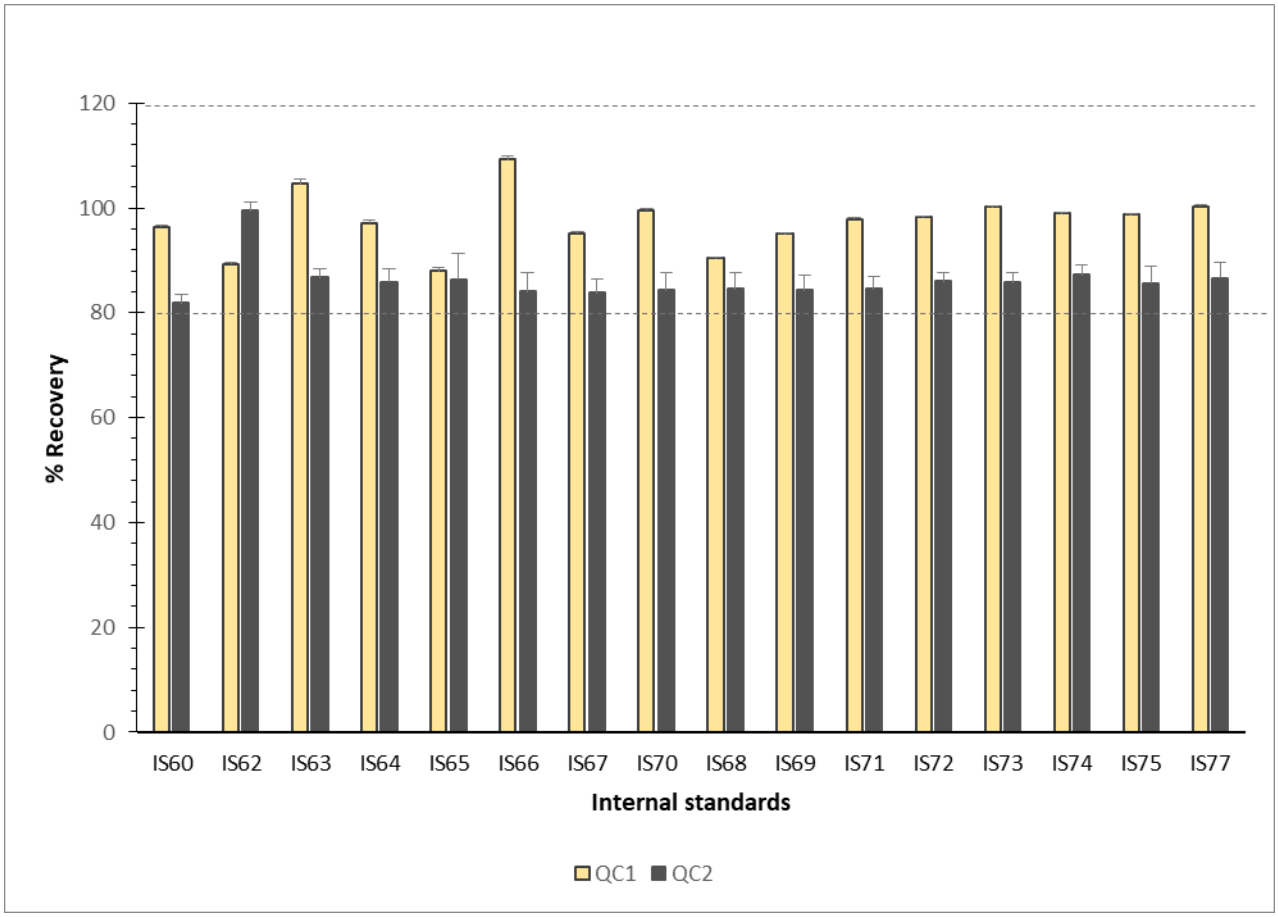


Figure 3-2: Average percentage recoveries of deuterated internal standards in test stool samples

Recoveries were within acceptable range (±20%) at both QC1 (10 nM) and QC2 (500 nM) concentrations. Error bars are standard deviation of replicates (n=3). For internal standard ID codes, the reader is referred to Table 3-3.

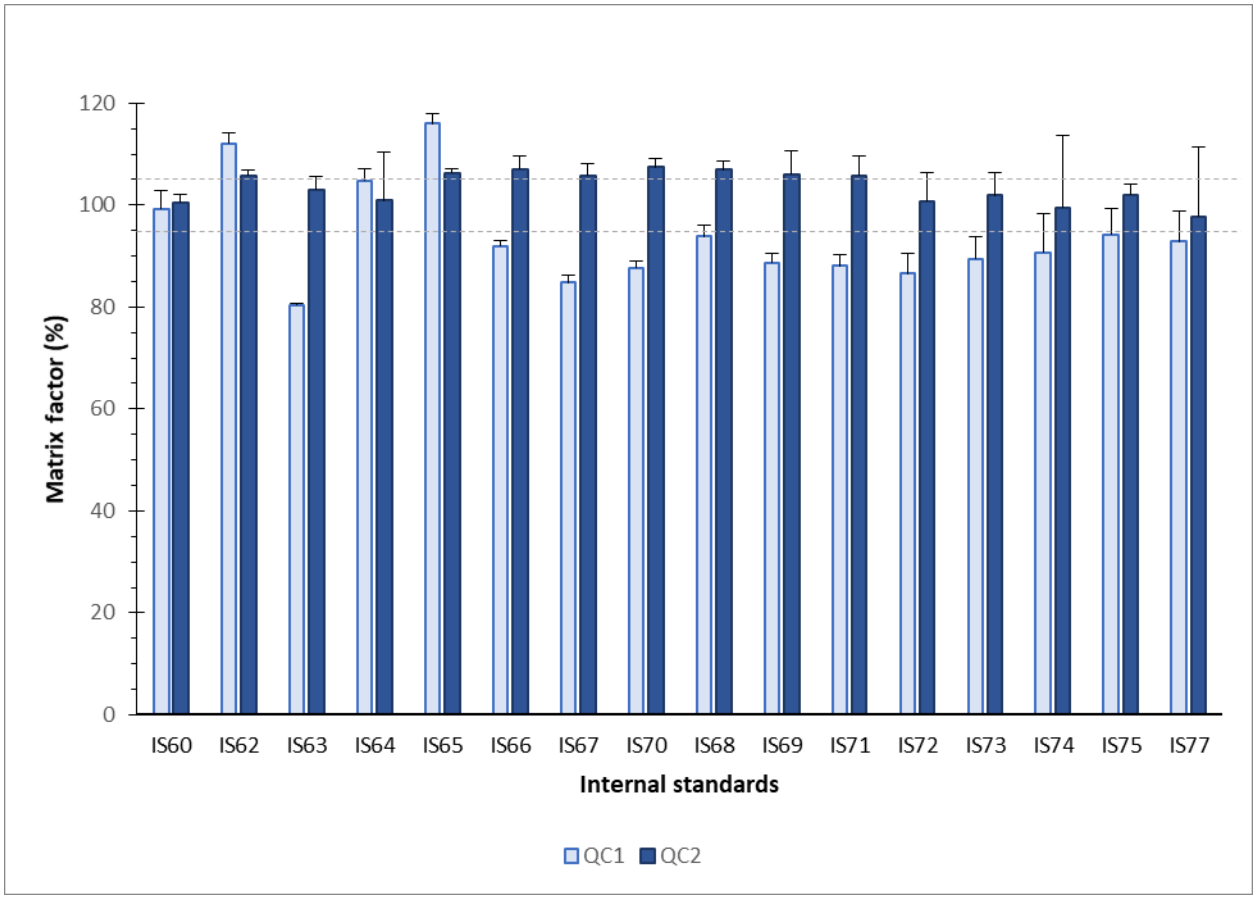


Figure 3-3: Average matrix factors (%) observed in test stool samples.

Internal standard responses of over-spiked samples at QC1 (10 nM) and QC2 (500 nM) concentrations were expressed as a percentage of single blank reference solutions prepared at same concentrations. Horizontal dash lines indicate ±5% acceptance range limits. QC1 was observed to be affected at a greater extent by the presence of matrix ions compared to QC2. Error bars are standard deviation of replicates (*n* = 3). For internal standard ID codes, the reader is referred to Table 3-3.

3.4.6 Application to clinical samples

The method was subsequently applied to the analysis of faecal samples (*n* = 131) obtained from a clinical cohort of healthy adult vegans and omnivores. Samples were randomised and processed according to the sample preparation procedure in Section 3.3.10. A biological "QC" sample was prepared by pooling equal volumes of all faecal filtrates for injection at a regular interval during analysis to monitor assay performance (Sangster, Major, Plumb, Wilson, & Wilson, 2006). However, due to results from validation experiments which revealed that endogenous BAs were present in faecal extracts at widely varying concentrations, the clinical samples were extracted without being pre-spiked with DIS. The biological sample was in this case prepared and analysed before embarking on the whole sample set in order to determine how much dilution will be required to get concentrations within acceptable range. Samples were then diluted as required with 2:1:1 solution containing DIS (*i.e.* samples were post-spiked). Final DIS concentration was 500 nM.

All extracts were diluted by 10-fold to reduce the effect of contaminant peaks so that DIS peaks could be reliably integrated. A total of 53 BAs were detected in this sample set at widely varying concentrations. Several BAs were present in high concentrations exceeding the ULOQ of the analytical range. To quantify these "major BAs", comprising of BA23, BA05, BA04, BA06, BA19, BA22, BA24, BA32, BA35 and BA52; aliquots of extracts were further diluted by 100-fold to be able to get their concentrations within the analytical range. Other BAs detected at lower concentrations ("minor BAs") were analysed at the initial dilution, resulting in two sets of analyses.

The MS data was analysed in TargetLynx to obtain DIS-normalised concentrations. Ion suppression was observed in the x10 dilution sample set for IS60 and IS64 which are deuterated forms of BA05 and BA23 respectively. This was not observed in the x100 dilution, which were used for the quantification of these BAs as they were highly abundant. Concentrations that were indeterminate (too low to be quantified by the TargetLynx algorithm), with a signal-to-noise ratio less than 5, or less than LLOQ were set to 0. Values above the ULOQ (+20%) were removed and treated as missing data. Subsequently, concentrations were adjusted for dilutions and divided by the starting faecal mass. Final concentrations are presented per mg dry stool mass. Total faecal BA concentration was then calculated as the sum of all BA concentrations and added to the dataset as a variable.

For univariate analyses, BAs detected in less than 70% of samples were excluded from statistical tests to reduce noise in the analyses. This left only 24 BAs to be included in the analyses. Differences between groups were investigated by parametric independent samples t-tests or non-parametric Mann Whitney tests as appropriate. *P* values were adjusted for

Table 3-5: Faecal BA concentrations in vegans vs omnivores.

		Vegans	Omnivores	p_{raw}	p_{adj}
	n	**68**	**63**		
BA04	*3-Ketocholanic acid*	3.73(5.99)	6.11(7.05)	0.113	0.294
BA05	***Lithocholic acid***	**52.54(72.01)**	**48.18(79.56)**	**0.809**	**0.919**
BA06	***Isolithocholic acid***	**10.15(13.38)**	**13.45(16.57)**	**0.204**	**0.319**
BA07	*Allolithocholic acid*	0.78(0.94)	0.92(1.35)	0.048	0.154
BA09	*3α,12α, 23-Nordeoxycholic Acid*	0.01(0.02)	0.02(0.03)	0.001	0.028
BA10	*9(11), (5β)-Cholenic acid-3α-ol-12-one*	0.02(0.02)	0.03(0.03)	0.152	0.305
BA13	*3,6-Diketocholanic acid*	0.46(0.63)	0.77(1.50)	0.049	0.154
BA18	*3α-hydroxy-7-ketolithocholic acid*	0.08(0.20)	0.06(0.29)	0.928	0.961
BA19	***3α-hydroxy-12-ketolithocholic acid***	**6.38(8.92)**	**12.58(22.49)**	**0.020**	**0.099**
BA22	*5β-Cholanic acid-3β, 12α-diol*	3.30(4.95)	5.40(9.93)	0.045	0.154
BA23	***Deoxycholic acid***	**72.98(102.98)**	**116.39(211)**	**0.014**	**0.099**
BA24	*Chenodeoxycholic acid*	1.72(4.15)	1.43(6.20)	0.958	0.961
BA25	*Hyodeoxycholic acid*	0.08(0.11)	0.10(0.20)	0.225	0.330
BA31	*12-Dehydrocholic acid*	0.73(1.43)	0.59(1.87)	0.961	0.961
BA32	*Ursodeoxycholic acid*	1.31(1.42)	1.88(3.37)	0.012	0.099
BA34	β-Muricholic acid	0.01(0.02)	0.02(0.03)	0.129	0.294
BA35	*Cholic acid*	0.63(2.15)	0.75(3.11)	0.601	0.716
BA41	*Glycochenodeoxycholic acid*	0.50(1.15)	0.45(0.71)	0.560	0.700
BA42	*Glycodeoxycholic acid*	0.29(0.43)	0.19(0.71)	0.497	0.654
BA43	*Glycoursodeoxycholic acid*	0.02(0.04)	0.02(0.05)	0.126	0.294
BA46	*Glycocholic acid*	0.59(1.44)	0.41(0.89)	0.171	0.305
BA52	***Taurochenodeoxycholic acid***	**6.80(13.1)**	**7.39(16.10)**	**0.163**	**0.305**
BA53	*Taurodeoxycholic acid*	0.06(0.12)	0.06(0.30)	0.200	0.319
BA59	*Taurocholic acid*	0.08(0.13)	0.07(0.14)	0.457	0.635
Total BA	***Total BA***	**196.26(255.62)**	**284.21(317.23)**	**0.018**	**0.099**

All concentrations are ng per mg dry stool. 24 BAs with >70% non-null values, as well as the total faecal bile acid concentrations, are displayed. Results are displayed as median (interquartile range). "Major" bile acids detected at higher concentrations are highlighted in bold font. Groups were compared by Mann Whitney U test. Raw p values were adjusted by Benjamini-Hochberg false discovery rate correction method. p_{raw} = raw p values, p_{adj} = adjusted *p* values.

multiple testing by applying Benjamini-Hochberg false discovery rate (FDR) correction. Results from these are presented in Tables 3-5 and 3-6.

In both groups, the unconjugated secondary BAs, BA23 (deoxycholic acid, subsequently referred to as DCA) and BA05 (lithocholic acid, subsequently referred to as LCA) accounted for about 70% of total faecal BA concentrations (Figure 3-4). Omnivores tended to have a higher concentration of total faecal BA (p_{raw} = 0.026, Table 3-5, Figure 3-5) than vegans, driven mostly by DCA, which was also higher (p_{raw} = 0.014, Table 3-5, Figure 3-5) in omnivores than in vegans, although these differences were not found to be statistically significant following multiple-testing corrections by FDR (p_{adj} > 0.05).

Proportionally, DCA represented a higher percentage of the omnivore BA profile (p = 0.060, Table 3-6, Figure 3-6) while vegans had a much higher percentage of LCA (p = 0.003, Table 3-6, Figure 3-6).

Table 3-6: Percentage compositions of BA in vegans and omnivores.

	Vegans	***Omnivores***	***p_{raw}***
n	68	63	
DCA%	34.63(±18.2)	41.06(±20.63)	0.060
LCA%	28.74(17.89)	21.69(23.16)	0.003
Unconjugated BA%	94.06(14.68)	96.02(8.25)	0.525
Conjugated BA%	5.94(14.68)	3.98(8.25)	0.525

Data is presented as mean(standard deviation) or median (IQR) as appropriate. DCA%, LCA%, unconjugated% and conjugated% are deoxycholic acid, lithocholic acid, total unconjugated bile acid and total conjugated bile acid concentrations respectively expressed as percentages of total faecal bile acid concentration.

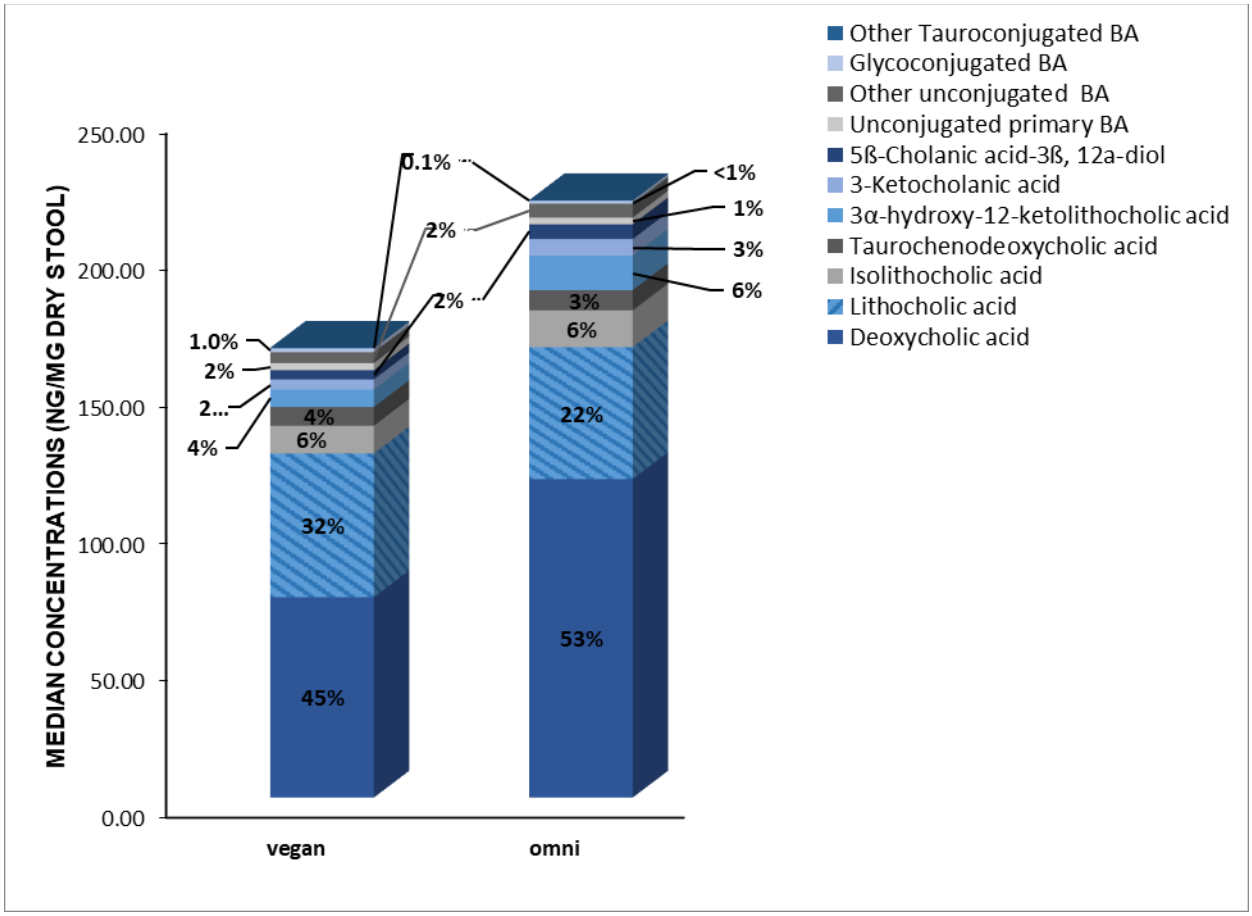


Figure 3-4: Faecal bile acid profiles of vegans and omnivores displayed in a stacked bar plot.

Panel totals correspond to median total BA concentrations (ng per mg dry stool). Omnivores tended to have higher faecal BA but differences in total faecal BA was not found to be statistically significant between groups ($p > 0.05$, Mann Whitney U test, Benjamini-Hochberg FDR correction). Both groups had similar profiles, except that vegans tended to have more LCA than DCA in their faeces ($p = 0.003$).

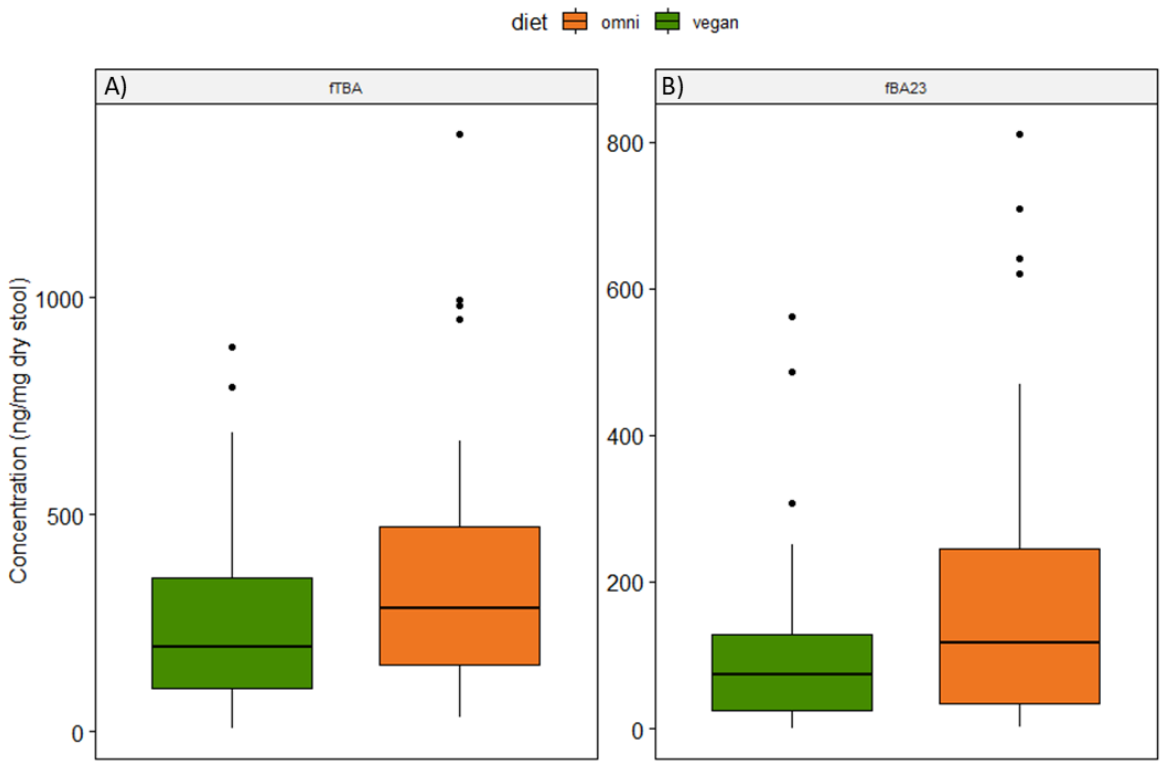


Figure 3-5: Bile acid concentrations (ng/mg dry stool) in vegans (green) and omnivores (orange).

Boxplots are showing a) total faecal bile acid concentrations and b) faecal deoxycholic acid (fBA23) concentrations. In both cases, omnivores tended to have higher concentrations than vegans, although this difference did not retain statistical significance following correction for multiple testing (Benjamini-Hochberg false discovery rate correction). Outliers were not excluded as their inclusion did not alter FDR-corrected significance/p-value.

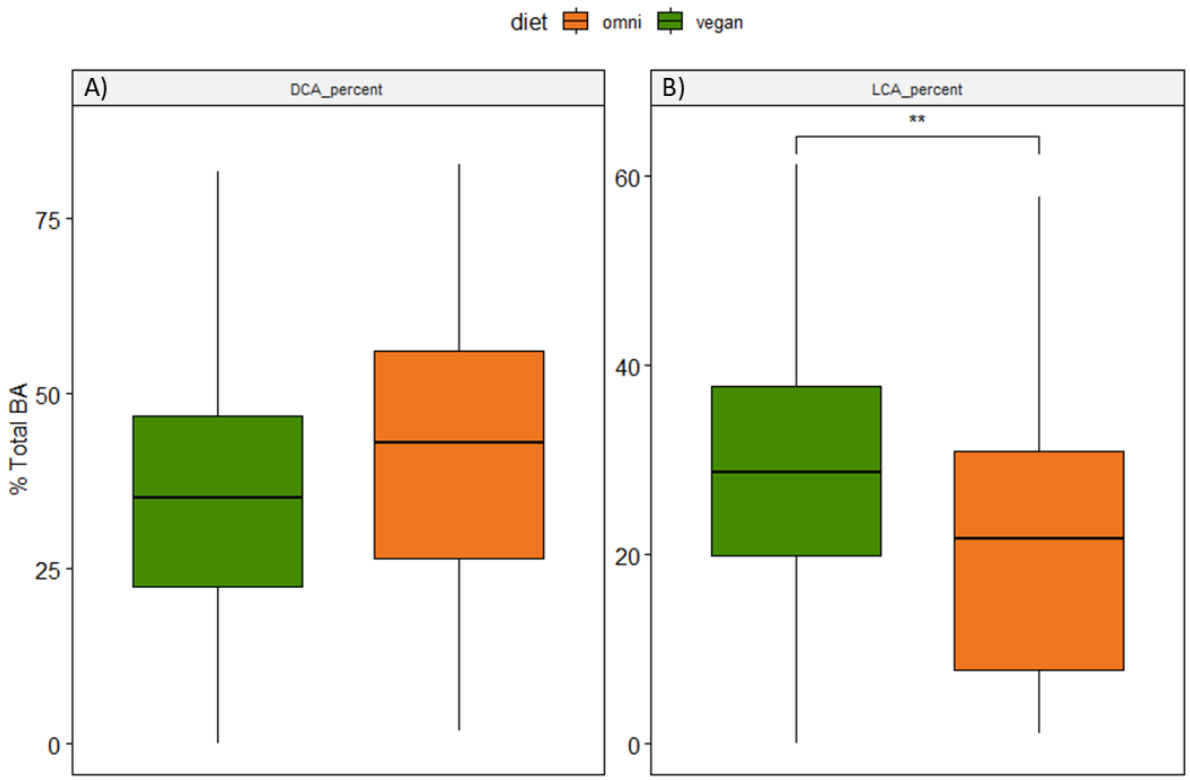


Figure 3-6: Boxplots showing percentage compositions of a) deoxycholic acid (BA23) and b) lithocholic acid (BA05) in faecal samples of vegans and omnivores.

Both bile acids were the top most abundant in faeces, together accounting for about 70% of total bile acid concentration. However, while both groups had almost comparable percentages of deoxycholic acid, vegans had a much higher lithocholic acid fraction.

Similar to results from univariate statistics, multivariate analyses revealed no differences between vegans and omnivores. Principal Component Analysis (PCA) plots (Figure 3-7) showed that the variation in the dataset was not explained by diet, as there was no obvious clustering by diet groups (Figure 3-7A). As observed in Figure 3-7B, PC1 explained for 63.3% of the variation along the horizontal axis, driven by DCA (BA23) concentration. The PC2 (along the vertical axis) explained for 10.3% of the variation observed in the dataset, which was driven by LCA (BA05) concentrations (Figure 3-7C). Exclusion of BAs with less than 70% non-null values did not significantly alter the PCA model (cumulative R2 of PC1 and PC2 of model for all 53 detected BA plus total BA = 0.767 *vs* cumulative R2 of model for only 24 BAs plus total BA = 0.768).

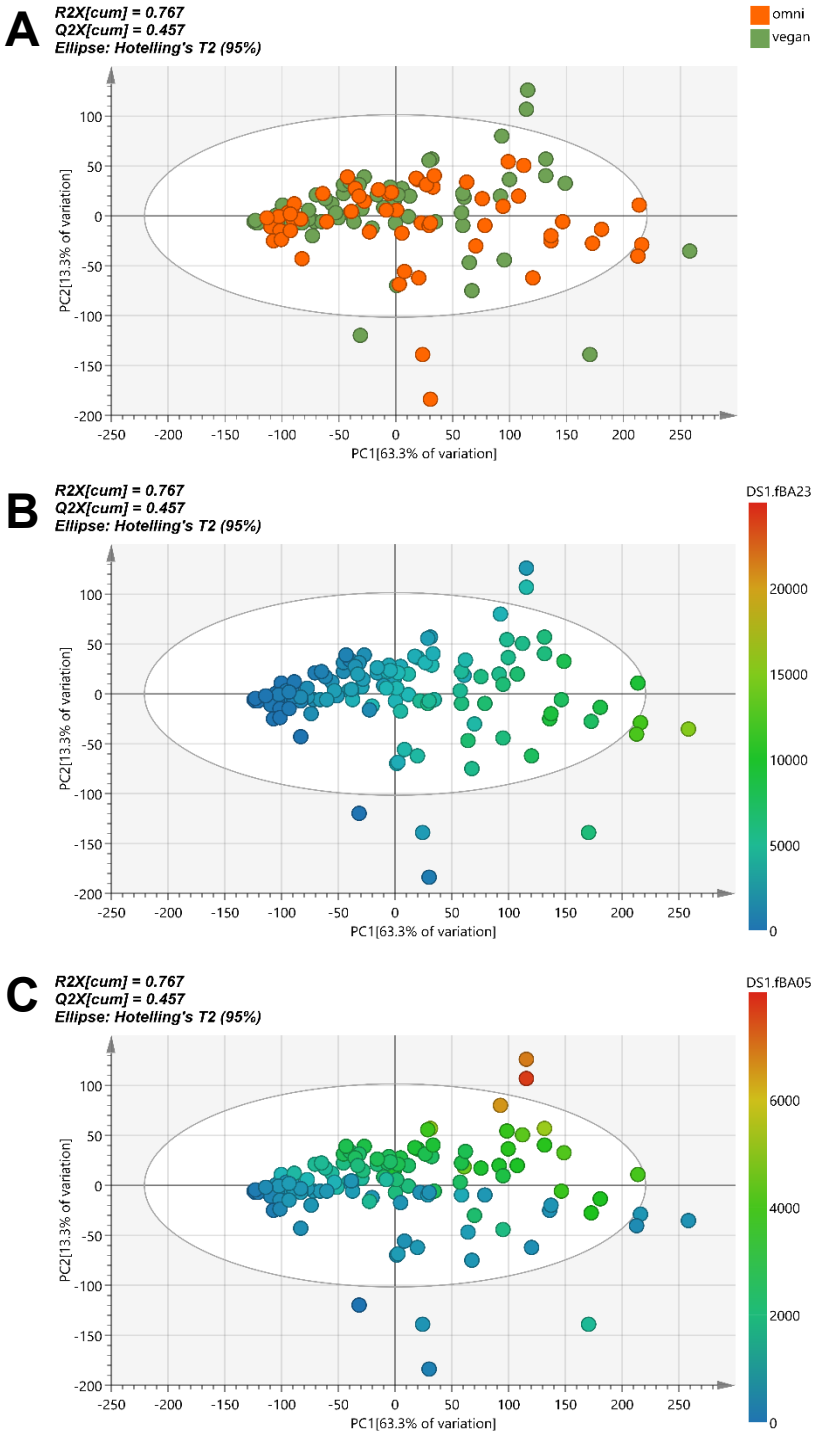


Figure 3-7: PCA Scores plots showing bile acid profiles of vegans and omnivores (excluding n = 6 extreme outliers).

PC1 explained 63.3% of variation in the data while PC2 could explain 13.3% of variation observed in the dataset. As observed in plot(A), there was no separation based on diet grouping. Plots (B) and (C) show that variations along PC1 and PC2 are largely driven by DCA/BA23 and LCA/BA05 respectively. Model was built on all detected BAs (n=53). Data was pareto-scaled.

3.5 Discussion

3.5.1 Summary of findings

In summary, the following were the findings presented in this chapter:

- A UPLC-MS method was optimised and validated for quantifying BA in faecal samples. The method was successfully validated for accuracy, precision, linearity and selectivity.
- Efficiency of sample extraction was assessed by recovery experiments. Good recoveries (>80%) were obtained.
- Matrix factor was also investigated. Faecal extract matrix was found to minimally impact signal of DIS/BA ions.
- The method was applied to faecal samples from a cohort of vegans and omnivores. While vegans tended to have lower BA concentrations than omnivores, differences between groups were not statistically significant following correction for multiple comparisons.

3.5.2 General discussion

The main aim of this chapter was to perform a partial validation of the EL method. Based on in-house laboratory and FDA recommendations, partial validations were required to be carried out when there were modifications to a previously validated bioanalytical method. In this particular case, the changes were a new biofluid/matrix within the same species (from human plasma and urine to human faecal extracts, the primary reason for the validation), a necessary method transfer between laboratories and instruments (transfer from one site to another to facilitate analyses, using similar Waters MS instruments and software), and a consequent change in sample volume (to better detect analytes in the new matrix).

The method was validated in terms of analytical range, linearity, intra-day accuracy and precision and selectivity. Analytical ranges, linearity, accuracy and precision were confirmed to be within acceptable limits, in accordance with the previous report from the initial validation of the UPLC-MS method in other biological fluids (Sarafian et al., 2015). In addition to those experiments, the selectivity of the method which was not previously investigated by Sarafian and colleagues was explored in this validation exercise. Analyte-analyte interferences were observed between BAs with similar structures, such as those only differing in the presence/absence of a functional group. For example, cross-interfering BAs BA04 (3-

ketocholanic acid) and BA01 (ursocholanic acid) differ only in the ketone group on C3 in the former; BA10 (9(11), (5β)-Cholenic acid-3α-ol-12-one) and BA19 (3α-hydroxy-12-ketolithocholic acid) differ only in the C9=C11 double bond in the former; and BA25 (hyodeoxycholic acid) and BA36 (hyocholic acid) differ in the presence of -OH group on C7 in the latter (Figure 3-8). Isomers were also observed to interfere with the other, for example BA24 (chenodeoxycholic acid) with BA32 (ursodeoxycholic acid, a 7-beta isomer of BA24) (Figure S3-1x). In other cases, interference was observed from the unconjugated forms of some of the BAs suggesting a possible in-source loss of the attached taurine or glycine. Interfering responses were observed for glycoconjugates BA42 (glycodeoxycholic acid) , BA43 (glyco-ursodeoxycholic acid) , BA44 (3,7,12-glycodehydrocholic acid) and BA46 (glycocholic acid) at the RT of their unconjugated moieties BA23 (deoxycholic acid), BA32 (ursodeoxycholic acid), BA26 (3,7,12-dehydrocholic acid) and BA35 (cholic acid) respectively. Similarly, tauroconjugates BA47 (tauroursocholanic acid), BA51 (taurohyodeoxycholic acid), BA54 (3,7,12-taurodehydrocholic acid) and BA56 (tauro-β-muricholic acid) had interfering responses at the RT of their unconjugated parts BA01 (ursocholanic acid), BA25 (hyodeoxycholic acid), BA26 (3,7,12-dehydrocholic acid) and BA34 (β-muricholic acid) respectively. (All BA structures may be found in Appendix 3).

BAs are a class of structurally similar compounds and as such, the relatively low interference (therefore conversely, high selectivity) observed, in addition to the other validation results, shows that the described method is sufficiently specific and sensitive for the reliable quantification of BAs detected. With only a few exceptions, the observed interferences were at much lower intensities than the LLOQ concentration of the respective BAs and were therefore not expected to impact on the results. Nevertheless, the LLOQs for BAs with interferences greater than their (initial) LLOQ levels were increased as appropriate, leading to a truncation of the CC of the affected BA; and all values below LLOQ were excluded during the data analysis. Truncations were carried out while ensuring that all CCs had at least 6 valid calibration points on the calibration curve.

Also, interference from matrix contaminants in the SRM windows for the DIS of unconjugated BAs was observed as unidentified peaks, eluting close to or at the RT of labelled standards, that sometimes swamped the signal from the DIS ions, especially at low concentrations. Extra care had to be taken to ensure the right peaks were integrated by the TargetLynx algorithm. This observation was supported by the results of the matrix factor experiments. The acceptable matrix factor range was between 95% - 105% (±2%). All DIS passed at QC2, with the exception of IS60 (lithocholic acid-d4), IS64 (deoxycholic acid-d4), IS68 (glycochenodeoxycholic acid-d4) and IS75 (taurodeoxycholic acid-d4), which failed at QC1.

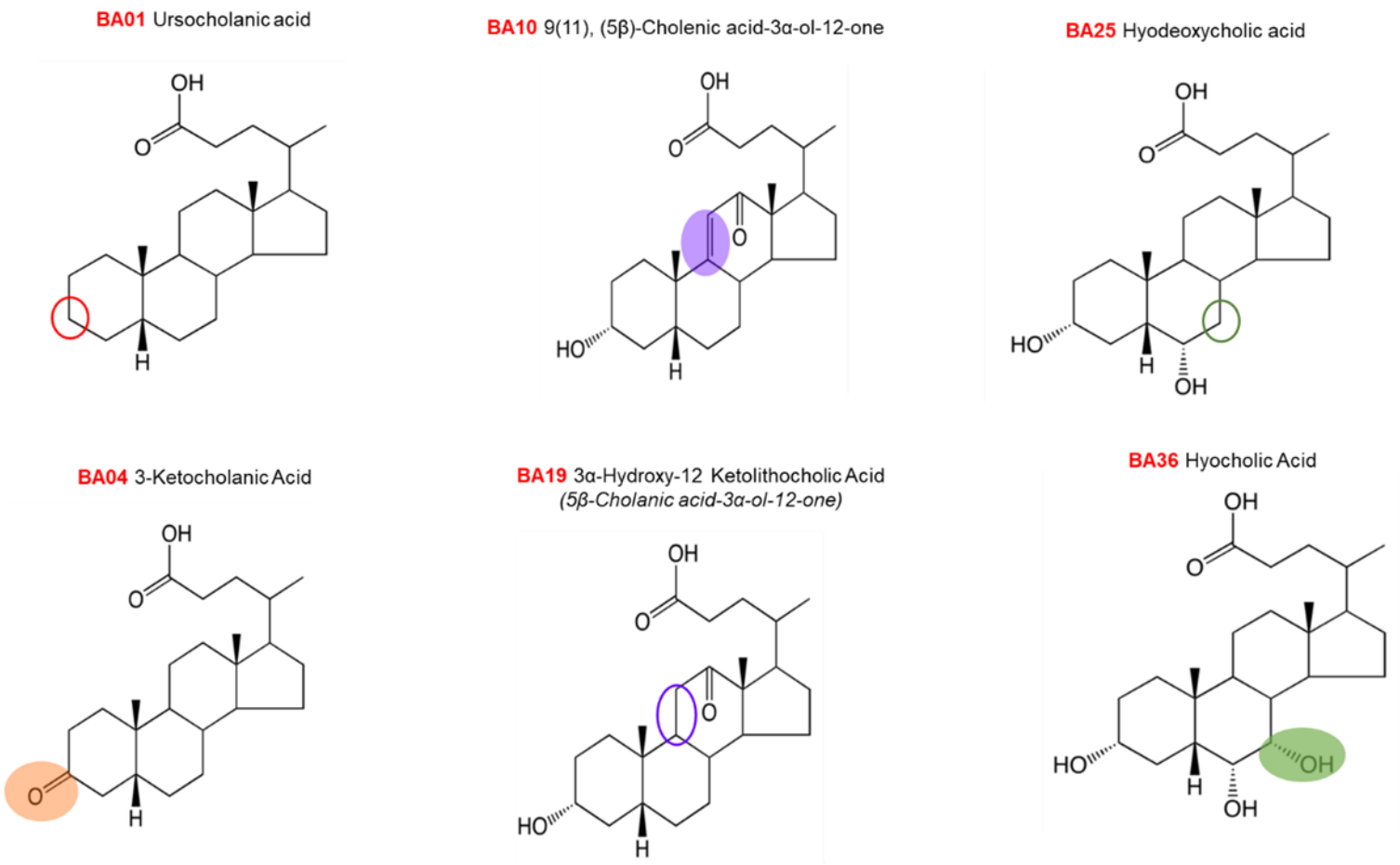


Figure 3-8: Structures of examples of cross-interfering bile acid species differing in the presence or absence of a functional group.

Thus, it was obvious that better analytical accuracy may be achieved by spiking samples with DIS at concentrations higher than Q1. For the application analysis, a decision was made to dilute extracts and spike in DIS at QC2 (500nM) in order to reduce matrix effect during analysis.

Overall, the method showed stability across the run. The CVs for all the BAs in the biological QCs injected at regular intervals across the run ranged between 6.6 – 15.9%, indicating that the method was suitable for analysing large numbers of clinical samples. Compared to previous methods, the sample preparation procedure described coupled to the 15-min UPLC-MS method provides a quick analytical pipeline for analysing faecal BA and covers a wide range of BAs detected in human biofluids. The stool matrix is undoubtedly a complex biofluid. However, as a result of the targeted SRM method, an intensive sample preparation could be avoided without sacrificing on selectivity, sensitivity and coverage.

Analyte recovery in faecal extracts were also investigated, with most recoveries being >80%. These results demonstrated the efficiency of the extraction protocol and LC-MS method for analysing BA in faecal extracts as well as previously described biofluids of serum, plasma and urine.

Lastly, the method was applied to clinical samples as a proof of concept. A total of 53 BAs were detected in the clinical samples, demonstrating the diversity of BA species due to multiple transformations mediated by members of the gut microbiota community. The faecal BA pool was comprised of >95% unconjugated BAs in both vegans and omnivores, most abundant of which were the secondary BAs, DCA and LCA. This finding is in alignment with results from previous studies which have reported that DCA and LCA are the most abundant BAs in healthy adult human faeces (Ashok Kumar Batta et al., 1999; Danielsson, Eneroth, Hellstrom, Lindstedt, & Sjovall, 1963; Franco et al., 2019; Kakiyama et al., 2014; Korpela et al., 1988, 1986; S. Reddy, Sanders, Owen, & Thompson, 1998; K. D. Setchell et al., 1983; van Faassen et al., 1993). The other top BAs detected were isolithocholic acid (iso-LCA), taurochenodeoxycholic acid (TCDCA), 3α-hydroxy-12-ketolithocholic acid, 3-ketocholanic acid, chenodeoxycholic acid (CDCA), cholic acid (CA), 5ß-cholanic acid-3ß,12a-diol (3ß-isomer of DCA) and ursodeoxycholic acid (UDCA). It was surprising to find TCDCA among the topmost abundant BA, it was in fact the most abundant tauro-conjugated BA detected in stool, while the other tauro-BAs were detected in trace amounts. For glyco-conjugated BA species, glycocholic, glycodeoxycholic and glycochenodeoxycholic acids were detected in similar low concentrations. These results are similar to results obtained in our lab from infants meconium samples (data available upon request). The pattern seem to be retained in adulthood.

Faecal BA profiles in vegans have been seldomly investigated prior to this study. A very recent study investigating faecal BA profiles in vegans and omnivores reported similar results as obtained in this study, where vegans had significantly lower BAs than omnivores, despite that the researchers analysed wet stool mass without accounting for differences in faecal water content (Trefflich et al., 2019). Whereas in the method described in this chapter, all samples were freeze-dried in order to control for the differences in stool consistency and moisture content. Also, an earlier study investigating the effect of 20-day vegan, lacto-ovo-vegetarian and omnivorous diets in a crossover randomised controlled trial reported significant decreases in concentrations of total and all BAs on the vegan diet compared to both the lacto-ovo-vegetarian and omnivorous diets (van Faassen et al., 1987). These results are somewhat similar to the observations from this study. However, while vegans tended to have lower concentrations of BAs than omnivores, these differences were not statistically significant following corrections for multiple comparisons. This may be because while vegans in this study were self-declared to have been adhering to veganism for at least a year, participants recruited onto the study conducted by the van Faassen group were fed a vegan diet for 20 days only(van Faassen et al., 1987). It is possible that the drastic fall in faecal BA concentrations induced by an "acute" exposure to a meat-less diet under controlled conditions may not be as strong over a longer "chronic" period of time.

On the other hand, more studies have been conducted in vegetarians and omnivores where similar results to those obtained in this study were also reported. Reddy *et al.* (S. Reddy et al., 1998), Korpela *et al.* (Korpela et al., 1988) and van Faassen *et al.* (van Faassen et al., 1993) all reported no significant differences in total BA and DCA concentrations between vegetarians and omnivores, even though concentrations in the former tended to be lower. In contrast to these studies, Nestel and colleagues (Nestel, Billington, & Smith, 1981) found a significant difference in total BA excreted by lacto-ovo-vegetarians versus omnivores. These results may be interpreted with caution as the sample size was small, with only 7 lacto-ovo-vegetarians vs 6 non-vegetarians. In fact, the previous studies were conducted in fewer participants than this study, hence, may be underpowered.

3.5.3 Strengths and Limitations of the study

However, it is important to note that the method described herein makes use of stable isotope labelled standards for normalisation of analyte concentrations. As listed on Tables 3-2 and 3-3, only 16 of the isotopically labelled standards could be obtained, hence BAs were matched to IS analogues or those closest in structure and RT, resulting in relative quantification for some BAs. In addition, it has been argued that stable isotope labelled standards may not be

optimal for ESI-LC-MS tandem experiments (Wieling, 2002). Previous reports have shown significant ion suppression of internal standards in the presence of the endogenous analyte (H. R. Liang, Foltz, Meng, & Bennett, 2003; Remane, Wissenbach, Meyer, & Maurer, 2010). This was also observed in this study for the two most abundant BAs, DCA (BA23, DIS= DCA-d4, IS64) and LCA (BA05, DIS= LCA-d4, IS60). However, the suppression of the DIS ions was observed only in the x10 dilutions, as such this did not impact the quantification of these BAs because their x1000 dilutions were used for that. This may be due to the fact that endogenous concentrations were very high and possibly saturating the detector signal. This also brings up the question about whether using DIS recoveries as a measure of extraction efficiency is optimal. An alternative would have been to use spike standards into charcoal-stripped “blank” samples. However, charcoal-stripping may remove other components of the matrix other than the analytes of interest (in this case, BAs) and as such, may not be a representative reference for comparing recoveries (Griffiths & Sjövall, 2010). Stable-isotope labelled standards such as deuterated internal standards used in these experiments are generally considered good practice (N. R. Reddy, 2017).

The strengths of this study include the first-time application of the previously validated LC-MS method in human faecal samples, the first-time investigation of faecal BA composition in habitual vegans in a sample size as large as this and the wide coverage of the LC-MS method utilized for this investigation.

3.6 Conclusion

As part of the investigations into the influence of diet on PDC metabolism, a partial validation of an existing UPLC-MS method for BA analysis was carried out. The method which combines a rapid high throughput analytical LC-MS cycle (15minutes) with a relatively simple faecal sample preparation protocol was successfully validated for accuracy, precision, linearity, selectivity, matrix factor and analyte recovery. Also, the method was successfully applied to clinical samples. The novelty of this study includes the determination of the analyte selectivity of the method, the first-time application of the validated targeted method to human faecal matrix and the first-time comparative analysis of faecal BA compositions in habitual vegans and omnivores in a cohort of this size.

Not only did the validation results demonstrate that the method was reliable for detecting and quantifying BAs ranging from low to very high abundance in human faeces, the method was also found to be stable across a large run of hundreds of samples facilitating the analysis of large numbers of clinical samples in a short amount of time. Furthermore, the method was applied to samples from a cohort of healthy vegans and omnivores. Although total BA concentrations did not differ significantly (post FDR) between groups, BA compositions in vegans and omnivores were found to be significantly different, especially in the percentage composition of LCA present.

With particular reference to the primary aim of this thesis, the validation results revealed that the method was suitable for the reliable quantification of DCA, the monomer of PDC, which is the primary analyte of interest. Therefore, the method development process for PDC quantification could proceed. The next chapter will describe how the method was further optimised for quantifying PDC in human faecal samples.

4. QUANTIFYING POLYDEOXYCHOLIC ACID: DEVELOPMENT OF A HIGH THROUGPUT WORKFLOW FOR QUANTIFYING POLYDEOXYCHOLIC ACID IN HUMAN FAECAL SAMPLES

4.1 Introduction

Given the potential role for polydeoxycholic acid (PDC) in human physiology and health, a high-throughput analytical method was developed to accurately detect and quantify PDC in relevant biofluids. The method development process was performed in two major steps. First, a targeted reverse phase ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS)-based method was validated (subsequently referred to as EL method) for analysing monomeric bile acids (BA) in biological samples of interest (*i.e.* faecal samples). This initial step was discussed in detail in the preceding chapter.

With respect to the PDC monomer, deoxycholic acid (DCA, short code ID = BA23 from previous chapter, *m/z* 391.2, RT 10.36mins, Figure 4-1), the validation results revealed that DCA could be reliably detected and measured on the EL method to a high degree of accuracy and precision (Table 4-1). In addition, the sample preparation procedure was optimal for efficient recovery of the deuterated DCA standard (DCA-d4) with no significant impact of matrix ions (Table 4-2). Analyte-analyte selectivity indicated that glycodeoxycholic acid (short code ID = BA42 from previous chapter) may fragment into unconjugated DCA (interference = 4.6x LLOQ signal), suggesting that this may minimally inflate the DCA signal. To diminish this potential bias in concentrations the lower limit of quantification for DCA was adjusted to 5 nM (from 1 nM). The final calibration range for DCA was 5 nM – 2500 nM, over which linearity was confirmed (R^2 = 0.9950).

Since these results confirmed that the EL method was “fit for purpose” by the high efficiency of extraction, detection and analysis of DCA, the next step taken was to then optimise the method for the analysis of PDC. To achieve this, PDC standards were synthesised and used to facilitate the method optimisation for PDC quantification.

Table 4-1: Relative accuracy and precision for detecting and quantifying deoxycholic acid (BA23) on the EL method

DCA (BA23)	LLOQQC	LQC	MQC	HQC	ULOQQC
Relative accuracy, %	105.61	87.74	111.19	93.39	94.36
Precision, %	11.97	13.07	1.22	0.78	0.86

Mean values (*n* = 6) for relative accuracy and precision obtained from validation experiments.

Table 4-2:DCA-d4 (deuterated internal standard for deoxycholic acid) recovery and matrix factor results from the validation of EL method

DCA-d4 (IS64)	Recovery	Matrix Factor
QC1 (10nM)	97.1(±0.6)%	104.8(±2.3)%
QC2 (500nM)	85.9(±2.5)%	100.9(±9.5)%

Mean (±standard deviation) recovery and matrix factor values obtained from validation experiments. (*n* = 3)

In summary, this chapter describes how the validated analytical (EL) method was further optimised for PDC quantification in faecal samples. Finally, the developed method and workflow was applied to clinical samples from the same vegan and omnivore cohort previously described in the preceding chapter.

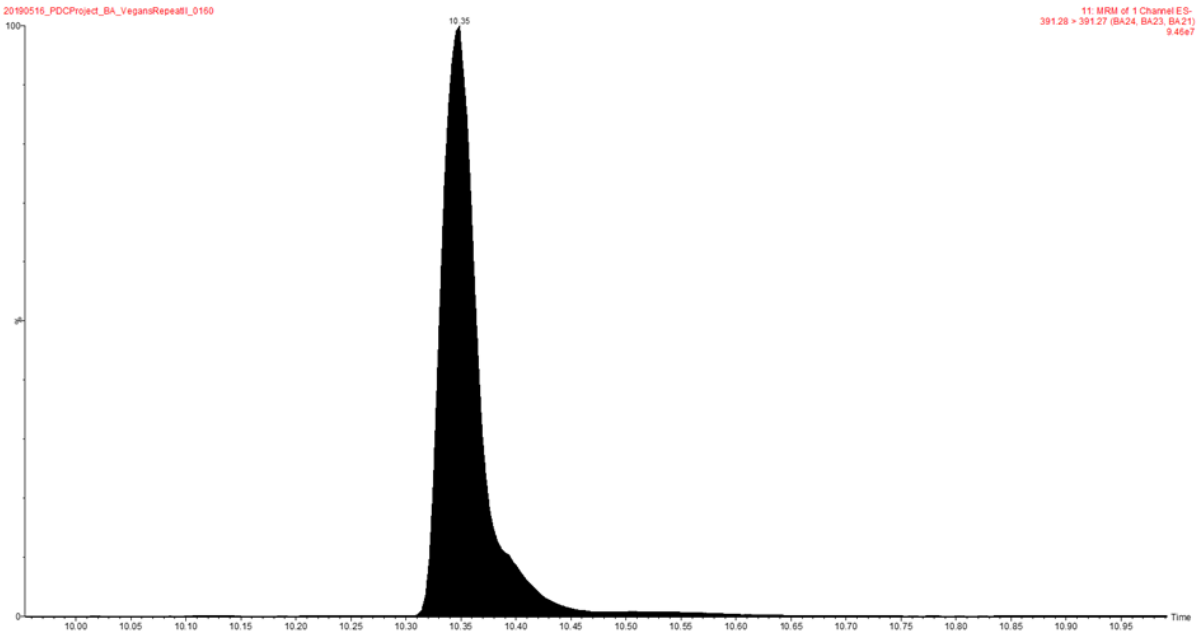


Figure 4-1: Chromatogram of deoxycholic acid peak; m/z 391.3, RT 10.36mins on the EL method

4.2 Aim and Objectives

4.2.1 Aim

To develop a suitable extraction, detection and quantification workflow to recover and measure polydeoxycholic acid (PDC) in the biological sample of interest (*i.e.* faeces).

4.2.2 Objectives

- To synthesise PDC using classical synthetic chemistry approaches
- To explore various available analytical platforms (namely and confirm synthesis using UPLC-MS, proton nuclear magnetic resonance (^{1}H-NMR) and matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS)) for PDC detection in biological samples
- To optimise existing sample preparation procedures for high-throughput analysis of PDC in faecal samples
- To develop a workflow for PDC analysis in human faecal samples on targeted LC-MS
- To apply developed workflow in clinical samples: For this part of the analysis, the *a priori* hypothesis was that “PDC concentrations will differ significantly between vegans and omnivores.”

4.3 Methods

4.3.1 Chemicals and reagents

Deoxycholic acid standard (DCA), p-toluenesulfonic acid (PTSA), toluene, sodium hydroxide (NaOH), hydrochloric acid (37% purity, HCl), chloroform ($CHCl_3$) and deuterated chloroform ($CDCl_3$) for synthesis of PDC standards were purchased from Sigma-Aldrich. Deoxycholic acid-d4 (DCA-d4) was purchased from Steraloids Inc., Newport (Rhode Island, USA).

UHPLC-MS grade Chromasolv^{TM} solvents acetonitrile (ACN), 2-propanol (IPA) and methanol (MeOH) from Honeywell International Inc (USA) and UHPLC-UV grade water (H_2O) purchased from Fischer Scientific (Loughborough, UK) were used for preparation of mobile phases and sample extraction solutions. Eluent additives ammonium acetate and acetic acid (used for adjusting pH of EL method mobile phase A) were purchased from Sigma-Aldrich now Merck (Dorset/Gillingham, UK).

4.3.2 Chemical synthesis of PDC standards for method development

PDC standards were synthesised using the method previously described by Benson *et al.* (Benson et al., 1993) using a classical Dean-Stark distillation set-up. The DCA standard (1 g) and PTSA catalyst (24.5 mg) were weighed into a round-bottom flask to which boiling chips and a magnetic flea were added. The flask was coupled to a Dean-Stark trap connected to a condenser as pictured in Figure 4-2. The Dean-Stark column was insulated with cotton wool to keep rising vapour from condensing before reaching the condenser. The flask was placed in a heating block and heated on a magnetic hotplate for the entirety of the experiment. The reaction mixture was brought to boil at 150 °C and was continuously stirred at 320 rpm. Samples were collected at time, t = 0, 30, 60, 90, 120, 150, 180, 240, 300 and 360 mins. Solvent was removed on a rotary evaporator, and the dried powder was stored in airtight borosilicate glass bottles at room temperature.

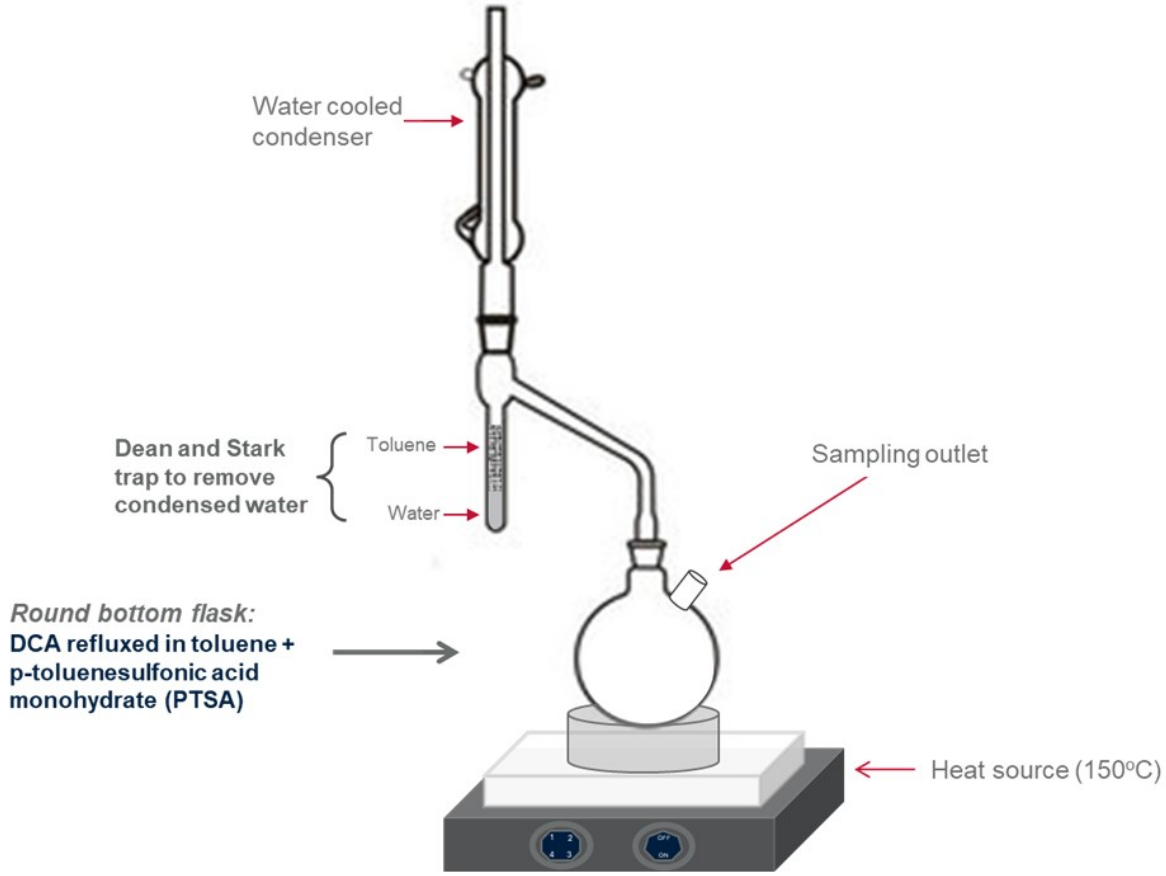


Figure 4-2: A Dean-Stark set-up for the synthesis of polydeoxycholic acid standards.

Deoxycholic acid is refluxed in toluene in the presence of p-toluenesulfonic acid monohydrate catalyst for 6 hours. Samples are taken out at specified times throughout the reaction.

The synthesis experiment was carried out in the Department of Chemistry laboratory, Imperial College London with adequate safety measures in place. The reaction mixture was brought to boil in ~10mins. Boiling chips were added to prevent bumping. Samples were taken at the stated times across 6hours. The reaction mixture at t = 360mins (6hrs) was cooled and filtered into a pre-weighed flask using a Buchner filter to remove anti-bumping granules. The toluene was removed, and the dried powder obtained was weighed and transferred to an air-tight borosilicate glass bottle.

4.3.3 UPLC-MS Experimental Conditions

Samples (in toluene) from the synthesis reaction were monitored by UPLC-MS. UPLC-MS analyses were performed on a Waters ACQUITY UHPLC instrument (Waters Ltd., Elstree, UK) coupled to a Xevo triple quadrupole TQ-S mass spectrometer (Waters, Manchester, UK) and a Waters 2777C external CTC autosampler (Waters, Wilmslow, UK). The UPLC-MS system was as previously described in Chapter 3 with minor changes to the LC gradient and wash cycles to minimise carryover. The adjusted gradient is given in Table 4-3. The needle wash cycle was extended to three alternating sets of weak-strong-weak, followed by a similar wash cycle for the injector valve.

Apart from the slight change in LC gradient and wash cycles, the LC-MS settings remained the same as previously described in Chapter 3.

Table 4-3: Adjusted gradient for chromatographic separation for analysing bile acids using the "EL" UPLC-MS based method.

TIME (min)	FLOW (mL/min)	%A	%B	Curve
0.00	0.6	90	10	6
0.10	0.6	90	10	6
9.25	0.6	65	35	6
11.50	0.65	15	85	6
11.80	0.8	0	100	6
12.00	0.95	0	100	6
12.10	1	0	100	6
12.40	1	0	100	6
14.40	1	0	100	6
14.45	0.85	45	55	6
14.50	0.85	90	10	6
14.60	0.8	90	10	6
14.70	0.7	90	10	6
14.80	0.6	90	10	6
16.65	0.6	90	10	6

4.3.4 Proton-NMR Experimental Conditions

Samples of synthesised PDC standards were analysed by ^{1}H NMR spectroscopy. DCA and synthesised dried PDC standards were dissolved in $CDCl_3$ (99.8% atom) containing the internal standard 3-trimethylsilyl-1-[2,2,3,3-$^{2}H4$] propionic acid (TSP). The samples were vortexed and centrifuged after which supernatant was transferred to a 5 mm NMR tube for analysis. One dimensional spectra were acquired on a 600 MHz Bruker NMR spectrometer operating at 300 K using a standard pulse sequence (recycle delay (RD)-90°-t_1–90°-t_m-90°-acquire free induction decay (FID)). For this experiment, 8 dummy scans were followed by 32 scans collected in 64K data points with a spectral window set to 20 ppm. Line broadening was set at 3 Hz, relaxation delay 1s, mixing time 0.1s, acquisition time 2.1s and pulse length 9.820 µs. Spectra were automatically calibrated to the TSP resonance at δ 0.0, and phase and baseline corrected in Topspin (Bruker).

4.3.5 MALDI-MS Experimental Conditions

Samples collected from the synthesis experiment were analysed at the Imperial College Department of Chemistry Mass Spectrometry facility by MALDI-MS for the confirmation of the synthesis. The analyst was provided with the expected molecular masses of the PDC oligomers. MALDI-MS analyses were carried out on a micromass MALDI-time of flight (ToF) MS instrument equipped with a 337 nm laser operating in positive mode. MS settings were as follows: reflectron 5 kV, pulse 2100 V and detector 2300 V. The samples were tested using different matrices and in different detection modes.

4.3.6 Preparation of working solutions

Working solutions of DCA and DCA-d4 were prepared from the parent 1 mg/ml stock solutions by diluting as required with H_2O:ACN:IPA solution in 2:1:1 ratio (subsequently referred to as the 2:1:1 solution) to produce the calibration curve (CC).

4.3.7 Collection of clinical faecal samples from Vegans vs Omnivores Study

Clinical samples were obtained from a cross-sectional study of vegans and omnivores already described in Chapter 3. Briefly, volunteers were recruited onto the study through various print and online media. Eligible volunteers had to meet the following criteria: aged between 18-65 years, weight-stable for a minimum of two months prior to the study, were not pregnant, lactating or following a medically prescribed diet, had no known gastrointestinal disease and

had not taken antibiotics within 3 months prior to the study. Vegan volunteers were eligible if they had adhered to a vegan diet for at least 1 year. Enrolled study volunteers were provided with Sarstedt sampling tubes to collect their faecal samples at home. Samples were kept at -18°C and transferred on dry ice within 48 hours for storage at -80°C until analysis. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethical Committee for the Capital Region of Denmark (ethical approval ref: j no. H-3-2012-145). All participants gave their written informed consent prior to inclusion in the study.

4.3.8 Preparation of stool samples

DCA concentrations in healthy human stool were found to be very high (reaching detector saturation) during the method validation described in the preceding chapter. Faecal extracts were required to be diluted x1000-fold to get the DCA concentrations within the analytical range. This led to a decision to reduce the mass of starting material in the optimisation experiments. Following several tests with 10 mg, 15 mg and 25 mg; an eventual decision was made to start with 10 mg (±2 mg) of lyophilised stool and to dilute the final extract by x20-fold just before LC-MS analysis. Also, this informed the concentration of DCA-d4 to spike in at the beginning of the extraction/sample preparation process to ensure that the final concentration could be detected by the MS. For "PDC-as-DCA" quantification in clinical samples, hydrolysed samples were post-spiked with DCA-d4 at appropriate concentration following the dilution of extracts (*i.e.* immediately prior to LC-MS analysis) so that internal standard-normalised "PDC" concentrations could be obtained.

BA extraction was carried out as already described using the sample preparation protocol described in Chapter 3. Briefly, 10 mg of freeze-dried faecal sample was extracted in 2:1:1 solution by bead-beating (to homogenise sample), centrifuging and filtering the supernatant to further clean up the sample extracts. Filtrates were then analysed for DCA by LC-MS.

For PDC quantification, an alkaline hydrolysis was introduced into sample preparation process. After centrifugation, the pellet obtained was further extracted by adding 1 ml of $CHCl_3$ and bead-beating again for 3x30s and centrifuged as before. The $CHCl_3$ supernatant was collected and allowed to dry overnight in the fume hood. The dried-down $CHCl_3$ were then resuspended in 1 ml 1 M NaOH and incubated at 95°C in a water bath overnight (~15 hours). 2 ml Eppendorf® Safe-Lock microcentrifuge tubes were found to be optimal for this process in order to avoid tubes from popping open during the heating process and to prevent the consequent sample loss. As an added precautionary measure, microcentrifuge tube cap-locks (StarLab, UK) were used to ensure no sample loss. Samples of synthesized PDC standards were also subjected to the alkaline hydrolysis to validate the saponification treatment. The

hydrolysed samples were cooled and neutralised with 1M HCl after which the neutralised hydrolysate was extracted with an equal volume of 1:1 IPA: ACN solution, filtered using microcentrifuge filters and analysed by LC-MS.

Biological quality control (QC) samples were prepared from a pool of equal volumes of the filtrates. These were injected at regular intervals during the analysis to monitor assay performance.

4.3.9 MS Data Analyses

MS data from the targeted LC-MS analysis was acquired on MassLynx software 4.1 and pre-processed using TargetLynx 4.2. Linearity of DCA curve was assessed over a concentration range of 5 nM – 2.5 µM visually on TargetLynx as well as an R^2 value greater than 0.99. Each CC point was included in the curve if signal-to-noise ratio was greater than 5 and %deviation from nominal concentration was <25%. Statistical analyses on processed MS data were carried out in R v3.5.2 using *tidyverse* v1.2.1 collection of packages and Microsoft Excel for Office 365.

Monomer DCA concentration (i.e. DCA extracted into aqueous fraction) was determined as well as PDC concentration (i.e. DCA obtained from alkaline hydrolysis of PDC organic extracts). Concentrations that were indeterminate (too low to be quantified by the TargetLynx algorithm), with a signal-to-noise ratio less than 5, or less than lowest limit of quantification were set to 0. Values above the upper limit of quantification were removed and treated as missing data. Final concentrations were adjusted for dilution and normalised by mass to give final values per mg dry stool mass. Total faecal DCA (total fDCA) concentration was calculated as sum of both PDC and DCA concentrations (total fDCA = DCA + PDC) and included as a variable in analyses. In addition, PDC was expressed as a percentage of total fDCA to evaluate how much of the total DCA was contained in the PDC-fraction.

4.4 Results and Discussion

4.4.1 Synthesis of PDC oligomers

Masses of PDC oligomers were estimated by calculating the expected masses of different oligomers as follows:

Expected mass of *n*-chain oligomer = *n* * (mass of one DCA molecule) – (*n*-1) * (mass of one water molecule)

Alternatively, the addition (difference in mass between consecutive oligomers) was estimated to be 374 g/mol (Table 4-4).

Table 4-4: Predicted molecular weights for PDC oligomers up to the decamer (10 DCA monomer units).

No. DCA units	Predicted mass (g/mol)
1	392.29
2	766.57
3	1140.86
4	1515.14
5	1889.42
6	2263.70
7	2637.99
8	3012.27
9	3386.55
10	3760.83

4.4.2 LC-MS detection of PDC oligomers:

The EL method was explored for the possibility of detecting PDC oligomers in the samples (in toluene) collected throughout the synthesis. An *m/z* scan for the deprotonated (M-H) ions of each of the oligomers up to the limit of the detector range (*m/z* 2048) was carried out with minimal collision energy set at 5 eV. The DCA monomer (*m/z* 391.3) eluted as usual at RT 10.4 mins while the PDC dimer (*m/z* 765.6) was found to elute at 11.6 mins. When peak areas for both DCA monomer and PDC dimer were plotted, there was a gradual decrease in DCA and an increase in the PDC-dimer over time (Figure 4-3). No other oligomer was detected on the EL method.

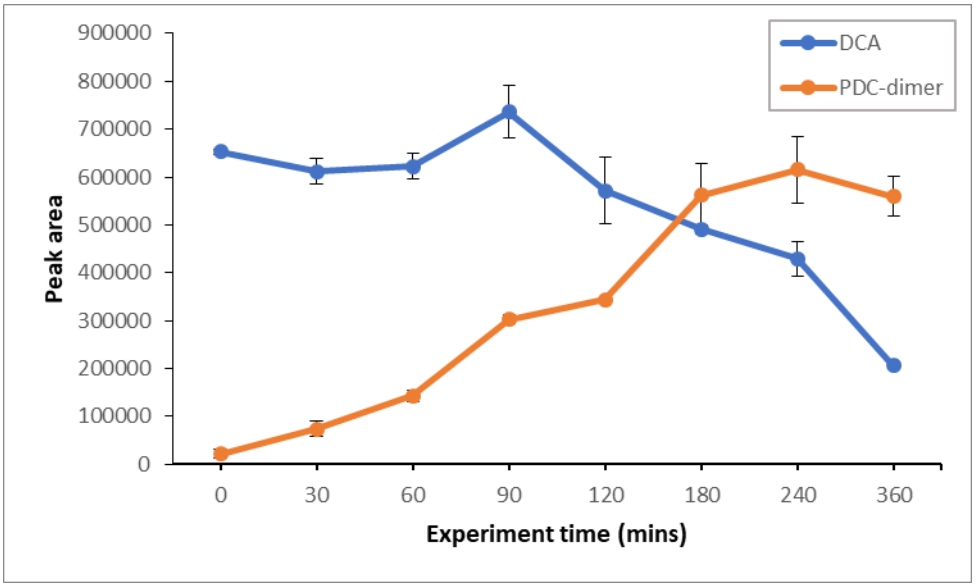


Figure 4-3: Peak areas of deoxycholic acid (DCA) and polydeoxycholic acid dimer (PDC-dimer) over the course of the synthesis experiment.

Mean integrated peak areas are displayed. Error bars are standard deviations of integrated peak areas. Integration was carried out using Waters® TargetLynx™ software.

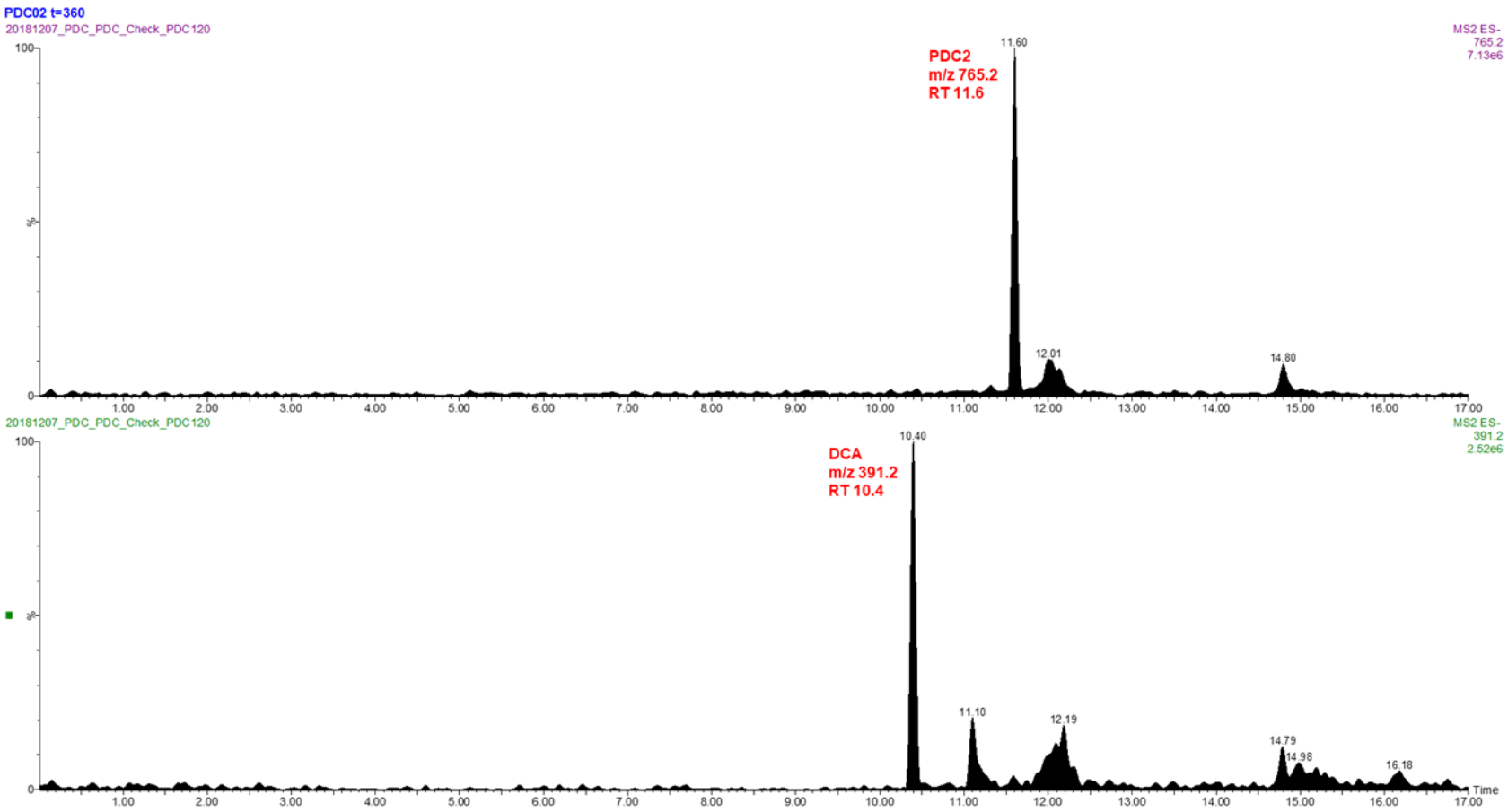


Figure 4-4: Extracted mass chromatograms of DCA and PDC-dimer on the EL method.

DCA eluted at 10.4mins while the PDC-dimer was found to elute at 11.6mins. No longer oligomers could be detected on this method. Peak smoothing was carried using mean (3x2).

A more lipophilic LC gradient (subsequently referred to as “OLIG” method) was then created, following quick tests on both Waters ACQUITY C8 and C18 columns with different mobile phases. The final mobile phase was a binary solvent system made up of 100% UPLC-MS grade H_2O (mobile phase A) and 100% MeOH (mobile phase B). The final LC gradient for this OLIG method is presented in Table 4-5. The MS parameters remained the same as on the EL method.

Elution of the standards up to the pentamer was achieved on a C8 column at a constant flow rate of 0.45µL/min. Extracted mass chromatograms from sample *t* = 360mins may be observed in Figure 4-5.

The peak areas for all of the detected oligomers were integrated and plotted over the course of the synthesis experiment. This showed the gradual decrease in DCA and complementing increases in the oligomers (Figure 4-6). The trimer was not reliably detected until at *t* = 60mins, the tetramer at *t* = 180mins and the pentamer at *t* = 360mins.

Table 4-5:Gradient for chromatographic separation for analysing synthesised PDC oligomers using the "OLIG" UPLC-MS based method

Time(min)	Flow Rate (mL/min)	%A	%B	Curve
Initial	0.45	20.0	80.0	Initial
2.00	0.45	1.0	99.0	6
8.50	0.45	1.0	99.0	6
8.51	0.45	20.0	80.0	6
10.00	0.45	20.0	80.0	6

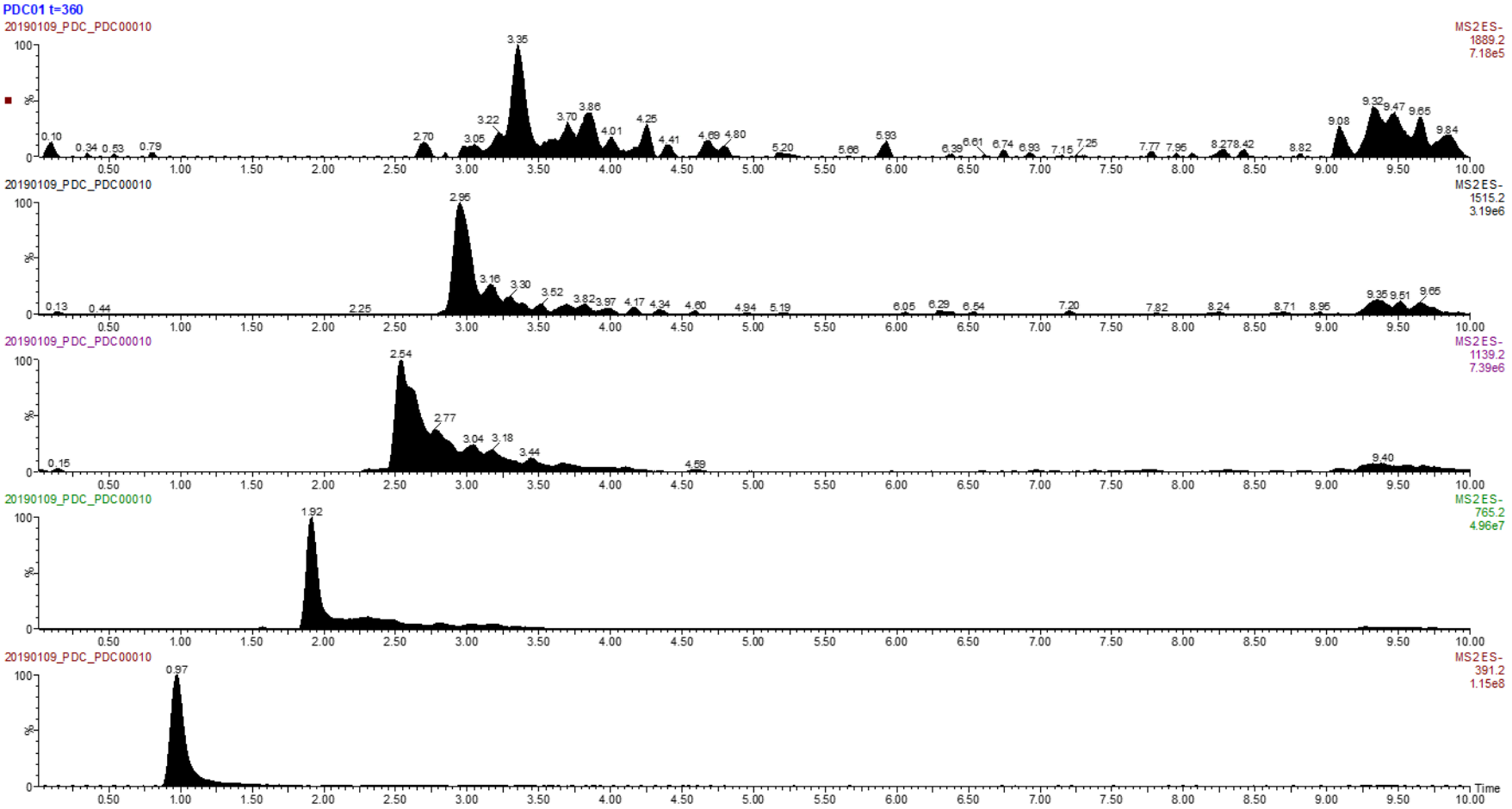


Figure 4-5: Extracted mass chromatograms showing peaks (after smoothing using mean 3x2) of DCA and PDC oligomers from synthesised standard samples (t = 360mins) on the OLIG method.

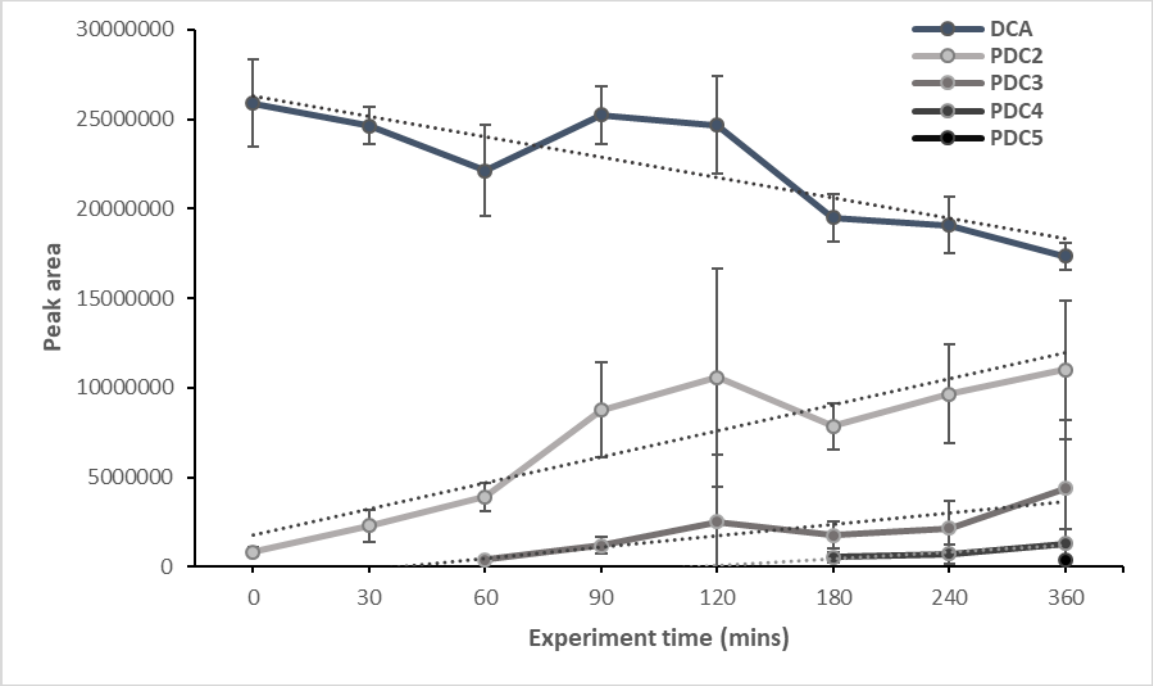


Figure 4-6: Peak areas of deoxycholic acid (DCA) and polydeoxycholic acid oligomers [-dimer (PDC2), -trimer(PDC3), -tetramer(PDC4) and -pentamer(PDC5)] detected on OLIG method.

PDC2 (dimer) was detected in minimal amounts at t=0, PDC3 (trimer) not until t=60, PDC4 (tetramer) not until t=180 and t=5 (pentamer) only at t=360. The limit of the detector was *m/z* 2048 therefore, no further oligomers may have been detected. Linear trends indicate a consistent decrease and increase in DCA and PDC respectively as the experiment proceeded. Error bars are standard deviation of peak areas.

4.4.3 NMR detection of PDC oligomers:

To gain structural information on standards synthesised, dried PDC powder (*t* = 360 mins) was analysed by NMR spectroscopy. One dimensional (1D) spectra from ^{1}H NMR confirmed the presence of PDC oligomers. Three spectral peaks were identified to provide information for the oligomers. The peak at 3.69 ppm arose from the terminal OH-group at the C-3-position (see in Figure 4-7). The peak at 4.75 ppm corresponded to the internal OH-group at the C3-position that arises when the terminal OH-group is used to create an oligomer (see Figure 4-7). Accordingly, this internal OH-group peak was absent in the DCA spectrum (Figure 4-7). An additional peak was observed for the β-OH group at the C12-position. Mean polymer length was estimated from the ratio of the integrals of the ^{1}H signals from the internal and terminal hydrogens as follows:

Divide internal by terminal to obtain a ratio, round up to nearest whole, and add 1 (original DCA).

For example,

If signal for internal was 9 and terminal was 3, 9/3 = 3; 3+1 = 4; length is 4 DCA subunits.

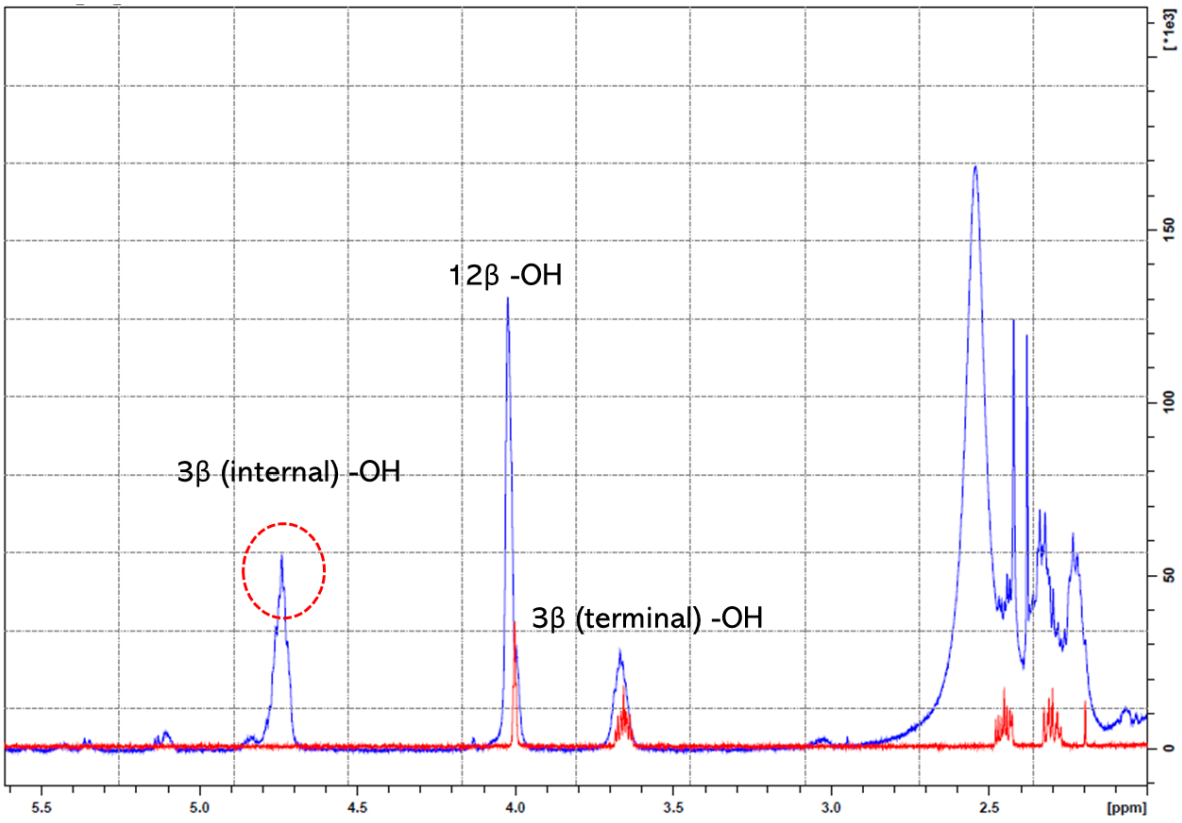


Figure 4-7: 1D Proton NMR spectra of synthesized PDC standards.

Red spectrum = DCA, blue spectrum = PDC. Mean polymer length was estimated from the ratio of the integrals of the ^{1}H signals from the internal and terminal hydrogens. In this case, signal from terminal is approximately 30, while internal is approximately 60. Mean polymer length is estimated to be 60/30 = 2+1 = 3 i.e. PDC trimer.

4.4.4 MALDI-MS detection of PDC oligomers

Finally, synthesised PDC standards were analysed by MALDI-MS for a more comprehensive detection. Oligomers up to the heptamer could be detected in linear positive mode using 2,5-dihydroxybenzoic acid (DHB, sodiated masses observed) (Figure 4-8) and nitroaniline (Figure 4-9). Both matrices confirmed the formation of PDC oligomers by the presence of repeating units of DCA minus the loss of one H_2O molecule (*m/z* ~374).

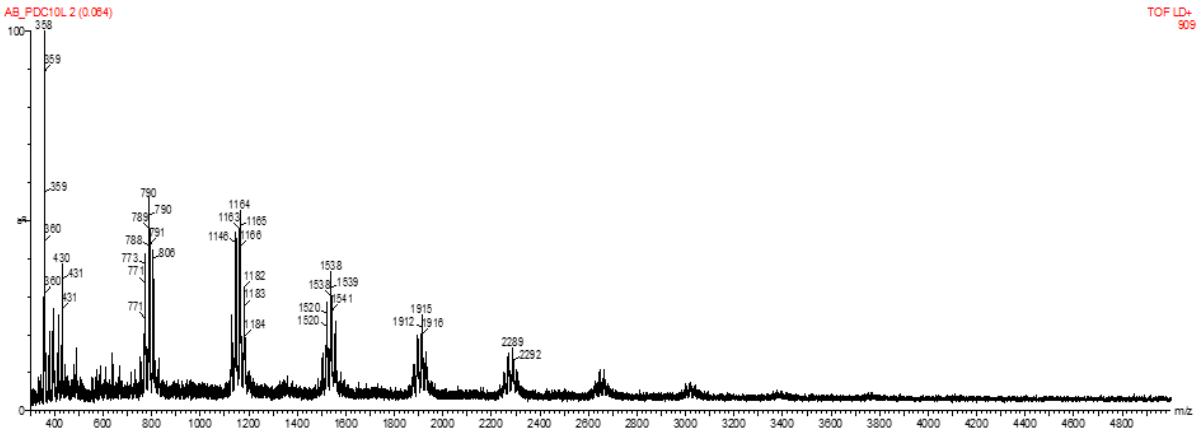


Figure 4-8: MALDI-MS chromatogram of synthesized PDC standards using 2,5-Dihydroxybenzoic acid as matrix.

The predominant DCA ion observed was the $[M+H–nH_2O]^+$ form with m/z 358. Sodiated versions $[M+Na]^+$ of the oligomers up to the heptamer are also observed but at low intensities. A difference of m/z 374 is observed between consecutive signals, corresponding to the mass of DCA minus the mass of one water molecule.

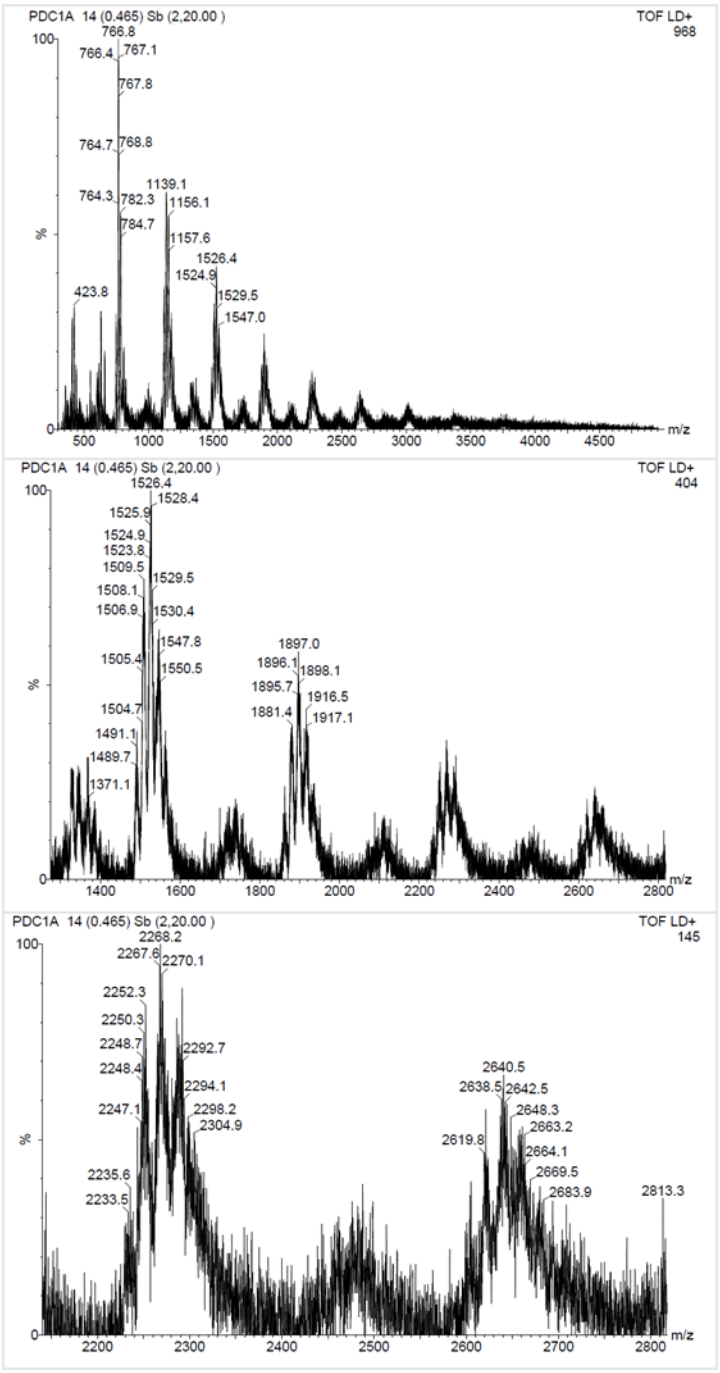


Figure 4-9: MALDI-MS chromatogram of synthesized PDC standards using nitroaniline as matrix.

Oligomers up to the heptamer are observed but at low intensities. A difference of m/z 374 is observed between consecutive signals, corresponding to the mass of DCA minus the mass of one water molecule.

4.4.5 Optimisation of PDC extraction

The extraction solvent was optimised. As toluene was not an optimal solvent being dangerous and highly volatile, other solvents were tested so as to enable high-throughput biological sample extraction and analysis. Dried down PDC standards were dissolved in different mixtures of water, IPA, ACN, MeOH and $CHCl_3$. Based on visual examination, solutions made in $CHCl_3$ provided the most complete dissolution of PDC. On analysis, oligomers up to the pentamer were detected in the $CHCl_3$ solution. As a result, $CHCl_3$ was chosen for subsequent extraction analyses.

To confirm the presence of PDC oligomers in stool, aliquots of the faecal $CHCl_3$ extracts were analysed using the OLIG method. A high level of noise was observed from the initial scan which prompted the creation of multiple reaction monitoring (MRM) windows to monitor and better detect the M-H ions of each oligomer. Also, it was expected that bigger oligomers may fragment into the smaller subunits, therefore transitions were created to monitor these as well. It should be noted that MRM windows could only be created up to the expected *m/z* for the pentamer because the mass limit for the TQ-S was *m/z* 2048 (Table 4-6). The SIM transitions for the M-H ions were found to be most stable and reproducible (Figures 4-10 – 4-12). However, peak shapes were not optimal, particularly those of DCA and the PDC-dimer (not shown). As this was only a qualitative test to confirm presence of oligomers in the faecal extracts, the obtained results were deemed adequate to confirm the presence of these ions.

Table 4-6: MRM transitions created for polydeoxycholic acid oligomers up to the pentamer for analysis on the OLIG method

PDC oligomer	Calculated mass (g/mol)	M-H (m/z ratio)	SIM/MRM transitions	RT (min)
Monomer	392.3	391.3	391.3 > 391.2	0.98
Dimer	766.57	765.2	765.21 > 765.2	1.90
			765.2 > 391.2	
Trimer	1140.86	1139.2	1139.2 > 1139.1	2.52
			1139.1 > 765.2	
			1139.15 > 391.2	
Tetramer	1515.14	1514.2	1515.2 > 1515.1	2.90
			1515.15 > 1139.2	
			1515.11 > 765.2	
			1515 > 391.2	
Pentamer	1889.42	1888.2	1888.8 > 1888.7	3.48
			1888.75 > 1515.1	
			1888.72 > 1139.1	
			1888.83 > 765.2	
			1888.86 > 391.2	

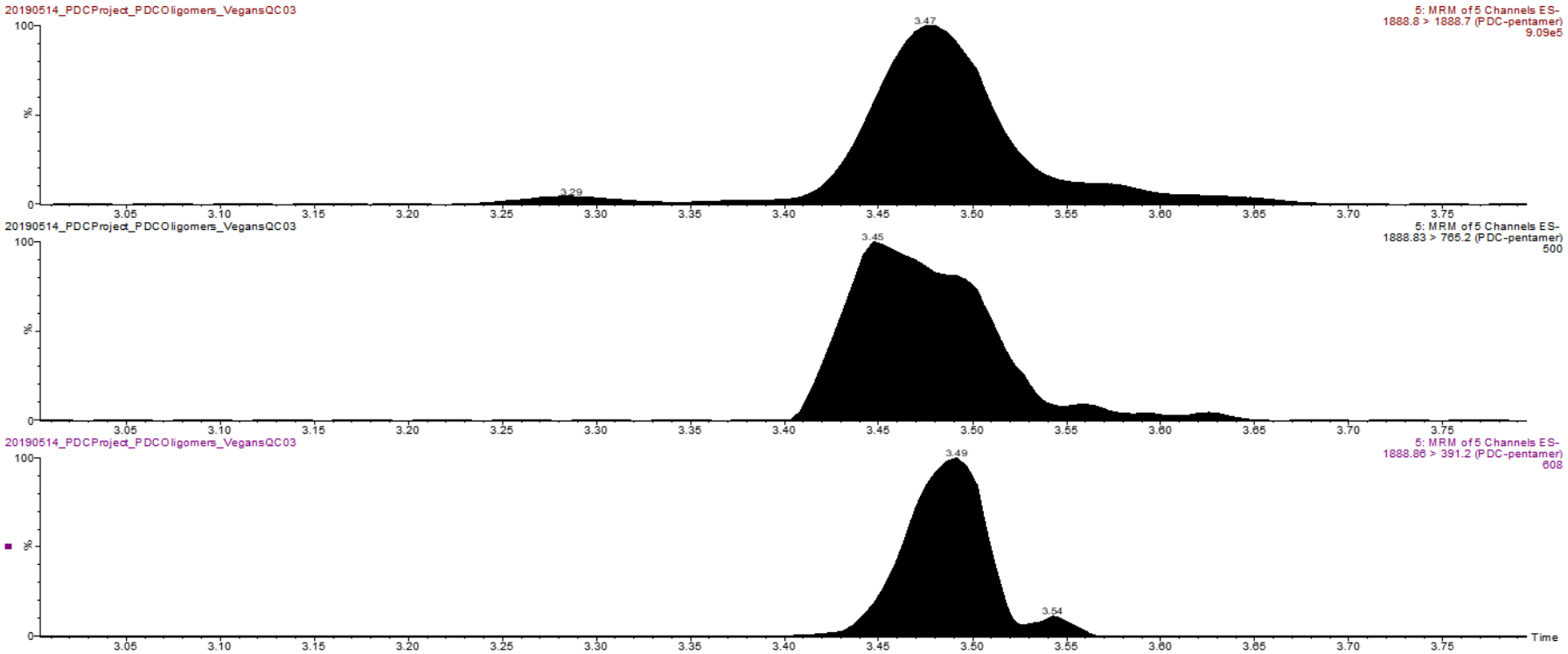


Figure 4-10: Mass chromatograms showing MRM transition peaks (after smoothing using mean 3x2) of the PDC-pentamer in faecal extracts.

The transition for the deprotonated mass was found to be most stable signal and was observed at a high intensity. Transitions of the dimer and monomer fragments were observed but at much lower intensities. No signal was detected from the two other transitions (tetramer fragment 1888.8>1515.1, and trimer fragment 1888.8>1139.1).

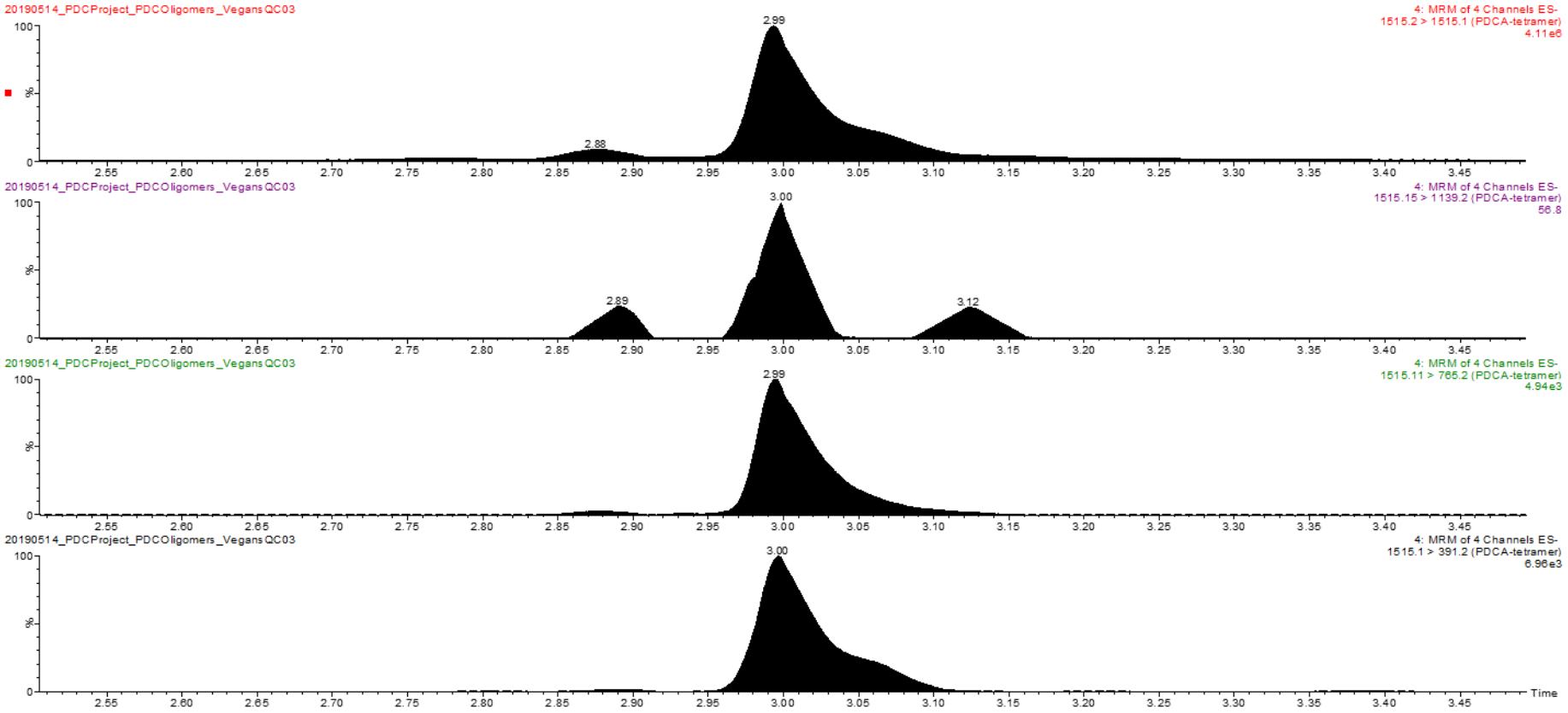


Figure 4-11: Mass chromatograms showing MRM transition peaks (after peak smoothing using mean 3x2) of the PDC-tetramer in faecal extracts.

The transition for the deprotonated mass was found to be most stable signal and was observed at a high intensity. Transitions of the trimer, dimer and monomer fragments were observed but at much lower intensities. The trimer fragment was observed at lowest intensity.

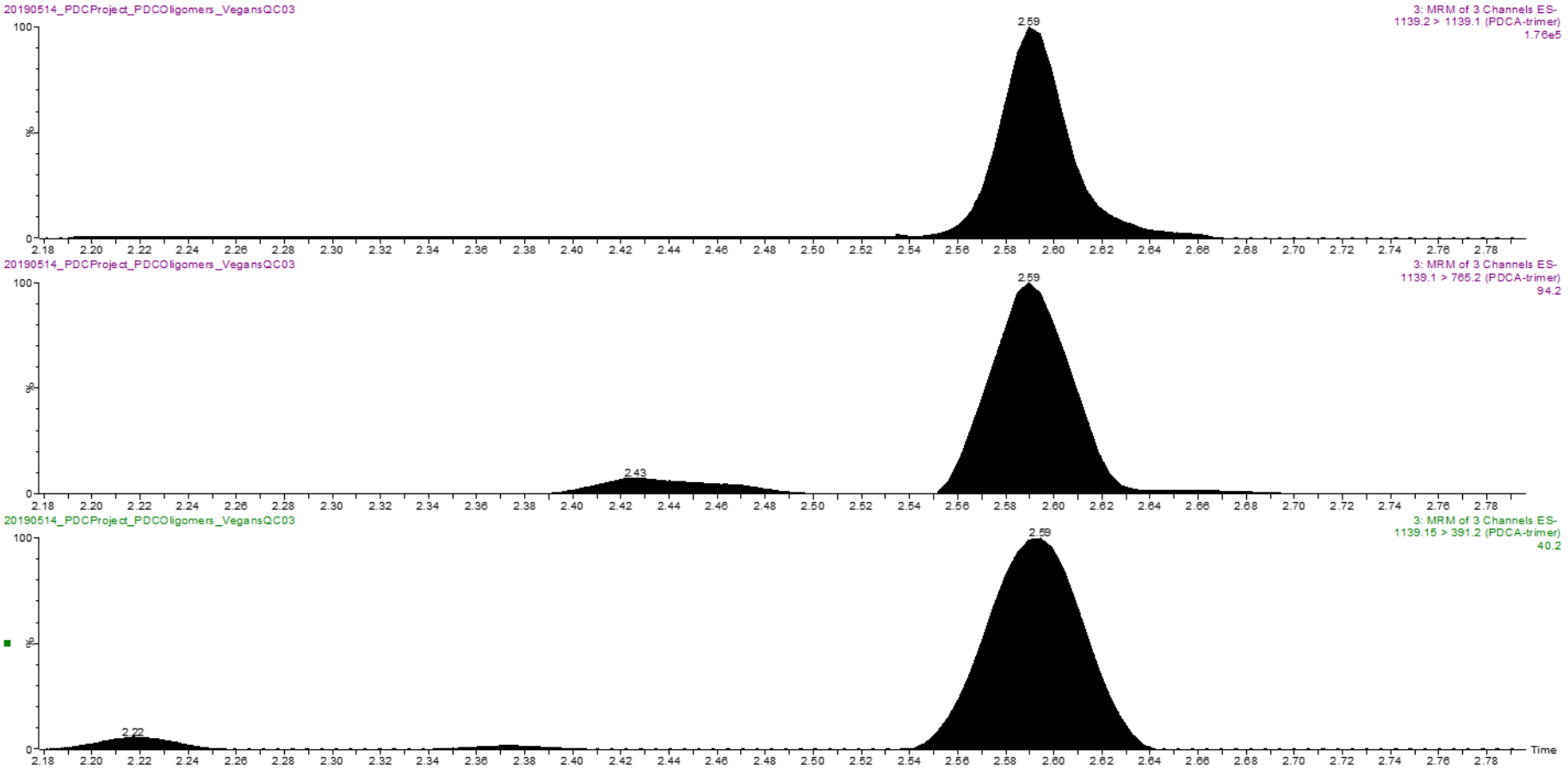


Figure 4-12: Mass chromatograms showing MRM transition peaks (after peak smoothing using mean 3x2) of the PDC-trimer in faecal extracts.

The transition for the deprotonated mass was found to be most stable signal and was observed at a high intensity. Transitions of the dimer and monomer fragments were also detected but at much lower intensities.

4.4.6 DCA and PDC Recovery in Stool Samples

Due to the unavailability of a labelled PDC standard, DCA-d4 was used as a proxy to determine the efficiency of PDC extraction in samples. Initially, DCA-d4 recoveries from a single extraction step resulted in an average 77.7(4.0)% recovery from test stool samples, while an average of 5.7% DCA-d4 was recovered from the $CHCl_3$ fraction (collected following the initial aqueous extraction of monomeric DCA, and was dried down, resuspended in 2:1:1 solution and quantified on the EL method for DCA). This suggested that some of the monomer DCA may be extracted into the organic phase. This potential bias was corrected by introducing a second 2:1:1 extraction before organic extraction yielding in a total DCA-d4 recovery (Table 4-7). indicating that the DCA quantified after $CHCl_3$ extraction as "PDC" would not be inflated by monomeric DCA.

Table 4-7: Mean % Recoveries (± standard deviation) of DCA-d4 from test faecal samples.

	AQ	ORG	AQ(x2wash)	ORG(x2wash)	%Inc in recovery
Test samples	77.72 (±4.00)	5.74 (±1.60)	109.03(±12.9)	0(0)	40.29

AQ =aqueous, ORG = organic %Inc= % increase

4.4.7 Validation of Hydrolysis of PDC Oligomers into DCA

Following a successful synthesis of PDC standards as confirmed by UPLC-MS, ^{1}H NMR and MALDI-MS, solutions of the standards were subjected to alkaline hydrolysis in order to validate the hydrolysis process. The alkaline hydrolysis successfully saponified the PDC ester bonds. This was evidenced by a consequent disappearance of oligomers and a concomitant increase in DCA concentration following an average of 15 hours of heating at 95^{o}C (Figure 4-13).

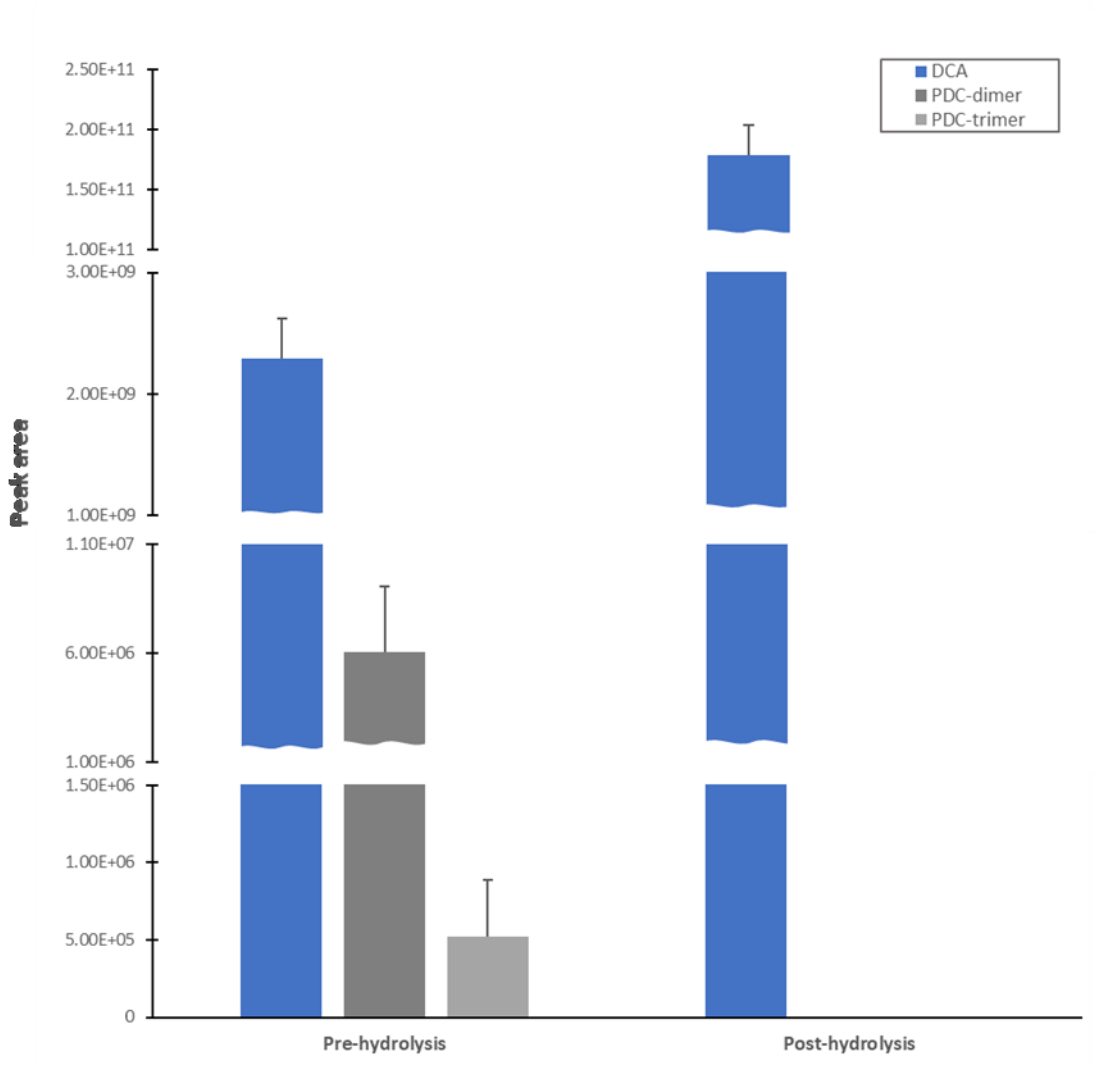


Figure 4-13: Bar graph of mean DCA and PDC peak areas before (pre-) and after (post-) PDC extraction including alkaline hydrolysis step.

The disappearance of oligomer signal and increase in DCA signal indicated that the alkaline hydrolysis broke the ester bonds in the polymers to yield DCA units. Error bars are standard deviation (SD).

4.4.8 Final Sample Preparation Workflow for Analysing PDC in Stool Samples

The final sample preparation workflow was organised into the following steps:

1. Freeze-dry faecal sample in suitable tubes.
2. Prepare 2:1:1 H_2O:ACN:IPA solution. Also prepare 2:1:1 solution containing DCA-d4 at known concentration.
3. Weigh 10-25 mg of dry sample into labelled 2 ml tubes with screw caps containing 1.0 mm Zirconia beads.
4. Add 1ml of 2:1:1 + DCA-d4 solution.
5. Ensure cap is crewed on tightly and bead beat for 3 x 30 s (with 5s breaks in between if not using pre-set program).
6. Centrifuge for 10mins at 16 000 *g* and 4°C.
7. Collect supernatant.
8. Resuspend pellet in 1 ml of 2:1:1 solution only and repeat steps 5-7. Pool supernatants from each sample. Keep pellet for PDC extraction.
9. Transfer 500 μL of pooled (resuspended) supernatant into labelled spin filters (0.45 μM pore, nylon membrane, Corning CoStar). Spin for 10 mins, 16 000 *g* , 4 °C.
10. Transfer filtrate into tube for storage at -20 °C until UPLC-MS analysis.
11. Dilute as required on day of LC-MS analysis.
12. After 2nd wash, resuspend pellet in 1 ml chloroform. Repeat steps 4-6, but centrifuge for 20 mins.
13. Collect organic supernatant. Be sure to carefully collect only the organic layer.
14. Transfer 500 μL into labelled 2 ml Eppendorf Safe Lock tubes (or suitable substitutes) to dry down overnight in fume hood.
15. Resuspend dried down residue in 250 μL 1M NaOH prepared with UPLC-grade water. Shake for 15-20 mins.
16. Incubate in water bath at 95 °C overnight (12-18 hrs). *(NB: Cap locks may be used to prevent caps from opening during incubation)*
17. Cool tubes to room temperature. Add 250 μL 1M HCl to neutralise excess alkali. Vortex.
18. Prepare 1:1 ACN:IPA solution. Add 500 μL of this to each sample. Shake for 10 mins.
19. Centrifuge for 20 mins, 16 000 *g* , 4 °C.
20. Collect supernatant and store at -20°C for UPLC-MS analysis.
21. Dilute as required on day of analysis with 2:1:1 solution containing DCA-d4 at known concentration.

4.4.9 Application to clinical samples: PDC in vegans and omnivores

The method was then applied to the same samples used in Chapter 3. A total of n = 120 faecal samples (n_{vegans} = 62, $n_{omnivores}$ = 58) were analysed for PDC. For data analysis, n = 2 and n = 1 samples were excluded because their respective DCA and PDC concentrations were above ULOQ/detector levels (*i.e.* detector saturation).

PDC concentrations were found to be comparable between the groups (p = 0.252, Figure 4-15). However, omnivores were found to have significantly higher concentrations of monomeric DCA and of total faecal DCA (fDCA) than vegans (p < 0.05, Table 4-8). Percentagewise, vegans tended to have more of their DCA as PDC (this measure was expressed as PDC%) than omnivores, but this difference did not attain statistical significance (Table 4-8).

Table 4-8: DCA and PDC measures in vegans and omnivores.

	Vegans	***n***	**Omnivores**	***n***	***P***
PDC, ng/mg dry stool	60.49(52.23)	61	71.83(43.57)	58	0.252
DCA, ng/mg dry stool	56.70(58.28)	62	82.96(133.47)	56	0.020
Total fDCA, ng/mg dry stool	125.96(88.87)	61	169.23(123.95)	56	0.015
PDC%	56.34(42.41)	61	43.89(47.46)	56	0.122

Data is presented as median (interquartile range). DCA=monomeric deoxycholic acid, PDC = Polydeoxycholic acid, total fDCA= total faecal deoxycholic acid, PDC% percentage PDC of total faecal deoxycholic acid. Differences between groups were assessed statistically by Mann Whitney U tests. p values from analyses including all data are reported because removal of outliers did not alter significance of differences.

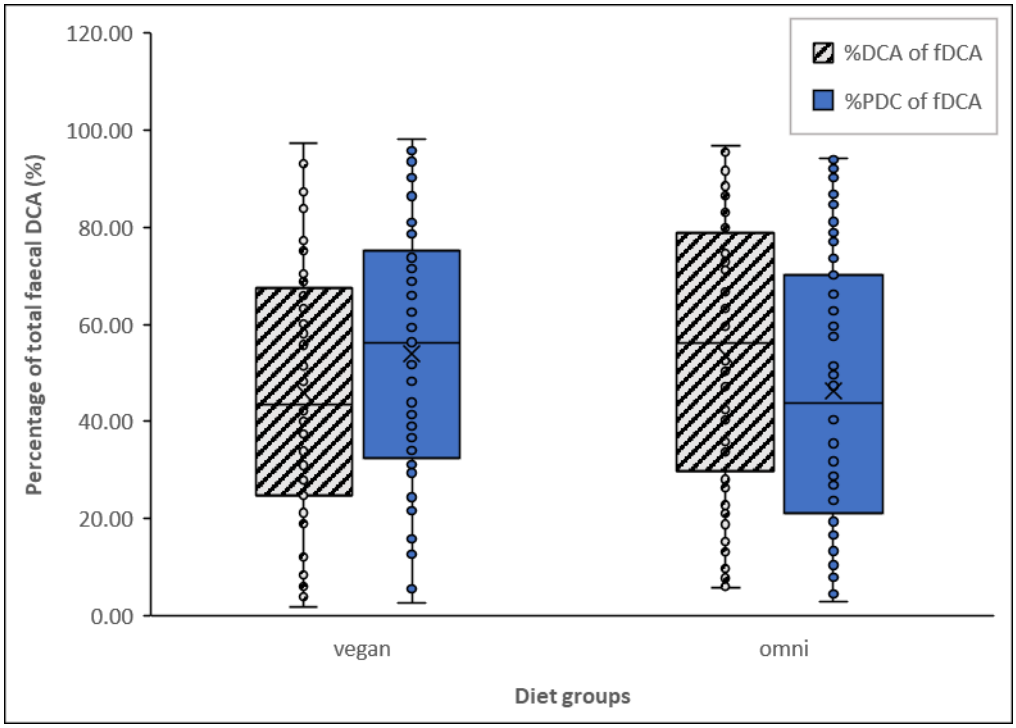


Figure 4-14: Boxplots showing percentage of total faecal deoxycholic acid (fDCA) in monomeric (DCA) and polymeric forms (PDC) in vegans vs omnivores.

Vegans tended to have more polymerised deoxycholic acid compared to omnivores, however there was no significant difference ($p>0.05$) in PDC or DCA proportions between groups.

4.5 Discussion

4.5.1 Summary of results

- Synthesis of PDC standards was successfully carried out and reliably confirmed on three analytical platforms.
- Sample preparation protocols were adapted and optimised for PDC extraction. Efficiency of extraction was assessed and found to be reliable and of high efficiency.
- A workflow was developed and validated for quantifying PDC in faecal samples.
- The workflow was successfully applied to clinical samples obtained from vegans and omnivores.

4.5.2 General discussion

Following on from a successful validation, optimisation and application of the EL method in faecal samples, the next step taken to further facilitate method development for PDC quantification was to synthesise PDC standards. The synthesis of PDC oligomers was successfully carried out using a simple distillation set-up in which the monomer substrate, DCA, was refluxed in toluene in the presence of heat and PTSA catalyst. The catalyst facilitated the condensation polymerisation of DCA into PDC oligomers, of which up to the heptamer could be reliably detected on the available analytical platforms namely UPLC-MS, ^{1}H-NMR spectroscopy and MALDI-MS.

Firstly, the EL method was explored for the possibility of detecting PDC in the samples (in toluene) collected throughout the synthesis. The PDC dimer (*m/z* 765.6) which was found to elute at 11.6 mins was the only oligomer that could be detected with this method. This was expected as PDC solubility should decrease as chain length increases, and the chromatography system in the EL method was adapted for less lipophilic compounds. However, the PDC dimer trace over time was encouraging, indicating that DCA was being converted to PDC (dimer) over time. Furthermore, the detection of the dimer on the EL method indicated that the dimer may be easily included in conventional routine analyses. However, this was not explored further as there was no purified standard for the PDC-dimer to perform an extensive method development.

Subsequently, the chromatographic system was refined to improve the detection of PDC polymers. A more lipophilic gradient was formulated (*i.e.* the OLIG method), on which PDC oligomers up to the pentamer were detected. The OLIG method was formulated with a more

lipophilic LC gradient elution considering the hydrophobicity of the PDC molecules. As PDC is highly lipophilic, MeOH may have performed better than H_2O/ACN/IPA system (of the EL method) by allowing for a quicker elution at a high organic concentration (99% MeOH, refer to Table 4-5 for gradient). No further oligomer was detected in the standard mix, although this may be because the limit of mass range on the mass spectrometer was *m/z* 2048. The traces for all of the detected oligomers showed the gradual decrease in DCA and complementing increases in the oligomers. The response signals of DCA and the PDC-dimer on the OLIG method were different from the ones obtained on the EL method. This may reflect better ionisation of DCA and the PDC-dimer with the EL-method. Unlike on the EL method where the mobile phase was H_2O/ACN/IPA (+ additives) in ~2:1:1 ratio, the OLIG method was almost 100% MeOH across the 10-minute LC run. This suggests that a more hydrophilic medium may be the preferential solvent for the DCA monomer and dimer (for future analyses). This was confirmed by the poor detection peak shape observed for the two analytes on the OLIG method. A complete/better dissolution would result in better ionisation and thus better detection of an analyte.

Faecal extracts were also analysed using the OLIG LC-gradient. To improve detection, an MS method using MRM windows were created to help filter out noise (other ions in the sample). Ions corresponding to the expected *m/z* and RTs of the PDC oligomer standards were observed in faecal extracts. The SIM transitions for the deprotonated ions were found to be most stable and reproducible. However, peak shapes were not optimal. As this was only to confirm presence of oligomers in the organic $CHCl_3$ faecal extracts, the method was only used for a quick detection and not for quantification of the detected PDC oligomers.

It is important to mention here that the UPLC-MS platform was foreseeably limited, as these oligomers were expected to be highly hydrophobic, hence expected to not be easily ionisable. It is possible that even the trimer-pentamer were only partially ionised, but the information obtained was sufficient to indicate the presence or absence of these compounds in the injected standard/faecal extract. This would suffice for the subsequent optimisation and validation of the sample preparation procedure (extraction and alkaline hydrolysis) to break the polymer into constituent DCA moieties.

More evidence of a successful polymerization was observed on the ^{1}H NMR spectra. However, the NMR could only provide information on the mean length of PDC present in the sample.

Finally, the samples from the synthesis were sent for MALDI-MS analyses. This was the most suitable of the three platforms for reliable detection and confirmation of PDC oligomers. The presence of PDC was confirmed using two different matrices. The results from MALDI-MS experiments proved undoubtedly that the PDC polymers were formed, evidenced by the

detection of signals corresponding to the molecular ions of the PDC oligomers. Also, the MALDI-ToF chromatogram showed the existence of a repeating DCA unit less the mass of one water molecule (difference of *m/z* 374).

Following successful synthesis and confirmation of the PDC standards, these compounds were used to refine the sample preparation procedure, introducing an alkaline hydrolysis step before BA extraction. The alkaline hydrolysis, also known as saponification, breaks the ester bonds between DCA subunits in the PDC polymer. The amount of DCA contained within polymer can then be quantified using the EL method providing a measure of PDC.

Analyte recovery was used as a measure of extraction efficiency. As previously done in Chapter 3, due to a lack of a true "blank" DCA-free matrix, the recovery of a spiked-in deuterated DCA internal standard (DCA-d4, *m/z* 395.06, RT 10.36mins) was investigated. However, because a PDC internal standard was unavailable, PDC recovery was ascertained indirectly by a "negative recovery" (*i.e.* absence) of DCA-d4 in the sample.

The method was then applied to clinical samples. As these were the same samples which were analysed in the validation step preceding this set of experiments, it was possible to compare the effect of the modifications made to the sample preparation on extraction efficiency. The modifications were the reduction of the starting material and the addition/inclusion of an extra 2:1:1 extraction step. Concentrations obtained from both extraction methods were significantly dissimilar ($p < 0.001$, Wilcoxon match pairs rank test, Figure 4-16A). Although concentrations from the EL validation experiments tended to be higher, concentrations obtained from both extractions were in a similar order of magnitude (micromolar range) and median-normalised analyses revealed no significant differences between "fBA23" as extracted by the earlier method (X1), and "DCA" as extracted by the modified method (X2) ($p = 0.357$, Wilcoxon matched pairs test, Figure 4-16B). Interestingly as well, similar samples were found to be the topmost and least abundant in DCA across both methods.

Furthermore, the modifications did not result in differential effects in the diet groups as observed by Wilcoxon tests performed on differences between extraction yields from the different protocols ($p = 0.622$, Mann Whitney U tests, Figure 4-17). This could also be observed in the similarity in the results obtained from both extractions; DCA was found to be significantly higher in omnivores than vegans ($p_{x1} = 0.014$ vs $p_{x2} = 0.019$, unadjusted p values, Mann Whitney U tests) even though number of samples varied. However, the coefficient of variation for the x2 extraction was lower ($CV_{x1} = 9.2\%$ vs $CV_{x2} = 4.1\%$) as observed in the biological QCs. Overall, both methods had high extraction efficiencies, gave reproducible

results to high standards and increased the validity of the results obtained (i.e. that omnivores tended to have a higher faecal DCA concentration than vegans).

The difference in per mg concentration observed in some samples may have been due to the reduction of the starting material. This was similar to observation from the tests carried out before a final choice of 10mg was made. Nevertheless, while it is important to note that as observed, inferences/conclusions on precise absolute DCA concentrations may not be possible, the results indicate with adequate confidence that DCA exists in the micromolar concentrations per dry weight in healthy human faeces in the cohort being studied.

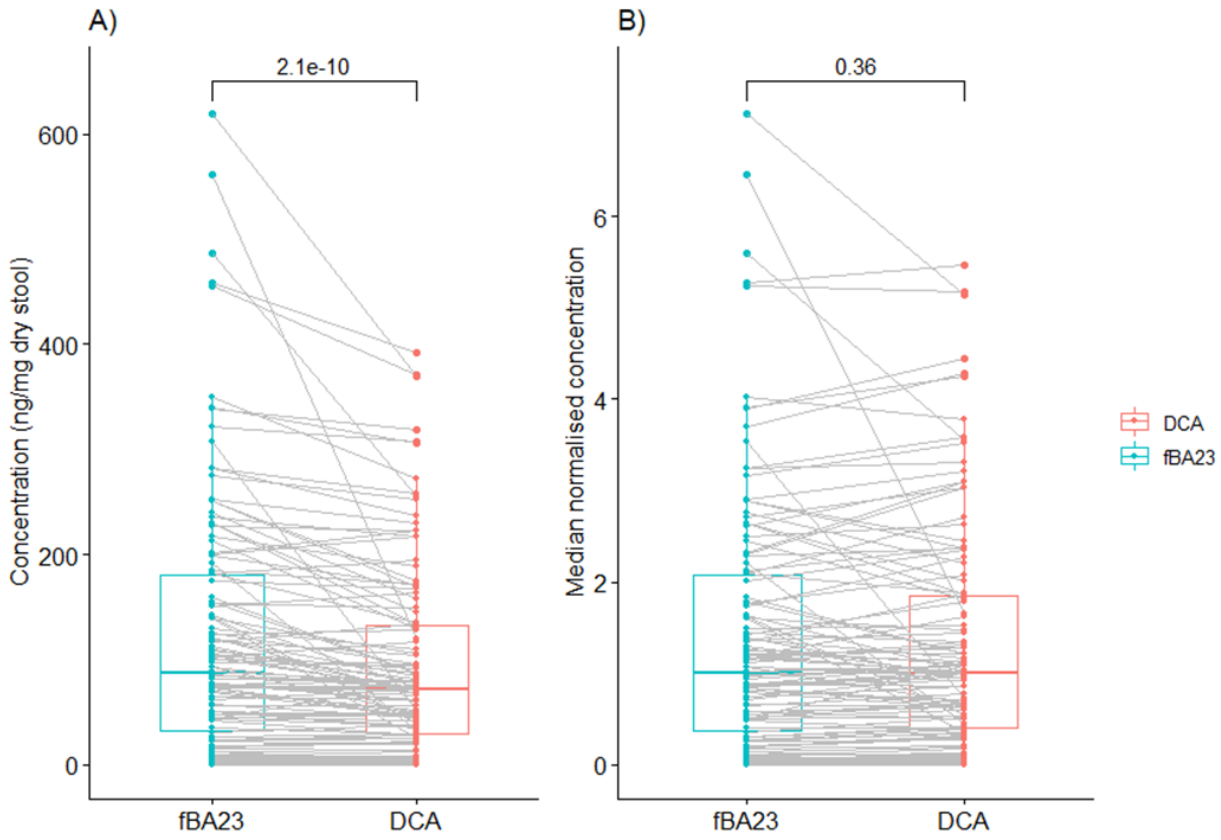


Figure 4-15: Boxplots showing DCA concentrations obtained from EL validation experiments described in the Chapter 3 ("fBA23") and modified extraction protocol from optimisation experiments described in the chapter 4 ("DCA").

The previous extraction protocol extracted 50mg of dried stool in 1ml of extraction solvent, while the modified protocol extracted 10mg of dried stool in 2ml of extraction solvent. All concentrations have been adjusted for mass of stool and dilutions. Boxplot A) shows that absolute DCA concentrations obtained from the different methods are observed to be significantly different ,although in similar range (order of magnitude). Boxplot B) shows that median normalised concentrations from either extraction are not different from the other.

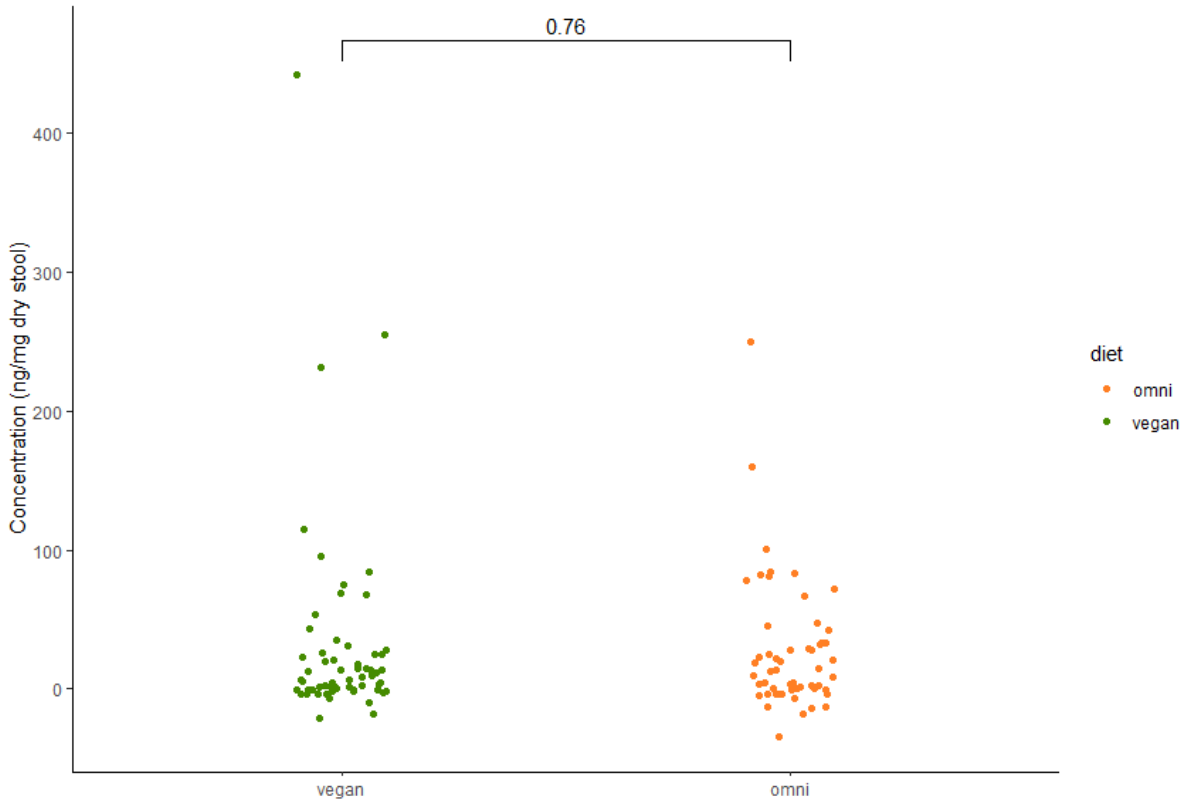


Figure 4-16: A dot plot showing the differences in DCA concentrations obtained from old extraction and modified protocols in vegans and omnivores.

This further confirms that the modifications to the extraction did not alter results differently between the diet groups.

For the first time, “PDC” was measured in a large cohort of healthy adult humans. PDC was found to make up a substantial part of total faecal DCA output (range in vegans = 2.7% - 98.3%, range in omnivores = 3.0% - 94.2%). However, there was no significant difference in PDC concentrations and PDC% proportions between vegans and omnivores. These results reflect observations made by Korpela and colleagues in vegetarian and omnivore postmenopausal women (n = 20, 10 in either group), although they reported concentrations of “saponifiable DCA” and not PDC specifically. While no statistically significant difference in PDC or %PDC was found between vegans and omnivores in this study ($p > 0.05$), a similar but wide range was observed across both groups. It would be interesting to investigate how participants with low faecal PDC and PDC% differ from those with high PDC and PDC% content.

4.5.3 Strengths and Limitations of this work

A limitation of this method development was that structural elucidation of the PDC standards was not carried out. As a result, it was not possible to ascertain if the standards formed were linear or branched as the polymerisation reaction may not be stereoselective. Although previous reports indicate that is a greater tendency for the linear polymer to be formed (polymerisation at the less sterically hindered C3-OH) (Benson et al., 1993). Benson and co reported the trace (~4%) detection by ^{1}H NMR of esterification at the C12 -OH group.

Also, PDC was not quantified at the same time as other faecal BAs, or with a similar extraction method. As observed by the differences in DCA quantified on both methods, it was therefore not possible to combine the results and estimate how much of the total BA was PDC. However, as faecal DCA is the most abundant BA in the faecal BA pool, it is safe to say that total DCA is an indicator/predictor of total BA. A very high positive correlation observed between DCA and total BA (Spearman rho, $\rho = 0.792$, $p_{raw} < 0.001$, Figure 4-17 and Figure 4-18) buttresses this assertion. Thus, by expressing PDC as a fraction of total fDCA, the resultant PDC% may be indicative of the proportion of the entire total faecal BA concentration that is PDC. Also, by computing PDC%, this would prevent the tendency to conflate participants who may have “low” PDC concentrations as a result of generally having low BA concentrations with those who have “low” PDC concentrations even though they may have high total faecal BA concentrations. This was particularly important because PDC concentrations were not found to correlate with monomer DCA concentrations (Spearman rho, $\rho = 0.035$, $p_{raw} = 0.718$, Figure 4-17 and Figure 4-19), indicating that polymerisation occurred regardless of how much DCA was present as starting material.

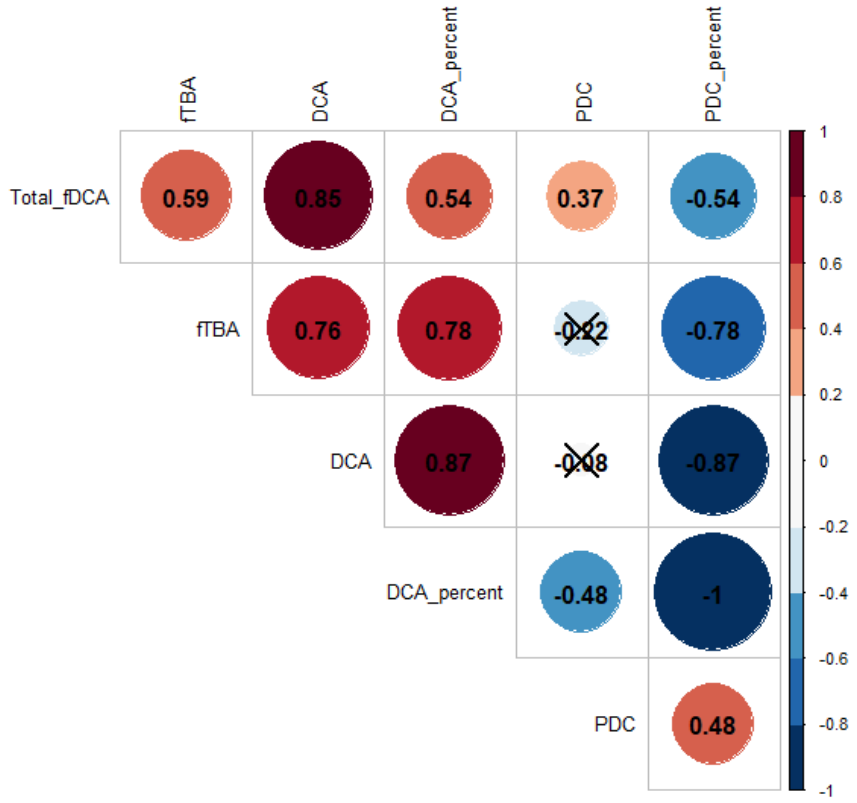


Figure 4-17: Correlation heatmap showing Spearman correlations between PDC, PDC%, DCA, DCA%, total faecal DCA (DCA + PDC) and total faecal BA in whole cohort.

Correlation coefficients are displayed for all correlations. Statistically insignificant ($p > 0.05$) correlations not represented by a colour and are crossed out with an "X".

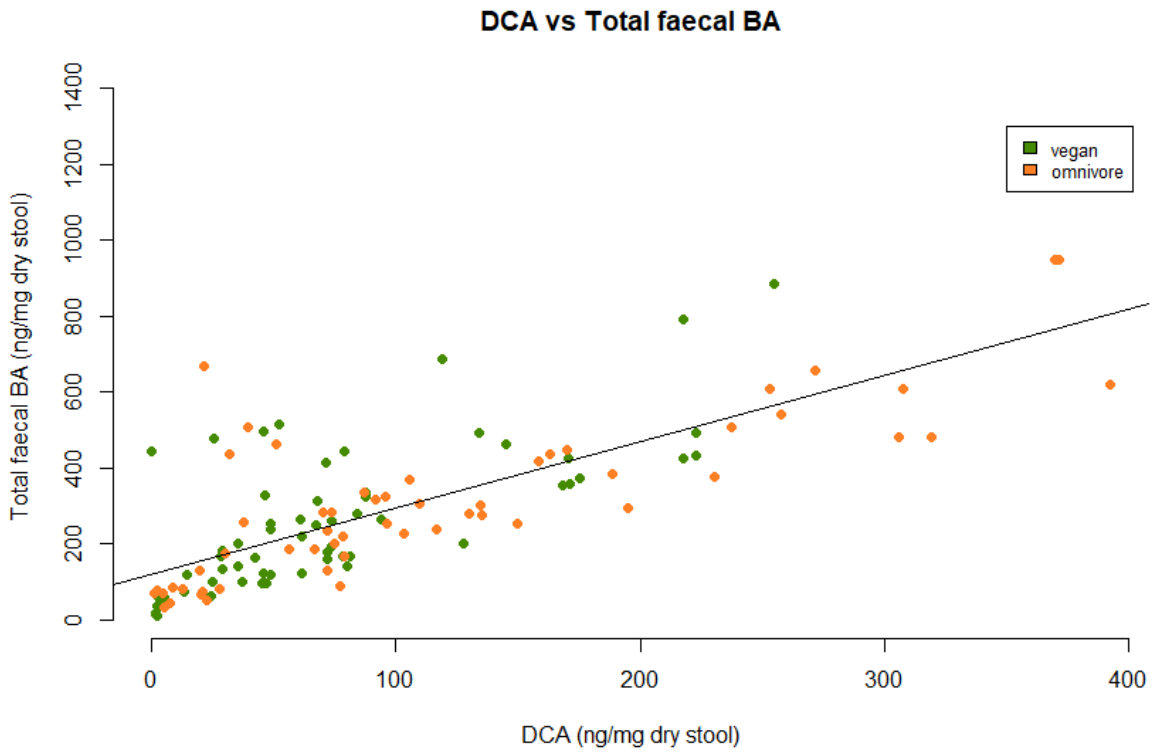


Figure 4-18: Scatterplot showing a strong correlation (Spearman rho, $\rho = 0.792$, $p_{raw} < 0.001$) between DCA and total faecal bile acid concentrations in vegans and omnivores.

A strong correlation is observed between DCA concentrations as measured by X2 method and total faecal BA as measured by X1 method.

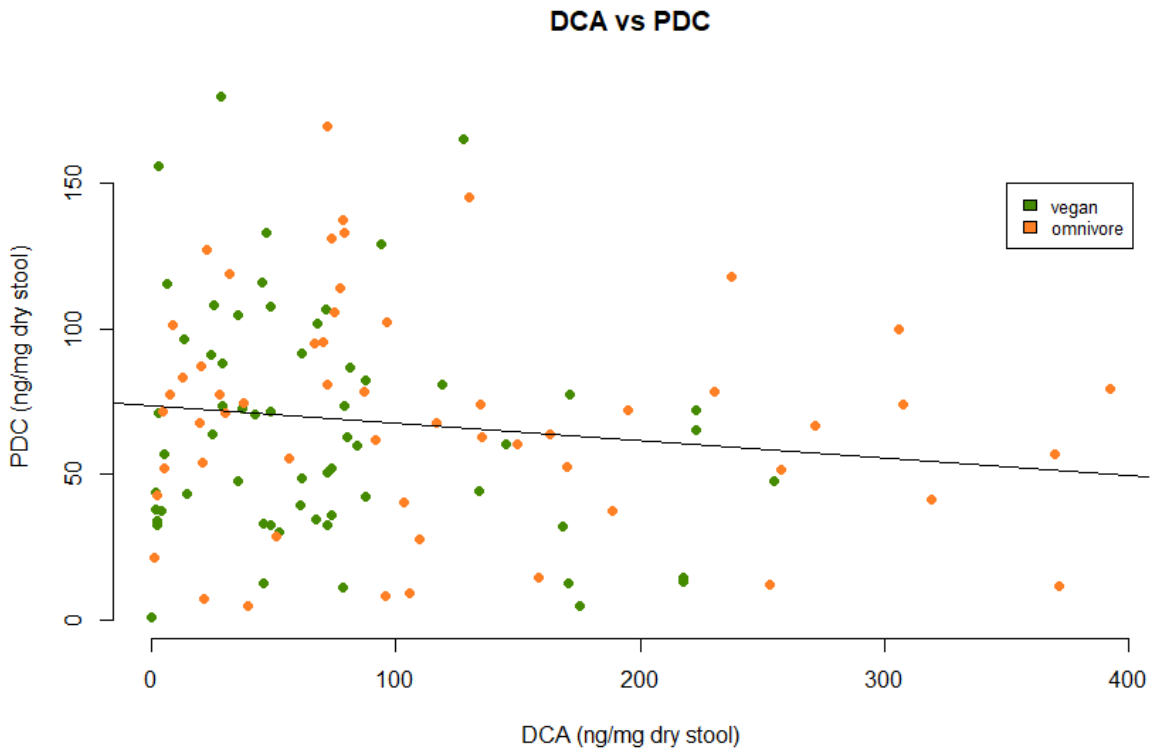


Figure 4-19: Scatterplot showing no correlation (Spearman rho, $\rho = 0.035$, $p_{raw} = 0.718$) between PDC and DCA concentrations in vegans and omnivores.

There is no correlation observed between PDC concentrations and DCA concentrations, i.e. PDC is independent of DCA.

It is important to acknowledge that one of the justifications for developing a new method was to shorten the sample preparation time. The sample preparation process as it currently stands is spread over 3 days and thus may seem counterintuitive. However, this is because the removal of $CHCl_3$ is done overnight in the fume hood at room temperature for safety reasons. In a similar vein, overnight incubation was found to be practical for application in real-life. Apart from these two steps, the protocol is therefore relatively short and not cumbersome.

Furthermore, it is possible that some of the PDC-dimer was extracted into the aqueous fraction and may not have been quantified. Solvent partitioning was considered during the method optimisation process, however the amount of DCA extracted into the $CHCl_3$ fraction was greater than that obtained from a stepwise solvent extraction, hence the latter was chosen. And since PDC-dimer could not be adequately quantified on the LC-MS method, this was accepted as a limitation of the method for further improvement.

Lastly, other saponifiable BAs were not investigated, as this was out of the scope of the primary hypothesis of this PhD.

Regardless of these limitations, the results obtained showed that a substantial part of total faecal DCA was polymerised, observing as much as 98% in some of the samples. This suggests that PDC may play a significant role in normal physiology and metabolism of the host or/and the gut microbiome. Also, an important strength of this study is that it described the development of a highly sensitive and efficient workflow for detecting and measuring PDC in human faecal samples and went on to apply this workflow in clinical samples for the first-time. Concentrations of PDC (as DCA) in a cohort of vegans and omnivores were reported for the first time. While previous reports have analysed “saponifiable BA”, none has characterised and quantified PDC before. The original paper where PDC was first reported (Benson et al., 1993), as well as a subsequent investigation by the Zhang group (J.-X. Zhang et al., 1993), conducted their investigations mainly in hamsters. Apart from the analysis by the Benson group in $n = 3$ samples as a proof of concept, this was the first time PDC was investigated in clinical samples to the best of the author’s knowledge.

4.6 Conclusion

In conclusion, PDC standards were successfully synthesised and detected on multiple analytical platforms. The synthesised standards were then used to facilitate the optimisation of an existing targeted UPLC-MS method for detecting and quantifying PDC (measured as DCA) in stool samples for the first time. Quantification of PDC was achieved by introducing an alkaline hydrolysis step to the sample preparation process in order to hydrolyse PDC polymers into constituent DCA moieties which could then be quantified on the validated UPLC-MS method. The final workflow method combines a thoroughly efficient sample preparation with a highly sensitive UPLC-MS analysis of DCA. High analyte recoveries demonstrated the efficiency of the developed workflow. Finally, the method was applied to clinical samples obtained from an observational study of vegans and omnivores.

PDC has been hypothesised to be produced by resident microbiota in the gut and may have physiological effects on the host. Although, no statistically significant difference was found between vegans and omnivores in this study, detected PDC concentrations and proportions support a possible physiological role in host or/and gut microbiota metabolism. The intestines being not as readily accessible, stool serves as the desired non-invasive proxy in most cases to enable investigations of the workings of intestinal microbial processes. However, stool as a biological sample for clinical investigation has posed a great deal of challenge due to its many complexities as a biofluid. By providing a simple workflow that involves relatively minimal sample processing, combined with a highly selective, accurate and precise LC-MS method, the study of this compound and how it may influence both host physiology and gut microbial composition may now be carried out in earnest.

Also discussed was the development of a method that would detect and quantify PDC in its oligomeric states. Both NMR and MS platforms show great promise but would need further development to be able to adapt to high-throughput analysis of clinical samples.

In furtherance of the primary aims of this PhD research work, the next chapter will discuss the further investigation of the role of PDC in cholesterol and general bile acid metabolism using data from the vegans and omnivores cohort. Further along, the method was also applied to investigate PDC in a group of healthy mildly hypercholesterolemic adults and to explore the intestinal formation of PDC *in vitro* using a validated gut model.

5. EXPLORING POLYDEOXYCHOLIC ACID *IN VIVO* IN HEALTHY HUMAN ADULT COHORT OF VEGANS AND OMNIVORES

5.1. Introduction

Plant-based diets, particularly vegetarian diets, have been shown to be associated with lower blood concentrations of cholesterol and lipoproteins (D. Li, 2011; F. Wang et al., 2015). Different forms of vegetarianism involve a high intake of dietary fibre (DF), which has been shown to lower blood cholesterol and increase the faecal excretion of bile acids (BA) (L. Brown et al., 1999). As such, it is hypothesised that fibre intake may be responsible for the hypocholesterolemic effects of plant-based/plant-majority diets. Although numerous *in vitro* studies, as well as human studies, have investigated the relationship between DF intake, circulating cholesterol levels and BA excretion, the mechanisms underlying this DF-induced hypocholesterolemic effect remain poorly understood.

Polydeoxycholic acid (PDC) is a constituent of saponifiable BA fractions that was demonstrated to inversely correlate with plasma cholesterol levels in hamsters (Benson et al., 1993). Furthermore, Korpela *et al.* reported that saponifiable BA fractions are higher in the faeces of vegetarians compared to omnivores and colorectal cancer patients (Korpela et al., 1988) and that production of these saponifiable BA could be modulated in humans by the intake of high fibre containing foods (Korpela et al., 1992). Despite the lack of significant differences in PDC concentrations and proportions between vegans and omnivores, that was reported following a successful method development and application in the previous chapter, the Vegan cohort presented an invaluable opportunity to explore the relationship of PDC with circulating cholesterol in healthy humans and the effect of diet (particularly DF intake) on its production and to identify associated gut microbial members as potential candidate PDC producers or consumers. Thus, further analyses were carried out in this cohort with an aim to better understand the role of PDC in healthy adults.

5.2. Aim and Objectives

5.2.1 Aim

To explore the relationship between polydeoxycholic acid (PDC) and the diet, gut microbiome, circulating cholesterol and bile acid (BA) metabolism in healthy adult humans.

5.2.2 Objectives

- To investigate the differences in serum lipoproteins, dietary intake and serum, urine and faecal BA signatures between participants with low-PDC and high-PDC;
- To investigate the differences in serum lipoproteins, dietary intake and serum, urine and faecal BA signatures between participants with low-PDC% and high-PDC%;
- To explore the relationship between PDC and serum cholesterol, dietary habit, dietary intake, faecal microbial composition and serum, urine and faecal BA signatures.

5.3. Methods

5.3.1 Sample collection and analysis

Study design and sample analysis for faecal BA and PDC have already been described in Chapters 3 and 4. All other data were obtained from collaborators with permission for the purposes of the research questions explored in this body of work. Specific details on how samples and data were collected, prepared and pre-processed may be found in Appendix 5.

5.3.2 Data and Statistical analyses

MS data was generated and pre-processed as previously described in Sections 3.3.12 and 4.4.8. All other data obtained were "post-processed" where necessary and combined after matching by sample ID. Data were checked for sparsity and variables with <70% non-zero or missing values (*i.e.* "sparse variables") were excluded from analyses. Analyses were further refined to only participants with complete datasets (*i.e.* have data for all of serum lipoproteins, dietary intakes, BA analyses on all three biofluids and PDC data). This was termed the complete cases (CC) set. This was done so as to avoid the unequal numbers in the datasets being integrated. The very few missing data left in the final CC dataset obtained were followed up where possible or/and imputed as zero based on the assumption that the reason for such missingness was that the analyte of interest was not detected. Data clean-up is summarised in Figure S5-1.

For descriptive statistics, categorical variables are expressed as frequency and percentages, while continuous variables are presented depending on their data distribution. Normality of data was analysed by Shapiro-Wilk test. Normally distributed data are expressed as mean (±standard deviation) while non-normally distributed data are expressed as median (interquartile range). Differences between groups were assessed by parametric independent samples t-tests or non-parametric Mann Whitney U tests as appropriate. Associations of PDC with measured variables were investigated by Spearman's rank order correlation and by complete linkage hierarchical clustering, results from which were displayed as heatmaps and dendrograms respectively. The Benjamini-Hochberg method (FDR) was employed for multiple testing correction in analyses performed on more than 12 variables.

All statistical analyses were carried out in R v3.5.2 with the following packages; *tidyverse* v1.2.1, *ltm* v1.1-1 and *purrr* v0.3.2. Plots and graphs were produced using Microsoft Excel for Office 365 or the following R packages; *ggpubr* v0.2, *ggplot2* v3.1.1, *corrplot* v0.84 and *RColorBrewer* v1.1-2.

5.4. Results

5.4.1 Summary of Patient Characteristics

A total of 150 participants were enrolled onto the study, 75 each of vegans and omnivores. However, following post-processing and the final dataset contained n = 106 participants (n_{vegans} = 57, $n_{omnivores}$ = 49). Comparisons between the "master" (M) dataset of all available data and the CC dataset revealed that there were no significant differences between the group ($p > 0.05$ in all cases, Mann Whitney U test). When compared with the excluded values ("NA" dataset), CC dataset did not differ significantly from them either save for in wholegrain intake. Results are summarised in Table 5-1.

As reported in the Chapter 4, there was no significant differences in PDC concentrations or PDC proportions (measured as PDC% of total faecal DCA) between both diet groups. However, it is important to highlight that there were other differences between vegans and omnivores, summarised in Table 5-2. Vegans had a lower BMI (p = 0.017) as well as lower total serum cholesterol ($p < 0.001$) and serum LDL cholesterol ($p < 0.001$) levels than their omnivore counterparts. There were no significant differences in serum concentrations of HDL-cholesterol, VLDL-cholesterol or triglycerides between groups.

Although average daily energy intakes were similar between groups, vegan and omnivore diets had dissimilar macronutrient profiles (Table 5-2). Vegans were shown to consume more carbohydrates but less protein and fat than omnivores. Also, vegans consumed significantly greater daily amounts of fruits, vegetables, nuts and legumes than omnivores ($p \leq 0.001$ in all cases). However, daily wholegrain intakes did not differ between both groups.

In all three biofluids analysed, there were no significant between-group differences in total BA concentrations detected (Table 5-2). Similar to faecal BA concentrations observed in the entire group (n = 131) as presented in Chapter 3, omnivores in the subset data tended to have a higher concentration of total faecal BA than vegans however in this case, the difference did not attain statistical significance. Finally, there was no significant difference between vegans and omnivores in their PDC concentrations or PDC proportions of total faecal DCA (total fDCA). Compared to the entire group (n = 117), there was a trend for omnivores in the subset to have higher total fDCA concentrations than vegans however this did not reach significance (p = 0.053).

Table 5-1: Demographic and clinical characteristics of study participants in the subset data (complete cases "CC" dataset) compared to the dataset including all participants (master "M" dataset) and excluded values ("NA" dataset).

	M dataset	*N*	***CC dataset***	***n***	*NA*	*n*	*P1*	*P2*
Diet habit, no of vegans (%)	75 (50%)	150	**57(53.8%)**	**106**	18(40.9%)	44		
Sex, no. Female (%)	80 (53.3%)	150	**55 (51.9%)**	**106**	25(56.8%)	44		
Age, years	28.7(10.5)	150	**29.3(11.1)**	**106**	27.8(7.5)	44	0.628	0.243
BMI, kg/m²	21.3(2.4)	150	**21.3(2.6)**	**106**	21.3(2.4)	44	0.850	0.648
Waist circumference, cm	76.0(8.8)	150	**76.5(9.8)**	**106**	75.5(8.0)	44	0.677	0.315
Body fat percentage, %*	20.0(8.8)	150	**20.1(8.7)**	**106**	19.7(8.9)	44	0.867	0.687
Total cholesterol, mmol/L	4.1(1.1)	150	**4.0(1.2)**	**106**	4.2(0.9)	44	0.899	0.760
HDL cholesterol, mmol/L	1.4(0.4)	150	**1.3(0.4)**	**106**	1.4(0.4)	44	0.804	0.549
LDL cholesterol, mmol/L	2.3(0.9)	150	**2.3(0.9)**	**106**	2.4(1.0)	44	0.857	0.499
Triglycerides, mmol/L	0.9(0.4)	150	**0.9(0.4)**	**106**	0.8(0.4)	44	0.686	0.331
VLDL cholesterol, mmol/L	0.4(0.2)	150	**0.4(0.2)**	**106**	0.4(0.2)	44	0.727	0.400
Energy, kJ	9726.5(4042.9)	142	**9799.8(3853.1)**	**106**	9693.8(3864.4)	36	0.743	0.390
Carbohydrate, g/day	278.4(142.3)	142	**281.6(139.9)**	**106**	252.7(133.7)	36	0.633	0.211
Protein, g/day	76.7(33.3)	142	**78.2(31.5)**	**106**	67.6(34.8)	36	0.462	0.054
Fat, g/day	75.1(39.0)	142	**73.7(40.6)**	**106**	78.7(33.5)	36	0.919	0.789
Fibre, g/day	34.7(19.0)	142	**35.7(19.1)**	**106**	30.5(14.2)	36	0.639	0.219
Fibre, g/1000kj	3.7(1.6)	142	**3.7(1.6)**	**106**	3.8(1.5)	36	0.825	0.562
Fruits, g/day	206.9(228.5)	142	**205.5(204.6)**	**106**	215.1(214.7)	36	0.693	0.300
Nuts and legumes, g/day	44.4(67.5)	142	**43.7(70.1)**	**106**	45.9(67.7)	36	0.740	0.384
Vegetables, g/day	255.4(224.1)	142	**251.8(197.3)**	**106**	266.3(279.8)	36	0.867	0.661
Wholegrains, g/day	145.9(126.3)	142	**157.7(130.8)**	**106**	108.4(123.5)	36	0.184	<0.001
Total faecal bile acids, ng/mg dry stool	253.3(309.0)	131	**244.5(285.4)**	**106**	330.3(375.3)	25	0.786	0.404
Total serum bile acids, µM	38.1(39.5)	147	**40.6(40.2)**	**106**	30.5(33.1)	41	0.481	0.081
Total urine bile acids, µM	1.6(3.2)	144	**1.6(3.3)**	**106**	1.8(3.0)	38	0.912	0.777
PDC, ng/mg dry stool	67.4(51.6)	119	**69.2(49.9)**	**106**	63.0(72.0)	13	0.950	0.795
DCA, ng/mg dry stool	71.8(103.9)	118	**70.9(97.3)**	**106**	82.5(114.9)	12	0.829	0.352
Total faecal DCA, ng/mg dry stool	150.6(106.2)	117	**143.8(106.3)**	**106**	183.0(68.4)	11	0.809	0.277
PDC% (of total faecal DCA)	49.6(46.7)	117	**51.6(46.5)**	**106**	32.5(36.7)	11	0.823	0.315

Data is displayed as median (interquartile range) and significance testing was carried out by Mann Whitney U test. P1 = p-value for M dataset vs CC dataset, P2 = p-value for CC dataset vs NA dataset.

Table 5-2: Demographic and clinical characteristics of vegans and omnivores in the subset data (n=106).

	Vegans	*Omnivores*	*P*
No of samples, *n*	57	49	
Sex, no. Female (%)	30 (52.6)	25 (51.0%)	
Age, years	28.1(10.6)	31.2(10.1)	0.353
BMI, kg/m^2	20.9(2.2)	22.1(1.8)	0.017
Waist circumference, cm*	77.4(±6.7)	76.6(±6.8)	0.526
Body fat percentage, %*	19.4(±6.1)	19.3(±5.4)	0.898
Total cholesterol, mmol/L*	3.9(±0.7)	4.5(±0.9)	<0.001
HDL cholesterol, mmol/L	1.3(0.4)	1.4(0.4)	0.148
LDL cholesterol, mmol/L*	2.1(±0.6)	2.6(±0.8)	<0.001
Triglycerides, mmol/L	1.0(0.4)	0.8(0.4)	0.150
VLDL cholesterol, mmol/L	0.4(0.2)	0.4(0.2)	0.226
Energy, kJ*	10265.2(±2862.2)	10051.4(±2780.5)	0.698
Carbohydrate, g/day	312.1(148.6)	253.7(95.2)	0.001
Protein, g/day	68.5(30.0)	86.3(41.1)	<0.001
Fat, g/day	67.0(36.1)	90.2(42.4)	0.001
Fibre, g/day	43.1(22.7)	29.1(11.7)	<0.001
Fibre, g/1000kj	4.5(1.2)	3.0(0.8)	<0.001
Fruits, g/day	267.3(294.6)	168.8(137.9)	<0.001
Nuts and legumes, g/day	74.3(100.8)	21.3(37.0)	<0.001
Vegetables, g/day	307.4(254)	217.9(139.3)	<0.001
Wholegrains, g/day	175.0(130.0)	144.6(132.4)	0.284
Total faecal bile acids, ng/mg dry stool	200.4(252.4)	281.6(290.3)	0.217
Total serum bile acids, µM	43.9(34.9)	38.6(44.0)	0.689
Total urine bile acids, µM	1.6(2.0)	1.5(4.7)	0.864
PDC, ng/mg dry stool*	65.8(±38.2)	71.7(±38.8)	0.429
DCA, ng/mg dry stool	61.1(55.0)	79.0(130.7)	0.085
Total faecal DCA, ng/mg dry stool	122.5(88.9)	161.6(121.2)	0.053
PDC% (of total faecal DCA)	56.3(40.6)	49.6(48.2)	0.284

Data is displayed as median (interquartile range) and significance testing was carried out by Mann Whitney U test unless otherwise stated. *Data is displayed as mean(± standard deviation) and significance testing was carried out by independent samples t-test.

5.4.2 Investigating lower and upper quartiles of PDC

As observed in Figure 5-1, PDC concentrations of participants belonging to both vegan and omnivore diet groups were found to be similar ($p > 0.05$) and covered the same range (range; 0.9 – 179.9 ng/mg dry stool). To further understand the distinguishing characteristics of participants who had higher faecal PDC concentrations from those with lower PDC concentrations, a post hoc extreme group analysis (EGA) was performed. In these analyses, both diet groups were combined and divided into quartiles based on PDC concentrations, the lower and upper quartile samples were re-categorised into "low PDC" (*i.e.* ≤ Q1 PDC) and "high PDC" (*i.e.* ≥ Q3 PDC) groups respectively and were compared.

The low-PDC group comprised of 17 vegans and 10 omnivores, while the high-PDC group had 14 vegans and 13 omnivores. In addition to having a higher total faecal DCA concentration and percentage PDC, the latter were also more than twice as likely to be male and had higher caloric and dietary protein intakes ($p < 0.05$, Table 5-3). They also tended to have a higher dietary fat intake ($p = 0.057$). The high PDC group also had higher concentrations of allolithocholic acid (allo-LCA) in their faeces ($p_{adj} < 0.001$). All other variables were similar across the groups (Table 5-3). Similar results were observed in the master dataset.

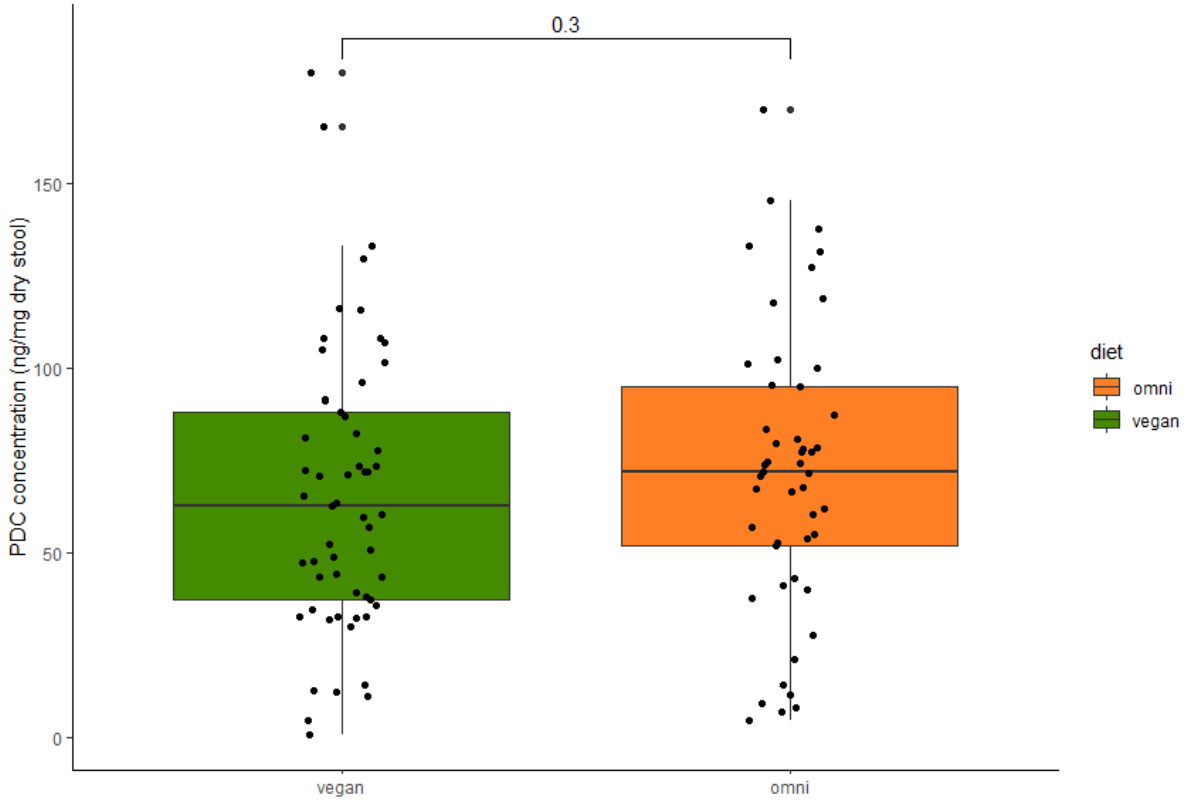


Figure 5-1: Boxplot showing the spread of concentrations of PDC concentrations (ng/mg dry stool) in vegans (n = 57) and omnivores(n = 49).

A similar trend and range are observed in both diet groups ($p > 0.05$, Independent sample t test).

Table 5-3: Demographic, clinical and lab data for lower ("low PDC") and upper ("high PDC") quartile samples for PDC concentrations (ng per mg dry stool).

	Low PDC	***High PDC***	***P***
No of participants, n	27	27	
Sex, no. Female (%)	18 (66.7%)	8 (29.6%)	
Diet habit, no of vegans (%)	17 (63%)	14 (51.9%)	
PDC, ng/mg dry stool	**27.6(22.4)**	**115.6(30.7)**	**<0.001**
DCA, ng/mg dry stool	72(91.3)	66.7(48.6)	0.473
Total faecal DCA, ng/mg dry stool	**102.0(107.0)**	**169.8(63.7)**	**<0.001**
PDC% (of total faecal DCA)	**24.9(30.1)**	**64.0(20.5)**	**<0.001**
Age, years	28.3(8.0)	27.8(9.7)	0.353
*BMI, kg/m²**	21.4(±2.0)	21.0(±1.7)	0.411
*Waist circumference, cm**	75.6(±6.6)	78.5(±7.2)	0.130
Body fat percentage, %	21.1(9.6)	17.6(10.8)	0.117
Total cholesterol, mmol/L	3.9(0.9)	3.9(1.4)	0.762
HDL cholesterol, mmol/L	1.4(0.4)	1.3(0.4)	0.303
*LDL cholesterol, mmol/L**	2.3(±0.5)	2.3(±0.9)	0.924
Triglycerides, mmol/L	0.9(0.3)	0.9(0.6)	0.723
VLDL cholesterol, mmol/L	0.4(0.2)	0.4(0.3)	0.797
Energy, kJ	**9007.3(3001.3)**	**10381.9(2166.4)**	**0.026**
Carbohydrate, g/day	262.4(113)	302.5(107.6)	0.294
Protein, g/day	**72.2(27.4)**	**86.9(46.2)**	**0.025**
Fat, g/day	**67.0(37.3)**	**92.9(37.1)**	**0.057**
Fibre, g/day	36.0(18.3)	35.9(14.8)	0.471
Fibre, g/1000kj	3.8(1.2)	3.5(1.6)	0.655
Fruits, g/day	202.3(224.6)	168.8(172.4)	0.856
Nuts and legumes, g/day	45.7(61.6)	48.5(50.3)	0.556
Vegetables, g/day	270.8(219.1)	268.8(151.1)	0.503
Wholegrains, g/day	149.0(105.6)	188.8(103.2)	0.177
Total faecal bile acids, ng/mg dry stool	323.5(270.7)	202.0(174.6)	0.097
Total serum bile acids, μM	52.3(44.0)	38.2(32.1)	0.472
Total urine bile acids, μM	1.4(1.5)	1.6(3.1)	0.599

Results are presented as median (interquartile range) and between-group differences were assessed for significance by Mann Whitney U tests except otherwise indicated. *Data was normally distributed and thus presented as mean(±standard deviation) and tested using parametric Independent samples t-tests. Variables that were significantly different between the two quartiles are highlighted in bold.

5.4.3 Investigating lower and upper quartiles of PDC%

As observed in Figure 5-2, PDC% were found to be similar between vegans and omnivores over a similar range = 2.7 - 98.3 %PDC). To investigate drivers for increasing the amount of faecal DCA in PDC-form, the diet groups were combined and divided into quartiles based on PDC proportions expressed as %PDC of total fDCA in an EGA. The lower and upper quartile samples were recategorised into "low PDC%" (*i.e.* ≤ Q1 %PDC) or "high PDC%" (*i.e.* ≥ Q3 %PDC) groups respectively and were compared. This investigation was particularly important because PDC proportion controlled for the difference in total faecal DCA which was observed to vary widely across individuals and largely determine the total faecal BA concentration (results obtained in Chapter 3 demonstrated that DCA was the most abundant faecal BA and as such contributed mostly to the total faecal BA pool).

The low-PDC% group comprised of 10 vegans and 17 omnivores, while high-PDC% group was composed of 16 vegans and 11 omnivores. Total fDCA and total faecal BA ($p_{adj} < 0.001$ in both cases, Figures 5-3 and 5-4 respectively) were significantly higher in low-PDC% group. In contrast, the high-PDC% group had significantly higher faecal concentrations of PDC ($p < 0.001$, Figure 5-3A) and intake of nuts and legumes ($p = 0.004$, Figure 5-5). Fibre intake per 1000 kJ tended to be significantly higher in the high PDC% group ($p = 0.052$, Figure 5-6A). A tendency to significance was observed in vegetable intakes between the two groups ($p = 0.081$,), however, on closer investigation, this was found to be driven largely by outliers in the high PDC% group. In BAs where there was a significant difference between the low- and high-PDC% groups, the former were found to have significantly higher concentrations of all BAs in all three biofluids especially of DCA, taurine and glycine conjugated DCA, as well as 5β-cholanic acid-3β, 12α-diol (a 3β-isomer of DCA. No other statistically significant differences were observed between the two extreme quartiles.

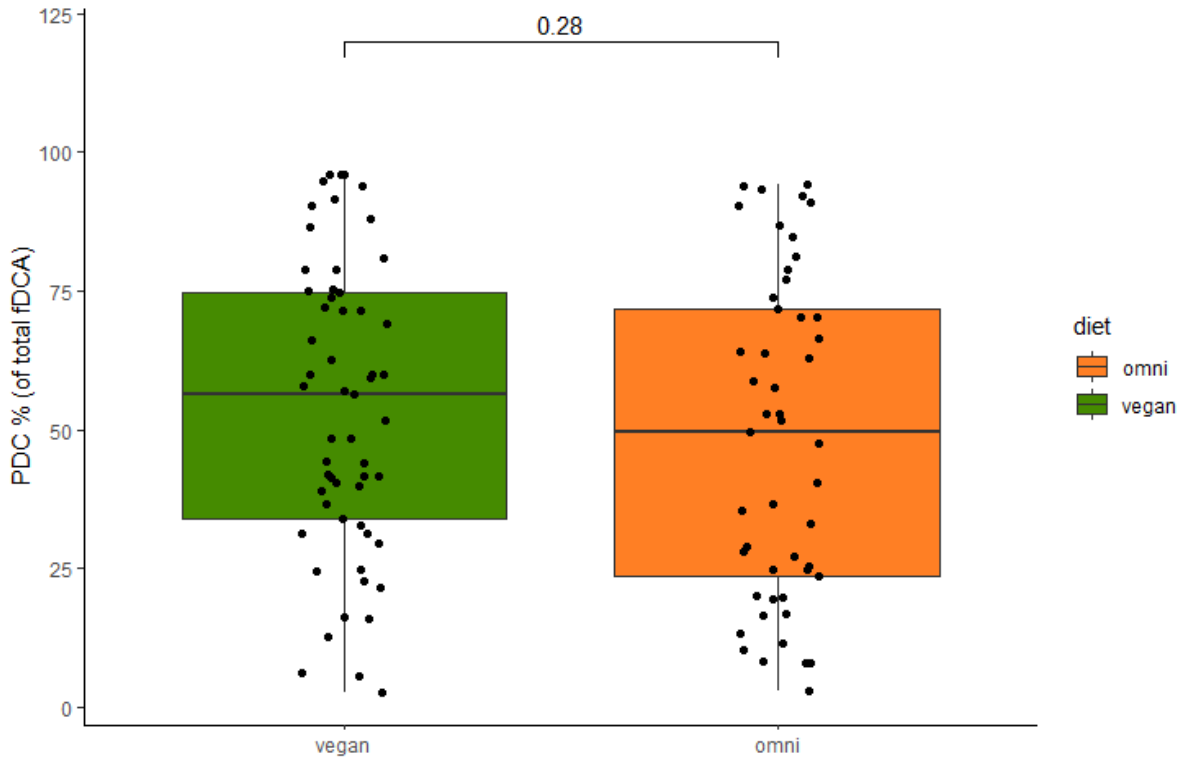


Figure 5-2: Boxplot showing PDC percentage of total faecal DCA output in vegans and omnivores.

No significant difference was found between the two groups ($p>0.05$, Mann Whitney U tests), however, a wide and similar range was observed across both groups.

Table 5-4: Demographic, clinical and lab data for lower ("low PDC%") and upper ("high PDC%") quartile samples for percentage PDC proportions.

	Low PDC %	High PDC %	*P*
No. of participants, n	27	27	
Sex, no. Female (%)	15(55.6%)	12(44.4%)	
Diet habit, no of vegans (%)	10(37.0%)	16(59.3%)	
PDC, ng/mg dry stool	**37.6(53.9)**	**77.5(59.6)**	**<0.001**
DCA, ng/mg dry stool	**194.9(141.5)**	**8.6(20.3)**	**<0.001**
Total faecal DCA, ng/mg dry stool*	**238.9(±122.9)**	**91.5(±50.0)**	**<0.001**
PDC% (of total faecal DCA)	**16.6(13.8)**	**87.9(14.8)**	**<0.001**
Age, years	31.6(11.5)	30.5(8.1)	0.891
BMI, kg/m²*	21.7(±1.8)	21.1(±1.9)	0.204
Waist circumference, cm*	75.9(±5.7)	78.1(±6.2)	0.175
Body fat percentage, %*	19.8(±6.1)	18.9(±6.0)	0.577
Total cholesterol, mmol/L*	4.4(±0.8)	4.2(±0.9)	0.287
HDL cholesterol, mmol/L	1.5(0.4)	1.3(0.3)	0.359
LDL cholesterol, mmol/L*	2.4(±0.6)	2.4(±0.8)	0.848
Triglycerides, mmol/L	1.0(0.4)	0.8(0.4)	0.191
VLDL cholesterol, mmol/L	0.4(0.2)	0.4(0.2)	0.184
Energy, Kj	10453.4(4659.4)	9308.4(1985.4)	0.655
Carbohydrate, g/day	275.1(150.8)	283.2(129.9)	0.986
Protein, g/day	86.0(40.3)	73.4(22.1)	0.216
Fat, g/day	70.6(43.0)	68.8(27.9)	0.864
Fibre, g/day	32.2(21.4)	35.9(17.8)	0.235
Fibre, g/1000kj*	**3.4(±1.2)**	**4.0(±0.9)**	**0.052**
Fruits, g/day	169.3(260.4)	191.0(241.7)	0.299
Nuts and legumes, g/day	**15.3(45.6)**	**49.3(82.4)**	**0.004**
Vegetables, g/day	193.0(188.4)	252.3(138.2)	0.087
Wholegrains, g/day*	175.9(±101.4)	189.8(±97.3)	0.609
Total faecal bile acids, ng/mg dry stool	**446.2(243.9)**	**70.2(74.5)**	**<0.001**
Total serum bile acids, µM	45.3(42.5)	40.0(28.3)	0.862
Total urine bile acids, µM	1.5(3.2)	1.8(4.7)	0.856

Results are expressed as median(interquartile range) and between-group differences were assessed for significance by Mann Whitney U tests except otherwise indicated. *Results are expressed as mean(±standard deviation) and group differences were assessed by Independent samples t-tests. Cells of variables that were significantly different between the two quartiles are highlighted in grey.

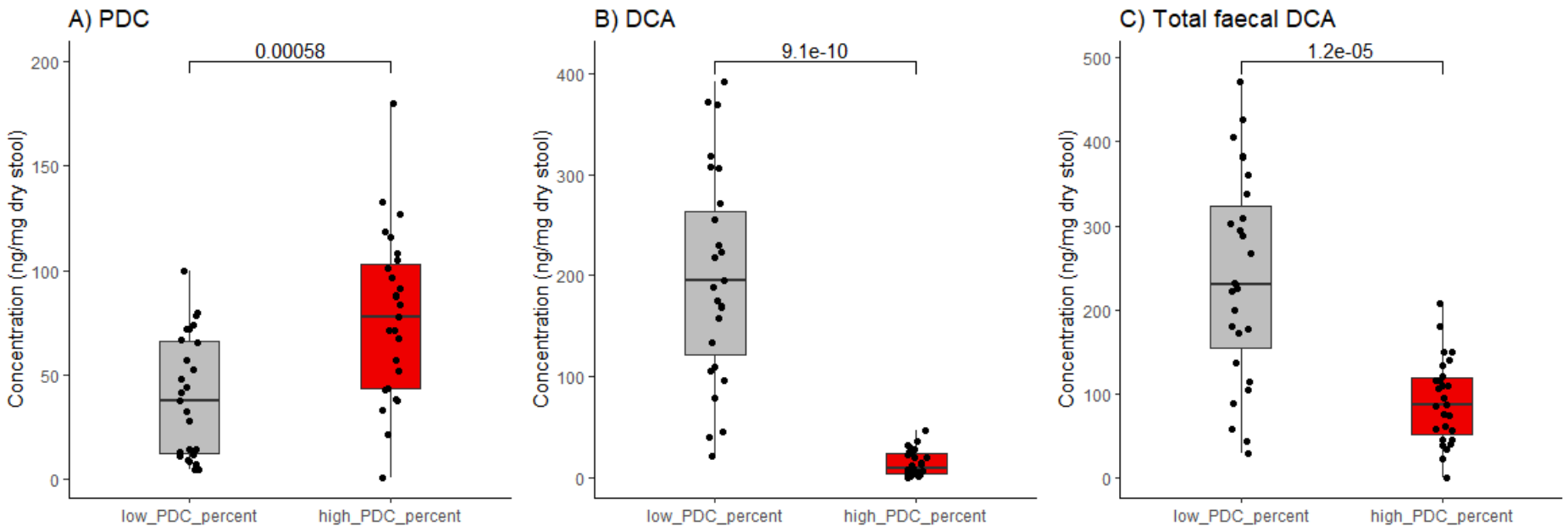


Figure 5-3: Boxplots showing A)PDC B)DCA and C)total faecal DCA concentrations in low-%PDC and high-%PDC groups.

Extreme group analyses show differences in PDC, DCA and total faecal DCA concentrations (ng/mg dry stool) between low (Q1) PDC% and high (Q4) PDC% quartiles. PDC and DCA concentrations were tested using Mann Whitney U statistical tests while difference in total faecal DCA was tested by independent samples t-test (significance set at 0.05).

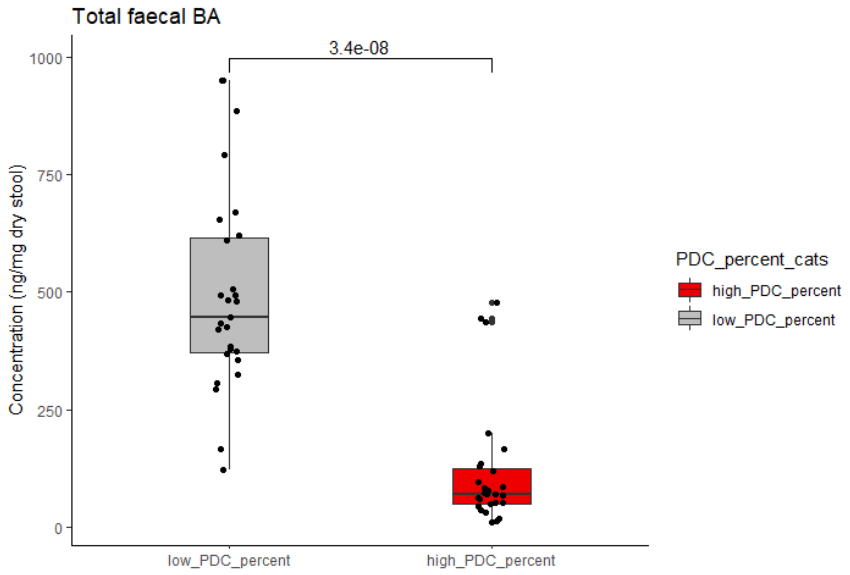


Figure 5-4: Boxplots showing total faecal bile acid concentrations between low %PDC and high %PDC groups.

Removal of outliers did not alter statistical significance of difference observed between low-PDC% and high-PDC% groups (significance set at 0.05, Mann Whitney U).

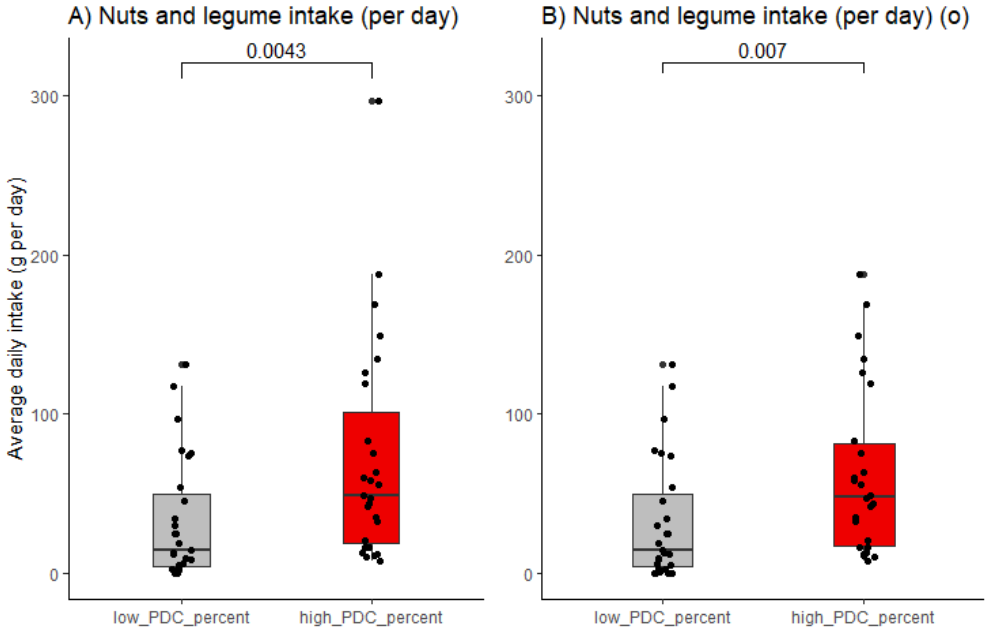


Figure 5-5: Boxplots showing average nut and legume intakes A)with and B)without outliers in low-%PDC and high-%PDC groups.

Removal of outliers did not alter significance of difference detected between the groups being compared (significance set at 0.05, Mann Whitney U).

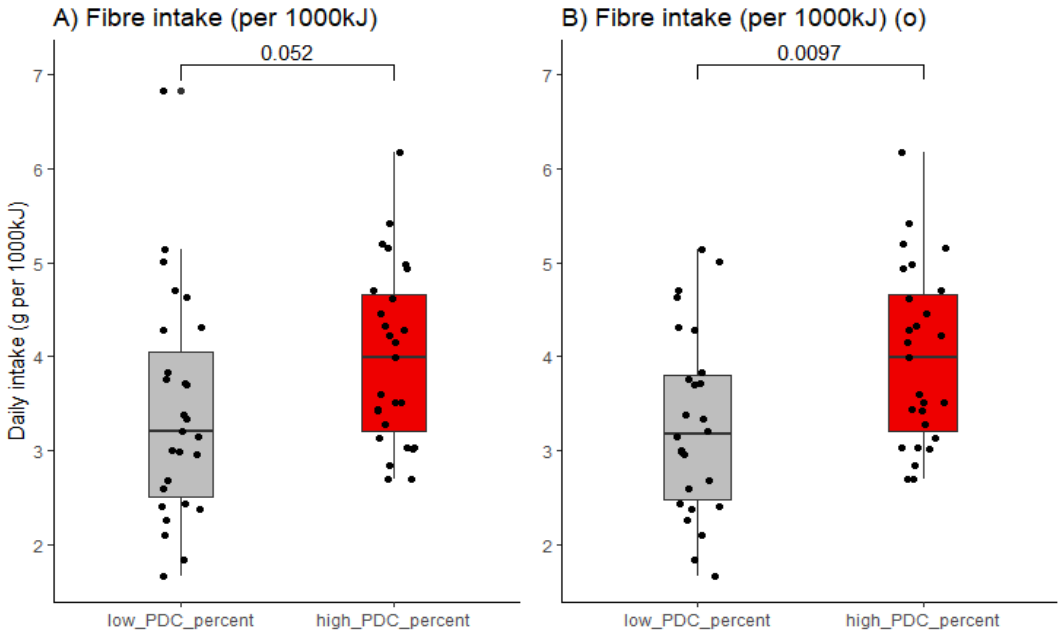


Figure 5-6: Boxplots showing fibre intakes per 1000kJ energy consumed A)with and B)without outliers in low-%PDC and high-%PDC groups.

Removal of outliers in the low PDC% group resulted in a statistically significant difference between the two groups. However, clinically, fibre intake per 1000kJ in that range may be expected, therefore unadjusted *p* value is reported and discussed. Group differences were assessed by independent samples t-test (significance level set as 0.05).

5.4.4 Investigating relationship of PDC with cholesterol, diet and other biofluid-specific BAs

Spearman correlations were performed to investigate the relationship between PDC and other measures namely diet, blood cholesterol and BA in faecal, serum and urine samples. Although there was no significant difference in PDC concentrations and proportions (as measured by PDC%) between the vegan and omnivore groups, however there were significant differences between the two diet groups in factors hypothesized to influence the production of PDC, such as the diet (see Table 5-2). Therefore, Spearman correlations were performed in the whole cohort as well as in the different diet groups to investigate whether the differential intakes of potentially pertinent nutrients and foods would affect their associations with PDC/PDC%.

5.4.4.1 *Relationship of PDC with serum cholesterol*

To determine the relationship between PDC and blood cholesterol, Spearman correlations were performed between all five measures of DCA, PDC, total fDCA, DCA% and PDC% and serum lipoproteins analysed (namely total, LDL, HDL and VLDL cholesterol and triglycerides). No significant correlation was observed between any of the five PDC/DCA measures and serum lipoproteins analysed in the whole dataset as well as in either diet group.

5.4.4.2 *Relationship of PDC with diet*

To determine the relationship between PDC and the diet, dietary variables were sub-grouped as follows;

- “major nutrients subgroup” comprising of total intakes of energy and the macronutrients namely carbohydrate, protein and fat;
- “minor nutrients subgroup” comprising of intakes of micronutrients like minerals and vitamins, fatty acids and amino acids; and
- “fibre foods subgroup” comprising of total dietary fibre intake, relative dietary fibre intake (intake per 1000 kJ) and average intakes of specific fibre-rich food groups based on the Dietary Approaches to Stop Hypertension (DASH) diet. The DASH food groups are namely fruits, vegetables, nuts and legumes (excluding vegetables) and wholegrains.

Only dietary variables with at least one statistically significant correlation with one of the five PDC/DCA measures and with an absolute value equal to or greater than 0.1 (no correlation coefficient obtained was higher than 0.5) were presented in a correlation heatmap and discussed below.

i. Major nutrients group

PDC, DCA and total fDCA were found to be positively associated with energy and macronutrient intake (Figure 5-7). In the whole group, all three measures were positively associated with dietary protein intake ($p < 0.05$) especially total fDCA ($p < 0.01$). Total fDCA was also found to significantly positively correlate with energy and fat intakes. Interestingly, total fDCA was more strongly associated with carbohydrate intake in vegans but with fat intake in omnivores ($p < 0.05$ in both cases). Hierarchical clustering revealed that PDC and PDC% clustered separately further away from all macronutrient intakes than DCA, DCA% and total fDCA in the whole cohort as well as the diet subgroups.

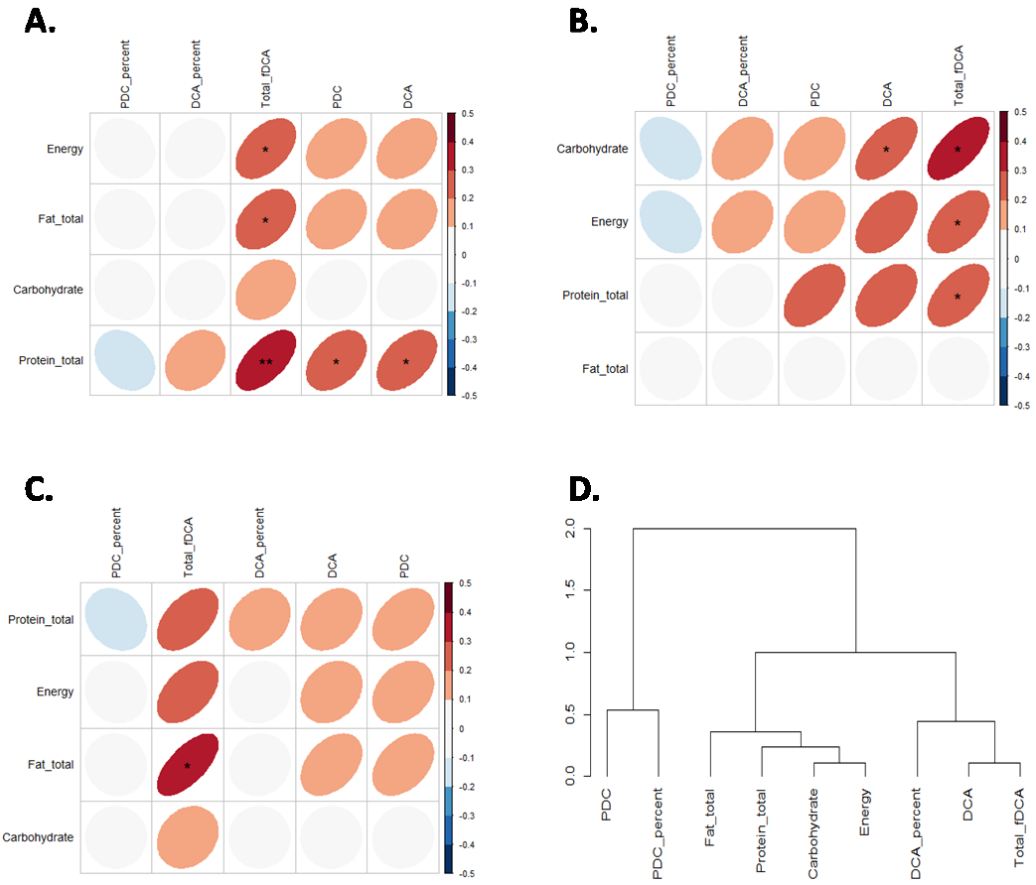


Figure 5-7: Diagrammatic summary of Spearman correlations between PDC, DCA, total faecal DCA, PDC% and DCA% and dietary variables.

Heatmaps showing correlations with intakes of total energy and macronutrients in A)whole cohort, B) vegans only and C) omnivores only and dendrogram shows clustering pattern of these dietary variables with PDC, DCA, total faecal DCA, PDC% and DCA%. Clustering pattern is same across whole group and in both diet groups. Correlations with an absolute value ≥0.1 are displayed. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

ii. Minor nutrients group

To avoid spurious findings, correlations with “minor” nutrients were further corrected for multiple testing by the FDR method (significance set at $p = 0.05$). Results are presented in Figures 5-8.

In the whole cohort, total fDCA was positively associated with many of the dietary variables, particularly amino acids (phenylalanine, serine, leucine, tryptophan, tyrosine, histidine and proline), fatty acids (various saturated fatty acids like stearic acid and monounsaturated fatty acids like palmitoleic acid), dietary minerals and trace elements (sodium, selenium and iodine) and vitamins (retinol, biotin, niacin, pantothenic acid and alpha-tocopherol). PDC, DCA and total fDCA concentrations were significantly positively correlated with total nitrogen intakes while PDC% was found to be positively associated with alpha-tocopherol intake.

In vegans, DCA was positively associated with intake of starch and simpler glucose units (such as the monosaccharides glucose and fructose, and the disaccharide maltose) while PDC was strongly positively associated with added sugar intake and negatively associated with L-ascorbic acid intake (Figure S5-3).

In omnivores, DCA and total fDCA were negatively associated with intakes of vitamins A, C and E as well as of the trace mineral copper while PDC was positively correlated with these variables (Figure S5-3).

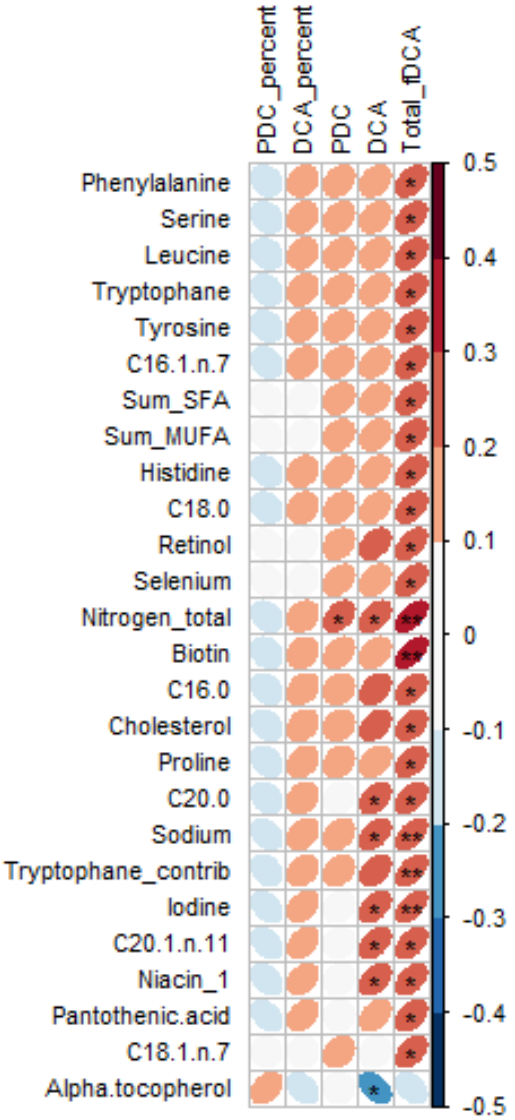


Figure 5-8: Correlation heatmap showing only statistically significant Spearman correlations of an absolute value ≥0.1 between PDC, DCA, total faecal DCA, PDC% and DCA% and intakes of "minor" nutrients in entire cohort.

Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

iii. Fibre and DASH food groups

Spearman correlations performed with the fibre/DASH food subgroup identified a positive association between PDC% (and conversely negative correlation with %DCA) and mean intake of nuts and legumes in the whole group (Figure 5-9A). Also, PDC% was positively associated with vegetable intake and fibre intake per 1000 kJ in the whole group, but these did not attain statistical significance ($p > 0.05$, (Figure S5-2). However, all of these associations were stronger and achieved statistical significance in omnivores (Figure 5-11C). In addition, positive associations were observed between PDC and the average intakes of nuts and legumes and of vegetables in the omnivores. Conversely, DCA, DCA% and total fDCA were negatively associated with these variables. Hierarchical clustering revealed that PDC and PDC% clustered distinctly with all of the fibre-rich food groups in the whole cohort and in omnivores but not in vegans (Figures 5-9B and 5-9C).

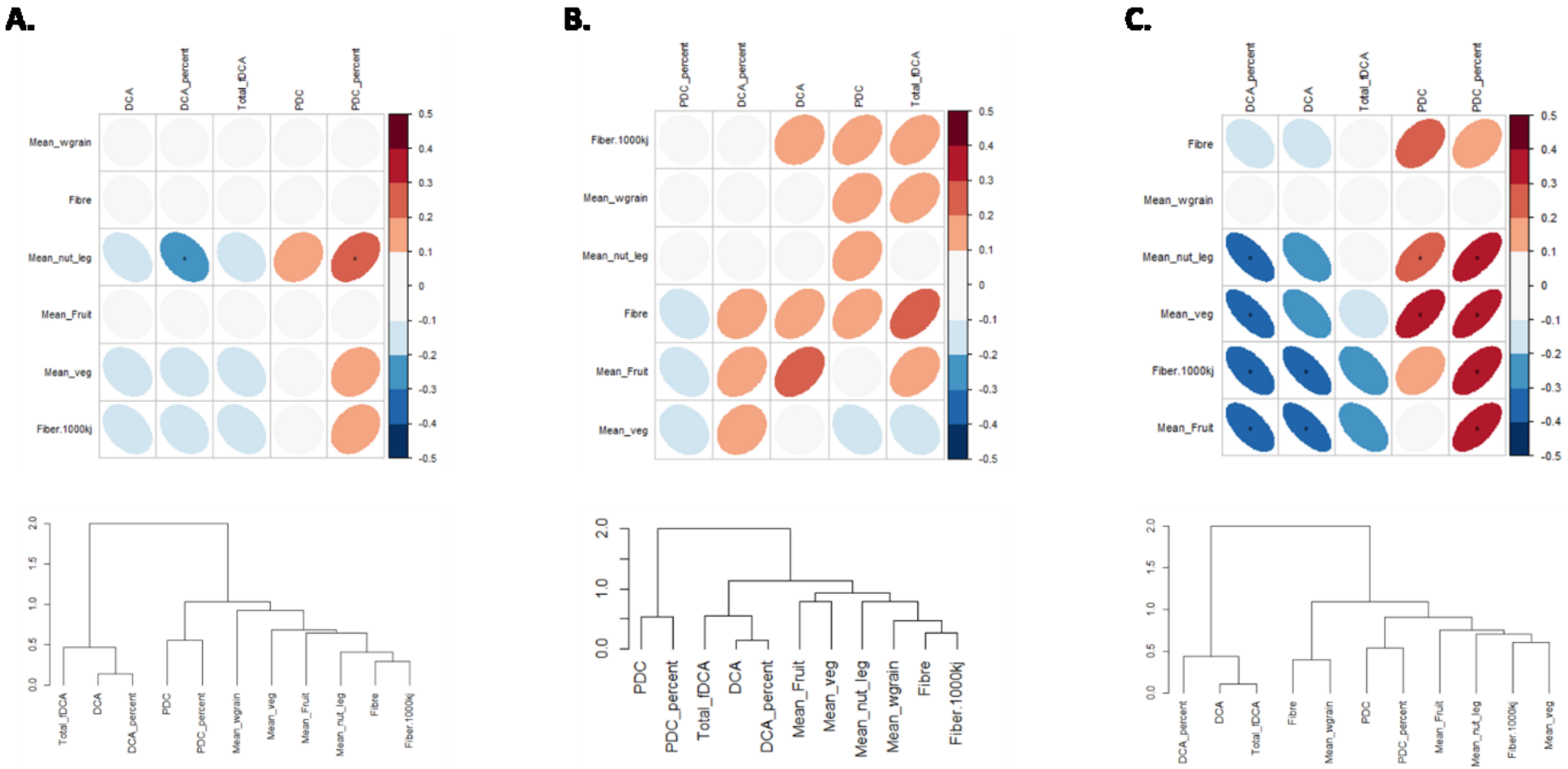


Figure 5-9: Correlation heatmaps and corresponding dendrograms showing Spearman correlations of an absolute value ≥0.1 between PDC, DCA, total faecal DCA, PDC% and DCA% and intakes of fibres in A) the whole cohort, B) vegans only and C) omnivores only.

Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

5.4.4.3 Relationship of PDC with faecal bile acids

Spearman correlations were performed between PDC measures and faecal BAs obtained in Chapter 3. Correlation results were further corrected using the FDR method. Stronger correlations were observed, so cut-off was increased to 0.2. Statistically significant correlations with an absolute value equal to or above 0.2 were displayed as a heatmap (Figure 5-10). While DCA, total fDCA and DCA% were found to have strong positive associations with most of the faecal BAs, PDC was found to correlate negatively with β-muricholic acid, (βMCA) in whole group as well as in either vegans or omnivores, chenodeoxycholic acid, (CDCA) in the whole group and omnivores but not in vegans and 3α-hydroxy-7-ketolithocholic acid (7KLCA) in whole group but not in either sub-group analyses; and positively with allolithocholic acid (allo-LCA) in the whole group as well as both vegans and omnivores, and isolithocholic acid (iso-LCA) in the whole group but not in either diet sub-group. PDC% was negatively associated with all the faecal BAs except with allo-LCA. In cluster analyses, PDC and PDC% clustered distinctly with allo-LCA only in whole group and sub-group analyses (Figure S5-5).

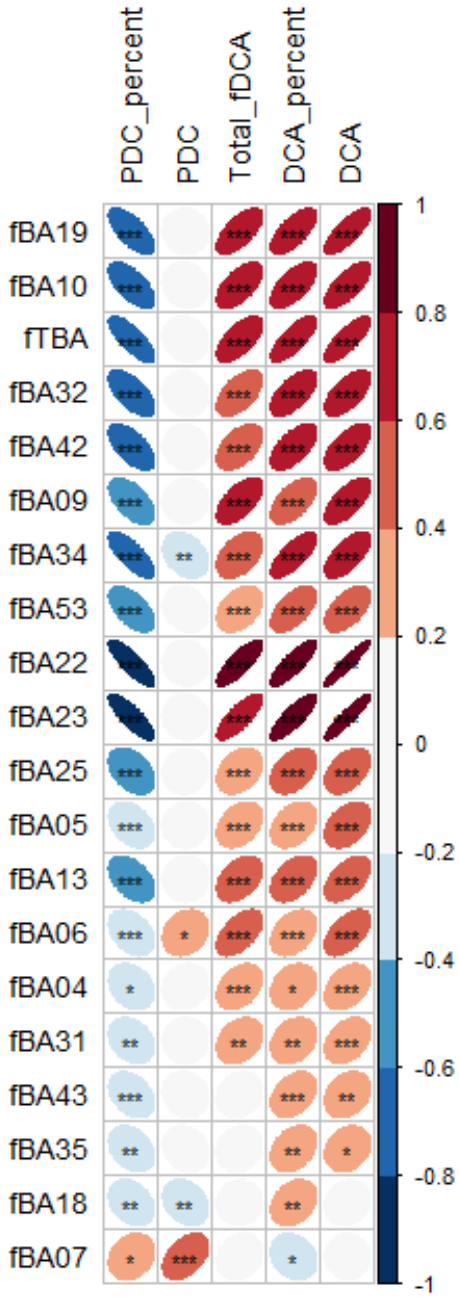


Figure 5-10: Heatmap showing correlations of PDC, DCA and total faecal DCA concentrations and percentages with faecal bile acids in whole cohort (n=106).

Spearman correlations were performed and corrected by Benjamini-Hochberg correction to filter out false positives. Variables with at least one significant correlation with an absolute value ≥0.2 with one of the PDC measures were further displayed in a heatmap. fBA19 = 3α-hydroxy-12-ketolithocholic acid, fBA10 = 9(11), (5β)-cholenic acid-3α-ol-12-one, fTBA = faecal total bile acids, fBA32 = ursodeoxycholic acid, fBA42 = glycodeoxycholic acid, fBA09 = 3α,12α, 23-nordeoxycholic acid, fBA34 =β-muricholic acid, fBA53: taurodeoxycholic acid, fBA22 = 5β-cholanic acid-3β,12α-diol, fBA23 = deoxycholic acid, fBA25 = hyodeoxycholic acid, fBA05 = lithocholic acid, fBA13 = 3,6-diketocholanic acid, fBA06 = isolithocholic acid, fBA04 = 3-ketocholanic acid, fBA31 = 12-dehydrocholic acid, fBA43 = glycoursodeoxycholic acid, fBA35 = cholic acid, fBA18 = 3α-hydroxy-7-ketolithocholic acid, fBA07 = allolithocholic acid. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

5.4.4.4 *Relationship of PDC with serum bile acids*

Similarly, Spearman correlations were calculated between serum BAs and PDC measures. PDC was positively correlated with isolithocholic acid (iso-LCA) and negatively correlated with glycoursodeoxycholic acid (GUDCA) and tauroursodeoxycholic acid (TUDCA), while PDC% was negatively associated with most other serum BAs (Figure 5-11). Specifically, in vegans, PDC and PDC% were negatively associated with total BA. Cluster analyses revealed that PDC and PDC% were clustered very separately from the serum BAs (Figure S5-6).

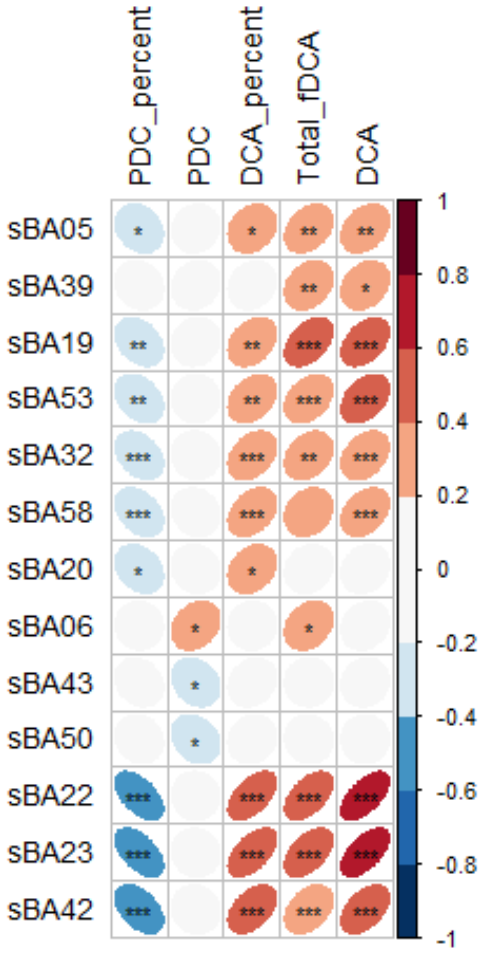


Figure 5-11: Heatmap showing correlations between PDC, DCA and total faecal DCA concentrations and percentages with serum bile acids.

Spearman correlations were performed and corrected by Benjamini-Hochberg correction to filter out false positives. Significant correlations with an absolute value ≥0.2 with one of the PDC measures were further displayed as a correlation heatmap. sBA05 = lithocholic acid, sBA39 = glycolithocholic acid, sBA19 = 3α-hydroxy-12-ketolithocholic acid, sBA53: taurodeoxycholic acid, sBA32 = ursodeoxycholic acid, sBA58 = tauro ω-muricholic acid, sBA20 = murocholic acid, sBA06 = isolithocholic acid, sBA43 = glycoursodeoxycholic acid, sBA50 = tauroursodeoxycholic acid, sBA22 = 5β-cholanic acid-3β,12α-diol, sBA23 = deoxycholic acid, sBA42 = glycodeoxycholic acid. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

5.4.4.5 *Relationship of PDC with urine bile acids*

Results from Spearman correlations with urine BAs showed that PDC was negatively associated with hyocholic acid (HCA) and glycochenodeoxycholic acid (GCDCA) (Figure 5-12). However, cluster analyses demonstrated that HCA, a 6α-epimer of CA, was more similar to DCA than PDC (Figure S5-7). PDC% likewise was negatively associated with HCA but positively with taurocholic acid (TCA). PDC and PDC% clustered more similarly with glyco- and tauro- conjugates of primary BAs, namely CDCA, CA and HCA.

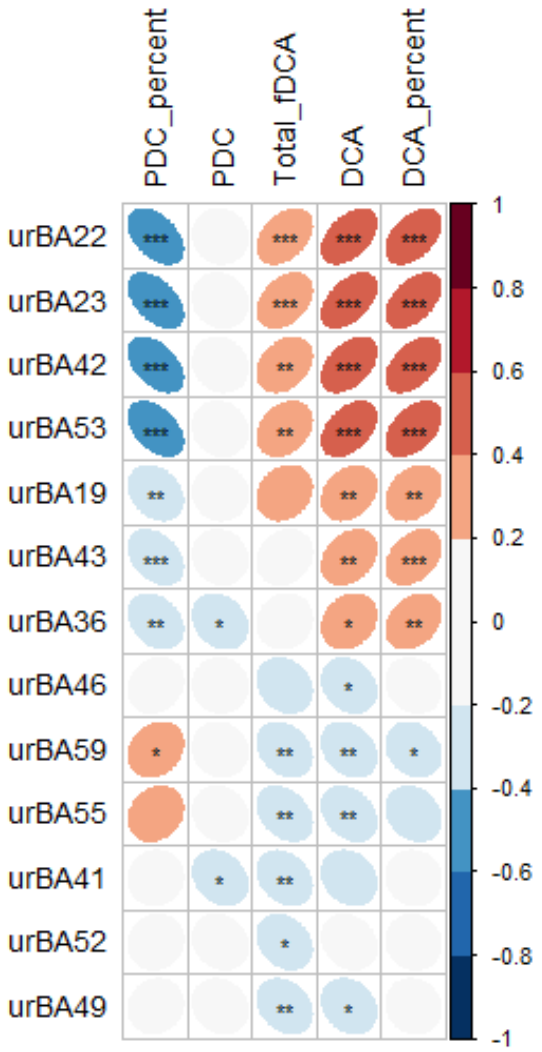


Figure 5-12: Heatmap showing correlations between PDC, DCA and total faecal DCA concentrations and percentages with urine bile acids.

Spearman correlations were performed and corrected by Benjamini-Hochberg correction to filter out false positives. Significant correlations with an absolute value ≥0.2 with one of the PDC measures were further displayed as a correlation heatmap. urBA22 = 5β-cholanic acid-3β,12α-diol, urBA23 = deoxycholic acid, urBA42 = glycodeoxycholic acid, urBA53: taurodeoxycholic acid, urBA19 = 3α-hydroxy-12-ketolithocholic acid, urBA43 = glycoursodeoxycholic acid, urBA36 = hyocholic acid, urBA46 = glycocholic acid, urBA59 = taurocholic acid, urBA55 = taurohyocholic acid, urBA41 = glycochenodeoxycholic acid, urBA52 = taurochenodeoxycholic acid, urBA49 = glycohyocholic acid. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

5.4.4.6 Relationship of PDC with faecal bacterial species

Metagenomic bacterial species (MSP) were restricted to those with a prevalence above 10% (n = 629). Spearman correlations were calculated between each qualifying MSP and each of the 5 PDC/DCA measures. MSPs with at least one statistically significant correlation ($p < 0.05$, n = 222) were arranged in descending order of absolute magnitude and the top 20 were further displayed as heatmaps. All identifiable top 20 MSPs belonged to the order, *Clostridia*, except for *Eggerthella lenta* (previously classified as *Eubacterium lentum*) which belongs to the *Coriobacteriales* order in the *Actinobacteria* phylum (Kageyama, Benno, & Nakase, 1999; Wade et al., 1999).

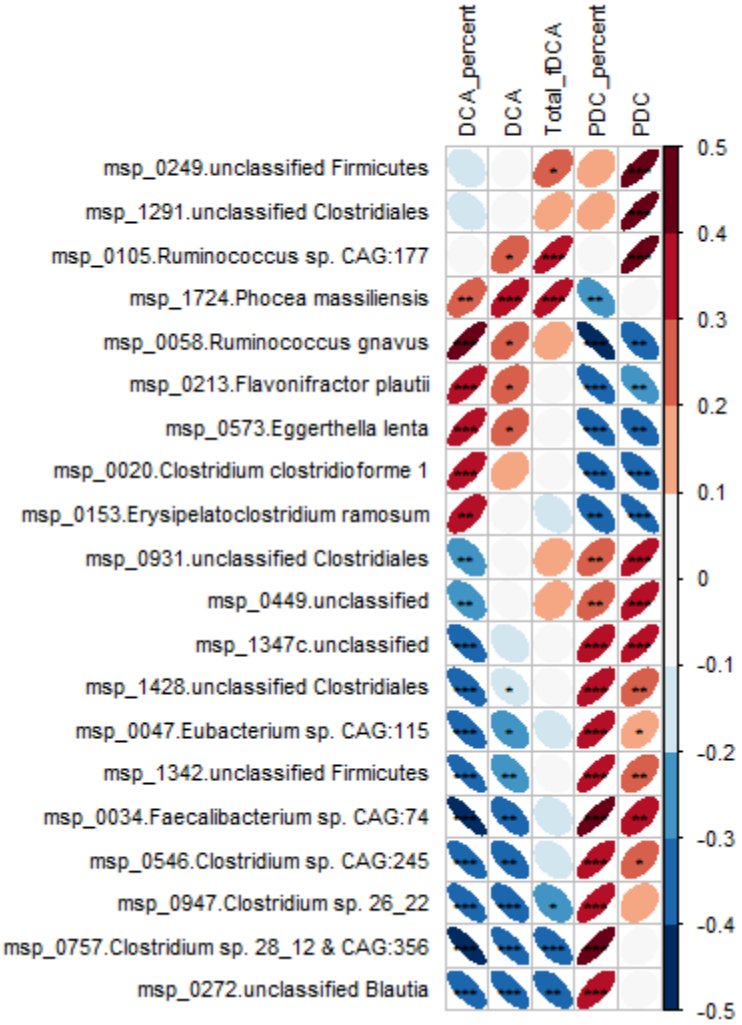


Figure 5-13: Heatmap showing top 20 significant correlations between PDC, DCA and total faecal DCA concentrations and percentages with metagenomic species (MSP).

Spearman correlations were performed and the first 20 MSPs with the strongest significant correlations with one of the five PDC measures were further displayed as a correlation heatmap. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

5.5. Discussion

5.5.1 Summary of findings

- Participants with higher PDC were found to also have higher concentrations of allolithocholic acid in their faecal samples when compared to those with lower PDC concentrations. They also tended to have a higher intake of total calories, protein and, to a slightly lesser extent, fat in their diet.
- Participants with a higher fraction of their total faecal DCA output in the polymerised form (as PDC) were found to have significantly lower concentrations of BAs especially DCA, glyco-DCA, tauro-DCA and "3β-isoDCA" in their serum, urine and faecal samples. They also had lower total faecal DCA and faecal BA concentrations. Lastly, they were found to have a higher intake of nuts and legumes, and of dietary fibre with respect to their total caloric consumption.
- All three of PDC, DCA and total fDCA were found to be positively associated with total protein intake. These observations with macronutrient intakes were reflected in associations with minor nutrients as well. Significant positive correlations were observed between all three and total nitrogen intake which is a strong marker of protein intake. Similarly, all three PDC, DCA and total fDCA positively correlated with biotin, which is sourced from foods that contain appreciable amounts of protein such as liver, other organ meats, meats, salmon, nuts and seeds (although only total fDCA attained statistical significance in this case). Results obtained from post hoc extreme group analyses of the lower and upper PDC quartiles were in alignment with these observations, as upper quartile PDC members were found to consume significantly higher amounts of protein than lower quartile samples.
- A weak but statistically significant positive correlation was found between PDC% and mean intake of nuts and legumes when Spearman correlations were performed in the whole group. PDC and PDC% correlated and clustered with intake of nuts and legumes and alpha tocopherols. Correlations of PDC/PDC% with intake of nuts, legumes, vegetables and fruits were stronger in omnivores.
- There was no significant association between PDC/PDC% and serum lipoproteins.
- PDC/PDC% correlated with microbiota-transformed BAs such as isolithocholic acid, glycoursodeoxycholic acid, tauroursodeoxycholic acid and particularly allo-LCA in at least two biofluids. However, hierarchical clustering did not sometimes agree with these observations.
- PDC and PDC% were found to be strongly correlated mostly with members of the *Clostridia* class. This included negative correlations with *Ruminicoccus gnavus*,

Clostridium clostridioforme, *Flavonifractor plautii* and *Erysipelatoclostridium ramosum* and positive correlations with species belonging to *Clostridium, Ruminococcus, Eubacterium, Faecalibacterium and Blautia* genera.

5.5.2 General discussion

By expressing PDC as a fraction of total fDCA, this prevented the tendency to group participants who may have "low" PDC concentrations as a result of generally having low BA concentrations with those who have "low" PDC concentrations even though they may have high total faecal BA concentrations. Thus, %PDC maybe a more accurate measure of PDC as it controls for total DCA present in the faeces of the individual, and by extension total BA since DCA is the most abundant and major faecal BA. However, although the proportional computation (*i.e.* PDC%) may be a more accurate variable than concentrations (PDC), extra care is taken to interpret results obtained from correlations performed with respect to its proportional relationship with DCA% (an association with PDC% will result in an opposite association with DCA%, and vice versa). Nevertheless, PDC and PDC% were observed to vary similarly in most cases with other variables investigated in this analysis. In instances when variation differs, results are interpreted with all five measures in context.

To overcome the complexities of missing data across matrices in the entire dataset, an analysis was performed using only those individuals with complete data across all matrices. Various strategies have been employed to handle missing data to varying degrees of efficiency and accuracy, although no single method is perfect (N. J. Horton & Kleinman, 2007; Kang, 2013). Substitution/imputation is one of the most common methods for dealing with missing data, however there is great tendency to introduce bias into the data (N. J. Horton & Kleinman, 2007). Due to the first-time exploratory nature of this analysis, introduction of statistical artefacts and potential bias may not be desirable. Data from this study was missing largely due to the absence of samples for analysis (samples were not collected or were exhausted), therefore data was assumed to be randomly missing and analyses were thus restricted to complete cases only. Although commonly used, complete case analysis is sometimes advised against because of the potential loss in power. However, a power calculation was not carried out *a priori* and therefore there were no restrictions. Nevertheless, "sensitivity" analyses were carried out by comparing results from complete case analysis to those obtained from available case analyses (using pairwise-deletion in master "M" dataset). Results from both were found to be comparable (see Appendix 5).

5.5.2.1 Serum lipoproteins

There was no correlation between PDC, DCA and serum lipoproteins in the Vegan cohort. These results are in contrast with a recently published study by Breuninger and colleagues (Breuninger et al., 2019) where they found a positive association between faecal DCA and serum triglycerides. It is important to note that the Breuninger cohort was a large cohort of 1,387 adults with a wider range of serum lipoprotein concentrations from normocholesterolaemic to hypercholesterolaemic levels, compared to the vegan study cohort which was of a more moderate size and had healthy participants with cholesterol levels within the normal healthy limit. While there was a significant difference in total and LDL cholesterol between vegans and omnivores, the narrowness of the ranges in serum lipoproteins may not have been adequate for disentangling the relationship between PDC and serum lipoproteins.

5.5.2.2 Dietary intakes

Although relatively weak (rho/correlation coefficient < 0.5), the Spearman correlation results obtained from the analysis with dietary variables revealed a consistent pattern across the "major" and "minor" nutrients and the fibre-rich food groups investigated. Results consistently demonstrated that PDC/PDC%, more than the other variables, correlated with intake of protein and fibres. PDC positively correlated with protein intake, while PDC% correlated with intake of nuts and legumes, as well as total fibres. The nuts and legumes DASH group comprises of foods such as tree nuts, seeds, beans, chickpeas and lentils which are good sources of both protein and fibre. A positive correlation of PDC% with the vitamin E isoform, alpha tocopherol, which may be obtained from eating nuts and seeds, may have also strengthened the link between PDC and this food group. In addition to nuts and legumes, PDC and/or PDC% were positively correlated with intakes of fruits and vegetables in sub-group analyses in omnivores. More so, PDC and/or PDC% positively correlated with total intake of fibre, both when and when not adjusted for caloric intake (*i.e.* per 1000kJ). These results lend credit to the main hypothesis of this thesis that fibres may play a role in the production of PDC in the colon, however the association with protein was novel. Post hoc extreme group analyses of the lower and upper PDC% quartiles corroborated these observations by revealing that members of the high (upper quartile) PDC% group recorded significantly higher intakes of nuts and legumes as well as fibre per 1000 kJ, thus strongly suggesting a selective link between PDC and the intake of dietary protein and fibre.

Furthermore, stronger associations were observed in omnivores and not vegans. This was a particularly interesting observation because one would expect stronger associations in vegans as they consumed greater amounts of these foods ($p < 0.001$).

On the other hand, DCA and total fDCA were in distinctly contrasting associations with these dietary variables. In fact, DCA and total fDCA were anticorrelated with fibre intake per energy consumed (g/1000kJ); although the associations only approached significance (p = 0.056 and p = 0.086 for DCA and total fDCA respectively) in the whole group, while they were significant when investigated in omnivores but not in vegans.

Total fDCA, but not PDC or DCA, was also significantly associated with total caloric and fat intakes, with a corresponding pattern observed with the fatty acids (mostly made up of monounsaturated fatty acids and saturated fatty acids). These results suggested that while total fat and macronutrient intake may be influential on the total DCA produced or excreted via the faecal route, but not PDC specifically.

While there was no data on intake of specific types of fibres available, wholegrain intake was distinctively not associated with PDC/PDC%. Vegetables, fruits and legumes are known to contain more soluble than insoluble fibres. Soluble and not insoluble fibres are known to increase BA faecal output (McPherson-Kay, 1987). Thus, this may suggest that insoluble fibres may not contribute to PDC production as much as soluble fibres (i.e. fermentable carbohydrates). However, there was insufficient data to arrive at a more definitive conclusion.

In summary, results from investigations of the relationship of PDC measures with dietary variables suggest that while the production of total fDCA may be related to total caloric, protein and fat intakes, PDC formation appears to be specifically favoured by a simultaneous intake of foods rich in dietary fibre, particularly vegetables, fruits, nuts and legumes. Furthermore, these observed associations with specific plant-based foods were intensified in omnivores. Omnivores in this cohort were observed to have consumed greater amounts of fat and protein than vegans. Thus, these results indicate that a high fibre intake combined with concurrent high intakes of dietary protein and fat may promote the production of PDC compared to a high fibre but low fat/protein diet.

5.5.2.3 Faecal BA results

PDC was found to be positively associated with allo-LCA (an epimer of LCA, with the C5-OH in *α* rather than *β* conformation) and iso-LCA (an epimer of LCA, with the C3 in the *β* rather than *α* configuration) and negatively with βMCA (an epimer of CA with the C6-OH group in β-conformation) and 7KLCA (a keto-derivative of LCA). On the other hand, PDC% negatively correlated with iso-LCA but positively with allo-LCA. Not much is known about allo-LCA. It has been detected in less evolved organisms and seldomly in adult human faecal samples (Shiffka, Kane, & Swaan, 2017). Similarly, βMCA is hardly detected in humans but is found in considerable amounts in mice where it is a dominant primary BA (Russell, 2003), and can be

produced from the hydroxylation of the C6 in UDCA (S. Takahashi et al., 2016). 7KLCA is an oxo-bile acid obtained from bacterial transformation of CDCA or UDCA (Fromm, Carlson, Hofmann, Farivar, & Amin, 1980; Odermatt et al., 2011). Iso-LCA has been previously detected in human faeces and bile (Danielsson et al., 1963; Norman & Palmer, 1964; van Faassen et al., 1987) and demonstrated to be produced from the transformation of LCA by some members of the *Clostridium* genus (Ashok K Batta, Salen, & Shefer, 1985; Borriello & Owen, 1982). It may also be synthesized in the liver as demonstrated by *in vitro* experiments in liver isolates (Björkhem, Einarsson, & Hellers, 1973; Yoshiki, Wataru, Akio, Toshikazu, & Kazuya, 1985). Iso-LCA was detected mostly in the saponifiable fractions obtained from human faecal samples (Norman & Palmer, 1964) which, upon characterisation, comprised of fatty-acid esters of iso-LCA (Kelsey, Molina, Huang, & Hwang, 1980).

Iso-LCA, allo-LCA and DCA are all produced from gut bacterial transformation of primary BAs, so it may be logical to observe PDC (which is formed from DCA) to be positively correlated with these BAs. This is particularly observed in the clustering of iso-LCA with LCA, GLCA, DCA and others. However, what is interesting is that DCA was, on the contrary, negatively associated with allo-LCA. Also, the extreme group (Q1 v Q3) comparisons showed that high PDC samples contained significantly higher concentrations of allo-LCA. Although allo-LCA was detected in comparatively low concentrations in this study (allo-LCA <200 nM/mg dry stool) and the strength of association observed was weak (Spearman rho ρ <0.4), the significance was quite high ($p < 0.01$) suggesting a potent association with PDC and indicating the possibility of a common microbial pathway/producer. However, while allo-LCA positively correlated with PDC% with respect to its proportion of total fDCA, indicating that regardless of DCA or total fDCA concentrations, a higher PDC proportion corresponded to a higher concentration of allo-LCA in the faeces, there was no significant difference in allo-LCA concentrations between low- and high- PDC% groups.

5.5.2.4 Serum BA results

In serum samples, a similar trend was observed with PDC positively associating with iso-LCA, along with DCA and total fDCA, whereas, there was a very weak, negative and statistically insignificant association between PDC% and the same analyte (iso-LCA). Iso-LCA has seldomly been reported in the serum before. It is considered to be a less toxic form of LCA (Devlin & Fischbach, 2015) and may be reabsorbed from the colon after bacterial transformation. It was reported to be detected in serum samples of patients with colorectal adenomas (Bayerdorffer et al., 1995) and was only detected in the nanomolar range in the present cohort being studied. The similarity in direction of association of PDC, DCA and total

fDCA with serum iso-LCA suggest, just as observed with faecal BAs, that the similarity in microbial origins of 7α-dehydroxylation of unconjugated primary BAs may contribute to the association observed.

On the other hand, PDC (as well as PDC%) was negatively associated with TUDCA and GUDCA, taurine and glycine conjugated forms of UDCA respectively. UDCA is thought to be synthesised by epimerisation of CDCA in the colon in relatively small amounts by *Clostridium absonum* (Ridlon et al., 2016) and may be reabsorbed into circulation and amido-conjugated in the liver (A F Hofmann, 1994).

However, these associations may not be very telling, as hierarchical clustering revealed contrasting results showing that PDC and PDC% clustered separately from the serum BAs, unlike DCA, DCA% and total fDCA which clustered with all other BAs, predominantly the unconjugated BAs as well as GLCA. Nevertheless, these are all BAs transformed by the colonic bacteria before absorption before re-entry into the portal venous circulation, and as such the observed associations may be explained by their similarity in gut microbial origins.

5.5.2.5 *Urine BA results*

Urinary excretion is a minor pathway for BA elimination in healthy humans, with preferential elimination of sulphated and glucuronidated BAs, as well as TCA and cholic acid metabolites (Barnes, Gollan, & Billing, 1977; Wilson, Burckhardt, Murer, Rumrich, & Ullrich, 1981; Zollner, Marschall, Wagner, & Trauner, 2006). LCA metabolites are mostly detected in urine in sulphated and glucuronidated forms. Unlike faecal and serum matrices, neither iso-LCA nor allo-LCA was reported in urine samples in this study. It is possible that LCA and its metabolites were not detected because sulphates and glucuronides were not included in the LC-MS method.

However, complete linkage hierarchical clustering patterns revealed that although PDC and PDC% cluster together distinctly from all other analytes, the primary BAs (TCA, GCA, THCA, GCDCA and TCDCA) were more similar to PDC/PDC% than they were to DCA, DCA% and total fDCA. This is in contrast to the negative association observed with faecal primary BAs. The presence of the conjugated BAs in urine may indicate a reduced efficiency of BA reabsorption in the ileum, along the colon and in the glomerular tubules, and as such, a positive association of these BAs with PDC may reflect similarities in mechanisms that results in their formation (i.e. reduced reabsorption).

5.5.2.6 Metagenomic species

PDC and PDC% were found to be positively associated with various members of phylum *Firmicutes*, particularly those belonging to the class *Clostridia* such as *Ruminococcus sp, Clostridium sp, Faecalibacterium sp*., *Eubacterium sp.* and *Blautia sp*. Several unclassified MSPs belonging to *Firmicutes* were also found to be positively associated with PDC or/and PDC%. Conversely, PDC and PDC% anticorrelated with the *Phocea massiliensis*, *Ruminococcus gnavus, Flavonifractor plautii, Eggerthella lenta, Clostridium clostridioforme* and *Erysipelatoclostridium ramosum*. Most of these microbes, particularly members of the genera *Clostridium* and *Rumminococcus* as well as the prevalent gut microbe *E. lenta* have been shown to possess bile salt hydrolases which deconjugate BAs and hydroxysteroid dehydrogenases which epimerize and oxidize unconjugated BAs at the C3, C7 and C12 carbon atoms (Devlin & Fischbach, 2015; Hirano & Masuda, 1981; Lee et al., 2013; MacDonald, Jellett, Mahony, & Holdeman, 1979; Ridlon et al., 2006).

Interestingly, all of the positively associated microbes are themselves or are closely related to known butyrate-producers found in the human gut (C. Liu et al., 2015; Pryde, Duncan, Hold, Stewart, & Flint, 2002; K. Takahashi et al., 2016; Vital, Howe, & Tiedje, 2014). Butyrate is one of the main products of macronutrient fermentation in the gut (Byrne, Chambers, Morrison, & Frost, 2015; Cummings, Pomare, Branch, Naylor, & Macfarlane, 1987; Morales et al., 2016; Oliphant & Allen-Vercoe, 2019; C. K. Yao, Muir, & Gibson, 2016) and has been shown to have anti-inflammatory effects. Thus, butyrate-producing gut microbes are thought to be beneficial to host health. For instance, certain *Eubacterium* species together with *Roseburia spp* form a major group of butyrate-producers (Duncan et al., 2006; P. Louis & Flint, 2009, 2017) that have been in previous studies found to be inversely associated with atherosclerosis (Karlsson et al., 2012; Kasahara et al., 2018) and inflammatory diseases such as Crohn's disease (CD) (K. Takahashi et al., 2016), irritable bowel syndrome (IBS) (Chassard et al., 2012) and ulcerative colitis (UC) (Kumari, Ahuja, & Paul, 2013). *Blautia spp* has also been shown to be inversely associated with CD (K. Takahashi et al., 2016) as well as with visceral fat accumulation (Ozato et al., 2019), diabetes (Larsen et al., 2010) and liver cirrhosis (Kakiyama et al., 2013).

Conversely, the members of the gut microbial community found to be negatively with PDC/PDC% have been previously identified to be proinflammatory. *R. gnavus* is a prevalent gut microbe (Qin et al., 2010) that has been linked to a myriad of inflammatory diseases including CD, UC, eczema, diverticulitis and spondylarthritis (Breban et al., 2017; Hall et al., 2017; S. G. K. Hansen, Skov, & Justesen, 2013; Henke et al., 2019; Joossens et al., 2011; Levine & Burakoff, 2011; Machiels et al., 2017; Nishino et al., 2018; Png et al., 2010; Willing et al., 2010; Zheng et al., 2016). *P. massiliensis* is a relatively new bacterial species identified

from human faeces (Ndongo, Lagier, Fournier, Raoult, & Khelaifia, 2016). It was found to have 92% sequence identity with an *Anaerotruncus colihominis* strain which was isolated from human faeces and has been previously implicated in bacteraemia (Lau et al., 2006; Lawson et al., 2004). *E. lenta* which is part of normal human gut microbiota and can be altered by consumption of cruciferous vegetables (F. Li, Hullar, Schwarz, & Lampe, 2009) has also been reported to be implicated in fatal bacteraemia (Gardiner et al., 2015). *F. plautii* was found to be enriched in patients with irritable bowel syndrome and infections (Berger et al., 2018; Hollister et al., 2019; Yixuan Liu et al., 2016) while *C. clostridioforme*, although found to be associated with long term vegetarian diet (Matijašić et al., 2014), is an antimicrobial resistant microbe long thought to be involved in bacteraemia (Finegold et al., 2005). Lastly, *E. ramosum* has been associated with metabolic syndrome and a myriad of infections (Woting, Pfeiffer, Loh, Klaus, & Blaut, 2014; Zakham et al., 2019).

5.5.3 Strengths and limitations

A major limitation was the fact that results were obtained from an observational study, as such causality may not be established. As well, correlations do not imply causality, therefore more research is warranted to investigate the relationship of PDC with these gut microbes. Furthermore, results may be interpreted with caution as relative abundances of gut bacteria and not actual amounts were explored. Nevertheless, useful information was gleaned about the potential role of PDC in host physiology. The sample size was reasonably large and the results may be generalisable albeit interpreted with caution.

5.6. Conclusion

PDC/PDC% was found to be associated with a specific dietary pattern, high in the intake of fat, protein and fibre, particularly foods belonging to the nut and legumes DASH food group. While a high fat and protein-rich diet may induce the production of more DCA in total (Bortolotti et al., 2009), it was observed that a high legume and nut intake may favour polymerisation of monomer DCA into PDC. A high protein, high fibre diet has been previously shown to improve metabolic health in overweight and obese women in a 10-week intervention (Morenga, Williams, Brown, & Mann, 2010). It will be interesting to further explore the role of PDC in mechanisms underlying the beneficial effects of this particular diet pattern. It will be particularly interesting to further investigate the effect of consumption of fibre-rich foods, particularly legumes and/or seeds, on PDC production in a prospective study design. The next chapters will take these observations further by investigating the effect of an increase in legume intake on PDC production in a crossover randomised controlled trial.

6. EXPLORING THE EFFECT OF AN INCREASED INTAKE OF LEGUMES ON POLYDEOXYCHOLIC ACID IN HEALTHY ADULT HUMANS

6.1 Introduction

Following on from findings reported in Chapter 5, it was observed that the consumption of the Dietary Approaches to Stop Hypertension (DASH) “nuts and legumes” food group positively correlated with PDC concentrations and proportions. As the vegan study was an observational study, it was not possible to establish a causal relationship. To further investigate this interesting observation, the effect of a controlled legume intake on PDC production was investigated in a double-blind crossover randomised controlled trial (RCT).

The DASH “nuts and legumes” food group comprised of nuts and seeds such as almonds, and pecans, as well as legumes and leguminous vegetables such as peas, beans and lentils (L.B. et al., 2009; US Department of Health and Human Services, National Institutes of Health, National Heart, Lung, 2003). Legumes and nuts are commonly featured in cardioprotective diets due to their exceptional nutritional profile being good sources of protein, dietary fibre, vitamins, minerals and nutraceuticals; and their widely documented beneficial effects on cardiometabolic health, particularly on blood lipids (Becerra-Tomás, Papandreou, & Salas-Salvadó, 2019; Becerra-Tomás, Paz-Graniel, et al., 2019; Ha et al., 2014; Viguiliouk et al., 2019).

In particular, the pea (*Pisum sativum* L.) is one of the most important grain legumes consumed globally (Figueira, Curtain, Beck, & Grafenauer, 2019; Mitchell, Lawrence, Hartman, & Curran, 2009; Mudryj et al., 2012; Ratnayake, Hoover, & Warkentin, 2002; Thompson, 2019). Peas are a great source of good quality protein (~25%), dietary fibres (~10% insoluble and ~5-6% soluble) and starches (Dahl, Foster, & Tyler, 2012; Millar, Gallagher, Burke, McCarthy, & Barry-Ryan, 2019; Tosh & Yada, 2010). In addition to documented hypocholesterolemic effects of dietary soluble fibres, pea proteins have been shown to have an additive cholesterol-lowering effect in rats and humans (Busnelli, Manzini, Sirtori, Chiesa, & Parolini, 2018; Parolini et al., 2013; Sirtori et al., 2012). Similar to the fibres, the underlying mechanisms are currently not well-understood. However, pea protein intake was demonstrated to stimulate faecal excretion of bile acids with a concurrent depletion of hepatic cholesterol and upregulation of the CYP7A1, the rate-limiting enzyme in the synthesis pathway of bile acids (Spielmann, Stangl, & Eder, 2008) indicating that pea proteins may also interrupt normal BA metabolism.

This further supported the hypothesis that the intake of peas may significantly modulate the production of PDC.

There were two strains of peas employed in this RCT, namely the wild-type (*RR*) with a smooth seed surface and the wrinkled (*rr*) pea with a wrinkled seed surface (Ratnayake et al., 2002). The *rr* pea is a result of a naturally occurring mutation, primarily at the *rugosus (r)* locus on chromosome 7 but may also be found at the *rb* and *rug* loci (M. Bhattacharyya, Martin, & Smith, 1993). The two lines differ not only in seed surface smoothness but also in chemical composition (Kooistra, 1962). The *rr* peas contain less starch than *RR* peas and have considerably higher amylose to amylopectin ratios, which are the two main components of pea starch (M. K. Bhattacharyya, Smith, Ellis, Hedley, & Martin, 1990; Kooistra, 1962). Amylose has a linear molecular structure composed mainly of α-(1→4) linked D-glucopyranosyl residues; this prevents it from being easily degraded by α-amylase, an enzyme that breaks down starch in the body (Maier et al., 2017). Whereas, amylopectin has a branched molecular structure composed of linear chains of (1→4)-α-D-glucose residues connected through (1→6)-α-glycosidic bonds which are more susceptible to α-amylase degradation (Ratnayake et al., 2002). Larger amounts of amylose thus leads to greater resistance to digestion, and as such confers a high resistant starch content on the *rr* pea. Resistant starch (RS) as coined by Englyst and colleagues in 1982 (Englyst, Wiggins, & Cummings, 1982) is classified as a component of dietary fibre and may be beneficial to host health, particularly in lowering blood cholesterol and its associated risk for developing cardiovascular diseases (Maier et al., 2017; Ratnayake et al., 2002). In addition to higher amounts of RS, the *rr* pea type also has a higher sucrose (Kooistra, 1962) and total lipid content (Kosson, Czuchajowska, & Pomeranz, 1994). While the pea lines differ in these components their protein content remains comparable (Klamczynska, Czuchajowska, & Baik, 2001; Kosson et al., 1994). All of these nutritive components can be affected by processing and cooking methods (N. Wang, Hatcher, & Gawalko, 2008).

The RCT was originally designed to investigate the impact of the consumption of pea products (pea hummus and mushy peas) on β-cell function and pulsatile insulin secretion via a gut microbiome-dependent mechanism. A subset of the data (n = 13) was explored for the purpose of this PhD research, that is, to explore the influence of pea intake on faecal PDC production in healthy humans.

6.2 Aim and Objectives

6.2.1 Aim

To investigate the impact of a controlled intake of peas on the production of polydeoxycholic acid (PDC), the gut microbiome, cholesterol and bile acid (BA) metabolism in a crossover randomised controlled trial (RCT).

6.2.2 Objectives

- To analyse the differences in PDC production before and after wrinkled "*rr*" pea-based dietary intervention.
- To analyse the differences in PDC production before and after wild-type "*RR*" pea-based dietary intervention.
- To compare the changes in PDC production between the two interventions
- To explore the relationship between PDC, serum lipoproteins and faecal microbial composition in these participants at baseline and over the course of the interventions.

6.3 Methods

6.3.1 Sample collection

Samples were obtained from a crossover double-blind RCT carried out by Dr Aikaterini Petropoulou and has already been described (Petropoulou, 2019). Briefly, volunteers for this RCT were recruited from the Healthy Volunteer Database maintained by the NIHR Imperial Clinical Research Facility at Hammersmith Hospital, London. Healthy adults between the ages 55 – 70 years with a BMI between 20-35 kg/m^2 and were weight-stable (defined as change in weight ≤ 3 kg in the 2 months before screening) were eligible for the study. Exclusion criteria were pregnancy, diabetes, cancer as well as cardiovascular, gastrointestinal, liver and kidney diseases. Also excluded were current smokers and volunteers who consumed excess alcohol, had ongoing/past substance abuse, or who used antibiotics, anti-inflammatory drugs and other medications that were likely to interfere with energy metabolism, hormonal balance and appetite regulation. The study protocol was approved by the South East Coast, Surrey, REC (REC reference number 15/LO/0184). All participants gave their written informed consent.

Volunteers were screened, enrolled if found eligible and allocated to two 28-day dietary intervention phases in random order. The intervention phases were separated by a 28-day washout period. During either intervention phase, study participants were provided with food products made from wrinkled *rr* or wild-type *RR* peas. As this was a double-blind study, it was ensured that the food products were similar except for the pea line from which they were made from, and both participants and researchers were blinded to the intervention allocation. Randomisation and allocation were carried out by an independent online service provider (Sealed Envelope Ltd, London, UK; sealedenvelope.com). Participants were advised to refrain from altering their lifestyles during and in-between the intervention phases to prevent confounding and bias. Fasting blood samples for lipoprotein measurements and stool samples for bile acid and microbial analyses were obtained before and after both interventions. The study design and procedures are summarised in Figure 6-1.

6.3.2 Sample analysis

Fasting blood samples were obtained, serum extracted and immediately transferred to the Pathology Laboratory, Hammersmith Hospital for lipoprotein analysis. 16S rRNA gene sequencing was carried out on Illumina MiSeq (Illumina Inc., Saffron Walden, UK) using Illumina's 16S Metagenomic Sequencing Library Preparation Protocol with minor alterations. Stool samples were analysed for BA using the method developed and validated in Chapter 4.

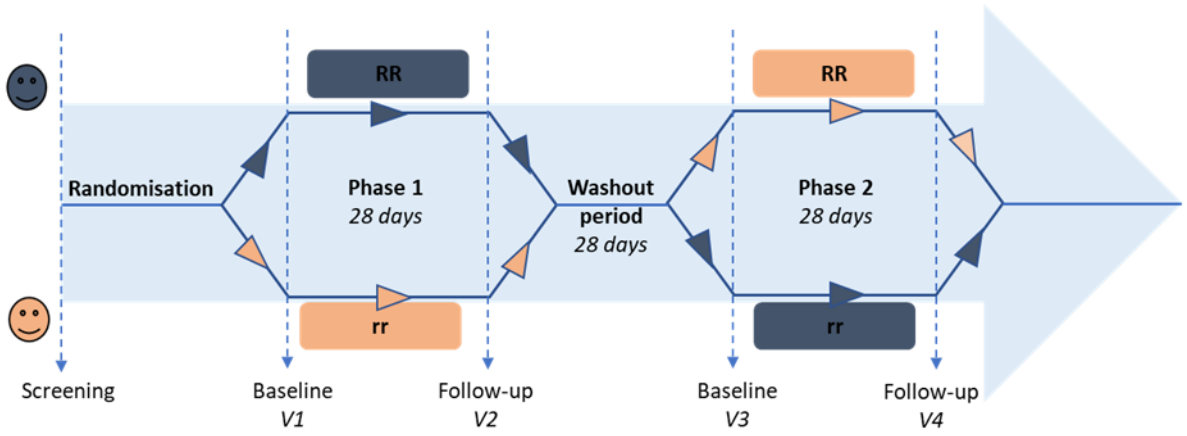


Figure 6-1: Study design of crossover randomised controlled trial.

Eligible and consented volunteers were randomised to start on one of the two 28-day intervention arms and then move on to the other intervention arm after an adequate washout period of 28 days. The different colours show the different possible randomisations and the arrows show the participant's progress through the trial. Samples were obtained at baseline and follow-up visits which were carried out before and after either intervention respectively. The interventions were pea-derived food products (pea hummus and mushy peas). V1-4 = Visits 1-4.

6.3.3 Data and Statistical analyses

Only participants with stool samples from all four study visits were included in the BA analyses (*n* = 13). MS data generated during BA analysis was pre-processed as previously described in Chapter 4. Other datasets, namely participant anonymised clinical (meta)data, lipoprotein concentrations and amplicon sequence variants (ASVs) tables along with assigned taxonomies were obtained, pre-processed where necessary and matched by sample ID.

Final number of samples with complete data (*i.e.* all of the BA, lipoprotein and sequencing datasets) was *n* = 11. Due to the small number of samples available, the best-case strategy was adopted in dealing with missing data in this case. Hence, all *n* = 13 samples were included in the analyses.

Data obtained at Visit 1 (V1) were taken as baseline measurements. Delta changes between baseline and follow-up concentrations of lipoproteins, PDC, DCA and total faecal DCA(total fDCA) were computed as the fractional increase between the two visits, *i.e.*:

$$\Delta\text{PDC}/\Delta\text{DCA}/\Delta\text{total fDCA} = \frac{V_F - V_{BL}}{V_{BL}}$$

Where V_{BL} and V_F were the variable concentrations observed at baseline/initial and follow-up visits respectively.

Delta changes in PDC%, DCA% and microbial relative abundances were computed as the difference($V_F - V_{BL}$) in proportions observed in the respective samples.

Based on normality as assessed by Shapiro- Wilk test, normally distributed data are expressed as mean (±SD) while non-normally distributed data are expressed as median (IQR) as appropriate. Similarly, differences between groups were assessed by parametric paired Student's t-tests or non-parametric Wilcoxon signed rank tests as appropriate. Correlations between data were evaluated by Spearman's rank order correlation. The Benjamini-Hochberg method (FDR) was employed for multiple testing correction where required. All statistical analyses were carried out in R software environment v3.5.2 using the *base* and *tidyverse* v1.2.1 packages. Plots and graphs were produced using the following R packages; *ggplot2* v3.1.1, *corrplot* v0.84 and *RColorBrewer* v1.1-2.

6.4 Results

6.4.1 Participant Characteristics

Participant characteristics are summarised in Table 6-1. Participants ages and BMI ranged from 44 – 67 years and 19.9 – 32.3 kg/m^2 respectively. Adequacy of the washout period was confirmed by comparing V1 and V3 measurements ($p > 0.05$ in all cases, paired *t* tests or Wilcoxon signed rank tests as appropriate, Appendix 6) to ensure all parameters returned to baseline levels after the first intervention phase.

There was a significant decrease in serum total, LDL and non-HDL cholesterol as well as total:HDL cholesterol ratio following the *RR* intervention. There was no change in any other parameter on this intervention, neither were there any significant changes observed on the *rr* intervention.

Table 6-1: Patient characteristics from before and after both RR and rr interventions

	***RR* baseline**	***RR* follow-up**	***p RR***	***rr* baseline**	***rr* follow-up**	***p rr***
No of participants	13			13		
Age*	58.7 (±6.7)			58.7 (±6.7)		
Female/Male	9/4			9/4		
Weight	71.3 (10.6)	71.3 (11.1)	0.076	72.4 (10.8)	72.7 (12.2)	0.401
BMI (kg/m^2)*	24.9 (±3.4)	25.2 (±3.3)	0.060	25.0 (±3.3)	25.0 (±3.3)	0.906
Total cholesterol (mmol/L)*	5.2 (±0.6)	4.9 (±0.7)	0.018	5.3 (±0.6)	5.0 (±0.7)	0.245
LDL cholesterol (mmol/L)*	3.2 (±0.5)	2.8 (±0.6)	0.003	3.2 (±0.7)	3.0 (±0.6)	0.388
HDL cholesterol (mmol/L)	1.5 (0.4)	1.5 (0.5)	0.519	1.5 (0.3)	1.6 (0.3)	0.552
Non-HDL cholesterol (mmol/L)*	3.7 (±0.6)	3.3 (±0.7)	0.003	3.6 (±0.8)	3.5 (±0.7)	0.331
Triglycerides (mmol/L)	0.8 (0.4)	0.9 (0.6)	0.117	1.0 (0.6)	0.9 (0.5)	0.542
Total Cholesterol:HDL	3.2 (0.9)	3.0 (0.9)	0.001	3.2 (0.9)	3.2 (0.7)	0.916
Shannon diversity	4.4 (0.4)	4.2 (0.4)	0.077	4.3 (0.7)	4.1 (0.4)	0.414

Data is displayed as median (interquartile range) and significance testing was carried out by Wilcoxon signed rank test unless otherwise stated. *Data is displayed as mean(± standard deviation) and significance testing was carried out by paired samples t-test.

6.4.2 DCA and PDC on both *rr* and *RR* interventions

There were no significant differences observed in PDC concentrations and proportions following either of the interventions (Figure 6-2). While there seemed to be a tendency for PDC concentrations to decrease following the wild type *RR* intervention phase, this did not attain statistical significance. The reduction in PDC may be because there was an average reduction in total faecal DCA across the participants ($p = 0.05$, Appendix 6) in the *RR* intervention arm. Indeed, there was a net increase in PDC% following both diets (Figure 6-2D), however neither achieved statistical significance. Compliance was assessed by measuring *N*-methylnicotinic acid (trigonelline) in the urine (Posma et al., 2017), but no difference was found in trends between compliant and non-compliant participants in either phase (Figure 6-3). There was also no significant difference in ΔPDC/ΔPDC% between the *rr* phase vs the *RR* phase (Appendix 6).

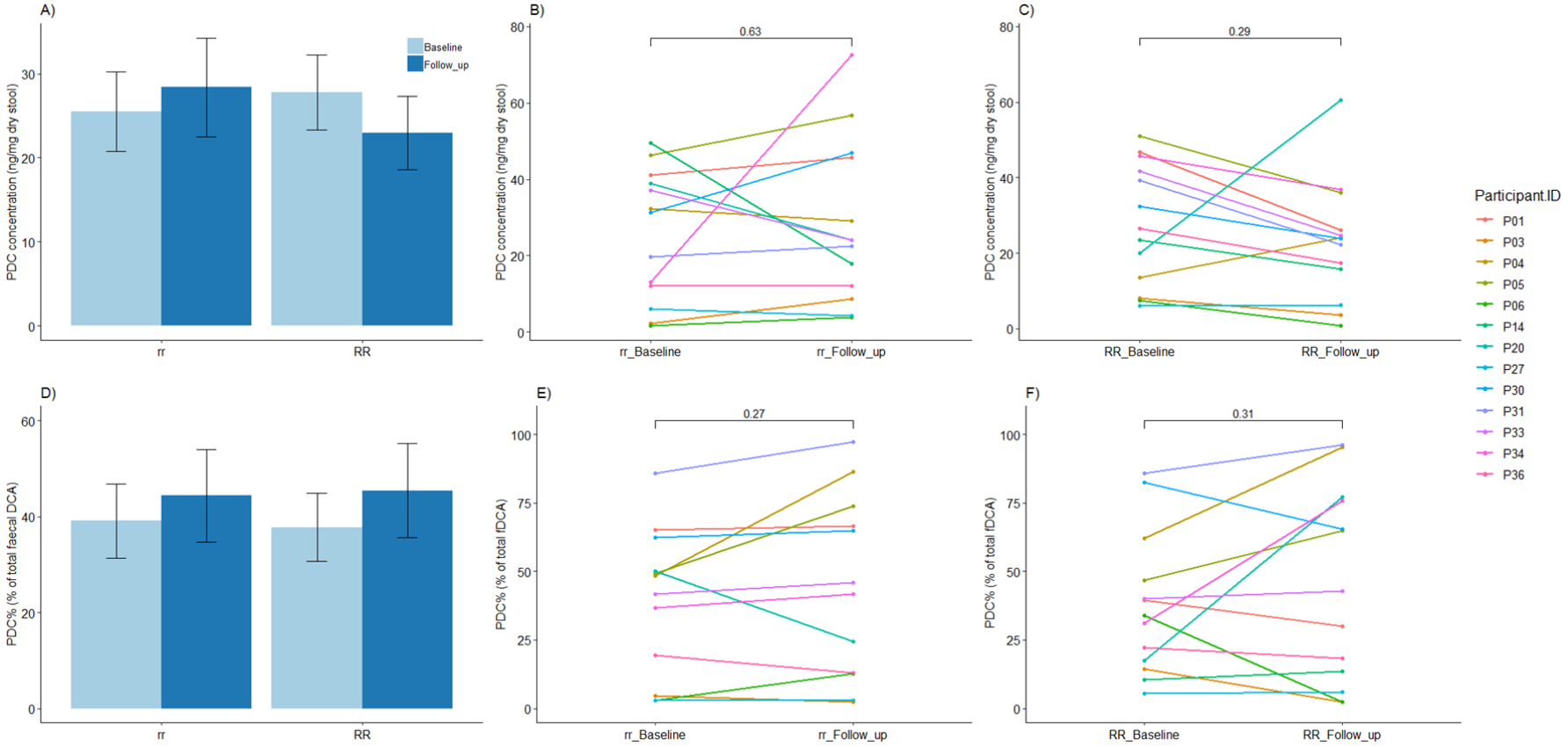


Figure 6-2: PDC and PDC% at baseline and follow-up visits on both rr and RR interventions.

Plots A), B), C) and D), E), F) show bar charts and line graphs of participants in the rr and RR phases respectively. Error bars are standard error of mean. Differences were tested by Wilcoxon signed rank test (for non-normally distributed data) and paired samples t-test (for normally-distributed data) as appropriate.

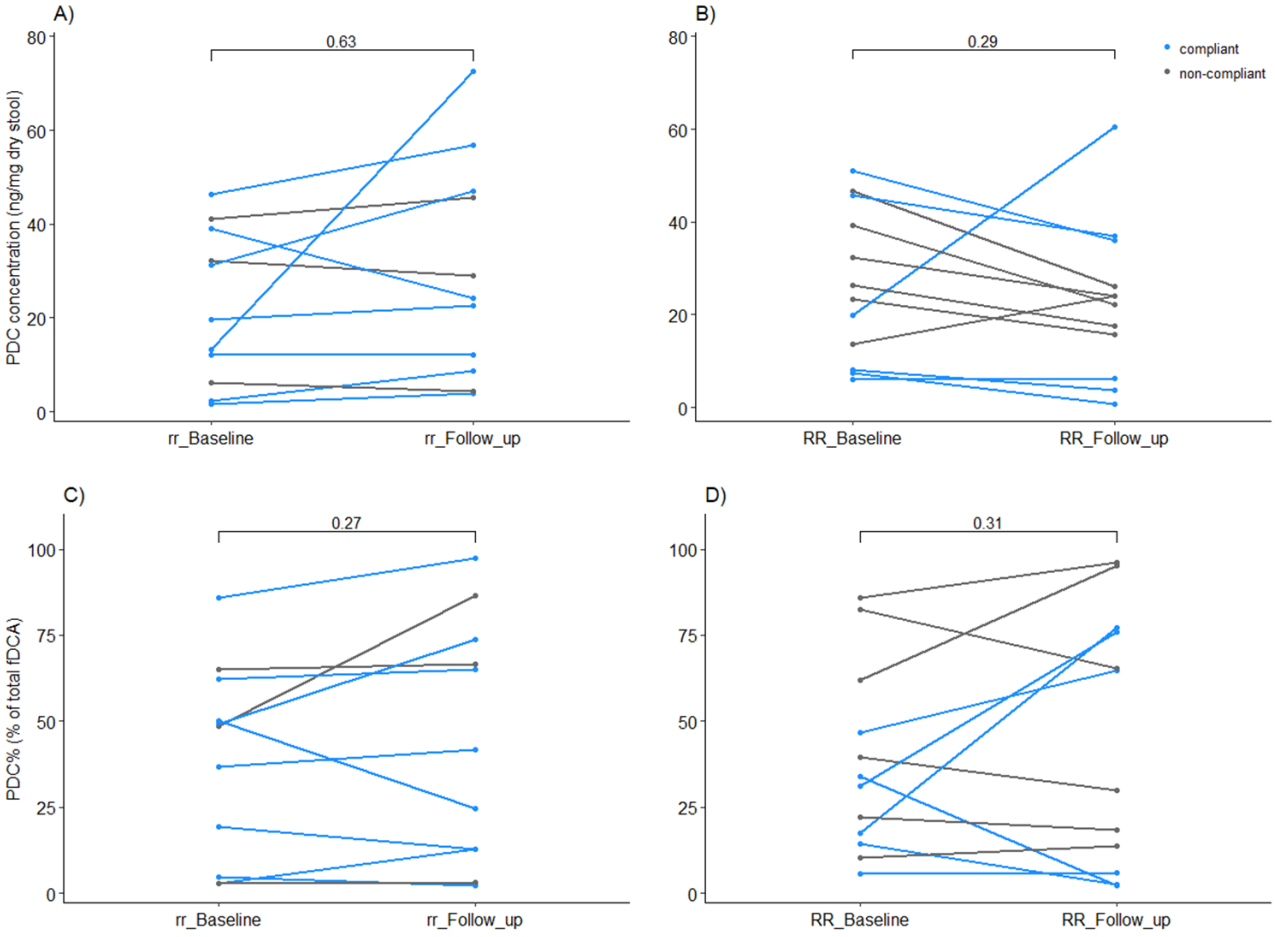


Figure 6-3: Line graphs of participants showing compliant and non-compliant participants on both interventions.

Compliance was monitored by the presence of trigonelline in the urine. No difference was observed in trend between complaint and non-compliant participants. Differences were tested by Wilcoxon signed rank test (for non-normally distributed data) and paired samples t-test (for normally-distributed data) as appropriate.

6.4.3 Correlations between blood lipoproteins and PDC/PDC%

Spearman correlations were carried out to investigate the relationship between PDC/PDC% and blood lipoproteins. At baseline (Visit 1), there were weak negative correlations (Spearman constant, rho range: 0.2 – 0.5) observed between the lipoproteins and PDC, DCA and total faecal DCA, but these were not statistically significant ($p > 0.05$). When delta changes were compared, fractional changes in total, LDL and non-HDL cholesterol were positively associated with PDC on the *rr* intervention and with DCA on the *RR* intervention. No other significant association was observed.

6.4.4 Correlations between gut microbiota and PDC/PDC%

ASVs were restricted to those that had an annotated phylum and a prevalence of at least 10% (n = 385). Spearman correlations were performed between the qualifying ASV relative abundances and each of the 5 PDC/DCA measures as previously done in Chapter 5. ASVs with at least one statistically significant correlation ($p < 0.05$) were arranged in descending order of absolute magnitude of correlation coefficient and the top 20 (final n may differ slightly due to ties in magnitude) were further displayed as heatmaps and discussed in this chapter.

In summary, the ASVs with the strongest Spearman correlations with PDC/PDC% at V1 were ASVs belonging to the three phyla, *Firmicutes* (n = 15), *Actinobacteria* (n = 4) and *Bacteroidetes* (n = 1). Similar to results obtained from the Vegan study cohort, the top 20 genera that were strongly associated with the five PDC/DCA measures predominantly belonged to the *Clostridia* class. Some non-clostridial groups were also found to correlate to these measures, this included *Bifidobacterium* (belonging to *Actinobacteria* class) and *Bacteroides* (belonging to *Bacteroidia* class). Also, similar to results previously obtained from the vegan study cohort, PDC/PDC% was observed to be negatively associated with the genus *Eggerthella* while being positively associated with butyrate-producing genera *Roseburia* and *Blautia* (Figure 6-4). In addition to these, PDC was found to be positively associated with *Christensenellaceae R-7 group* and *Bifidobacterium* and PDC% with *Bacteroides*.

Comparing delta changes (i.e. ΔPDC, ΔDCA, Δtotal fDCA, ΔPDC% and ΔDCA%) on both interventions, ASVs interacted with PDC/DCA measures differently between the pea lines. Nevertheless, the top genera that correlated with ΔPDC/ΔDCA measures on either intervention belonged to one of four phyla namely; *Firmicutes* (n = 16 on either intervention), *Bacteroidetes* (n = 5 on *rr*, n = 3 on *RR*), *Actinobacteria* (n = 2 on either intervention) and Proteobacteria (n = 1 on *rr*, n = 0 on *RR*). Members of the *Clostridia* class were found to be the most dominant *Firmicutes,* belonging predominantly to the *Ruminococcaceae* or

Lachnospiraceae families. All *Bacteroidetes* that significantly correlated with PDC/DCA measures on both interventions belonged to the *Bacteroides* genus. In the phylum *Actinobacteria*, the *Bifidobacterium* genus (*Actinobacteria* class) was common to both interventions. However, the *Eggerthella* genus from the *Actinobacteria* class was significantly associated with ΔPDC/ΔDCA on the *RR* intervention, while the *Collinsella* genus, also from the *Actinobacteria* class, was significantly associated with ΔPDC(%)/ΔDCA(%) on the *rr* intervention (Figure 6-5).

In the *rr* intervention phase, genus-level relative abundance changes in *Bifidobacterium*, *Christensenellaceae R-7 group*, *Bacteroides, Alistipes* and *Ruminococcaceae* were positively associated with ΔPDC/ΔPDC% while changes in *Haemophilus*, *Colinsella*, *Blautia*, *Roseburia*, *Intestinibacter* and *Clostridium* sensu stricto genera were negatively associated with ΔPDC/ΔPDC% (Figure 6-5A). Whereas, in the *RR* phase, an increase in the relative abundance in the genera *Christensenellaceae R-7 group, Ruminococcaceae*, *Coprococcus, Eggerthella* and *Fusicatenibacter* were significantly associated with corresponding increases in PDC/PDC% while an increase in *Dorea*, *Catenibacterium*, *Butyricoccus*, *Bacteroides, Bifidobacterium* as well as some unclassified members of *Lachnospiraceae* and *Ruminococcaceae* was associated with a decrease in PDC/PDC% (Figure 6-5B). ASVs belonging to the *Family XIII AD3011 group* but resolved into distinct sequence variants were found to be both positively and inversely associated with ΔPDC/ΔPDC% on the *RR* diet.

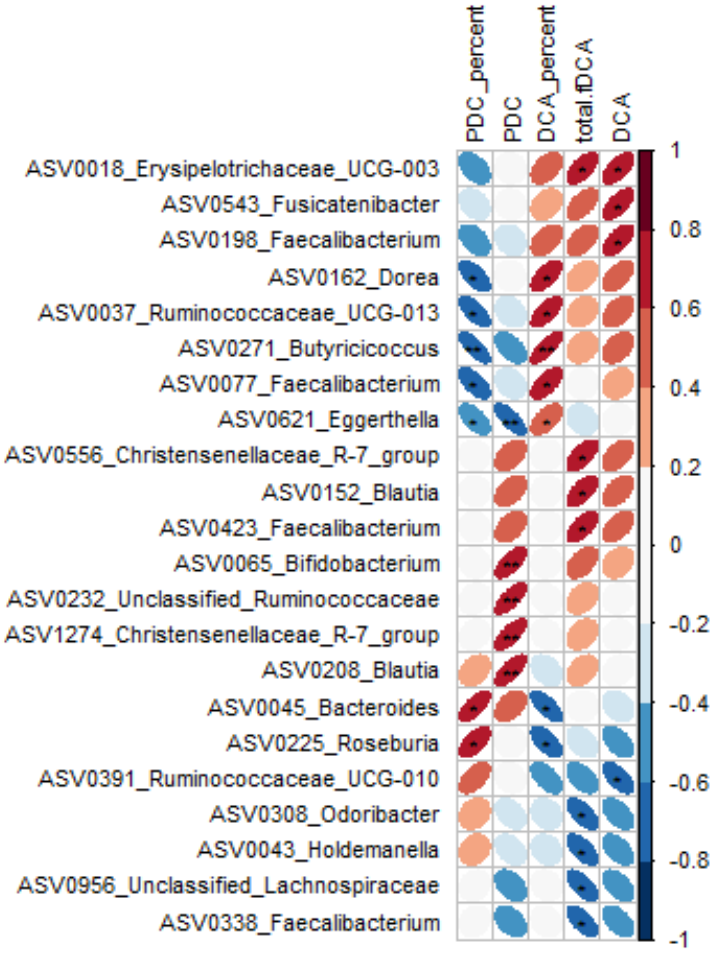


Figure 6-4: Spearman correlations between metagenomic species and each of PDC, PDC%, DCA, DCA% and total faecal DCA.

Statistically significant associations observed were strong (ρ >0.6). Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

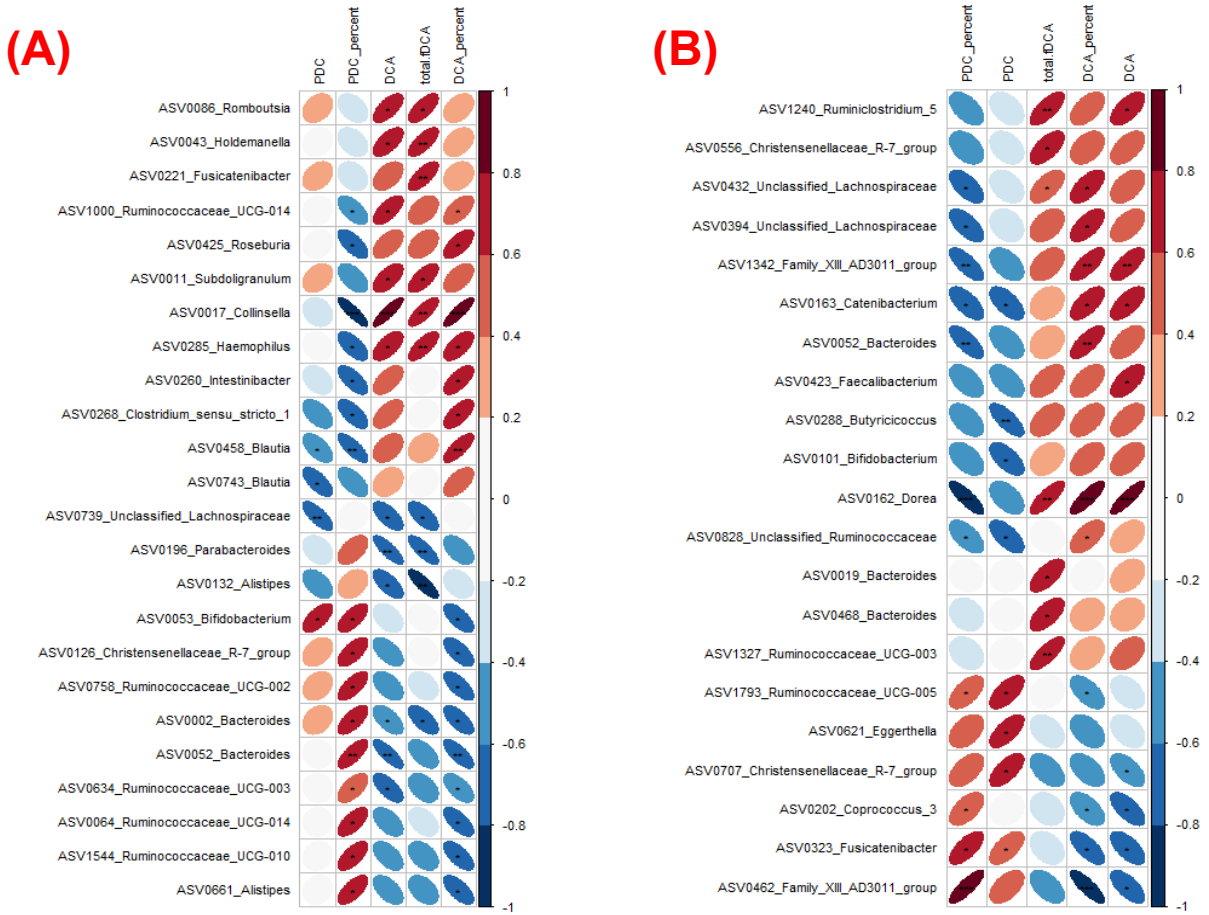


Figure 6-5: Spearman correlations of delta changes in A) rr and B) RR phases.

Fractional changes in PDC, DCA and total faecal DCA and changes in relative proportions of PDC(%), DCA(%) and ASVs were correlated. The top 20 statistically significant correlations were plotted. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

6.5 Discussion

6.5.1 Summary of findings

No significant changes in PDC concentrations or proportions were observed following either of the two legume-based interventions investigated in a double-blind crossover RCT. The interventions were food products derived from two different genotypes of the pea plant *Pisum sativum* L., namely the wrinkled *rr* pea and wild-type *RR* pea.

While there was no significant correlation between blood lipoproteins and PDC measures at baseline, changes in total, LDL and non-HDL cholesterol observed on the *rr* and *RR* interventions were positively associated with changes in PDC and DCA respectively. This indicated that changes in these variables followed similar trends (*e.g.,* a reduction in serum total cholesterol was associated with a reduction in PDC).

Bacteria from the *Clostridia* class were again found to be most associated with PDC, consistent with the observations from the Vegan study cohort. Prominent gut microbiota genera like *Bacteroides*, *Bifidobacterium*, *Roseburia* and *Christensenellaceae R7 group* were also found in significant associations with PDC/PDC% in the cohort under study.

6.5.2 General discussion

Peas are a major legume crop that is consumed all over the world (Świątecka, Kostyra, & Świątecki, 2010). Intake of peas has been well-documented to be beneficial to health, particularly with respect to its effect on blood lipids(Alizadeh, Gharaaghaji, & Gargari, 2014; Dubois et al., 1993; Gormley, Kevany, O'Donnell, & McFarlane, 1979; Sandström, Hansen, & Sørensen, 1994), however there is a paucity of reports on its effect on bile acid metabolism in humans.

From results obtained from the vegan study, it was observed that PDC production was associated with high intake of legumes, with participants with high PDC proportions (PDC%) found to consume significantly greater amounts of this food group. However, in the present study, no statistically significant change was observed in PDC concentrations following both legume interventions explored. While on average the *rr* intervention caused an increase and the *RR* intervention a decrease in PDC (as well as DCA and total faecal DCA, see Appendix 6), investigations into individual responses revealed that the averages may not accurately reflect the effects of the interventions. As such, the individual responses were presented in the Results section above. However, the PDC response to the *RR* intervention appeared more homogenous than the response to the *rr* intervention, which was more variable. This was an

interesting finding, as the diets were designed to appear identical (participants were blinded to interventions). The higher susceptibility of the *rr* responses to individual variability could not be explained by adherence (or lack of) to the prescribed diet, as compliance did not differ between the diet interventions. In the *RR* phase, all of the participants, except *n* = 2 had a reduction in PDC concentrations. These two individuals did not differ from the other group members in terms of compliance and did not have notable changes in their total faecal DCA or PDC/DCA% . However, it is possible that the participants may have altered their overall diets differentially between the intervention phases and this may have caused the different BA responses observed. The diet intervention was designed in such a way that the intervention food products were to replace habitual food items with little to no change in daily caloric and macronutrient intake. Participants may have inadvertently changed their caloric intake or reduced their intake of specific foods and nutrients (like dietary lipids or protein) without a change in total energy intake. The latter is possible if the intervention foods replaced or displaced animal protein food sources in their usual diet. This may have resulted in a generally reduced secretion of DCA since BAs, particularly DCA, are preferably stimulated to a greater extent by the presence of animal protein in the duodenum (David et al., 2014). Even though pea protein may increase faecal excretion (Spielmann et al., 2008), it is possible that the mechanism of increase does not affect and may therefore not offset reduced BA secretion from the liver. However, due to an unavailability of data on their dietary intakes during the intervention study, it was impossible to ascertain if the participants inadvertently changed their diet outside of the intervention food products, and if they did so differentially in either phase.

Due to the global effect of the interventions on absolute PDC, DCA and thus total faecal DCA concentrations, it may be more suitable to consider the changes in PDC proportions (*i.e.* PDC%) which control for fluctuations in total faecal DCA production. There was a trend for both interventions to increase PDC% suggesting that both the *rr* and *RR* interventions could positively induce polymerisation of faecal DCA in a similar manner. However, this difference was not statistically significant. The lack of significance may result from inadequate statistical power due to the low number of participants. These initial results provide preliminary data to support the need for a larger study exploring this effect.

Interestingly, there was also no difference in PDC proportions following intake of the two pea lines. This may indicate that the differences in composition of the two pea lines, such as RS content, may not result in differential promotion of PDC production in the colon. However, the results may also not have been significant due to the low number of participants recruited.

As with the vegan study cohort, there were no direct significant correlations between PDC and plasma lipoproteins at baseline. However, changes in PDC, DCA, DCA% and total faecal DCA

were found to correlate positively with changes in all lipoproteins on both interventions. Specifically, changes in all three of total cholesterol (ΔTC), LDL cholesterol (ΔLDL) and non-HDL cholesterol (ΔnHDL) were significantly associated with changes in PDC (ΔPDC) on *rr* and changes in DCA (ΔDCA) on the *RR* interventions. This is interesting because all of these measures (ΔTC , ΔLDL, ΔnHDL, ΔDCA and ΔPDC) were not statistically significant on either intervention, except for the three lipoproteins in the *RR* phase. Also, this was contrary to the *a priori* hypothesis, which would have been demonstrated by an inverse relationship where an increase in faecal DCA and PDC would result in or correlate with a reduction in serum (total/LDL) cholesterol. This suggests that the diets induced a reduction in serum lipoproteins as well as a reduction in the faecal excretion of steroids. This opens up an avenue for further exploration.

Consistent with results from the Chapter 4 and reports from literature (Ridlon et al., 2006), the ASVs most associated with PDC/PDC% metabolism were members of the *Clostridiales* order. Most of the microbial species belonging to this order have been previously shown to possess, or have phylogenetic proximity to species that possess, BA metabolizing enzymes such as 7α-dehydrogenases (Ridlon, Alves, Hylemon, & Bajaj, 2013; Ridlon et al., 2016, 2006; Vital, Rud, Rath, Pieper, & Schlüter, 2019). The results of Spearman correlations conducted at baseline (V1) revealed interesting associations, some of which were similar to those observed in the Vegan study cohort. PDC/PDC% were found to be negatively associated with *Dorea* belonging to the *Lachnospiraceae* family, *Eggerthella* belonging to the *Eggerthellaceae* family as well as *Ruminococcaceae_UCG-013*, *Butyricicoccus* and *Faecalibacterium* belonging to the *Ruminococcaceae* family. On the other hand, PDC/PDC% were in positive association with *Bifidobacterium*, *Christensenellaceae R7*, *Blautia*, *Bacteroides* and *Roseburia*.

Over the course of the interventions, ASVs such as *Christensenellaceae R7* were found to change consistently (*i.e.* positively correlated) with PDC/PDC% on both diets. *Christensenellaceae* is a model family related to optimum health in humans. The family has not only been consistently demonstrated to be inversely associated with BMI and adiposity, but is also depleted with inflammation and negatively correlates with serum triglycerides, total and LDL cholesterol (Waters & Ley, 2019). Found in higher proportions in omnivores compared to vegetarians (Barrett et al., 2018), it has been linked to diets high in dietary fibre and protein (Waters & Ley, 2019), the latter preferably from animal sources (Beaumont et al., 2017; David et al., 2014; Manor et al., 2018; Roager et al., 2016). This is similar to the correlation observed in the Vegan study cohort between PDC/PDC% and intakes of dietary fat and protein in the omnivorous diet (refer to Chapter 4). Discovered fairly recently (Morotomi, Nagai, & Watanabe, 2012), not much functional information is available on *Christensenellaceae*, however, all indications from research conducted so far are that the

family is a marker of health. A strong positive association of this family with PDC and PDC% at baseline (V1) and on both interventions thus further lends credence to PDC having a beneficial role in inflammation or/and in lipid metabolism.

The *Bacteroides* genus makes up about one-quarter of all anaerobic microbes found in the gut (Wexler, 2007) and is comprised of commensals that are found in the gut from within the first few weeks of life (Mackie, Sghir, & Gaskins, 1999). PDC/PDC% was found to be positively associated with this genus on the *rr* diet and negatively associated on the *RR* diet. Some *Bacteroides* species(Bennett, McKnight, & Coleman, 2003; Stellwag & Hylemon, 1976; Wexler, 2007) but not all(L. Yao et al., 2018) possess bile salt hydrolases (BSH) and hydroxysteroid dehydrogenases (HSDH) which help them deconjugate and oxidise primary BAs respectively. As PDC is considered to be formed following the deconjugation and subsequent dehydroxylation of conjugated primary BAs (*e.g.* TCA/GCA → DCA → PDC), it is logical that more BSH-possessing *Bacteroides* may lead to the increased production of DCA and PDC. This may explain the positive association between PDC and some members of this genus. However, this may not be the case if PDC is deleterious to some of the ASVs within this genus which may be why some were found to negatively correlate with PDC/PDC%. Although PDC may not be biologically active due to its polymeric nature, future work may investigate PDC effects on bacterial as well as host colonic cells to explore its toxicity.

Unlike the Bacteroides which are more heavily involved in deconjugation, the Clostridium group is established and known for 7α-dehydroxylation. The *Clostridium sensu stricto* genus was negatively associated with PDC/PDC%. *Clostridium sensu stricto* refers to a phylogenetic cluster of *Clostridium* species (Gupta & Gao, 2009; Wiegel, Tanner, & Rainey, 2006), some of which are the major 7α-dehydroxylating species that have been hitherto discovered and characterised (Gustafsson, Midtvedt, & Norman, 1966; Hayakawa & Hattori, 1970; Hirano, Nakama, Tamaki, Masuda, & Oda, 1981; Stellwag & Hylemon, 1979; Takamine & Imamura, 1995; Wells, Berr, Thomas, Dowling, & Hylemon, 2000; Wells & Hylemon, 2000), and are thus established to be primary producers of DCA. Since PDC production requires DCA monomers, this may explain for the inverse relationship of these DCA producers with PDC/PDC%. It may also suggest that polymerisation may proceed at the expense of this group of microbes or PDC in itself may be directly harmful to them.

Bifidobacteria are established pro-healthy gut bacteria, thus, significant associations with PDC/PDC% may be important for host physiology and gut health. A *Bifidobacterium* ASV was found to be positively associated with DCA and total fDCA but negatively associated with ΔPDC% on the *RR* diet. As *Bifidobacteria* species are also known to contain BSH genes (Cheng, Zhu, & Faden, 2019), this may explain the positive association with DCA. However,

not all *Bifidobacterium* species contain BSH (Ferrari, Pacini, & Canzi, 1980). Another *Bifidobacterium* ASV was significantly associated with PDC% on the *rr* intervention. This finding was more interesting because this particular *Bifidobacterium* was only positively associated with ΔPDC/ΔPDC% (conversely negatively with ΔDCA%) but with no correlation with ΔDCA or Δtotal fDCA. These results suggest that the specific *Bifidobacterium* may be involved only in PDC production and not DCA production and may therefore be a unique candidate PDC producer.

The *Ruminococcaceae* family is a large family of microbiota found in human gut. This family of obligate anaerobes together with the *Lachnospiraceae* and *Peptostreptococcaceae* families make up a large proportion of the total faecal bacteria in healthy adults (Ishiguro, Haskey, & Campbell, 2018). Changes in the relative abundance of genera belonging to the *Ruminococcaceae* family were found to be significantly associated with changes in PDC concentrations and proportions. An increase in *Ruminococcaceae UCG-002, Ruminococcaceae UCG-003, Ruminococcaceae UCG-010* and *Ruminococcaceae UCG*-014 was associated with an increase in PDC proportion (%) on the *rr* diet, while only *Ruminococcaceae UCG-005* was similarly associated with PDC and PDC% on the *RR* diet. Conversely, *Butyricoccus* also belonging to the *Ruminococcaceae* family was inversely associated with PDC. While *Lachnospiraceae* such as *Dorea* and *Coprococcus* were found to be inversely and *Fusicatenibacter* positively associated with PDC/PDC% on the *RR* diet, *Roseburia* and *Blautia* were found to be inversely associated with PDC/PDC% on the *rr* diet. *Intestinibacter* (from the *Peptostreptococcaceae* family) was inversely associated with PDC% on the *rr* diet only. These three families comprise bacteria that possess bile acid inducing (*bai*) genes (Vital et al., 2019). The *bai* operon codes for the enzymes that catalyse the various biotransformation processes (*e.g.* 7α/β-dehydroxylation, epimerization at the different carbon atoms) that convert unconjugated primary BAs to secondary BAs (Ridlon et al., 2006). As such, a positive association with PDC/PDC% may be due to these microbes contributing to the formation of DCA which is formed from the 7α-dehydroxylation of cholic acid to DCA, the monomer "precursor" of PDC. This is further demonstrated by the observed positive association found in some of these genera with DCA and total faecal DCA as well.

Changes in family *Erysipelotrichaceae*, to which the *Catenibacterium* and *Holdemanella* genera belong, have been linked to host lipid and cholesterol metabolism as well as the diet consumed (Kaakoush, 2015; I. Martínez et al., 2009; C. Zhang et al., 2010), for instance, a low fibre diet is thought to increase their abundance in mice (Trompette et al., 2014; Turnbaugh, Bäckhed, Fulton, & Gordon, 2008). In particular, the *Catenibacterium* genus was higher in humanised mice fed a high fat diet (Turnbaugh et al., 2009) as well as humans consuming a high fat diet (Shin, Sim, Lee, & Shin, 2016). DCA and DCA% and total faecal

DCA were observed to be positively associated while PDC% was negatively associated with these genera, further implicating PDC in a hypocholesterolaemic role.

Furthermore, an increase in PDC/PDC% corresponded with a decrease in *Colinsella* and *Hemophilus*. *Colinsella* has been previously shown to be increased in atherosclerosis (Karlsson et al., 2012). The genus, which is also capable of BA biotransformation by the presence of a HSDH (L. Liu, Aigner, & Schmid, 2011) was also found to decrease following a high protein diet (Walker et al., 2011) and a high fibre diet (Candela et al., 2016; Gomez-Arango et al., 2018). The observed inverse association with PDC/PDC% may thus further strengthen the indication that PDC is promoted by a high protein and high fibre diet. *Hemophilus* is an opportunistic pathogen implicated in wide-ranging cases of human infections, biliary cirrhosis and cholecystitis (Lv et al., 2016; Petrov et al., 2018). An inverse correlation with PDC/PDC% contributes to the narrative of a beneficial effect of PDC production.

6.5.3 Strengths and Limitations of this study

While the crossover RCT ranks highly on the hierarchy of evidence, the current study was limited in some instances, particularly in study design. As these were pilot investigations carried out in a subset of a larger study, it was not possible to conduct power calculations and the study may have been underpowered. This may have contributed to the lack of statistical significance of results obtained. A lack of significant change in PDC following either of the interventions may also be because the duration of intervention was not long enough to promote significant PDC production. Lack of information on the participants' dietary habits also hinder inferences from these results. Participants in the vegan study cohort consumed high amounts of dietary fibre, and in fact, a higher PDC% was observed in people with higher fibre intake per 1000 kJ as well as a high intake of fat and protein. Although the daily portions of pea-products were designed to supply 3.5-5.6 g of fibre per 100 g of peas daily, these amounts are not comparable to those observed in the vegan study cohort. A lack of dietary information from the participants restricted the ability to make more robust inferences from the results obtained. It is likely for instance that the "background" diet may have had an additive or depletive impact on the potential effect of pea intake on PDC production in this trial. In the absence of additional dietary information on their general diet within the period, it is impossible to ascertain whether these participants consumed less than average amounts of fibre in their overall diet, or whether they had a very different nutritional profile compared to the vegan study participants. Future investigations should, as much as possible, include a measure of habitual or total dietary intake.

It is also probable that the intervention was not adequate, for instance, the chosen legume may affect blood cholesterol levels via a different mechanism to those found within the diet in the Danish vegan cohort. Although previously demonstrated to have cholesterol-lowering effects, it may be possible that peas may exert this effect via another mechanism. Rather than the effect of a sole food source. PDC production may more likely have been promoted by the sum total of the dietary habit. Also, the food vehicles (pea hummus and mushy peas) may have impacted the effect, as food processing is known to modify digestibility of foods, and as such the accessibility to gut microbiota.

It is also important to note that relative abundances are to be interpreted with caution, as a change in relative abundance of a bacterial group does not always mean a corresponding/similar change in absolute abundance or functional capacity.

Finally, participants in this trial were mildly hypercholesterolemic and as such a significant effect was expected. Although a lowering effect on TC and LDL was observed in the *RR* intervention, PDC was not significantly reduced/affected in that phase. This may seem to weaken the hypothesis for a hypocholesterolemic role for PDC, however, PDC was found to be associated with gut microbiota that have been implicated in cholesterol reduction. More sophisticated analyses, such as mediation analyses, may better delineate the seemingly complex interplay between PDC, cholesterol metabolism, the gut microbiome and the diet in future investigations.

Nevertheless, the crossover RCT controlled for possible confounders and reduced inter-individual variability as each participant was their own control. This is known to vastly reduce the amount of bias in results obtainable from an intervention study in humans. The study was successfully carried out and internally validated, as the washout period was adequate, judging by the fact that there was no significant differences in factors between the two “baseline” visits. With regard to the *a priori* objectives set for this investigation, the results obtained from the study substantiated observations made in the vegan study that PDC production has strong correlations with key gut bacteria that are linked to health and may be beneficial to the host. Novel associations with beneficial microbes such as *Bifidobacterium* and *Christensenellaceae* corroborated this as well.

6.6 Conclusion

In conclusion, PDC was again detected in healthy human adult faeces. A slight increase in faecal PDC proportions was observed following 4-week interventions consuming pea-derived products made from two different pea genotypes, namely the wrinkled *rr* and the wild-type *RR* peas, but the observed increase was not significant statistically. There was also no difference in PDC production between these two interventions. Similar to results obtained from the vegan study cohort reported earlier in this thesis, PDC/PDC% correlated positively with microbial markers of health such as butyrate-producing and anti-inflammatory bacteria and negatively with pro-inflammatory genera. While there was no statistically significant link with serum lipoproteins directly, PDC/PDC% was found in positive associations with gut bacteria that had been previously demonstrated to correlate inversely with dyslipidaemia. Furthermore, PDC/PDC% was linked with bacteria that have been shown to be enriched by high protein and high fibre diets further substantiating the link with this type of diet as was observed in the vegan study results.

7. *IN VITRO* INVESTIGATIONS OF THE COLONIC PRODUCTION OF POLYDEOXYCHOLIC ACID

7.1 Introduction

Dietary soluble fibres (also referred to as fermentable carbohydrates in this text) have been well-documented to have hypocholesterolemic effects in humans and animal models (McRae, 2017). The mechanism underlying this observed effect is not well-understood but has been shown to involve a disruption of bile acid metabolism that leads to the increased faecal excretion of bile acids. The more bile acid is excreted from the body, the higher the *de novo* hepatic synthesis of cholesterol and consequent reduction of cholesterol in circulation. The main hypothesis of this PhD project is that the presence of soluble fibres in the gut may encourage certain members of resident gut microbiota to polymerize secondary bile acid deoxycholic acid (DCA) to make insoluble polydeoxycholic acid (PDC) which cannot be reabsorbed and is thus excreted from the body.

Given this beneficial role hypothesised for PDC, it was imperative to explore the mechanisms that were involved in its formation in the colon. Indigestible intestinal contents have been shown to collect in and travel along the different parts of the colon for up to 24 hours (Thomas et al., 2001; Torsoli et al., 1971). However, the colon is not easily accessible for sampling. Instead, an *in vitro* gut model of the proximal, transverse and distal colon would be suitable for this investigation. A validated three-stage continuous fermentation model (Macfarlane et al., 1998) was considered in this approach. However, first batch fermentation cultures, which are a simpler and quicker type of *in vitro* gut model (Williams et al., 2015), were set up to serve as a screening tool to explore preferred precursors for PDC formation and to investigate the effect of fibres on PDC production.

Of all soluble fibres, β-glucans found in oat(Hoang V T Ho et al., 2016) and barley(H V T Ho et al., 2016) have been extensively demonstrated to moderately reduce total and LDL cholesterol. Cereal β-glucans found in barley and oat are linear homopolysaccharides comprising of blocks of β (1←4)-linked glucose moieties separated by β (1←3) linkages (Lazaridou & Biliaderis, 2007). Several food regulatory bodies worldwide have since approved health claims on barley- and oat- containing foods for cholesterol-lowering effects (EFSA Panel on Dietetic Products, 2011; Health Canada, n.d.-a, n.d.-b; US Food and Drug Association, 2003, 2006, 2008). It is generally accepted that consuming at least 3g per day β-glucan soluble fibre from either whole oats or barley, or a combination of whole oats and barley would reduce blood cholesterol by 5-6% (FDA, 2008).

To explore the processes involved in PDC production *in vitro*, batch culture fermentation models of the gut were supplemented with cereal β-glucans from both oat and barley. It was hypothesised that these would serve as fuel to faecal microbiota that may be involved in the PDC production process and encourage DCA polymerisation *in vitro*.

7.2 Aim and Objectives

7.2.1 Aim

To characterise processes that favour intestinal PDC formation using *in vitro* batch culture fermentations inoculated with human faeces.

7.2.2 Objectives

To achieve this aim, two sets of batch culture experiments were conducted with detail as follows;

Experiment I

Hypothesis: *PDC will be produced in the presence of physiologically excess amounts of precursors and will be preferentially produced by a specific precursor bile acid.*

This set of experiments was set up as a screening for the best precursor for PDC. Batch cultures inoculated with faeces from individual humans were supplemented with the primary conjugated bile acids taurocholic acid (TCA), glycocholic acid (GCA), and the unconjugated form, cholic acid (CA).

Experiment II

Hypothesis: *"PDC will be produced in the presence of β-glucans in the colon".*

This set of experiments explored the effects of the presence of β-glucans on PDC production *in vitro*, using inulin and cellulose as positive and negative controls respectively. Inulin was selected as a positive control because it is a fermentable carbohydrate with no known hypocholesterolemic effect while cellulose was used as a negative control as it is non-fermentable and also has no known hypocholesterolemic property.

7.3 Materials and Methods

7.3.1 Raw materials for buffers, reagents and culture media

Peptone water, bile salts (composed of 24.6% sodium taurocholate, 11.5% sodium taurodeoxycholate, 10.2% sodium glycocholate, 9.4% sodium cholate and 2.1% sodium glycodeoxycholate) and phosphate buffered saline (PBS) tablets were purchased from Oxoid (Basingstoke, Hampshire, UK). Yeast extract, sodium chloride (NaCl), potassium phosphate monobasic (KH_2PO_4), potassium phosphate dibasic (K_2HPO_4), magnesium sulphate heptahydrate ($MgSO_4.7H_2O$), calcium chloride hexahydrate ($CaCl_2.6H_2O$), sodium bicarbonate ($NaHCO_3$), polyoxyethylenesorbitan monooleate (TWEEN® 80), hemin, Vitamin K, L-cysteine hydrochloride, resazurin sodium salt, hydrochloric acid (HCl), sodium hydroxide (NaOH), α-amylase from *Bacillus licheniformis*, pepsin from porcine gastric mucosa, calcium chloride ($CaCl_2$), pancreatin from porcine pancreas, and bovine bile were purchased from Sigma-Aldrich (Dorset, UK).

7.3.2 Substrates

For Experiment I, cholic, glycocholic and taurocholic acids were purchased from Sigma-Aldrich (Dorset, UK). For Experiment II, oat β-glucan (OBG) (70%, OGLUCAN70) fibre sample was kindly supplied by Garuda International Inc. (Exeter, California, USA). Barley grains, as a source of barley β-glucan (BBG), were kindly gifted by Hodmedod's (Suffolk, UK). Inulin (Beneo Orafti® HP, DKSH, London, UK) and cellulose (ACI Group Ltd, Slough, UK) were used in Experiment II as positive and negative controls respectively.

7.3.3 Preparation of basal nutrient medium

Basal nutrient medium for the batch culture experiments was prepared in required quantities for the experiments. Components for the medium are listed in Table 7-1 as have been previously described (Martín-Peláez et al., 2008). After preparation, medium was dispensed into 150 mL Duran bottles, autoclaved and stored at 4°C. The medium was sterilised again for 30 minutes before inoculation on the day of analysis.

Table 7-1: Components of basal nutrient medium for batch culture experiments.

BASAL NUTRIENT MEDIA	***/L***
Peptone water	2 g
Yeast extract	2 g
NaCl	0.1 g
K_2HPO_4	0.04 g
KH_2PO_4	0.04 g
$MgSO_4.7H_2O$	0.01 g
$CaCl_2.6H_20$	0.01 g
$NaHCO_3$	2 g
Tween 80	2 mL
Hemin (dissolve in a few drops of 1M NaOH)	0.05 g
Vitamin K (cut the tip off to ease dispensing & remove prior autoclaving)	10 µL
L-Cysteine HCL	0.5 g
*Bile salts**	0.5 g
Resazurin solution (0.025g/100ml)	4 mL

*In Experiment I, basal medium was modified to exclude bile salts in some of the fermentation vessels to explore the effect of the individual bile salt supplemented. Bile salts are composed of 24.6% sodium taurocholate, 11.5% sodium taurodeoxycholate, 10.2% sodium glycocholate, 9.4% sodium cholate and 2.1% sodium glycodeoxycholate.

7.3.4 Collection and preparation of faecal sample

Fresh faecal samples were collected from healthy volunteers (*n* = 6 in Experiment I and *n* = 3 in Experiment II) who had no known metabolic and gastrointestinal diseases (*e.g.,* diabetes, ulcerative colitis, Crohn's disease, irritable bowel syndrome, peptic ulcers and cancer) and had not taken antibiotics in the 3 months prior. All volunteers provided their informed consent. This study was approved by The University of Reading Research Ethics Committee (UREC 15/20). Volunteers were provided with detailed instructions on how to collect samples and were given an anaerobic jar (AnaeroJarTM 2.5 L, Oxoid Ltd) with a gas-generating kit (AnaeroGenTM, Oxoid) to keep the chamber environment anaerobic. Samples were retrieved within an hour of being voided. A 10% faecal slurry was prepared by homogenising sample (Stomacher 400, Seward, West Sussex, UK) in 0.1x PBS solution (pH 7.4) (Martín-Peláez et al., 2008). 15 mL of resulting 10% (w/v) slurry was used to inoculate the batch culture vessels.

7.3.5 Simulated *in vitro* upper gut digestion

The human upper gastrointestinal tract digestion process was simulated *in vitro* in different stages. First, barley grains were weighed out and boiled in water only. Weights before and after cooking were recorded. The cooked barley (100 g) was mixed with 150 ml of water and homogenised briefly in the stomacher (Stomacher 400, Seward, West Sussex, UK) for 5 minutes at 250 rpm to mimic mastication in the mouth. 20 mg of α-amylase was made into solution in 1 mM $CaCl_2$ (6.25 mL, pH 7.0) and was added to the "masticated" food to mimic the chewing and enzymatic activity in the mouth. The mixture was incubated at 37 ˚C for 30 minutes on a shaker to ensure thorough mixture and to maximise enzyme activity. The pH was then adjusted to 2.0 in readiness for the stomach/gastric phase. 2.7g of pepsin was dissolved in 25 ml of 0.1 M of HCl and added to the "bolus-mixture" obtained from the mouth phase. The resulting mixture was then incubated at 37 ºC for 2 hours on a shaker. Moving on to the small intestinal phase, 560 mg of pancreatin and 3.5 g of bile were dissolved in 125 ml of 0.5 M $NaHCO_3$ and added to the "digesta" for digestion of lipids. The pH was then adjusted to 7.0 and the digesta incubated at 37 ºC for 3 hours. The sample solution was afterwards transferred to a 500 Da molecular weight cut-off regenerated cellulose dialysis tubing and dialysed overnight against 0.01 M NaCl at 5 ºC to remove low molecular mass digestion products. Dialysis was allowed to proceed for a total of 17 hours with the dialysis fluid changed at 15 hours. The resulting dialysate was freeze-dried and stored until batch culture experiments were carried out. To ensure that all substrates were treated similarly, the pure fibre substrates were also mixed with water into pastes and pre-digested as outlined above.

At the point of inoculation, the dialysate was resuspended in medium and used to supplement the batch cultures at 1% (w/v) final constitution.

7.3.6 Batch culture fermentations

Two main batch culture fermentation experiments were performed to explore the set objectives in Section 7.2.2. The experiments were carried out as previously described by Cueva and colleagues (Cueva et al., 2013) with slight modifications.

In the first experiment (Experiment I), batch cultures were carried out to determine the preferred precursor for PDC formation. Batch culture vessels were supplemented with one of the three possible precursors for DCA namely taurocholic, glycocholic or cholic acids at 0.5 mM concentration. This concentrations was informed by previous reports of cecal bile acid concentrations obtained from victims of sudden death (Hamilton et al., 2007). To prevent confounding effect of other bile salts that are normally part of the basal nutrient medium, two sets of fermentations were carried out with a modified basal nutrient medium excluding bile salts (set A) and with a normal basal nutrient medium including bile salts (set B). Details of the experiment are summarised in Table 7-2.

In the second experiment (Experiment II), the effect of β-glucan on *in vitro* PDC production was investigated. This was carried out by supplementing the test vessels with pre-digested OBG and BBG and the control vessels with inulin and cellulose. Details of this experiment are presented in Table 7-3.

In both experiments, sterile batch culture fermentation vessels (150 mL working volume) were aseptically filled with 135 mL of pre-reduced sterile basal nutrient medium. The fermentation vessels were then sparged overnight with O_2-free nitrogen gas to ensure anaerobic conditions. This sparging continued throughout the fermentation experiment, as well as continuous stirring by a magnetic flea. Temperature was controlled by connecting the vessels to a circulating water bath set at 37 ºC, while pH was controlled by using 0.5 M HCl or 0.5 M NaOH connected to pH meter controllers (Electrolab260, Electrolab Ltd., Tewkesbury, UK). pH for the first and second set of experiments was set between 6.7-6.9 to mimic the distal colon and between 5.4-5.6 to mimic the proximal colon respectively. The vessels were inoculated with faecal slurry from a single individual and supplemented with 1.5 g of the fibre substrates, and the fermentations were run for 48 hours. In total, 6 donors (3 vegetarians and 3 omnivores) were investigated in experiment I and 3 donors (all omnivores) were investigated in experiment II. Each donor vessel served as a biological replicate. Batch culture supernatant (1 mL) was collected at 0, 5, 10, 24 and 48 hours and frozen at -20 ºC for PDC analyses.

Table 7-2: Schematic summary of the different sets of batch culture fermentations carried out in Experiment I investigations.

EXPERIMENT I				
BILE ACID	**BLANK (no added BA)**	**0.5 mM TAUROCHOLIC ACID**	**0.5 mM GLYCOCHOLIC ACID**	**0.5 mM CHOLIC ACID**
Modified basal medium (- BS)	 V0-A	 V1-A	 V2-A	 V3-A
Basal medium (+ BS)	 V0-B	 V1-B	 V2-B	 V3-B

Experiment I was designed to explore the effects of the different primary bile acid precursors of DCA on the production of PDC in the presence or absence of other bile salts (BS, a mixture of sodium taurocholate, sodium taurodeoxycholate, sodium glycocholate, sodium cholate and sodium glycodeoxycholate). BA: bile acid, BS: bile salts. V0= no bile acid substrate, V1 = taurocholic acid substrate, V2 = glycocholic acid substrate, V3 = cholic acid substrate. Set A = modified basal nutrient medium, Set B = normal basal nutrient medium

Table 7-3: Schematic summary of the different sets of batch culture fermentations carried out in Experiment II investigations.

EXPERIMENT II					
	BLANK (no CHO)	**OAT β-GLUCAN 1.5 g**	**BARLEY β-GLUCAN 1.5 g**	**INULIN (+ve control) 1.5 g**	**CELLULOSE (-ve control) 1.5 g**
Basal media + CHO					

Experiment II was designed to explore the effects of oat and barley β-glucans on PDC production using inulin and cellulose as positive and negative controls respectively. CHO: carbohydrate source.

7.3.7 PDC analysis

Batch culture samples were extracted for PDC analysis as described by McDonald and colleagues (J. A. K. McDonald et al., 2018) with slight modifications. 50 µL of batch culture sample was diluted with 150 µL ultra-grade water, then 600 µL of ice-cold methanol was added for protein denaturation. After incubation at -20 ºC for 3 hours, samples were centrifuged at 13 000 rpm, 4 ºC for 20 minutes and supernatant removed for monomer DCA analysis.

To quantify PDC, another 50 µL aliquot of batch culture sample was first saponified with 50 µL of 2 M NaOH at 95 ºC overnight, neutralised with 100 µL 1 M HCl before protein denaturation and extraction. Extracts were analysed for PDC using the ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS) based method developed and validated in Chapter 4.

7.3.8 Data and Statistical analyses

MS data generated during bile acids analysis was pre-processed using TargetLynx as described in previous chapters. “PDC” was computed as the difference between DCA concentrations of unsaponified aliquots and saponified aliquots of the same sample. Differences between unsaponified and saponified samples were tested by unpaired student’s t tests. DCA concentrations across sampling points were analysed by one-way analysis of variance (ANOVA). Pairwise comparisons of significant ANOVA results were further examined using Tukey Honest Significant Differences (Tukey HSD). All statistical analyses were carried out in R environment v3.5.1 using the R *stats* package. Significance was determined as *p*-value $p < 0.05$ unless otherwise stated. All data is expressed as mean ± standard error of mean (SEM) except when otherwise indicated. All graphs were plotted using Microsoft Excel for Office 365.

7.4 Results

7.4.1 Investigating preferred precursors for PDC production *in vitro* (Experiment I)

DCA, CA, GCA and TCA were monitored throughout the experiments and results are summarised in Figures 7-1 and 7-2. There was no difference observed between vegetarian (n = 3) and omnivore donors (n = 3) in any of the measured bile acids, hence the results were combined and reported as so.

In experiment I, most of the bile acids present at T0 in the blank vessel-(A) (*i.e.* inoculated with faeces but no bile acids) were unconjugated CA and DCA. This reflects the residual bile acids present in the faecal samples used to inoculate the vessels. Minimal change was observed in the proportions of the four monitored bile acids as the fermentation progressed in this vessel. DCA was present in low concentrations throughout the fermentation compared to the other vessels supplemented with the different bile acids. This confirmed that no bile salt or bile acid was present in the vessel, except that which was present in the faecal slurry. All four of DCA, CA, GCA and TCA were detected at T0 in blank vessel-(B). These vessels were not supplemented with the PDC precursor bile acids, but the unmodified basal nutrient medium did contain the bile salts TCA, GCA, CA, TDCA and GDCA. At T0 of the respective vessels, the substrate bile acid was detected at highest proportions confirming successful inoculation.

Results from Experiment I showed that most deconjugation of GCA and TCA had occurred by t = 10 hours and most of the CA was converted to DCA by t = 48 hours in all vessels (Figure 7-1).

Set A (modified medium without bile salts) vessels were also compared with vessels from set B (normal medium with bile salts). In set A vessels where the individual primary bile acid supplemented into the vessel was the sole substrate, analyses of DCA concentrations at T0, T10 and T48 revealed that there was no difference between DCA concentrations before and after saponification (alkaline hydrolysis), indicating that no PDC was produced (Figure 7-2). In contrast, there were statistically significant differences in DCA concentrations at T0 between unsaponified and saponified samples ($p < 0.05$) from all set B vessels with the normal basal medium (containing the mix of bile salts as well as bile acid substrate). Other timepoint analysed showed no differences between unsaponified and saponified samples.

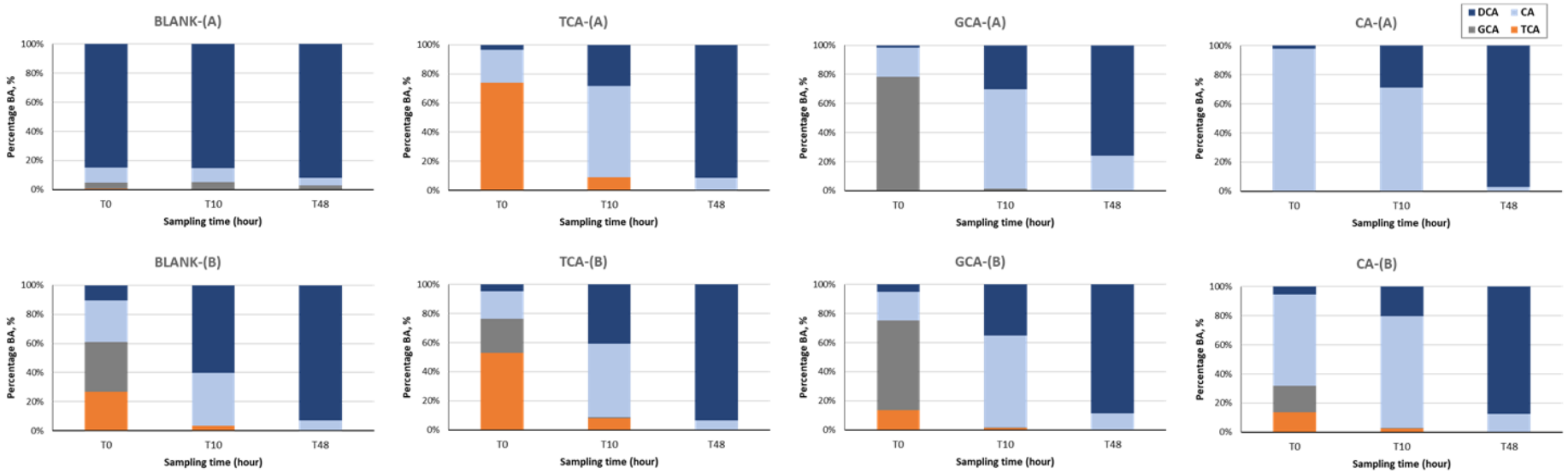


Figure 7-1: Bar plots showing average taurocholic acid (TCA), glycocholic acid (GCA), cholic acid (CA) and deoxycholic acid (DCA) percentages in different batch culture vessels throughout Experiment I.

Experiments A: no bile salts in basal nutrient medium; Experiments B: bile salts in basal medium. Bile salts are composed of 24.6% sodium taurocholate, 11.5% sodium taurodeoxycholate, 10.2% sodium glycocholate, 9.4% sodium cholate and 2.1% sodium glycodeoxycholate. Number of samples n = 6.

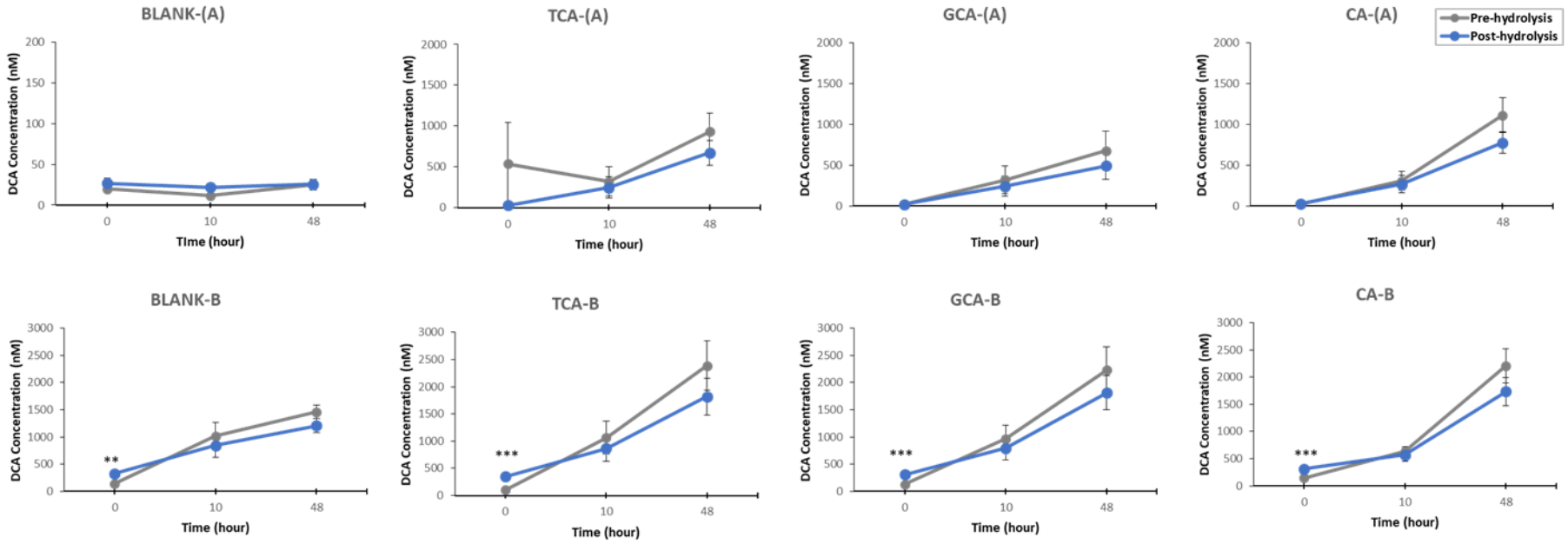


Figure 7-2: Line graphs showing DCA concentrations of batch culture samples from Experiment I. No difference in DCA concentrations before (grey) and after (blue) alkaline hydrolysis indicates that no PDC was produced.

Experiments A: no bile salts in basal nutrient medium; Experiments B: bile salts in basal medium. Bile salts are composed of 24.6% sodium taurocholate, 11.5% sodium taurodeoxycholate, 10.2% sodium glycocholate, 9.4% sodium cholate and 2.1% sodium glycodeoxycholate.
Error bars are standard error of mean. Number of samples, $n = 6$. Significance was tested by unpaired student's t tests. Significance labels *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Except in the blank vessel-(A), where there was no bile acid substrate or bile salt in the medium, DCA concentration gradually increased in all the vessels as the experiment proceeded from *t* = 0 through 48 hours. Table 7-4 summarises the significant *p*-values of ANOVA comparisons as well as pairwise comparisons between the timepoints analysed (*i.e.* T0, T10 and T48). Most of the significant increases occurred after *t* = 10 hours.

Table 7-4: P-values of analysis of variance conducted on DCA concentrations (unsaponified samples) in Experiment I.

Vessel ID	Substrate	**ANOVA**	**T10-T0**	**T48-T0**	**T48-T10**
BLANK-A	-	NS	NS	NS	NS
TCA-A	TCA only	NS	NS	NS	NS
GCA-A	GCA only	0.047	NS	0.038	NS
CA-A	CA only	<0.001	NS	<0.001	0.003
BLANK-B	BS*	<0.001	0.005	<0.001	NS
TCA-B	TCA + BS	0.001	NS	<0.001	0.025
GCA-B	GCA + BS	<0.001	NS	<0.001	0.018
CA-B	CA + BS	<0.001	NS	<0.001	<0.001

Only significant p-values ($p < 0.05$) are displayed. Pairwise comparisons are adjusted p-values from Tukey's method of multiple comparisons. T0, T10, T48 are samples obtained at time t = 0, 10 and 48 hours respectively. Experiments A: no bile salts in basal nutrient medium; Experiments B: with bile salts in basal medium. ns: not significant. *BS: bile salts in basal nutrient medium. Bile salts are composed of 24.6% sodium taurocholate, 11.5% sodium taurodeoxycholate, 10.2% sodium glycocholate, 9.4% sodium cholate and 2.1% sodium glycodeoxycholate. Number of samples, *n* = 6.

7.4.2 Effect of fibres on PDC production *in vitro* (Experiment II)

Results from experiment I are summarised in Figure 7-3. In a similar manner to experiment I, the bile acids DCA, CA, GCA and TCA were monitored throughout the experiments. While the blank (no CHO) vessel was primarily dominated by unconjugated CA and DCA at T0, other vessels that were supplemented with pre-digested fibres, had all four bile acids in considerable proportions. Deconjugation of GCA and TCA had occurred by *t* = 10 hours but surprisingly most of the CA remained unconverted at *t* = 24 hours in all vessels.

Similar to results from Experiment I Set B, DCA concentrations detected in all the vessels, except the blank, were significantly higher in saponified samples than in unsaponified samples at T0. No difference was observed at any other timepoint. In contrast to experiment I, DCA concentrations were only slightly altered throughout experiment II (Figure 7-3). DCA concentrations were only significantly altered by OBG and inulin supplemented vessels (Table 7-5, one-way ANOVA).

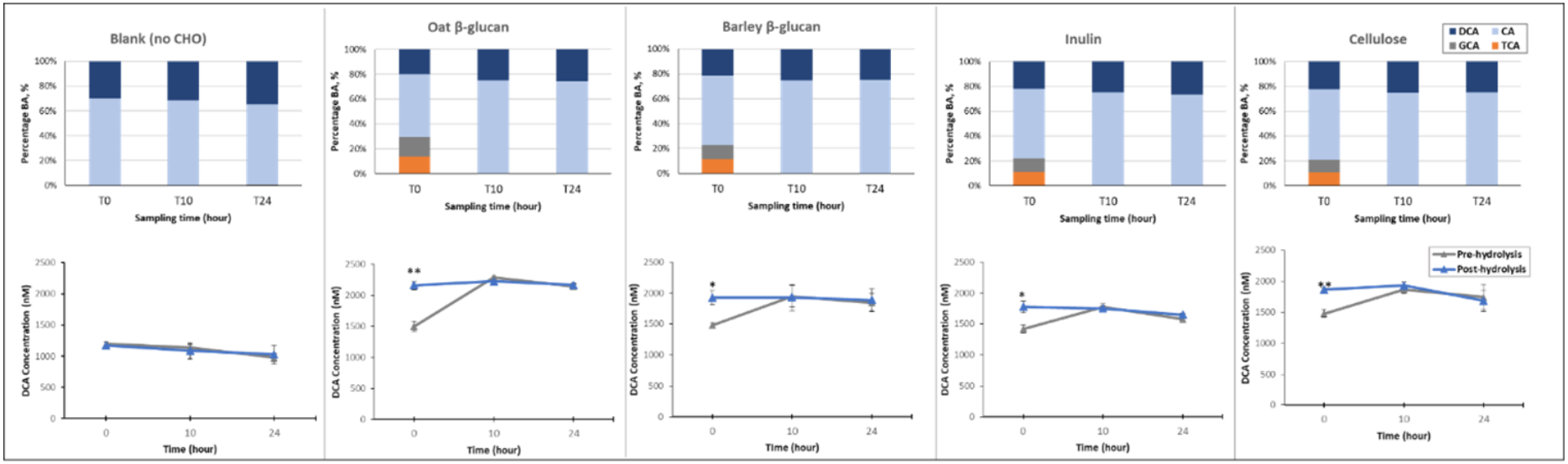


Figure 7-3: Summary charts of bile acid concentrations and proportions from Experiment II batch cultures.

Bar charts show percentages of taurocholic, glycocholic, cholic and deoxycholic acids in each vessel (chart title) throughout T0-24hours while line graphs show deoxycholic acid concentrations throughout T0-24hours. Chart titles are substrates supplemented into the respective vessels. Number of samples *n* = 3. Error bars are standard error of mean. CHO: carbohydrate.

Table 7-5: P-values of analysis of variance conducted on DCA concentrations (unsaponified samples) in Experiment II.

Vessel/Substrate	ANOVA	T10-T0	T24-T0	T24-T10
Blank (no CHO)	0.051[#]	Ns	0.048	ns
Oat β-glucan	<0.001	<0.001	<0.001	ns
Barley β-glucan	0.102[#]	0.104[#]	Ns	ns
Inulin	0.006	0.005	Ns	0.061[#]
Cellulose	Ns	Ns	Ns	ns

Only significant p-values ($p < 0.05$) are displayed. Pairwise comparisons are adjusted p-values from Tukey's method of multiple comparisons. T0, T10, T24 are samples obtained at time t = 0, 10 and 24 hours respectively ns: not significant. Number of samples, $n = 3$.[#]*p*-values are displayed because they tended towards significance. Number of samples, $n = 3$. CHO = carbohydrate source

7.5 Discussion

7.5.1 Summary of key findings

- "PDC", defined as the difference between unsaponified and saponified aliquots of samples, was detected at T0 in some vessels, however no PDC was produced in either set of batch culture experiments performed.
- DCA gradually increased as experiment I progressed whereas DCA was not seen to be produced in most of the vessels in Experiment II.

7.5.2 General discussion

This *in vitro* work was undertaken to gain further understanding of the processes and conditions favouring the production of PDC in the colon. To this end, the distal and proximal sections of the colon were modelled in the first (Experiment I) and second (Experiment II) sets of experiments respectively.

Experiment I explored the hypothesis that PDC would be produced in the presence of physiologically excess amounts of bile acids in the distal colon. The different primary bile acid precursors (namely GCA, TCA and CA) were supplemented as substrates into the batch cultures to observe if any would be a preferred precursor for DCA. The basal nutrient medium was used because the physiological digesta was expected to have been depleted of most nutrients at this point in the colon. This simple medium also allowed for a clearer observation of the effect of the substrate on metabolite and microbial constituents of the fermentation system. As the basal nutrient medium is normally prepared with salts of TCA, GCA, CA, TDCA and GDCA a modified medium was prepared excluding these compounds so that their confounding effects could be avoided. Thus, the primary bile acid substrate was investigated both in the absence (using the modified medium, designated as "set A") and presence (using normal medium, designated as "set B") of the bile salts. However, regardless of the medium or substrate, there was a rapid conversion of GCA or TCA to CA (deconjugation) and of CA to DCA (7α-dehydroxylation) in all vessels. This was expected and has been reported previously (Martin et al., 2018). Indeed, all vessels showed a gradual increase in DCA concentrations up to 48 hours after inoculation, and there was no difference in DCA trend between vessels A and B of the same substrate throughout the fermentation.

Experiment II explored the hypothesis that PDC may be produced in the presence of fermentable carbohydrates. To test this hypothesis, β-glucans were chosen because they are the most commonly investigated soluble fibres with well-documented hypocholesterolemic effects. Although OBG has had a disproportionately greater attention in this area of research,

BBG has also been reported to reduce blood cholesterol in human trials, albeit not without controversies. Both sources of soluble β-glucan were supplemented into the batch fermentation vessels, along with control vessels which were supplemented with inulin and cellulose. The basal nutrient medium was again used in these experiments to delineate the effect of the fibre substrates from the effect of other carbohydrate sources that are found in more complex media. Compared to the blank (no fibre) vessel in this experiment, the vessels supplemented with pre-digested fibres were observed to contain higher concentrations of GCA, TCA and DCA at T0. This indicates that the fibres may have bound some of the bile salts from the pancreatic phase of the simulated *in vitro* upper digestion process.

Interestingly, an almost constant level of DCA was maintained throughout Experiment II, particularly from *t* = 10 hours through *t* = 24 hours (and up to 48 hours for OBG and BBG, Appendix 7). Furthermore, percentage bar plots revealed that there was an initial and rapid unconjugation of TCA and GCA however no further DCA was produced (*i.e.* no biotransformation of CA to DCA) throughout the sampling period in Experiment II. This was in stark contrast to Experiment I where DCA was increasingly produced from the 7α-dehydroxylation of CA, even in the blank vessels with no substrates. This difference may have been due to a microbial preference for CHO as a substrate than for BA, as the latter was the only source of carbon in Experiment I while the bacteria had more options in Experiment II.

An alternative explanation for the observed difference in trend of DCA production may be due to the different pH values for both experiments. In Experiment I, pH was maintained at 6.8 (6.7 - 6.9) while in Experiment II pH was set at 5.5 (5.4 - 5.6) which were representative of the distal and proximal colons respectively. This may suggest that the latter pH did not favour the production of DCA. Although, 7α-dehydroxylation is suspected to occur in the proximal colon (Nair, Gordon, & Reback, 1967) as well as distal colon (Stellwag & Hylemon, 1979), it occurred at a much slower rate in the former than in the latter in these experiments. It was interesting to observe that the presence of fibres in the proximal colon models, which would serve as prebiotic fuel for the gut microbes that facilitate 7α-dehydroxylation, could not counteract this pH effect.

No PDC was detected in any of these models. PDC was analysed using an earlier version of the extraction methods developed and reported in Chapter 4 as the *in vitro* investigations described in this chapter were conducted prior to the clinical studies. PDC was thus computed as the difference between DCA concentrations of unsaponified/unhydrolysed samples and saponified aliquots. A consistent observation from Experiments I and II was the detection of "PDC" at T0 but only in vessels with the normal basal nutrient media (*i.e.* Experiment I Set B and Experiment II). This prompted a deeper examination of the methods and results. The

basal nutrient medium contains a mixture of bile salts comprising GDCA and TDCA as well as other bile acids. It was hypothesised that the alkaline hydrolysis procedure may induce deconjugation of taurine or glycine in the conjugated bile acid. To further confirm this, GDCA and TDCA standards were subjected to alkaline hydrolysis in a test, and they were observed to be deamidated by the hydrolysis process. Therefore, the seeming increase in DCA at T0 in saponified samples from these particular vessels was attributed to an inflation from DCA molecules split by alkaline hydrolysis from TDCA/GDCA present in the media. By T10, it is likely that the GDCA and TDCA may have been bacterially deconjugated and could no longer contribute to the difference in DCA concentrations, hence the lack of difference between unsaponified and saponified samples at later timepoints.

This observation led to the optimisation of the extraction method, the final version of which was reported in Chapter 4, where an aqueous extraction was first performed to remove all conjugated and unconjugated bile acid monomers, followed by an organic extraction for PDC oligomers. The new method was tested and confirmed to remove all GDCA and TDCA in the two-step aqueous extraction before PDC was quantified and was then applied to the samples obtained from the clinical studies already reported in previous chapters in this thesis. As it was obvious that no PDC was formed (due to the lack of difference between unsaponified and saponified from T10 onward) in the batch culture fermentation experiments, the samples from these *in vitro* experiments were not reanalysed with the optimised extraction method. While this is acknowledged as a limitation of this study, the results obtained therein provided additional insight to the mechanisms involved in PDC production and as such have been included in this thesis. Results from both experiments demonstrated that PDC production may not occur in the proximal colon, whether or not prebiotic/fermentable carbohydrates were present. However, they revealed that while the production of DCA was inhibited in the proximal colon model, this may then facilitate the delivery of unconjugated CA to the transverse and distal sections of the colon where 7α-dehydroxylation, and hypothetically PDC production, may be better favoured. While a lack of PDC production in the Experiment I vessels (particularly the ones supplemented with CA) may argue against this supposition, it may be crucial to point out that the lack of effect may have been due to a lack of fibres present in these models. Furthermore, results from the human studies revealed that PDC production is associated with conditions favouring the simultaneous presence of high amounts of bile acid and fermentable carbohydrates in the colon (*i.e.* high fat and high fibre diets). Therefore, further investigations modelling these conditions may provide better insight into the mechanisms of colonic PDC production.

7.5.3 Strengths and Limitations of this study

A major limitation of *in vitro* models is that they do not fully reflect the complex processes of the real physiological system. As a result, observations made may have limited clinical relevance. Nevertheless, these models may supply invaluable insight into otherwise inaccessible processes in the body, provided they are set up with as close as possible resemblance to the real systems and results obtained are reasonably interpreted within context.

In Experiment I, a concentration of 0.5 mM of the substrate was used based on reports by Hamilton *et al.* (Hamilton et al., 2007). However, this may have been inadequate for the production of PDC. An extensive dose-finding study may have been more appropriate to determine the best possible precursor bile acid and the optimal concentration for PDC production.

Similarly, the amount of carbohydrate substrate used in Experiment II may have been inadequate for any significant effects to be observed. The fibre substrate is usually supplemented into the static batch cultures at 1% w/v concentration which is representative of the estimated physiological equivalent in the gut of an average healthy free-living individual . However, PDC, as observed from subsequently acquired results from the clinical studies, may be produced in conditions involving high fibre consumption, hence a higher concentration of the fibre substrate may have been more suitable for the purposes of this PhD work. For instance, the final concentration of the OBG in Experiment II was 1.05 g (70 % purity of OGLUCAN product) whereas the barley grains contained 4 % β-glucan which translated to a modest 0.06 g in the vessels. Realistically, 1% β-glucan would have meant supplementing the vessels with 37.1 g of digested barley which would have overloaded the vessel. Future work must carefully select suitable substrates for the *in vitro* investigations to avoid these complexities.

7.6 Conclusion

In conclusion, PDC was not produced *in vitro* in the presence of selected fibre substrates in the proximal colon, neither was it detected when bile acid precursors were supplemented into models of the distal colon. Hence, PDC could not be produced in the presence of either one of these factors alone. These results further contribute to the observations made from the *in vivo* studies which suggest that PDC is produced in conditions of concurrent high amounts of fats (which would lead to higher secretion of BA) and fibres. Further work will thus investigate PDC production in the presence of fibres and “excess” bile acids, particularly in the transverse and distal parts of the colon where DCA production may be better promoted.

8. GENERAL DISCUSSION

8.1 Summary of work and future recommendations

To recap, the main objective of this PhD project was to explore the colonic formation of polydeoxycholic acid (PDC) in humans. It was hypothesised that PDC production was gut-microbial modulated and that the presence of fibres in the gut would modulate its production. It was further hypothesised PDC production, or its promotion, was a possible mechanism for soluble fibre-mediated lowering of blood cholesterol, by increasing the amount of bile acids (BA) excreted and consequently triggering compensatory mechanisms in the liver to replace lost bile acids (BA) in faeces.

To achieve set objectives, it was imperative to develop a high throughput pipeline using modern and cutting edge metabolomic tools for a more sensitive detection and quantification of PDC. A method development was thus carried out. In parallel, *in vitro* investigations were designed to explore the conditions that may favour PDC production in the colon using validated batch culture gut models. To investigate the effect of a high fibre diet on PDC production *in vivo*, samples were obtained from a cross-sectional cohort-study of vegans and omnivores.

Following a successful development of a suitable analytical pipeline for PDC detection and quantification (chapters 3 and 4), samples obtained from the *in vivo* human study and the *in vitro* batch culture experiments were analysed for PDC. Consequent to the results from these analyses, that implicated legume intake in PDC production, a further investigation was carried out in samples obtained from a prospective randomised controlled trial with pea-based interventions.

The following paragraphs will proceed to discuss briefly the limitations of each component of this thesis and recommendations for future work:

Firstly, an ultra-performance liquid chromatography-based method was optimised for the quantification as part of a pipeline that involved the extraction, saponification and quantification of monomer deoxycholic acid (DCA) units as an indirect measure of PDC . A major limitation of the method developed was that it did not measure PDC directly but measured in terms of constituent DCA. This may limit the clinical relevance of the results obtained. However, it provided the opportunity to compute the proportion of DCA that was polymerised. Another limitation of the method is that there was no control for other bile acid specie that may have been esterified to DCA. As such, the results obtained may better reflect “saponifiable DCA” than “polymerised DCA”. Although, PDC was detected in stool samples by targeted UPLC-

MS, it is possible for the final concentration of DCA post hydrolysis to have been inflated by any of these species if they exited in the samples.

Furthermore, the developed method may be considered rigorous, the analyst having to resort to two dilutions and injections in order to make a comprehensive quantification of all BAs in faeces. While faecal BA concentrations may largely depend on the clinical cohort being studied (for example, diseased vs healthy), another alternative may be to consider polarity switching during the LC-MS analysis (G. Chen, Zhong, Zhou, & Zhang, 2018; Pursley et al., 2014). This was applied to ESI-LC-MS analysis of metabolites from the tryptophan-kynurenine pathway, where the highly abundant tryptophan was analysed in ESI- compared to the other less abundant analytes which were analysed in ESI+ (Whiley et al., 2019). The change in polarity enabled the analysis of all the analytes in one injection, although made the assay run slightly longer in time (+2mins). BAs, particularly free unconjugated species, have in most cases been investigated in negative mode when using ESI ionization because the deprotonated ion elicits a better signal (Qiao et al., 2012). However, the analyst may investigate the performance of abundant faecal BA ions like DCA and LCA in ESI+ mode to consider maximizing the benefit of dual polarity data acquisition in a single run.

Following on from the successful synthesis of PDC standards and detection in biological samples, a method that would reliably quantify PDC oligomers with as minor modification to structure may be considered. PDC molecules are large hydrophobic molecules, insoluble in water and methanol but soluble in chloroform, although solubility is suspected to decrease as the chain length increases. These properties would need to be factored in when considering the various analytical platforms.

From the experiments presented in Chapter 4, the 1D ^{1}H-NMR spectra revealed that the platform was ill-suited for quantifying specific PDC oligomers in samples as it only gives a spectrum for the mean chain lengths of polymers present in the sample. PDC is expected to occur in mixtures in biological samples. The 1D NMR could at best confirm the presence of a PDC oligomer, but information on amount and intensity of different oligomers was unobtainable. An alternative strategy may be Diffusion Ordered Spectroscopy (DOSY)-NMR, a form of 2D NMR that measures the translational diffusion coefficient of molecules (Gu, Onorato, Xiao, Luscombe, & Loo, 2018). The diffusion coefficient is proportional to the hydrodynamic radius and thus molecular size of a molecule (Augé et al., 2009). As such, DOSY-NMR which, with the required expertise, can be carried out on most NMR machines (Lewinski, Sosnowski, Kazmierski, & Penczek, 2015) can be and has been used to study polymers (Groves, 2017).

On the other hand, the MS platforms employed gave more promising results in delineating the various chain lengths albeit with attendant challenges. For instance, the polymers (up to the 5mer) were detected on UPLC- (ESI)-MS analysis on the TQ-S in both scan and MRM mode, with the latter helping to filter out noise by selecting and measuring selected precursor ions. The SIM (for the M-H) proved to be more stable than the MRM transitions into the smaller DCA moieties, possibly due to the "soft" ESI approach employed by using a low collision energy and voltage. An obvious weakness of this method was the limited mass range of the mass analyser and the consequent inability to ascertain if longer chain PDC oligomers were present in a sample. Also, column chemistry may be an important factor to consider although in this case, the Acquity BEH C8 column was found to be adequate for the analyses described herein. Indisputably, the most positive results were obtained on the MALDI-MS which could detect up to the heptamer albeit at very low intensities. A possible reason for this may be the polydispersity of the polymer sample (i.e. heterogeneity of sizes of molecules present in a sample). Low-mass components have been observed to be over-represented compared to high-mass counterparts in polymer samples with high polydispersity indices (Hanton, 2001). The ionizability of the oligomers may also have been a challenge as a suitable and practical solvent for complete dissolution, extraction and mobile phase would be required for successful detection of oligomers by LC-MS while the choice of matrix would determine the quality of the resultant MALDI spectra. Alternative strategies to explore may include ion mobility and size exclusion chromatography coupled to an MS platform (Gies, 2012), or solvent-less techniques such as desorption-ESI (DESI) (Takáts, Wiseman, Gologan, & Cooks, 2004) and surface-assisted desorption/ionization (SALDI) (Sunner, Dratz, & Chen, 1995).

PDC was investigated in two human cohorts. PDC was detected and found to be a considerable proportion of total DCA in the faeces. The proportion of DCA found in the polymerised ("saponifiable") fraction suggested that it was an important process, but not yet fully elucidated by the results of this project.

PDC was observed to be positively correlated with predominantly beneficial gut microbe as well as others that may have bile acid metabolising properties or yet unknown functions. With regard to identifying potential PDC producers, a positive correlation was reported in chapter 5 with several members of the *Clostridia* order, as well as gut microbiota-transformed BAs, particularly those BAs produced by members of this order. *Clostridium* species and other identified *Firmicutes* are already established members of the gut microbial community that convert CA (primary BA) to DCA (secondary BA) and other tertiary BA and may understandably be potential manufacturers of PDC in the gut.

Furthermore, PDC/PDC% was found to be positively associated with anti-inflammatory gut microbes and negatively associated with pro-inflammatory gut microbes. The positive correlation of PDC/PDC% with known anti-inflammatory butyrate-producing microbes and the negative correlation with pro-inflammatory microbial groups jointly suggest the possibility of a protective role in inflammation. Indeed, BAs have recently been recognised as playing an important role in inflammation (M. L. Chen et al., 2019). However, this association may be weakened by the fact that it is the mucosal microbiota, and not the luminal microbiota, which is closer to faecal microbiota, that are more involved in inflammation processes. Regardless, *in vitro* work such as those exploring the effect of PDC on colonic cell lines and colon-derived organoids may provide more insight into this role.

Lastly, PDC/PDC% had no association with serum lipoproteins. This may be because the cohorts from which samples were obtained were predominantly healthy with well-managed cholesterol levels. Perhaps a measurement of 7α-hydroxylase converting cholesterol to BA may have shed more light on the relationship between serum levels of total and LDL cholesterol with PDC/PDC%. However, low bile acid excretion has been associated with atherosclerosis and higher CVD incidence.(G Charach et al., 1998; Gideon Charach et al., 2018). It is possible that bile acid excretion and PDC may have a role in reducing CVD risk but not via lipoproteins. This may have been evidenced by the association with gut microbiota that have been preciously linked with cardioprotective and hypocholesterolemic properties. Further investigation is warranted.

9. REFERENCES

Alexander, W. (2013). The uphill path to successful clinical trials: keeping patients enrolled. *P & T : A Peer-Reviewed Journal for Formulary Management*, *38*(4), 225–227.

Alizadeh, M., Gharaaghaji, R., & Gargari, B. P. (2014). The effects of legumes on metabolic features, insulin resistance and hepatic function tests in women with central obesity: a randomized controlled trial. *International Journal of Preventive Medicine*, *5*(6), 710–720.

Alrefai, W. A., & Gill, R. K. (2007). Bile Acid Transporters: Structure, Function, Regulation and Pathophysiological Implications. *Pharmaceutical Research*, *24*(10), 1803–1823. https://doi.org/10.1007/s11095-007-9289-1

Anderson, J. W., Allgood, L. D., Lawrence, A., Altringer, L. A., Jerdack, G. R., Hengehold, D. A., & Morel, J. G. (2000). Cholesterol-lowering effects of psyllium intake adjunctive to diet therapy in men and women with hypercholesterolemia: meta-analysis of 8 controlled trials. *The American Journal of Clinical Nutrition*, *71*(2), 472–479. https://doi.org/10.1093/ajcn/71.2.472

Anderson, J. W., Davidson, M. H., Blonde, L., Brown, W. V, Howard, W. J., Ginsberg, H., … Weingand, K. W. (2000). Long-term cholesterol-lowering effects of psyllium as an adjunct to diet therapy in the treatment of hypercholesterolemia. *The American Journal of Clinical Nutrition*, *71*(6), 1433–1438. https://doi.org/10.1093/ajcn/71.6.1433

Annesley, T. M. (2003). Ion Suppression in Mass Spectrometry. *Clinical Chemistry*, *49*(7), 1041–1044. https://doi.org/10.1373/49.7.1041

Arena, M. P., Caggianiello, G., Fiocco, D., Russo, P., Torelli, M., Spano, G., & Capozzi, V. (2014). Barley β-glucans-containing food enhances probiotic performances of beneficial bacteria. *International Journal of Molecular Sciences*, *15*(2), 3025–3039. https://doi.org/10.3390/ijms15023025

Augé, S., Schmit, P.-O., Crutchfield, C. A., Islam, M. T., Harris, D. J., Durand, E., … Delsuc, M.-A. (2009). NMR Measure of Translational Diffusion and Fractal Dimension. Application to Molecular Mass Measurement. *The Journal of Physical Chemistry B*, *113*(7), 1914–1918. https://doi.org/10.1021/jp8094424

Axelson, M., Björkhem, I., Reihnér, E., & Einarsson, K. (1991). The plasma level of 7α-hydroxy-4-cholesten-3-one reflects the activity of hepatic cholesterol 7α-hydroxylase in man. *FEBS Letters*, *284*(2), 216–218. https://doi.org/https://doi.org/10.1016/0014-5793(91)80688-Y

Baigent, C., Blackwell, L., Emberson, J., Holland, L. E., Reith, C., Bhala, N., … Collins, R.

(2010). Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet (London, England)*, *376*(9753), 1670–1681. https://doi.org/10.1016/S0140-6736(10)61350-5

Bajor, A., Gillberg, P.-G., & Abrahamsson, H. (2010). Bile acids: short and long term effects in the intestine. *Scandinavian Journal of Gastroenterology*, *45*(6), 645–664. https://doi.org/10.3109/00365521003702734

Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics*, *18*(2), 114. https://doi.org/10.1186/s12864-017-3501-4

Banerjee, S., & Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, *2012*, 282574. https://doi.org/10.1155/2012/282574

Barnes, S., Gollan, J. L., & Billing, B. H. (1977). The role of tubular reabsorption in the renal excretion of bile acids. *The Biochemical Journal*, *166*(1), 65–73. https://doi.org/10.1042/bj1660065

Barrett, H. L., Gomez-Arango, L. F., Wilkinson, S. A., McIntyre, H. D., Callaway, L. K., Morrison, M., & Dekker Nitert, M. (2018). A Vegetarian Diet Is a Major Determinant of Gut Microbiota Composition in Early Pregnancy. *Nutrients*, *10*(7), 890. https://doi.org/10.3390/nu10070890

Batta, Ashok K, Salen, G., & Shefer, S. (1985). Transformation of bile acids into iso-bile acids by clostridium perfringens: Possible transport of 3β-hydrogen via the coenzyme. *Hepatology*, *5*(6), 1126–1131. https://doi.org/10.1002/hep.1840050611

Batta, Ashok Kumar, Salen, G., Rapole, K. R., Batta, M., Batta, P., Alberts, D., & Earnest, D. (1999). Highly simplified method for gas-liquid chromatographic quantitation of bile acids and sterols in human stool. *Journal of Lipid Research* , *40*(6), 1148–1154. Retrieved from http://www.jlr.org/content/40/6/1148.abstract

Bayerdorffer, E., Mannes, G. A., Ochsenkuhn, T., Dirschedl, P., Wiebecke, B., & Paumgartner, G. (1995). Unconjugated secondary bile acids in the serum of patients with colorectal adenomas. *Gut*, *36*(2), 268–273. https://doi.org/10.1136/gut.36.2.268

Beaumont, M., Portune, K. J., Steuer, N., Lan, A., Cerrudo, V., Audebert, M., … Blachier, F. (2017). Quantity and source of dietary protein influence metabolite production by gut microbiota and rectal mucosa gene expression: a randomized, parallel, double-blind trial

in overweight humans. *The American Journal of Clinical Nutrition*, *106*(4), 1005–1019. https://doi.org/10.3945/ajcn.117.158816

Becerra-Tomás, N., Papandreou, C., & Salas-Salvadó, J. (2019). Legume Consumption and Cardiometabolic Health. *Advances in Nutrition*, *10*(Supplement_4), S437–S450. https://doi.org/10.1093/advances/nmz003

Becerra-Tomás, N., Paz-Graniel, I., W.C. Kendall, C., Kahleova, H., Rahelić, D., Sievenpiper, J. L., & Salas-Salvadó, J. (2019). Nut consumption and incidence of cardiovascular diseases and cardiovascular disease mortality: a meta-analysis of prospective cohort studies. *Nutrition Reviews*, *77*(10), 691–709. https://doi.org/10.1093/nutrit/nuz042

Begley, M., Gahan, C. G. M., & Hill, C. (2005). The interaction between bacteria and bile. *FEMS Microbiology Reviews*, *29*(4), 625–651. https://doi.org/10.1016/j.femsre.2004.09.003

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood. Education and Practice Edition*, *98*(6), 236–238. https://doi.org/10.1136/archdischild-2013-304340

Beiko, R. G. (2015). Microbial Malaise: How Can We Classify the Microbiome? *Trends in Microbiology*, *23*(11), 671–679. https://doi.org/https://doi.org/10.1016/j.tim.2015.08.009

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., … Muntner, P. (2017). Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation*, *135*(10), e146–e603. https://doi.org/10.1161/CIR.0000000000000485

Bennett, M. J., McKnight, S. L., & Coleman, J. P. (2003). Cloning and Characterization of the NAD-Dependent 7α-Hydroxysteroid Dehydrogenase from Bacteroides fragilis. *Current Microbiology*, *47*(6), 475–484. https://doi.org/10.1007/s00284-003-4079-4

Benson, G. M., Haskins, N. J., Eckers, C., Moore, P. J., Reid, D. G., Mitchell, R. C., … Suckling, K. E. (1993). Polydeoxycholate in human and hamster feces: a major product of cholate metabolism. *Journal of Lipid Research* , *34*(12), 2121–2134. Retrieved from http://www.jlr.org/content/34/12/2121.abstract

Berg, J., Tymoczko, J., & Stryer, L. (2002). The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels. In *Biochemistry* (5th ed.). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK22336/

Berger, F. K., Schwab, N., Glanemann, M., Bohle, R. M., Gärtner, B., & Groesdonk, H. V.

(2018). Flavonifractor (Eubacterium) plautii bloodstream infection following acute cholecystitis. *IDCases*, *14*, e00461–e00461. https://doi.org/10.1016/j.idcr.2018.e00461

Bertolotti, M., Del Puppo, M., Gabbi, C., Corna, F., Carulli, L., Pellegrini, E., … Carulli, N. (2008). Correlation between plasma levels of 7α-hydroxy-4-cholesten-3-one and cholesterol 7α-hydroxylation rates in vivo in hyperlipidemic patients. *Steroids*, *73*(11), 1197–1202. https://doi.org/https://doi.org/10.1016/j.steroids.2008.05.011

Bhat, B. G., Rapp, S. R., Beaudry, J. A., Napawan, N., Butteiger, D. N., Hall, K. A., … Keller, B. T. (2003). Inhibition of ileal bile acid transport and reduced atherosclerosis in apoE−/− mice by SC-435. *Journal of Lipid Research* , *44*(9), 1614–1621. https://doi.org/10.1194/jlr.M200469-JLR200

Bhattacharyya, M. K., Smith, A. M., Ellis, T. H. N., Hedley, C., & Martin, C. (1990). The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*, *60*(1), 115–122. https://doi.org/https://doi.org/10.1016/0092-8674(90)90721-P

Bhattacharyya, M., Martin, C., & Smith, A. (1993). The importance of starch biosynthesis in the wrinkled seed shape character of peas studied by Mendel. *Plant Molecular Biology*, *22*(3), 525–531. https://doi.org/10.1007/BF00015981

BHF. (2020). British Heart Foundation Heart Statistics Factsheet UK. Retrieved April 10, 2020, from https://www.bhf.org.uk/what-we-do/our-research/heart-statistics

Björkhem, I., Einarsson, K., & Hellers, G. (1973). Metabolism of Mono-and Dihydroxylated Bile Acids in Preparations of Human Liver. *European Journal of Clinical Investigation*, *3*(6), 459–465. https://doi.org/10.1111/j.1365-2362.1973.tb02215.x

Borriello, S. P., & Owen, R. W. (1982). The metabolism of lithocholic acid and lithocholic acid-3-α-sulfate by human fecal bacteria. *Lipids*, *17*(7), 477–482. https://doi.org/10.1007/BF02535328

Bortolotti, M., Kreis, R., Debard, C., Cariou, B., Faeh, D., Chetiveaux, M., … Tappy, L. (2009). High protein intake reduces intrahepatocellular lipid deposition in humans. *The American Journal of Clinical Nutrition*, *90*(4), 1002–1010. https://doi.org/10.3945/ajcn.2008.27296

Boyer, J. L. (2013). Bile formation and secretion. *Comprehensive Physiology*, *3*(3), 1035–1078. https://doi.org/10.1002/cphy.c120027

Breban, M., Tap, J., Leboime, A., Said-Nahal, R., Langella, P., Chiocchia, G., … Sokol, H. (2017). Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Annals of*

the Rheumatic Diseases, *76*(9), 1614 LP – 1622. https://doi.org/10.1136/annrheumdis-2016-211064

Brestoff, J. R., & Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. *Nature Immunology*, *14*(7), 676–684. https://doi.org/10.1038/ni.2640

Breuninger, T. A., Wawro, N., Meisinger, C., Artati, A., Adamski, J., Peters, A., … Linseisen, J. (2019). Associations between fecal bile acids, neutral sterols, and serum lipids in the KORA FF4 study. *Atherosclerosis*, *288*, 1–8. https://doi.org/https://doi.org/10.1016/j.atherosclerosis.2019.06.911

Brown, L., Rosner, B., Willett, W. W., & Sacks, F. M. (1999). Cholesterol-lowering effects of dietary fiber: a meta-analysis. *The American Journal of Clinical Nutrition*, *69*(1), 30–42. https://doi.org/10.1093/ajcn/69.1.30

Brown, M. S., & Goldstein, J. L. (1986). A receptor-mediated pathway for cholesterol homeostasis. *Science (New York, N.Y.)*, *232*(4746), 34–47. https://doi.org/10.1126/science.3513311

Bruins, A. P. (1998). Mechanistic aspects of electrospray ionization. *Journal of Chromatography A*, *794*(1), 345–357. https://doi.org/https://doi.org/10.1016/S0021-9673(97)01110-2

Brum, J., Ramsey, D., McRorie, J., Bauer, B., & Kopecky, S. L. (2018). Meta-Analysis of Usefulness of Psyllium Fiber as Adjuvant Antilipid Therapy to Enhance Cholesterol Lowering Efficacy of Statins. *The American Journal of Cardiology*, *122*(7), 1169–1174. https://doi.org/https://doi.org/10.1016/j.amjcard.2018.06.040

Buchwald, H., Varco, R. L., Matts, J. P., Long, J. M., Fitch, L. L., Campbell, G. S., … Chalmers, T. C. (1990). Effect of Partial Ileal Bypass Surgery on Mortality and Morbidity from Coronary Heart Disease in Patients with Hypercholesterolemia. *New England Journal of Medicine*, *323*(14), 946–955. https://doi.org/10.1056/NEJM199010043231404

Bundgaard-Nielsen, C., Hagstrøm, S., & Sørensen, S. (2018). Interpersonal Variations in Gut Microbiota Profiles Supersedes the Effects of Differing Fecal Storage Conditions. *Scientific Reports*, *8*(1), 17367. https://doi.org/10.1038/s41598-018-35843-0

Busnelli, M., Manzini, S., Sirtori, C. R., Chiesa, G., & Parolini, C. (2018). Effects of Vegetable Proteins on Hypercholesterolemia and Gut Microbiota Modulation. *Nutrients*, *10*(9), 1249. https://doi.org/10.3390/nu10091249

Byrdwell, W. C. (2001). Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids. *Lipids*, *36*(4), 327–346. https://doi.org/10.1007/s11745-001-0725-5

Byrne, C. S., Chambers, E. S., Morrison, D. J., & Frost, G. (2015). The role of short chain fatty acids in appetite regulation and energy homeostasis. *International Journal of Obesity*, *39*(9), 1331–1338. https://doi.org/10.1038/ijo.2015.84

Cajka, T., & Fiehn, O. (2014). Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry*, *61*, 192–206. https://doi.org/https://doi.org/10.1016/j.trac.2014.04.017

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Candela, M., Biagi, E., Soverini, M., Consolandi, C., Quercia, S., Severgnini, M., … Brigidi, P. (2016). Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet. *The British Journal of Nutrition*, *116*(1), 80–93. https://doi.org/10.1017/S0007114516001045

Carlson, J. L., Erickson, J. M., Hess, J. M., Gould, T. J., & Slavin, J. L. (2017). Prebiotic Dietary Fiber and Gut Health: Comparing the in Vitro Fermentations of Beta-Glucan, Inulin and Xylooligosaccharide. *Nutrients*, *9*(12). https://doi.org/10.3390/nu9121361

Carruthers, L. V, Moses, A., Adriko, M., Faust, C. L., Tukahebwa, E. M., Hall, L. J., … Lamberton, P. H. L. (2019). The impact of storage conditions on human stool 16S rRNA microbiome composition and diversity. *PeerJ*, *7*, e8133. https://doi.org/10.7717/peerj.8133

Caruso, V., Song, X., Asquith, M., & Karstens, L. (2019). Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. *MSystems*, *4*(1), e00163-18. https://doi.org/10.1128/mSystems.00163-18

Casdorph, H. R. (1967). Hypercholesteremia. Treatment with cholestyramine, a bile acid sequestering resin. *California Medicine*, *106*(4), 293–295.

Center for Drug Evaluation and Research. (2013). Bioanalytical Method Validation Guidance for Industry Bioanalytical Method Validation. *FDA Guidance for Industry*, (May), 1–22.

Retrieved from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htmand/orhttp://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm

Charach, G, Rabinovich, P. D., Konikoff, F. M., Grosskopf, I., Weintraub, M. S., & Gilat, T. (1998). Decreased fecal bile acid output in patients with coronary atherosclerosis. *Journal of Medicine*, *29*(3–4), 125–136.

Charach, Gideon, Argov, O., Geiger, K., Charach, L., Rogowski, O., & Grosskopf, I. (2018). Diminished bile acids excretion is a risk factor for coronary artery disease: 20-year follow up and long-term outcome. *Therapeutic Advances in Gastroenterology*, *11*, 1756283X17743420. https://doi.org/10.1177/1756283X17743420

Chassard, C., Dapoigny, M., Scott, K. P., Crouzet, L., Del'homme, C., Marquet, P., … Bernalier-Donadille, A. (2012). Functional dysbiosis within the gut microbiota of patients with constipated-irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, *35*(7), 828–838. https://doi.org/10.1111/j.1365-2036.2012.05007.x

Chen, G., Zhong, W., Zhou, Z., & Zhang, Q. (2018). Simultaneous determination of tryptophan and its 31 catabolites in mouse tissues by polarity switching UHPLC-SRM-MS. *Analytica Chimica Acta*, *1037*, 200–210. https://doi.org/https://doi.org/10.1016/j.aca.2018.02.026

Chen, M. L., Takeda, K., & Sundrud, M. S. (2019). Emerging roles of bile acids in mucosal immunity and inflammation. *Mucosal Immunology*, *12*(4), 851–861. https://doi.org/10.1038/s41385-019-0162-4

Cheng, S., Zhu, L., & Faden, H. S. (2019). Interactions of bile acids and the gut microbiota: learning from the differences in Clostridium difficile infection between children and adults. *Physiological Genomics*, *51*(6), 218–223. https://doi.org/10.1152/physiolgenomics.00034.2019

Chiang, J. Y. L. (2013). Bile acid metabolism and signaling. *Comprehensive Physiology*, *3*(3), 1191–1212. https://doi.org/10.1002/cphy.c120023

Chong, J., & Xia, J. (2017). Computational Approaches for Integrative Analysis of the Metabolome and Microbiome. *Metabolites*, *7*(4). https://doi.org/10.3390/metabo7040062

Cueva, C., Sánchez-Patán, F., Monagas, M., Walton, G. E., Gibson, G. R., Martín-Álvarez, P. J., … Moreno-Arribas, M. V. (2013). In vitro fermentation of grape seed flavan-3-ol fractions by human faecal microbiota: changes in microbial groups and phenolic metabolites. *FEMS Microbiology Ecology*, *83*(3), 792–805. https://doi.org/10.1111/1574-

6941.12037

Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P., & Macfarlane, G. T. (1987). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut*, *28*(10), 1221–1227. https://doi.org/10.1136/gut.28.10.1221

Daggy, B. P., O'Connell, N. C., Jerdack, G. R., Stinson, B. A., & Setchell, K. D. (1997). Additive hypocholesterolemic effect of psyllium and cholestyramine in the hamster: influence on fecal sterol and bile acid profiles. *Journal of Lipid Research*, *38*(3), 491–502.

Dahl, W. J., Foster, L. M., & Tyler, R. T. (2012). Review of the health benefits of peas (Pisum sativum L.). *British Journal of Nutrition*, *108*(S1), S3–S10. https://doi.org/DOI: 10.1017/S0007114512000852

Dai, D., Rhoads, W. J., Edwards, M. A., & Pruden, A. (2018). Shotgun Metagenomics Reveals Taxonomic and Functional Shifts in Hot Water Microbiome Due to Temperature Setting and Stagnation . *Frontiers in Microbiology* , Vol. 9, p. 2695. Retrieved from https://www.frontiersin.org/article/10.3389/fmicb.2018.02695

Daniels, T. F., Killinger, K. M., Michal, J. J., Wright, R. W. J., & Jiang, Z. (2009). Lipoproteins, cholesterol homeostasis and cardiac health. *International Journal of Biological Sciences*, *5*(5), 474–488. https://doi.org/10.7150/ijbs.5.474

Danielsson, H., Eneroth, P., Hellstrom, K., Lindstedt, S., & Sjovall, J. (1963). On the turnover and excretory products of cholic and chenodeoxycholic acid in man. *The Journal of Biological Chemistry*, *238*, 2299–2304.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

Davidson, R. M., & Epperson, L. E. (2018). *Microbiome Sequencing Methods for Studying Human Diseases BT - Disease Gene Identification: Methods and Protocols* (J. K. DiStefano, Ed.). https://doi.org/10.1007/978-1-4939-7471-9_5

Dawson, P. A., Lan, T., & Rao, A. (2009). Bile acid transporters. *Journal of Lipid Research*, *50*(12), 2340–2357. https://doi.org/10.1194/jlr.R900012-JLR200

de Aguiar Vallim, T. Q., Tarling, E. J., & Edwards, P. A. (2013). Pleiotropic Roles of Bile Acids in Metabolism. *Cell Metabolism*, *17*(5), 657–669. https://doi.org/https://doi.org/10.1016/j.cmet.2013.03.013

Defesche, J. C., Gidding, S. S., Harada-Shiba, M., Hegele, R. A., Santos, R. D., & Wierzbicki,

A. S. (2017). Familial hypercholesterolaemia. *Nature Reviews Disease Primers*, *3*(1), 17093. https://doi.org/10.1038/nrdp.2017.93

Dell, R. B., Holleran, S., & Ramakrishnan, R. (2002). Sample Size Determination. *ILAR Journal*, *43*(4), 207–213. https://doi.org/10.1093/ilar.43.4.207

Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., … Rossen, J. W. A. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *Journal of Biotechnology*, *243*, 16–24. https://doi.org/https://doi.org/10.1016/j.jbiotec.2016.12.022

Devkota, S., Wang, Y., Musch, M. W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., … Chang, E. B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature*, *487*(7405), 104–108. https://doi.org/10.1038/nature11225

Devlin, A. S., & Fischbach, M. A. (2015). A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nature Chemical Biology*, *11*(9), 685–690. https://doi.org/10.1038/nchembio.1864

Dona, A. C., Kyriakides, M., Scott, F., Shephard, E. A., Varshavi, D., Veselkov, K., & Everett, J. R. (2016). A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Computational and Structural Biotechnology Journal*, *14*, 135–153. https://doi.org/10.1016/j.csbj.2016.02.005

Dubois, C., Cara, L., Armand, M., Borel, P., Senft, M., Portugal, H., … Lairon, D. (1993). Effects of pea and soybean fibre on postprandial lipaemia and lipoproteins in healthy adults. *European Journal of Clinical Nutrition*, *47*(7), 508–520.

Duncan, S. H., Aminov, R. I., Scott, K. P., Louis, P., Stanton, T. B., & Flint, H. J. (2006). Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *56*(Pt 10), 2437–2441. https://doi.org/10.1099/ijs.0.64098-0

Dutta, M., Cai, J., Gui, W., & Patterson, A. D. (2019). A review of analytical platforms for accurate bile acid measurement. *Analytical and Bioanalytical Chemistry*, *411*(19), 4541–4549. https://doi.org/10.1007/s00216-019-01890-3

Eastwood, M., & Mowbray, L. (1976). The binding of the components of mixed micelle to dietary fiber. *The American Journal of Clinical Nutrition*, *29*(12), 1461–1467. https://doi.org/10.1093/ajcn/29.12.1461

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., … Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, *308*(5728), 1635–1638. https://doi.org/10.1126/science.1110591

EFSA Panel on Dietetic Products, N. and A. (NDA). (2011). Scientific Opinion on the substantiation of health claims related to beta-glucans from oats and barley and maintenance of normal blood LDL-cholesterol concentrations (ID 1236, 1299), increase in satiety leading to a reduction in energy intake (ID 851, 852), reduction of post-prandial glycaemic responses (ID 821, 824), and “digestive function” (ID 850) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. *EFSA Journal*, *9*(6), 2207. https://doi.org/10.2903/j.efsa.2011.2207

Ellegård, L., & Andersson, H. (2007). Oat bran rapidly increases bile acid excretion and bile acid synthesis: an ileostomy study. *European Journal of Clinical Nutrition*, *61*(8), 938–945. https://doi.org/10.1038/sj.ejcn.1602607

Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., … Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, *9*(7). https://doi.org/10.3390/metabo9070123

Englyst, H., Wiggins, H. S., & Cummings, J. H. (1982). Determination of the non-starch polysaccharides in plant foods by gas - Liquid chromatography of constituent sugars as alditol acetates. *The Analyst*, *107*(1272), 307–318. https://doi.org/10.1039/AN9820700307

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing*, *12*(1), 77–84. https://doi.org/10.1046/j.1365-2702.2003.00662.x

Expert Panel on Detection, Evaluation, and T. of H. B. C. in A. (2001). Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA*, *285*(19), 2486–2497. https://doi.org/10.1001/jama.285.19.2486

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., … Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, *352*(6285), 560 LP – 564. https://doi.org/10.1126/science.aad3503

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)*,

246(4926), 64–71. https://doi.org/10.1126/science.2675315

Fernandez, M. L. (1995). Distinct mechanisms of plasma LDL lowering by dietary fiber in the guinea pig: specific effects of pectin, guar gum, and psyllium. *Journal of Lipid Research*, *36*(11), 2394–2404.

Fernandez, M. L. (2001). Soluble fiber and nondigestible carbohydrate effects on plasma lipids and cardiovascular risk. *Current Opinion in Lipidology*, *12*(1), 35–40. https://doi.org/10.1097/00041433-200102000-00007

Fernandez, M. L., Ruiz, L. R., Conde, A. K., Sun, D. M., Erickson, S. K., & McNamara, D. J. (1995). Psyllium reduces plasma LDL in guinea pigs by altering hepatic cholesterol homeostasis. *Journal of Lipid Research*, *36*(5), 1128–1138.

Ferrari, A.-M., Pacini, N., & Canzi, E. (1980). A Note on Bile Acids Transformations by Strains of Bifidobacterium. *Journal of Applied Bacteriology*, *49*(2), 193–197. https://doi.org/10.1111/j.1365-2672.1980.tb05117.x

Figueira, N., Curtain, F., Beck, E., & Grafenauer, S. (2019). Consumer Understanding and Culinary Use of Legumes in Australia. *Nutrients*, *11*(7), 1575. https://doi.org/10.3390/nu11071575

Figuerola, F., Hurtado, M. L., Estévez, A. M., Chiffelle, I., & Asenjo, F. (2005). Fibre concentrates from apple pomace and citrus peel as potential fibre sources for food enrichment. *Food Chemistry*, *91*(3), 395–401. https://doi.org/https://doi.org/10.1016/j.foodchem.2004.04.036

Finegold, S. M., Song, Y., Liu, C., Hecht, D. W., Summanen, P., Könönen, E., & Allen, S. D. (2005). Clostridium clostridioforme: a mixture of three clinically important species. *European Journal of Clinical Microbiology and Infectious Diseases* , *24*(5), 319–324. https://doi.org/10.1007/s10096-005-1334-6

Fiorucci, S., Mencarelli, A., Palladino, G., & Cipriani, S. (2009). Bile-acid-activated receptors: targeting TGR5 and farnesoid-X-receptor in lipid and glucose disorders. *Trends in Pharmacological Sciences*, *30*(11), 570–580. https://doi.org/10.1016/j.tips.2009.08.001

Franco, P., Porru, E., Fiori, J., Gioiello, A., Cerra, B., Roda, G., … Roda, A. (2019). Identification and quantification of oxo-bile acids in human faeces with liquid chromatography–mass spectrometry: A potent tool for human gut acidic sterolbiome studies. *Journal of Chromatography A*, *1585*, 70–81. https://doi.org/https://doi.org/10.1016/j.chroma.2018.11.038

Fromm, H., Carlson, G. L., Hofmann, A. F., Farivar, S., & Amin, P. (1980). Metabolism in man of 7-ketolithocholic acid: precursor of cheno- and ursodeoxycholic acids. *The American Journal of Physiology*, *239*(3), G161-6. https://doi.org/10.1152/ajpgi.1980.239.3.G161

Furtado, R. H. M., & Giugliano, R. P. (2020). What Lessons Have We Learned and What Remains to be Clarified for PCSK9 Inhibitors? A Review of FOURIER and ODYSSEY Outcomes Trials. *Cardiology and Therapy*. https://doi.org/10.1007/s40119-020-00163-w

Gallo, L. C., Penedo, F. J., Carnethon, M., Isasi, C. R., Sotres-Alvarez, D., Malcarne, V. L., … Talavera, G. T. (2014). The Hispanic Community Health Study/Study of Latinos Sociocultural Ancillary Study: sample, design, and procedures. *Ethnicity & Disease*, *24*(1), 77–83.

Gao, H., Ho, S., & Williams, J. (2013, September 16). LC-MS Bioanalysis of Drugs in Tissue Samples. *Handbook of LC-MS Bioanalysis*, pp. 297–306. https://doi.org/doi:10.1002/9781118671276.ch23

Garcia-Diez, F., Garcia-Mediavilla, V., Bayon, J. E., & Gonzalez-Gallego, J. (1996). Pectin feeding influences fecal bile acid excretion, hepatic bile acid and cholesterol synthesis and serum cholesterol in rats. *The Journal of Nutrition*, *126*(7), 1766–1771. https://doi.org/10.1093/jn/126.7.1766

García, J. L., Uhía, I., & Galán, B. (2012). Catabolism and biotechnological applications of cholesterol degrading bacteria. *Microbial Biotechnology*, *5*(6), 679–699. https://doi.org/10.1111/j.1751-7915.2012.00331.x

Gardiner, B. J., Tai, A. Y., Kotsanas, D., Francis, M. J., Roberts, S. A., Ballard, S. A., … Korman, T. M. (2015). Clinical and Microbiological Characteristics of Eggerthella lenta Bacteremia. *Journal of Clinical Microbiology*, *53*(2), 626 LP – 635. https://doi.org/10.1128/JCM.02926-14

Gérard, P. (2013). Metabolism of cholesterol and bile acids by the gut microbiota. *Pathogens (Basel, Switzerland)*, *3*(1), 14–24. https://doi.org/10.3390/pathogens3010014

Gharahdaghi, F., Kirchner, M., Fernandez, J., & Mische, S. M. (1996). Peptide-Mass Profiles of Polyvinylidene Difluoride-Bound Proteins by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry in the Presence of Nonionic Detergents. *Analytical Biochemistry*, *233*(1), 94–99. https://doi.org/https://doi.org/10.1006/abio.1996.0012

Gibson, G. R., Hutkins, R., Sanders, M. E., Prescott, S. L., Reimer, R. A., Salminen, S. J., … Reid, G. (2017). Expert consensus document: The International Scientific Association for

Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nature Reviews. Gastroenterology & Hepatology*, *14*(8), 491–502. https://doi.org/10.1038/nrgastro.2017.75

Gies, A. P. (2012). Ionization Techniques for Polymer Mass Spectrometry. In *Mass Spectrometry in Polymer Chemistry* (pp. 33–56). https://doi.org/10.1002/9783527641826.ch2

Golomb, B. A., & Evans, M. A. (2008). Statin adverse effects : a review of the literature and evidence for a mitochondrial mechanism. *American Journal of Cardiovascular Drugs : Drugs, Devices, and Other Interventions*, *8*(6), 373–418. https://doi.org/10.2165/0129784-200808060-00004

Gomez-Arango, L. F., Barrett, H. L., Wilkinson, S. A., Callaway, L. K., McIntyre, H. D., Morrison, M., & Dekker Nitert, M. (2018). Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. *Gut Microbes*, *9*(3), 189–201. https://doi.org/10.1080/19490976.2017.1406584

Gormley, T. R., Kevany, J., O'Donnell, B., & McFarlane, R. (1979). Effect of Peas on Serum Cholesterol Levels in Humans. *Irish Journal of Food Science and Technology*, *3*(2), 101–109. Retrieved from http://www.jstor.org/stable/25557982

Gorzelak, M. A., Gill, S. K., Tasnim, N., Ahmadi-Vand, Z., Jay, M., & Gibson, D. L. (2015). Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool. *PLOS ONE*, *10*(8), e0134802. Retrieved from https://doi.org/10.1371/journal.pone.0134802

Greaves, J., & Roboz, J. (2014). Instrumentation. In *Mass Spectrometry for the Novice* (1st editio, pp. 35–110). https://doi.org/https://doi.org/10.1201/b15436

Griffiths, W. J., & Sjövall, J. (2010). Bile acids: analysis in biological fluids and tissues. *Journal of Lipid Research*, *51*(1), 23–41. https://doi.org/10.1194/jlr.R001941-JLR200

Griffiths, W. J., & Wang, Y. (2017). Sterolomics: State of the art, developments, limitations and challenges. *Biochimica et Biophysica Acta. Molecular and Cell Biology of Lipids*, *1862*(8), 771–773. https://doi.org/10.1016/j.bbalip.2017.03.001

Grimaldi, R., Cela, D., Swann, J. R., Vulevic, J., Gibson, G. R., Tzortzis, G., & Costabile, A. (2017). In vitro fermentation of B-GOS: impact on faecal bacterial populations and metabolic activity in autistic and non-autistic children. *FEMS Microbiology Ecology*, *93*(2). https://doi.org/10.1093/femsec/fiw233

Groves, P. (2017). Diffusion ordered spectroscopy (DOSY) as applied to polymers. *Polymer Chemistry*, *8*(44), 6700–6708. https://doi.org/10.1039/C7PY01577A

Gu, K., Onorato, J., Xiao, S. S., Luscombe, C. K., & Loo, Y.-L. (2018). Determination of the Molecular Weight of Conjugated Polymers with Diffusion-Ordered NMR Spectroscopy. *Chemistry of Materials*, *30*(3), 570–576. https://doi.org/10.1021/acs.chemmater.7b05063

Gumel, A. M., Annuar, M. S. M., & Chisti, Y. (2013). Recent Advances in the Production, Recovery and Applications of Polyhydroxyalkanoates. *Journal of Polymers and the Environment*, *21*(2), 580–605. https://doi.org/10.1007/s10924-012-0527-1

Gunness, P., & Gidley, M. J. (2010). Mechanisms underlying the cholesterol-lowering properties of soluble dietary fibre polysaccharides. *Food & Function*, *1*(2), 149–155. https://doi.org/10.1039/c0fo00080a

Gupta, R. S., & Gao, B. (2009). Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus Clostridium sensu stricto (cluster I). *International Journal of Systematic and Evolutionary Microbiology*, *59*(Pt 2), 285–294. https://doi.org/10.1099/ijs.0.001792-0

Gustafsson, B. E., Midtvedt, T., & Norman, A. (1966). ISOLATED FECAL MICROORGANISMS CAPABLE OF 7 α-DEHYDROXYLATING BILE ACIDS . *Journal of Experimental Medicine*, *123*(2), 413–432. https://doi.org/10.1084/jem.123.2.413

Ha, V., Sievenpiper, J. L., de Souza, R. J., Jayalath, V. H., Mirrahimi, A., Agarwal, A., … Jenkins, D. J. A. (2014). Effect of dietary pulse intake on established therapeutic lipid targets for cardiovascular risk reduction: a systematic review and meta-analysis of randomized controlled trials. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *186*(8), E252–E262. https://doi.org/10.1503/cmaj.131727

Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., … Huttenhower, C. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Medicine*, *9*(1), 103. https://doi.org/10.1186/s13073-017-0490-5

Hamilton, J. P., Xie, G., Raufman, J.-P., Hogan, S., Griffin, T. L., Packard, C. A., … Hofmann, A. F. (2007). Human cecal bile acids: concentration and spectrum. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *293*(1), G256–G263. https://doi.org/10.1152/ajpgi.00027.2007

Hammersley, D., & Signy, M. (2017). Ezetimibe: an update on its clinical usefulness in specific patient groups. *Therapeutic Advances in Chronic Disease*, *8*(1), 4–11.

https://doi.org/10.1177/2040622316672544

Handelsman, J. (2004). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, *68*(4), 669 LP – 685. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Hansen, S. G. K., Skov, M. N., & Justesen, U. S. (2013). Two Cases of Ruminococcus gnavus Bacteremia Associated with Diverticulitis. *Journal of Clinical Microbiology*, *51*(4), 1334 LP – 1336. https://doi.org/10.1128/JCM.03382-12

Hansen, T. H., Kern, T., Bak, E. G., Kashani, A., Allin, K. H., Nielsen, T., … Pedersen, O. (2018). Impact of a vegan diet on the human salivary microbiota. *Scientific Reports*, *8*(1), 5847. https://doi.org/10.1038/s41598-018-24207-3

Hansen, T. H., Madsen, M. T. B., Jørgensen, N. R., Cohen, A. S., Hansen, T., Vestergaard, H., … Allin, K. H. (2018). Bone turnover, calcium homeostasis, and vitamin D status in Danish vegans. *European Journal of Clinical Nutrition*, *72*(7), 1046–1054. https://doi.org/10.1038/s41430-017-0081-y

Hanton, S. D. (2001). Mass Spectrometry of Polymers and Polymer Surfaces. *Chemical Reviews*, *101*(2), 527–570. https://doi.org/10.1021/cr9901081

Hayakawa, S., & Hattori, T. (1970). 7α-dehydroxylation of cholic acid by Clostridium bifermentans strain ATCC 9714 and Clostridium sordellii strain NCIB 6929. *FEBS Letters*, *6*(2), 131–133. https://doi.org/https://doi.org/10.1016/0014-5793(70)80020-5

Health Canada. (n.d.-a). *Summary of Assessment of a Health Claim about Oat Products and Blood Cholesterol Lowering*. (November 2010).

Health Canada. (n.d.-b). *Summary of Health Canada's Assessment of a Health Claim about Barley Products and Blood Cholesterol Lowering*. (July 2012), 1–6.

Heinken, A., Ravcheev, D. A., Baldini, F., Heirendt, L., Fleming, R. M. T., & Thiele, I. (2019). Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome*, *7*(1), 75. https://doi.org/10.1186/s40168-019-0689-3

Hendriks, M. M. W. B., Eeuwijk, F. A. va., Jellema, R. H., Westerhuis, J. A., Reijmers, T. H., Hoefsloot, H. C. J., & Smilde, A. K. (2011). Data-processing strategies for metabolomics studies. *TrAC Trends in Analytical Chemistry*, *30*(10), 1685–1698. https://doi.org/https://doi.org/10.1016/j.trac.2011.04.019

Henke, M. T., Kenny, D. J., Cassilly, C. D., Vlamakis, H., Xavier, R. J., & Clardy, J. (2019).

Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences*, *116*(26), 12672 LP – 12677. https://doi.org/10.1073/pnas.1904099116

Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., … Sanders, M. E. (2014). 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*, *11*(8), 506–514. https://doi.org/10.1038/nrgastro.2014.66

Hillenkamp, F., Karas, M., Beavis, R. C., & Chait, B. T. (1991). Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers. *Analytical Chemistry*, *63*(24), 1193A-1203A. https://doi.org/10.1021/ac00024a716

Hirano, S., & Masuda, N. (1981). Transformation of bile acids by Eubacterium lentum. *Applied and Environmental Microbiology*, *42*(5), 912–915. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6947718

Hirano, S., Nakama, R., Tamaki, M., Masuda, N., & Oda, H. (1981). Isolation and characterization of thirteen intestinal microorganisms capable of 7 alpha-dehydroxylating bile acids. *Applied and Environmental Microbiology*, *41*(3), 737 LP – 745. Retrieved from http://aem.asm.org/content/41/3/737.abstract

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., … Tai, H. L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist. Reviews*, *24*(1), 3–12.

Ho, H V T, Sievenpiper, J. L., Zurbau, A., Mejia, S. B., Jovanovski, E., Au-Yeung, F., … Vuksan, V. (2016). A systematic review and meta-analysis of randomized controlled trials of the effect of barley β-glucan on LDL-C, non-HDL-C and apoB for cardiovascular disease risk reductioni-iv. *European Journal of Clinical Nutrition*, *70*(11), 1239–1245. https://doi.org/10.1038/ejcn.2016.89

Ho, Hoang V T, Sievenpiper, J. L., Zurbau, A., Blanco Mejia, S., Jovanovski, E., Au-Yeung, F., … Vuksan, V. (2016). The effect of oat β-glucan on LDL-cholesterol, non-HDL-cholesterol and apoB for CVD risk reduction: a systematic review and meta-analysis of randomised-controlled trials. *The British Journal of Nutrition*, *116*(8), 1369–1382. https://doi.org/10.1017/S000711451600341X

Hofmann, A F. (1994). Pharmacology of Ursodeoxycholic Acid, an Enterohepatic Drug.

Scandinavian Journal of Gastroenterology, *29*(sup204), 1–15. https://doi.org/10.3109/00365529409103618

Hofmann, Alan F. (1999). The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Archives of Internal Medicine*, *159*(22), 2647–2658. https://doi.org/10.1001/archinte.159.22.2647

Hollister, E. B., Oezguen, N., Chumpitazi, B. P., Luna, R. A., Weidler, E. M., Rubio-Gonzales, M., … Versalovic, J. (2019). Leveraging Human Microbiome Features to Diagnose and Stratify Children with Irritable Bowel Syndrome. *The Journal of Molecular Diagnostics*, *21*(3), 449–461. https://doi.org/https://doi.org/10.1016/j.jmoldx.2019.01.006

Horton, J. D., Cuthbert, J. A., & Spady, D. K. (1994). Regulation of hepatic 7 alpha-hydroxylase expression by dietary psyllium in the hamster. *The Journal of Clinical Investigation*, *93*(5), 2084–2092. https://doi.org/10.1172/JCI117203

Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, *61*(1), 79–90. https://doi.org/10.1198/000313007X172556

Hudson, J. I., Pope, H. G. J., & Glynn, R. J. (2005). The Cross-Sectional Cohort Study: An Underutilized Design. *Epidemiology*, *16*(3). Retrieved from https://journals.lww.com/epidem/Fulltext/2005/05000/The_Cross_Sectional_Cohort_Study__An_Underutilized.14.aspx

Hylemon, P. B., Zhou, H., Pandak, W. M., Ren, S., Gil, G., & Dent, P. (2009). Bile acids as regulatory molecules. *Journal of Lipid Research*, *50*(8), 1509–1520. https://doi.org/10.1194/jlr.R900007-JLR200

Imray, C. H., Radley, S., Davis, A., Barker, G., Hendrickse, C. W., Donovan, I. A., … Neoptolemos, J. P. (1992). Faecal unconjugated bile acids in patients with colorectal cancer or polyps. *Gut*, *33*(9), 1239–1245. https://doi.org/10.1136/gut.33.9.1239

Inagaki, T., Moschetta, A., Lee, Y.-K., Peng, L., Zhao, G., Downes, M., … Kliewer, S. A. (2006). Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3920 LP – 3925. https://doi.org/10.1073/pnas.0509592103

Ishiguro, E., Haskey, N., & Campbell, K. (2018). *Chapter 3 - Gut Microbiota Throughout the Lifespan* (E. Ishiguro, N. Haskey, & K. B. T.-G. M. Campbell, Eds.). https://doi.org/https://doi.org/10.1016/B978-0-12-810541-2.00003-8

Islam, K. B. M. S., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., … Yokota, A. (2011). Bile Acid Is a Host Factor That Regulates the Composition of the Cecal Microbiota in Rats. *Gastroenterology*, *141*(5), 1773–1781. https://doi.org/https://doi.org/10.1053/j.gastro.2011.07.046

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12. https://doi.org/https://doi.org/10.1016/0197-2456(95)00134-4

Javitt, N. B. (2002). Cholesterol, Hydroxycholesterols, and Bile Acids. *Biochemical and Biophysical Research Communications*, *292*(5), 1147–1153. https://doi.org/https://doi.org/10.1006/bbrc.2001.2013

Jia, E.-T., Liu, Z.-Y., Pan, M., Lu, J.-F., & Ge, Q.-Y. (2019). Regulation of bile acid metabolism-related signaling pathways by gut microbiota in diseases. *Journal of Zhejiang University. Science. B*, *20*(10), 781–792. https://doi.org/10.1631/jzus.B1900073

Jia, W., Li, H., Zhao, L., & Nicholson, J. K. (2008, February). Gut microbiota: a potential new territory for drug targeting. *Nature Reviews. Drug Discovery*, Vol. 7, pp. 123–129. https://doi.org/10.1038/nrd2505

Jiang, D., Armour, C. R., Hu, C., Mei, M., Tian, C., Sharpton, T. J., & Jiang, Y. (2019). Microbiome Multi-Omics Network Analysis: Statistical Considerations, Limitations, and Opportunities. *Frontiers in Genetics*, *10*, 995. https://doi.org/10.3389/fgene.2019.00995

Jo, J.-H., Kennedy, E. A., & Kong, H. H. (2016). Research Techniques Made Simple: Bacterial 16S Ribosomal RNA Gene Sequencing in Cutaneous Research. *Journal of Investigative Dermatology*, *136*(3), e23–e27. https://doi.org/https://doi.org/10.1016/j.jid.2016.01.005

Jones, J. M. (2014). CODEX-aligned dietary fiber definitions help to bridge the “fiber gap”. *Nutrition Journal*, *13*, 34. https://doi.org/10.1186/1475-2891-13-34

Jones, M. L., Martoni, C. J., & Prakash, S. (2012). Cholesterol lowering and inhibition of sterol absorption by Lactobacillus reuteri NCIMB 30242: a randomized controlled trial. *European Journal of Clinical Nutrition*, *66*(11), 1234–1241. https://doi.org/10.1038/ejcn.2012.126

Jones, M., Tett, S., Peeters, G. M. E. E., Mishra, G. D., & Dobson, A. (2017). New-Onset Diabetes After Statin Exposure in Elderly Women: The Australian Longitudinal Study on Women’s Health. *Drugs & Aging*, *34*(3), 203–209. https://doi.org/10.1007/s40266-017-0435-0

Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal*, *20*(5), 453 LP – 458. https://doi.org/10.1136/emj.20.5.453

Jones, B. V, Begley, M., Hill, C., Gahan, C. G. M., & Marchesi, J. R. (2008). 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*, *105*(36), 13580–13585. https://doi.org/10.1073/pnas.0804437105

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., … Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*, *60*(5), 631 LP – 637. https://doi.org/10.1136/gut.2010.223263

Jose, J. (2016). Statins and its hepatic effects: Newer data, implications, and changing recommendations. *Journal of Pharmacy & Bioallied Sciences*, *8*(1), 23–28. https://doi.org/10.4103/0975-7406.171699

Jovanovski, E., Yashpal, S., Komishon, A., Zurbau, A., Blanco Mejia, S., Ho, H. V. T., … Vuksan, V. (2018). Effect of psyllium (Plantago ovata) fiber on LDL cholesterol and alternative lipid targets, non-HDL cholesterol and apolipoprotein B: a systematic review and meta-analysis of randomized controlled trials. *The American Journal of Clinical Nutrition*, *108*(5), 922–932. https://doi.org/10.1093/ajcn/nqy115

Joyce, S. A., & Gahan, C. G. M. (2016). Bile Acid Modifications at the Microbe-Host Interface: Potential for Nutraceutical and Pharmaceutical Interventions in Host Health. *Annual Review of Food Science and Technology*, *7*(1), 313–333. https://doi.org/10.1146/annurev-food-041715-033159

Kaakoush, N. O. (2015). Insights into the Role of Erysipelotrichaceae in the Human Host. *Frontiers in Cellular and Infection Microbiology*, *5*, 84. https://doi.org/10.3389/fcimb.2015.00084

Kageyama, A., Benno, Y., & Nakase, T. (1999). Phylogenetic evidence for the transfer of Eubacterium lentum to the genus Eggerthella as Eggerthella lenta gen. nov., comb. nov. *International Journal of Systematic Bacteriology*, *49 Pt 4*, 1725–1732. https://doi.org/10.1099/00207713-49-4-1725

Kakiyama, G., Muto, A., Takei, H., Nittono, H., Murai, T., Kurosawa, T., … Bajaj, J. S. (2014). A simple and accurate HPLC method for fecal bile acid profile in healthy and cirrhotic subjects: validation by GC-MS and LC-MS. *Journal of Lipid Research*, *55*(5), 978–990.

https://doi.org/10.1194/jlr.D047506

Kakiyama, G., Pandak, W. M., Gillevet, P. M., Hylemon, P. B., Heuman, D. M., Daita, K., … Bajaj, J. S. (2013). Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *Journal of Hepatology*, *58*(5), 949–955. https://doi.org/https://doi.org/10.1016/j.jhep.2013.01.003

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, *60*(20), 2299–2301. https://doi.org/10.1021/ac00171a028

Karlsson, F. H., Fåk, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., … Nielsen, J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nature Communications*, *3*(1), 1245. https://doi.org/10.1038/ncomms2266

Kasahara, K., Krautkramer, K. A., Org, E., Romano, K. A., Kerby, R. L., Vivas, E. I., … Rey, F. E. (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. *Nature Microbiology*, *3*(12), 1461–1471. https://doi.org/10.1038/s41564-018-0272-x

Kelsey, M. I., Molina, J. E., Huang, S. K., & Hwang, K. K. (1980). The identification of microbial metabolites of sulfolithocholic acid. *Journal of Lipid Research* , *21*(6), 751–759. Retrieved from http://www.jlr.org/content/21/6/751.abstract

Kho, Z. Y., & Lal, S. K. (2018). The Human Gut Microbiome – A Potential Controller of Wellness and Disease . *Frontiers in Microbiology* , Vol. 9, p. 1835. Retrieved from https://www.frontiersin.org/article/10.3389/fmicb.2018.01835

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., & Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *Journal of the American Society for Mass Spectrometry*, *11*(11), 942–950. https://doi.org/10.1016/S1044-0305(00)00163-X

Klamczynska, B., Czuchajowska, Z., & Baik, B.-K. (2001). Composition, soaking, cooking properties and thermal characteristics of starch of chickpeas, wrinkled peas and smooth peas. *International Journal of Food Science & Technology*, *36*(5), 563–572. https://doi.org/10.1046/j.1365-2621.2001.00486.x

Kooistra, E. (1962). On the differences between smooth and three types of wrinkled peas.

Euphytica, *11*(3), 357–373. https://doi.org/10.1007/BF00031985

Korpela, J. T., Adlercreutz, H., & Turunen, M. J. (1988). Fecal free and conjugated bile acids and neutral sterols in vegetarians, omnivores, and patients with colorectal cancer. *Scandinavian Journal of Gastroenterology*, *23*(3), 277–283. https://doi.org/10.3109/00365528809093865

Korpela, J. T., Fotsis, T., & Adlercreutz, H. (1986). Multicomponent analysis of bile acids in faeces by anion exchange and capillary column gas-liquid chromatography: Application in oxytetracycline treated subjects. *Journal of Steroid Biochemistry*, *25*(2), 277–284. https://doi.org/https://doi.org/10.1016/0022-4731(86)90429-2

Korpela, J. T., Korpela, R., & Adlercreutz, H. (1992). Fecal bile acid metabolic pattern after administration of different types of bread. *Gastroenterology*, *103*(4), 1246–1253. https://doi.org/10.5555/uri:pii:0016508592915112

Kosson, R., Czuchajowska, Z., & Pomeranz, Y. (1994). Smooth and wrinkled peas. 2. Distribution of protein, lipid, and fatty acids in seed and milling fractions. *Journal of Agricultural and Food Chemistry*, *42*(1), 96–99. https://doi.org/10.1021/jf00037a015

Kreisberg, R. A., & Oberman, A. (2002). Lipids and Atherosclerosis: Lessons Learned from Randomized Controlled Trials of Lipid Lowering and Other Relevant Studies. *The Journal of Clinical Endocrinology & Metabolism*, *87*(2), 423–437. https://doi.org/10.1210/jcem.87.2.8057

Kristensen, N. B., Madsen, M. L., Hansen, T. H., Allin, K. H., Hoppe, C., Fagt, S., … Pedersen, O. (2015). Intake of macro- and micronutrients in Danish vegans. *Nutrition Journal*, *14*(1), 115. https://doi.org/10.1186/s12937-015-0103-3

Kritchevsky, D., & Story, J. A. (1974). Binding of bile salts in vitro by nonnutritive fiber. *The Journal of Nutrition*, *104*(4), 458–462. https://doi.org/10.1093/jn/104.4.458

Kumari, R., Ahuja, V., & Paul, J. (2013). Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. *World Journal of Gastroenterology*, *19*(22), 3404–3414. https://doi.org/10.3748/wjg.v19.i22.3404

L.B., G. A., D., L. A., A., B. R., Dana, D., M., L. J., L., R. B., … J., M.-D. E. (2009). Association Between the Dietary Approaches to Hypertension Diet and Hypertension in Youth With Diabetes Mellitus. *Hypertension*, *53*(1), 6–12. https://doi.org/10.1161/HYPERTENSIONAHA.108.116665

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical

trials. *Controlled Clinical Trials*, *2*(2), 93–113. https://doi.org/https://doi.org/10.1016/0197-2456(81)90001-5

Lane, A. N. (2012). Principles of NMR for Applications in Metabolomics. In T. M. Fan, A. Lane, & R. Higashi (Eds.), *The Handbook of Metabolomics. Methods in Pharmacology and Toxicology.* https://doi.org/https://doi.org/10.1007/978-1-61779-618-0_6

Lange, M., Ni, Z., Criscuolo, A., & Fedorova, M. (2019). Liquid Chromatography Techniques in Lipidomics Research. *Chromatographia*, *82*(1), 77–100. https://doi.org/10.1007/s10337-018-3656-4

Larsen, N., Vogensen, F. K., van den Berg, F. W. J., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., … Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PloS One*, *5*(2), e9085. https://doi.org/10.1371/journal.pone.0009085

Lau, S. K. P., Woo, P. C. Y., Woo, G. K. S., Fung, A. M. Y., Ngan, A. H. Y., Song, Y., … Yuen, K. (2006). Bacteraemia caused by Anaerotruncus colihominis and emended description of the species. *Journal of Clinical Pathology*, *59*(7), 748–752. https://doi.org/10.1136/jcp.2005.031773

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, *307*(1), 80–86. https://doi.org/10.1111/j.1574-6968.2010.01965.x

Lawson, P. A., Song, Y., Liu, C., Molitoris, D. R., Vaisanen, M.-L., Collins, M. D., & Finegold, S. M. (2004). Anaerotruncus colihominis gen. nov., sp. nov., from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *54*(Pt 2), 413–417. https://doi.org/10.1099/ijs.0.02653-0

Lazaridou, A., & Biliaderis, C. G. (2007). Molecular aspects of cereal β-glucan functionality: Physical properties, technological applications and physiological effects. *Journal of Cereal Science*, *46*(2), 101–118. https://doi.org/https://doi.org/10.1016/j.jcs.2007.05.003

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., … consortium, M. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, *500*(7464), 541–546. https://doi.org/10.1038/nature12506

Lee, J.-Y., Arai, H., Nakamura, Y., Fukiya, S., Wada, M., & Yokota, A. (2013). Contribution of the 7β-hydroxysteroid dehydrogenase from Ruminococcus gnavus N53 to ursodeoxycholic acid formation in the human colon. *Journal of Lipid Research* , *54*(11),

3062–3069. https://doi.org/10.1194/jlr.M039834

Lenth, R. V. (2001). Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician*, *55*(3), 187–193. https://doi.org/10.1198/000313001317098149

Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, *14*(7), 641–642. https://doi.org/10.1038/nmeth.4346

Levine, J. S., & Burakoff, R. (2011). Extraintestinal manifestations of inflammatory bowel disease. *Gastroenterology & Hepatology*, *7*(4), 235–241.

Lewinski, P., Sosnowski, S., Kazmierski, S., & Penczek, S. (2015). l-Lactide polymerization studied by 1H NMR with diffusion-ordered spectroscopy (DOSY): a “One NMR Tube Experiment” providing data on monomer conversion, polymer structure, Mn and Mw. *Polymer Chemistry*, *6*(24), 4353–4357. https://doi.org/10.1039/C5PY00455A

Li, D. (2011). Chemistry behind Vegetarianism. *Journal of Agricultural and Food Chemistry*, *59*(3), 777–784. https://doi.org/10.1021/jf103846u

Li, F., Hullar, M. A. J., Schwarz, Y., & Lampe, J. W. (2009). Human Gut Bacterial Communities Are Altered by Addition of Cruciferous Vegetables to a Controlled Fruit- and Vegetable-Free Diet. *The Journal of Nutrition*, *139*(9), 1685–1691. https://doi.org/10.3945/jn.109.108191

Li, G., Cao, Q., Liu, Y., DeLaney, K., Tian, Z., Moskovets, E., & Li, L. (2019). Characterizing and alleviating ion suppression effects in atmospheric pressure matrix-assisted laser desorption/ionization. *Rapid Communications in Mass Spectrometry : RCM*, *33*(4), 327–335. https://doi.org/10.1002/rcm.8358

Li, T., & Chiang, J. Y. L. (2014). Bile Acid Signaling in Metabolic Disease and Drug Therapy. *Pharmacological Reviews*, *66*(4), 948 LP – 983. https://doi.org/10.1124/pr.113.008201

Liang, H. R., Foltz, R. L., Meng, M., & Bennett, P. (2003). Ionization enhancement in atmospheric pressure chemical ionization and suppression in electrospray ionization between target drugs and stable-isotope-labeled internal standards in quantitative liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM*, *17*(24), 2815–2821. https://doi.org/10.1002/rcm.1268

Liang, Y., Dong, T., Chen, M., He, L., Wang, T., Liu, X., … Xia, Y. (2020). Systematic Analysis of Impact of Sampling Regions and Storage Methods on Fecal Gut Microbiome and Metabolome Profiles. *MSphere*, *5*(1), e00763-19. https://doi.org/10.1128/mSphere.00763-19

Linton, M., Yancey, P., Davies, S., & Al., E. (2000). *The Role of Lipids and Lipoproteins in Atherosclerosis. [Updated 2019 Jan 3].* (K. Feingold, B. Anawalt, A. Boyce, & E. Al., Eds.). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK343489/

Liu, C., Li, J., Zhang, Y., Philip, A., Shi, E., Chi, X., & Meng, J. (2015). Influence of glucose fermentation on CO2 assimilation to acetate in homoacetogen Blautia coccoides GA-1. *Journal of Industrial Microbiology & Biotechnology*, *42*(9), 1217–1224. https://doi.org/10.1007/s10295-015-1646-1

Liu, G., & Aubry, A.-F. (2013, September 16). Best Practices in Biological Sample Preparation for LC-MS Bioanalysis. *Handbook of LC-MS Bioanalysis*, pp. 165–184. https://doi.org/doi:10.1002/9781118671276.ch14

Liu, L., Aigner, A., & Schmid, R. D. (2011). Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7β-hydroxysteroid dehydrogenase from Collinsella aerofaciens. *Applied Microbiology and Biotechnology*, *90*(1), 127–135. https://doi.org/10.1007/s00253-010-3052-y

Liu, Yanan, Rong, Z., Xiang, D., Zhang, C., & Liu, D. (2018). Detection technologies and metabolic profiling of bile acids: a comprehensive review. *Lipids in Health and Disease*, *17*(1), 121. https://doi.org/10.1186/s12944-018-0774-9

Liu, Yixuan, Zhang, L., Wang, X., Wang, Z., Zhang, J., Jiang, R., … Duan, L. (2016). Similar Fecal Microbiota Signatures in Patients With Diarrhea-Predominant Irritable Bowel Syndrome and Patients With Depression. *Clinical Gastroenterology and Hepatology*, *14*(11), 1602-1611.e5. https://doi.org/https://doi.org/10.1016/j.cgh.2016.05.033

Lorenzo-Zúñiga, V., Bartolí, R., Planas, R., Hofmann, A. F., Viñado, B., Hagey, L. R., … Gassull, M. A. (2003). Oral bile acids reduce bacterial overgrowth, bacterial translocation, and endotoxemia in cirrhotic rats. *Hepatology*, *37*(3), 551–557. https://doi.org/10.1053/jhep.2003.50116

Louis, P., & Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters*, *294*(1), 1–8. https://doi.org/10.1111/j.1574-6968.2009.01514.x

Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, *19*(1), 29–41. https://doi.org/10.1111/1462-2920.13589

Louis, T. A., Lavori, P. W., Bailar, J. C. 3rd, & Polansky, M. (1984). Crossover and self-controlled designs in clinical research. *The New England Journal of Medicine*, *310*(1),

24–31. https://doi.org/10.1056/NEJM198401053100106

Lv, L.-X., Fang, D.-Q., Shi, D., Chen, D.-Y., Yan, R., Zhu, Y.-X., … Li, L.-J. (2016). Alterations and correlations of the gut microbiome, metabolism and immunity in patients with primary biliary cirrhosis. *Environmental Microbiology*, *18*(7), 2272–2286. https://doi.org/10.1111/1462-2920.13401

MacDonald, I. A., Jellett, J. F., Mahony, D. E., & Holdeman, L. V. (1979). Bile salt 3 alpha- and 12 alpha-hydroxysteroid dehydrogenases from Eubacterium lentum and related organisms. *Applied and Environmental Microbiology*, *37*(5), 992–1000.

Macfarlane, G. T., Macfarlane, S., & Gibson, G. R. (1998). Validation of a Three-Stage Compound Continuous Culture System for Investigating the Effect of Retention Time on the Ecology and Metabolism of Bacteria in the Human Colon. *Microbial Ecology*, *35*(2), 180–187. https://doi.org/10.1007/s002489900072

Machiels, K., Sabino, J., Vandermosten, L., Joossens, M., Arijs, I., de Bruyn, M., … Vermeire, S. (2017). Specific members of the predominant gut microbiota predict pouchitis following colectomy and IPAA in UC. *Gut*, *66*(1), 79 LP – 88. https://doi.org/10.1136/gutjnl-2015-309398

Mackie, R. I., Sghir, A., & Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *The American Journal of Clinical Nutrition*, *69*(5), 1035s-1045s. https://doi.org/10.1093/ajcn/69.5.1035s

Maier, T. V, Lucio, M., Lee, L. H., VerBerkmoes, N. C., Brislawn, C. J., Bernhardt, J., … Jansson, J. K. (2017). Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome. *MBio*, *8*(5), e01343-17. https://doi.org/10.1128/mBio.01343-17

Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., … Evans, R. M. (1995). The nuclear receptor superfamily: The second decade. *Cell*, *83*(6), 835–839. https://doi.org/https://doi.org/10.1016/0092-8674(95)90199-X

Manor, O., Zubair, N., Conomos, M. P., Xu, X., Rohwer, J. E., Krafft, C. E., … Magis, A. T. (2018). A Multi-omic Association Study of Trimethylamine N-Oxide. *Cell Reports*, *24*(4), 935–946. https://doi.org/10.1016/j.celrep.2018.06.096

Mao, S.-Y., Huo, W.-J., & Zhu, W.-Y. (2016). Microbiome–metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. *Environmental Microbiology*, *18*(2), 525–541. https://doi.org/10.1111/1462-2920.12724

Marteau, P., Pochart, P., Doré, J., Béra-Maillet, C., Bernalier, A., & Corthier, G. (2001). Comparative Study of Bacterial Groups within the Human Cecal and Fecal Microbiota. *Applied and Environmental Microbiology*, *67*(10), 4939 LP – 4942. https://doi.org/10.1128/AEM.67.10.4939-4942.2001

Martín-Peláez, S., Gibson, G. R., Martín-Orúe, S. M., Klinder, A., Rastall, R. A., La Ragione, R. M., … Costabile, A. (2008). In vitro fermentation of carbohydrates by porcine faecal inocula and their influence on Salmonella Typhimurium growth in batch culture systems. *FEMS Microbiology Ecology*, *66*(3), 608–619. https://doi.org/10.1111/j.1574-6941.2008.00610.x

Martin, G., Kolida, S., Marchesi, J. R., Want, E., Sidaway, J. E., & Swann, J. R. (2018). In Vitro Modeling of Bile Acid Processing by the Human Fecal Microbiota . *Frontiers in Microbiology* , Vol. 9, p. 1153. Retrieved from https://www.frontiersin.org/article/10.3389/fmicb.2018.01153

Martínez, I., Wallace, G., Zhang, C., Legge, R., Benson, A. K., Carr, T. P., … Walter, J. (2009). Diet-induced metabolic improvements in a hamster model of hypercholesterolemia are strongly linked to alterations of the gut microbiota. *Applied and Environmental Microbiology*, *75*(12), 4175–4184. https://doi.org/10.1128/AEM.00380-09

Martínez, N., Hidalgo-Cantabrana, C., Delgado, S., Margolles, A., & Sánchez, B. (2019). Filling the gap between collection, transport and storage of the human gut microbiota. *Scientific Reports*, *9*(1), 8327. https://doi.org/10.1038/s41598-019-44888-8

Matijašić, B. B., Obermajer, T., Lipoglavšek, L., Grabnar, I., Avguštin, G., & Rogelj, I. (2014). Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. *European Journal of Nutrition*, *53*(4), 1051–1064. https://doi.org/10.1007/s00394-013-0607-6

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., … Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, *6*(3), 610–618. https://doi.org/10.1038/ismej.2011.139

McDonald, J. A. K., Mullish, B. H., Pechlivanis, A., Liu, Z., Brignardello, J., Kao, D., … Marchesi, J. R. (2018). Inhibiting Growth of Clostridioides difficile by Restoring Valerate, Produced by the Intestinal Microbiota. *Gastroenterology*, *155*(5), 1495-1507.e15. https://doi.org/10.1053/j.gastro.2018.07.014

McPherson-Kay, R. (1987). Fiber, stool bulk, and bile acid output: Implications for colon cancer

risk. *Preventive Medicine*, *16*(4), 540–544. https://doi.org/https://doi.org/10.1016/0091-7435(87)90069-7

McRae, M. P. (2017). Dietary Fiber Is Beneficial for the Prevention of Cardiovascular Disease: An Umbrella Review of Meta-analyses. *Journal of Chiropractic Medicine*, *16*(4), 289–299. https://doi.org/10.1016/j.jcm.2017.05.005

Mendis, S., Puska, P., & Norrving, B. (2011). *Global Atlas on Cardiovascular Disease Prevention and Control.*

Michael, H. P., N., P. P., & A., M. F. (2008). Evaluating the Evidence. *Circulation*, *118*(16), 1675–1684. https://doi.org/10.1161/CIRCULATIONAHA.107.721357

Millar, K. A., Gallagher, E., Burke, R., McCarthy, S., & Barry-Ryan, C. (2019). Proximate composition and anti-nutritional factors of fava-bean (Vicia faba), green-pea and yellow-pea (Pisum sativum) flour. *Journal of Food Composition and Analysis*, *82*, 103233. https://doi.org/https://doi.org/10.1016/j.jfca.2019.103233

Mills, E. J., Chan, A.-W., Wu, P., Vail, A., Guyatt, G. H., & Altman, D. G. (2009). Design, analysis, and presentation of crossover trials. *Trials*, *10*, 27. https://doi.org/10.1186/1745-6215-10-27

Mitchell, D. C., Lawrence, F. R., Hartman, T. J., & Curran, J. M. (2009). Consumption of Dry Beans, Peas, and Lentils Could Improve Diet Quality in the US Population. *Journal of the American Dietetic Association*, *109*(5), 909–913. https://doi.org/https://doi.org/10.1016/j.jada.2009.02.029

Molinero, N., Ruiz, L., Sánchez, B., Margolles, A., & Delgado, S. (2019). Intestinal Bacteria Interplay With Bile and Cholesterol Metabolism: Implications on Host Physiology . *Frontiers in Physiology* , Vol. 10, p. 185. Retrieved from https://www.frontiersin.org/article/10.3389/fphys.2019.00185

Momozawa, Y., Deffontaine, V., Louis, E., & Medrano, J. F. (2011). Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. *PloS One*, *6*(2), e16952. https://doi.org/10.1371/journal.pone.0016952

Monte, M. J., Marin, J. J. G., Antelo, A., & Vazquez-Tato, J. (2009). Bile acids: chemistry, physiology, and pathophysiology. *World Journal of Gastroenterology*, *15*(7), 804–816. https://doi.org/10.3748/wjg.15.804

Morales, P., Fujio, S., Navarrete, P., Ugalde, J. A., Magne, F., Carrasco-Pozo, C., …

Gotteland, M. (2016). Impact of Dietary Lipids on Colonic Function and Microbiota: An Experimental Approach Involving Orlistat-Induced Fat Malabsorption in Human Volunteers. *Clinical and Translational Gastroenterology*, *7*(4), e161–e161. https://doi.org/10.1038/ctg.2016.20

Morenga, L. Te, Williams, S., Brown, R., & Mann, J. (2010). Effect of a relatively high-protein, high-fiber diet on body composition and metabolic risk factors in overweight women. *European Journal of Clinical Nutrition*, *64*(11), 1323–1331. https://doi.org/10.1038/ejcn.2010.163

Morotomi, M., Nagai, F., & Watanabe, Y. (2012). Description of Christensenella minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *62*(Pt 1), 144–149. https://doi.org/10.1099/ijs.0.026989-0

Moundras, C., Behr, S. R., Rémésy, C., & Demigné, C. (1997). Fecal losses of sterols and bile acids induced by feeding rats guar gum are due to greater pool size and liver bile acid secretion. *The Journal of Nutrition*, *127*(6), 1068–1076. https://doi.org/10.1093/jn/127.6.1068

Mudryj, A. N., Yu, N., Hartman, T. J., Mitchell, D. C., Lawrence, F. R., & Aukema, H. M. (2012). Pulse consumption in Canadian adults influences nutrient intakes. *British Journal of Nutrition*, *108*(S1), S27–S36. https://doi.org/DOI: 10.1017/S0007114512000724

Mullish, B. H., Pechlivanis, A., Barker, G. F., Thursz, M. R., Marchesi, J. R., & McDonald, J. A. K. (2018). Functional microbiomics: Evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods*, *149*, 49–58. https://doi.org/https://doi.org/10.1016/j.ymeth.2018.04.028

Mysara, M., Vandamme, P., Props, R., Kerckhof, F.-M., Leys, N., Boon, N., … Monsieurs, P. (2017). Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiology Ecology*, *93*(4). https://doi.org/10.1093/femsec/fix029

Nagana Gowda, G. A. (2011). NMR spectroscopy for discovery and quantitation of biomarkers of disease in human bile. *Bioanalysis*, *3*(16), 1877–1890. https://doi.org/10.4155/bio.11.152

Nair, P. P., Gordon, M., & Reback, J. (1967). The Enzymatic Cleavage of the Carbon-Nitrogen Bond in 3α,7α,12α-Trihydroxy-5β-cholan-24-oylglycine. *Journal of Biological Chemistry* , *242*(1), 7–11. Retrieved from http://www.jbc.org/content/242/1/7.abstract

Ndongo, S., Lagier, J.-C., Fournier, P.-E., Raoult, D., & Khelaifia, S. (2016). “Phocea massiliensis” a new bacterial species isolated from the human gut. *New Microbes and New Infections*, *13*, 67–68. https://doi.org/10.1016/j.nmni.2016.06.009

Nestel, P. J., Billington, T., & Smith, B. (1981). Low density and high density lipoprotein kinetics and sterol balance in vegetarians. *Metabolism*, *30*(10), 941–945. https://doi.org/https://doi.org/10.1016/0026-0495(81)90090-1

Nicholson, J. K., & Lindon, J. C. (2008). Metabonomics. *Nature*, *455*(7216), 1054–1056. https://doi.org/10.1038/4551054a

Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., … Andoh, A. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *Journal of Gastroenterology*, *53*(1), 95–106. https://doi.org/10.1007/s00535-017-1384-4

Norman, A. (1964). Faecal excretion products of cholic acid in man. *British Journal of Nutrition*, *18*(1), 173–186. https://doi.org/DOI: 10.1079/BJN19640017

Norman, A., & Palmer, R. H. (1964). Metabolites of lithocholic acid-24-C^{14} in human bile and feces. *The Journal of Laboratory and Clinical Medicine*, *63*(6), 986–1001. https://doi.org/10.5555/uri:pii:0022214364901635

Odermatt, A., Da Cunha, T., Penno, C. A., Chandsawangbhuwana, C., Reichert, C., Wolf, A., … Baker, M. E. (2011). Hepatic reduction of the secondary bile acid 7-oxolithocholic acid is mediated by 11β-hydroxysteroid dehydrogenase 1. *Biochemical Journal*, *436*(3), 621–629. https://doi.org/10.1042/BJ20110022

Oliphant, K., & Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome*, *7*(1), 91. https://doi.org/10.1186/s40168-019-0704-8

Ozato, N., Saito, S., Yamaguchi, T., Katashima, M., Tokuda, I., Sawada, K., … Nakaji, S. (2019). Blautia genus associated with visceral fat accumulation in adults 20–76 years of age. *Npj Biofilms and Microbiomes*, *5*(1), 28. https://doi.org/10.1038/s41522-019-0101-x

Pandak, W. M., & Kakiyama, G. (2019). The acidic pathway of bile acid synthesis: Not just an alternative pathway. *Liver Research*, *3*(2), 88–98. https://doi.org/https://doi.org/10.1016/j.livres.2019.05.001

Parolini, C., Manzini, S., Busnelli, M., Rigamonti, E., Marchesi, M., Diani, E., … Chiesa, G. (2013). Effect of the combinations between pea proteins and soluble fibres on

cholesterolaemia and cholesterol metabolism in rats. *British Journal of Nutrition*, *110*(8), 1394–1401. https://doi.org/DOI: 10.1017/S0007114513000639

Patterson, E., Cryan, J. F., Fitzgerald, G. F., Ross, R. P., Dinan, T. G., & Stanton, C. (2014). Gut microbiota, the pharmabiotics they produce and host health. *The Proceedings of the Nutrition Society*, *73*(4), 477–489. https://doi.org/10.1017/S0029665114001426

Petković, M., Schiller, J., Müller, M., Benard, S., Reichl, S., Arnold, K., & Arnhold, J. (2001). Detection of Individual Phospholipids in Lipid Mixtures by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry: Phosphatidylcholine Prevents the Detection of Further Species. *Analytical Biochemistry*, *289*(2), 202–216. https://doi.org/https://doi.org/10.1006/abio.2000.4926

Petropoulou, A. (2019). *The role of plant-based food structures on gastrointestinal digestion, colonic fermentation and glucose homeostasis*. Imperial College London.

Petrov, V. A., Fernández-Peralbo, M. A., Derks, R., Knyazeva, E. M., Merzlikin, N. V, Sazonov, A. E., … Saltykova, I. V. (2018). Biliary microbiota and bile acids composition in cholelithiasis. *BioRxiv*, 469825. https://doi.org/10.1101/469825

Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical Biochemist. Reviews*, *30*(1), 19–34.

Png, C. W., Linden, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., … Florin, T. H. J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *The American Journal of Gastroenterology*, *105*(11), 2420–2428. https://doi.org/10.1038/ajg.2010.281

Poole, C. F. (2003). *Chapter 9 - Spectroscopic Detectors for Identification and Quantification* (C. F. B. T.-T. E. of C. Poole, Ed.). https://doi.org/https://doi.org/10.1016/B978-044450198-1/50022-7

Posma, J. M., Garcia-Perez, I., Heaton, J. C., Burdisso, P., Mathers, J. C., Draper, J., … Nicholson, J. K. (2017). Integrated Analytical and Statistical Two-Dimensional Spectroscopy Strategy for Metabolite Identification: Application to Dietary Biomarkers. *Analytical Chemistry*, *89*(6), 3300–3309. https://doi.org/10.1021/acs.analchem.6b03324

Prabha, V., & Ohri, M. (2006). Review: Bacterial transformations of bile acids. *World Journal of Microbiology and Biotechnology*, *22*(2), 191–196. https://doi.org/10.1007/s11274-005-9019-y

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., & Levin, E. (2020).

Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLOS ONE*, *15*(1), e0227434. Retrieved from https://doi.org/10.1371/journal.pone.0227434

Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S., & Flint, H. J. (2002). The microbiology of butyrate formation in the human colon. *FEMS Microbiology Letters*, *217*(2), 133–139. https://doi.org/10.1111/j.1574-6968.2002.tb11467.x

Pursley, J., Shen, J. X., Schuster, A., Dang, O. T., Lehman, J., Buonarati, M. H., … Arnold, M. E. (2014). LC–MS/MS determination of apixaban (BMS-562247) and its major metabolite in human plasma: an application of polarity switching and monolithic HPLC column. *Bioanalysis*, *6*(15), 2071–2082. https://doi.org/10.4155/bio.14.66

Qiao, X., Ye, M., Liu, C., Yang, W., Miao, W., Dong, J., & Guo, D. (2012). A tandem mass spectrometric study of bile acids: Interpretation of fragmentation pathways and differentiation of steroid isomers. *Steroids*, *77*(3), 204–211. https://doi.org/https://doi.org/10.1016/j.steroids.2011.11.008

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., … Consortium, M. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Ramezani, A., & Raj, D. S. (2014). The gut microbiome, kidney disease, and targeted interventions. *Journal of the American Society of Nephrology : JASN*, *25*(4), 657–670. https://doi.org/10.1681/ASN.2013080905

Rangel, I., Sundin, J., Fuentes, S., Repsilber, D., de Vos, W. M., & Brummer, R. J. (2015). The relationship between faecal-associated and mucosal-associated microbiota in irritable bowel syndrome patients and healthy subjects. *Alimentary Pharmacology & Therapeutics*, *42*(10), 1211–1221. https://doi.org/10.1111/apt.13399

Ratnayake, W. S., Hoover, R., & Warkentin, T. (2002). Pea Starch: Composition, Structure and Properties — A Review. *Starch - Stärke*, *54*(6), 217–234. https://doi.org/10.1002/1521-379X(200206)54:6<217::AID-STAR217>3.0.CO;2-R

Reddy, N. R. (2017). “Stable Labeled Isotopes as Internal Standards: A Critical Review.” *Modern Applications in Pharmacy & Pharmacology*, *1*(2), 1–4. https://doi.org/10.31031/mapp.2017.01.000508

Reddy, S., Sanders, T. A. B., Owen, R. W., & Thompson, M. H. (1998). Faecal pH, bile acid and sterol concentrations in premenopausal Indian and white vegetarians compared with white omnivores. *British Journal of Nutrition*, *79*(6), 495–500. https://doi.org/DOI: 10.1079/BJN19980087

Remane, D., Wissenbach, D. K., Meyer, M. R., & Maurer, H. H. (2010). Systematic investigation of ion suppression and enhancement effects of fourteen stable-isotope-labeled internal standards by their native analogues using atmospheric-pressure chemical ionization and electrospray ionization and the relevance for multi-anal. *Rapid Communications in Mass Spectrometry*, *24*(7), 859–867. https://doi.org/10.1002/rcm.4459

Rideout, T. C., Harding, S. V, Jones, P. J., & Fan, M. Z. (2008). Guar gum and similar soluble fibers in the regulation of cholesterol metabolism: current understandings and future research priorities. *Vascular Health and Risk Management*, *4*(5), 1023–1033. https://doi.org/10.2147/vhrm.s3512

Rideout, T. C., Yuan, Z., Bakovic, M., Liu, Q., Li, R.-K., Mine, Y., & Fan, M. Z. (2007). Guar gum consumption increases hepatic nuclear SREBP2 and LDL receptor expression in pigs fed an atherogenic diet. *The Journal of Nutrition*, *137*(3), 568–572. https://doi.org/10.1093/jn/137.3.568

Ridlon, J. M., Alves, J. M., Hylemon, P. B., & Bajaj, J. S. (2013). Cirrhosis, bile acids and gut microbiota: unraveling a complex relationship. *Gut Microbes*, *4*(5), 382–387. https://doi.org/10.4161/gmic.25723

Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J., & Hylemon, P. B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes*, *7*(1), 22–39. https://doi.org/10.1080/19490976.2015.1127483

Ridlon, J. M., Kang, D.-J., & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research* , *47*(2), 241–259. https://doi.org/10.1194/jlr.R500013-JLR200

Ridlon, J. M., Kang, D. J., Hylemon, P. B., & Bajaj, J. S. (2014). Bile acids and the gut microbiome. *Current Opinion in Gastroenterology*, *30*(3), 332–338. https://doi.org/10.1097/MOG.0000000000000057

Ringel, Y., Maharshak, N., Ringel-Kulka, T., Wolber, E. A., Sartor, R. B., & Carroll, I. M. (2015). High throughput sequencing reveals distinct microbial populations within the mucosal and luminal niches in healthy individuals. *Gut Microbes*, *6*(3), 173–181.

https://doi.org/10.1080/19490976.2015.1044711

Roager, H. M., Hansen, L. B. S., Bahl, M. I., Frandsen, H. L., Carvalho, V., Gøbel, R. J., … Licht, T. R. (2016). Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nature Microbiology*, *1*(9), 16093. https://doi.org/10.1038/nmicrobiol.2016.93

Roberts, L. D., McCombie, G., Titman, C. M., & Griffin, J. L. (2008). A matter of fat: An introduction to lipidomic profiling methods. *Journal of Chromatography B*, *871*(2), 174–181. https://doi.org/https://doi.org/10.1016/j.jchromb.2008.04.002

Rockwood, A. L., Kushnir, M. M., & Clarke, N. J. (2018). *2 - Mass Spectrometry* (N. Rifai, A. R. Horvath, & C. T. B. T.-P. and A. of C. M. S. Wittwer, Eds.). https://doi.org/https://doi.org/10.1016/B978-0-12-816063-3.00002-5

Rose, G., & Shipley, M. (1986). Plasma cholesterol concentration and death from coronary heart disease: 10 year results of the Whitehall study. *British Medical Journal (Clinical Research Ed.)*, *293*(6542), 306–307. https://doi.org/10.1136/bmj.293.6542.306

Rosen, M. J., Callahan, B. J., Fisher, D. S., & Holmes, S. P. (2012). Denoising PCR-amplified metagenome data. *BMC Bioinformatics*, *13*(1), 283. https://doi.org/10.1186/1471-2105-13-283

Russell, D. W. (2003). The Enzymes, Regulation, and Genetics of Bile Acid Synthesis. *Annual Review of Biochemistry*, *72*(1), 137–174. https://doi.org/10.1146/annurev.biochem.72.121801.161712

Sabatine, M. S., Giugliano, R. P., Keech, A. C., Honarpour, N., Wiviott, S. D., Murphy, S. A., … Pedersen, T. R. (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *The New England Journal of Medicine*, *376*(18), 1713–1722. https://doi.org/10.1056/NEJMoa1615664

Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A., & Hendriks, M. M. W. B. (2014). Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*, *10*(3), 361–374. https://doi.org/10.1007/s11306-013-0598-6

Sampson, U. K., Fazio, S., & Linton, M. F. (2012). Residual cardiovascular risk despite optimal LDL cholesterol reduction with statins: the evidence, etiology, and therapeutic challenges. *Current Atherosclerosis Reports*, *14*(1), 1–10. https://doi.org/10.1007/s11883-011-0219-7

Sandström, B., Hansen, L. T., & Sørensen, A. (1994). Pea fiber lowers fasting and

postprandial blood triglyceride concentrations in humans. *The Journal of Nutrition*, *124*(12), 2386–2396. https://doi.org/10.1093/jn/124.12.386

Sangster, T., Major, H., Plumb, R., Wilson, A. J., & Wilson, I. D. (2006). A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst*, *131*(10), 1075–1078. https://doi.org/10.1039/B604498K

Sarafian, M. H., Lewis, M. R., Pechlivanis, A., Ralphs, S., McPhail, M. J. W., Patel, V. C., … Nicholson, J. K. (2015). Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry. *Analytical Chemistry*, *87*(19), 9662–9670. https://doi.org/10.1021/acs.analchem.5b01556

Sauter, G., Berr, F., Beuers, U., Fischer, S., & Paumgartner, G. (1996). Serum concentrations of 7alpha-hydroxy-4-cholesten-3-one reflect bile acid synthesis in humans. *Hepatology (Baltimore, Md.)*, *24*(1), 123–126. https://doi.org/10.1053/jhep.1996.v24.pm0008707250

Schwartz, G. G., Steg, P. G., Szarek, M., Bhatt, D. L., Bittner, V. A., Diaz, R., … Zeiher, A. M. (2018). Alirocumab and Cardiovascular Outcomes after Acute Coronary Syndrome. *The New England Journal of Medicine*, *379*(22), 2097–2107. https://doi.org/10.1056/NEJMoa1801174

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, *14*(8), e1002533. Retrieved from https://doi.org/10.1371/journal.pbio.1002533

Setchell, K. D., Lawson, A. M., Tanida, N., & Sjövall, J. (1983). General methods for the analysis of metabolic profiles of bile acids and related compounds in feces. *Journal of Lipid Research* , *24*(8), 1085–1100. Retrieved from http://www.jlr.org/content/24/8/1085.abstract

Setchell, K. D. R., Ives, J. A., Cashmore, G. C., & Lawson, A. M. (1987). On the homogeneity of stools with respect to bile acid composition and normal day-to-day variations: a detailed qualitative and quantitative study using capillary column gas chromatography-mass spectrometry. *Clinica Chimica Acta*, *162*(3), 257–275. https://doi.org/https://doi.org/10.1016/0009-8981(87)90045-3

Shiffka, S. J., Kane, M. A., & Swaan, P. W. (2017). Planar bile acids in health and disease. *Biochimica et Biophysica Acta. Biomembranes*, *1859*(11), 2269–2276. https://doi.org/10.1016/j.bbamem.2017.08.019

Shin, J.-H., Sim, M., Lee, J.-Y., & Shin, D.-M. (2016). Lifestyle and geographic insights into the distinct gut microbiota in elderly women from two different geographic locations.

Journal of Physiological Anthropology, *35*(1), 31. https://doi.org/10.1186/s40101-016-0121-7

Silverman, M. G., Ference, B. A., Im, K., Wiviott, S. D., Giugliano, R. P., Grundy, S. M., … Sabatine, M. S. (2016). Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-analysis. *JAMA*, *316*(12), 1289–1297. https://doi.org/10.1001/jama.2016.13985

Silverman, R. B., & Holladay, M. W. (2014). *Chapter 8 - Drug Metabolism* (R. B. Silverman & M. W. B. T.-T. O. C. of D. D. and D. A. (Third E. Holladay, Eds.). https://doi.org/https://doi.org/10.1016/B978-0-12-382030-3.00008-8

Sirtori, C. R., Triolo, M., Bosisio, R., Bondioli, A., Calabresi, L., De Vergori, V., … Arnoldi, A. (2012). Hypocholesterolaemic effects of lupin protein and pea protein/fibre combinations in moderately hypercholesterolaemic individuals. *British Journal of Nutrition*, *107*(8), 1176–1183. https://doi.org/DOI: 10.1017/S0007114511004120

Slavin, J. (2013). Fiber and prebiotics: mechanisms and health benefits. *Nutrients*, *5*(4), 1417–1435. https://doi.org/10.3390/nu5041417

Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., … Li, J. (2019). Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome*, *7*(1), 9. https://doi.org/10.1186/s40168-019-0628-3

Spielmann, J., Stangl, G. I., & Eder, K. (2008). Dietary pea protein stimulates bile acid excretion and lowers hepatic cholesterol concentration in rats. *Journal of Animal Physiology and Animal Nutrition*, *92*(6), 683–693. https://doi.org/10.1111/j.1439-0396.2007.00766.x

Stellwag, E. J., & Hylemon, P. B. (1976). Purification and characterization of bile salt hydrolase from Bacteroides fragilis subsp. fragilis. *Biochimica et Biophysica Acta*, *452*(1), 165–176. https://doi.org/10.1016/0005-2744(76)90068-1

Stellwag, E. J., & Hylemon, P. B. (1979). 7alpha-Dehydroxylation of cholic acid and chenodeoxycholic acid by Clostridium leptum. *Journal of Lipid Research* , *20*(3), 325–333. Retrieved from http://www.jlr.org/content/20/3/325.abstract

Steuer, R., Morgenthal, K., Weckwerth, W., & Selbig, J. (2007). *A Gentle Guide to the Analysis of Metabolomic Data BT - Metabolomics: Methods and Protocols* (W. Weckwerth, Ed.). https://doi.org/10.1007/978-1-59745-244-1_7

Story, J. A., & Kritchevsky, D. (1978). Bile acid metabolism and fiber. *The American Journal*

of Clinical Nutrition, *31*(10 Suppl), S199–S202. https://doi.org/10.1093/ajcn/31.10.S199

Studer, E., Zhou, X., Zhao, R., Wang, Y., Takabe, K., Nagahashi, M., … Hylemon, P. B. (2012). Conjugated bile acids activate the sphingosine-1-phosphate receptor 2 in primary rodent hepatocytes. *Hepatology (Baltimore, Md.)*, *55*(1), 267–276. https://doi.org/10.1002/hep.24681

Sudhop, T., Lütjohann, D., Kodal, A., Igel, M., Tribble, D. L., Shah, S., … von Bergmann, K. (2002). Inhibition of intestinal cholesterol absorption by ezetimibe in humans. *Circulation*, *106*(15), 1943–1948. https://doi.org/10.1161/01.cir.0000034044.95911.dc

Sunner, J., Dratz, E., & Chen, Y. C. (1995). Graphite surface-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides and proteins from liquid solutions. *Analytical Chemistry*, *67*(23), 4335–4342. https://doi.org/10.1021/ac00119a021

Swann, J. R., Want, E. J., Geier, F. M., Spagou, K., Wilson, I. D., Sidaway, J. E., … Holmes, E. (2011). Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proceedings of the National Academy of Sciences*, *108*(Supplement 1), 4523 LP – 4530. https://doi.org/10.1073/pnas.1006734107

Świątecka, D., Kostyra, H., & Świątecki, A. (2010). Impact of glycated pea proteins on the activity of free-swimming and immobilised bacteria. *Journal of the Science of Food and Agriculture*, *90*(11), 1837–1845. https://doi.org/10.1002/jsfa.4022

Swidsinski, A., Loening–Baucke, V., Verstraelen, H., Osowska, S., & Doerffel, Y. (2008). Biostructure of Fecal Microbiota in Healthy Subjects and Patients With Chronic Idiopathic Diarrhea. *Gastroenterology*, *135*(2), 568-579.e2. https://doi.org/https://doi.org/10.1053/j.gastro.2008.04.017

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., … Sugimoto, M. (2016). Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion*, *93*(1), 59–65. https://doi.org/10.1159/000441768

Takahashi, S., Fukami, T., Masuo, Y., Brocker, C. N., Xie, C., Krausz, K. W., … Gonzalez, F. J. (2016). Cyp2c70 is responsible for the species difference in bile acid metabolism between mice and humans. *Journal of Lipid Research*, *57*(12), 2130–2137. https://doi.org/10.1194/jlr.M071183

Takamine, F., & Imamura, T. (1995). Isolation and Characterization of Bile Acid 7-Dehydroxylating Bacteria from Human Feces. *Microbiology and Immunology*, *39*(1), 11–

18. https://doi.org/10.1111/j.1348-0421.1995.tb02162.x

Takáts, Z., Wiseman, J. M., Gologan, B., & Cooks, R. G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science (New York, N.Y.)*, *306*(5695), 471–473. https://doi.org/10.1126/science.1104404

Talati, R., Baker, W. L., Pabilonia, M. S., White, C. M., & Coleman, C. I. (2009). The effects of barley-derived soluble fiber on serum lipids. *Annals of Family Medicine*, *7*(2), 157–163. https://doi.org/10.1370/afm.917

Tan, A., & Awaiye, K. (2013, September 16). Use of Internal Standards in LC-MS Bioanalysis. *Handbook of LC-MS Bioanalysis*, pp. 217–227. https://doi.org/doi:10.1002/9781118671276.ch17

Tennent, D. M., Siegel, H., Zanetti, M. E., Kuron, G. W., Ott, W. H., & Wolf, F. J. (1960). Plasma cholesterol lowering action of bile acid binding polymers in experimental animals. *Journal of Lipid Research* , *1*(5), 469–473. Retrieved from http://www.jlr.org/content/1/5/469.abstract

Theriot, C. M., Koenigsknecht, M. J., Carlson, P. E. J., Hatton, G. E., Nelson, A. M., Li, B., … Young, V. B. (2014). Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature Communications*, *5*, 3114. https://doi.org/10.1038/ncomms4114

Theuwissen, E., & Mensink, R. P. (2008). Water-soluble dietary fibers and cardiovascular disease. *Physiology & Behavior*, *94*(2), 285–292. https://doi.org/10.1016/j.physbeh.2008.01.001

Thiese, M. S. (2014). Observational and interventional study design types; an overview. *Biochemia Medica*, *24*(2), 199–210. https://doi.org/10.11613/BM.2014.022

Thomas, L. A., Veysey, M. J., French, G., Hylemon, P. B., Murphy, G. M., & Dowling, R. H. (2001). Bile acid metabolism by fresh human colonic contents: a comparison of caecal versus faecal samples. *Gut*, *49*(6), 835 LP – 842. https://doi.org/10.1136/gut.49.6.835

Thompson, H. J. (2019). Dietary Bean Consumption and Human Health. *Nutrients*, *11*(12), 3074. https://doi.org/10.3390/nu11123074

Threapleton, D. E., Greenwood, D. C., Evans, C. E. L., Cleghorn, C. L., Nykjaer, C., Woodhead, C., … Burley, V. J. (2013). Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ (Clinical Research Ed.)*, *347*, f6879. https://doi.org/10.1136/bmj.f6879

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *The Biochemical Journal*, *474*(11), 1823–1836. https://doi.org/10.1042/BCJ20160510

Torsoli, A., Ramorino, M. L., Ammaturo, M. V, Capurso, L., Paoluzi, P., & Anzini, F. (1971). Mass movements and intracolonic pressures. *The American Journal of Digestive Diseases*, *16*(8), 693–696. https://doi.org/10.1007/BF02239591

Tosh, S. M., & Yada, S. (2010). Dietary fibres in pulse seeds and fractions: Characterization, functional attributes, and applications. *Food Research International*, *43*(2), 450–460. https://doi.org/https://doi.org/10.1016/j.foodres.2009.09.005

Trefflich, I., Marschall, H.-U., Giuseppe, D. R., Ståhlman, M., Michalsen, A., Lampen, A., … Weikert, C. (2019). Associations between Dietary Patterns and Bile Acids—Results from a Cross-Sectional Study in Vegans and Omnivores. *Nutrients* , Vol. 12. https://doi.org/10.3390/nu12010047

Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, *489*(7415), 242–249. https://doi.org/10.1038/nature11552

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., … Marsland, B. J. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Medicine*, *20*(2), 159–166. https://doi.org/10.1038/nm.3444

Turley, S. D., Daggy, B. P., & Dietschy, J. M. (1991). Cholesterol-lowering action of psyllium mucilloid in the hamster: sites and possible mechanisms of action. *Metabolism: Clinical and Experimental*, *40*(10), 1063–1073. https://doi.org/10.1016/0026-0495(91)90131-f

Turnbaugh, P. J., Bäckhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe*, *3*(4), 213–223. https://doi.org/https://doi.org/10.1016/j.chom.2008.02.015

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine*, *1*(6), 6ra14 LP-6ra14. https://doi.org/10.1126/scitranslmed.3000322

US Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and B. I. (2003). Your Guide to Lowering Blood Pressure With DASH. Retrieved February 24, 2020, from https://www.nhlbi.nih.gov/files/docs/public/heart/hbp_low.pdf

US Food and Drug Association. (2003). Food labeling: health claims; soluble dietary fiber from certain foods and coronary heart disease. Final rule. *Federal Register*, *68*(144), 44207–44209.

US Food and Drug Association. (2006). Food labeling: health claims; soluble dietary fiber from certain foods and coronary heart disease. Final rule. *Federal Register*, *71*(98), 29248–29250.

US Food and Drug Association. (2008). Food labeling: health claims; soluble fiber from certain foods and risk of coronary heart disease. Interim final rule. *Federal Register*, *73*(37), 9938–9947.

Van den Abbeele, P., Van de Wiele, T., Verstraete, W., & Possemiers, S. (2011). The host selects mucosal and luminal associations of coevolved gut microorganisms: a novel concept. *FEMS Microbiology Reviews*, *35*(4), 681–704. https://doi.org/10.1111/j.1574-6976.2011.00270.x

van Faassen, A., Bol, J., van Dokkum, W., Pikaar, N. A., Ockhuizen, T., & Hermus, R. J. (1987). Bile acids, neutral steroids, and bacteria in feces as affected by a mixed, a lacto-ovovegetarian, and a vegan diet. *The American Journal of Clinical Nutrition*, *46*(6), 962–967. https://doi.org/10.1093/ajcn/46.6.962

van Faassen, A., Hazen, M. J., van den Brandt, P. A., van den Bogaard, A. E., Hermus, R. J., & Janknegt, R. A. (1993). Bile acids and pH values in total feces and in fecal water from habitually omnivorous and vegetarian subjects. *The American Journal of Clinical Nutrition*, *58*(6), 917–922. https://doi.org/10.1093/ajcn/58.6.917

Velichko, N. V, & Pinevich, A. V. (2019). Classification and Identification Tasks in Microbiology: Mass Spectrometric Methods Coming to the Aid. *Microbiology*, *88*(5), 534–547. https://doi.org/10.1134/S0026261719050151

Viguiliouk, E., Glenn, A. J., Nishi, S. K., Chiavaroli, L., Seider, M., Khan, T., … Sievenpiper, J. L. (2019). Associations between Dietary Pulses Alone or with Other Legumes and Cardiometabolic Disease Outcomes: An Umbrella Review and Updated Systematic Review and Meta-analysis of Prospective Cohort Studies. *Advances in Nutrition (Bethesda, Md.)*, *10*(Suppl_4), S308–S319. https://doi.org/10.1093/advances/nmz113

Vital, M., Howe, A. C., & Tiedje, J. M. (2014). Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data. *MBio*, *5*(2), e00889-14. https://doi.org/10.1128/mBio.00889-14

Vital, M., Rud, T., Rath, S., Pieper, D. H., & Schlüter, D. (2019). Diversity of Bacteria Exhibiting

Bile Acid-inducible 7α-dehydroxylation Genes in the Human Gut. *Computational and Structural Biotechnology Journal*, *17*, 1016–1019. https://doi.org/10.1016/j.csbj.2019.07.012

Wade, W. G., Downes, J., Dymock, D., Hiom, S. J., Weightman, A. J., Dewhirst, F. E., … Coleman, B. (1999). The family Coriobacteriaceae: reclassification of Eubacterium exiguum (Poco et al. 1996) and Peptostreptococcus heliotrinreducens (Lanigan 1976) as Slackia exigua gen. nov., comb. nov. and Slackia heliotrinireducens gen. nov., comb. nov., and Eubacterium . *International Journal of Systematic Bacteriology*, *49 Pt 2*, 595–600. https://doi.org/10.1099/00207713-49-2-595

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., … Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ismej.2010.118

Wang, F., Zheng, J., Yang, B., Jiang, J., Fu, Y., & Li, D. (2015). Effects of Vegetarian Diets on Blood Lipids: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Journal of the American Heart Association*, *4*(10), e002408. https://doi.org/10.1161/JAHA.115.002408

Wang, N., Hatcher, D. W., & Gawalko, E. J. (2008). Effect of variety and processing on nutrients and certain anti-nutrients in field peas (Pisum sativum). *Food Chemistry*, *111*(1), 132–138. https://doi.org/https://doi.org/10.1016/j.foodchem.2008.03.047

Wang, Qing, Wang, K., Wu, W., Giannoulatou, E., Ho, J. W. K., & Li, L. (2019). Host and microbiome multi-omics integration: applications and methodologies. *Biophysical Reviews*, *11*(1), 55–65. https://doi.org/10.1007/s12551-018-0491-7

Wang, Qiong, Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261 LP – 5267. https://doi.org/10.1128/AEM.00062-07

Waterhous, D. V, Barnes, S., & Muccio, D. D. (1985). Nuclear magnetic resonance spectroscopy of bile acids. Development of two-dimensional NMR methods for the elucidation of proton resonance assignments for five common hydroxylated bile acids, and their parent bile acid, 5 beta-cholanoic acid. *Journal of Lipid Research* , *26*(9), 1068–1078. Retrieved from http://www.jlr.org/content/26/9/1068.abstract

Waters, J. L., & Ley, R. E. (2019). The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biology*, *17*(1), 83.

https://doi.org/10.1186/s12915-019-0699-4

Watson, E.-J., Giles, J., Scherer, B. L., & Blatchford, P. (2019). Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure. *Scientific Reports*, *9*(1), 16831. https://doi.org/10.1038/s41598-019-53183-5

Wells, J. E., Berr, F., Thomas, L. A., Dowling, R. H., & Hylemon, P. B. (2000). Isolation and characterization of cholic acid 7α-dehydroxylating fecal bacteria from cholesterol gallstone patients. *Journal of Hepatology*, *32*(1), 4–10. https://doi.org/https://doi.org/10.1016/S0168-8278(00)80183-X

Wells, J. E., & Hylemon, P. B. (2000). Identification and Characterization of a Bile Acid 7α-Dehydroxylation Operon in Clostridium sp. Strain TO-931, a Highly Active 7α-Dehydroxylating Strain Isolated from Human Feces. *Applied and Environmental Microbiology*, *66*(3), 1107 LP – 1113. https://doi.org/10.1128/AEM.66.3.1107-1113.2000

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews*, *20*(4), 593–621. https://doi.org/10.1128/CMR.00008-07

Whiley, L., Nye, L. C., Grant, I., Andreas, N., Chappell, K. E., Sarafian, M. H., … Wilson, I. D. (2019). Ultrahigh-Performance Liquid Chromatography Tandem Mass Spectrometry with Electrospray Ionization Quantification of Tryptophan Metabolites and Markers of Gut Health in Serum and Plasma—Application to Clinical and Epidemiology Cohorts. *Analytical Chemistry*, *91*(8), 5207–5216. https://doi.org/10.1021/acs.analchem.8b05884

Whitehead, A., Beck, E. J., Tosh, S., & Wolever, T. M. S. (2014). Cholesterol-lowering effects of oat β-glucan: a meta-analysis of randomized controlled trials. *The American Journal of Clinical Nutrition*, *100*(6), 1413–1421. https://doi.org/10.3945/ajcn.114.086108

Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2016). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research*, *25*(3), 1057–1073. https://doi.org/10.1177/0962280215588241

WHO. (2018). World Health Organisation. The Top 10 Causes of Death. Retrieved April 10, 2020, from https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death

Wiegel, J., Tanner, R., & Rainey, F. A. (2006). *An Introduction to the Family Clostridiaceae BT - The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria* (M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt, Eds.).

https://doi.org/10.1007/0-387-30744-3_20

Wieling, J. (2002). LC-MS-MS experiences with internal standards. *Chromatographia*, *55*(1), S107–S113. https://doi.org/10.1007/BF02493365

Williams, C. F., Walton, G. E., Jiang, L., Plummer, S., Garaiova, I., & Gibson, G. R. (2015). Comparative Analysis of Intestinal Tract Models. *Annual Review of Food Science and Technology*, *6*(1), 329–350. https://doi.org/10.1146/annurev-food-022814-015429

Williamson, L. N., & Bartlett, M. G. (2007). Quantitative liquid chromatography/time-of-flight mass spectrometry. *Biomedical Chromatography : BMC*, *21*(6), 567–576. https://doi.org/10.1002/bmc.844

Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., … Engstrand, L. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology*, *139*(6), 1844-1854.e1. https://doi.org/10.1053/j.gastro.2010.08.049

Wilson, F. A., Burckhardt, G., Murer, H., Rumrich, G., & Ullrich, K. J. (1981). Sodium-coupled taurocholate transport in the proximal convolution of the rat kidney in vivo and in vitro. *The Journal of Clinical Investigation*, *67*(4), 1141–1150. https://doi.org/10.1172/jci110128

Winston, J. A., & Theriot, C. M. (2020). Diversification of host bile acids by members of the gut microbiota. *Gut Microbes*, *11*(2), 158–171. https://doi.org/10.1080/19490976.2019.1674124

Wong, S. F., Meng, C. K., & Fenn, J. B. (1988). Multiple charging in electrospray ionization of poly(ethylene glycols). *The Journal of Physical Chemistry*, *92*(2), 546–550. https://doi.org/10.1021/j100313a058

Woting, A., Pfeiffer, N., Loh, G., Klaus, S., & Blaut, M. (2014). Clostridium ramosum promotes high-fat diet-induced obesity in gnotobiotic mouse models. *MBio*, *5*(5), e01530-14. https://doi.org/10.1128/mBio.01530-14

Wu, W.-K., Chen, C.-C., Panyod, S., Chen, R.-A., Wu, M.-S., Sheen, L.-Y., & Chang, S.-C. (2019). Optimization of fecal sample processing for microbiome study — The journey from bathroom to bench. *Journal of the Formosan Medical Association*, *118*(2), 545–555. https://doi.org/https://doi.org/10.1016/j.jfma.2018.02.005

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, *17*(1), 135. https://doi.org/10.1186/s12859-016-0992-y

Yao, C. K., Muir, J. G., & Gibson, P. R. (2016). Review article: insights into colonic protein fermentation, its modulation and potential health implications. *Alimentary Pharmacology & Therapeutics*, *43*(2), 181–196. https://doi.org/10.1111/apt.13456

Yao, L., Seaton, S. C., Ndousse-Fetter, S., Adhikari, A. A., DiBenedetto, N., Mina, A. I., … Devlin, A. S. (2018). A selective gut bacterial bile salt hydrolase alters host metabolism. *ELife*, *7*, e37182. https://doi.org/10.7554/eLife.37182

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., … Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635–645. https://doi.org/10.1038/nrmicro3330

Yetley, E. A., MacFarlane, A. J., Greene-Finestone, L. S., Garza, C., Ard, J. D., Atkinson, S. A., … Wells, G. A. (2017). Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. *The American Journal of Clinical Nutrition*, *105*(1), 249S-285S. https://doi.org/10.3945/ajcn.116.139097

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., … Glöckner, F. O. (2013). The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), D643–D648. https://doi.org/10.1093/nar/gkt1209

Yoshiki, A., Wataru, Y., Akio, M., Toshikazu, H., & Kazuya, H. (1985). Reduction of 3-keto-5β-cholanoic acid to lithocholic and isolithocholic acids by human liver cytosol in vitro. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, *837*(1), 20–26. https://doi.org/https://doi.org/10.1016/0005-2760(85)90081-5

Yoshimoto, S., Loo, T. M., Atarashi, K., Kanda, H., Sato, S., Oyadomari, S., … Ohtani, N. (2013). Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature*, *499*(7456), 97–101. https://doi.org/10.1038/nature12347

Yu, E., Malik, V. S., & Hu, F. B. (2018). Cardiovascular Disease Prevention by Diet Modification: JACC Health Promotion Series. *Journal of the American College of Cardiology*, *72*(8), 914–926. https://doi.org/10.1016/j.jacc.2018.02.085

Zakham, F., Pillonel, T., Brunel, A.-S., Zambelli, P.-Y., Greub, G., Croxatto, A., & Bertelli, C. (2019). Molecular diagnosis and enrichment culture identified a septic pseudoarthrosis due to an infection with Erysipelatoclostridium ramosum. *International Journal of Infectious Diseases*, *81*, 167–169.

https://doi.org/https://doi.org/10.1016/j.ijid.2019.02.001

Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., … Zhao, L. (2010). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *The ISME Journal*, *4*(2), 232–241. https://doi.org/10.1038/ismej.2009.112

Zhang, J.-X., Hallmans, G., Adlercreutz, H., Åman, P., Westerland, E., Lundin, E., & Stenling, R. (1993). Effects of oat rye fractions on biliary and faecal bile acid profiles in Syrian golden hamsters (Mesocricetus auratus). *British Journal of Nutrition*, *70*(2), 525–536. https://doi.org/DOI: 10.1079/BJN19930146

Zhao, L. (2013, September). The gut microbiota and obesity: from correlation to causality. *Nature Reviews. Microbiology*, Vol. 11, pp. 639–647. https://doi.org/10.1038/nrmicro3089

Zheng, H., Liang, H., Wang, Y., Miao, M., Shi, T., Yang, F., … Li, D.-K. (2016). Altered Gut Microbiota Composition Associated with Eczema in Infants. *PloS One*, *11*(11), e0166026. https://doi.org/10.1371/journal.pone.0166026

Zhou, H., & Hylemon, P. B. (2014). Bile acids are nutrient signaling hormones. *Steroids*, *86*, 62–68. https://doi.org/https://doi.org/10.1016/j.steroids.2014.04.016

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. L., & de Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Applied and Environmental Microbiology*, *68*(7), 3401–3407. https://doi.org/10.1128/aem.68.7.3401-3407.2002

Zollner, G., Marschall, H.-U., Wagner, M., & Trauner, M. (2006). Role of Nuclear Receptors in the Adaptive Response to Bile Acids and Cholestasis: Pathogenetic and Therapeutic Considerations. *Molecular Pharmaceutics*, *3*(3), 231–251. https://doi.org/10.1021/mp060010s

10.APPENDICES

(Appendices are titled according to Chapter number; some chapters may not have supplementary information).

Appendix 1

Permissions obtained for reusing figures printed in chapter 1:

- Permission obtained for reusing Figure 1-5 to be attached after page 288.

Appendix 2

Permissions obtained for reusing figures printed in chapter 2:

- Permission for Figure 2-3*:

Copyright

Copyright © 2012 Shibdas Banerjee and Shyamalava Mazumdar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Obtained from: https://www.hindawi.com/journals/ijac/2012/282574/#copyright

- Permissions obtained for reusing figures 2-1, 2-4 and 2-6 to be attached after page 288.

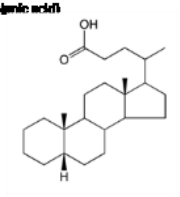
Appendix 3

Supplementary information for Chapter 3

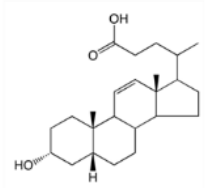
Table showing intra-day accuracy and precision obtained from partial validation of UPLC-MS based method

BA ID	NAME	FINAL ACCURACY					FINAL PRECISION					R2	LLOQ	ULOQ
		LLOQ	LQC	MQC	HQC	ULOQQC	LLOQ	LQC	MQC	HQC	ULOQQC			
BA01	Ursocholanic Acid	104.97	96.58	114.00	108.89	106.59	7.29	1.04	2.03	2.21	4.78	0.9911	1	500
BA03	5-Cholenic Acid-3β-ol	98.94	103.02	106.44	110.40	110.55	16.94	4.82	1.13	1.15	1.31	0.9903	0.5	1000
BA02	Lithocholenic Acid	98.60	98.66	92.58	91.10	87.93	11.89	5.78	4.76	5.75	2.43	0.9904	1	2500
BA04	3-Ketocholanic Acid	101.79	97.11	98.28	97.37	103.23	6.94	3.01	1.62	2.93	1.88	0.9978	0.5	1000
BA07	Allolithocholic Acid	102.79	100.50	98.05	94.79	89.07	12.48	2.31	0.83	1.52	3.81	0.9920	0.5	1000
BA06	Isolithocholic Acid	94.81	96.74	93.15	95.28	96.48	4.19	2.07	1.70	0.96	1.12	0.9919	2.5	1000
BA05	Lithocholic acid	117.68	109.96	103.43	93.53	91.55	4.60	3.99	0.93	1.67	2.14	0.9907	0.5	2500
BA09	3α,12α, 23-Nordeoxycholic Acid	109.62	106.34	100.61	97.67	112.83	7.68	4.18	1.96	1.95	2.13	0.9924	0.5	5000
BA10	9(11), (5β)-Cholenic Acid-3α-ol-12-one	108.67	96.37	88.78	86.98	100.73	1.68	6.01	1.38	1.73	1.93	0.9984	0.5	5000
BA11	5α-Cholanic Acid-3, 6-dione	99.05	93.72	96.08	89.30	94.09	12.86	10.04	7.48	7.84	7.01	0.9943	0.5	5000
BA12	3,7-Diketocholanic Acid	98.55	95.67	103.03	87.79	86.82	7.43	3.73	4.07	5.57	5.16	0.9918	0.5	5000
BA15	8(14),(5β)-Cholenic Acid-3α, 12α-diol	111.98	92.50	86.49	88.48	92.08	4.80	8.50	3.79	3.48	4.22	0.9955	0.5	5000
BA16	5β-Cholenic Acid-7α-ol-3-one	103.60	98.82	104.77	102.46	106.99	2.91	3.22	2.85	2.02	3.25	0.9981	1	2500
BA17	5α-Cholanic Acid-3α-ol-6-one	95.87	88.64	91.09	90.42	103.70	14.24	2.72	7.24	5.85	3.30	0.9916	0.5	5000
BA18	3α-Hydroxy-7 Ketolithocholic Acid	116.88	93.97	89.34	86.77	100.94	6.15	5.82	2.33	2.08	2.30	0.9980	0.5	5000
BA19	3α-Hydroxy-12 Ketolithocholic Acid	94.96	89.22	109.85	90.52	91.89	14.97	3.90	1.02	0.33	1.27	0.9906	0.5	5000
BA20	Murocholic Acid	91.98	87.95	84.68	85.23	88.76	3.55	7.84	1.35	8.26	4.08	0.9922	0.5	2500
BA32	Ursodeoxycholic acid	111.73	85.60	106.74	89.55	91.26	8.21	10.07	0.90	0.35	1.33	0.9912	0.5	5000
BA25	Hyodeoxycholic acid	115.97	102.98	94.91	92.52	107.48	8.07	5.03	7.32	3.27	3.64	0.9919	0.5	2500
BA22	5β-Cholanic Acid-3β, 12α-diol	104.78	92.25	87.74	84.80	97.13	3.69	0.77	1.77	1.88	1.89	0.9978	0.5	2500
BA24	Chenodeoxycholic Acid	116.32	95.55	109.04	90.46	91.60	12.67	6.59	0.83	0.77	1.32	0.9912	2.5	5000
BA23	Deoxycholic Acid	105.61	87.74	111.19	93.39	94.36	11.97	13.07	1.22	0.78	0.86	0.9950	5	2500
BA21	Isodeoxycholic Acid	105.37	92.57	99.31	105.48	106.83	6.34	2.75	1.34	1.58	1.30	0.9913	1	1000
BA26	3,7,12 Dehydrocholic acid	84.15	88.47	109.99	91.96	94.14	4.17	4.18	1.94	0.34	1.48	0.9905	0.5	5000
BA27	3α-Hydroxy-7,12-DiketoCholanic Acid	108.61	112.68	103.50	95.92	106.72	6.12	2.16	2.96	0.89	4.87	0.9902	0.5	5000
BA28	3α-Hydroxy-6,7-DiketoCholanic Acid	87.34	88.78	106.33	87.58	88.80	8.17	3.62	1.11	0.32	1.40	0.9903	0.5	2500
BA29	5β-Cholanic Acid-3α, 6α-diol-7-one	116.28	104.17	107.37	98.10	108.87	4.77	0.98	2.48	2.38	4.25	0.9906	0.5	5000
BA31	12 Dehydrocholic Acid	98.05	93.10	102.06	91.89	101.78	11.80	12.81	6.10	5.29	7.09	0.9901	2.5	5000
BA30	3 Dehydrocholic Acid	94.34	90.22	103.81	86.25	88.05	13.50	7.00	0.96	0.53	1.41	0.9907	0.5	5000
BA33	ω Muricholic Acid	120.84	106.19	93.39	84.54	93.46	2.12	4.34	3.42	2.17	4.50	0.9947	0.5	2500
BA37	α Muricholic Acid	117.35	102.70	92.06	90.86	93.85	3.40	2.42	3.66	3.18	4.69	0.9939	0.5	2500
BA34	β Muricholic Acid	116.18	98.74	89.21	88.39	91.25	5.10	4.33	3.78	3.14	3.67	0.9922	0.5	2500
BA36	Hyocholic acid	113.00	109.92	110.74	99.80	109.43	2.00	5.18	3.38	2.84	4.49	0.9913	0.5	5000
BA35	Cholic acid	93.99	96.72	103.98	87.17	89.65	9.91	5.38	0.98	0.46	1.22	0.9911	0.5	2500
BA38	Glyco-ursocholanic Acid	91.13	91.27	103.01	104.72	105.38	8.38	2.72	1.40	0.70	0.63	0.9903	2.5	2500
BA39	Glycolithocholic Acid	100.49	95.46	102.38	87.58	89.38	7.93	3.33	0.90	0.42	1.19	0.9933	1	5000
BA41	Glycochenodeoxycholic Acid	111.42	86.78	104.44	89.29	92.20	13.78	8.35	1.22	0.75	1.71	0.9911	1	5000
BA42	Glycodeoxycholic acid	101.76	90.51	103.66	89.01	91.45	10.19	8.33	1.41	0.35	1.20	0.9911	0.5	5000
BA43	Glycoursodeoxycholic Acid	116.35	104.88	90.70	85.59	98.28	5.58	1.86	1.95	0.83	0.68	0.9993	0.5	5000
BA40	Glycohyodeoxycholic Acid	111.24	105.13	112.68	96.49	97.28	4.31	3.13	1.75	0.45	1.25	0.9912	0.5	5000
BA44	3,7,12 Glycodehydrocholic acid	92.29	91.87	101.03	87.21	90.11	5.77	1.86	0.82	0.42	1.54	0.9940	0.5	5000
BA49	Glycohyocholic Acid	94.51	94.26	102.47	86.17	87.88	10.02	3.28	1.06	0.38	1.37	0.9920	1	2500
BA46	Glycocholic Acid	93.51	86.29	90.83	94.77	96.17	7.95	3.88	1.53	0.72	0.99	0.9925	0.5	5000
BA47	Tauro-ursocholanic Acid	108.40	99.82	110.27	87.96	86.39	7.05	2.33	1.25	0.63	1.45	0.9908	1	2500
BA48	Taurolithocholic Acid	106.20	98.14	111.52	93.82	96.72	5.77	1.62	1.11	0.14	1.63	0.9921	5	5000
BA50	Tauro-ursodeoxycholic Acid	107.86	105.63	112.60	94.60	96.65	9.73	2.25	1.03	0.57	1.46	0.9930	2.5	2500
BA51	Taurohyodeoxycholic Acid	110.13	103.95	114.42	94.95	96.99	10.61	2.40	1.26	0.18	1.24	0.9910	2.5	2500
BA52	Taurochenodeoxycholic Acid			109.23	95.10	98.83			3.02	1.19	1.77	0.9932	5	5000
BA53	Taurodeoxycholic Acid	104.25	94.15	106.30	92.29	95.36	8.91	4.65	0.78	0.33	1.43	0.9929	0.5	5000
BA54	3,7,12 Taurodehydrocholic acid	100.60	91.32	107.03	91.81	93.93	9.35	4.51	1.01	0.27	1.13	0.9932	0.5	5000
BA58	Tauro ω-Muricholic Acid sodium salt	109.57	107.93	113.89	89.63	87.90	8.38	6.64	2.14	3.22	4.27	0.9930	0.5	5000
BA57	Tauro-α Muricholic Acid sodium salt	106.02	104.64	101.87	85.54	87.44	3.80	1.99	1.12	1.48	1.56	0.9927	2.5	2500
BA55	Taurohyocholic Acid Sodium salt	90.39	83.58	87.51	91.75	92.43	4.27	2.52	1.46	0.93	1.20	0.9918	0.5	2500
BA59	Taurocholic Acid	100.32	92.46	104.02	91.67	95.00	7.47	4.45	1.08	0.27	1.55	0.9930	0.5	2500

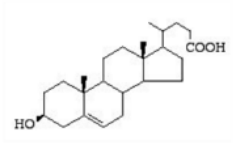
BA01 Unocholanic acid
(5β-Cholanic acid)



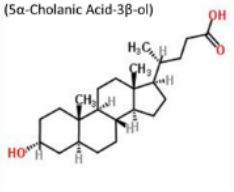
BA02 Litocholanic Acid
(3(β)-Cholanic acid-3α-ol)



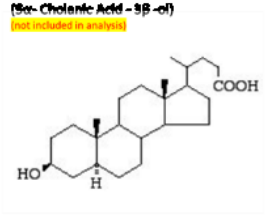
BA03 5-Cholanic acid-3β-ol



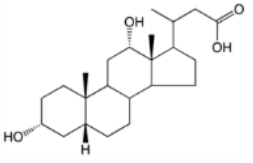
BA07 Allolithocholic acid
(5α-Cholanic Acid-3β-ol)



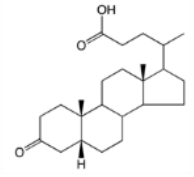
BA08 Isoallolithocholic acid
(5α- Cholanic Acid - 3β -ol)
(not included in analysis)



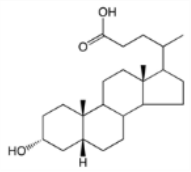
BA09 3α,12α, 23-Nordeoxycholic Acid



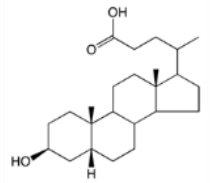
BA04 3-Ketocholanic Acid



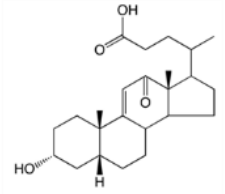
BA06 Lithocholic acid
(5β-cholanic acid-3α-ol)



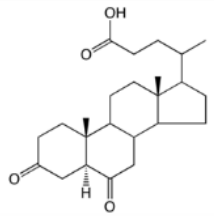
BA08 Isolithocholic acid
(5β-Cholanic Acid-3β-ol)



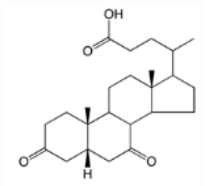
BA10 9(11), (5β)-Cholenic Acid-3α-ol-12-one



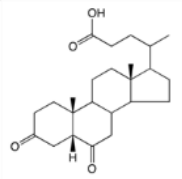
BA11 5α-Cholanic Acid-3, 6-dione



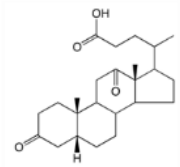
BA12 3,7-Diketocholanic Acid
(5β-Cholanic Acid-3,7-dione)



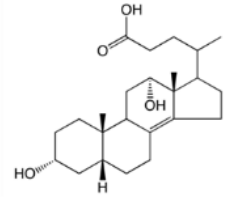
BA13 3,6-Diketocholanic Acid
(5β-Cholanic Acid-3,6-dione)



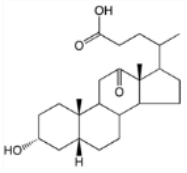
BA14 3,12-Diketocholanic Acid



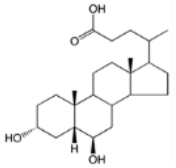
BA15 8(14),(5β)-Cholenic Acid-3α, 12α-diol



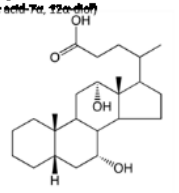
BA16 3α-Hydroxy-12 Ketolithocholic Acid
(5β-Cholanic Acid-3α-ol-12-one)



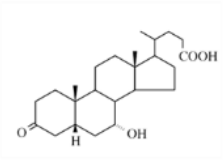
BA20 Murocholic acid
(5β-cholanic acid-3α, 6β-diol)



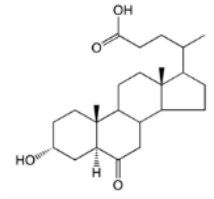
BA21 Isodeoxycholic acid
(5β-cholanic acid-7α, 12α-diol)



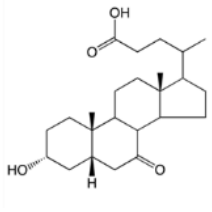
BA18 5β-Cholanic Acid-7α-ol-3-one



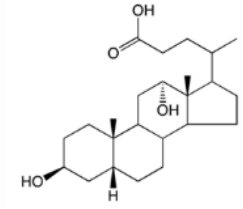
BA17 5α-Cholanic Acid-3α-ol-6-one



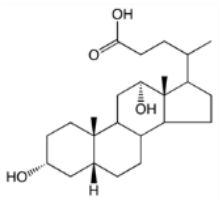
BA19 3α-Hydroxy-7 Ketolithocholic Acid
(5β-Cholanic Acid-3α-ol-7-one)



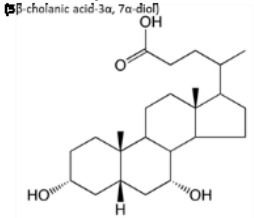
BA22 5β-Cholanic Acid-3β, 12α-diol



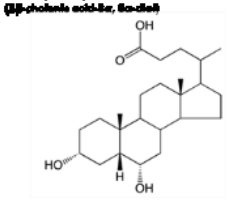
BA23 Deoxycholic acid
(5β-cholanic acid-3α, 12α-diol)



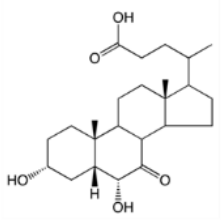
BA24 Chenodeoxycholic acid
(5β-cholanic acid-3α, 7α-diol)



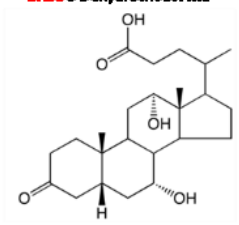
BA28 Hyodeoxycholic acid (3β-cholanic acid-6α, 6α-diol)



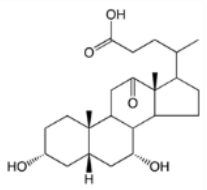
BA29 5β-Cholanic Acid-3α, 6α-diol-7-one



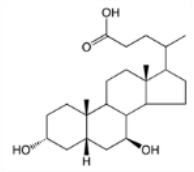
BA30 3 Dehydrocholic Acid



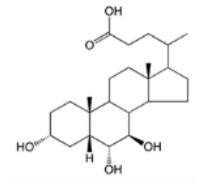
BA31 12 Dehydrocholic Acid



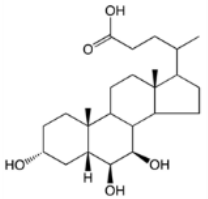
BA32 Ursodeoxycholic acid



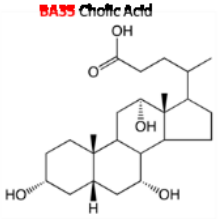
BA33 Ω-MURICHOLIC ACID



BA34 β Muricholic Acid



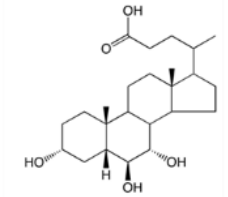
BA35 Cholic Acid



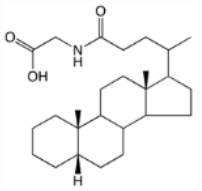
BA36 Hyocholic Acid



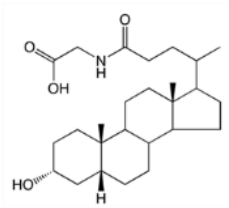
BA37 α Muricholic Acid



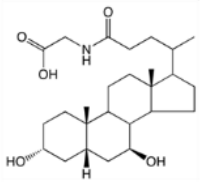
BA38 Glycoursocholanic acid



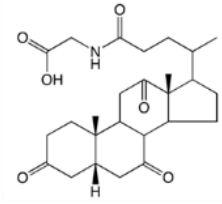
BA39 Glycolithocholic acid



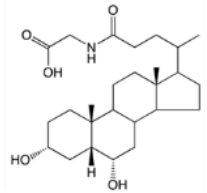
BA43 Glycoursodeoxycholic Acid



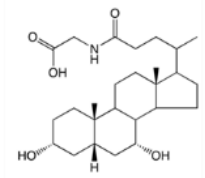
BA44 3,7,12-Glycodehydrocholic acid



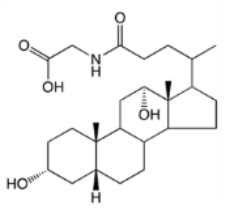
BA40 Glycohyodeoxycholic Acid



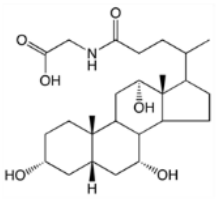
BA41 Glycochenodeoxycholic Acid



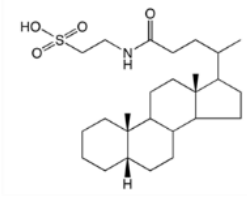
BA42 Glycodeoxycholic Acid



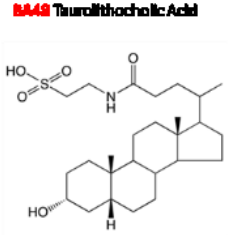
BA46 Glycocholic acid



BA47 Taurursocholanic acid



BA48 Taurolithocholic Acid



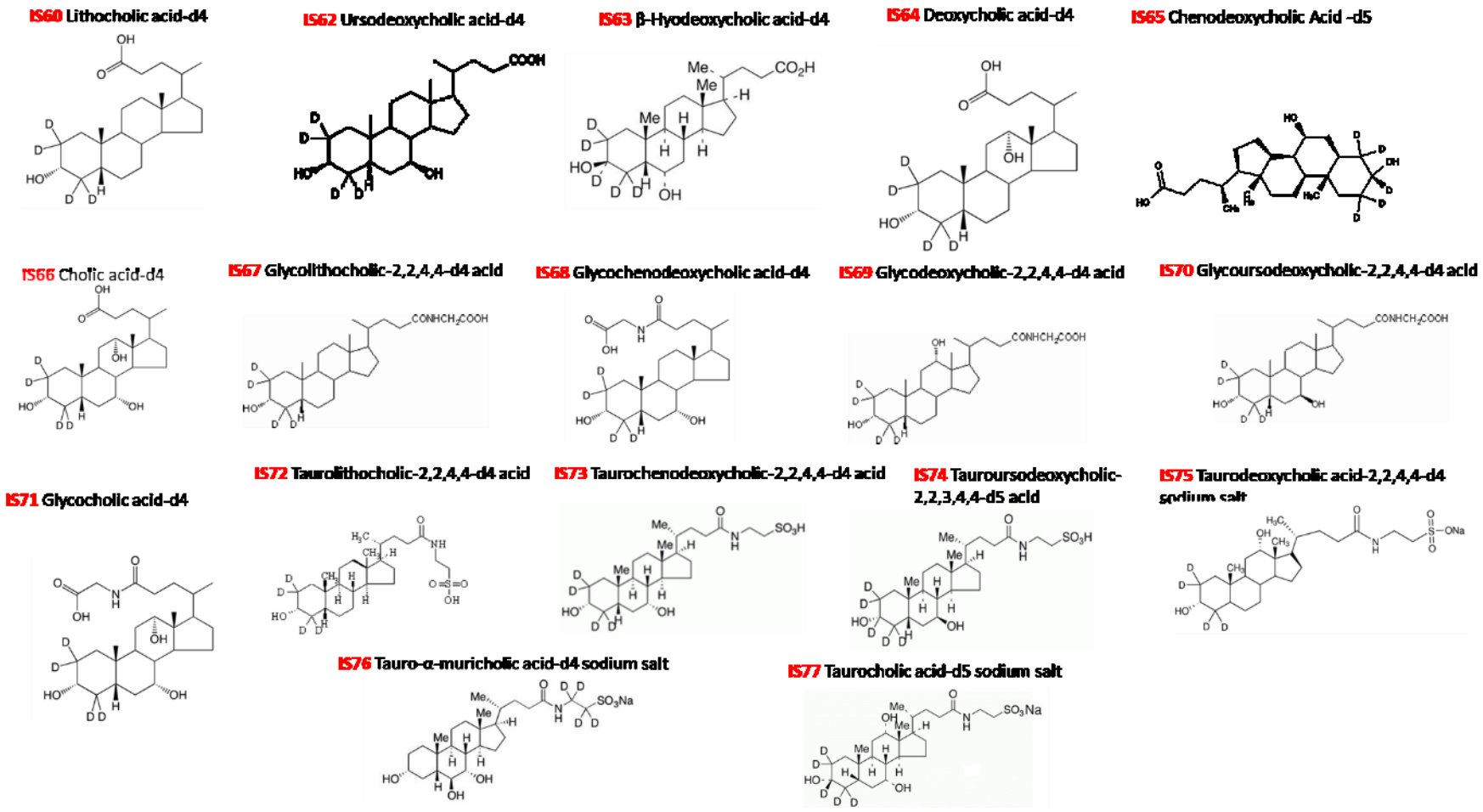


Figure S3-1: Molecular structures of bile acids in ultra (high) performance liquid chromatography mass spectrometry method (UPLC-MS) method described in Chapter 3.

Appendix 4

No supplementary information for Chapter 4

Appendix 5

Vegan Study Methods For Sample Collection And Data Processing Obtained From Collaborators

Study design and sample collection

[Text provided by Prof Tue Hansen]

Participants and examination

150 healthy volunteers, 75 vegans and 75 omnivores, were recruited from urban areas in Denmark by advertisements in print, online and social media from November 2013 to November 2014. Inclusion criteria were as follows:

- 18 – 65 years old
- weight-stable (±1 kg, assessed by interview) for a minimum of 2 months.
- In addition, vegan volunteers were eligible for inclusion in the study if they had been adherent to a vegan diet for a minimum of 1 year.

Exclusion criteria were as follows: Received antibiotic treatment within 3 months, had known gastrointestinal disease or reported gastrointestinal symptoms at the time of the study, or followed a medically prescribed diet, pregnant and lactating women.

The study was approved by the Ethical Committees of the Capital Region of Denmark (H-3-2012-145) and conducted in accordance with the Helsinki declaration. Informed consent was obtained from all volunteers prior to inclusion.

Study participants were examined in the morning following a 10-hour overnight fast with sampling of blood and second void urine; both of which were immediately stored at -80°C until analysis. Participants were weighed on an electronic scale (TANITA WB-110MA, Tanita Corporation of America, Arlington Heights, Illinois, USA) without shoes, dressed in light indoor clothing or underwear after having emptied their bladder. Height was measured to the nearest 0.5 cm without shoes, using a wall-mounted stadiometer (ADE MZ10023, ADE, Hamburg, Germany). Body-mass index (BMI) was calculated as weight in kg divided by the square of height in meters. Body composition was assessed using a bioelectrical impedance analyser (BIA310e, Biodynamics Corp., Seattle, Washington, USA). Blood pressure was recorded as

the mean of triplicate measurements on the non-dominant arm, in reclined position following a 5-minute rest.

Dietary intakes

Information on habitual diet was acquired using a 4-day weighed food record, including two working days and two weekend days within one week. Foods were quantified to the nearest 0.1 g using a calibrated precision scale (ProScale XC-2000, HBI Europe, Erkelenz, Germany). Volunteers were instructed in filling in the record by qualified medical staff. Nutrient composition of the recorded diets were assessed using the Dankost Pro software (version 1.5.49.21), which is based on the food database at the Danish Food Composition Databank containing 1,049 food items (www.frida.fooddata.dk). Vegan recipes not included in the database were constructed based on foods with complete validity in the database. Average daily intake (ADI) of macro- and micronutrients was calculated as,

$$ADI = ((\text{average on working days} \times 4) + (\text{average on weekend days} \times 3)) / 7$$

The vegan subjects were asked to bring their dietary supplements on the day of examination at which point every supplemented nutrient was noted with exact daily dose.

Faecal sample collection

Faecal samples were collected on the last day of food recording or the day thereafter following standardized procedures, including home sampling with immediate freezing at –18°C and transfer on dry ice for final storage at –80°C within 48 h. Sarstedt tube x 2 collected by the participants, part specimen. Imperial got 1 of the 2 tubes, non-aliquoted, non-homogenized.

Lipoprotein analyses

Plasma triglyceride (TG), total cholesterol (TC), and high-density lipoprotein (HDL) were analysed on a Vitros 5600 system (CV 14.6%, 11.6%, and 17.0%, respectively). Very-low-density lipoprotein (VLDL) was calculated as VLDL=0.45×TG. Low-density lipoprotein was calculated as LDL=TC−HDL−VLDL. Non-HDL was calculated as Non-HDL=TC−HDL.

Bile acids analyses

Urine and serum bile acids were analysed according to Sarafian et al 2015 (exisitng lab method that was then optimised for PDC quantification).

Metagenomic analyses

[Text provided by Ms Florence Thirion]

DNA extraction, library preparation and sequencing for shotgun metagenomics

Faecal samples from 75 vegans and 75 omnivores were selected for shotgun metagenomics. From each sample a frozen aliquot of 200 mg of faeces was suspended in 250 µL of guanidine thiocyanate, 0.1 M tris(hydroxymethyl)aminomethane (pH 7.5), and 40 µL of 10% *N*-lauroyl sarcosine, after which DNA was extracted as previously described The concentration and integrity of extracted DNA were estimated by spectrophotometry (NanoDrop, Thermo Fisher Scientific Inc., MA USA) and agarose gel electrophoresis, respectively. DNA library preparation followed standardized steps, shearing of DNA, quantification of fragmented DNA, end repairing, ligation of adapters containing one of 96 unique barcodes for each sample processed, size selection, amplification by PCR (8 cycles), purification, and quantification. Quantification libraries ensures the amount of DNA obtained (4000 pM desired, sample rejected below 2000 pM) and the size of the libraries. Libraries were sequenced on an IonProton platform (Thermo Fisher Scientific Inc., MA USA). Libraries were loaded on the sequencers until a minimum of 20M high quality reads were generated, resulting in a median 20,940,000 (Q_1-Q_3, 20,380,000-22,000,000) reads per sample.

Gene count table

To create the gene count table, METEOR software was used: first, reads were filtered for low-quality using AlienTrimmer (with params '-k 10 -l 45 -m 5 -p 40 -q 20'). Reads that aligned Human genome (GRCh38 assembly version) (identity > 95%) were also discarded using Bowtie2. Remaining reads were mapped onto the Integrated Gut Catalogue (IGC) comprising 9.9 million of genes, with Bowtie2 First, the unique mapped reads (reads mapped to a unique gene in the catalogue) were attributed to their corresponding genes. Second, the shared reads (reads that mapped with the same alignment score to multiple genes in the catalogue) were attributed according to the ratio of their unique mapping counts of the captured genes. The resulting count table was further processed using the MetaOMineR R package. It was downsized at 12 million mapped reads to take into account differences in sequencing depth across samples. Then downsized matrix was normalized for gene length and transformed into a frequency matrix (FPKM normalization).

Metagenomic Species (MGS) profiles

The IGC was organized into 1776 MetaGenomic Species (MGS) using MSPminer. Briefly, 1262 samples were used to cluster co-abundant genes from catalogue into MGS (clusters of

at least 100 genes). Taxonomical annotation of MGS was performed using an in-house pipeline. First, all genes were aligned on public databases (ncbi) using Blast. An MGS was annotated with the lowest taxonomical rank (from species to super-kingdom) that brought consensus in at least 50% of its genes. To avoid misleading annotations due to error in databases, for each gene the 20 first hits were being considered. Relative abundance of an MGS was computed as the mean abundance of its 50 'marker' genes (that is, the genes that correlate the most altogether in terms of abundance through the 1262 samples described above). If less than 10% of 'marker' genes were seen in a sample, the abundance of the MGS was set to 0. Relative abundances at higher taxonomical ranks were computed as the sum of the MGS that belong to a given taxa. MGS count was assessed as the number of MGS present in a sample (that is, whose abundance is strictly positive).

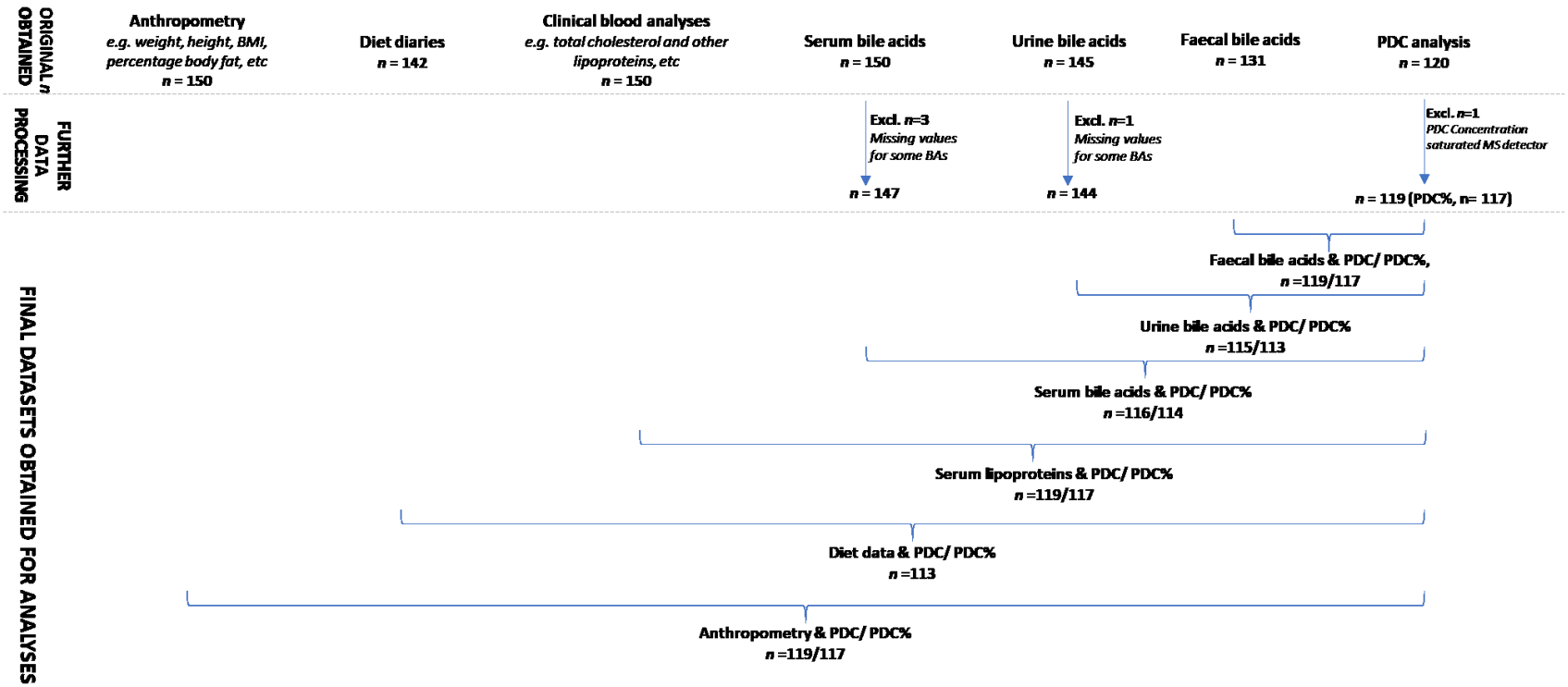
Microbiota functions

Genes from the IGC catalogue were mapped with diamond onto KEGG Orthologous (KO) from KEGG database (version 8.2) . Each gene was assigned to the best-ranked KO among hits with e-value < 10^{e}-05 and bit score > 60. Then we assessed presence of KEGG modules and Gut-Metabolic Modules (GMM) in a pair sample-MGS. For the module to be present in the pair, we checked (i) if the MGS was present in the sample and (ii) if at least 90% KOs from the module were present in the intersection of sample genes and MGS genes. Finally, to assess module potential in a sample, we summed up abundances of all MGS that were found to carry this module for this sample.

Statistical analysis

All statistical analyses were performed with R and Rstudio. Contrasts in quantitative variables were performed using Mann-Whitney test if 2 groups, Kruskal-Wallis if more than 2 groups. Correlations between variables (either metagenomic variables or clinical variables) were performed using Spearman's correlations.

Supplementary Figures and Results for Chapter 5



*No of participants with all datasets, *n* = 106

Figure S5-1: Data treatment workflow. BA concentrations from n = 3 serum samples were recoded as missing because one entry had a total serum BA concentration equal to zero, and two entries only recorded BA concentrations for three BA which were in consecutive order. Similarly, n=1 urine sample was recoded as missing because concentrations were recorded for only three consecutive BAs. Final numbers of participants with samples matched with PDC data are given. Final number of participants with all data was 106.

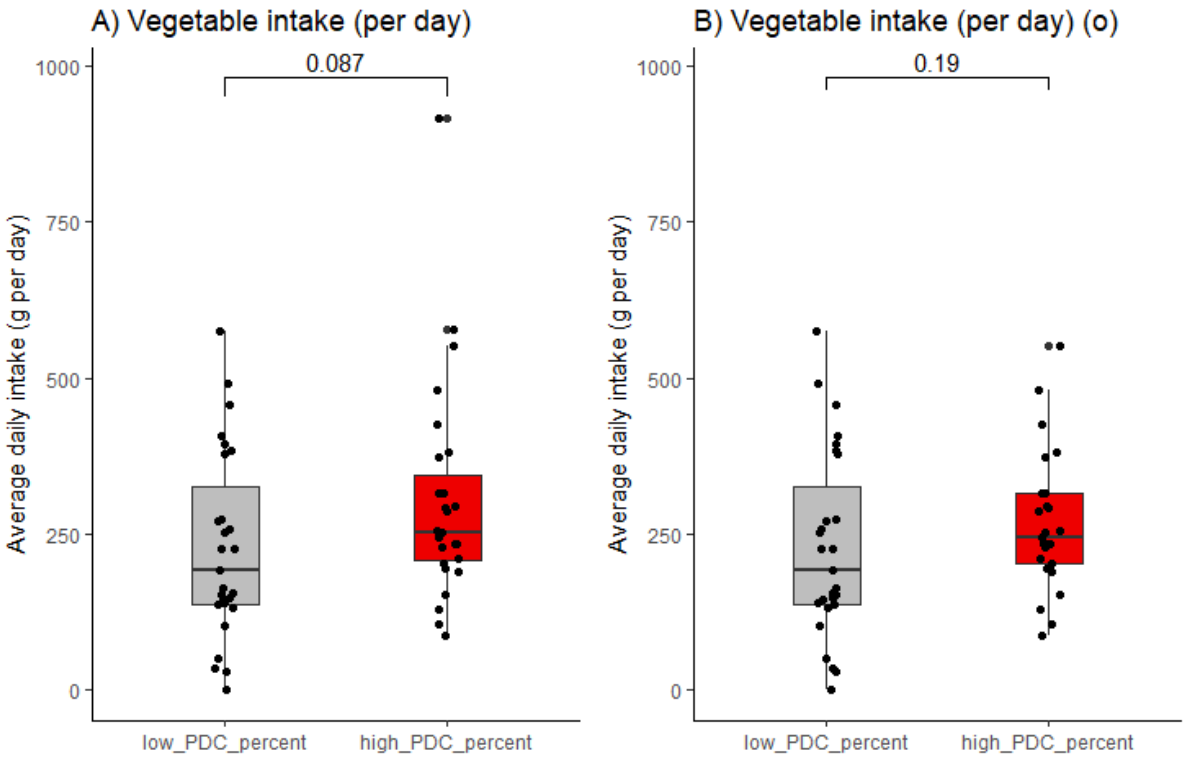


Figure S5-2: Boxplots showing average daily vegetable intakes A)with and B)without outliers in low-%PDC and high-%PDC groups. There was a tendency to significance observed which was mainly driven by outliers in the high PDC% group. No significant difference was observed between both groups (Mann Whitney U test) with and without the outliers removed.

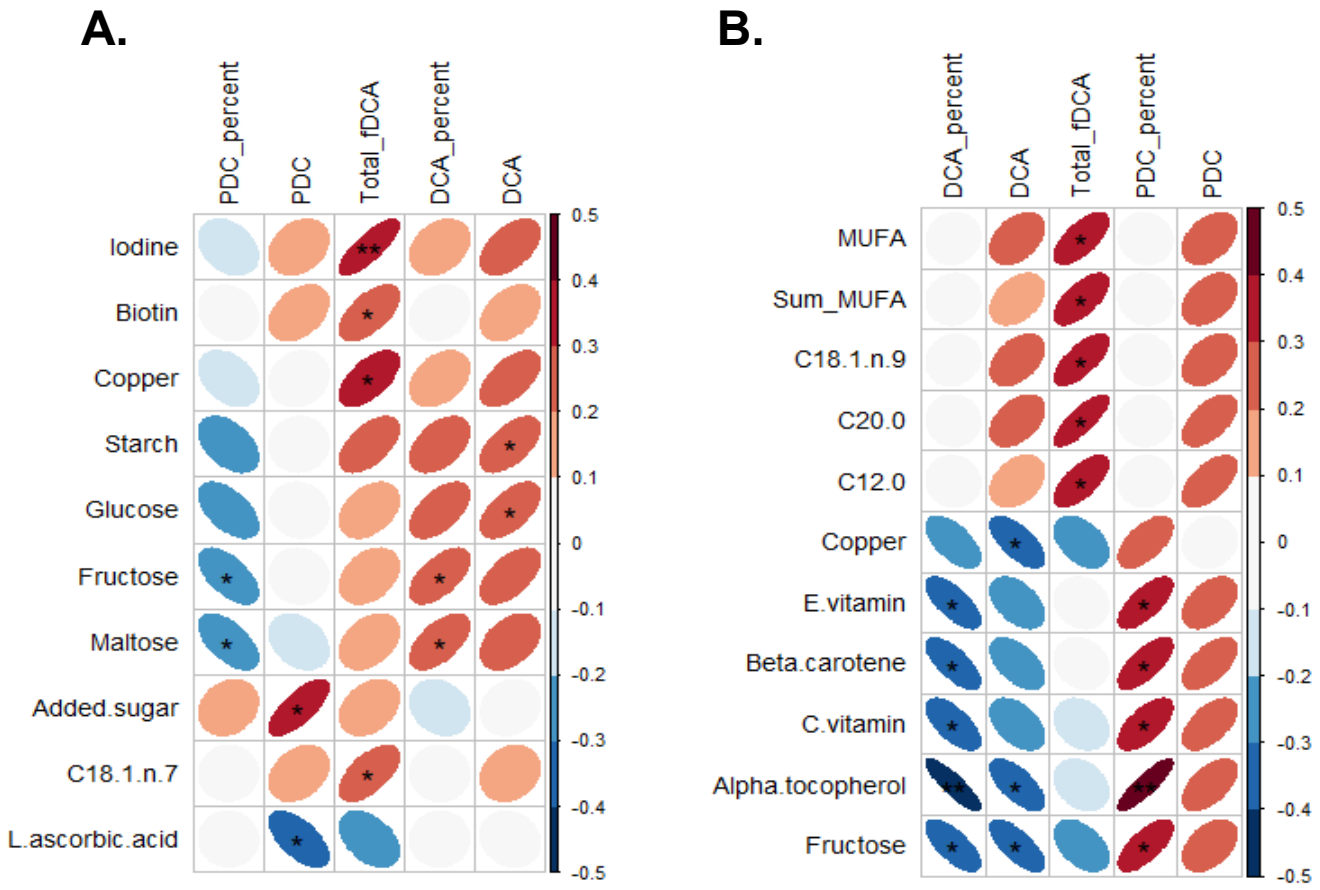


Figure S5-3: Correlation heatmap showing only statistically significant Spearman correlations of an absolute value ≥0.1 between PDC, DCA, total faecal DCA, PDC% and DCA% and intakes of “minor” nutrients in A)vegans only and B)omnivores only. Signicance levels 0.001, 0.01 and 0.05 are labelled as ***, ** and * respectively.

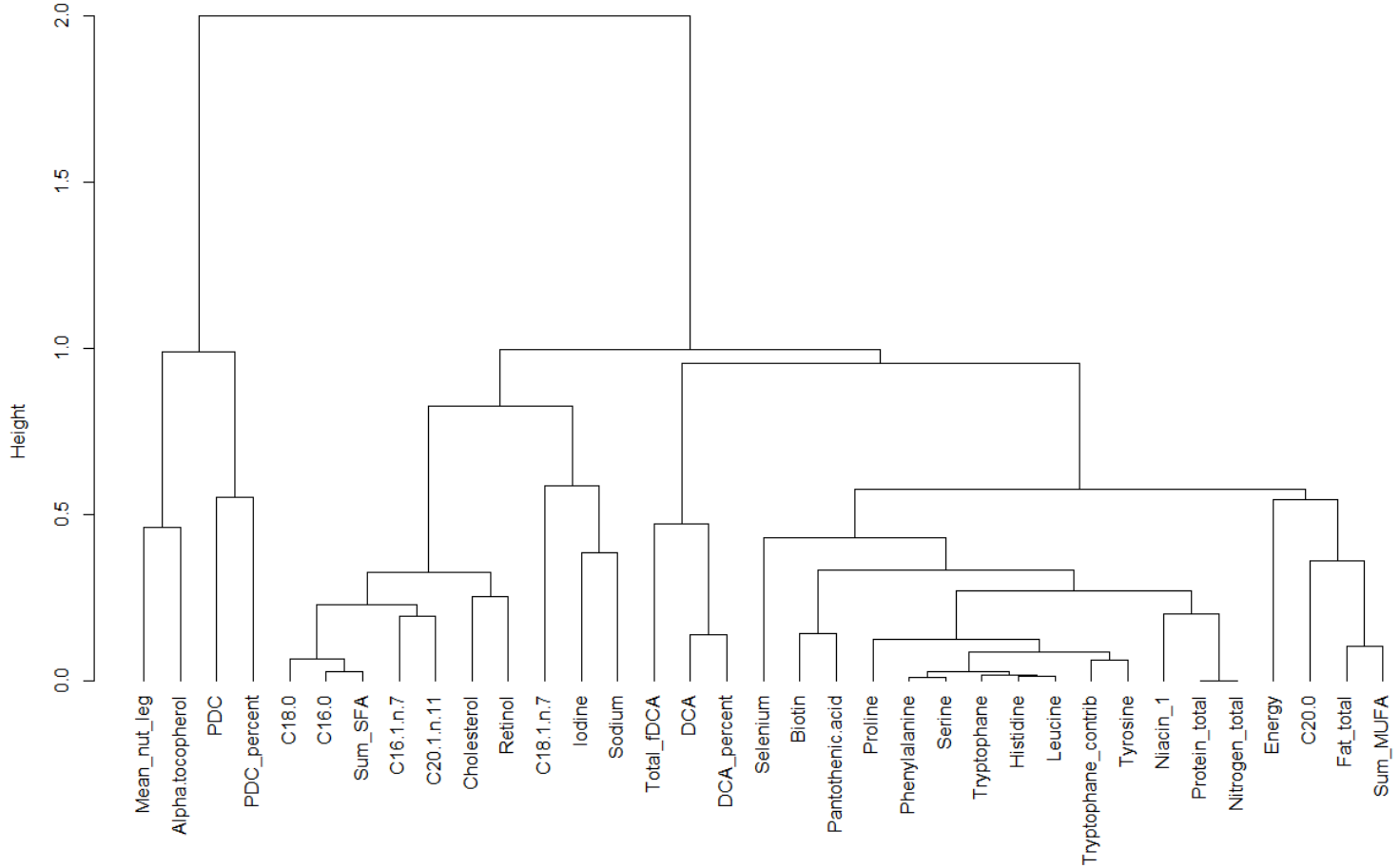


Figure S5-4: Similarity dendrogram showing hierarchical clustering pattern of PDC measures and diet variables. Significant Spearman correlations were clustered by complete-linkage method (most similar variables are clustered together). PDC and PDC% are observed to distinctly cluster with intakes of nut and legume foods and the micronutrient alpha-tocopherol.

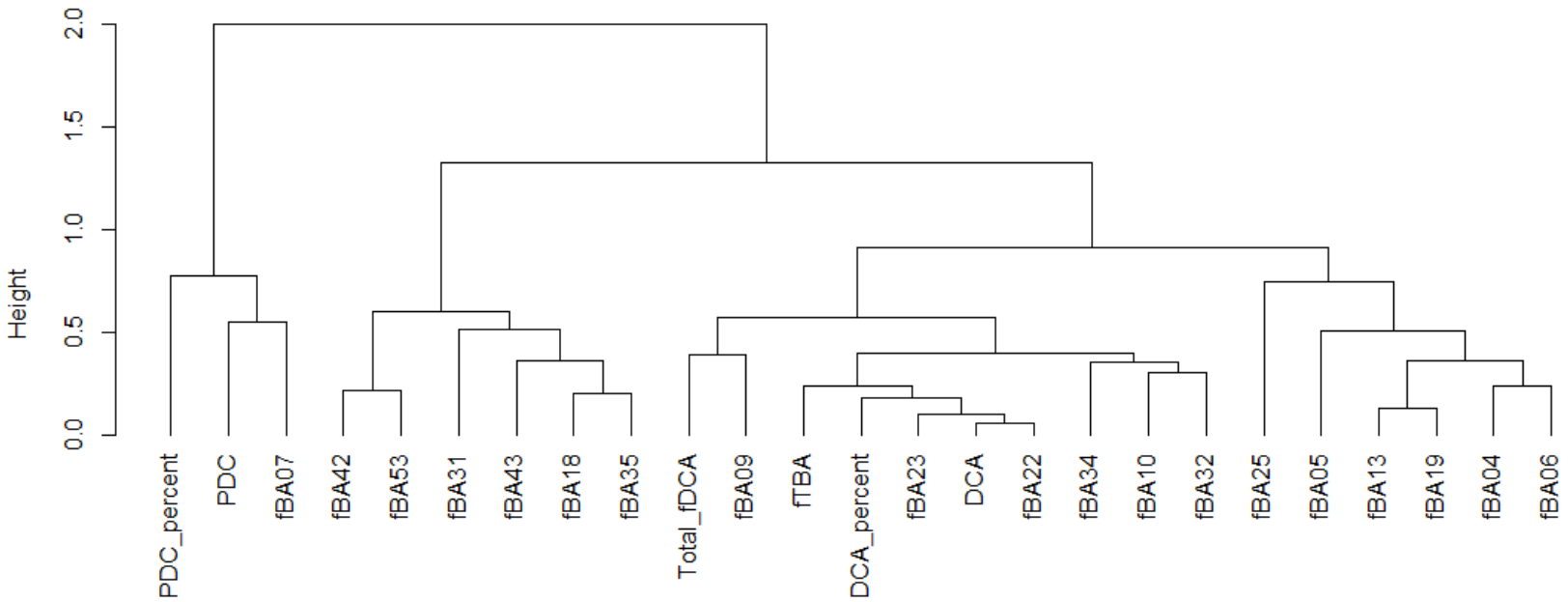


Figure S5-5: Similarity dendrogram showing hierarchical clustering pattern of PDC measures and faecal bile acids. Significant Spearman correlations were clustered by complete-linkage method (most similar variables are clustered together). fBA07 = allo-lithocholic acid, fBA42 = glycodeoxycholic acid, fBA53: taurodeoxycholic acid, fBA31 = 12-dehydrocholic acid, fBA43 = glycoursodeoxycholic acid, fBA18 = 3α-hydroxy-7-ketolithocholic acid, fBA35 = cholic acid, fBA09 = 3α,12α, 23-nordeoxycholic acid, fTBA = faecal total bile acids, fBA23 = deoxycholic acid,fBA22 = 5β-cholanic acid-3β,12α-diol, fBA34 = β-muricholic acid, fBA10 = 9(11), (5β)-cholenic acid-3α-ol-12-one, fBA32 = ursodeoxycholic acid, fBA25 = hyodeoxycholic acid, fBA05 = lithocholic acid, fBA13 = 3,6-diketocholanic acid, fBA19 = 3α-hydroxy-12-ketolithocholic acid, fBA04 = 3-ketocholanic acid, fBA06 = isolithocholic acid.

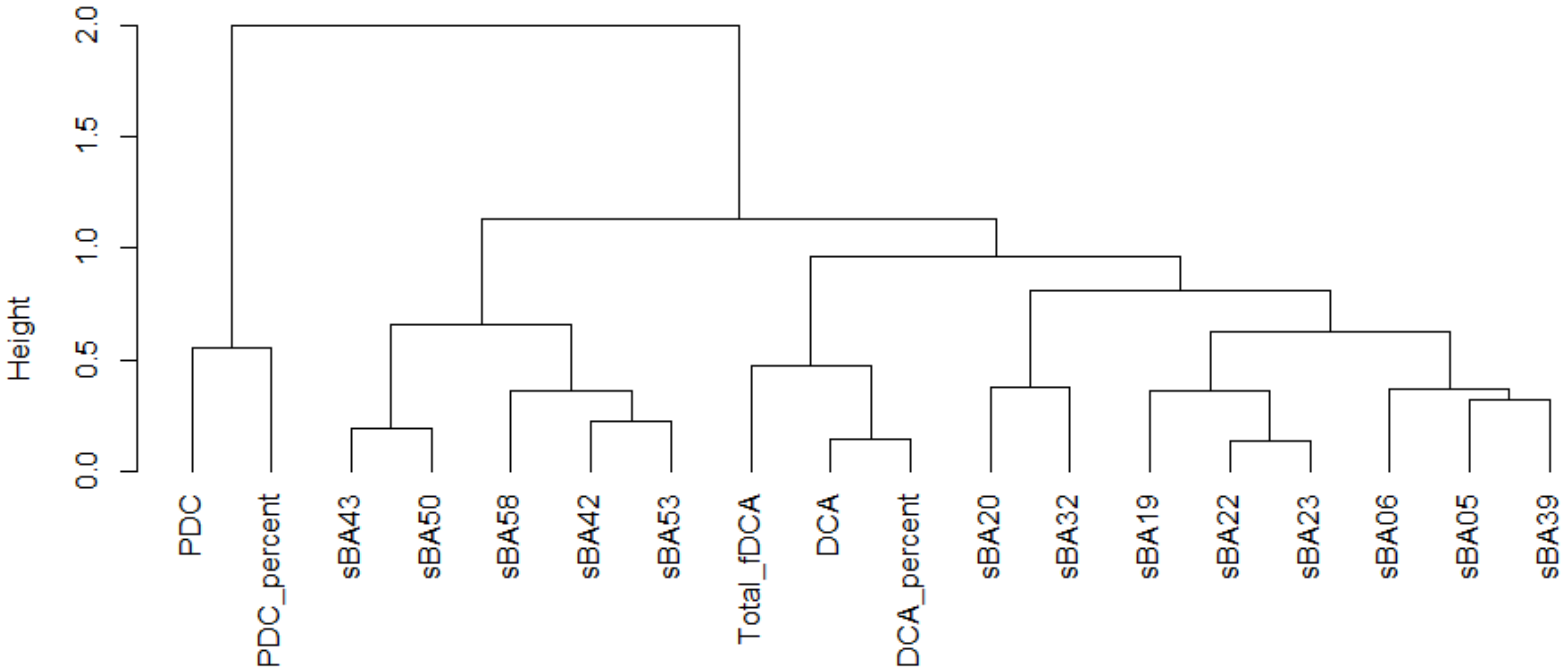


Figure S5-6: Similarity dendrogram showing hierarchical clustering pattern of PDC measures and serum bile acids. Significant Spearman correlations were clustered by complete-linkage method (most similar variables are clustered together). sBA43 = glycoursodeoxycholic acid, sBA50 = tauroursodeoxycholic acid, sBA58 = tauro ω-muricholic acid, sBA42 = glycodeoxycholic acid, sBA53 = taurodeoxycholic acid, sBA20 = murocholic acid, sBA32 = ursodeoxycholic acid, sBA19 = 3α-hydroxy-12-ketolithocholic acid, sBA22 = 5β-cholanic acid-3β, 12α-diol, sBA23 = deoxycholic acid, sBA06 = isolithocholic acid, sBA05 = lithocholic acid, sBA39 = glycolithocholic acid.

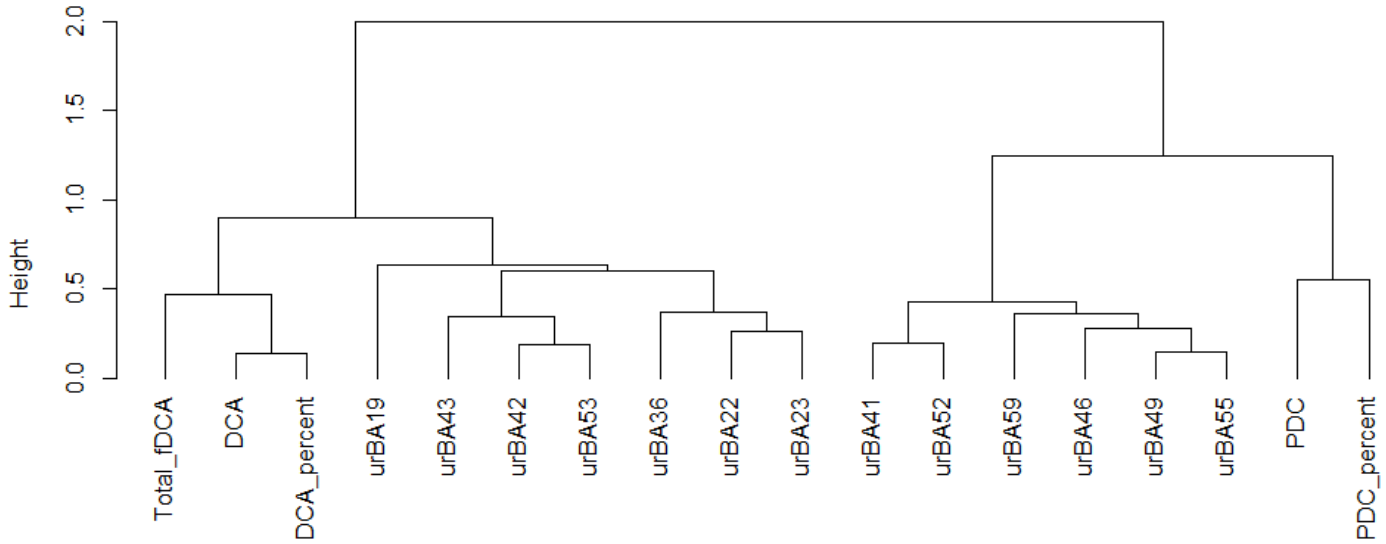


Figure S5-7: Similarity dendrogram showing hierarchical clustering pattern of PDC measures and urine bile acids. Significant Spearman correlations were clustered by complete-linkage method (most similar variables are clustered together). urBA19 = 3α-hydroxy-12-ketolithocholic acid, urBA43 = glycoursodeoxycholic acid, urBA42 = glycodeoxycholic acid, urBA53: taurodeoxycholic acid, urBA36 = hyocholic acid, urBA22 = 5β-cholanic acid-3β,12α-diol, urBA23 = deoxycholic acid, urBA41 = glycochenodeoxycholic acid, urBA52 = taurochenodeoxycholic acid, urBA59 = taurocholic acid, urBA46 = glycocholic acid, urBA49 = glycohyocholic acid, urBA55 = taurohyocholic acid.

Table S5-1: Table showing *p* values of comparisons made based on diet (vegan vs omnivore), PDC quartile categories (low PDC vs high PDC) and PDC% quartile categories (low PDC% vs high PDC%).

		Vegan vs Omnivore		Low PDC vs High PDC		Low PDC% vs High PDC%	
		CC dataset	M dataset	CC dataset	M dataset	CC dataset	M dataset
1	**DCA**	0.085	0.019	0.481	0.562	p < 0.0001	p < 0.0001
2	**PDC**	0.429	0.252	p < 0.0001	p < 0.0001	p < 0.0001	p < 0.0001
3	**Total fDCA**	0.053	0.015	p < 0.0001	p < 0.0001	p < 0.0001	p < 0.0001
4	**PDC%**	0.284	0.122	p < 0.0001	p < 0.0001	p < 0.0001	p < 0.0001
5	**Height**	0.708	0.594	0.001	p < 0.0001	0.243	0.340
6	**Weight**	0.079	0.044	0.050	0.014	0.956	0.982
7	**Waist**	0.526	0.683	0.130	0.108	0.175	0.137
8	**Body fat Percent**	0.898	0.979	0.117	0.092	0.577	0.574
9	**Age**	0.353	0.694	0.353	0.197	0.891	0.701
10	**BMI**	0.017	0.019	0.411	0.706	0.204	0.128
12	**Systolic BP**	0.990	0.784	0.001	p < 0.0001	0.124	0.197
13	**Diastolic BP**	0.997	0.861	0.060	0.094	0.169	0.386
14	**BMR**	0.365	0.334	0.012	p < 0.0001	0.643	0.769
18	**Vegan Duration**	NA	NA	0.183	0.099	0.103	0.047
19	**Total cholesterol**	p < 0.0001	p < 0.0001	0.762	0.824	0.287	0.260
20	**HDL**	0.148	0.070	0.303	0.520	0.359	0.336
21	**LDL**	p < 0.0001	p < 0.0001	0.924	0.784	0.848	0.820
22	**Triglyceride**	0.150	0.155	0.723	0.535	0.191	0.252
23	**VLDL**	0.226	0.174	0.797	0.539	0.184	0.244
24	**Mean Fruit intake g/day**	p < 0.0001	p < 0.0001	0.856	0.770	0.299	0.410
25	**Mean veg intake g/day**	p < 0.0001	p < 0.0001	0.503	0.243	0.087	0.154
26	**Mean nut & legume intake g/day**	p < 0.0001	p < 0.0001	0.556	0.482	0.004	0.009
27	**Mean wholegrain intake g/day**	0.284	0.492	0.177	0.183	0.609	0.467
28	**Alanine**	p < 0.0001	p < 0.0001	0.066	0.038	0.353	0.239
29	**Alpha tocopherol**	p < 0.0001	p < 0.0001	0.797	0.943	0.014	0.010
30	**Arginine**	0.012	0.004	0.078	0.037	0.607	0.516
31	**Aspartic acid**	p < 0.0001	p < 0.0001	0.057	0.030	0.514	0.345
32	**Vitamin A**	0.001	0.001	1.000	0.893	0.595	0.951
33	**Vitamin B6**	0.051	0.015	0.574	0.638	0.829	0.652
34	**Beta.carotene**	0.013	0.001	0.471	0.277	0.009	0.048
35	**Biotin**	0.001	p < 0.0001	0.034	0.009	0.202	0.177
36	**C12.0**	p < 0.0001	p < 0.0001	0.723	0.502	0.959	0.932
37	**C14.0**	p < 0.0001	p < 0.0001	0.287	0.144	0.268	0.222
38	**C16.0**	p < 0.0001	p < 0.0001	0.154	0.038	0.091	0.107
39	**C16.1.n.7**	p < 0.0001	p < 0.0001	0.191	0.144	0.102	0.090
40	**C18.0**	p < 0.0001	p < 0.0001	0.046	0.026	0.105	0.228
41	**C18.1.n.7**	0.050	0.029	0.039	0.078	0.578	0.913

42	**C18.1.n.9**	0.001	0.006	0.079	0.088	0.559	0.902
43	**C18.2.n.6**	0.002	p < 0.0001	0.197	0.350	0.295	0.222
44	**C18.3.n.3**	0.586	0.951	0.316	0.162	0.628	0.913
45	**C20.0**	0.010	0.029	0.416	0.358	0.462	0.809
46	**C20.1.n.11**	p < 0.0001	p < 0.0001	0.392	0.274	0.049	0.044
47	**C22.0**	0.282	0.231	0.924	0.917	0.404	0.821
48	**Calcium**	p < 0.0001	p < 0.0001	0.228	0.177	0.656	0.555
49	**Cholesterol**	p < 0.0001	p < 0.0001	0.297	0.232	0.136	0.134
50	**Chromium**	0.004	0.006	0.762	0.731	0.836	0.950
51	**Vitamin C**	p < 0.0001	p < 0.0001	0.775	0.558	0.057	0.055
52	**Cystine**	0.072	0.049	0.147	0.060	0.525	0.458
53	**Energy (total intake)**	0.698	0.473	0.026	0.048	0.655	0.578
54	**Vitamin E**	p < 0.0001	p < 0.0001	0.619	0.855	0.025	0.018
55	**Fat % of total energy**	p < 0.0001	p < 0.0001	0.177	0.089	0.906	0.877
56	**Fat (total intake)**	0.001	0.004	0.057	0.050	0.864	0.914
57	**Folate**	p < 0.0001	p < 0.0001	0.897	0.492	0.144	0.216
58	**Phosphorous**	0.149	0.077	0.023	0.013	0.918	0.817
59	**Fructose**	0.011	0.003	0.869	0.878	0.481	0.589
60	**Glucose**	0.004	0.001	0.784	0.804	0.932	0.769
61	**Glutamic acid**	p < 0.0001	p < 0.0001	0.061	0.038	0.409	0.293
62	**Glycine**	p < 0.0001	p < 0.0001	0.081	0.043	0.439	0.300
63	**Histidine**	p < 0.0001	p < 0.0001	0.050	0.025	0.235	0.163
64	**Isoleucine**	p < 0.0001	p < 0.0001	0.101	0.050	0.310	0.198
65	**Iron**	0.004	p < 0.0001	0.172	0.281	0.257	0.253
66	**Iodine**	p < 0.0001	p < 0.0001	0.439	0.480	0.008	0.016
67	**Vitamin K1**	p < 0.0001	p < 0.0001	0.364	0.307	0.022	0.035
68	**Potassium**	0.004	p < 0.0001	0.481	0.432	0.797	0.841
69	**Copper**	0.002	p < 0.0001	0.876	0.854	0.802	0.519
70	**Fibre (total intake)**	p < 0.0001	p < 0.0001	0.471	0.444	0.235	0.215
71	**Fiber per 1000kj**	p < 0.0001	p < 0.0001	0.655	0.663	0.052	0.033
72	**Carbohydrate % of total energy**	p < 0.0001	p < 0.0001	0.416	0.290	0.574	0.769
73	**Carbohydrate (total intake)**	0.001	p < 0.0001	0.294	0.442	0.986	0.700
74	**Lactose**	p < 0.0001	p < 0.0001	0.449	0.268	0.086	0.078
75	**L-ascorbic acid**	p < 0.0001	p < 0.0001	0.655	0.405	0.215	0.182
76	**L-dehydroascorbic acid**	0.009	0.010	0.324	0.326	0.048	0.031
77	**Leucine**	p < 0.0001	p < 0.0001	0.087	0.046	0.302	0.188
78	**Lysine**	p < 0.0001	p < 0.0001	0.116	0.058	0.310	0.198
79	**Magnesium**	0.001	p < 0.0001	0.381	0.298	0.616	0.384
80	**Maltose**	0.042	0.006	0.835	0.867	0.331	0.361
81	**Manganese**	0.002	0.002	0.910	0.886	0.972	0.816
82	**Methionine**	p < 0.0001	p < 0.0001	0.152	0.080	0.190	0.124
83	**MUFA**	0.002	0.012	0.124	0.120	0.436	0.709
84	**SFA**	p < 0.0001	p < 0.0001	0.190	0.050	0.337	0.354
85	**Sodium**	p < 0.0001	p < 0.0001	0.055	0.178	0.037	0.074
86	**Niacin_1**	p < 0.0001	p < 0.0001	0.268	0.149	0.054	0.075
87	**Niacin_2**	0.019	0.015	0.381	0.471	0.184	0.370

88	**Nickel**	p < 0.0001	p < 0.0001	0.918	0.596	0.409	0.449
89	**Pantothenic acid**	0.001	p < 0.0001	0.691	0.213	0.037	0.040
90	**Phenylalanine**	p < 0.0001	p < 0.0001	0.081	0.045	0.353	0.239
91	**PUFA**	p < 0.0001	p < 0.0001	0.235	0.451	0.172	0.124
92	**Proline**	p < 0.0001	p < 0.0001	0.142	0.116	0.184	0.124
93	**Protein % of total energy**	p < 0.0001	p < 0.0001	0.392	0.188	0.207	0.256
94	**Protein (total intake)**	p < 0.0001	p < 0.0001	0.025	0.016	0.216	0.211
95	**Retinol**	p < 0.0001	p < 0.0001	0.138	0.133	0.126	0.118
96	**Riboflavin**	p < 0.0001	p < 0.0001	0.073	0.040	0.659	0.715
97	**Saccharose**	0.421	0.485	0.631	0.442	0.492	0.387
98	**Selenium**	p < 0.0001	p < 0.0001	0.033	0.031	0.263	0.246
99	**Serine**	p < 0.0001	p < 0.0001	0.094	0.041	0.327	0.210
100	**Starch**	0.001	0.024	0.341	0.341	0.302	0.253
101	**Sum_MUFA**	p < 0.0001	p < 0.0001	0.087	0.095	0.426	0.709
102	**Sum_SFA**	p < 0.0001	p < 0.0001	0.159	0.071	0.184	0.188
103	**Sum_n.3**	0.011	0.014	0.169	0.071	0.945	0.840
104	**Sum_n.6**	0.002	p < 0.0001	0.186	0.337	0.303	0.228
105	**Sum_PUFA**	0.017	0.001	0.085	0.151	0.368	0.297
106	**Thiamine**	0.016	0.008	0.610	0.666	0.647	0.554
107	**Threonine**	p < 0.0001	p < 0.0001	0.057	0.029	0.481	0.322
108	**Nitrogen (total)**	p < 0.0001	p < 0.0001	0.024	0.015	0.239	0.231
109	**Tryptophane_contrib**	p < 0.0001	p < 0.0001	0.078	0.042	0.138	0.113
110	**Tryptophane**	p < 0.0001	p < 0.0001	0.097	0.048	0.263	0.172
111	**Tyrosine**	p < 0.0001	p < 0.0001	0.124	0.065	0.271	0.172
112	**Valine**	p < 0.0001	p < 0.0001	0.069	0.034	0.419	0.279
113	**Water**	0.320	0.067	0.847	0.889	0.961	0.822
114	**Zink**	p < 0.0001	p < 0.0001	0.132	0.083	0.272	0.287
115	**(stool) 3-Ketocholanic acid**	0.639	0.938	0.037	0.089	0.283	0.156
116	**(stool) Lithocholic acid**	0.365	0.812	0.522	0.932	0.586	0.449
117	**(stool) Isolithocholic acid**	0.335	0.404	0.272	0.129	0.095	0.092
118	**(stool) Allolithocholic acid**	0.023	0.036	0.036	0.033	0.756	0.876
119	**(stool) 3a,12a, 23-Nordeoxycholic acid**	0.829	0.154	0.426	0.426	0.006	0.002
120	**(stool) 9(11), (5ß)-Cholenic Acid-3a-ol-12-one**	0.207	0.094	0.278	0.128	0.012	0.002
121	**(stool) 3,6-Diketocholanic acid**	0.408	0.624	0.341	0.234	p < 0.0001	p < 0.0001
122	**(stool) 3a-hydroxy-7 ketolithocholic acid**	0.864	0.602	0.429	0.309	p < 0.0001	p < 0.0001
123	**(stool) 3a-hydroxy-12 ketolithocholic acid**	0.403	0.281	0.619	0.781	0.680	0.422
124	**(stool) 5ß-Cholanic acid-3ß, 12a-diol**	p < 0.0001	p < 0.0001	0.586	0.767	0.993	0.613
125	**(stool) Deoxycholic acid**	0.007	0.002	0.460	0.476	0.002	p < 0.0001
126	**(stool) Chenodeoxycholic acid**	0.083	0.036	0.011	0.009	0.850	0.688
127	**(stool) Hyodeoxycholic acid**	0.042	0.005	0.460	0.322	0.835	0.628
128	**(stool) 12--Dehydrocholic acid**	0.403	0.385	0.333	0.413	0.359	0.191
129	**(stool) Ursodeoxycholic acid**	0.034	0.008	0.607	0.411	0.607	0.805
130	**(stool) ß-Muricholic acid**	0.140	0.078	0.294	0.274	p < 0.0001	p < 0.0001
131	**(stool) Cholic acid**	1.000	0.975	0.020	0.016	0.046	0.010
132	**(stool) Glycochenodeoxycholic acid**	0.747	0.767	0.837	0.958	0.294	0.449
133	**(stool) Glycodeoxycholic acid**	0.751	0.971	0.842	0.611	0.768	0.575

134	**(stool) Glycoursodeoxycholic acid**	0.022	0.002	0.015	0.009	0.595	0.610
135	**(stool) Glycocholic acid**	0.106	0.007	0.012	0.005	0.182	0.076
136	**(stool) Taurochenodeoxycholic acid**	0.776	0.924	0.619	0.394	0.643	0.723
137	**(stool) Taurodeoxycholic acid**	0.899	0.944	0.551	0.505	0.004	0.001
138	**(stool) Taurocholic acid**	0.917	0.760	0.033	0.030	0.932	0.975
139	**Total faecal BA**	0.916	0.886	0.009	0.010	0.001	p < 0.0001
140	**(serum)Ursocholanic Acid**	0.632	0.473	0.684	0.808	0.467	0.524
141	**(serum)5-Cholenic Acid-3β-ol**	0.355	0.240	0.157	0.110	0.559	0.758
142	**(serum)Lithocholic acid**	0.022	0.018	0.092	0.192	0.411	0.501
143	**(serum)Isolithocholic Acid**	0.070	0.137	0.363	0.436	0.043	0.070
144	**(serum)3α-Hydroxy-12 Ketolithocholic Acid**	0.862	0.676	0.539	0.594	p < 0.0001	p < 0.0001
145	**(serum)Murocholic Acid**	0.038	0.072	0.327	0.336	p < 0.0001	p < 0.0001
146	**(serum)5β-Cholanic Acid-3β, 12α-diol**	0.189	0.348	0.917	0.743	0.397	0.376
147	**(serum)Deoxycholic Acid**	0.437	0.463	0.219	0.272	0.141	0.212
148	**(serum)Chenodeoxycholic Acid**	0.149	0.011	0.184	0.180	0.945	0.987
149	**(serum)3,7,12 Dehydrocholic acid**	0.551	0.186	0.006	0.012	0.429	0.425
150	**(serum)Ursodeoxycholic acid**	0.229	0.816	0.048	0.173	0.006	0.011
151	**(serum)Cholic acid**	p < 0.0001	p < 0.0001	0.051	0.005	0.390	0.446
152	**(serum)Hyocholic acid**	0.001	p < 0.0001	0.179	0.091	p < 0.0001	p < 0.0001
153	**(serum)Glycolithocholic Acid**	0.085	0.069	0.236	0.174	0.001	p < 0.0001
154	**(serum)Glycochenodeoxycholic Acid**	p < 0.0001	p < 0.0001	0.897	0.408	0.242	0.289
155	**(serum)Glycodeoxycholic acid**	0.039	0.029	0.153	0.192	0.813	0.611
156	**(serum)Glycoursodeoxycholic Acid**	p < 0.0001	p < 0.0001	0.208	0.124	0.390	0.444
157	**(serum)Glycocholic Acid**	p < 0.0001	0.001	0.276	0.123	0.401	0.265
158	**(serum)Taurolithocholic Acid**	0.057	0.089	0.250	0.350	p < 0.0001	p < 0.0001
159	**(serum)Glycohyocholic Acid**	0.071	0.116	0.436	0.441	0.081	0.082
160	**(serum)Tauro-ursodeoxycholic Acid**	0.291	0.349	0.539	0.451	0.119	0.098
161	**(serum)Taurochenodeoxycholic Acid**	0.864	0.198	0.503	0.684	0.771	0.795
162	**(serum)Taurodeoxycholic Acid**	0.265	0.113	0.120	0.123	0.013	0.008
163	**(serum)Taurohyocholic Acid Sodium salt**	0.514	0.809	0.824	0.719	0.002	p < 0.0001
164	**(serum)Tauro ω-Muricholic Acid sodium salt**	0.530	0.204	0.023	0.012	p < 0.0001	p < 0.0001
165	**(serum)Taurocholic Acid**	0.122	0.048	p < 0.0001	p < 0.0001	0.156	0.063
166	**Total serum BA**	0.015	0.001	0.639	0.795	p < 0.0001	p < 0.0001
167	**(urine)Lithocholic acid**	0.894	0.152	0.332	0.318	p < 0.0001	p < 0.0001
168	**(urine)3α-Hydroxy-12 Ketolithocholic Acid**	0.082	0.049	0.945	0.994	p < 0.0001	p < 0.0001
169	**(urine)5β-Cholanic Acid-3β, 12α-diol**	0.694	0.928	0.023	0.023	0.007	0.002
170	**(urine)Deoxycholic Acid**	0.089	0.020	0.719	0.889	p < 0.0001	p < 0.0001
171	**(urine)Chenodeoxycholic Acid**	0.276	0.045	0.514	0.602	p < 0.0001	p < 0.0001
172	**(urine)3 Dehydrocholic Acid**	0.124	0.014	0.439	0.542	p < 0.0001	p < 0.0001
173	**(urine)12 Dehydrocholic Acid**	0.490	0.958	0.092	0.139	0.019	0.006
174	**(urine)Cholic acid**	0.985	0.225	0.446	0.347	0.001	0.001
175	**(urine)Hyocholic acid**	0.590	0.994	0.945	0.976	0.019	0.005
176	**(urine)Glycochenodeoxycholic Acid**	0.097	0.012	0.583	0.644	p < 0.0001	p < 0.0001
177	**(urine)Glycodeoxycholic acid**	0.686	0.129	0.007	0.012	p < 0.0001	p < 0.0001
178	**(urine)Glycoursodeoxycholic Acid**	0.718	0.601	0.116	0.123	0.003	0.001

179	**(urine)Glycocholic Acid**	0.355	0.560	0.179	0.374	0.009	0.010
180	**(urine)Taurolithocholic Acid**	0.175	0.497	0.303	0.478	p < 0.0001	p < 0.0001
181	**(urine)Glycohyocholic Acid**	0.342	0.126	0.145	0.221	0.001	0.001
182	**(urine)Taurochenodeoxycholic Acid**	0.196	0.171	0.693	0.971	0.041	0.072
183	**(urine)Taurodeoxycholic Acid**	0.470	0.163	0.021	0.026	0.069	0.053
184	**(urine)Taurohyocholic Acid Sodium salt**	0.657	0.200	0.019	0.035	p < 0.0001	p < 0.0001
185	**(urine)Taurocholic Acid**	0.783	0.457	0.112	0.112	0.152	0.153
186	**Total urine BA**	0.251	0.026	0.202	0.177	p < 0.0001	p < 0.0001

Appendix 6

Table S6-1: Table showing *p* values of differences between V1 and V3 of Peas study parameters. This demonstrated that all levels returned to baseline values after washout period.

Measurement	**V1 vs V3**
BMI	0.700
Weight	0.778
Breath Hydrogen	0.918
Glucose (AUC)	0.218
Insulin (AUC)	0.920
Fructosamine	0.317
C-Peptide	0.626
Total cholesterol	0.537
Triglycerides	0.521
HDL	0.488
LDL	0.878
Total cholesterol:HDL ratio	0.778
Non-HDL cholesterol	0.898
PDC	0.840
DCA	0.979
total fDCA	0.689
PDC%	0.979
DCA%	0.979

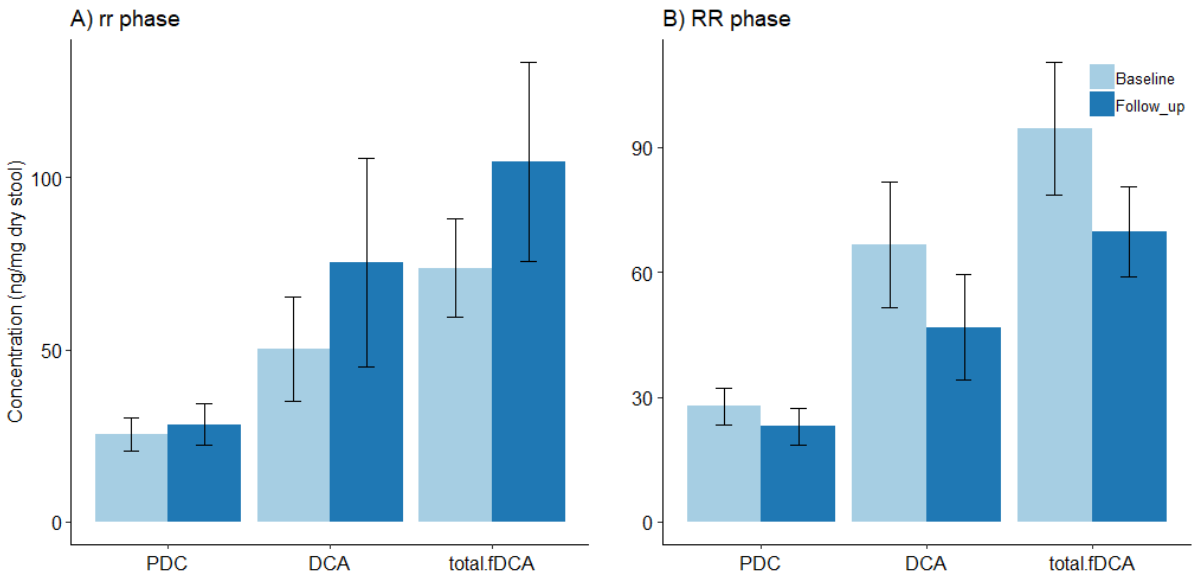


Figure S6-1: Bar charts showing mean baseline and follow-up PDC, DCA and total faecal DCA concentrations on A) *rr* and B) *RR* interventions

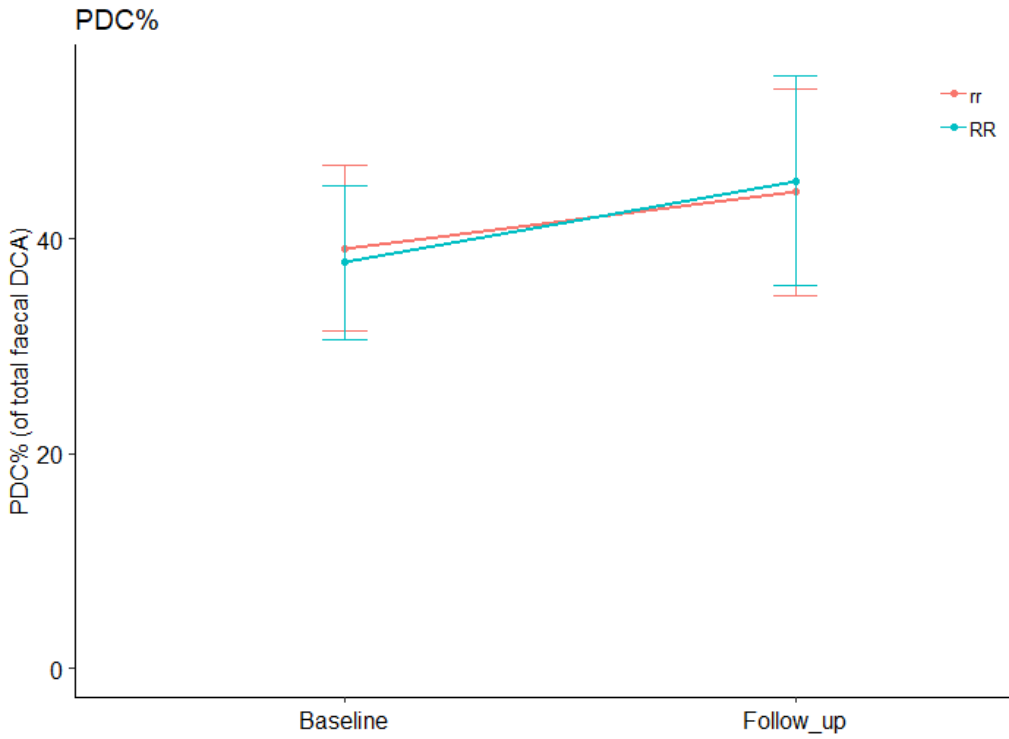


Figure S6-2: Line graph showing average baseline and follow-up PDC% on *rr* and *RR* interventions.

Appendix 7

No supplementary information for Chapter 7

Appendix 8

No supplementary information for Chapter 8

(Permissions from Appendices 1 and 2 to follow)

Copyright Clearance Center Marketplace™

American Soc for Biochemistry & Molecular Biology - License Terms and Conditions

This is a License Agreement between Adesola Bello ("You") and American Soc for Biochemistry & Molecular Biology ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Soc for Biochemistry & Molecular Biology, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	08-Jun-2020	**Type of Use**	Republish in a thesis/dissertation
Order license ID	1040301-1		
ISSN	1539-7262	**Publisher**	AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC.
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Journal of lipid research : JLR	**Country**	United States of America
Author/Editor	American Society for Biochemistry and Molecular Biology.	**Rightsholder**	American Soc for Biochemistry & Molecular Biology
Date	01/01/1959	**Publication Type**	e-Journal
Language	English	**URL**	http://www.jlr.org/

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the disabled?**	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	**Incidental promotional use?**	No
Duration of Use	Life of current edition	**Currency**	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Investigating polymerisation of deoxycholic acid	**Institution name**	Imperial College London
		Expected presentation date	2020-06-08
Instructor name	Adesola Bello		

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	Adesola Bello

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 8	**Title of the article/chapter the portion is from**	Polydeoxycholate in Human and Hamster Feces: A Major Product of Cholate Metabolism
Editor of portion(s)	N/A	**Author of portion(s)**	American Society for Biochemistry and Molecular Biology.
Volume of serial or monograph	34	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	2132	**Publication date of portion**	1993-01-01

PUBLISHER TERMS AND CONDITIONS

Journal of Biological Chemistry permissions policies: • If you are an author of the content for which you are seeking permission, or if you are not an author but are requesting permission to copy, distribute, transmit and adapt the work for noncommercial purposes (e.g. reproduction of a figure for educational purposes such as schoolwork, or appending a reprinted article to a PhD dissertation), you do not need to seek permission using the options listed below, as long as any reuse includes the credit line in the reuse policies listed above. • Parties who are not authors on the article who wish to reuse content for commercial purposes such as reproducing a figure in a book, journal, or coursepack published by a commercial publisher, do need permission and should request permission by completing the form below. For more information please see Journal of Biological Chemistry: http://www.jbc.org/site/misc/edpolicy.xhtml#copyright

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of

the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

    8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

    8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

    8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

    8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

    8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Apr 08, 2020

This Agreement between Imperial College London -- Adesola Bello ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4804121061123
License date	Apr 08, 2020
Licensed content publisher	Oxford University Press
Licensed content publication	The American Journal of Clinical Nutrition
Licensed content title	Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group
Licensed content author	Yetley, Elizabeth A; MacFarlane, Amanda J
Licensed content date	Dec 7, 2016
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	PhD Student
Publisher of your work	Imperial College London

Expected publication date	Apr 2020
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	PhD Student
Institution name	Imperial College London
Expected presentation date	Apr 2020
Portions	Figure 1 Hierarchy of Evidence pyramid
Requestor Location	Imperial College London Sir Alexander Fleming Building Exhibition road London, SW7 2AZ United Kingdom Attn: Imperial College London
Publisher Tax ID	GB125506730
Total	0.00 USD

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press’s written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 08, 2020

This Agreement between Imperial College London -- Adesola Bello ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4804120956003
License date	Apr 08, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Microbiology
Licensed Content Title	Classification and Identification Tasks in Microbiology: Mass Spectrometric Methods Coming to the Aid
Licensed Content Author	N. V. Velichko et al
Licensed Content Date	Sep 30, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	PhD Student
Institution name	Imperial College London
Expected presentation date	Apr 2020
Portions	Figure 2
Requestor Location	Imperial College London Sir Alexander Fleming Building Exhibition road London, SW7 2AZ United Kingdom Attn: Imperial College London
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also

seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that

affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 08, 2020

This Agreement between Imperial College London -- Adesola Bello ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4804130640321
License date	Apr 08, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Microbiology
Licensed Content Title	Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences
Licensed Content Author	Pablo Yarza et al
Licensed Content Date	Aug 14, 2014
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	PhD Student
Institution name	Imperial College London
Expected presentation date	Apr 2020
Portions	Figure 1: Variable regions of the 16S ribosomal RNA.
Requestor Location	Imperial College London Sir Alexander Fleming Building Exhibition road London, SW7 2AZ United Kingdom Attn: Imperial College London
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any

other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.