The gut microbiota characteristics of children with Coeliac Disease and the effect of treatment

Ms Konstantina Zafeiropoulou, BSc

A thesis submitted for the degree of MSc (by Research)

To

The College of Medicine, Veterinary and Life Sciences

University of Glasgow

December 2017

From research conducted at

The Department of Human Nutrition

University of Glasgow

New Lister Building

Glasgow Royal Inifirmary

ABSTRACT

Objectives and study

The incidence of coeliac disease (CD) has been significantly increasing over the past decades in Western society [1]. The rapid increase in incidence, along with the relatively low prevalence of the disease among genetic susceptible individuals suggest a strong contribution of environmental factors in the development of CD, including changes in the gut microbiota. Recent literature indicates that the gut microbiota is altered in CD. However, the evidence is inconclusive and it still has to be confirmed whether a disturbed microbiota is contributing to disease development, or is merely a consequence of disease activity and treatment with gluten free diet (GFD). The current study evaluated microbiota compositional and functional characteristics in treated CD children (TCD), newly diagnosed untreated CD (UCD), their siblings and healthy controls (HC).

Methods

Faecal samples were collected from TCD on GFD for at least one year, UCD, their siblings and HC. UCD were followed prospectively for six and 12 months after the initiation of treatment with GFD. Dietary intake characteristics were evaluated using food frequency questionnaire (FFQ), short chain fatty acids (SCFA) were measured using gas chromatography (GC) and microbial community structure and composition using 16S sequencing. Compliance with GFD was evaluated based on the Biagi score, clinicians' evaluation, serum tissue transglutaminase (tTG) immunoglobulin A (IgA) levels, as well as the novel biomarker faecal gluten immunogenic peptide (GIP).

Results

145 participants had faecal samples collected (45 TCD, 20 UCD, 23 siblings of the CD children and 57 HC). Thirteen UCD patients provided paired samples at baseline, six and 12 months on GFD. TCD had significantly lower concentrations of propionic, butyric, valeric, and caproic acids per total faecal output than HC (all $p < 0.05$). Moreover, TCD microbiota clustered separately from HC ($R^2 = 2.56$, $p = 0.013$) in the non-metric multidimensional scaling (NMDS) plot for the Bray-Curtis dissimilarity index at genus level. There was no significant difference in the microbiota structure and functionality between UCD and HC children. In the prospective cohort, acetic acid significantly increased [(72.5 (1.4)) Vs (65.5 (1.3), $p = 0.003$)] and butyric acid significantly decreased [(9.7 (1.4)) Vs (13.6 (1.04), $p = 0.013$)] at six months on GFD. The relative abundance of OTU_908 *Anaerostipes,* a butyrate-producing bacterium, also decreased at six months on GFD. Furthermore, the relative abundance of acetic acid was significantly associated with GIP levels in faeces of CD children who were on GFD for at least one year. Increased GIP levels in faeces, so decreased compliance with GFD, were strongly associated with decreased levels of acetic ($p < 0.05$), indicating that changes in the microbiota metabolites concentration were significantly associated with the GFD adherence. Similarly, compared to baseline values, isobutyric and isovaleric acids significantly decreased at 12 months on GFD ($p < 0.05$).

Significant differences in the relative abundance of OTU_259 *Erysipelatoclostridium* and OTU_98 *Ruminococcaceae* were observed in faeces of CD children after treatment with GFD, although they were of different direction in the prospective and cross-sectional study. Compared to baseline values, the relative abundance of OTU_259 and OTU_98 significantly decreased at six months on GFD, whereas the relative abundance of both OTU were significantly higher in TCD than UCD children, in the cross-sectional study.

Conclusion

The findings of the present study indicate that alterations in microbiota composition and SCFA in patients with CD are most likely a secondary consequence of adherence to a GFD rather than a primary disease effect. Although no UCD microbiota "dysbiosis" was observed, faeces of UCD children had significantly lower relative abundance of 31 distinct OTU. This may be used to unravel our knowledge on the disease process, but these alterations observed in this study need replication in larger cohorts of CD children. Finally, the altered microbiota functionality of patients with CD after the initiation of GFD could be used to assess compliance with GFD, monitor disease improvement but also unravel the mechanisms through which GFD exhibits its beneficial effect.

TABLE OF CONTENTS

ABSTRACT ..... **II**

TABLE OF CONTENTS ..... **IV**

TABLE LIST ..... **VII**

FIGURE LIST ..... **X**

ACKNOWLEDGMENTS ..... **XIII**

AUTHOR'S DECLARATION ..... **XIV**

ABBREVIATION LIST ..... **XV**

CHAPTER 1 COELIAC DISEASE, GLUTEN FREE DIET AND THE GUT MICROBIOTA **1**

1.1 Coeliac disease **1**

1.1.1 Prevalence 1
1.1.2 Pathogenesis 1
1.1.2.1 Genetic factors 2
1.1.2.2 Environmental factors 2
1.1.3 Treatment 3

1.2 The role of gut microbiota in Coeliac disease **3**

1.2.1 Gut microbiota 3
1.2.2 Is the gut microbiota involved in the development of CD? 4
1.2.3 Literature review of studies on faecal microbiota in CD 4
1.2.4 Faecal microbiota in CD. How good is the evidence? What is missing? 10
1.2.5 Microbiota metabolic activity 13
1.2.5.1 Short chain fatty acids 13
1.2.5.2 Lactate 14
1.2.5.3 Sulphide 14
1.2.5.4 Ammonia 15
1.2.6 Literature review of studies on microbiota metabolites in CD 15
1.2.7 Microbiota metabolites in CD. How good is the evidence? What is missing? 19

1.3 General conclusions and the purpose of this thesis **20**

CHAPTER 2 SUBJECTS AND METHODS **22**

2.1 Chapter outline **22**

2.2 Study design **22**

2.3 Participants & exclusion criteria **22**

2.4 Initial appointment **22**
2.4.1 Anthropometric data 23
2.4.2 Dietary assessment 23
2.4.3 Gastrointestinal symptoms 23
2.4.4 Biagi score 23

2.5 Laboratory measurements **24**
2.5.1 Faecal sample collection 24
2.5.2 Faecal water content 24
2.5.3 Faecal pH 24
2.5.4 SCFA 25
2.5.5 Sulphide 25
2.5.6 Lactate 26
2.5.7 Gluten immunogenic peptide (GIP) 26
2.5.8 Serological tests 26
2.5.9 16S rRNA gene sequencing 27

2.6 Bioinformatics **28**

2.7 Statistical analysis **28**

2.8 Ethical considerations **30**

CHAPTER 3 METHODOLOGY CHAPTER: THE EFFICACY OF DIFFERENT METHODS TO ASSESS THE COMPLIANCE WITH GFD **31**

3.1 Chapter outline **31**

3.2 Introduction **31**

3.3 Subjects and methods **32**

3.4 Statistical analysis **32**

3.5 Results **32**
3.5.1 Agreement between the Biagi score and faecal GIP analysis 32
3.5.2 Agreement between clinicians' report and faecal GIP analysis 33
3.5.3 Agreement between the serum tTG and faecal GIP analysis 33

3.6 Conclusion **34**

CHAPTER 4 GUT MICROBIOTA STRUCTURE AND FUNCTIONALITY IN CD PATIENTS **35**

4.1 Chapter outline **35**

4.2 Cross – sectional study **35**
4.2.1 Dietary intake 36
4.2.2 Compliance with GFD 37
4.2.3 Serological tests 37
4.2.4 Gastrointestinal symptoms (PedsQL-GSS questionnaire) 38
4.2.5 Faecal sample characteristics 40

4.2.6 Sulphide ... 40
4.2.7 Lactate ... 42
4.2.8 SCFA ... 43
4.2.9 Faecal microbiota ... 48
4.2.9.1 Faecal microbiota community structure ... 48
4.2.9.2 Bacterial taxon relative abundance ... 58

4.3 Prospective study ... 76
4.3.1 Dietary intake ... 76
4.3.2 Compliance with GFD ... 77
4.3.3 Serological markers ... 79
4.3.4 Gastrointestinal symptoms (PedsQL-GSS questionnaire) ... 80
4.3.5 Faecal sample characteristics ... 81
4.3.6 Sulphide ... 82
4.3.7 Lactate ... 82
4.3.8 SCFA ... 83
4.3.9 Faecal microbiota ... 87
4.3.9.1 Faecal microbiota community structure ... 87
4.3.9.2 Bacterial taxon relative abundance ... 93

4.4 Association between microbiota SCFA and metadata ... 103
4.4.1.1 Predictors of faecal SCFA levels ... 117

CHAPTER 5 GENERAL DISCUSSION AND CONCLUSIONS ... 130

5.1 Study strengths and limitations ... 136

5.2 Conclusion ... 137

REFERENCES ... 139

APPENDIX: PEDSQL-GSS QUESTIONNAIRE ... 144

TABLE LIST

Table 1.1: Potential factors influencing early gut colonization (left side), and the cross-talk between the microbiome and host genetics, immunity, metabolism and nutrition (right side). Table adapted from Houghteling et al., 2015 [34]. 4
Table 1.2: Primary evidence from human studies exploring the faecal microbiota composition in patients with Coeliac disease 6
Table 1.3: Strengths and weaknesses of the techniques used to characterise the gut microbiota. Table adapted from Fraher et al (2012) [52] 11
Table 1.4: Primary evidence of human studies exploring the gut microbiota metabolic activity in individuals with Coeliac Disease (CD) 17
Table 1.5: Microbiota metabolites in CD (overall presentation) 19
Table 2.1: PCR specific conditions 27
Table 2.2: Interpretation of Kappa values to evaluate the strength of agreement between two methods 29
Table 3.1: Frequency of compliant and non-compliant CD children according to faecal GIP concentration, Biagi score, clinicians' report and serum tTG levels 34
Table 4.1: Demographic and anthropometric characteristics of HC, UCD, TCD children and siblings 35
Table 4.2: Dietary intake of HC, UCD, TCD children and siblings 36
Table 4.3: Biagi score, serological antibody tests, and faecal GIP levels of UCD and TCD children 38
Table 4.4: Frequency (n (%)) of self-reported GI symptoms and PedsQL-GSS score of HC, UCD, TCD children and siblings 39
Table 4.5: Faecal sample characteristics (total sample weight, pH, water content and ammonia) of HC, UCD, TCD children and siblings 40
Table 4.6: Faecal sulphide concentration (μmol/g) of HC, UCD, TCD children and siblings 41
Table 4.7: Faecal D-, L-lactate concentration (g/ 100g) of HC, UCD, TCD children and siblings 42
Table 4.8: Faecal SCFA concentration (μmol/g) of HC, UCD, TCD children and siblings 44
Table 4.9: Faecal BCFA concentration (μmol/g) of HC, UCD, TCD children and siblings 45
Table 4.10: Relative abundance (%) of faecal SCFA and BCFA of HC, UCD, TCD children and siblings 46
Table 4.11: Alpha diversity indices in faecal samples of UCD, TCD and HC 48
Table 4.12: PERMANOVA analysis to evaluate the variance using Bray-Curtis and unweighted Unifrac distance matrices at OTU and genus level in TCD, UCD and HC (clustering using USEARCH) 49
Table 4.13: PERMANOVA analysis to evaluate the variance using Bray-Curtis and unweighted Unifrac distance matrices at OTU and genus level in TCD, UCD and HC (clustering using DADA2) 50
Table 4.14: OTU with significantly different relative abundance in faecal samples of UCD, TCD and HC 58
Table 4.15: Genera with significantly different relative abundance in faecal samples of UCD, TCD & HC 64

Table 4.16: OTU with significantly different relative abundance in both comparisons UCD Vs TCD and group 1 (GIP < 0.156 μg/g wet matter) Vs group 2 (GIP ≥ 0.156 μg/g wet matter) 69
Table 4.17: Anthropometric characteristics of CD children at diagnosis and at six and 12 months on GFD 76
Table 4.18: Dietary intake of CD children at diagnosis and at six and 12 months on GFD 77
Table 4.19: Biagi score & GIP concentration in faecal samples of CD children at diagnosis and at six and 12 months on GFD 78
Table 4.20: CD serological markers in CD children at diagnosis and at six and 12 months on GFD 79
Table 4.21: Frequency of self-reported GI symptoms in CD children at diagnosis and at six and 12 months on GFD 80
Table 4.22: Stool samples' characteristics of CD children at diagnosis and at six and 12 months on GFD 81
Table 4.23: Faecal sulphide concentration (μmol/g) of CD children at diagnosis and at six and 12 months on GFD 82
Table 4.24: Faecal D-, L-lactate concentration (g/ 100g matter) of CD children at diagnosis and at six and 12 months on GFD 83
Table 4.25: Faecal SCFA concentration (μmol/g) of CD children at diagnosis and at six and 12 months on GFD 83
Table 4.26: Faecal BCFA concentration (μmol/g) of CD children at diagnosis and at six and 12 months on GFD 84
Table 4.27: Faecal SCFA and BCFA relative abundance (%) of CD children at diagnosis and at six and 12 months on GFD 85
Table 4.28: Alpha diversity indices in faecal samples of CD children at diagnosis, six and 12 months on GFD 87
Table 4.29: OTU with significantly different relative abundance in faecal samples of (a) paired data from 13 CD children at diagnosis and at six and 12 months after the initiation of GFD (b) independent data from UCD patients (n=20) and TCD patients (n=45) 94
Table 4.30: Genera with significantly different relative abundance in faecal samples of (a) paired data from 13 CD children at diagnosis and at six and 12 months after the initiation of GFD (b) independent data from UCD patients (n=20) and TCD patients (n=45) (OTU clustering using USEARCH) 97
Table 4.31: Spearman rank correlation between the absolute concentration (μmol/g dry matter) / relative abundance (%) of faecal SCFA and the age (years), BMI z-score (SD), faecal GIP, serum tTG, PedsQL-GSS score, energy intake and dietary intake of fat of CD patients on GFD at least one year 105
Table 4.32: Spearman rank correlation between the absolute concentration (μmol/g dry matter) / relative abundance (%) of faecal SCFA and the dietary intake of fat (%TDEI), protein (g, %TDEI, %RNI), carbohydrates (g, %TDEI) and sugars (%TDEI) of CD patients on GFD at least one year 107
Table 4.33: Spearman rank correlation between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and the dietary intake of starch (%TDEI), NMES (%TDEI),

NSP (g, %TDEI), dietary fibre (g), the faecal pH and water content (%) of CD patients on GFD at least one year.....109
Table 4.34: Association and Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of faecal SCFA and the categorical variables (a) gender and (b) compliance with GFD based on faecal GIP levels of CD patients on GFD at least one year 111
Table 4.35: Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and age (years), BMI z-score, GIP (μg/g wet matter), tTG (U/ mL), PedsQL-GSS score, energy intake (%EAR) and fat intake (gr) of CD patients on GFD at least one year 121
Table 4.36: Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and dietary intake of fat (%TDEI), protein (g, %TDEI, % RNI), carbohydrates (g, %TDEI) and sugars (%TDEI) of CD patients on GFD at least one year 123
Table 4.37: Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and dietary intake of starch (% TDEI), NMES (% TDEI), NSP (g, % TDEI) and dietary fibre (g), and faecal water content (%) of CD patients on GFD at least one year 125
Table 4.38: Predictors of SCFA concentration (μmol/g) in dry matter of CD patients on GFD at least one year (n=60), using multiple linear regression analysis (Beta Coefficient, *p-value*) 127
Table 4.39: Predictors of SCFA concentration (μmol/g) in wet matter of CD patients on GFD at least one year (n=58), using multiple linear regression analysis (Beta Coefficient, *p-value*).....128
Table 4.40: Predictors of the relative abundance of SCFA (%) of CD patients on GFD at least one year (n=58), using multiple linear regression analysis (Beta Coefficient, *p-value*).....129

FIGURE LIST

Figure 2.1: Biagi questionnaire, adapted from Biagi et al (2012) [91] 24
Figure 4.1: Faecal free sulphide concentration (μmol/g) of HC, UCD and TCD children measured in dry and wet matter 41
Figure 4.2: Faecal L-lactate concentration (g/ 100g) of HC, UCD and TCD children measured in dry, wet matter and per total faecal output 43
Figure 4.3: Faecal D-lactate concentration (g/ 100g wet matter) of HC, UCD and TCD children 43
Figure 4.4: Relative abundance (%) of faecal (a) acetic, (b) butyric and (c) valeric acid in samples of children with TCD, UCD & HC 47
Figure 4.5: Alpha diversity measures at OTU and ASV level using (a) USEARCH and (b) DADA2 algorithms respectively, for TCD, UCD and HC 51
Figure 4.6: Alpha diversity measures at genus level using (a) USEARCH and (b) DADA2 algorithms for TCD, UCD and HC 52
Figure 4.7: Non-metric multidimensional scaling (NMDS) plot at OTU level using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for TCD, UCD and HC (OTU clustering using USEARCH) 53
Figure 4.8: Non-metric multidimensional scaling (NMDS) plot at ASV level, using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for TCD, UCD and HC (ASV using DADA2) 54
Figure 4.9: Non-metric multidimensional scaling (NMDS) plot at genus level for the Bray-Curtis dissimilarity index using (a) USEARCH and (b) DADA2 algorithms for TCD, UCD and HC 55
Figure 4.10: Local contribution of β-diversity (LCBD) analysis at (a) OTU and (b) genus level for TCD, UCD and HC (% of total community dispersion) 56
Figure 4.11: Beta dispersion measured as the distance of each sample from the respective group ellipse centroid in (a) HC and UCD patients, (b) UCD and TCD and (c) HC and TCD patients (OTU clustering using USEARCH) 57
Figure 4.12: Log-relative abundance of OTU that differentiated UCD children from HC 66
Figure 4.13: Log-relative abundance of genera that differentiated UCD children from HC 67
Figure 4.14: Log-relative abundance of OTU that differentiated TCD from UCD children 71
Figure 4.15: (a) Log-relative abundance of genera that differentiated TCD from UCD; (b) Kendall rank correlation of the OTU that differentiated the TCD from UCD and faecal GIP concentration, PedsQL-GSS score, dietary intake of NMES (%TDEI), NSP (g, %DRV), Starch (%TDEI) and PUFA (%TDEI) 72
Figure 4.16: Log-relative abundance of OTU that differentiated TCD from HC children 74
Figure 4.17: (a) Log-relative abundance of genera that differentiated TCD from HC; (b) Kendall rank correlation of OTU that differentiated TCD from HC and acetic (%), propionic (μmol/ total faecal output), butyric (%, μmol/ total faecal output), valeric acid (%, μmol/ total faecal output), caproic (μmol/gr wet matter, μmol/ total faecal output), isobutyric (μmol/ total faecal output), isovaleric

(μmol/ total faecal output), isocaproic (μmol/gr wet matter, μmol/ total faecal output), and the PedsQL-GSS score 75
Figure 4.18: Classification of CD children into three groups, based on faecal GIP levels at diagnosis and at six and 12 months on GFD 79
Figure 4.19: SCFA and BCFA in 13 CD children at diagnosis and at six and 12 months on GFD; (a) relative abundance (%) of acetic acid, (b) relative abundance (%) of butyric acid, (c) concentration of butyric acid in dry matter (μmol/g), (d) relative abundance (%) of isobutyric (left) and isovaleric (right) acid 86
Figure 4.20: Alpha diversity measures at OTU and ASV level using (a) USEARCH and (b) DADA2 algorithms respectively, for CD patients at diagnosis, six and 12 months on GFD 88
Figure 4.21: Alpha diversity measures at genus level using (a) USEARCH and (b) DADA2 algorithms for CD patients at diagnosis, six and 12 months on GFD 89
Figure 4.22: Non-metric multidimensional scaling (NMDS) plot at OTU level, using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for CD patients at diagnosis, six and 12 on GFD (OTU clustering using USEARCH) 90
Figure 4.23: Non-metric multidimensional scaling (NMDS) plot at genus level, using Bray-Curtis dissimilarity index for CD patients at diagnosis, six and 12 months on GFD (OTU clustering using USEARCH) 91
Figure 4.24: Local contribution of β-diversity (LCBD) analysis at (a) OTU and (b) genus level for CD patients at diagnosis, six and 12 months on GFD (% of total community dispersion) 92
Figure 4.25: Log-relative abundance of OTU that changed significantly at six months on GFD compared to the diagnosis 99
Figure 4.26: Log-relative abundance of genera that changed significantly at six months on GFD compared to the diagnosis 100
Figure 4.27: Log-relative abundances of (a) OTU and (b) genera that changed significantly at 12 months on GFD compared to the diagnosis 101
Figure 4.29: Concentration in dry matter (μmol/g) of (a) acetic, (b) propionic, (c) butyric acid, (d) total SCFA in female and male CD children on GFD at least one year 113
Figure 4.30: Spearman rank correlation between the PedsQL-GSS score and the concentration in dry matter of faecal (a) acetic acid, (b) total SCFA, as well as between total daily energy intake (%EAR) and the concentration in dry matter of (a) acetic acid, (d) total SCFA in CD children (n=60) on GFD at least one year 114
Figure 4.31: Spearman rank correlation between faecal pH and the (a) concentration of total SCFA (μmol/g) in wet matter, (b) relative abundance of acetic acid, (c) concentration of propionic acid (μmol/g) in wet matter, (d) relative abundance of butyric acid of CD children (n=60) on GFD at least one year 115

Figure 4.32: Spearman rank correlation between the faecal water content (%) and the concentration of (a) acetic acid in dry matter (µmol/g), (b) propionic acid in dry matter (µmol/g), relative abundance of (c) isobutyric acid (%), and (d) isovaleric acid (%) of CD children (n=60) on GFD at least one year...116

ACKNOWLEDGMENTS

First and foremost, I wish to express my sincere gratitude to my supervisors Dr Konstantinos Gerasimidis, Professor Christine Edwards and Dr Umer Zeeshan Ijaz for offering me this opportunity and for their help, support, encouragement, unlimited patience and supervision throughout the duration of this project.

I would like to thank Alex and Ben for helping during this bioinformatic “adventure” and all my lab mates, colleagues and friends in the New Lister Building for being always there and supporting me – Vaios, Michael, Eleni, Margarita and Rodanthi, thank you!

Of course, a huge thanks to my family, who have been incredibly supportive, my friends Neesha and Danae, and my “flatmates” Konstantina and Marianna for making my day, and smoothie...

AUTHOR'S DECLARATION

I declare that the recruitment, collection of samples, measurement of faecal water content, pH, ammonia, hydrogen sulphide, and D-, L-lactate, genomic DNA extraction, amplification of the 16S rRNA gene and analysis of the dietary habits were part of the work of a previous PhD student. During my MSc (by Research), I extracted and measured faecal SCFA and GIP. After following a three-week course on metagenomics for microbial communities, I performed bioinformatic analysis to produce bacterial taxonomic assignments (OTU table) that I later used in RStudio for the multivariate statistical analysis. This dissertation has not been submitted for any other degree at the University of Glasgow or any other institution.

Konstantina Zafeiropoulou, BSc

I certify that the work reported in this thesis has been performed by Konstantina Zafeiropoulou and that during the period of study she has fulfilled the conditions of the ordinances and regulations governing the Degree of MSc (by Research), University of Glasgow.

Dr Konstantinos Gerasimidis, BSc, MSc, PhD, FHEA

ABBREVIATION LIST

CD	Coeliac disease
UCD	Untreated coeliac disease
TCD	Treated coeliac disease
HC	Healthy controls
GFD	Gluten free diet
GCD	Gluten containing diet
SCFA	Short chain fatty acid
BCFA	Branched chain fatty acids
GIP	Gluten immunogenic peptide
PedsQL-GSS	Paediatric quality of life - gastrointestinal symptom scale (score)
FFQ	Food frequency questionnaire
Ttg	Tissue transglutaminase
HLA	Human leukocyte antigen
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
RCT	Randomised clinical trial
GI	Gastrointestinal
BMI	Body mass index
LOQ	Level of quantification
OTU	Operational taxonomic unit
ASV	Amplicon sequencing variants
GLM	General linear model
NMDS	Non-metric multidimensional scaling plot
LCBD	Local contribution for β diversity
EAR	Estimated average requirement
TDEI	Total daily energy intake
RNI	Reference nutrient intake
DRV	Dietary reference value
SFA	Saturated fatty acids
PUFA	Polyunsaturated fatty acids
MUFA	Monounsaturated fatty acids
NMES	Non-milk extrinsic sugars
NSP	Non-starch polysaccharides

CHAPTER 1 Coeliac Disease, Gluten Free Diet and the Gut Microbiota

1.1 Coeliac disease

Coeliac disease (CD) is an immune-mediated systemic disorder triggered by gluten ingestion in genetically susceptible individuals [2]. Gluten is a protein complex in the endosperm of the grains wheat, barley and rye, and mainly consists of gliadin monomers and glutenin polymers [3]. Gliadin' s fractions (rich in proline and glutamine, and hence resistant to enzymatic digestion) accumulate in the small intestine and trigger an abnormal innate and adaptive immune response in genetically predisposed subjects [4].

CD is characterized by destruction of the small intestinal villi [5]. According to the ESPGHAN guidelines (2012), CD is diagnosed based on CD specific antibodies, human leukocyte antigen (HLA)-DQ2 and HLA-DQ8 haplotypes, small bowel biopsy findings and gluten dependent enteropathy. The clinical manifestations of CD may be either gastrointestinal (GI) or extra-intestinal. The typical GI symptoms are diarrhoea and abdominal distension [6]. The prevalence of atypical extra-intestinal symptoms, such as headache, ataxia and psychiatric disorders, anaemia, encephalopathy, brain stem dysfunction, decreased bone density and neuropathy, have been increasing over the past decades [7], with up to 8.3% of patients exhibiting short stature in the absence of GI symptoms [8].

1.1.1 Prevalence

CD is relatively common and prevalence in the general population is close to 1% [9]. Estimation of the exact prevalence of CD is complicated by the fact that many patients suffer from subclinical CD that often remains undiagnosed. However, a novel serogenetic approach has recently been suggested [10]. Several other conditions, ranging from autoimmune diseases, such as Type 1 Diabetes Mellitus, to chromosomal aberration disorders (Turner syndrome or Down syndrome), are also associated with CD. The estimated prevalence of CD among people with Type 1 Diabetes Mellitus, based on 21 studies in children and 7 studies in adults, has been reported to range from 1- 12% when diagnosed using serology and from 1- 11% when diagnosed using intestinal biopsy [9].

1.1.2 Pathogenesis

The exact pathogenesis of CD is not clear yet. It is evident that the aetiology of CD is multifactorial, with CD being the outcome of the interaction between both genetic and environmental factors [11]. However, further research is needed to explore the exact interactions between factors implicated in the development of CD.

1.1.2.1 Genetic factors

There is a strong genetic component based on the major histocompatibility class II HLA-DQA and DQB genes, in the short arm of chromosome 6 [2]. More than 95% of CD patients share the HLA-DQ2 heterodimer, while the majority of the rest share the HLA-DQ8 [2]. Subjects homozygous for HLA-DQ2 exhibit a 5-fold greater risk of developing CD compared to those that are heterozygous [12]. Moreover, genome wide association studies have indicated 39 non-HLA loci to be associated with CD development to a lesser extent; they explain only 5% of the risk for CD [13], while the HLA loci account for 35% of the risk [14]. Taken as a whole, the genetic susceptibility for CD has been reported in 30-40% of general population in the Western world, although this frequency is population dependent [15]. However, only a small percentage of genetically susceptible individuals develop CD, indicating that environmental factors play an important role.

1.1.2.2 Environmental factors

The involvement of environmental factors in the development of CD is further illustrated by the relatively short time frame in which the incidence of CD has been increasing over the past decades, particularly in the Western world [1]. The slow long-term changes in the human genome cannot explain this increase, nor can changes in gluten consumption [1]. Likewise, it seems that, other than gluten intake, additional environmental factors are involved in the development of CD.

Indeed infant feeding, weaning practices, and modalities of gluten introduction in diet (timing and amount) have been identified as potential risk factors for CD [11]. A protective effect of breastfeeding [16], and cautious introduction of gluten in diet at four to six months [17] was suggested approximately ten years ago, but recent studies (i.e. Prevent CD study [18]) do not support this hypothesis.

A possible involvement of viral infections, especially GI infections, in the development of CD has also been suggested [11]. Infectious episodes in the GI tract may increase the gut permeability, and expand antigen penetration, increasing the risk for development of CD [11]. Children with a history of repeated infections were characterised by an increased risk to develop CD, regardless of infant feeding practices and socioeconomic status [19].

Finally, increasing evidence supports a role of the gut microbiota in the pathogenesis of CD. Many of the multifactorial diseases, which show a similar increasing incidence [i.e. intestinal bowel disease (IBD)], have been associated with distinct changes in the microbial community structure and composition of the gut microbiota (termed “dysbiosis”). Thus over the past decade research has also explored the role of the microbiota in the development of CD. However, the results (discussed in detail in section 1.2) are still inconclusive.

1.1.3 Treatment

A lifelong gluten-free diet (GFD) is the only proven effective treatment for CD [20], but the presence of gluten in many food products, even in trace amounts, makes strict adherence to GFD a challenge for CD patients. Currently, gluten-free products, as defined by the revised Codex Alimentarius (2008), include items with concentrations below 20 ppm or "foods specially processed to reduce gluten content to a level above 20 to 100 mg/kg". However, even though determination of gluten levels in food products is feasible [21] and thresholds have been set with regard to gluten-free products, gluten contamination is a frequent issue. Overall, a wide range from 17- 80% of CD patients do not comply to a GFD [22].

1.2 The role of gut microbiota in Coeliac disease

1.2.1 Gut microbiota

The gut microbiota is an extremely dense and diverse ecosystem consisting of 10^{14} metabolically active bacterial cells, which has evolved with us [23, 24]. The gut microbiota plays an important role in health and disease via continual "cross-talk" between the bacteria and intestinal epithelial cells [25]. It participates in the host's metabolism by improving energy yield from food and ferments non-digestible carbohydrates including cellulose, xylans and resistant starch, which reach the colon undigested, to short chain fatty acids (SCFA) [26]. The gut microbiota enhances the host's colonization resistance, competing with pathogens for space, nutrients and host receptors [27]. It influences the immune system, and the development and maturation of gut homeostasis. This was clearly illustrated by comparative studies of germ-free and conventionally colonized animals. Colonizing the intestine of gnotobiotic animals with single species affected both epithelial cell function and the composition of gut-associated lymphoid tissue [28]. Collectively the bacterial genome (microbiome) consists of at least 100 times [29] as many genes as our own genome (20,000 genes coding proteins), and provides us with traits and features we have not evolved on our own [30].

The gut microbiota is composed of strict anaerobes that dominate the facultative anaerobes and aerobes by two to three orders of magnitude [31]. Bacteroidetes and Firmicutes are the two main phyla, while Proteobacteria, Actinobacteria, Fusobacteria and Cyanobacteria explain most of the remaining diversity [32]. The number of bacteria increases progressively along the GI tract from stomach (10^{1}- 10^{3} bacteria per gram of content) to small intestine (10^{4}- 10^{7} respectively) and then colon (10^{11}- 10^{12}) [33], while the microbial composition is different in the intestinal mucosa compared with intestinal lumen [31].

Gut microbial composition is determined by environmental and genetic factors, with infant colonization setting the stage for the adult microbiome [34]. Table 1.1 lists (on the left side) the influencing factors of early gut colonization, including environmental factors (i.e. delivery mode, breastfeeding) and genetic effects. Accordingly, on the right side of the table 1.1 factors that "cross-talk" with the gut microbiota in an interdependent manner are shown.

Table 1.1: Potential factors influencing early gut colonization (left side), and the cross-talk between the microbiome and host genetics, immunity, metabolism and nutrition (right side). Table adapted from Houghteling et al., 2015 [34].

Initial exposures	Ongoing and interdependent "cross-talk"
Genetics	Gene expression
Prenatal maternal exposures	Mucus and barrier function
Delivery mode	Immune function
Breast feeding	Metabolism
Perinatal antibiotics	Nutrition
Weaning process and timing	Nutrients for mucosa

1.2.2 Is the gut microbiota involved in the development of CD?

Recent evidence points to the possible role of the gut microbiota in the development of CD. The gut microbiota, as mentioned above, influences the mammalian immune system, and disturbances in its structure and complexity result in dysregulation of the immune system, that may underlie disorders such as IBD [35]. In order to clarify whether the gut microbiota is also involved in the proinflammatory pathways of CD, the effect of faeces of both treated (TCD) and untreated CD (UCD) patients on cytokine production in peripheral blood mononuclear cells in vitro was determined. Samples from TCD and UCD patients induced significantly higher (proinflammatory) TNF-α production and significantly lower (anti-inflammatory) IL-10 production compared with faecal samples of healthy controls (HC) [36]. These findings suggest a possible contribution of the microbiota of CD patients to the inflammatory milieu characteristic of the disease. However, further studies are needed to explore this effect and evaluate the role of gut microbiota in the multifactorial aetiology of the development of CD.

1.2.3 Literature review of studies on faecal microbiota in CD

In this thesis, a thorough literature search was carried out in MEDLINE (PubMed), using the search engine NCBI, to identify original articles characterising gut microbiota composition and metabolic activity in faeces of CD patients. The search terms used were (*Coeliac* OR *Celiac*) AND (*microbiota* OR *microbiome* OR *bacteria* OR *microflora* OR *microb** OR *bacter** OR *metabol**), and the study inclusion criteria were (1) human studies with participants of any age and gender, (2) original articles of prospective, cross-sectional or interventional design, and (3) studies that investigated faecal microbiota composition and/ or metabolic activity in CD. Table 1.2 summarises studies looking at the faecal microbiota of CD patients, while Table 1.4 (discussed in detail in section 1.2.6) summarises the studies looking at the microbiota metabolic activity of CD patients. Three studies [37-39] that characterised both faecal microbiota composition and metabolites were included in both Tables 1.2 and 1.4.

The faecal microbiota of UCD patients and HC were characterized by nine research groups. All but four studies reported an "unhealthy" UCD microbiota with reduced putative beneficial species like *Bifidobacterium* and *Lactobacillus* [38, 40-43]. One study found no difference between the two groups in *Bifidobacterium* and *Lactobacillus* numbers [44], while another study found that there was no difference in *Lactobacillus* diversity between UCD and HC, but that *B. bifidum* was significantly increased in UCD [37]. The remaining two studies, from the same research group, did not measure bacteria belonging to *Lactobacillus* and *Bifidobacterium* [45, 46]. UCD had not only significantly lower numbers [38, 41, 42] and relative abundance [43] of *Bifidobacteria* than HC, but also lower *Bifidobacterium* diversity than HC [40]. Similarly, UCD had significantly fewer cultivable lactic acid bacteria [38] and *L. casei* than HC [40].

Characterisation of the bacterial composition of UCD and HC revealed increased numbers of *Staphylococci* [42, 44, 46] and *Clostridia* in UCD compared with HC [38, 42, 44]. Collado et al (2007) showed that UCD had significantly higher numbers of *C. histolyticum* than HC [44]. In contrast, De Palma et al (2010) found a significantly lower relative abundance of *C. histolyticum* and *C. lituseburense* in faecal samples of UCD than HC [43]. As a whole, the bacterial taxa belonging to *Bacteroidetes* were significantly higher in UCD compared with HC, with more *Bacteroides* and *Prevotella* species [38, 44].

The faecal microbiota of TCD patients had significantly lower numbers [42] and diversity [46] of *Staphylococci* than UCD, suggesting the effect of treatment with GFD. However, this evidence has been ambiguous, since Di Cagno et al (2009) found significantly higher *Staphylococcus* numbers in TCD than UCD [38]. *Bifidobacterium* and *Lactobacillus* were significantly lower in TCD than UCD [37, 41], indicating a microbiota with reduced beneficial species in CD subjects after treatment with GFD. Only *B. longum* numbers were significantly increased in TCD compared to UCD patients [41], along with increased numbers of *Enterobacteriaceae* [38].

Three studies reported significantly lower numbers of *Bifidobacteria* and *Lactobacilli* in faecal samples of TCD compared with HC [37, 38, 47], confirming the decline of beneficial bacteria after treatment with GFD. Conversely, three other studies found significantly increased numbers of *Bifidobacteria* and *Lactobacilli* in TCD compared with HC [39, 42, 48], showing once more that the results of studies on faecal microbiota in CD are inconclusive.

The evidence from studies evaluating the effect of GFD on *Staphylococcus* species has also been inconsistent. Sanchez et al (2012) found that *S. epidermidis* was more frequently recovered in TCD than HC, while *S. Warneri* was less frequently isolated in TCD than HC [46]. As a whole *Staphylococcus* diversity [46] and bacterial counts did not differ between the two groups [42], but Di Cagno et al (2011) observed significantly higher numbers of *Staphylococci*/ *Micrococci* in TCD than HC [39]. The same research group found that TCD had significantly lower numbers of *Staphylococci*/ *Micrococci* than HC in another study [38]. Likewise, the evidence of studies looking at the faecal microbiota in CD and the effect of GFD is inconclusive, and the need for further research is evident.

Table 1.2: Primary evidence from human studies exploring the faecal microbiota composition in patients with Coeliac disease

Authors, year	Study design, population	Methods	Outcome measurements	Results
Pisarello et al., 2015 [47]	Cross – sectional; Children; (1) Asymptomatic TCD (negative serology & normal histology) on GFD min 6 months: n=15; (2) HC: n=15	(1) CMT for total aerobes & anaerobes; (2) Isolation of *Lactobacillus* strains (Lb) using cross-streak method & biochemical identification of Lb	(1) CFU/ g faeces; (2) identified strains using its biochemical profile	(1) TCD: ↓ Lb **Vs** HC; (2) Isolation of 5 Lb strains with different autoaggregation, hydrophobicity & viability values
Olivares et al., 2014 [49]	RCT for 3 months; Children; Symptomatic UCD: n=33; Group 1: on GFD + *Bifidobacterium longum* CECT 7347 (10^9 CFU/capsule per day): n=17; Group 2: on GFD + placebo (skimmed milk with 30% sucrose & 0.5% vitamin C): n=16	Real-time qPCR for total bacteria, *Bacteroides fragilis* group, Enterobacteriaceae, *Lactobacillus* group, *Bifidobacterium* spp., *Clostridium coccoides* group, *Clostridium leptum* group;	log gene copies/ g faeces	<u>Differences within group</u> (1) GFD + placebo, t=3 months: ↑ *Bacteroides fragilis* group, ↑ *Enterobacteriaceae* & ↓ (*Lactobacillus* group + *Bifidobacterium* spp. / *B. fragilis* + *Enterobacteriaceae)* **Vs** GFD + placebo, t=0; <u>Differences between group, t=3 months</u> (2) GFD + placebo: ↑ *B. fragilis* group **Vs** GFD + *B. longum* CECT 7347
Golfetto et al., 2014 [48]	Cross – sectional; Adults; (1) Asymptomatic TCD on GFD min 2 years: n=14; (2) HC: n=42	(1) CMT for *Bifidobacterium*; (2) Measurement of faecal pH	(1) CFU/g faeces; (2) faecal pH	(1) TCD: ↑ CFU *Bifidobacterium* /g faeces **Vs** HC; (2) faecal pH: no significant difference between TCD and HC groups
Nistal et al., 2012 [37][a]	Cross – sectional; Adults; (1) Symptomatic UCD: n=10; (2) TCD (negative serology & normal histology) on GFD min 2 years: n=11;	PCR-DGGE using primers for total bacteria, *Lactobacillus* & *Bifidobacterium*	(1) Number of PCR amplicon bands visualised on gel; (2) Diversity (based on the number of	(1) UCD & HC on GCD: ↑ *Lactobacillus sakei* & ↑ *Lactobacillus* diversity **Vs** TCD; (2) UCD: ↑ *B. bifidum* **Vs** HC on GCD; (3) UCD: ↑ *B. catenulatum* **Vs** TCD; (4) HC on GCD: ↑ *Bifidobacterium* diversity **Vs** TCD;

	(3) Asymptomatic HC (negative serology & non DQ2/ DQ8 HLA-DQ phenotype on GCD): n=11 (10/11 HC: on GFD for 1 week)		bands visualised on gel); (3) Community clustering	(5) Faecal microbiota community structure of HC clustered separately to UCD
Sanchez et al., 2012 [46][b]	Cross – sectional; Children; (1) UCD: n=20; (2) TCD on GFD min 2 years: n=20; (3) HC: n=20	(1) CMT; (2) Identification of *Staphylococcus* spp. using (i) multiplex PCR, (ii) sequencing of 883bp fragment (by *dnaJ* primers), or (iii) 16S RNA sequencing	Abundance of isolated species	(1) TCD & UCD: ↑ *S. epidermidis,* ↓ *E. faecium,* **Vs** HC; (2) UCD: ↑ *S. haemolyticus* **Vs** HC; (3) UCD: ↑ *Staphylococcus* spp. diversity, ↓ *S. aureus,* ↓ *Enterococcus spp.* **Vs** TCD & HC; (4) TCD: ↓ *S. warneri* **Vs** UCD & HC;
Di Cagno et al., 2011 [39][c]	Cross – sectional; Children; (1) Asymptomatic TCD on GFD min 2 years: n=19; (2) Control group (patients with negative for CD endoscopy & further GI investigation): n=15	(1) CMT for total anaerobes *Lactobacillus*, *Bifidobacterium*, Staphylococcus & micrococci, enterococci, *Bacteroides, Porphyromonas, Prevotella, Clostridium* & enterobacteria; (2) PCR-DGGE using primers for total bacteria, *Bifidobacterium* & *Lactobacillus*	(1) Community (UPGMA) clustering; (2) log CFU/g faeces	(1) No microbiota clustering between TCD & HC; (2) *Lactobacillus*: All but one TCD clustered separately to HC; (3) *Bifidobacterium*: most of TCD clustered separately to HC; (4) TCD: ↓ total anaerobes, ↓ *Bifidobacterium*, ↓ *Lactobacillus*, ↓ enterococci, ↑ *Bacteroides,* ↑ *Porphyromonas,* ↑ *Prevotella,* ↑ Staphylococcus / micrococci, ↑ *Enterobacteria* **Vs** HC;
De Palma et al., 2010 [43][b]	Cross – sectional; Children; (1) Symptomatic UCD: n=24; (2) Asymptomatic TCD (negative serology & normal histology), on GFD min 2 years: n=18; (3) HC: n=20	FISH & Flow cytometry for total bacteria, *Bifidobacterium*, *Lactobacillus/ Enterococcus, Bacteroides/ Prevotella, Escherichia coli, C. histolyticum, C. Lituseburense, F. Prausnitzii,*	Relative abundance (Proportion: bacterial cells hybridising with group-specific probes / total bacteria hybridising with EUB probe 338);	UCD: ↓ Gram-positive bacteria, ↓ *Bifidobacterium*, ↓ *C. histolyticum,* ↓ *C. Lituseburense*, ↓ *F. Prausnitzii,* ↑ *Bacteroides - Prevotella* **Vs** HC;

		SRB & *Staphylococcus*		
Collado et al., 2009 [42][b]	Cross – sectional; Children; (1) Symptomatic UCD: n=30; (2) TCD (negative serology & normal histology) on GFD min 2 years: n=18; (3) HC: n=30	Real-time qPCR for: total bacteria, *Bifidobacterium* group, *Bacteroides* group, *Clostridium coccoides* group, *Clostridium leptum* group, *E. coli, Staphylococcus, Lactobacillus* group, *Akkermansia municiphila*	(1) prevalence (positive amplification/ total samples analysed by PCR); (2) bacterial counts (log cells/g faeces);	<u>Bacterial prevalence</u> (1) UCD & TCD: ↑ *E. coli* **Vs** HC; (2) UCD: ↑ Staphylococcus **Vs** HC; <u>Bacterial counts</u> (3) UCD & TCD: ↑ total bacteria, ↑ Bacteroides, ↑ C. *Leptum*, ↓ *Bifidobacterium* **Vs** HC; (4) UCD: ↑ *E. coli* **Vs** HC; (5) UCD: ↑ *Staphylococcus* **Vs** HC & TCD; (6) TCD: ↑ *Lactobacillus* group **Vs** HC
Di Cagno et al., 2009 [38][c]	Cross – sectional; Children; (1) Symptomatic UCD: n=7; (2) Asymptomatic TCD, on GFD min 2 years: n=7; (3) HC (siblings of TCD): n=7	(1) CMT for total anaerobes, *Lactobacillus*, *Bifidobacterium*, Staphylococcus & micrococci, enterococci, *Bacteroides, Porphyromonas, Prevotella, Clostridium* & enterobacteria; (2) PCR-DGGE using primers for total bacteria, *Bifidobacterium*, *Lactobacillus*, *Enterococcus, Clostridium* & *Bacteroides*	(1) Community (UPGMA) clustering; (2) log CFU/ gr faeces	<u>Microbial community structure</u> (1) No clear microbiota clustering; (2) *Lactobacillus* community: TCD & HC clustered together (in 3 clusters), UCD did not group with any cluster; (3) *Bifidobacterium* community: All but one of HC & UCD grouped together <u>Bacterial CFU</u> (4) TCD & UCD: ↓ *Bifidobacterium,* ↑ *Bacteroides,* ↑ *Clostridium* **Vs** HC; (5) Staphylococcus/ micrococci: UCD < TCD < HC; (6) *Enterobacteriaceae*: UCD < HC < TCD ; (7) *Lactobacillus*: HC > TCD > UCD
Collado et al., 2008 [41][b]	Cross – sectional; Children; (1) UCD: n=30; (2) TCD on GFD min 2 years: n=18; (3) HC (biopsy examination: normal	Real-time qPCR	(1) Prevalence (number of positive amplification from total samples analysed by	<u>Prevalence</u> (1) UCD & HC: ↓ *B. adolescentis* **Vs** TCD; (2) TCD: ↑ *B. dentium* **Vs** HC; <u>Bacterial counts</u>

	villous structure): n=30		PCR); (2) bacterial counts (log cells/ gr faeces)	(3) UCD & TCD: ↓ *Bifidobacterium* group, ↓ *B. longum* **Vs** HC; (4) UCD: ↑ *B. adolescentis* **Vs** HC; & ↓ *B. longum*, ↑ *B. breve*, ↑ *B. bifidum* & ↓ *B. longum* **Vs** TCD (5) TCD: ↓ *B. breve* & ↓ *B. bifidum* **Vs** HC;
Sanchez et al., 2008 [45][b]	Cross – sectional; Children; (1) Symptomatic UCD: n=10; (2) Asymptomatic TCD on GFD 1-2 years: n= 10; (3) HC: n= 11	CMT for *Enterobacteriaceae*	Number of *E. coli* and *non - E. coli* colonies	TCD: ↑ non - *E. coli Enterobacteriaceae* colonies **Vs** HC
Collado et al., 2007 [44][b]	Cross – sectional; Children (max 4 years old); (1) Symptomatic UCD: n=26; (2) HC: n=23	(1) CMT for total anaerobes, *Bifidobacterium, Clostridium, Bacteroides, Enterococcus, Enterobacteriaceae, Staphylococcus, Lactobacillus;* (2) FISH	(1) log CFU/ g wet faeces (2) log cells/ g wet faeces	(1) UCD: ↑ log CFU/ g wet faeces of the genera *Bacteroides*, *Clostridium* & *Staphylococcus* **Vs** HC; (2) UCD: ↑ log cells/ g wet faeces of SRB, *Clostridium histolyticum* and the groups *Eubacterium rectale – C. coccoides*, *Bacteroides – Prevotella* & the Atopobium group **Vs** HC;
Sanz et al., 2007 [40][b]	Cross – sectional; Children (max 4 years old); (1) Symptomatic UCD: n=10; (2) HC: n=10	PCR-DGGE using universal primers for total bacteria, *Lactobacillus*, *Bifidobacterium* & sequencing of bands	(1) Number of bands visualised on gel; (2) Diversity (based on the number of bands visualised on gel)	(1) UCD: ↓ *Lactobacillus casei* group, ↑ *Lactobacillus curvatus,* ↑ *Leuconostoc mesenteroides,* ↑ *Leuconostoc cornasum,* ↓ *B. dentium* & ↓ *B. adolescentis* **Vs** HC; (2) UCD: ↑ Diversity of total bacteria & ↓ diversity of *Bifidobacteria* **Vs** HC

↑: significantly increased levels, ↓: significantly decreased levels (p-value< 0.05); UCD: untreated Coeliac disease; TCD: treated Coeliac disease; HC: healthy controls; GFD: gluten free diet; GCD: gluten containing diet; CMT: conventional microbiology techniques; (q)PCR: (quantitative) polymerase chain reaction; FISH: fluorescent *in situ* hybridization; DGGE: denaturing gradient gel electrophoresis; RAPD-PCR: randomly amplified polymorphic DNA-PCR; CFU: colony forming unit; SRB: sulphate reducing bacteria; [a, b, c]: Studies with the same superscript letter were published by the same or collaborating research groups

1.2.4 Faecal microbiota in CD. How good is the evidence? What is missing?

The evidence of studies looking at the gut microbiota community structure and complexity of faeces of patients with CD is inconclusive. This may be due to differences in the sample size, control group, study design, and technique used to characterise the gut microbiota in each study.

Indeed, looking at the sample size, there was a relatively wide range from 20 [40] to 78 participants [41, 42]. However, looking at differences in the control group all but one study that looked at the gut microbiota of CD patients compared the results with HC. Only in one study, the control group consisted of patients who were negative for CD endoscopy and underwent further investigation for GI disease [39], suggesting a small contribution of differences in the sample size and control group to differences in the results of the studies.

Overall, thirteen studies have evaluated the gut microbiota of faecal samples of CD children so far. All but one study were of cross-sectional design, which provides a snapshot of the gut microbiota structure and composition in UCD, TCD and HC subjects, but does not determine whether a disturbed microbiota in faeces of UCD patients has contributed to disease development or is merely a consequence of ongoing disease. Accordingly, this cannot determine whether an altered microbiota in faeces of TCD patients compared with UCD is due to compliance with GFD and disease improvement, since the use of independent UCD and TCD groups in the cross-sectional studies does not account for the inter-individual variability seen in the gut microbiota of distinct subjects [32]. Differences in the gut microbiota structure and composition of TCD compared with UCD patients may be due to various factors (i.e. differences in the age, delivery mode and breastfeeding practices), regardless of disease activity and treatment with GFD. Only prospective studies, with serial samples from the same CD subjects collected at diagnosis and throughout the treatment with GFD, could address these confounders.

So far, only one study has explored the effect of treatment with GFD on the gut microbiota of faeces of newly diagnosed CD children in a prospective manner; this was a randomised clinical trial (RCT) [49]. In this study, faecal samples of children with CD were collected at diagnosis and at three months on GFD. However, treatment with GFD for three months is not enough to achieve either normal villous height, which needs approximately (median) 3.8 years [50], nor normal serum tTG levels, which usually takes up to one year [51]. Likewise, larger prospective studies that follow CD cohorts from diagnosis to at least one year on GFD (required for resolution of inflammation) are needed to explore the effect of GFD and disease improvement on the microbiota community structure and composition. To the best of our knowledge, the study described in section 4.3 of this thesis is the first study where faecal samples of CD children were collected at six and 12 months on GFD to assess the effect of treatment with GFD on the gut microbiota.

Moreover, most of the studies that have looked at the faecal microbiota of patients with CD so far have used either culture dependent techniques or culture-independent, but primer- and probe-

dependent techniques to characterize the gut microbiota. Specifically, seven of thirteen studies used culture-based conventional microbiological techniques (CMT) that may be cheap, but are also labour intensive and offer a limited view of the microbiota diversity considering that less than 30% of the microbiota members have been cultured to date (Table 1.3) [32]. Three studies used real-time quantitative polymerase chain reaction (qPCR), which bypasses the need for lab cultivation, is fast and accurate to measure the total microbial load. However, qPCR is also dependent on the available primers, and hence is unable to identify unknown species [52]. Four studies used PCR combined with denaturing gradient gel electrophoresis (DGGE). This culture-independent technique is also fast, with multiple samples being analysed simultaneously, and offers a more detailed view of the diversity and abundance of the gut microbiota compared with qPCR. However, it does not offer direct phylogenetic identification, unless sequencing or probe hybridisation is applied (Table 1.3) [52]. Finally, two studies used fluorescent *in situ* hybridization (FISH), which is based on the hybridization of specific oligonucleotide probes, and hence unable to identify unknown species.

Table 1.3: Strengths and weaknesses of the techniques used to characterise the gut microbiota. Table adapted from Fraher et al (2012) [52]

Technique	Strengths	Weaknesses
Culture E.g. selective plating and Miles Misra dilution	○ cheap ○ semi-quantitative	○ labour intensive ○ limited view of the diversity; <30% of gut microbiota have been cultured so far
real-time qPCR	○ fast ○ quantitative ○ phylogenetic identification	○ primer-dependent ○ unable to identify unknown species
PCR-DGGE	○ fast ○ semi-quantitative	○ primer-dependent ○ no direct phylogenetic identification
FISH	○ phylogenetic identification ○ semi-quantitative	○ probe-dependent ○ unable to identify unknown species ○ if used with flow cytometry may have bias
16S rRNA sequencing	○ quantitative (bacteria in low abundance can be detected) ○ phylogenetic identification ○ identification of unknown species	○ expensive ○ needs intense bioinformatic analysis ○ need to avoid contamination in DNA extraction and amplification
Whole genome shotgun sequencing	○ sequencing of the whole community DNA ○ phylogenetic identification ○ quantitative	○ very expensive ○ need for intense bioinformatic analysis ○ need to avoid contamination in DNA extraction and amplification

Taken as a whole the techniques that have been used to characterize the gut microbiota of faeces of CD patients lack power to explore the whole microbiome and identify unknown species, due to culture-, primer- or probe- dependency. Sequencing of the microbiome is the current gold standard for taxonomic identification to species level [52]. Therefore, the need for studies that characterize the gut microbiota of CD patients using next-generation sequencing technologies (NGS), including 16S rRNA sequencing and shotgun sequencing, is apparent. 16S rRNA sequencing provides quantitative data and phylogenetic identification of both known and unknown species [52]. Accordingly, whole genome shotgun sequencing describes the whole community DNA, allowing us to study metabolic pathways in the community and predict the microbiota function [53] (Table 1.3).

The interpretation of this large amount of data generated by both 16S rRNA sequencing and whole genome shotgun sequencing requires computationally intense bioinformatic analysis (Table 1.3), which limits the usefulness of these techniques in non-specialist laboratories. However, these techniques provide the most powerful data to characterise in detail the gut microbiota of patients and potentially improve our understanding of various multifactorial diseases processes, including both the pathogenesis and treatment. Therefore, NGS technologies should be used in research laboratories, that bring together expertise of different disciplines (including bioinformaticians), to characterise the gut microbiota of faeces of patients with CD. Results from such a study (pursued and described in this thesis for first time) would increase our knowledge of the faecal microbiota characteristics in CD.

Nevertheless, in order for NGS technologies to allow us to unravel the role of the gut microbiota in the development of CD and to potentially establish causality, a suitable study should be designed. Cross-sectional study designs cannot determine the time ordering of any observed association, and hence we cannot establish any causality. Following healthy subjects into the future to record whether they develop CD, and then compare characteristics of the gut microbiota of faecal samples collected throughout the follow-up would ideally address this problem. Once differences in the gut microbiota were observed before the onset of CD, causality could be established. However, considering the prevalence of CD (approximately 1%), and the high cost to characterise the gut microbiota using NGS technologies, this kind of study would require an enormously large sample size and would be both time- and money- consuming.

Instead, the use of healthy siblings of patients with CD could reduce the need for such an expensive and challenging study. Siblings have not only a more similar genome than non-related people, but have also been exposed to similar environmental factors (i.e. diet). Once differences between independent UCD and HC individuals are replicated in the comparison between siblings, association between an altered microbiota and CD can be suggested more confidently. Moreover, siblings of patients with CD have 20 to 60 times higher risk for CD than general population [54]. This means that a smaller sample size would be required to pursue a prospective study (as described above) within a cohort of siblings of CD patients. By following siblings of patients with CD prospectively, we account not only for the inter-individual diversity, but also for genetic susceptibility

and, as mentioned above, environmental influences. Therefore, gradual changes in the gut microbiota structure and composition throughout the follow-up, until the onset of CD, would allow us to establish causality between the microbiota "dysbiosis" and the development of CD, whereas differences, compared with the baseline value, observed only after the onset of CD would confirm the effect of disease activity on the gut microbiota.

The observed microbiota "dysbiosis" could be used as a biomarker to diagnose CD either in subjects that suffer from subclinical CD, and usually remain undiagnosed, or in subjects that suffer from atypical extra-intestinal symptoms, and are usually diagnosed many years after the development of CD. Accordingly, alterations in the gut microbiota observed in a prospective study after the compliance with GFD could be used to evaluate the adherence to a GFD and monitor the success of the treatment.

1.2.5 Microbiota metabolic activity

1.2.5.1 Short chain fatty acids

SCFA are organic fatty acids (C2- 8) produced by bacterial fermentation of carbohydrates that reach the colon undigested. Acetic acid (C2) is predominant in the colon followed by propionic (C3) and butyric acid (C4) [55]. They are energy substrates for both the colonic epithelium and peripheral tissues [26]. The major carbohydrates available for bacterial fermentation in the colon are resistant starch, NSP, and a variety of unabsorbed sugars or oligosaccharides, such as fructose polymers [56]. However, SCFA, and in particular branched chain fatty acids (BCFA), are also produced from protein and endogenous substrates, such as mucus, pancreatic enzymes, and sloughed epithelial cells [57].

SCFA production is determined by three main factors: (a) the number and composition of gut microbiota, (b) the amount and type of available fermentable substrate, with different NSP leading to different (quantity- and quality- wise) SCFA production, and (c) the GI transit time [58]. Following production, SCFA are absorbed by colonic epithelial cells [59]. Butyric acid is oxidised by the epithelial cells and supplies them with energy, while propionic and acetic acid are transferred via the portal blood to the liver and then on to the peripheral tissues [26]. Propionic acid is gluconeogenic and affects lipid and cholesterol synthesis, while acetic is used for energy and de-novo lipogenesis [60]. All SCFA can be used as fuel by colonocytes but with less efficiency than butyric acid [59]. The three main SCFA can exert beneficial effects in colon carcinogenesis, such as stimulating apoptosis [61] and differentiation and inhibiting growth of cancer cells in vitro, with butyric being the most effective [62].

SCFA, especially propionic, also act as signalling molecules; they are ligands for two G protein-coupled receptors of free fatty acids, the Gpr41 and Gpr43 [63]. In the intestine, Gpr43 is involved in regulating appetite and insulin signalling, stimulating secretion of peptide YY, glucagon-like peptide-1 and glucagon-like peptide-2 from the enteroendocrine L cells of the GI tract [64]. Loss of Gpr41 is associated with faster GI transit time and reduced efficiency of energy harvest from the diet [63]. Finally, SCFA are intermediate factors in the diet-microbiota relationship [65]. Fibre consumption

increases SCFA production and lowers the colonic pH that, in turn, affects the gut microbiota composition and prevents overgrowth by pH-sensitive pathogenic bacteria like *Enterobacteriaceae* and *Clostridia* [66, 67].

1.2.5.2 Lactate

Lactate or 2-hydroxypropanoate is another fermentation product of undigested carbohydrates that is mainly produced by *Lactobacillus* and *Bifidobacterium* [68]. It is the simplest hydrocarboxylic acid and exists due to its asymmetric C2 atom in 2 stereoisomers, the D- isomer that is an exception and rotates light counterclockwise (levorotary), and the L-lactate that rotates light in the clockwise direction (dextrorotary) [68]. In mammalian cells, lactate is produced mainly in the latter form; it is only the methylglyoxal pathway through which minute amounts of D-lactate are produced. However, both D- and L-lactate are produced via microbial fermentation in the colon and are then further metabolized to other SCFA [69].

In the case of increased delivery of unabsorbed carbohydrates to the colon, such as in short bowel syndrome, faecal D-, L- lactate levels may be increased compared to HC [70]. Thus, colonic pH decreases while acid-resistant lactic acid producing bacteria increase, making the colonic pH more acid. However, under normal conditions, lactate produced by the bacterial fermentation does not pose an acid-base threat, since it is further bacterially converted to acetic and other SCFA [68].

1.2.5.3 Sulphide

Hydrogen sulphide (H_2S) was defined 100 years ago as a "decomposition product of many sulphur-containing compounds" formed by bacteria from both organic and inorganic sulphur containing substances [71]. In the colonic lumen, there are several bacteria capable of producing H_2S. Sulphate-reducing bacteria (SRB), such as *Desulfovibrio*, *Desulfobacter*, *Desulfomonas*, *Desulfobulbus* and *Desulfotomaculum* genera, combine oxidative phosphorylation with the reduction of sulphate to sulphide [72]. SRB are non-saccharolytic and so sulphur compounds like polysaccharide chondroitin sulphate and colonic mucins require degradation to release the substrate for SRB by clostridia, *Bifidobacterium*, *Bacteroides thetaiotaomicron* and other *Bacteroides fragilis* group members [73].

Faecal sulphide concentration results from the balance between microbial production and removal through detoxification mechanisms [74]. Detoxification occurs after conversion to thiosulfate, a non-volatile metabolite [75]. Therefore, increased concentration of sulphide in faeces may be due to a defect in detoxification mechanisms and not only increased production

H_2S is a toxic compound with potentially adverse effects in the colon that may play a role in the pathogenesis of ulcerative colitis [76], including inhibition of butyric oxidation within the colonocytes [74], and induction of mucosal hyperproliferation [77]. However, H_2S is unlikely to be produced in sufficient amounts to be toxic [78]. Beneficial effects on the GI tract and liver [79], and circulatory system homeostasis have also been demonstrated [73]. Intracolonic administration of a H_2S donor in

rats exerted an hypotensive effect, which lasted longer than parenteral infusions, with a more pronounced decrease in arterial blood pressure in hypertensive rats than normotensive rats [73].

1.2.5.4 Ammonia

Ammonia is one of the main products of bacterial protein breakdown in the large intestine; phenols and indoles are also produced [80]. Substrates may be of dietary origin or endogenous sources, including mucus, host proteins and bacterial proteins [80]. Gram-negative anaerobes, clostridia, enterobacteria and *Bacillus* spp. produce considerably higher amounts of ammonia than streptococci, micrococci and Gram- positive non-sporing anaerobes, e.g. *Bifidobacterium* [81]. Ammonia generated in the colon by bacteria is absorbed and converted into urea in the liver [82], or excreted through faeces.

1.2.6 Literature review of studies on microbiota metabolites in CD

Nine studies have measured the microbiota metabolites in faecal, urine and serum samples of CD subjects so far (including both TCD and UCD in four studies [37, 38, 83, 84], only TCD in two studies [39, 85], and only UCD in three studies [86-88], Table 1.4). Eight studies measured microbiota metabolites in faecal samples [37-39, 83, 84, 86-88], and one of these also measured metabolites in urine [39]. In the ninth study, SCFA were measured only in serum [85].

Microbiota metabolites were compared between UCD and HC in six studies [37, 38, 83, 84, 87, 88], and all but two studies found significantly higher levels of acetic, propionic, butyric, isobutyric, isovaleric and total SCFA in UCD patients than HC [37, 83, 84, 88]. One study found significantly lower total SCFA and higher alcohols, aldehydes, sulphur compounds and hydrocarbons levels in faecal samples of UCD children than HC [38]. The sixth study compared symptomatic and asymptomatic UCD patients for SCFA. It appeared that UCD had higher levels of acetic and total SCFA than HC regardless of symptoms. Symptomatic UCD had higher levels of isobutyric and isovaleric than HC, while asymptomatic UCD had significantly higher levels of isocaproic than HC [87]. Overall, all but one study reported significantly increased levels of bacterial SCFA in UCD compared with HC.

The effect of GFD on microbiota metabolites was evaluated in four observational studies and one RCT, where faecal samples of CD patients were collected at diagnosis and at 12 months on a (typical or enriched with 25-50 g oats per day) GFD. Among the observational studies, all but one found no significant difference between TCD and UCD groups [37, 83, 84]. One study found significantly lower levels of alcohols, aldehydes, sulphur compounds and hydrocarbons, and higher total SCFA in TCD than UCD patients [38]. In contrast, the RCT found that total SCFA significantly decreased one year after the initiation of a typical GFD [86], indicating that the evidence of studies looking at the microbiota metabolic activity of CD patients after the compliance with GFD is inconclusive, and more research is needed to clarify this.

Therefore, the effect of GFD was further evaluated comparing TCD with HC. When there was no difference between the two groups, it was considered that after compliance with GFD, SCFA had normalized back to healthy control levels. That was achieved in only two of six studies that compared TCD with HC [83, 85]. In particular, GFD was effective in restoring SCFA concentration after one year of treatment. Those on GFD for less than one year had significantly higher acetic, propionic, valeric, isobutyric, isovaleric and total SCFA than HC [83]. Results from the remaining four studies were conflicting. Three studies found significantly higher levels of acetic acid in TCD than HC [37, 39, 84]. Two found significantly higher levels of propionic, butyric, valeric, isobutyric, isovaleric and total SCFA in TCD than HC [37, 84], but the third study found significantly lower propionic, butyric, isocaproic and total SCFA in TCD than HC [39]. Finally, one study found significantly higher levels of ketones and lower levels of esters, except for ethyl acetate, ethyl propionate and octyl acetate in TCD compared with HC [38].

Table 1.4: Primary evidence of human studies exploring the gut microbiota metabolic activity in individuals with Coeliac Disease (CD)

Authors, year	Study design, population	Methods	Outcome measurements	Results
Tjellstrom et al., 2014 [f] [86]	RCT; UCD Children (N=71); Group 1: on GFD: n=37; Group 2: on GFD + oats (25-50g/ day): n=34	GLC – FID	Faecal SCFA concentration (mmol/Kg faeces), measured at baseline, 6 & 12 months later	(1) GFD, t=12 months: ↓ total SCFA Vs GFD, t=0; (2) GFD + oats, t=12 months: ↑ acetic, ↑ butyric acid & ↑ total SCFA Vs GFD, t=12months
Jakobsdottir et al., 2013 [85]	Cross – sectional; Adults; (1) TCD: n=15; (2) MC: n=12; (3) non-GI inpatients: n=21	GLC – FID	Serum SCFA concentration (µmol/L)	(1) TCD: no significant difference Vs MC & non-GI inpatients; (2) MC: ↑ valeric acid Vs non-GI inpatients
Tjellstrom et al., 2013 [f] [83]	Cross – sectional; Children; (1) UCD: n=53; (2) TCD on GFD < 1 year: n=74; (3) HC: n=54; Children + adults: (4) TCD on GFD > 1 year: n=25	GLC – FID	Faecal SCFA concentration (mmol/Kg faeces)	(1) UCD & TCD on GFD < 1 year: ↑ acetic, ↑ isobutyric, ↑ isovaleric acid & ↑ total SCFA Vs HC; (2) TCD on GFD < 1 year: ↑ propionic & ↑ valeric Vs HC; (3) TCD on GFD > 1 year: no significant difference Vs HC; (4) TCD on GFD < 1 year: ↑ acetic acid & ↑ total SCFA Vs TCD on GFD > 1 year
Nistal et al., 2012 [a] [37]	Cross – sectional; Adults; (1) Symptomatic UCD: n=10; (2) TCD (negative serology & normal histology) on GFD min 2 years: n=11; (3) Asymptomatic HC (negative serology & CD phenotype on GCD): n=11 (10/11 HC: on GFD for 1 week)	GLC – FID	Faecal SCFA concentration (mmol/Kg faeces)	(1) UCD & TCD: ↑ acetic, ↑ propionic, ↑ butyric acid & ↑ total SCFA Vs HC on GCD & HC on GFD; (2) TCD: ↑ n-valeric acid Vs HC on GFD
Di Cagno et al., 2011 [c] [39]	Cross – sectional; Children; (1) Asymptomatic TCD on GFD min 2 years: n=19;	GC – MS/ SPME; H-NMR;	(1) Faecal SCFA & VOC concentration (ppm); (2) urine metabolites	<u>Faecal samples:</u> (1) TCD: ↑ acetic, ↓ propionic, ↓ butyric, ↓ isocaproic, ↓ total SCFA, ↑ alcohols, ↓ aromatic organic compounds, ↑ proline,

	(2) Control group (patients with negative for CD endoscopy & further GI investigation): n=15	3 samples/ participant	concentration (ppm)	↑ tryptophane, ↑ histidine, ↑ asparagine, ↑ tyramine & ↑ Trimethylamine-N-ox Vs HC; <u>Urine samples:</u> (2) TCD: ↓ carnosine, ↓ glucose, ↓ glutamine, ↑ creatinine, ↑ methylamine, ↑ lysine, ↑ arginine Vs HC
Tjellstrom et al., 2010 [f] [87]	Cross – sectional; Children; (1) Asymptomatic UCD: n=15; (2) Symptomatic UCD: n=36; (3) HC: n=42	GLC – FID	Faecal SCFA concentration (mmol/Kg faeces)	(1) Asymptomatic UCD: ↑ acetic acid, ↑ isocaproic acid & ↑ total SCFA Vs HC; (2) Symptomatic UCD: ↑ acetic, ↑ isobutyric, ↑ isovaleric acid & ↑ total SCFA Vs HC; (3) Asymptomatic UCD: ↑ acetic & ↑ isocaproic acid Vs symptomatic UCD
Di Cagno et al., 2009 [c] [38]	Cross-sectional; Children; (1) UCD: n=7; (2) TCD, on GFD min 2 years: n=7; (3) HC - siblings of TCD: n=7	GC - MS / SPME; 3 samples / participant	Faecal SCFA & VOCs concentration (ppm)	(1) 3 samples/ participant: no intraindividual difference; (2) UCD: ↓ total SCFA, ↑ alcohols, ↑ aldehydes, ↑ sulphur compounds & ↑ hydrocarbons Vs TCD & HC; (3) HC: ↓ ketones & ↑ esters Vs UCD & TCD; exceptions: ethyl acetate, ethyl propionate & octyl acetate
Kopecny et al., 2008 [88]	Cross – sectional;Children; (1) UCD: n=49; (2) HC: n=n/a	GLC – FID	Faecal SCFA	(1) Total SCFA: UCD > HC; (2) pH: no difference between UCD & HC
Tjellstrom et al., 2005 [f] [84]	Cross – sectional; Children; (1) UCD: n=36; (2) TCD on GFD min 3 months: n=47; (3) HC: n=42	GLC – FID	Faecal SCFA concentration (mmol/Kg faeces)	(1) UCD & TCD: ↑ acetic, ↑ isobutyric, ↑ isovaleric & ↑ total SCFA Vs HC; (2) TCD: ↑ propionic & ↑ valeric acid Vs HC; (3) TCD Vs UCD: no significant difference

↑: significantly increased levels, ↓: significantly decreased levels (p-value < 0.05); CD: Coeliac disease; UCD: untreated Coeliac disease; TCD: treated Coeliac disease; HC: healthy controls; GFD: gluten free diet; GCD: gluten containing diet; G(L)C: gas (liquid) chromatography; FID: flame ionisation detection; MS: mass spectrometry; SPME: solid phase micro extraction analysis; NMR: H nuclear magnetic resonance GI: gastrointestinal; MC: microscopic colitis; [a, b, c]: Studies with the same superscript letter were published by the same or collaborating research groups

1.2.7 Microbiota metabolites in CD. How good is the evidence? What is missing?

Similar to the evidence of studies looking at the gut microbiota community structure and composition of faeces of patients with CD, the evidence of studies looking at the microbiota metabolic activity of CD is inconclusive. Table 1.5 shows all results from the nine studies that have measured microbiota metabolites in faecal, serum samples, and urine so far. It should be noted that four of nine studies came from the same research group (Tjellstrom et al). Most of studies suggest an increase in the microbiota metabolites of CD patients compared with HC, regardless of disease activity, but not all studies support this hypothesis. In particular, Di Cagno et al (2011) found significantly lower propionic, butyric and total SCFA in TCD than HC. More research is needed to determine the microbiota metabolic activity in CD.

Table 1.5: Microbiota metabolites in CD (overall presentation)

UCD	**TCD**
Versus HC 1. ↑ acetic [a], ↑ propionic, ↑ butyric, ↑isobutyric [b], ↑isovaleric [b], ↑ total SCFA [a] [37, 83, 84, 88] BUT ↓ total SCFA [38] 2. ↑ alcohols, ↑ aldehydes, ↑ sulphur compounds, ↑ hydrocarbons [38]	Versus UCD 1. no difference [37, 83, 84] 2. ↑ total SCFA [38] BUT ↓ **total SCFA** [86] 3. ↓ alcohols, ↓ aldehydes, ↓ sulphur compounds ↓ hydrocarbons [38] Versus HC 1. no difference (on GFD ≥ 1 year) [83, 85] 2. ↑ acetic [37, 39, 83, 84] 3. ↑ butyric [37, 84] BUT ↓ butyric [39] 4. ↑ propionic, ↑ total SCFA [37, 83, 84] BUT ↓ propionic, ↓ total SCFA [39] 5. ↑ isobutyric, ↑ valeric, ↑ isovaleric [37, 83, 84] 6. ↑ ketones, ↓ esters (except for ethyl acetate, ethyl propionate, octyl acetate) [38]

Tjellstrom et al (2010) [87] found that metabolites with superscript: ([a]) are significantly higher in UCD patients than HC regardless of symptoms, ([b]) are significantly higher only in symptomatic UCD than HC; results in bold refer to pairwise data

Looking at the study sample size, a wide range from 21 [38] to 206 subjects [83] was observed. This in conjunction with differences in the duration of treatment with GFD between TCD subjects could explain differences in the outcomes. GFD duration varied from three months [84] to more than two years [37-39], and, as discussed above, both compliance with treatment and disease activity are different at two years on GFD compared to three months on GFD.

The control group consisted of unrelated healthy subjects in five studies [37, 39, 83, 84, 88], healthy siblings of TCD children in one study [38] and non-GI inpatients in another [85]. Similarly, in the ninth study, the control group consisted of symptomatic UCD and was compared with

asymptomatic UCD patients [87]. Differences in the control group between studies may partially explain differences in the outcomes. Under medication, non-GI inpatients may have different microbiota compared with healthy subjects, and it is not ideal to use the former as a control group. However, the use of healthy sibling of patients with CD as a control group would allow us to explore the role of gut microbiota in CD in a more confident way, since, as mentioned above siblings have not only more similar genome than non-related people, but have also been exposed to similar environmental influences (i.e. diet). Therefore, observed differences in the microbiota (structure and functionality) between healthy siblings and patients with CD, would confirm the existence of an association between the gut microbiota and CD.

Taken as a whole, these studies are of cross-sectional design, except for one RCT where CD children were followed from diagnosis to 12 months on GFD. Therefore, we cannot ascertain whether differences in the microbiota metabolites are involved in the development of CD, or are a consequence of disease activity. Prospective studies are needed to address this question. The study presented in this thesis for first time measured targeted (SCFA) and untargeted metabolites of faecal samples of HC, siblings and patients with CD at diagnosis, and at six and 12 months on GFD. However, the detailed experimental description and results of the measurement of untargeted metabolites are out of the scope of the current thesis and will not be reported.

1.3 General conclusions and the purpose of this thesis

The aetiology of CD is clearly multifactorial, but the exact factors, as well as the interactions between factors implicated in the pathogenesis of CD have yet to be fully elucidated. Increasing evidence points to a role of the gut microbiota in the development of CD. However, the studies that have looked at the gut microbiota and bacterial metabolic activity in faecal samples of CD patients so far are heterogeneous in design, study populations and methods used to characterise the gut microbiota, and their results are inconsistent. Thus, it is still unclear whether alterations in the gut microbiota have been involved in the development of CD or are a secondary consequence of the ongoing disease and the adherence to treatment with GFD. So far, no study has explored the gut microbiota using NGS technologies, that would provide the most detailed characterisation of the whole microbiome.

Collectively the evidence of studies looking at faecal microbial community structure, bacterial taxon relative abundance and bacterial functionality in CD patients is inconsistent and unclear. Further research is needed to explore the gut microbiota in CD patients, compare it with healthy subjects, as well as with healthy siblings of the patients, and evaluate the effect of GFD. An observed microbiota "dysbiosis" in patients with CD could be used as a biomarker to diagnose CD in subjects that suffer from subclinical CD or atypical extra-intestinal symptoms, and either remain undiagnosed or are diagnosed many years after the development of CD. Such a biomarker would facilitate the exact estimation of the prevalence of CD and potentially monitor the treatment adherence and success.

Therefore, the aim of the current study was to explore the gut microbiota in CD, and the effect of GFD. The hypothesis tested was that CD patients are colonized by different gut microbiota to healthy subjects and this is explained by GFD. Finally, the main research objectives were to explore the gut microbiota composition in CD patients and healthy counterparts, evaluate faecal SCFA concentration in TCD, UCD and HC, assess compliance with GFD in treated CD patients and investigate potential correlations between faecal sample characteristics, microbiota composition and metabolic activity, diet composition and serological markers in TCD patients.

CHAPTER 2 Subjects and Methods

2.1 Chapter outline

This chapter will outline the study design, describing participant recruitment and research methodology, as well as bioinformatic approaches and statistical analysis.

2.2 Study design

This study has two cohorts; one of cross-sectional and another of prospective design. In the first cohort, TCD children on GFD for at least one year were compared with UCD on a gluten containing diet (GCD), siblings of CD children and HC. In the second cohort, a cohort of UCD patients was followed from diagnosis to six (GFD-6 months) and 12 months (GFD-12 months) on GFD.

2.3 Participants & exclusion criteria

Participants were under the care of the clinical team at the Royal Hospital for Sick Children (RHSC), Yorkhill, Glasgow and the Greater Glasgow and Clyde outreach NHS centres at the time of the study (2010-2012). For comparison purposes, healthy subjects were recruited; either siblings of CD children with no overt clinical symptoms and negative serology or healthy volunteers.

Subjects who reported use of antibiotics, probiotics or prebiotics within three months prior to faecal sampling, and patients with secondary autoimmune disease were excluded from the study. Children who were under investigation by the GI team for CD and had positive serology markers consistent with CD, but negative biopsy results, were also excluded.

The UCD patients were identified from referral letters to RHSC, whereas TCD children were recruited from the dietetic appointment list of the Yorkhill Hospital. HC and siblings of CD patients were recruited through the use of poster advertisements and by word of mouth. Before their appointment with the clinician at the CD clinic, all parents and children were posted an age specific letter explaining briefly the aim and the design of the study. The patients were approached by the researcher at the end of their clinical appointment and were asked to participate in the study. Participation in the study was voluntary and participants were free to discontinue at any time. Informed written consent was obtained from the parents/ carers and consent was given by the children as well.

2.4 Initial appointment

Once recruited into the study, parents/ carers and children were asked to jointly go through an interview to collect information through self- or researcher- completed questionnaires.

Anthropometric data, data concerning their daily diet, compliance with GFD and GI symptoms were collected during the first appointment.

2.4.1 Anthropometric data

Body weight was measured in kilograms (Kg), to the nearest gram, using a Tanita scale (TBF 300) and height in centimetres, to the nearest millimetre, using a fixed stadiometer.

2.4.2 Dietary assessment

The dietary habits of participants were evaluated by completion of self-completed Food Frequency Questionnaires (FFQ) by the children, with the help of their parents/ carers when necessary. The FFQ was an adaptation of an FFQ used for healthy children [89]. The FFQ included a list of 115 common foods, grouped under 12 categories, such as: a. Breakfast cereals, b. Potatoes, Rice and Pasta, c. Bread (including sandwiches and toast) etc. There were six response options (rarely/ never, 1-2 per month, 1 day/ week, 2-3 days per week, 4-6 days per week, every day). The total number of times per day they were consuming a specific food product was recorded too. The portion size of each food item was estimated for separate ages, based on the National Diet and Nutrition Surveys. Thus, for the dietary analysis, FFQ responses were converted into grams per day and analysed through the dietary analysis software WinDiets (version 2008; Univation Ltd, Aberdeen, Scotland). Energy was reported as a percentage of the estimated average requirement (%EAR), whereas macronutrients as g per day and/ or percentages of total daily energy intake (%TDEI).

2.4.3 Gastrointestinal symptoms

GI symptoms of the CD children were evaluated using an age specific GI symptom questionnaire (paediatric quality of life inventory™, PedsQL™ - gastrointestinal symptom scale (GSS), Version 6, Appendix). The questionnaires were handed to both parents/ carers and their children. Nine questions regarding pain in abdomen or stomach, diarrhoea, constipation, nausea, vomiting, discomfort in abdomen or stomach, flatulence – "passing wind", bloating and the frequency of not feeling hungry were included. The answers ranged from zero (never) to four (almost always), based on a 5-point Likert scale. The final score of the PedsQL-GSS questionnaire was extrapolated using the scoring procedure of the dimensions, where items are reversed scored and linearly transformed to a 0-100 scale as follows: 0=100, 1 =75, 2=50, 3=25, 4=0, and was equal to the sum of all the items over the number of items answered on all the Scales. Higher score indicates lower problems.

2.4.4 Biagi score

The Biagi questionnaire is a validated GFD adherence self-reported questionnaire that consists of few simple questions (i.e. "When you go out, do you tell the person who is cooking about your dietary needs?") forming a flow diagram (Figure 2.1) that leads to the final score [90]. Based on a 5-level scale, the Biagi score ranges from zero (0) to four (4). The scores zero (0) and one (1) indicate lack of

compliance. A score of two (2) indicates some compliance with GFD with important errors that require correction, while the scores three (3) and four (4) indicate strict adherence to GFD [91].

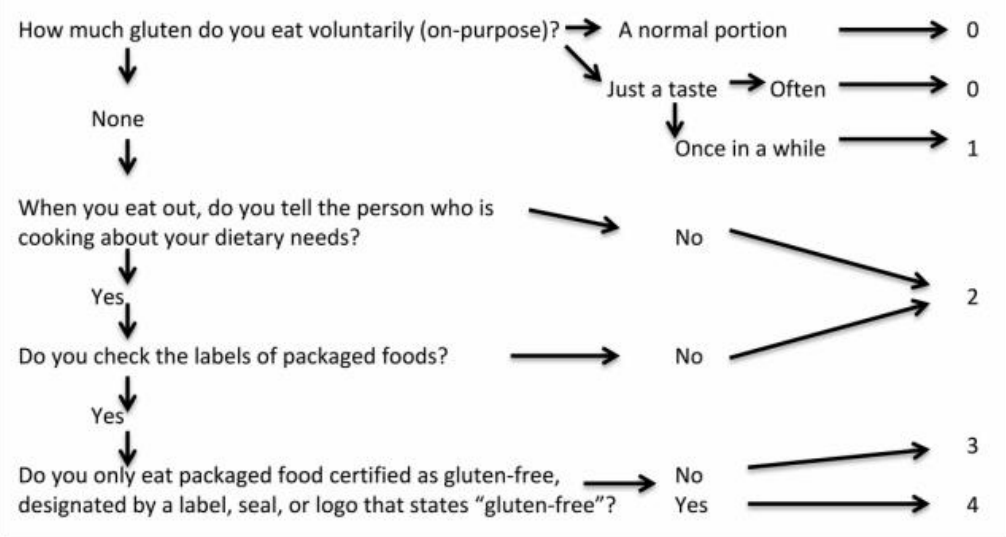


Figure 2.1: Biagi questionnaire, adapted from Biagi et al (2012) [91]

2.5 Laboratory measurements

2.5.1 Faecal sample collection

Sample collection was facilitated by a special kit. The collection device included a paper bedpan, a plastic single use container, a gas proof bag, a special anaerobic kit (Anaerocult® A, Merck, Germany), an indicator to check anaerobic conditions and an insulated bag in which the sample was transferred to the laboratory, within four hours. The sample was kept and transferred to the laboratory in the insulated bag along with frozen ice blocks. On arrival, the sample was mechanically blended with a domestic hand blender, until complete homogenization (2-3 min).

2.5.2 Faecal water content

The faecal water content was measured using the freeze drying method. Pre-weighed NaOH-stabilized stool samples were used. They were first kept in -80^{o} C, and when the freeze dryer (Freezer Dryer Micro Modulyo) was ready the samples were very quickly transferred in the apparatus and freeze dried for 36 hours. Then the dry sample weight was measured to two decimal places and the faecal water content was expressed as percentage of water per mass of stool sample.

2.5.3 Faecal pH

The measurement of faecal pH was performed on aqueous faecal slurries using an auto calibrated portable digital pH meter (Hanna HI 98140, Portugal). The slurries were made of 0.8 – 1.2 g of homogenized faecal sample suspended in threefold volume of distilled water (1:3 w/v).

2.5.4 SCFA

SCFA (acetic, propionic, butyric, valeric, caproic, heptanoic, octanoic acid) and BCFA (isobutyric, isovaleric, isocaproic acid) were measured using gas chromatography (GC) of ether extracts. SCFA are particularly volatile. In order to stabilize them, equal volumes of NaOH 1 M as the amount of the used faecal specimen were added. Thus the free carboxylic hydroxyl group of the acid was substituted by divalent ion and the compound was finally converted to their non-oxidizable salt form. In small bijoux tube (7ml), 0.8 – 1.5g of faecal sample was weighed and dispersed with the same amount of NaOH 1M. Four or five glass beads were added in the tube and then homogenised.

100μg of weighed dried sample was suspended in 300μL of distilled water, along with 100μL internal standard (72.65μmol/ L) and 100μL concentrated orthophosphoric acid. The mixture was homogenized by vortex for approximately 15s and extracted immediately three consecutive times with 1,5ml of diethyl ether, vortexed for 1 min each time, recovering and pooling the ether phase (supernatant) in one clean tube. The pooled extract, transferred in one vial, was ready for SCFA measurement.

SCFA were measured using a gas chromatograph (Agilent technologies 7820A G4350R), equipped with a flame ionisation detector (FID, 250^{o} C) and a DB – WAX UI glass column (15m * 0.535mm diam. * 1 μm film thickness). Nitrogen (30ml/ min) was used as the carrier gas. The pooled extract (1μl) was automatically injected onto the column (injector temperature 230^{o} C, splitless) and then the column temperature was held at 80^{o} C for 1 min, increasing by 15^{o} C per min to a final oven temperature of 210^{o} C. For the calibration, an external standard was used (185.83 μmol/ L acetic, 144.45 μmol/ L propionic, 97.31 μmol/ L isobutyric, 114.19 μmol/ L butyric, 87.03 μmol/ L isovaleric, 83.43 μmol/ L valeric, 52.64 μmol/ L isocaproic, 76.52 μmol/ L caproic, 65.80 μmol/ L heptanoic, 53.18 μmol/ L octanoic), with 6 dilutions to allow a standard curve. Each sample was extracted and analysed twice, in reverse order. As such, any time effect, due to evaporation of the very volatile SCFA, was accounted for. The results from the two extracts were averaged, unless the variance was wide. Measurements were repeated when the variance between the two replicates was higher than 10%. A quality control sample (well homogenized freeze-dried stock faecal material) was included at the beginning and the end of each run to assure repeatability of the assay and intra-assay comparison of the results between different runs. Measurements were repeated when the variance between the replicates of the quality control at the beginning and the end of each run were higher than 15%.

2.5.5 Sulphide

Free and total sulphide were measured in stool samples with a spectrophotometric assay in accordance to an in-house modified method reported by Strocchi et al. (1992) [92]. This is based on the principle of the methylene blue reaction first described by Cline et al. (1969) measuring sulphide in environmental specimens [93]. Typically, sulphide reacts with a diamine reagent in an acidic

environment under the oxidative effect of ferric chloride. The absorbance is measured at 670 nm and its intensity is proportional to the concentration of sulphide.

2.5.6 Lactate

D- and L-lactate were measured in freeze-dried faecal samples using a modified enzymatic spectrophotometric commercial assay (D, L Lactic Acid, UV Method, Boehringer Mannheim, Roche, Cat No; 11112821035). In the presence of lactate dehydrogenase, lactic acid is oxidized to pyruvate by NAD according to the reaction: Lactate + NAD ⇔Pyruvate + NADH + H

The equilibrium of these reactions lies on the side of lactate. By trapping pyruvate in a subsequent one-way reaction catalysed by the enzyme glutamate-pyruvate transaminase in the presence of L-glutamate, the equilibrium can be displaced in favour of pyruvate and NADH according to the reaction: Pyruvate + L-glutamate ⇒ L-alanine + 2-oxoglutarate

The amount of NADH formed in the above reactions is stoichiometric to the amount of D- and L-lactate respectively. The concentration of NADH is determined by means of its ultraviolet absorbance at 340 nm and equals to the expended lactate.

2.5.7 Gluten immunogenic peptide (GIP)

Faecal GIP concentration in CD patients was measured by a sandwich enzyme-linked immunosorbent assay (ELISA), using the iVYDAL *In Vitro* Diagnostics iVYLISA GIP-S kit (Biomedal S.L., Seville, Spain), following the manufacturer's guidelines. The method was split in two main steps; (a) the extraction of GIP from the stool matrix, followed by (b) the analysis of the sample and the quantification. Briefly, thawed stool sample was incubated with extraction solution in a ratio 1:9 for 60 min at 50^{o} C, after vortexing for 10 min. Along the 60-min incubation, the mixture was also vortexed for 10s every 15 min. After the incubation, the suspension was centrifuged for 10 min at 2,500g, and the supernatant was transferred to a clean tube. The extracted samples were then diluted 1:10 or 1:30 and incubated for 60 min in the provided microtitre plate coated with G12 antibodies, together with standards and assay's positive and negative controls. The wells were then washed with wash solution and incubated for another 60 min with horseradish peroxidase-conjugated G12 antibody. A second wash took place and the horseradish-peroxidase substrate was added in the well and left to incubate for 30 min. The last incubation was stopped with sulphuric acid and the absorbance was measured at 450 nm (Thermo Scientific, Multiscan Spectrum, 1500-859). All samples were measured in duplicate and the final concentrations, based on the standard curve were expressed in μg/g stool sample.

2.5.8 Serological tests

Anti-gliadin, anti-endomysial and anti-tissue transglutaminase antibodies were measured in serum samples of both TCD and UCD patients. In the prospective study, their levels were explored at six and 12 months after the initiation of GFD. Antibodies measured in the current study were: IgA anti-tissue

transglutaminase antibodies (tTG), total immunoglobulins IgG, IgA and IgM and IgA endomysium antibody (EMA IgA).

Tissue transglutaminase (tTG) is a Ca^{+} dependent enzyme responsible for post-translational modifications of specific polypeptide bound glutamines, by deamination and transamination [94]. Tissue transglutaminase has a double role in the pathogenesis of CD; (a) it is the target of disease-specific autoantibodies and (b) generates deamidated gliadin peptides that are recognized by CD4(+), DQ2-restricted T cells from the CD lesions [94]. Thus, in CD patients, anti-tissue transglutaminase IgA and IgG antibodies are produced and can be used to diagnose CD, as well as evaluate the effect of the compliance with the dietary treatment. However, since immunoglobulin IgA deficiency is more common in CD patients compared with HC there is a high possibility that the IgA-deficient CD patients will yield false-negative [95]. For that reason, total IgA were also used in this study

2.5.9 16S rRNA gene sequencing

Bacterial DNA was isolated using the chaotropic method [96]. 16S rRNA sequencing of the V4 region was performed on the MiSeq (Illumina) platform. The 16S ribosomal RNA is a component of the 30S small subunit of prokaryotic ribosomes, with approximate length of 1.5kb. The genes coding for it are referred to as 16S rDNA and are used in reconstructing phylogenies, since they are highly preserved. There has been a lot of discussion recently about which region(s) of the bacterial RNA are the best, in terms of length, given information and ease of amplification, especially when using Illumina [97], with the vast majority of studies indicating the V4 region. Thus, in our case the V4 region was amplified, using fusion adaptors barcoded on the reverse strand, while the used forward primer was the same for all our samples. The DNA was amplified using a PCR machine in specific cycle temperatures (Table 2.1), and then the produced amplicons were running on a 1% gel agarose. Using the Zymoclean Gel DNA Recovery Kit D4001, the DNA amplicons were extracted from the gel and pooled all together in the same concentration.

Table 2.1: PCR specific conditions

Step	Temperature	Duration	Purpose
I	95° C	5 minutes	Initial denaturation (1 cycle)
II	98° C	20 seconds	Denaturation
III	60° C	15 seconds	Annealing
IV	72° C	40 seconds	Extension
Repeat steps II, III and IV for 25 cycles			
V	72° C	1 minute	Final extension (1 cycle)

2.6 Bioinformatics

Quality trimming was done using Sickle by applying a sliding window approach and trimming regions where the average base quality (PHRED score) drops below 20 [98], after receiving the reverse and forward sequences from the sequence centre (microbesNG, Birmingham, UK). Assembling of the reverse and forward read was then done using PANDAseq [99]. Usearch was further used for operational taxonomic units (OTU) construction as described in https://docs.google.com/document /d/1BcZAk28k7Uycr7iKKAVSiZ0MB9jDs9bODpdPZtYFH3Y/pub#h.agz7rwlf8m6. It dereplicated the multiple sequences, annotated them with cluster sizes and sorted them by decreasing cluster size. It facilitated the de novo chimeras' detection, applying 97% of similarity, the reference based chimera's detection, using a gold database (http://drive5.com/uchime/uchime_download.html) that is derived from the ChimeraSlayer reference database in the Broad Microbiome Utilities (http://microbiomeutil.sourceforge.net/), as well as the OTU clustering. Amplicon sequence variants (ASV) were also identified using the Divisive Amplicon Denoising Algorithm (DADA2) [100]. The representative OTU were then taxonomically classified using either the Random Forests Algorithm or the Linear Disciminant Analysis. Phylogenetic distances between OTU were produced by first using MAFFT to align the OTU against each other [101] and then by FastTree on these alignments to generate an approximately-maximum-likelihood phylogenetic tree [102]. The OTU table, phylogenetic tree, taxonomic information and metadata were used in the multivariate statistical analysis.

2.7 Statistical analysis

For the statistical analysis, Minitab 17/ 18 was used. The data were Box-Cox transformed with optimal λ and parametric tests were applied; Box-Cox cannot perform with non-positive values, so z-scores of weight, height and BMI were not transformed. The pseudo code 10^{-11} was used for logarithmic purposes in data where the absolute value was zero. For all the statistical tests, the threshold *p-value was* up to 0.05, with anything less than 0.05 being significant.

In the cross-sectional designed study, General Linear Model (GLM) was used to compare TCD, UCD and HC, followed by pairwise comparison using Bonferroni correction.

In the prospective study, significant differences during follow-up were explored using GLM, accounted for the paired design of the study and followed by pairwise comparison with Bonferroni correction. Data are expressed as mean (SEM).

The agreement between the methods that evaluated the adherence to treatment with GFD was explored using Kappa statistics and the results were interpreted using the cut offs by Altman (1991)[103] (Table 2.2).

Table 2.2: Interpretation of Kappa values to evaluate the strength of agreement between two methods

Value of K	Strength of agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

Correlation between two variables was evaluated using Pearson coefficient when variables were normally distributed and Spearman rank-order correlation when variables were non-normally distributed. Normality of the distribution was evaluated using the Kolmogorov-Smirnov test.

Multivariate statistical analysis was performed in RStudio (R version 3.4.0); this mainly used packages Vegan, Phyloseq and DESeq2. Samples were rarefied to 5,000 reads to test for α-diversity. The diversity indices calculated for HC, TCD and UCD, both at diagnosis and during the follow-up, were species richness, Shannon H index and Simpson index.

Microbial compositional structure was assessed using a non-metric multidimensional scaling plot (NMDS) at genus and OTU level, which determined differences in communities of all samples of TCD, UCD and HC, as well as differences in communities of UCD samples during the follow-up. Herein, Bray-Curtis dissimilarity index and unweighted Unifrac distance analysis were used. The former considers bacterial taxon abundance, while the latter takes into account the phylogenetic distances of the bacterial taxa through presence/ absence, without accounting for their proportional representation. A covariance ellipse was added in the NMDS plot based on the standard error of the mean, with the centroid of each ellipse representing the group mean. The ellipse was defined by the covariance within each group; the bigger the ellipse, the more variability in community structure in samples within the group. For analysis of variance using distance matrices (Bray-Curtis/ unweighted Unifrac) PERMANOVA was used, with the strata command to take into account repeated sampling from UCD during the follow-up. Local contribution for β-diversity (LCBD) analysis was also performed to measure the contribution of each sample to the total OTU β-diversity; samples with high LCBD represent samples that are markedly different from the average β-diversity of all study samples. Differences in OTU, genus, family, class and phylum level between the groups were found using the DESeq2 package, identifying taxa that have log-fold changes between the groups. For correlations between discriminatory OTU and faecal, subjects' characteristics that were accordingly, significantly different between the groups, Kendall rank correlation was used. Benjamini-Hochberg correction was used for multiple testing in all.

2.8 Ethical considerations

The study was approved by the West of Scotland Research Ethics Committee and R&D NHS Greater and Clyde 11/WS/0006.

CHAPTER 3 Methodology Chapter: The efficacy of different methods to assess the compliance with GFD

3.1 Chapter outline

This thesis evaluated the efficacy of a novel biomarker, the faecal 33-mer GIP, to assess the compliance with GFD and compared it with the traditional methods used so far.

3.2 Introduction

Small bowel biopsy is the gold standard method used to assess mucosal healing and adherence to GFD. However, because of its invasiveness, relative risk, and cost, it is not practical to use this method routinely [104]. Anti-gliadin antibodies were the first to be used as a screening tool for CD [105]. Since then, serological markers, such as the anti-tTG immunoglobulin A (IgA), anti-gliadin IgA & anti-gliadin immunoglobulin G (IgG) antibodies, anti-endomysium IgA antibody (EmA-IgA) and anti-deamidated gliadin peptides IgA antibody, have been used for diagnosis, increasing diagnosis accuracy [2]. However, CD serology has not been always predictive of mucosal healing [106, 107], and hence Sharkey et al. (2013) recommended that serology should be used in follow-up only if validated against mucosal responses [108]. On the other hand, the use of self-reported questionnaires, such as the Biagi score that was developed based on the analysis of the strategy adopted by patients to avoid gluten consumption [91], is quite subjective and relies on patients' knowledge, awareness and honesty. Consequently, the need for an objective, specific and sensitive marker to monitor GFD adherence is still evident.

Recently, a novel marker, based on the detection of GIP in faeces or urine, has been found to be promising [5, 109]. Based on the identification and quantification of the GIP 33-mer peptide in human faeces, this method has 100% diagnostic specificity and 98.5% diagnostic sensitivity [22]. Gluten peptides, including α-gliadin and its 33-mer peptide, which are toxic for CD patients, reach the colon undigested and are subsequently excreted in significant amounts [110]. Therefore, the recovery of this fragment from human faeces indicates that dietary gluten has been consumed.

Among all gluten proteins, gliadin seems to play the most important role in the immune response cascade in CD. Gliadins are closely related proteins, categorised in four groups, α-, β-, γ- and ω- gliadin. They are all water soluble and rich in proline (20%) and glutamine (40%) [3]. The latter protects them from complete degradation and digestion [111], and hence they reach the large intestine partially undigested. The 33-mer peptide of α-gliadin was the first fragment of undigested peptides that was suggested as an initiator of the inflammatory response in CD patients [110]. This fragment has a span of 33 peptides in the molecule of α-gliadin and binds to tTG with substantially greater selectivity compared to other known natural substrates [110]. In the 33-mer peptide, there is a hexanoic epitope (QPQPLY), which is repeated three times and has been identified as three distinct T

cell epitopes [110], meaning that in association with proteins of the antigen-presenting cells, they are recognised by the gluten-specific HLA-DQ-restricted T cells leading to the immune response.

3.3 Subjects and methods

For the purposes of this sub-study, compliance with GFD was evaluated in 61 CD children on GFD for at least one year and 13 CD patients who were on GFD for six months (n=74). Patients were considered as compliant when (a) GIP was below the level of quantification (LOQ), according to the manufacturer's guidelines (GIP < 0.156 μg/g sample); (b) they had normal serum tTG levels, according to the laboratory's guidelines (tTG < 7 U/mL), (c) they had reported Biagi score equal to or greater than three (Biagi ≥ 3) [91], and (d) clinicians had reported "Good" adherence to GFD. When clinicians reported that compliance was "variable", patients were classified as non-compliant.

3.4 Statistical analysis

Since tTG and GIP levels were non-normally distributed, Spearman correlation was calculated as the most appropriate measure of their association. The agreement between all methods that evaluated the GFD adherence was explored using Kappa statistics and the results were interpreted using the cut offs by Altman (1991)[103] (Table 2.2). All samples with non-detectable faecal GIP levels were pseudo coding with 10^{-11} μg GIP/g sample, and for statistical tests, the threshold *p-value was* up to 0.05, with anything less than 0.05 being significant.

3.5 Results

3.5.1 Agreement between the Biagi score and faecal GIP analysis

Data from 65 CD children were used to evaluate agreement between the Biagi score and faecal GIP analysis. Fifty-two of 65 children were on GFD for at least one year and the remaining 13 were on GFD for six months. Fifty-one of 65 (78.5% of total) children were compliant according to both the Biagi score and GIP analysis (Table 3.1). Two of 65 (3.1% of total) were non-compliant according to both methods. Two CD children compliant by GIP analysis (3.8% of the compliant by GIP analysis) were non-compliant by the Biagi score, while 10 of 12 CD children deemed to be non-compliant by GIP analysis (83.3% of the non-compliant by GIP analysis) were compliant according to the Biagi score (Table 3.1). Overall, 96.2% of the compliant according to GIP analysis CD children were correctly identified as compliant by the Biagi score (sensitivity), while only 16.7% of the non-compliant by GIP analysis, were identified as non-compliant by the Biagi score (specificity). Moreover, 83.6% of CD children deemed to be compliant by the Biagi score were actually compliant by GIP analysis (positive predictive value-PPV), whereas only 50% of children that were non-compliant by the Biagi score were actually non-compliant by faecal GIP analysis (negative predictive value-NPV). Overall, the agreement between the two methods was poor (Kappa= 0.17, Table 3.1).

3.5.2 Agreement between clinicians' report and faecal GIP analysis

Data from 37 TCD children were used to assess the agreement between the clinicians' evaluation and faecal GIP analysis. Collectively 30 of 37 TCD (81.1% of total) were compliant based on both the clinicians' evaluation and faecal GIP analysis, while one of 37 (2.7% of total) were non-compliant based on these two methods. One non-compliant by clinician assessment CD child (50% of the non-compliant according to the clinicians' evaluation) was classified as compliant by GIP analysis, while five out of six CD children who were non-compliant based on the faecal GIP analysis (83.3% of the non-compliant by GIP analysis) were compliant according to the clinicians' report (Table 3.1). Taken as a whole, 96.8% of compliant by faecal GIP analysis CD children were correctly identified as compliant by clinicians (sensitivity), while similarly to the Biagi score only 16.7% of the non-compliant by faecal GIP analysis CD children were correctly identified as non-compliant by clinicians (specificity). Furthermore, 85.7% of CD children considered as compliant by clinicians were actually compliant based on faecal GIP analysis (PPV), whereas only 50% of CD children considered as non-compliant by clinicians were actually non-compliant based on the faecal GIP concentration (NPV). Likewise, the agreement between the two methods was poor (Kappa= 0.18, Table 3.1).

3.5.3 Agreement between the serum tTG and faecal GIP analysis

Data from 60 CD children who were on GFD for at least six months were used to evaluate the agreement between the serum tTG and faecal GIP analysis. Fifty-one of 60 children (85% of total) were on GFD for at least one year, while nine of 60 CD children (15% of total) were on GFD for six months. Thirty-seven of 60 children (61.7% of total) were classified as compliant using the results from both the serum tTG and faecal GIP analyses. Similarly, seven of 60 (11.7% of total) were non-compliant according to these two methods. Eleven of 49 CD children deemed to be compliant by faecal GIP analysis (22.9% of the compliant by GIP analysis) were non-compliant based on the serum tTG concentration, while five of 12 CD children deemed to be non-compliant by faecal GIP levels (41.7% of the non-compliant by GIP analysis), were compliant by serum tTG analysis. 77.1% of compliant by faecal GIP analysis CD children were correctly identified as compliant by serum tTG levels (sensitivity), whereas only 16.7% of the non-compliant by the faecal GIP analysis were correctly identified as non-compliant by serum tTG levels (sensitivity). Consequently, 88.1% of CD children deemed to be compliant by the serum tTG concentration were actually compliant by faecal GIP analysis, but only 8.3% of the CD who were considered as non-compliant by serum tTG levels, were actually non-compliant by faecal GIP analysis. There was no association between the serum tTG concentration and faecal GIP levels (Spearman Rho= 0.138, *p-value*= 0.293), and the agreement between the two methods was poor (Kappa= 0.17, Table 3.1).

Table 3.1: Frequency of compliant and non-compliant CD children according to faecal GIP concentration, Biagi score, clinicians' report and serum tTG levels

	Compliant (GIP < 0.156 μg/g wet matter)	**Non-compliant** (GIP > 0.156 μg/g wet matter)	**Total**
Compliant (Biagi ≥ 3)	51 (96.2)	10 (83.3)	61
Non-compliant (Biagi < 3)	2 (3.8)	2 (16.7)	4
Total	53	12	65
Compliant (Clinicians' report: Good)	30 (96.8)	5 (83.3)	35
Non-compliant (Clinicians' report: Variable)	1 (3.2)	1 (16.7)	2
Total	31	6	37
Compliant (tTG < 7 U/mL)	37 (77.1)	5 (83.3)	42
Non-compliant (tTG ≥ 7 U/mL)	11 (22.9)	1 (16.7)	12
Total	48	6	54

In brackets is shown the percentage (%) of column sum; Kappa (GIP Vs Biagi score) = 0.17; Kappa (GIP Vs clinicians' report)= 0.18; Kappa (GIP Vs tTG)= 0.17

3.6 Conclusion

The results of the present chapter show that faecal GIP analysis is the most accurate method to assess adherence to treatment with GFD. Almost all CD children deemed to be compliant by faecal GIP analysis were correctly identified as compliant by the Biagi score, clinicians and serum tTG method (sensitivity= 96.2%, 96.8% and 77.1% respectively). However, only 16.7% of non-compliant by faecal GIP analysis CD children were correctly identified as non-compliant by these three traditional methods, revealing important limitations compared to the faecal GIP method to identify non-compliant patients. Therefore, in the main study presented in this thesis, the adherence to treatment with GFD was evaluated using the novel biomarker faecal GIP.

CHAPTER 4 Gut Microbiota Structure and Functionality in CD patients

4.1 Chapter outline

This chapter presents the results of both the cross-sectional (section 4.2) and prospective study (section 4.3), with regard to the microbiota structure, composition, metabolic activity, along with patients metadata. Potential correlations of the produced bacterial SCFA and the subjects and faecal samples characteristics are explored in the section 4.4.

4.2 Cross – sectional study

In the cross-sectional study, 145 children were recruited, including 45 TCD children on GFD for at least one year, 20 UCD on GCD, 23 siblings of the patients and 57 HC. Mean age (SEM) of the participants at recruitment was 8.8 (0.3) years, with more than half of children being female (53.8%). HC were significantly younger (mean (SEM)) (7.8 (0.04)) than UCD (10.1 (0.7), p=0.017), and tended to be younger than TCD (9.3 (0.5), p= 0.065) (Table 4.1). Diagnosis of TCD children was made at a significantly younger mean age (SEM) (4.9 (0.5) years) than UCD patients (10.1 (0.7), p< 0.0001).

Mean (SEM) BMI z- score of children was 0.08 (0.09) SD, with three children (2.1% of total) classified underweight (BMI z- score <-2 SD) and ten (6.9% of total) obese (BMI z- score >2 SD). Mean (SEM) height z- score was 0.20 (0.09) SD, with three children (2.1% of total) presenting values less than -2 SD, suggesting short stature. Mean (SEM) weight z- score was 0.15 (0.10) SD, with eight children (5.5% of total) presenting values higher than 2 SD and five (3.5% of total) presenting values lower than -2 SD. The latter five children (3.5% of total) were classified underweight. As a whole, there was no difference in both absolute mean value and z-score of weight, height and BMI between the groups (Table 4.1).

Table 4.1: Demographic and anthropometric characteristics of HC, UCD, TCD children and siblings

	HC (57)	UCD (20)	TCD (45)	Siblings (23)	*p-overall*
Age (years)	7.8 (0.41)	10.1 (0.70) [a]	9.3 (0.47)	9.1 (0.65)	0.008
Gender (M/F)	27 / 30	10 / 10	20 / 25	10 / 13	-
Weight (Kg)	29.1 (1.63)	33.7 (2.87)	32.3 (1.7)	33.9 (3.3)	0.161
Weight z-score*	0.15 (0.15)	-0.23 (0.24)	0.16 (0.17)	0.44 (0.26)	0.371
< -2 SD [n (%)]	2 (3.5)	2 (10.0)	1 (2.2)	0 (0)	-
> 2 SD [n (%)]	2 (3.5)	0 (0)	3 (6.7)	3 (13.0)	-
Height (cm)	128.6 (2.9)	137.7 (4.1)	134.0 (2.8)	136.2 (4.4)	0.183
Height z-score #	0.29 (0.15)	-0.16 (0.22)	0.06 (0.16)	0.54 (0.25)	0.246
< -2 SD [n (%)]	1 (1.8)	2 (10.0)	0 (0)	0 (0)	-
> 2 SD [n (%)]	5 (8.8)	(0)	2 (4.4)	2 (8.7)	-

BMI (Kg/m²)	17.2 (0.34)	17.5 (0.52)	17.4 (0.29)	17.4 (0.59)	0.460
BMI z-score #	0.06 (0.15)	-0.23 (0.25)	0.18 (0.17)	0.20 (0.24)	0.373
< -2 SD [n (%)]	1 (1.8)	2 (10.0)	0 (0)	0 (0)	-
> 2 SD [n (%)]	4 (7.0)	0 (0)	3 (6.7)	3 (13.0)	-

Values expressed as mean (SEM); GLM for UCD, TCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; *: GLM without Box-Cox transformation; [a]: *p-value*= 0.017 compared to HC

4.2.1 Dietary intake

Participants' dietary intake was evaluated analysing self-reported FFQ, which was an adaptation of a validated questionnaire to include gluten free products, and analysed using the WinDiets software. Energy intake was expressed as a percentage to the estimated average requirements (%EAR). Mean (SEM) total daily energy intake (TDEI) was 92.1% (2.9), presenting no difference among the groups (p=0.854, Table 4.2). The mean (SEM) fat, protein and carbohydrates' consumption, expressed as a percentage to the TDEI, was 31.6% (0.50), 13.9% (0.40) and 52.3% (0.51) respectively, with no significant difference between groups (p= 0.249, 0.380 and 0.161 respectively, Table 4.2). TCD children consumed significantly less polyunsaturated fatty acids (PUFAs) (mean (SEM)), as a percentage to TDEI, (3.4 (0.18)) than HC (5.0 (0.20), p<0.0001) and UCD (4.5 (0.30), p= 0.002, Table 4.2). Protein intake (mean (SEM)) expressed as a percentage to the reference nutrient intake (RNI) was significantly different between TCD, UCD and HC (p= 0.021), with TCD consuming significantly less protein (%RNI) (179.8 (11.5)) than HC (219.1 (11.0), p= 0.035, Table 4.2). The mean (SEM) starch and NMES intake (%TDEI) were significantly lower in TCD [5.9 (0.6) and 12.8 (1.19) respectively] than HC [19.7 (0.70), p< 0.0001 and 15.3 (0.60), p= 0.009 respectively] and UCD [22.4 (1.10), p< 0.0001 and 16.0 (0.84), p= 0.012 respectively, Table 4.2]. finally, the mean (SEM) NSP intake (g, % DRV: dietary reference value) was significantly lower in TCD children (5.7 (0.40) and 32.4 (2.63) respectively) than HC (11.3 (0.69) and 71.0 (4.95) respectively) and UCD (9.9 (0.80) and 56.9 (4.84) respectively, p < 0.0001 in all cases, Table 4.2).

Table 4.2: Dietary intake of HC, UCD, TCD children and siblings

	HC (57)	UCD (20)	TCD (45)	Siblings (23)	*p-overall*
Energy (%EAR)	94.8 (4.8)	88.7 (5.1)	90.2 (4.2)	91.3 (6.2)	0.854
Fat (g)	62.1 (3.4)	58.5 (4.0)	59.4 (3.6)	59.6 (5.7)	0.808
Fat (%TDEI)	32.4 (0.69)	31.0 (0.85)	30.9 (0.96)	30.1 (1.3)	0.249
SFA (%TDEI)	13.7 (0.46)	13.1 (0.44)	13.3 (0.47)	13.1 (0.67)	0.734
PUFA (%TDEI)	5.0 (0.20)	4.5 (0.30)	3.4 (0.18)[a, b]	4.1 (0.31)	< 0.0001
MUFA (%TDEI)	7.6 (0.22)	7.3 (0.40)	6.9 (0.32)	7.2 (0.41)	0.129
Protein (g)	58.5 (3.3)	53.9 (3.4)	53.9 (2.8)	55.0 (3.4)	0.743
Protein (%TDEI)	13.9 (0.29)	13.2 (0.37)	14.0 (1.03)	13.3 (0.38)	0.380

Protein (%RNI)	219.1 (11.0)	176.7 (12.9)	179.8 (11.5)[c]	192.0 (15.1)	0.021
Carbohydrates (g)	229.7 (13.8)	230.4 (14.1)	235.9 (12.3)	238.0 (14.4)	0.627
Carbohydrates (%TDEI)	51.3 (0.71)	53.0 (1.05)	53.3 (0.93)	54.1 (1.34)	0.161
Sugars (%TDEI)	44.8 (2.55)	36.3 (3.26)	48.7 (3.80)	49.5 (5.34)	0.097
Starch (%TDEI)	19.7 (0.70)	22.4 (1.10)	5.9 (0.61)[d]	19.7 (0.78)	< 0.0001
NMES (%TDEI)	15.3 (0.60)	16.0 (0.84)	12.8 (1.19)[e, f]	18.5 (0.94)	0.003
NSP (g)	11.3 (0.69)	9.9 (0.80)	5.7 (0.40)[g]	9.6 (0.63)	< 0.0001
NSP (%DRV)	71.0 (4.95)	56.9 (4.84)	32.4 (2.63)[h]	55.2 (4.0)	< 0.0001
Dietary fibre (g)	10.4 (0.86)	10.8 (1.77)	14.4 (1.30)	10.9 (1.59)	0.014

EAR: estimated average requirement; RNI: reference nutrient intake; TDEI: total daily energy intake; DRV: dietary reference value; SFA: saturated fatty acids; PUFA: polyunsaturated fatty acids; MUFA: monounsaturated fatty acids; NMES: non-milk extrinsic sugars; NSP: non-starch polysaccharides; Values expressed as mean (SEM); GLM for UCD, TCD & HC; Box – Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: $p < 0.0001$ compared to HC; [b]: $p = 0.002$ compared to UCD; [c]: $p = 0.035$ compared to HC; [d, g, h]: $p < 0.0001$ compared to HC & UCD; [e]: $p = 0.009$ compared to HC; [f]: $p = 0.012$ compared to UCD

4.2.2 Compliance with GFD

Compliance with GFD was evaluated using all Biagi score, clinicians' report and faecal GIP method (Table 4.3). According to the Biagi score, 37 TCD children (90.2% of TCD) were strictly compliant with GFD (Biagi score = 3: 75.6% & Biagi score = 4: 14.6%). Biagi score was equal to zero (0) in two TCD children (2.5% of TCD), indicating lack of compliance with GFD (Biagi score = 0, Table 4.3), and equal to two (2) in the remaining three TCD children (7.3% of TCD), suggesting some compliance with GFD, but with important errors that require correction. GFD adherence, reported by clinician, was 'Good' in 37 out of 45 TCD (82.2% of TCD) and "variable" in three out of 45 TCD (6.7% of TCD) children; missing data of five children.

As expected mean (SEM) faecal GIP concentration was significantly higher in UCD children on GCD (3.5 (0.6)) than TCD on GFD (0.3 (0.06), $p < 0.0001$, Table 4.3). All but one UCD child had more than 0.156 μg of faecal GIP per gram of faecal wet matter. One UCD child had undetectable faecal GIP levels (GIP < 0.156 μg/g wet matter; LOQ of the used ELISA essay), suggesting that either this child was on GFD prior to CD diagnosis or this was a false negative measurement (Table 4.3). Thirty-eight of 45 TCD children (84.4% of TCD) had undetectable levels of faecal GIP, being compliant with GFD. GIP concentration in faecal samples of the remaining seven TCD (15.6% of TCD) was greater than or equal to 0.156 μg/g wet matter, indicating poor compliance with GFD.

4.2.3 Serological tests

Mean (SEM) anti-tTG IgA antibody concentration was 8-fold higher in UCD (64.8 (13.3)) than TCD group (7.9 (3.0), $p < 0.0001$), with 37 TCD children (79.1% of TCD) having normal tTG levels (<7 U/mL) and 10 UCD (90.1% of UCD) having increased levels of tTG (≥ 7 U/mL, Table 4.3).

Concentration of total IgA, IgG and IgM had no difference between the groups (Table 4.3). All 18 UCD children for whom EMA-IgA was evaluated, were EMA positive (Table 4.3).

Table 4.3: Biagi score, serological antibody tests, and faecal GIP levels of UCD and TCD children

	UCD (20)	TCD (45)	*p-value*
tTG (U/mL)	64.8 (13.3) [9]	7.9 (3.0) [2]	< 0.0001
< 7 [n (%)]	1 (9.1)	34 (79.1)	-
≥ 7 [n (%)]	10 (90.9)	9 (20.9)	-
IgA (g/L)	1.3 (0.08)	1.2 (0.09 [4]	0.524
IgG (g/L)	8.9 (0.5)	9.8 (0.5) [4]	0.285
IgM (g/L)	0.9 (0.06)	1.0 (0.06) [4]	0.558
EMA-IgA (positive/ negative)	18/ 0 [2]	8/ 1 [36]	-
Biagi score			
0 [n (%)]	-	2 (2.5)	-
1 [n (%)]	-	0 (0)	-
2 [n (%)]	-	3 (7.3)	-
3 [n (%)]	-	31 (75.6)	-
4 [n (%)]	-	6 (14.6)	-
GIP (µg/g wet matter)	3.5 (0.6) [1]	0.3 (0.06)	< 0.0001
< 0.156 [n (%)]	1 (5.3)	38 (84.4)	-
≥ 0.156 [n (%)]	18 (94.7)	7 (15.6)	-

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM for UCD Vs TCD; Box-Cox transformation with optimal λ

4.2.4 Gastrointestinal symptoms (PedsQL-GSS questionnaire)

GI symptoms were evaluated using the PedsQL-GSS questionnaire. Eleven UCD children (55% of UCD) reported high levels (3-4) of passing wind. Ten UCD children (50% of UCD) reported high levels (3-4) of abdominal pain and nine UCD (45% of UCD) reported raised levels of abdominal discomfort. Moreover, 20 to 30% of UCD reported high levels (3-4) of bloating, diarrhoea, constipation, nausea and not feeling hungry (Table 4.4).

GI symptoms were raised in fewer TCD than UCD. Passing wind was elevated (3-4) in 12 TCD children (26.7% of TCD), abdominal pain and discomfort was elevated (3-4) in five TCD children (11.1% of TCD), while constipation and bloating in three TCD children (6.7% of TCD). HC and siblings of CD patients considered their GI symptoms levels to range from zero (0) to two (2) (Table 4.4). Consequently, PedsQL-GSS score was significantly different between TCD, UCD and HC groups (*p-overall* < 0.0001, Table 4.4), with UCD having significantly lower mean value (SEM) (57.1 (4.8)), and hence more GI problems than TCD (77.5 (2.7)) and HC (91.4 (1.7), Table 4.4). Similarly mean (SEM)

PedsQL-GSS score was significantly lower in TCD than HC ($p< 0.0001$), indicating that not only UCD had significantly more GI symptoms than TCD and HC, but TCD children had significantly more GI symptoms than HC as well.

Table 4.4: Frequency (n (%)) of self-reported GI symptoms and PedsQL-GSS score of HC, UCD, TCD children and siblings

	HC (56)	UCD (20)	TCD (45)	Siblings (23)
Abdominal pain				
0	27 (48.2)	1 (5.0)	14 (31.1)	12 (52.2)
1	20 (35.7)	3 (15.0)	11 (24.4)	9 (39.1)
2	7 (12.5)	6 (30.0)	15 (33.3)	2 (8.7)
3	2 (3.6)	6 (30.0)	4 (8.9)	0
4	0	4 (20.0)	1 (2.2)	0
Diarrhoea				
0	30 (53.6)	6 (30.0)	24 (55.8)	13 (56.5)
1	22 (39.3)	4 (20.0)	14 (32.6)	5 (21.7)
2	4 (7.1)	6 (30.0)	5 (11.6)	4 (17.4)
3	0	3 (15.0)	0	1 (4.3)
4	0	1 (5.0)	0	0
Constipation				
0	33 (60.0)	8 (40.0)	26 (57.8)	15 (65.2)
1	17 (30.9)	3 (15.0)	8 (17.8)	5 (21.7)
2	3 (5.5)	4 (20.0)	8 (17.8)	2 (8.7)
3	2 (3.6)	3 (15.0)	0	1 (4.3)
4	0	2 (10.0)	3 (6.7)	0
Nausea				
0	39 (69.6)	5 (25.0)	20 (44.4)	14 (60.9)
1	14 (25.0)	7 (35.0)	13 (28.9)	8 (34.8)
2	3 (5.4)	3 (15.0)	12 (26.7)	1 (4.3)
3	0	4 (20.0)	0	0
4	0	1 (5.0)	0	0
Vomiting				
0	39 (69.9)	11 (55.0)	31 (68.9)	15 (65.2)
1	16 (28.6)	8 (40.0)	10 (22.2)	8 (34.8)
2	1 (1.8)	1 (5.0)	4 (8.9)	0
3	0	0	0	0
4	0	0	0	0
Abdominal				
discomfort	30 (53.6)	2 (10.0)	16 (35.6)	11 (47.8)
0	16 (28.6)	3 (15.0)	9 (20.0)	9 (39.1)
1	10 (17.9)	6 (30.0)	15 (33.3)	2 (8.7)
2	0	4 (20.0)	4 (8.9)	0
3	0	5 (25.0)	1 (2.2)	1 (4.3)
4				
Passing wind				
0	27 (48.2)	1 (5.0)	18 (40.0)	12 (52.2)
1	16 (28.6)	1 (5.0)	5 (11.1)	4 (17.4)
2	9 (16.1)	7 (35.0)	10 (22.2)	4 (17.4)
3	4 (7.1)	4 (20.0)	9 (20.0)	3 (13.0)
4	0	7 (35.0)	3 (6.7)	0

Not feeling hungry				
0	28 (50.0)	9 (45.0)	22 (48.9)	19 (82.6)
1	15 (26.8)	1 (5.0)	11 (24.4)	2 (8.7)
2	11 (19.6)	5 (25.0)	11 (24.4)	2 (8.7)
3	2 (3.6)	3 (15.0)	1 (2.2)	0
4	0	2 (10.0)	0	0
Bloating				
0	44 (78.6)	4 (20.0)	25 (55.6)	18 (81.8)
1	12 (21.4)	5 (25.0)	7 (15.6)	2 (9.1)
2	0	5 (25.0)	10 (22.2)	2 (9.1)
3	0	5 (25.0)	2 (4.4)	0
4	0	1 (5.0)	1 (2.2)	0
PedsQL-GSS score*	91.4 (1.7)	57.1 (4.8)[c]	77.5 (2.7)[a, b]	86.6 (2.6)

Values expressed as number of subjects (percentage out of the total); GLM for TCD, UCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; *p-overall*< 0.0001; [a]: *p*< 0.0001 compared to UCD; [b]: *p*< 0.0001 compared to HC; [c]: *p*< 0.0001 compared to HC

4.2.5 Faecal sample characteristics

All but one faecal sample characteristic had no difference between HC, UCD and TCD children. Faecal ammonia was significantly different among HC, UCD and TCD patients (*p-overall*< 0.0001), with faecal samples of TCD children having significantly higher mean (SEM) concentration of ammonia (19.6 (8.2)) than HC (11.5 (0.8)) and UCD (7.8 (0.8)) (*p*= 0.001 for both comparisons, Table 4.5). In contrast, total sample weight (i.e. faecal output), faecal pH and faecal water content had no difference between HC, UCD and TCD groups (*p-overall*= 0.399, 0.184 and 0.259 respectively, Table 4.5).

Table 4.5: Faecal sample characteristics (total sample weight, pH, water content and ammonia) of HC, UCD, TCD children and siblings

	HC (57)	UCD (18)	TCD (42)	Siblings (23)	*p-overall*
Total sample weight (g)	54.4 (5.1)	50.4 (10.3)	47.4 (6.1)	60.4 (9.5)	0.399
Faecal pH	6.9 (0.08)	6.7 (0.3)	7.1 (0.1)	6.9 (0.13)	0.184
Faecal water content (%)	67.8 (0.7)	66.3 (1.7)	69.2 (1.2)	65.5 (1.5)	0.259
Ammonia (*10^{-4} mg/g wet matter)	11.5 (0.8)	7.8 (0.8)	19.6 (8.2)[a]	11.4 (1.3)	< 0.0001

Values expressed as mean (SEM); GLM for TCD, UCD & HC; Box – Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*=0.001 compared to HC & UCD

4.2.6 Sulphide

UCD patients had significantly less free sulphide (mean (SEM)) in dry matter (0.18 (0.04)), wet matter (0.06 (0.01)) and per total faecal output (3.0 (0.84)) than HC (0.41 (0.05), *p*= 0.008, 0.13 (0.02), *p*= 0.009 and 9.4 (2.1), *p*= 0.040 respectively, Table 4.6, Figure 4.1). UCD patients tended to have less free sulphide (mean (SEM)) in dry matter than TCD (0.34 (0.06), *p*=0.070, Table 4.6), while there was no significant difference between UCD and TCD patients when free sulphide was measured in wet matter and per total faecal output (*p*= 0.149 and 0.593 respectively). Bound sulphide was significantly different between the three groups only when expressed per dry matter (*p-overall*= 0.036), with TCD

patients having a significantly higher mean value (SEM) (3.0 (0.29)) than HC (2.2 (0.29), *p*= 0.038, Table 4.6). Regarding total faecal sulphide, concentration in dry matter tended to differ among the groups, but this difference did not reach significance (*p-overall*= 0.055). Accordingly, there was no difference between the groups when total sulphide was measured in wet matter and per total faecal output (*p-overall*= 0.202 and 0.992 respectively).

Table 4.6: Faecal sulphide concentration (µmol/g) of HC, UCD, TCD children and siblings

	HC (57)	UCD (18)	TCD (42)	Siblings (23)	*p-overall*
Free sulphide (µmol/g)					
dry matter	0.41 (0.05)	0.18 (0.04) [a]	0.34 (0.06)	0.37 (0.07) [2]	0.011
wet matter	0.13 (0.02)	0.06 (0.01) [b]	0.10 (0.01)	0.11 (0.02)	0.012
total output	9.4 (2.1)	3.0 (0.84) [c]	4.3 (0.73)	5.6 (1.3)	0.035
Bound sulphide (µmol/g)					
dry matter	2.2 (0.29)	2.2 (0.37)	3.0 (0.29) [d]	2.8 (0.41) [2]	0.036
wet matter	0.71 (0.09)	0.77 (0.13)	0.94 (0.10)	0.98 (0.13)	0.115
total output	41.5 (7.8)	47.6 (12.6)	42.5 (7.1)	59.7 (15.1)	0.850
Total sulphide (µmol/g)					
dry matter	2.6 (0.32)	2.4 (0.37)	3.3 (0.31)	3.2 (0.47) (2)	0.055
wet matter	0.84 (0.10)	0.83 (0.13)	1.03 (0.11)	1.09 (0.14)	0.202
total output	50.5 (9.6)	50.6 (12.8)	46.7 (7.5)	65.3 (16.2)	0.992

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM for TCD, UCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*=0.008 compared to HC; [b]: *p*=0.009 compared to HC; [c]: *p*=0.040 compared to HC; [d]: *p*=0.038 compared to HC

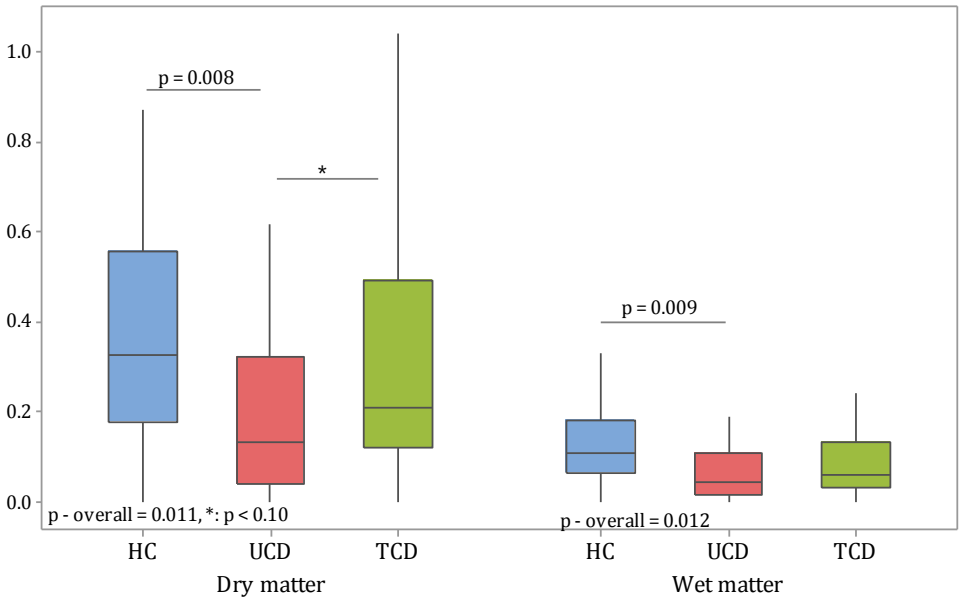


Figure 4.1: Faecal free sulphide concentration (µmol/g) of HC, UCD and TCD children measured in dry and wet matter

4.2.7 Lactate

Mean (SEM) faecal L-lactate concentration was significantly higher in UCD children [concentration in dry matter: 34.3 (6.7); in wet matter: 11.9 (2.4); per total faecal output: 5.9 (1.6)] than TCD [concentration in dry matter: 20.5 (1.5), *p*= 0.002; in wet matter: 6.0 (0.41), *p*= 0.003; per total faecal output: 3.4 (0.44), *p*= 0.005] and HC [concentration in dry matter: 30.4 (6.3), *p*= 0.011; in wet matter: 9.3 (1.8), *p*= 0.012; per total faecal output: 4.4 (0.55), *p*= 0.015, Table 4.7, Figure 4.2). Faecal D-lactate concentration in dry matter tended to differ between UCD, TCD and HC groups but this difference did not reach significance (*p-overall*= 0.092). However, the mean (SEM) D-lactate concentration in wet matter was significantly different between the three groups, with TCD children having significantly higher mean (SEM) concentration (32.1 (3.2) than UCD (19.1 (4.6), *p*= 0.037, Table 4.7, Figure 4.3. There was no difference in D-lactate concentration when expressed per total faecal output (*p-overall*= 0.455, Table 4.7).

Table 4.7: Faecal D-, L-lactate concentration (g/ 100g) of HC, UCD, TCD children and siblings

	HC (57)	UCD (18)	TCD (42)	Siblings (23)	*p-overall*
L-lactate					
(g/ 100g)	(*10^{-3})	(*10^{-3})	(*10^{-3})	(*10^{-3})	
dry matter	30.4 (6.3)	34.3 (6.7) [a, b]	20.5 (1.5)	20.3 (2.7)	0.003
wet matter	9.3 (1.8)	11.9 (2.4) [c, d]	6.3 (0.41)	10.2 (3.1)	0.004
total output	4.4 (0.55)	5.9 (1.6) [e, f]	3.4 (0.44)	6.8 (2.2)	0.005
D-lactate					
(g/ 100g)	(*10^{-3})	(*10^{-3})	(*10^{-3})	(*10^{-3})	
dry matter	41.6 (13.8)	19.1 (4.6)	32.1 (3.2)	32.7 (3.7)	0.092
wet matter	12.7 (4.1)	6.0 (1.4) [g]	10.1 (1.2)	13.9 (2.8)	0.033
total output	5.5 (0.76)	3.1 (0.83)	5.8 (1.0)	9.2 (2.2)	0.455

Values expressed as mean (SEM); GLM for TCD, UCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*= 0.011 compared to HC; [b]: *p*= 0.002 compared to TCD; [c]: *p*= 0.012 compared to HC; [d]: *p*= 0.003 compared to TCD; [e]: *p*= 0.015 compared to HC; [f]: *p*= 0.005 compared to TCD; [g]: *p*= 0.037 compared to TCD

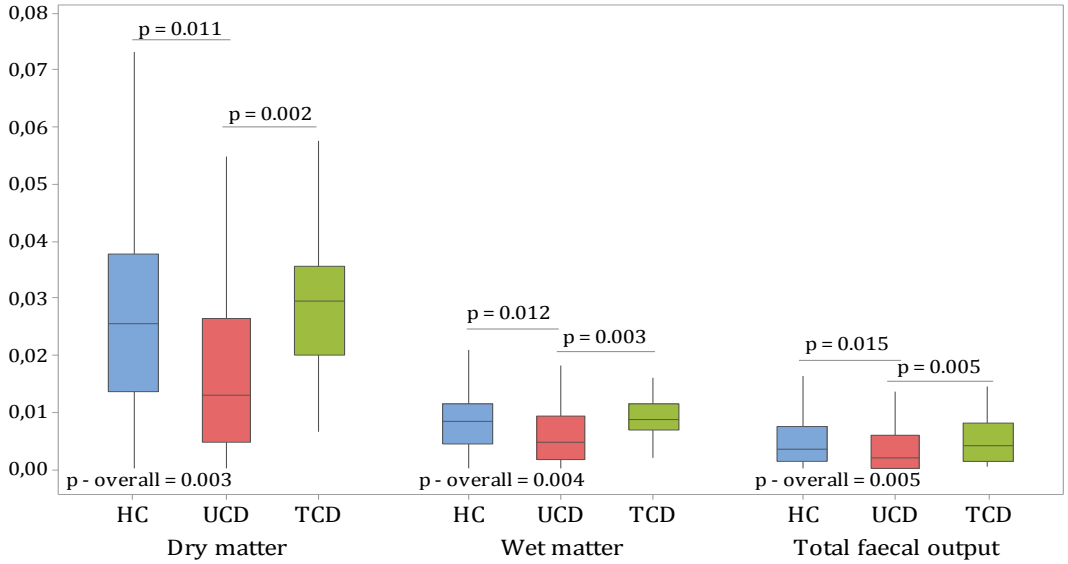


Figure 4.2: Faecal L-lactate concentration (g/ 100g) of HC, UCD and TCD children measured in dry, wet matter and per total faecal output

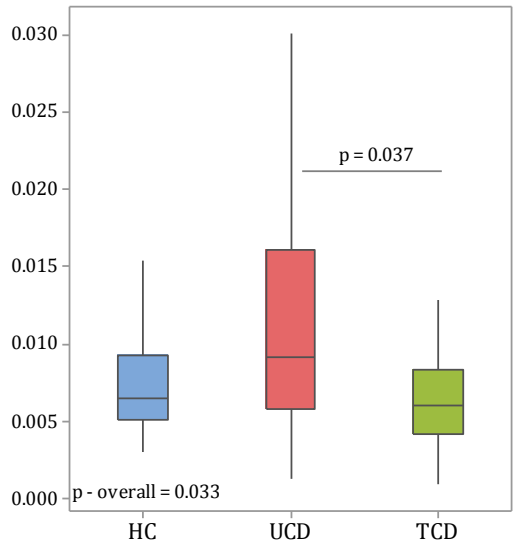


Figure 4.3: Faecal D-lactate concentration (g/ 100g wet matter) of HC, UCD and TCD children

4.2.8 SCFA

There was no difference in faecal SCFA concentration and relative abundance between HC and siblings of CD children. However, concentrations of propionic, butyric, valeric and caproic acids, expressed per total faecal output, were significantly different between TCD, UCD and HC groups (*p-overall*= 0.042, 0.015, 0.018 and 0.013 respectively, Table 4.8), with faecal samples of TCD children having significantly lower mean (SEM) concentration (µmol/g) (1096 (191), 994 (203), 113.3 (16.7) and 20.91 (3.86) respectively) than HC (1401 (157), *p*= 0.038, 1290 (240), *p*=0.015, 166.0 (17.2), *p*=0.018 and 42.69 (8.49), *p*=0.010 respectively, Table 4.8). Regarding SCFA concentration in wet matter, only caproic acid differed significantly among the three groups (*p-overall*= 0.035, Table 4.8), with TCD having significantly lower mean value (SEM) (0.43 (0.07)) than HC (0.68 (0.09), *p*= 0.030, Table 4.8). Finally, faecal SCFA concentration in dry matter had no difference between HC, UCD and TCD patients.

Regarding BCFA concentration, isobutyric and isovaleric acids differed significantly in faecal samples of HC, TCD and UCD children when expressed per total faecal output (*p-overall*= 0.038 and 0.044 respectively). In particular, mean (SEM) faecal concentration of isobutyric and isovaleric acids per total faecal output tended to be lower in TCD (147.6 (22.7) and 150.0 (22.9) respectively) than HC (189.6 (19.6), *p*= 0.051 and 188.7 (19.7), *p*= 0.064 respectively, Table 4.9). However, there was no difference in faecal concentration of isobutyric and isovaleric when expressed per dry and wet matter (*p*> 0.05, Table 4.9). In contrast, isocaproic acid concentration was significantly different between the three groups when expressed both per total faecal output and per gram of wet matter (*p-overall*= 0.003 and 0.040 respectively, Table 4.9). Specifically, TCD children had significantly lower mean (SEM) isocaproic acid concentration [concentration in wet matter: 0.08 (0.01); per total faecal output: 3.6 (0.72)] than HC [concentration in wet matter: 0.10 (0.01), *p*= 0.047; per total faecal output: (5.4 (0.79), *p*= 0.003]. Collectively, the concentration of total faecal SCFA per total faecal output tended to differ between TCD, UCD and HC groups, but this difference did not reach statistical significance (*p-overall*= 0.082, Table 4.8).

Table 4.8: Faecal SCFA concentration (µmol/g) of HC, UCD, TCD children and siblings

	HC (57)	UCD (18)	TCD (42)	Siblings (23)	*p-overall*
Acetic acid (µmol/g)					
dry matter	411.5 (19.8)	371.3 (35.1) [1]	432.9 (25.5)	355.0 (28.4)	0.400
wet matter	128.2 (5.2)	119.9 (10.0) [1]	124.6 (6.9) [2]	119.2 (7.8) [2]	0.656
Total output	6822 (671)	6806 (1351) [1]	5839 (860)	6151 (1218)	0.128
Propionic acid (µmol/g)					
dry matter	82.6 (5.6)	73.1 (9.5) [1]	79.8 (7.3)	70.1 (9.3)	0.584
wet matter	25.8 (1.6)	23.2 (2.7) [1]	23.1 (2.1) [2]	23.7 (2.7) [2]	0.308
Total output	1401 (157)	1484 (402) [1]	1096 (191) [a]	1271 (299)	0.042
Butyric acid (µmol/g)					
dry matter	78.7 (5.4)	74.4 (12.7) [1]	72.9 (8.8)	67.3 (8.6)	0.470
wet matter	24.4 (1.5)	23.3 (3.3) [1]	21.0 (2.5) [2]	22.2 (2.6) [2]	0.190
Total output	1290 (240)	1443 (371) [1]	994 (203) [b]	1078 (240)	0.015
Valeric acid (µmol/g)					
dry matter	9.6 (0.5)	9.4 (1.2)	8.14 (0.7)	8.8 (0.76)	0.172
wet matter	3.1 (0.18)	3.0 (0.37) [1]	2.5 (0.24) [2]	3.1 (0.29) [2]	0.098
Total output	166.0 (17.2)	194.1 (48.0) [1]	113.3 (16.7) [c]	162.7 (34.7)	0.018
Caproic acid (µmol/g)					
dry matter	2.3 (0.34)	1.7 (0.36) [1]	1.5 (0.24)	2.3 (0.7)	0.115
wet matter	0.68 (0.09)	0.51 (0.12) [1]	0.43 (0.07) [2] [d]	0.81 (0.35) [2]	0.035
Total output	42.7 (8.5)	25.9 (8.1) [1]	20.9 (3.9) [e]	39.6 (13.9)	0.013
Heptanoic acid					

(µmol/g)					
dry matter	0.19 (0.03)	0.19 (0.06) [1]	0.10 (0.02)	0.19 (0.07)	0.305
wet matter	0.06 (0.01)	0.06 (0.02)	0.03 (0.01) [2]	0.07 (0.04) [2]	0.311
Total output	3.7 (0.80)	3.1 (1.4)	1.3 (0.36)	3.0 (1.3)	0.178
Octanoic acid (µmol/g)					
dry matter	0.14 (0.05)	0.37 (0.12) [1]	0.33 (0.14)	0.37 (0.12)	0.539
wet matter	0.04 (0.01)	0.12 (0.04) [1]	0.09 (0.03) [2]	0.61 (0.57) [2]	0.505
Total output	3.0 (1.2)	6.7 (2.8) [1]	5.0 (2.0)	23.1 (19.8)	0.727
Total SCFA (µmol/g)					
dry matter	607.5 (29.3)	554.0 (57.4)	615.3 (40.3)	526.2 (46.7)	0.600
wet matter	189.5 (7.9)	177.7 (15.9)	177.9 (11.4)	177.0 (13.4)	0.491
Total output	10,113 (998)	10,421 (2,226)	8,372 (1,285)	9,146 (1,862)	0.081

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM for UCD, TCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p=0.038 compared to HC; [b]: p=0.015 compared to HC; [c]: p=0.018 compared to HC; [d]: p=0.030 compared to HC; [e]: p=0.010 compared to HC

Table 4.9: Faecal BCFA concentration (µmol/g) of HC, UCD, TCD children and siblings

	HC (57)	UCD (18)	TCD (42)	Siblings (23)	*p-overall*
Isobutyric acid (µmol/g)					
dry matter	11.1 (0.61)	11.6(1.3) [1]	9.8 (0.67)	10.6 (0.99)	0.242
wet matter	3.6 (0.22)	3.8 (0.46)	3.0 (0.25) [2]	3.7 (0.35) [2]	0.143
Total output	189.6 (19.6)	225.3 (51.5) [1]	147.6 (22.7)	206.1 (47.0)	0.038
Isovaleric acid (µmol/g)					
dry matter	11.0 (0.64)	11.7 (0.35) [1]	9.7 (0.72)	10.5 (0.87)	0.186
wet matter	3.6 (0.23)	3.9 (0.48) [1]	3.1 (0.28) [2]	3.7 (0.33)	0.151
Total output	188.7 (19.7)	227.9 (51.5) [1]	150.0 (22.9)	206.1 (46.7)	0.044
Isocaproic acid (µmol/g)					
dry matter	0.32 (0.03)	0.31 (0.04) [1]	0.27 (0.03)	0.28 (0.05)	0.121
wet matter	0.10 (0.01)	0.10 (0.01) [1]	0.08 (0.01) [2] [a]	0.09 (0.01) [2]	0.040
Total output	5.4 (0.79)	5.2 (1.2) [1]	3.6 (0.72) [b]	4.2 (1.0)	0.003

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM for UCD, TCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p=0.047 compared to HC; [b]: p=0.003 compared to HC

Mean (SEM) relative abundance of acetic acid was significantly higher in TCD group (72.3 (1.2)) than HC (68.1 (0.69), p= 0.009), while tended to be higher in TCD than UCD (68.1 (1.3), p= 0.081, Table 4.10, Figure 4.4a). The relative abundance of propionic acid did not differ between HC, UCD and TCD groups. However, the relative abundance of butyric and valeric acids was significantly different between the three groups (*p-overall*= 0.013 and 0.038 respectively, Table 4.10), with TCD having significantly lower mean value (SEM) (10.4 (0.72) and 1.4 (0.08) respectively) than HC (12.4 (0.47), p=

0.017 and 1.69 (0.09), *p*= 0.045 respectively, Figure 4.4b & 4.4c). Similarly, the relative abundance of isobutyric and isocaproic acid was significantly different among UCD, TCD and HC groups (*p-overall*= 0.044 and 0.031 respectively), while the relative abundance of isovaleric acid tended to differ between the groups (*p-overall*= 0.077). Mean (SEM) relative abundance of isobutyric and isovaleric acid tended to be lower in TCD (1.7 (0.10) and 1.8 (0.12) respectively) than UCD (2.2 (0.14), *p*= 0.068 and 2.22 (0.17), *p*= 0.060 respectively, Table 4.10), whilst there was no difference between HC and UCD groups.

Table 4.10: Relative abundance (%) of faecal SCFA and BCFA of HC, UCD, TCD children and siblings

	HC (57)	UCD (19)	TCD (45)	Siblings (23)	**p-overall*
Acetic acid (%)	68.1 (0.69)	68.1 (1.3)	72.3 (1.2) [a]	69.1 (1.4)	0.007
Propionic acid (%)	13.4 (0.38)	12.9 (0.85)	12.2 (0.54)	12.3 (0.73)	0.196
Butyric acid (%)	12.4 (0.47)	12.5 (0.86)	10.4 (0.72) [b]	11.8 (0.74)	0.013
Isobutyric acid (%)	2.0 (0.10)	2.2 (0.14)	1.7 (0.10)	2.1 (0.12)	0.044
Valeric acid (%)	1.7 (0.09)	1.7 (0.14)	1.4 (0.08) [c]	1.7 (0.09)	0.038
Isovaleric acid (%)	2.00 (0.11)	2.2 (0.17)	1.8 (0.12)	2.2 (0.14)	0.077
Caproic acid (%)	0.39 (0.05)	0.33 (0.07)	0.25 (0.03)	0.49 (0.16)	0.183
Isocaproic acid (%)	0.05 (0.004)	0.06 (0.008)	0.04 (0.004)	0.06 (0.008)	0.031
Heptanoic acid (%)	0.03 (0.006)	0.04 (0.012)	0.02 (0.004)	0.04 (0.016)	0.386
Octanoic acid (%)	0.02 (0.006)	0.07 (0.022)	0.05 (0.018)	0.26 (0.24)	0.532

Values expressed as mean (SEM); *: GLM for UCD, TCD & HC; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*=0.009 compared to HC; [b]: *p*=0.017 compared to HC; [c]: *p*=0.045 compared to HC

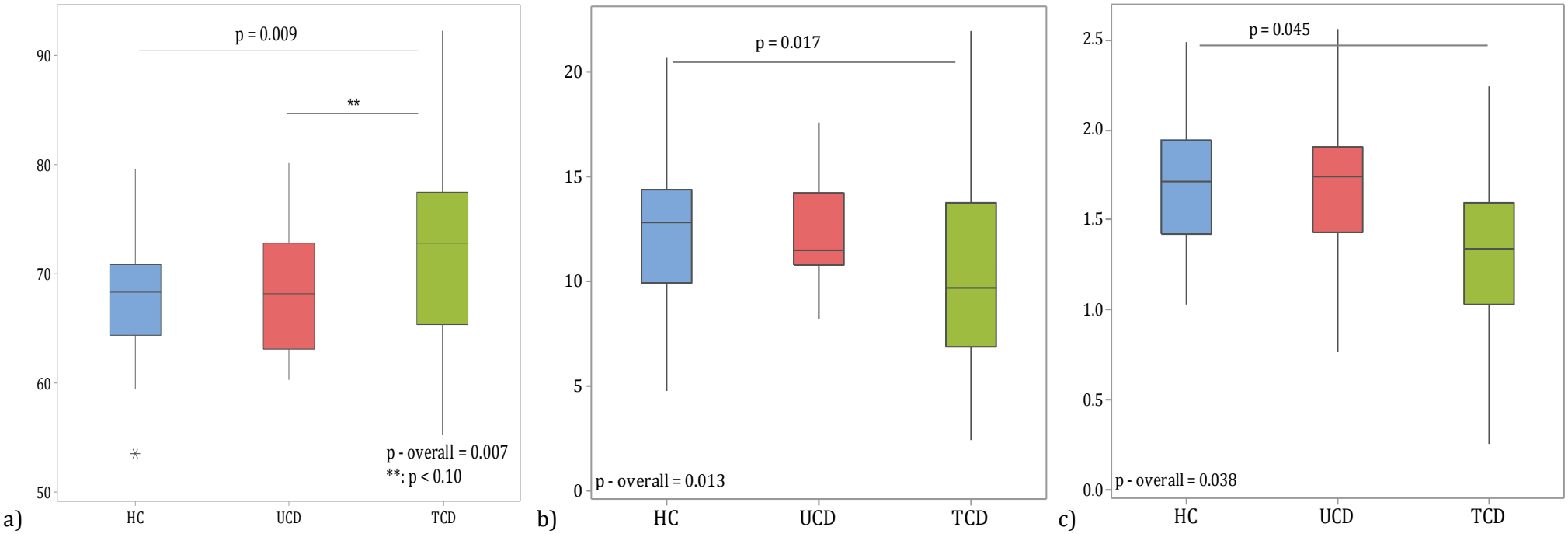


Figure 4.4: Relative abundance (%) of faecal (a) acetic, (b) butyric and (c) valeric acid in samples of children with TCD, UCD & HC

4.2.9 Faecal microbiota

4.2.9.1 Faecal microbiota community structure

There was no difference in alpha diversity, in terms of richness, Shannon H and Simpson index, at OTU (Figure 4.5a) and genus (Figure 4.6a) level between TCD, UCD and HC groups (p> 0.05 in all cases, Table 4.11). In contrast, their microbial community structure (β-diversity) differed significantly at OTU level using the Bray-Curtis dissimilarity index and unweighted Unifrac distance analysis. In the Bray-Curtis NMDS plot, 2.76% of the variation in community structure was explained by sample groups (p= 0.025, Figure 4.7a, Table 4.12). Similarly, in the NMDS plot for unweighted Unifrac, 2.49% of the variation in community structure was explained by sample groups (p= 0.027, Figure 4.7b, Table 4.12). In particular, the microbial community structure of TCD patients clustered separately to HC in the NMDS plot at OTU level for both the Bray-Curtis dissimilarity index and unweighted Unifrac distance analysis (p= 0.017 and 0.045 respectively, Table 4.12). TCD microbial community tended to differ compared with UCD patients in the NMDS plot for the unweighted Unifrac distance analysis at OTU level (p= 0.056, Table 4.12), and in the NMDS plot for the Bray-Curtis dissimilarity index at genus levels (p= 0.069, Table 4.12), while there was no difference between HC and UCD groups, neither in Bray-Curtis nor in unweighted Unifrac NMDS plot (Table 4.12).

Table 4.11: Alpha diversity indices in faecal samples of UCD, TCD and HC

Microbial diversity	*p-value* (USEARCH)	*p-value* (DADA2)
	OTU level	**ASV level**
Richness	0.44	0.28
Shannon H	0.65	0.20
Simpson index	0.57	0.32
	Genus level	
Richness	0.45	0.44
Shannon H	0.51	0.96
Simpson index	0.52	0.96

OTU: operational taxonomic unit, ASV: amplicon sequencing variants

LCBD analysis at OTU level confirmed that the gut microbiota structure of TCD individuals was significantly different compared with the microbiota of UCD and HC, with no difference between the latter two groups (Figure 4.10a). Specifically, TCD samples had significantly higher mean LCBD values than HC (p= 0.0017) and UCD (p= 0.0081), indicating that TCD were significantly more different from the average β-diversity of all study samples together (Figure 4.10a).

The effect of GFD on microbial community structure was evident at genus level as well. In the NMDS plot for the Bray-Curtis dissimilarity index, 3.0% of the variation was explained by sample groups (p= 0.026, Table 4.12, Figure 4.9a). There was no difference between HC and UCD patients (p= 0.515, R^2= 1.15%, Table 4.12). Nevertheless, the gut microbial community structure of TCD patients

was distinct from HC (p= 0.013, Table 4.12) and tended to be different to UCD (p= 0.069). LCBD analysis confirmed further that gut microbiota structure of TCD patients at genus level was different to HC (p=0.016, Figure 4.10b).

The variation in community structure among samples of the same participant group (β-dispersion) was measured as the distance of each sample from their respective group ellipse centroid. This was not different between the three groups neither at OTU nor at genus level when bacterial taxon abundance was taken into account (Bray-Curtis dissimilarity index). However, β-dispersion at OTU level in UCD group was significantly lower than β-dispersion in TCD group when phylogenetic distances were accounted in the unweighted Unifrac NMDS plot (p= 0.009, Figure 4.11b). Similarly, β-dispersion at OTU level in UCD patients tended to be lower than β-dispersion in HC group (p= 0.056, Figure 4.11a), indicating that the microbial community structure of UCD was more homogeneous than TCD and HC. Beta dispersion at OTU level in the unweighted Unifrac NMDS plot had no significant difference between the HC and TCD groups (p= 0.041, Figure 4.11c).

Table 4.12: PERMANOVA analysis to evaluate the variance using Bray-Curtis and unweighted Unifrac distance matrices at OTU and genus level in TCD, UCD and HC (clustering using USEARCH)

		Bray-Curtis	Unweighted Unifrac
		OTU level	
Overall		p = 0.025, R^2 = 2.76%	p = 0.027, R^2 = 2.49%
Group comparison			
UCD	HC	p = 0.506, R^2 = 1.21%	p = 0.125, R^2 = 1.77%
TCD	HC	p = 0.017, R^2 = 2.32%	p = 0.045, R^2 = 1.57%
	UCD	p = 0.106, R^2 = 2.35%	p = 0.056, R^2 = 2.46%
		Genus level	
Overall		p = 0.026, R^2 = 3.0%	n/a
Group comparison			
UCD	HC	p = 0.515, R^2 = 1.15%	n/a
TCD	HC	p = 0.013, R^2 = 2.56%	n/a
	UCD	p = 0.069, R^2 = 2.77%	n/a

OTU: operational taxonomic unit; n/a: not applicable

The results mentioned above were confirmed when sample sequences were inferred in a collection of amplicon sequencing variants (ASV) using DADA2 algorithm. Alpha diversity had no difference between the three groups (p> 0.05 in all cases, Table 4.11, Figure 4.6). Beta diversity was significantly different between TCD, UCD and HC at ASV and genus level, accounting for bacterial taxon abundance (Bray-Curtis dissimilarity index, p= 0.002 at ASV level and p= 0.001 at genus level, Table 4.13). In the NMDS plot for the Bray-Curtis dissimilarity index at genus level, 3,78% of the variation in community structure was explained by sample groups (p= 0.001, Table 4.13, Figure 4.9). Figure 4.9 shows clearly that gut microbial community structure of TCD patients clustered separately to UCD (p=

0.021, R^2= 3.01%) and HC (p= 0.001, R^2= 3.66%), while there was no difference between HC and UCD patients. Interestingly, this separation between the three groups was not significant when β-diversity was accounted for their phylogenetic relatedness (unweighted Unifrac at ASV level, p> 0.05, Table 4.13).

Table 4.13: PERMANOVA analysis to evaluate the variance using Bray-Curtis and unweighted Unifrac distance matrices at OTU and genus level in TCD, UCD and HC (clustering using DADA2)

		Bray-Curtis	Unweighted Unifrac
		ASV level	
Overall		p = 0.002, R^2 = 3.06%	p = 0.582, R^2 = 1.38%
Group comparison			
UCD	HC	p = 0.522, R^2 = 1.23%	p = 0.560, R^2 = 1.05%
TCD	HC	p = 0.001, R^2 = 2.74%	p = 0.434, R^2 = 0.09%
	UCD	p = 0.029, R^2 = 2.37%	p = 0.556, R^2 = 1.17%
		Genus level	
Overall		p= 0.001, R^2= 3.78%	n/a
Group comparison			
UCD	HC	p = 0.668, R^2 = 0.01%	n/a
TCD	HC	p= 0.001, R^2= 3.66%	n/a
	UCD	p= 0.021, R^2= 3.01%	n/a

ASV: amplicon sequencing variants; n/a: not applicable

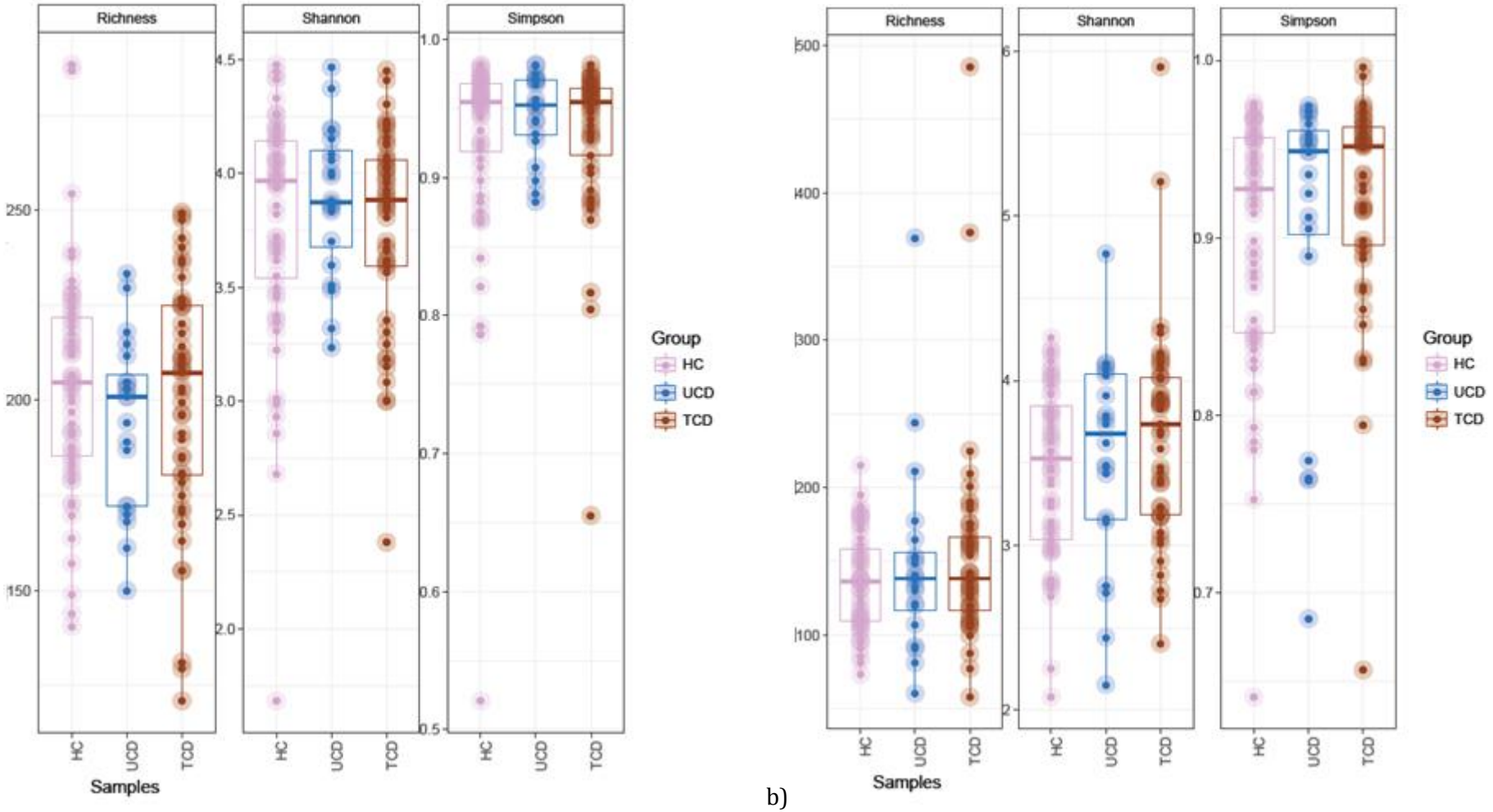


Figure 4.5: Alpha diversity measures at OTU and ASV level using (a) USEARCH and (b) DADA2 algorithms respectively, for TCD, UCD and HC

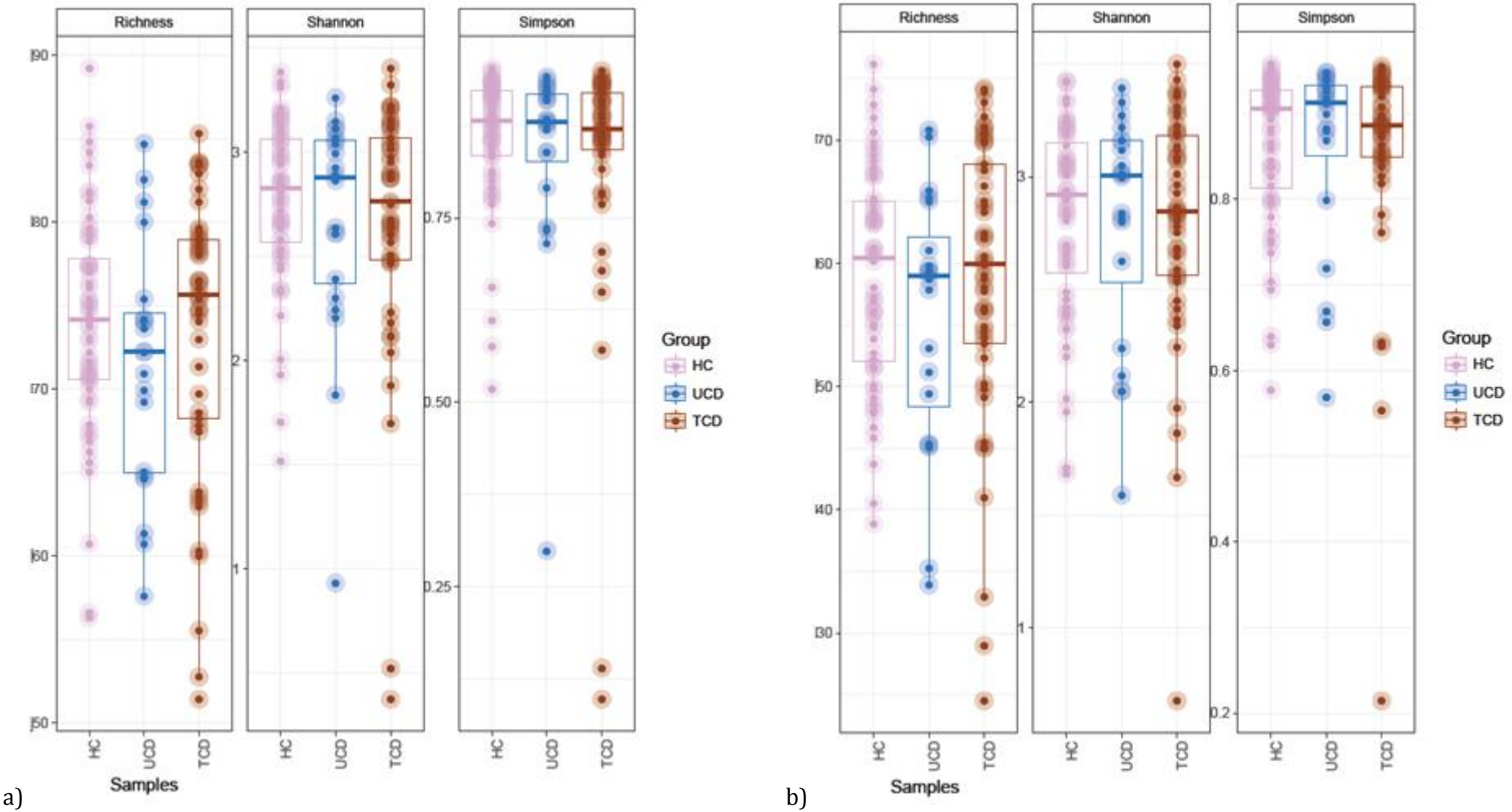


Figure 4.6: Alpha diversity measures at genus level using (a) USEARCH and (b) DADA2 algorithms for TCD, UCD and HC

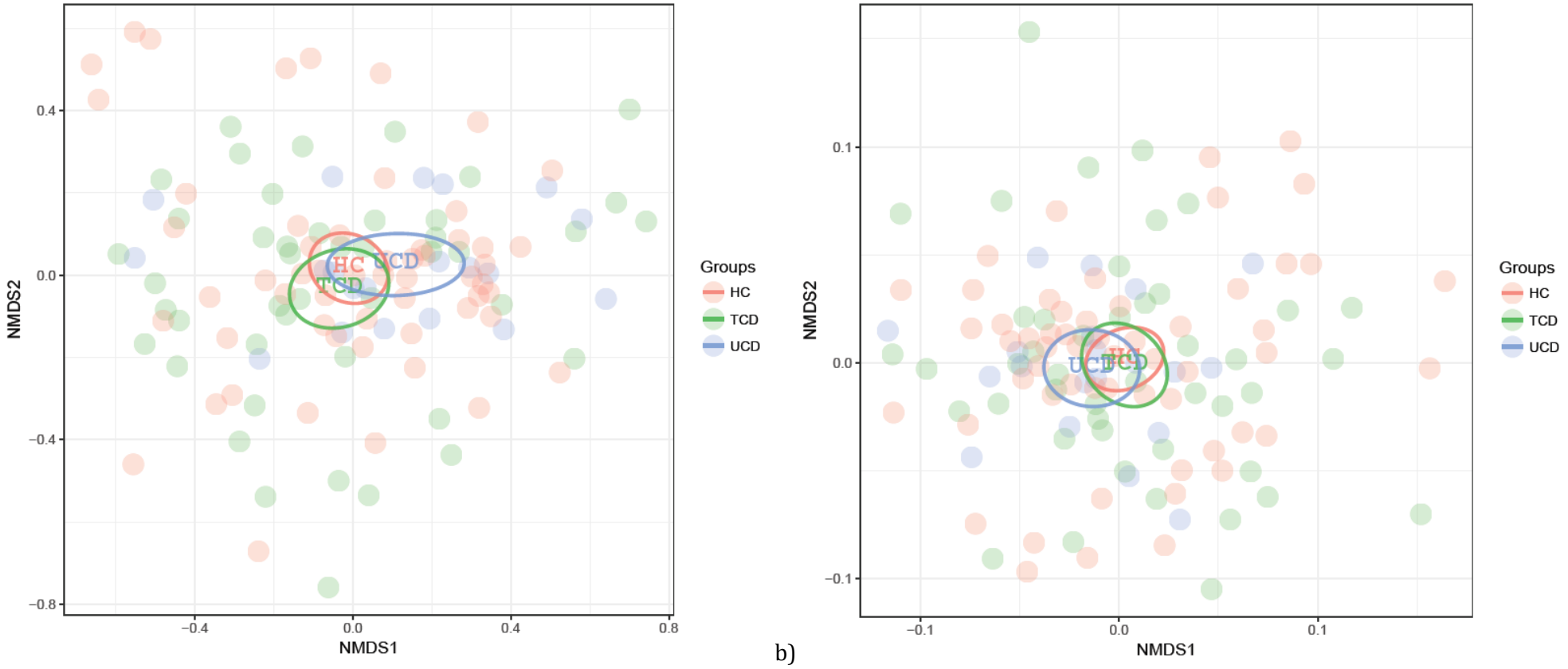


Figure 4.7: Non-metric multidimensional scaling (NMDS) plot at OTU level using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for TCD, UCD and HC (OTU clustering using USEARCH)

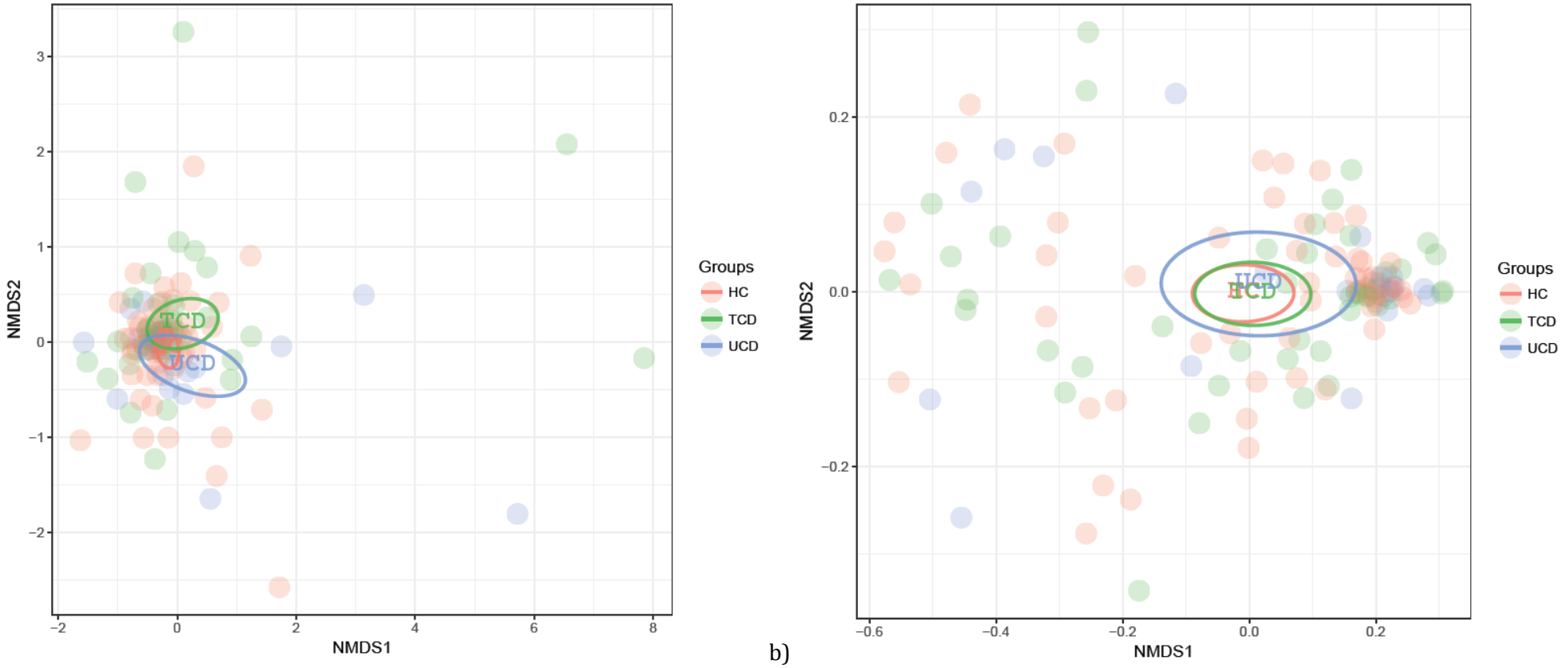


Figure 4.8: Non-metric multidimensional scaling (NMDS) plot at ASV level, using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for TCD, UCD and HC (ASV using DADA2)

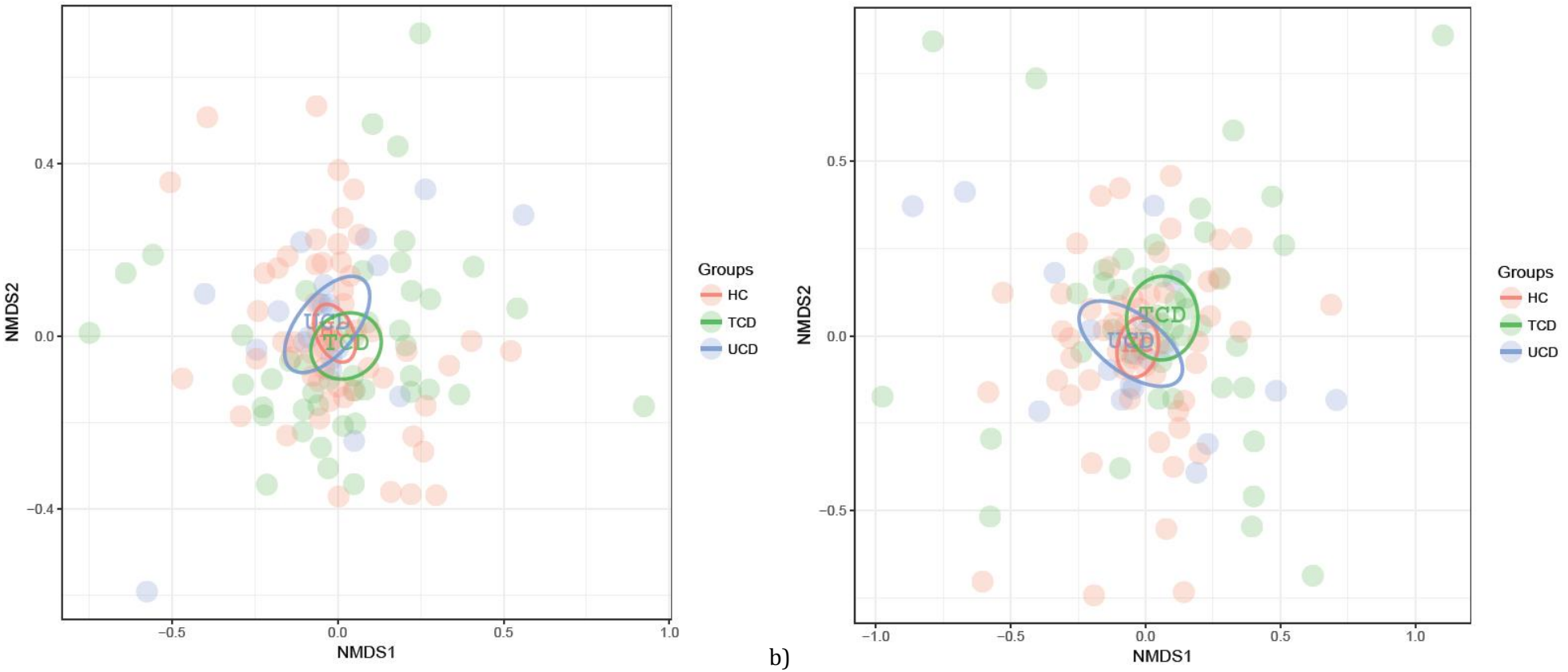


Figure 4.9: Non-metric multidimensional scaling (NMDS) plot at genus level for the Bray-Curtis dissimilarity index using (a) USEARCH and (b) DADA2 algorithms for TCD, UCD and HC

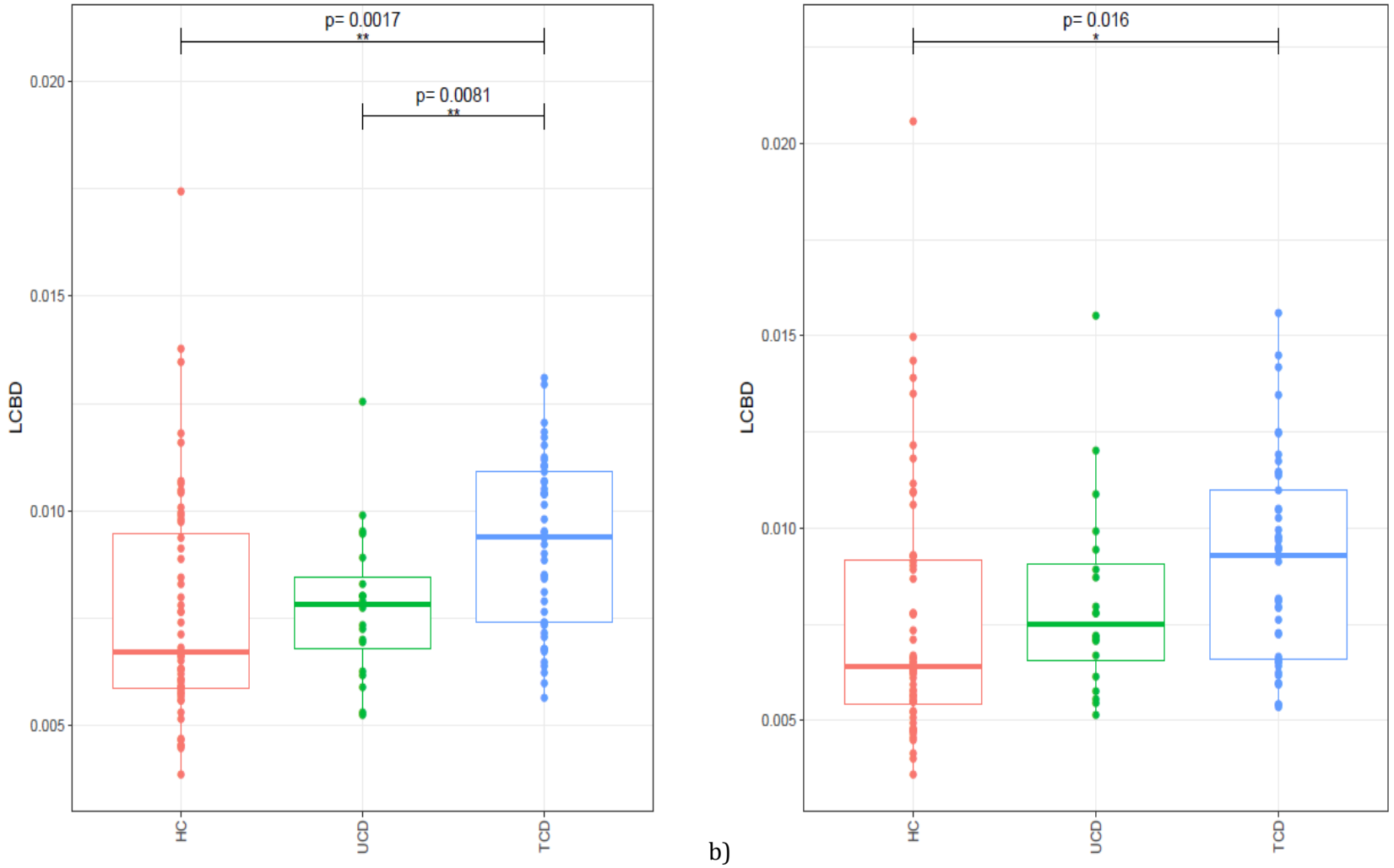


Figure 4.10: Local contribution of β-diversity (LCBD) analysis at (a) OTU and (b) genus level for TCD, UCD and HC (% of total community dispersion)

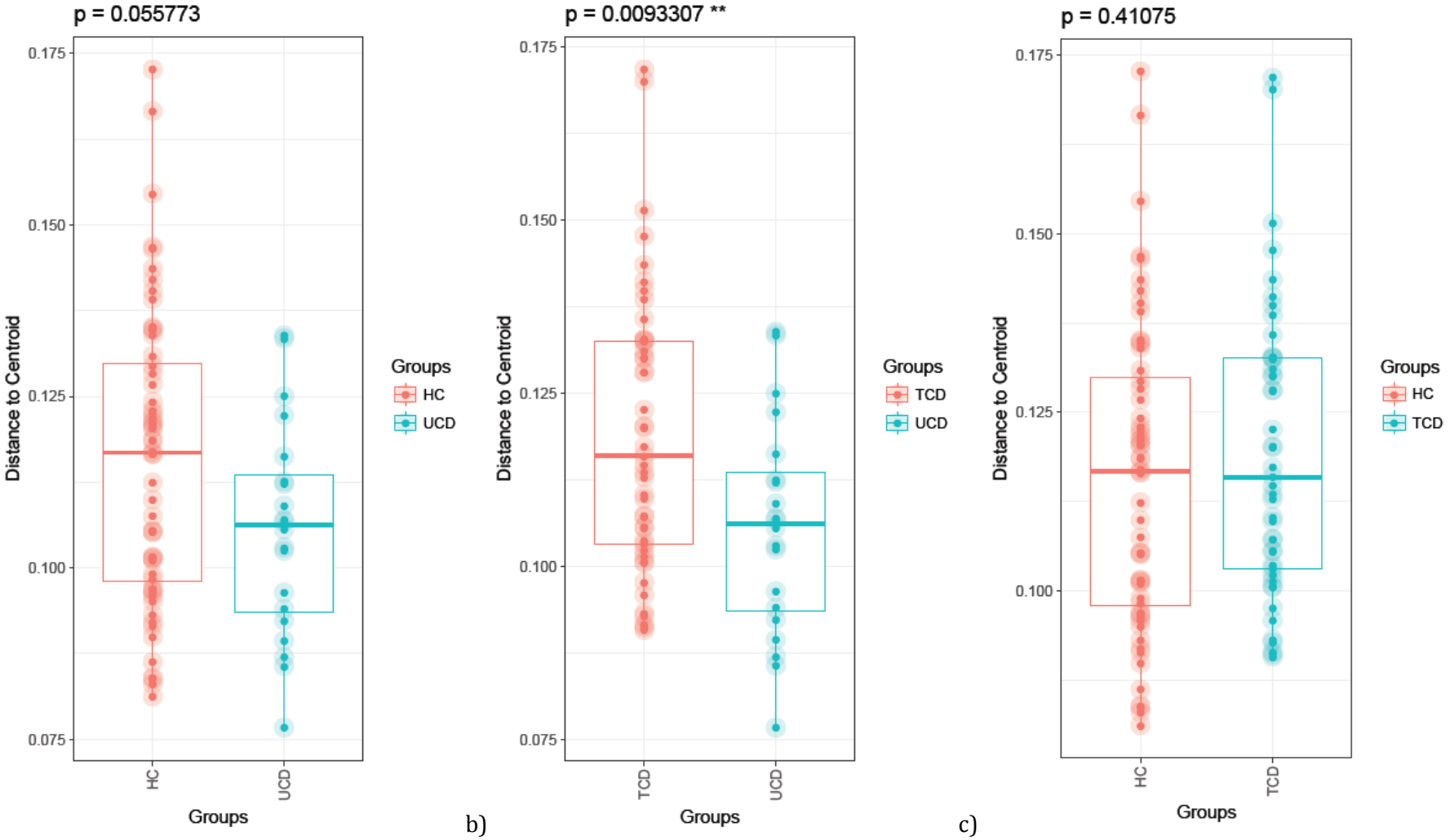


Figure 4.11: Beta dispersion measured as the distance of each sample from the respective group ellipse centroid in (a) HC and UCD patients, (b) UCD and TCD and (c) HC and TCD patients (OTU clustering using USEARCH)

4.2.9.2 Bacterial taxon relative abundance

Faecal samples of UCD, TCD and HC were characterised by a total of 1,082 distinct OTU, with each group being characterised by 795, 961 and 1,005 OTU respectively.

Differences between UCD Vs HC

This thesis explored the OTU that had significantly different relative abundance in faecal samples of UCD children and HC to evaluate whether the CD onset, accompanied by villous atrophy and crypt hyperplasia affects the composition of gut microbiota that populates the large bowel of the patients. This analysis revealed that the relative abundance of 31 of 1,082 OTU (2.87%) was significantly different between UCD and HC groups, with all 31 OTU having lower relative abundance in faecal samples of UCD children than HC (Table 4.14, Figure 4.12). Collectively, a reduction in relative abundance of OTU belonging predominantly to *Methanobacteriaceae* (*p-value (adjusted)*= 6.13E-05), as well as in the genera *Senegalimassilia, Methanobrevibacter, Alloprevotella, Slackia, Ruminococcaceae UCG-014, Parabacteroides, Sutterella, Holdemanella* and *Prevotella* 9 was observed in UCD patients compared with HC ($p< 0.05$, Table 4.15, Figure 4.13). Likewise, a respective decrease in *Methanobacteriales, Enterobacteriales and Euryarcheota* was observed in UCD compared with HC (*p-value (adjusted)*= 3.10E-05, 1.69E-02 and 1.40E-04 respectively).

The relative abundance of these 31 discriminatory OTU was tested against PedsQL-GSS score that was significantly different between HC and UCD groups ($p< 0.0001$, Table 4.4). Kendall rank correlation was used for that purpose and showed that there was no significant correlation between the OTU relative abundance and PedsQL score; neither within HC children nor within UCD group.

Table 4.14: OTU with significantly different relative abundance in faecal samples of UCD, TCD and HC

	Group comparison	BaseMean	log2Fold Change	*p-value*	*p-value (adjusted)*
OTU_60 *Bacteroides*	HC Vs UCD	7.06	-0.23	6.3E-01	7.83E-01
	HC Vs TCD	122.96	5.17	3.03E-23	2.17E-20
	UCD Vs TCD	80.26	4.15	1.43E-08	8.46E-07
OTU_143 *Ruminococcus 1*	HC Vs UCD	25.1	-2.08	8.85E-04	1.16E-02
	HC Vs TCD	18.11	-4.37	8.31E-23	2.98E-20
	UCD Vs TCD	3.30	-2.28	5.05E-08	2.44E-06
OTU_31 *Megamonas*	HC Vs UCD	39.6	-2.09	4.56E-03	3.55E-02
	HC Vs TCD	28.46	-4.72	3.43E-21	8.21E-19
	UCD Vs TCD	4.94	-2.64	3.99E-08	2.12E-06
OTU_244 *Ruminococcus 1*	HC Vs UCD	24.72	-1.37	3.2E-02	1.06E-01
	HC Vs TCD	17.06	-4.26	5.95E-21	1.07E-18
	UCD Vs TCD	2.15	-1.19	1.28E-03	9.45E-03
OTU_78 *Ruminiclostridium 5*	HC Vs UCD	2.71	-0.69	9.89E-02	2.34E-01
	HC Vs TCD	30.16	4.40	1.54E-20	2.21E-18

	UCD Vs TCD	44.97	5.09	1.97E-12	5.22E-10
OTU_42 *Methanobrevibacter*	HC Vs UCD	38.78	-3.84	1.33E-07	1.47E-05
	HC Vs TCD	29.62	-4.36	1.88E-17	2.24E-15
	UCD Vs TCD	2.84	-0.51	2.78E-01	4.25E-01
OTU_114 *Caprococcus 2*	HC Vs UCD	32.92	-0.25	7.13E-01	8.41E-01
	HC Vs TCD	38.73	-4.27	2.92E-17	2.99E-15
	UCD Vs TCD	4.24	-0.79	1.22E-01	2.32E-01
OTU_70 *Ruminicoccaceae*	HC Vs UCD	11.29	-3.43	1.28E-07	1.47E-05
UCG-014	HC Vs TCD	8.89	-3.36	5.94E-14	5.32E-12
	UCD Vs TCD	1.42	0.07	8.52E-01	n/a
OTU_120 *Lachnospiraceae*	HC Vs UCD	4.48	0.63	1.34E-01	2.97E-01
UCG-005	HC Vs TCD	69.03	3.59	7.21E-13	5.74E-11
	UCD Vs TCD	99.05	4.43	3.74E-11	3.43E-09
OTU_135 *Holdemanella*	HC Vs UCD	21.19	-1.04	1.10E-01	2.55E-01
	HC Vs TCD	14.82	-3.20	4.31E-12	3.09E-10
	UCD Vs TCD	5.52	-2.20	3.0E-06	7.23E-05
OTU_197 *Ruminococcaceae*	HC Vs UCD	8.55	1.18	1.27E-02	5.63E-02
UCG - 005	HC Vs TCD	22.96	2.76	1.12E-11	7.29E-10
	UCD Vs TCD	23.17	0.85	1.42E-01	2.56E-01
OTU_908 *Anaerostipes*	HC Vs UCD	12.70	-0.83	1.60E-01	3.32E-01
	HC Vs TCD	8.98	-2.66	4.46E-10	2.66E-08
	UCD Vs TCD	4.04	-1.83	2.57E-04	2.73-03
OTU_979 *Bifidobacterium*	HC Vs UCD	45.70	-0.18	7.36E-01	8.59E-01
Pseudocatenulatum	HC Vs TCD	31.04	-2.13	5.39E-09	2.74E-07
	UCD Vs TCD	19.91	-1.97	3.0E-05	5.69E-04
OTU_53 *Clostridium*	HC Vs UCD	200.98	-2.58	1.1E-05	4.04E-04
sensu stricto 1	HC Vs TCD	239.45	-2.77	5.72E-09	2.74E-07
	UCD Vs TCD	52.52	0.39	5.09E-01	6.45E-01
OTU_186	HC Vs UCD	1.35	-0.11	7.66E-01	n/a
Gastranaerophilales	HC Vs TCD	3.42	2.11	5.74E-09	2.74E-07
	UCD Vs TCD	4.53	2.22	7.91E-05	1.14E-03
OTU_84 *Enterorhabdus*	HC Vs UCD	7.93	-2.34	1.20E-04	2.54E-03
	HC Vs TCD	6.52	-2.32	3.29E-08	1.48E-06
	UCD Vs TCD	1.99	0.02	9.53E-01	9.71E-01
OTU_174 *Catenibacterium*	HC Vs UCD	1.56	0.70	2.74E-02	n/a
	HC Vs TCD	3.19	2.04	3.64E-08	1.54E-06
	UCD Vs TCD	4.47	1.32	1.81E-02	7.2E-02
OTU_336 *Holdemanella*	HC Vs UCD	7.99	-2.38	6.68E-05	1.76E-03
	HC Vs TCD	6.59	-2.30	6.27E-08	2.5E-06
	UCD Vs TCD	2.02	0.06	8.75E-01	9.27E-01
OTU_776 *Bifidobacterium*	HC Vs UCD	21.81	-2.25	5.29E-04	8.59E-03

Animalis	HC Vs TCD	17.47	-2.50	9.08E-08	3.26E-06
	UCD Vs TCD	4.47	1.32	1.81E-02	7.2E-02
OTU_88 *Bacteroides*	HC Vs UCD	281.31	0.02	9.79E-01	9.82E-01
	HC Vs TCD	863.13	2.49	1.02E-07	3.48E-06
	UCD Vs TCD	1,187.7	2.48	2.66E-04	2.77E-03
OTU_1002 *Lachnospira*	HC Vs UCD	8.97	-1.75	8.59E-04	1.16E-02
	HC Vs TCD	33.92	2.49	1.08E-07	3.51E-06
	UCD Vs TCD	45.22	3.23	1.54E-06	3.89E-05
OTU_28 *Bacteroides*	HC Vs UCD	151.07	-1.06	6.32E-02	1.7E-01
	HC Vs TCD	728.21	2.57	2.80E-07	8.72E-06
	UCD Vs TCD	552.60	3.13	1.08E-06	2.85E-05
OTU_1049	HC Vs UCD	53.35	0.60	3.14E-01	5.04E-01
Lachnoclostridium	HC Vs TCD	128.17	2.28	8.26E-07	2.28E-05
	UCD Vs TCD	180.41	1.68	1.09E-02	5.23E-02
OTU_125 *Parabacteroides*	HC Vs UCD	34.72	-1.85	9.65E-04	1.22E-02
	HC Vs TCD	116.40	2.28	9.30E-07	2.47E-05
	UCD Vs TCD	148.10	4.02	3.88E-11	3.43E-09
OTU_278 *Akkermansia*	HC Vs UCD	4.46	0.34	4.50E-01	6.47E-01
	HC Vs TCD	15.59	2.17	1.13E-06	2.88E-05
	UCD Vs TCD	20.63	2.37	1.87E-04	2.2E-03
OTU_537 *Dialister*	HC Vs UCD	18.0	-2.46	1.10E-04	2.43E-03
	HC Vs TCD	15.10	-2.20	1.37E-06	3.40E-05
	UCD Vs TCD	4.68	0.23	6.11E-01	7.29E-01
OTU_62 *Ruminococcaceae*	HC Vs UCD	23.89	-3.94	1.84E-08	1.01E-05
UCG-014	HC Vs TCD	11.97	-2.27	1.69E-06	3.90E-05
	UCD Vs TCD	3.30	0.88	7.0E-02	1.79E-01
OTU_752 *Subdoligranulum*	HC Vs UCD	43.81	-1.87	1.49E-03	1.68E-02
	HC Vs TCD	164.0	2.47	2.72E-06	5.57E-05
	UCD Vs TCD	122.66	3.48	5.24E-07	1.74E-05
OTU_99 *Senegalimassilia*	HC Vs UCD	73.24	-3.88	1.05E-07	1.47E-05
	HC Vs TCD	61.70	-2.47	9.10E-06	1.52E-04
	UCD Vs TCD	14.06	1.48	2.37E-02	8.5E-02
OTU_112 *Bacteroidetes*	HC Vs UCD	22.60	-3.07	1.16E-07	1.47E-05
	HC Vs TCD	31.79	-1.19	1.0E-02	4.2E-02
	UCD Vs TCD	13.92	2.42	1.43E-05	3.15E-04
OTU_1054 *Alistipes*	HC Vs UCD	17.29	-2.95	7.37E-07	6.78E-05
	HC Vs TCD	30.12	0.81	8.83E-02	2.18E-01
	UCD Vs TCD	28.41	3.74	1.82E-09	1.38E-07
OTU_146 *Slackia*	HC Vs UCD	7.95	-3.09	1.02E-06	8.05E-05
	HC Vs TCD	7.16	-1.73	1.27E-04	1.26E-03
	UCD Vs TCD	2.57	1.36	9.44E-03	4.91E-02

OTU_226 [*Eubacterium*]	HC Vs UCD	5.92	-2.79	2.55E-06	1.76E-04
oxideroducens group	HC Vs TCD	6.68	-0.48	2.6E-01	4.48E-01
	UCD Vs TCD	4.11	2.37	4.87E-05	8.33E-04
OTU_45 *Parabacteroides*	HC Vs UCD	56.41	-3.08	3.60E-06	2.21E-04
	HC Vs TCD	73.21	0.84	1.04E-01	2.47E-01
	UCD Vs TCD	69.9	3.44	3.06E-07	1.25E-05
OTU_895 *Barnesiella*	HC Vs UCD	58.0	-3.11	5.04E-06	2.77E-04
	HC Vs TCD	54.13	-1.48	5.16E-03	2.4E-02
	UCD Vs TCD	21.56	1.65	1.02E-02	5.03E-02
OTU_101 *Anaerotruncus*	HC Vs UCD	15.61	-3.11	5.53E-06	2.77E-04
	HC Vs TCD	17.95	-0.46	3.83E-01	5.67E-01
	UCD Vs TCD	11.03	2.64	4.55E-05	8.05E-04
OTU_572 *Family XVIII*	HC Vs UCD	9.54	-2.43	7.14E-06	3.16E-04
AD3011 group	HC Vs TCD	11.51	-0.19	6.55E-01	7.82E-01
	UCD Vs TCD	7.94	2.20	6.21E-05	9.70E-04
OTU_230 *Bacteroides*	HC Vs UCD	6.20	-2.79	7.43E-06	3.16E-04
	HC Vs TCD	6.79	-0.60	1.79E-01	3.62E-01
	UCD Vs TCD	4.02	2.18	1.13E-04	1.58E-03
OTU_65 *Bacteroides*	HC Vs UCD	9.79	-2.44	8.18E-06	3.23E-04
	HC Vs TCD	25.82	1.73	7.37E-05	1.28E-03
	UCD Vs TCD	29.9	4.19	1.17E-12	5.22E-10
OTU_49 *Alistipes*	HC Vs UCD	56.57	-2.4	1.86E-05	6.42E-04
	HC Vs TCD	174.05	1.50	3.31E-03	1.72E-02
	UCD Vs TCD	192.53	4.41	5.95E-12	1.05E-09
OTU_109 *Ruminococcus 2*	HC Vs UCD	6.0	-2.54	2.39E-05	7.75E-04
	HC Vs TCD	5.40	-1.57	2.12E-04	1.88E-03
	UCD Vs TCD	2.22	0.96	3.83E-02	1.17E-01
OTU_51 *Ruminococcaceae*	HC Vs UCD	13.48	-2.49	4.88E-05	1.50E-03
UCG-014	HC Vs TCD	16.15	-0.22	6.38E-01	7.66E-01
	UCD Vs TCD	11.22	2.25	1.20E-04	1.60E-03
OTU_111 *Blautia*	HC Vs UCD	23.54	-2.33	6.14E-05	1.7E-03
	HC Vs TCD	28.78	-0.13	7.87E-01	8.72E-01
	UCD Vs TCD	20.82	2.19	4.69E-04	4.61E-03
OTU_868 *Bacteroides*	HC Vs UCD	136.18	-2.64	7.33E-05	1.76E-03
	HC Vs TCD	293.71	1.33	1.42E-02	5.4E-02
	UCD Vs TCD	310.64	4.83	8.47E-12	1.12E-09
OTU_923 *Mavinbryantia*	HC Vs UCD	10.23	-2.0	7.03E-05	1.76E-03
	HC Vs TCD	13.06	0.07	8.63E-01	9.2E-01
	UCD Vs TCD	10.29	2.06	2.49E-04	2.7E-03
OTU_190 *Alistipes*	HC Vs UCD	6.26	-2.16	1.09E-04	2.43E-03
	HC Vs TCD	11.02	0.07	8.74E-01	9.26E-01

	UCD Vs TCD	13.20	3.35	5.57E-08	2.46E-06
OTU_168 *Alistipes*	HC Vs UCD	4.43	-2.01	4.11E-04	7.31E-03
	HC Vs TCD	3.92	-1.54	8.38E-05	9.40E-04
	UCD Vs TCD	1.75	0.45	2.5E-02	3.9E-01
OTU_3 *Akkermansia*	HC Vs UCD	2,041.4	-2.17	1.41E-03	1.63E-02
	HC Vs TCD	2,863.9	0.33	5.55E-01	7.09E-01
	UCD Vs TCD	2,466.0	1.45	5.12E-02	1.43E-01
OTU_1 *Prevotella 9*	HC Vs UCD	6,760.4	-2.68	3.77E-03	3.15E-02
	HC Vs TCD	5,953.8	-1.79	1.22E-02	4.84E-03
	UCD Vs TCD	193.2	0.85	2.36E-01	3.75E-01
OTU_302 *Eisenbergiella*	HC Vs UCD	26.32	-1.38	1.01E-02	4.74E-02
	HC Vs TCD	55.97	1.44	7.21E-04	5.23E-03
	UCD Vs TCD	97.89	3.49	2.47E-09	1.64E-07
OTU_259	HC Vs UCD	4.86	-1.71	6.88E-04	1.07E-02
Erysipelatoclostridium	HC Vs TCD	8.45	0.95	1.27E-02	5.0E-02
	UCD Vs TCD	8.6	2.61	4.31E-07	1.64E-05
OTU_448 *Ruminiclostridium 5*	HC Vs UCD	9.97	-1.66	5.36E-05	1.56E-03
	HC Vs TCD	15.06	0.61	5.6E-02	1.57E-01
	UCD Vs TCD	13.72	2.20	4.8E-07	1.7E-05
OTU_270 *Butyricimonas*	HC Vs UCD	5.25	-1.88	4.56E-04	7.63E-03
	HC Vs TCD	10.0	1.13	8.16E-03	3.52E-02
	UCD Vs TCD	10.47	3.0	7.49E-07	2.34E-05
OTU_248 *Clostridium*	HC Vs UCD	2.07	-0.77	7.32E-02	1.87E-01
sensu stricto 1	HC Vs TCD	4.37	1.57	2.49E-05	3.5E-04
	UCD Vs TCD	7.89	2.96	9.12E-07	2.69E-05
OTU_317 *Lachnospireaceae*	HC Vs UCD	3.81	-1.83	1.52E-04	2.89E-03
NK4A136 group	HC Vs TCD	7.29	1.56	4.05E-03	1.99E-02
	UCD Vs TCD	7.67	2.97	9.99E-07	2.79E-05
OTU_34 *Bacteroides*	HC Vs UCD	243.37	-1.75	4.11E-03	3.29E-02
	HC Vs TCD	425.59	0.96	3.8E-02	1.19E-01
	UCD Vs TCD	431.67	2.72	6.24E-06	1.44E-04
OTU_235 *Lachnospiraceae*	HC Vs UCD	6.69	-1.52	1.35E-03	1.60E-02
	HC Vs TCD	10.74	0.79	3.9E-02	1.21E-01
	UCD Vs TCD	10.62	2.31	1.95E-05	4.15E-04
OTU_1036 *Bacteroidetes*	HC Vs UCD	59.31	-1.43	3.60E-02	1.14E-01
	HC Vs TCD	76.45	0.23	6.5E-01	7.8E-01
	UCD Vs TCD	61.94	2.57	2.07E-05	4.22E-04
OTU_73 *Bacteroides*	HC Vs UCD	6.31	-1.26	8.05E-03	4.38E-02
	HC Vs TCD	11.31	1.08	7.24E-03	3.24E-02
	UCD Vs TCD	12.07	2.32	3.22E-05	5.89E-04
OTU_154 *Odoribacter*	HC Vs UCD	30.05	-0.95	1.06E-01	2.48E-01

	HC Vs TCD	61.85	1.48	1.13E-03	7.18E-03
	UCD Vs TCD	72.08	2.43	5.23E-05	8.68E-04
OTU_139 *Bilophila*	HC Vs UCD	22.05	-0.67	2.20E-01	3.99E-01
	HC Vs TCD	49.71	0.44	9.57E-05	1.01E-03
	UCD Vs TCD	59.99	2.37	5.65E-05	9.09E-04
OTU_77 *Lachnoclostridium*	HC Vs UCD	47.57	-0.75	1.61E-01	3.32E-01
	HC Vs TCD	120.45	1.11	1.82E-02	6.59E-02
	UCD Vs TCD	128.1	2.40	7.94E-05	1.14E-03
OTU_343 *Christensenellaceae*	HC Vs UCD	1.84	-0.42	2.68E-01	4.55E-01
	HC Vs TCD	4.18	1.82	4.55E-07	1.36E-05
	UCD Vs TCD	4.98	2.14	7.97E-05	1.14E-03
OTU_355 *Lachnoclostridium*	HC Vs UCD	24.25	-0.88	9.18E-02	2.20E-01
	HC Vs TCD	43.33	1.19	3.82E-03	1.9E-02
	UCD Vs TCD	46.44	2.01	1.19E-04	1.60E-03
OTU_98 *Ruminococcaceae*	HC Vs UCD	58.58	-0.55	3.0E-01	4.9E-01
	HC Vs TCD	118.5	1.54	1.46E-04	1.36E-03
	UCD Vs TCD	139.21	2.06	1.39E-04	1.80E-03
OTU_282 *Christensenellaceae*	HC Vs UCD	4.57	-0.49	3.34E-01	5.24E-01
	HC Vs TCD	11.23	1.37	1.8E-03	1.1E-02
	UCD Vs TCD	12.46	2.21	1.59E-04	1.92E-03
OTU_1104 *Akkermansia*	HC Vs UCD	8.04	-0.82	9.76E-02	2.32E-01
	HC Vs TCD	17.76	1.65	2.25E-04	1.97E-03
	UCD Vs TCD	21.41	2.46	1.56E-04	1.92E-03
OTU_46 *Incertae Sedis*	HC Vs UCD	192.02	-0.67	1.88E-01	3.69E-01
	HC Vs TCD	480.32	1.09	1.38E-02	6.17E-02
	UCD Vs TCD	512.68	2.34	1.94E-04	2.24E-03
OTU_63 *Blautia*	HC Vs UCD	154.0	-0.70	2.24E-01	4.02E-01
	HC Vs TCD	341.42	1.69	4.03E-04	3.14E-03
	UCD Vs TCD	415.46	2.33	2.2E-04	2.43E-03
OTU_176 *Lachnoclostridium*	HC Vs UCD	10.20	-0.71	1.39E-01	3.04E-01
	HC Vs TCD	19.53	1.37	8.38E-04	5.89E-03
	UCD Vs TCD	21.77	2.03	2.17E-04	2.43E-03
OTU_134 *Ruminiclostridium*	HC Vs UCD	30.41	-1.31	9.4E-03	4.59E-02
	HC Vs TCD	46.6	0.72	7.65E-02	1.94E-01
	UCD Vs TCD	45.40	2.02	2.77E-04	2.83E-03
OTU_136 [*Eubacterium*]	HC Vs UCD	117.66	-1.32	5.59E-02	1.57E-01
coprostanoligenes group	HC Vs TCD	158.56	0.38	4.9E-01	6.54E-01
	UCD Vs TCD	218.81	2.38	1.57E-03	1.13E-02
OTU_245 *Christensenellaceae*	HC Vs UCD	11.55	-0.64	2.71E-01	4.57E-01
R-7 group	HC Vs TCD	23.45	1.52	1.55E-03	9.45E-03
	UCD Vs TCD	25.55	2.0	2.23E-03	1.54E-02

n/a: not applicable

Table 4.15: Genera with significantly different relative abundance in faecal samples of UCD, TCD & HC

	Group comparison	BaseMean	log2Fold Change	*p-value*	*p-value (adjusted)*
Senegalimassilia	HC Vs UCD	68.25	-3.74	2.06E-07	3.11E-05
	HC Vs TCD	61.51	-2.48	7.48E-06	1.07E-04
	UCD Vs TCD	10.14	0.87	1.63E-01	4.25E-01
Methanobrevibacter	HC Vs UCD	36.76	-3.56	6.4E-07	4.83E-05
	HC Vs TCD	29.09	-4.26	7.33E-17	5.79E-15
	UCD Vs TCD	3.14	-0.65	1.67E-01	4.26E-01
Alloprevotella	HC Vs UCD	8.22	-3.14	1.02E-06	5.13E-05
	HC Vs TCD	6.79	-2.77	1.21E-10	4.79E-09
	UCD Vs TCD	1.48	0.35	3.65E-01	6.23E-01
Slackia	HC Vs UCD	18.05	-3.03	4.73E-06	1.79E-04
	HC Vs TCD	17.07	-1.57	2.14E-03	1.13E-02
	UCD Vs TCD	6.65	1.47	1.67E-02	9.36E-02
Parabacteroides	HC Vs UCD	118.6	-2.65	2.02E-05	5.09E-04
	HC Vs TCD	402.47	2.10	3.98E-05	4.49E-04
	UCD Vs TCD	477.18	4.67	5.51E-13	9.32E-11
Sutterella	HC Vs UCD	20.54	-2.21	8.31E-05	1.79E-03
	HC Vs TCD	21.17	-0.87	3.87E-02	1.18E-01
	UCD Vs TCD	18.81	2.09	1.93E-04	3.62E-03
Holdemanella	HC Vs UCD	61.04	-2.39	4.99E-04	9.43E-03
	HC Vs TCD	68.64	-1.11	4.98E-02	1.36E-02
	UCD Vs TCD	7.34	-2.05	1.75E-05	4.92E-04
Prevotella 9	HC Vs UCD	8,486.25	-2.74	1.88E-03	2.36E-02
	HC Vs TCD	8,421.87	-1.43	3.83E-02	1.18E-01
	UCD Vs TCD	250.21	-0.09	8.94E-01	9.65E-01
Megamonas	HC Vs UCD	39.27	-1.78	1.54E-02	8.3E-02
	HC Vs TCD	27.78	-4.70	2.91E-21	4.60E-19
	UCD Vs TCD	6.03	-3.05	7.67E-10	4.32E-08
Lachnospiraceae UCG - 005	HC Vs UCD	4.20	0.72	6.65E-02	2.05E-01
	HC Vs TCD	63.86	3.54	9.68E-13	5.10E-11
	UCD Vs TCD	89.19	4.29	1.08E-10	9.16E-09
Paraprevotella	HC Vs UCD	3.27	-0.54	2.16E-01	4.49E-01
	HC Vs TCD	10.31	2.36	7.13E-08	2.25E-06
	UCD Vs TCD	12.02	2.73	2.84E-05	6.86E-04
Desulfovibrio	HC Vs UCD	4.78	-0.39	4.26E-01	6.57E-01
	HC Vs TCD	18.59	1.79	1.58E-04	1.56E-03
	UCD Vs TCD	28.58	2.35	4.97E-04	7.64E-03

Eisenbergiella	HC Vs UCD	26.64	-1.25	1.26E-02	7.9E-02
	HC Vs TCD	84.62	1.41	1.57E-03	9.2E-03
	UCD Vs TCD	89.52	3.20	6.83E-09	2.89E-07
Butyricimonas	HC Vs UCD	1.14	0.14	7.28E-01	n/a
	HC Vs TCD	22.29	0.92	5.34E-02	1.43E-01
	UCD Vs TCD	21.79	3.28	3.68E-09	2.86E-07
Odoribacter	HC Vs UCD	26.93	-0.72	2.06E-01	4.49E-01
	HC Vs TCD	59.03	1.50	8.16E-04	5.6E-03
	UCD Vs TCD	67.49	2.28	1.27E-04	2.69E-03
Ruminococcaceae	HC Vs UCD	144.42	-2.91	9.6E-06	2.9E-04
UCG-014	HC Vs TCD	197.83	-1.36	9.38E-03	3.83E-02
	UCD Vs TCD	81.92	0.67	2.91E-01	5.29E-01

n/a: not applicable

Figure 4.12: Log-relative abundance of OTU that differentiated UCD children from HC

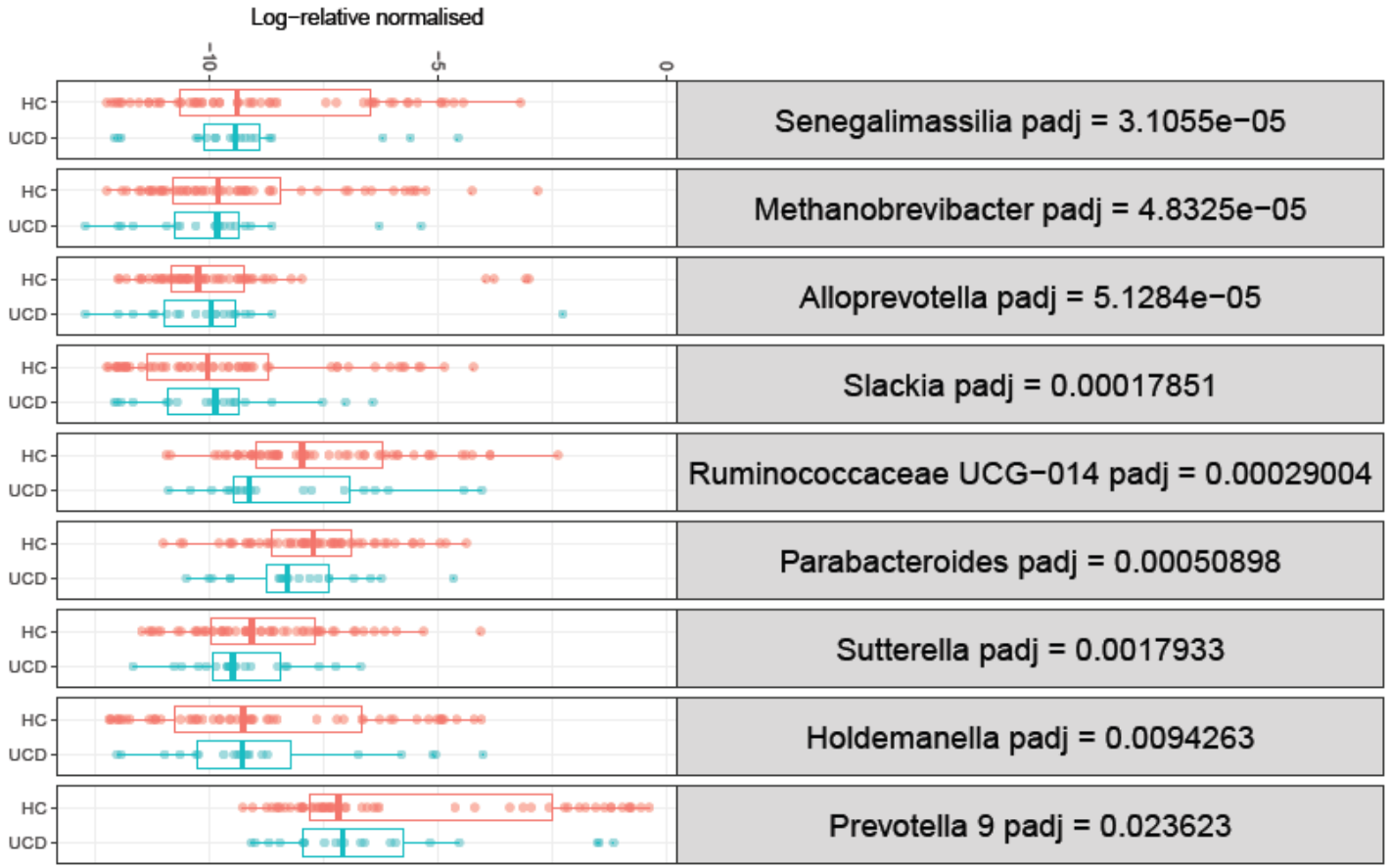


Figure 4.13: Log-relative abundance of genera that differentiated UCD children from HC

Differences between UCD Vs TCD

Likewise, this thesis explored the OTU that had significantly different relative abundance in faecal samples of UCD and TCD children to evaluate whether the treatment with GFD along with the respective disease improvement influences the faecal microbiota composition of CD children. It appeared that fifty-one of 1,082 OTU (4.7%) had significantly different relative abundance in faecal samples of TCD compared with UCD patients (Table 4.14, Figure 4.14). Three of 51 OTU (5.9%) had significantly lower relative abundance in TCD than UCD (OTU_31 *Megamonas*, OTU_143 *Ruminococcus* 1 and OTU_135 *Holdemanella*, Figure 4.14), with the remaining 48 OTU having significantly higher relative abundance in faecal samples of TCD than UCD. Accordingly, all but two genera had higher relative abundance in TCD than UCD patients (Table 4.15, Figure 4.15a). OTU belonging to *Megamonas* and *Holdemanella* had lower relative abundance in TCD than UCD (*p-value (adjusted)*= 4.32E-08 and 4.92E-04 respectively, Table 4.15), whereas OTU belonging to *Parabacteroides, Lachnospiraceae* UCG-005*, Eisenbergiella, Butyricimonas, Paraprevotella, Odoribacter, Sutterella* and *Desulfovibrio* had significantly higher relative abundance in faecal samples of TCD than UCD *(p-value (adjusted)*< 0.05, Table 4.15). As a whole, OTU belonging predominantly to *Gastranaerophilales* and *Cyanobacteria* were significantly more abundant in TCD than UCD patients (*p-value (adjusted)*= 2.41E-04 and 5.83E-05 respectively).

The relative abundance of the 51 discriminatory OTU was tested against PedsQL-GSS score, dietary intake of starch (%TDEI), NMES (%TDEI), NSP (g, %DRV), PUFA (%TDEI), faecal GIP and serum tTG concentration that were significantly different between UCD and TCD children (*p*< 0.05). Kendall rank correlation was used for this purpose, aiming to explore potential correlations of the OTU that differentiated TCD from UCD groups with subjects' clinical and dietary characteristics. Significant positive correlation was found within TCD group, between the dietary intake of starch as a percentage to TDEI and the relative abundance of OTU_186 *Gastranaerophilales* (Kendal Tau= 0.36, *p-value (adjusted)*= 1.64E-02), OTU_112 *Bacteroidetes* (Kendal Tau= 0.41, *p-value (adjusted)*= 6.14E-03) and OTU_73 *Bacteroides* (Kendal Tau= 0.35, *p-value (adjusted)*= 1.64E-02) (Figure 4.15b). Similarly, positive correlation, within TCD group, between the dietary intake of starch (%TDEI) and relative abundance of OTU_186, OTU_112 and OTU_73 was confirmed when Spearman correlation was applied (Spearman rho= 0.51, 0.61 and 0.51, *p-value (adjusted)*= 5.50E-04, 1.68E-05 and 4.75E-04 respectively).

However, there was no significant correlation between the PedsQL-GSS score and the relative abundance of the 51 OTU that differentiated UCD from TCD groups; neither within TCD nor within UCD group. Similarly no significant correlation between faecal GIP concentration and relative abundance of the 51 OTU was observed within the TCD or UCD group. The association of faecal GIP levels with the OTU relative abundance in UCD and TCD groups was further explored, classifying all UCD and TCD children into two groups according to their faecal GIP levels, regardless of disease activity. Group one (1) consisted of CD children with faecal GIP concentration up to 0.156 μg/ gr wet

matter (LOQ), and group two (2) consisted of patients with raised GIP (GIP ≥ 0.156 µg/g wet matter). In total, 121 OTU had significantly different relative abundance between group 1 and group 2. Eighteen of 121 OTU (14.9%) had significantly different relative abundance between UCD and TCD patients as well, suggesting that the difference in the relative abundance of 18 out of 51 OTU (35.3%) between UCD and TCD groups was explained by faecal GIP levels. Table 4.16 lists the 18 OTU that had significantly different relative abundance in both comparisons of UCD versus TCD groups and group 1 versus group 2.

Table 4.16: OTU with significantly different relative abundance in both comparisons UCD Vs TCD and group 1 (GIP < 0.156 µg/g wet matter) Vs group 2 (GIP ≥ 0.156 µg/g wet matter)

	Group comparison	BaseMean	log2Fold Change	*p-value*	*p-value (adjusted)*
OTU_868 *Bacteroides*	UCD Vs TCD	310.6	4.83	8.47E-12	1.12E-09
	group 1 Vs group 2	448.8	-7.0	8.50E-11	1.79E-08
OTU_60 *Bacteroides*	UCD Vs TCD	80.3	4.15	1.43E-08	8.46E-07
	group 1 Vs group 2	437.6	-7.43	8.85E-09	9.31E-07
OTU_28 *Bacteroides*	UCD Vs TCD	552.6	3.13	1.08E-06	2.85E-05
	group 1 Vs group 2	771.2	-5.01	1.27E+07	8.88E-06
OTU_1002 *Lachnospira*	UCD Vs TCD	45.2	3.23	1.54E-06	3.89E-05
	group 1 Vs group 2	63.1	-5.21	131E-06	7.91E-05
OTU_78	UCD Vs TCD	45.0	5.09	1.97E-12	5.22E-10
Ruminiclostridium 5	group 1 Vs group 2	65.2	-5.69	1.79E-06	9.42E-05
OTU_125	UCD Vs TCD	148.1	4.02	3.88E-11	3.43E-09
Parabacteroides	group 1 Vs group 2	211.3	-4.23	6.43E-06	2.70E-04
OTU_752	UCD Vs TCD	122.66	3.48	5.24E-07	1.74E-05
Subdoligranulum	group 1 Vs group 2	160.6	-3.75	5.69E-04	6.14E-03
OTU_101 *Anaerotruncus*	UCD Vs TCD	11.0	2.64	4.55E-05	8.05E-04
	group 1 Vs group 2	35.7	-3.78	1.01E-03	9.66E-03
OTU_1054 *Alistipes*	UCD Vs TCD	28.4	3.74	1.82E-09	1.38E-07
	group 1 Vs group 2	40.2	-3.11	1.58E-03	1.36E-02
OTU_88 *Bacteroides*	UCD Vs TCD	1,187.7	2.48	2.66E-04	2.77E-03
	group 1 Vs group 2	1,614.8	-3.14	2.06E-03	1.55E-02
OTU_278 *Akkermansia*	UCD Vs TCD	20.6	2.37	1.87E-04	2.2E-03
	group 1 Vs group 2	27.8	-3.16	2.04E-03	1.55E-02
OTU_302 *Eisenbergiella*	UCD Vs TCD	97.9	3.49	2.47E-09	1.64E-07
	group 1 Vs group 2	88.6	-2.49	3.16E-03	1.99E-02
OTU_65 *Bacteroides*	UCD Vs TCD	29.9	4.19	1.17E-12	5.22E-10
	group 1 Vs group 2	42.1	-2.60	4.36E-03	2.39E-02

OTU_120 *Lachnospiraceae*	UCD Vs TCD	99.1	4.43	3.74E-11	3.43E-09
UCG-005	group 1 Vs group 2	141.6	-3.07	4.36E-03	2.39E-02
OTU_136 [Eubacterium]	UCD Vs TCD	218.8	2.38	1.57E-03	1.13E-02
coprostanoligenes group	group 1 Vs group 2	288.1	-3.13	6.40E-03	2.97E-02
OTU_154 *Odoribacter*	UCD Vs TCD	72.1	2.43	5.23E-05	8.68E-04
	group 1 Vs group 2	97.3	-2.43	7.10E-03	3.08E-02
OTU_230 *Bacteroides*	UCD Vs TCD	4.0	2.18	1.13E-04	1.58E-03
	group 1 Vs group 2	8.1	-2.66	7.58E-03	3.19E-02
OTU_134	UCD Vs TCD	45.4	2.02	2.77E-04	2.83E-03
Ruminiclostridium 5	group 1 Vs group 2	59.7	-2.12	1.40E-02	4.7E-02

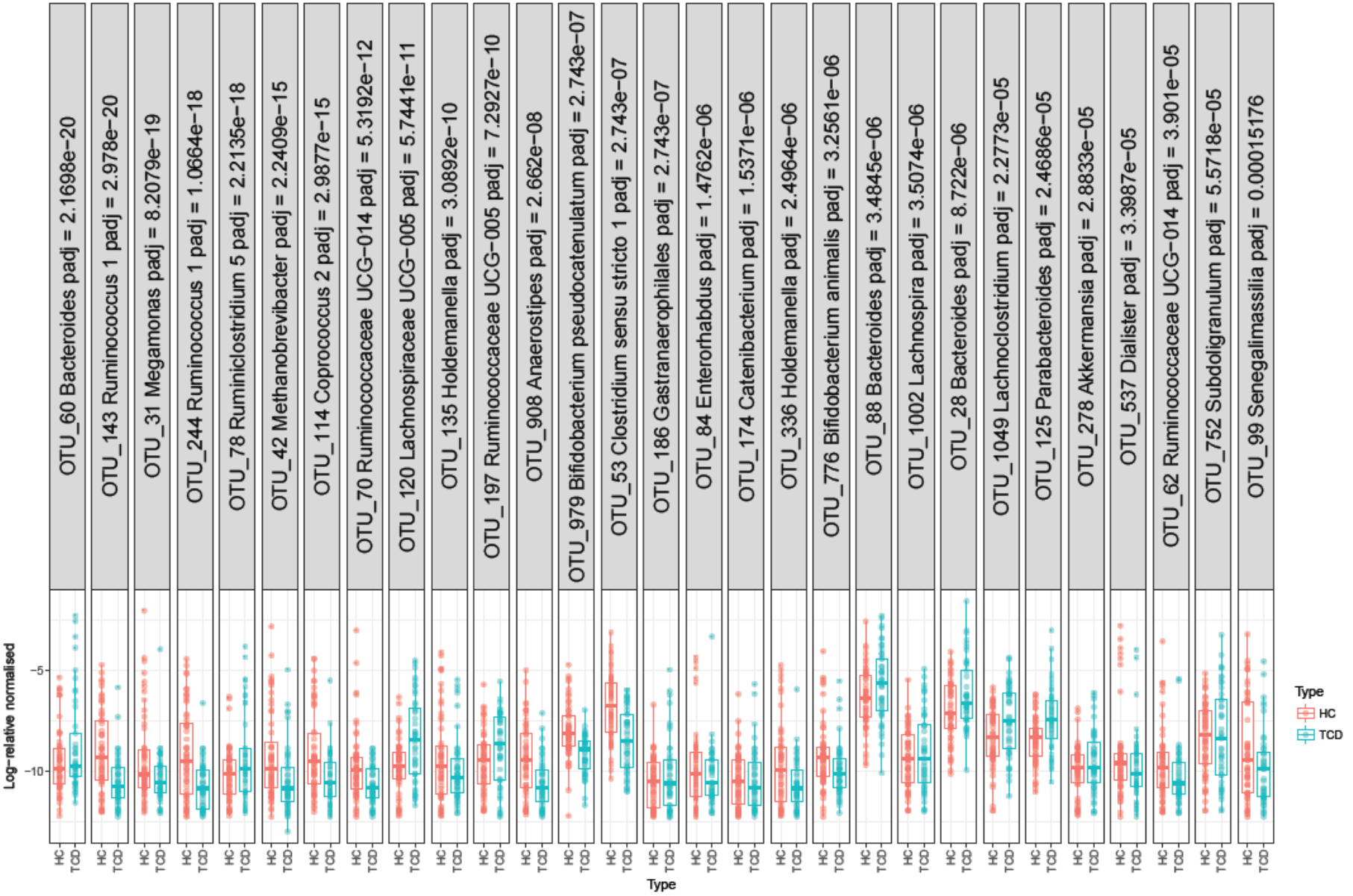
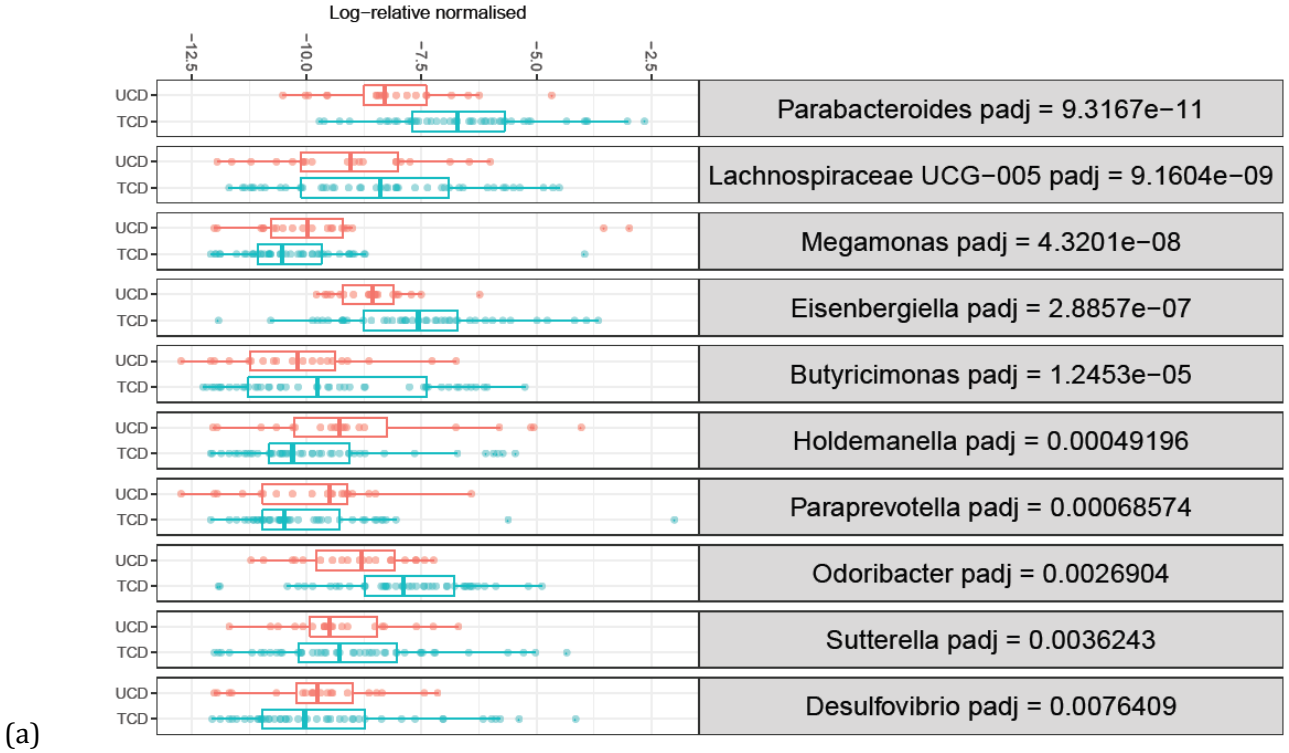
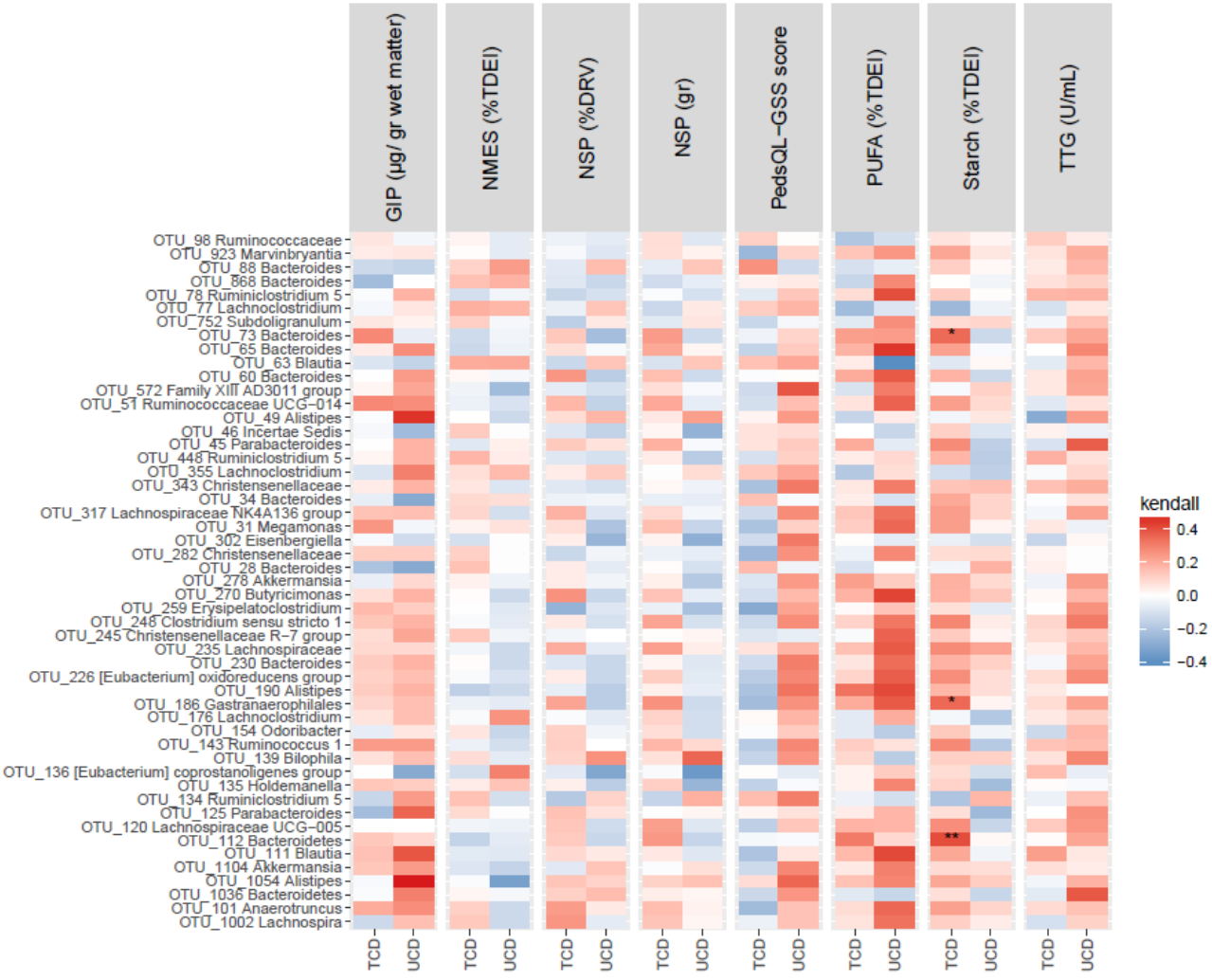


Figure 4.14: Log-relative abundance of OTU that differentiated TCD from UCD children



(a)



(b)

Figure 4.15: (a) Log-relative abundance of genera that differentiated TCD from UCD; (b) Kendall rank correlation of the OTU that differentiated the TCD from UCD and faecal GIP concentration, PedsQL-GSS score, dietary intake of NMES (%TDEI), NSP (g, %DRV), Starch (%TDEI) and PUFA (%TDEI)

Differences between HC Vs TCD

Furthermore, this thesis explored the OTU that had significantly different relative abundance in faecal samples of TCD and HC children, in order to evaluate whether the onset of CD and the compliance with GFD affect the composition of the gut microbiota. This analysis revealed that twenty-nine of 1082 OTU (2.7%) had significantly different relative abundance between HC and TCD groups (Figure 4.16, Table 4.14). Thirteen out of the 29 OTU (44.8%) had higher relative abundance in faecal samples of TCD than HC, while the remaining 16 OTU (55.2%) had significantly lower relative abundance in TCD than HC (Table 4.14). Similarly exploring differences at genus level it was found that OTU belonging to *Paraprevotella, Parabacteroides* and *Lachnospiraceae* UCG-005 had significantly higher relative abundance in TCD group than HC (*p-value (adjusted)*= 2.25E-06, 4.49E-04 and 5.10E-11 respectively, Table 4.15). In contrast, OTU belonging to *Megamonas, Methanobrevibacter, Alloprevotella* and *Senegalimassilia* had significantly higher relative abundance in HC than TCD (*p-value (adjusted)*< 0.05, Figure 4.17a, Table 4.15). A respective significant increase in relative abundance of OTU belonging widely to *Methanobacteriaceae, Methanobacteriales, Methanobacteria* and *Euryarcheota* was observed in HC compared with TCD, along with a decrease of OTU belonging to *Gastranaerophilales* and *Melainabacteria*.

The relative abundance of the 29 distinctive OTU was plotted, within both the TCD and HC group, against the concentration and relative abundance of SCFA that were significantly different between HC and TCD groups, as well as against the PedsQL-GSS score that differed significantly between the two groups (*p*< 0.0001, Table 4.4). Kendall rank correlation was applied between the relative abundance of the 29 OTU and (a) PedsQL-GSS score, (b) concentration of propionic, butyric, isobutyric, valeric, isovaleric, caproic and isocaproic acid expressed per total faecal output, (c) concentration of caproic and isocaproic acid in wet matter (μmol/g wet matter) and (d) relative abundance of acetic, butyric and valeric acid (%), and figure 4.17b shows that there was no significant correlation.

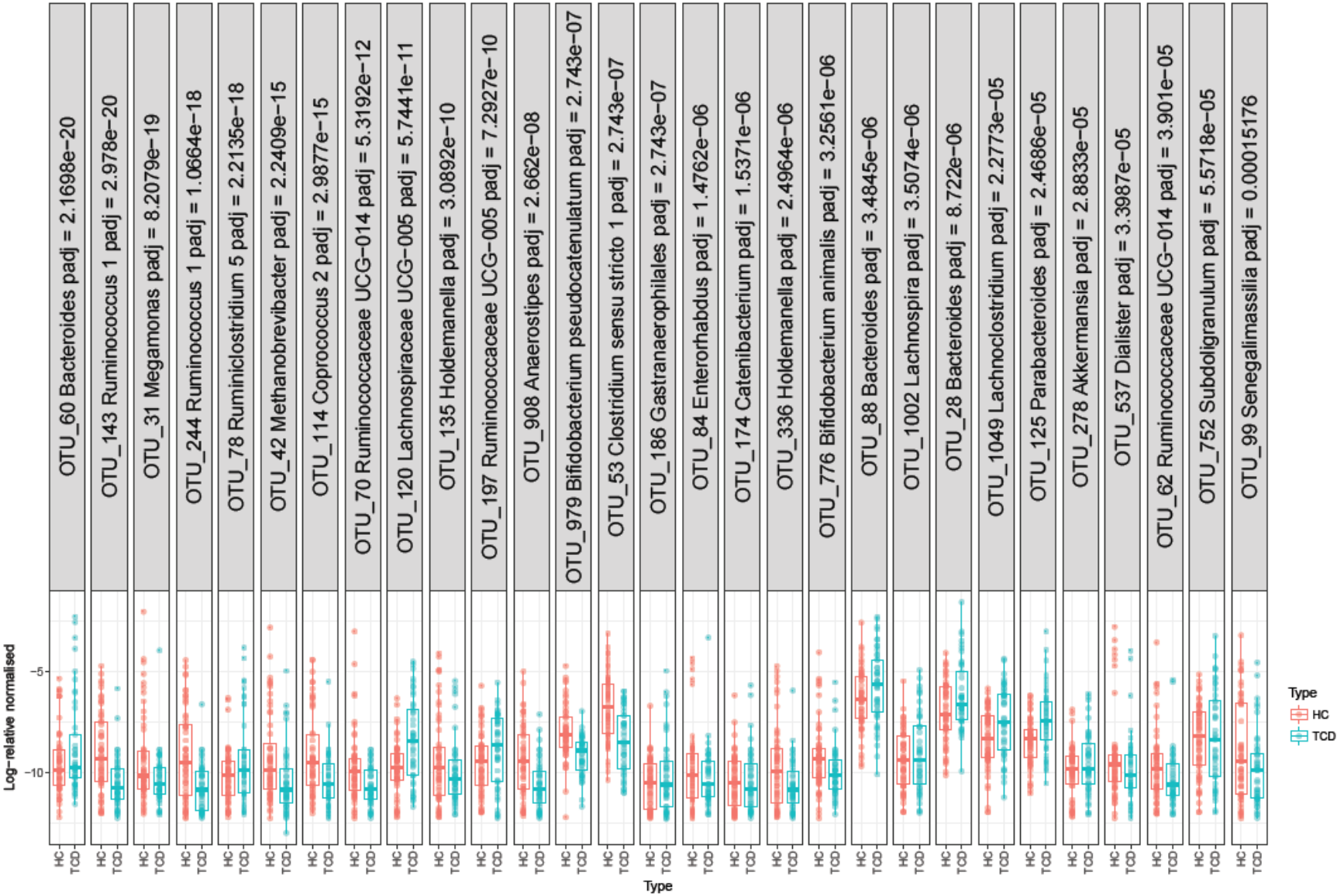
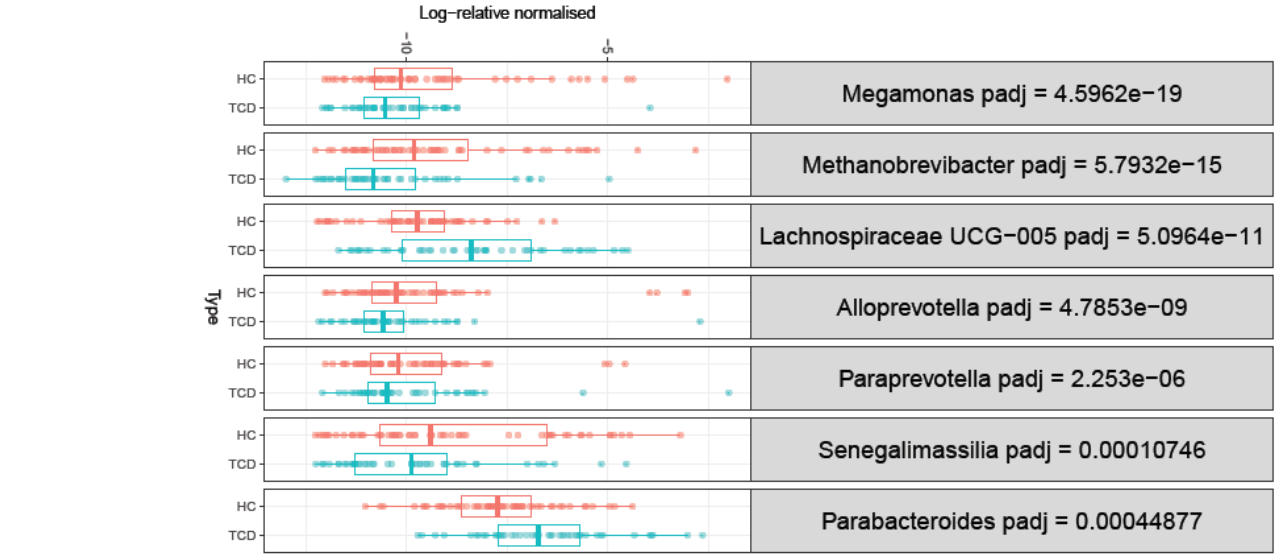
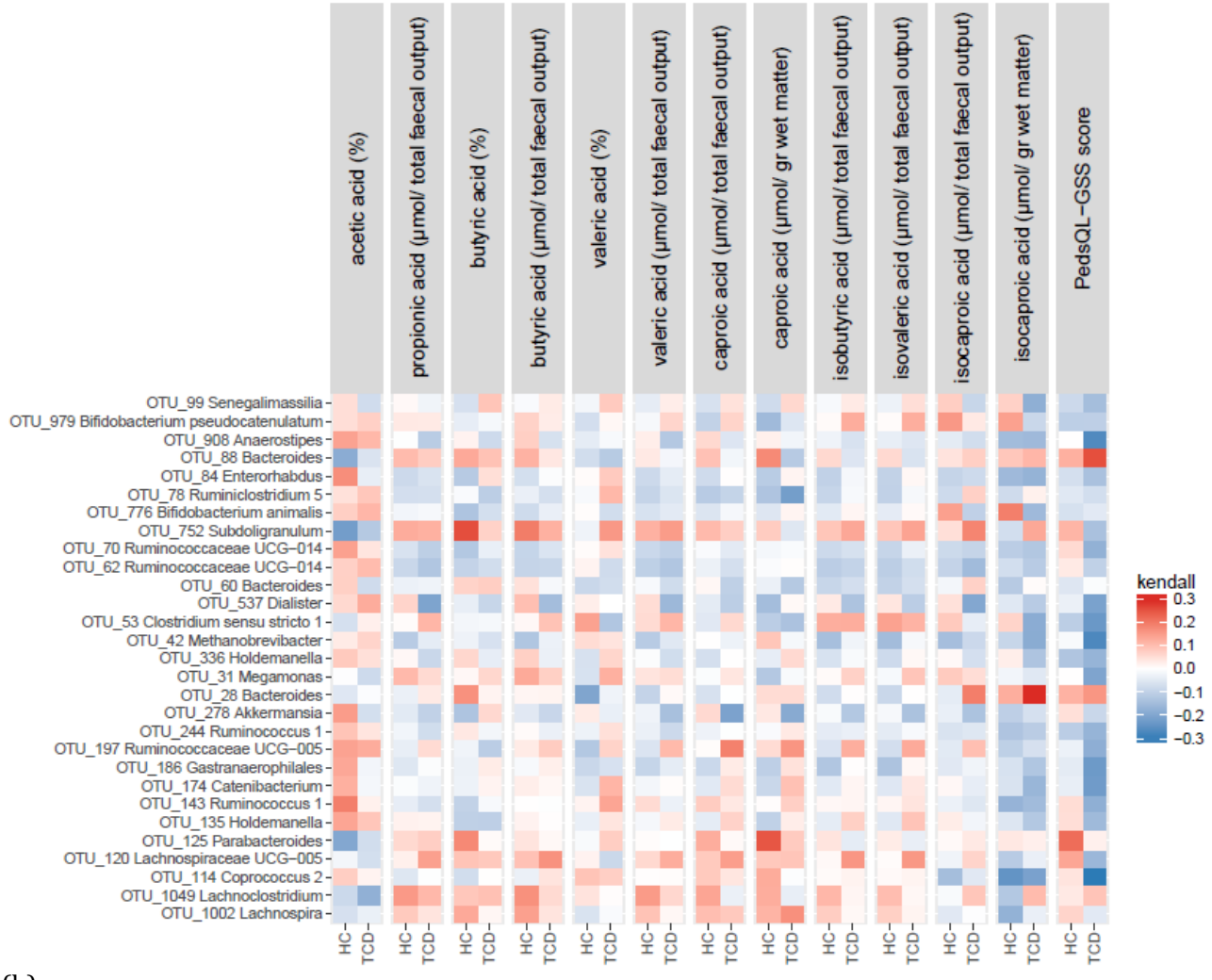


Figure 4.16: Log-relative abundance of OTU that differentiated TCD from HC children



(a)



(b)

Figure 4.17: (a) Log-relative abundance of genera that differentiated TCD from HC; (b) Kendall rank correlation of OTU that differentiated TCD from HC and acetic (%), propionic (µmol/ total faecal output), butyric (%, µmol/ total faecal output), valeric acid (%, µmol/ total faecal output), caproic (µmol/gr wet matter, µmol/ total faecal output), isobutyric (µmol/ total faecal output), isovaleric (µmol/ total faecal output), isocaproic (µmol/gr wet matter, µmol/ total faecal output), and the PedsQL-GSS score

4.3 Prospective study

All UCD patients went on GFD treatment, with 13 of 20 (65%) providing faecal samples at six and 12 months after the initiation of GFD. Table 4.17 shows that there was no significant difference in the z-score of weight, height and BMI of the 13 CD children during follow-up (p> 0.05).

Table 4.17: Anthropometric characteristics of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD-6 months (13)	GFD-12 months (13)	*p-overall*
Weight (Kg)	30.5 (3.1)[a]	32.5 (3.3)[b]	34.8 (3.5)	< 0.0001
Weight z-score	-0.34 (0.33)	-0.36 (0.30)	-0.29 (0.30)	0.712
< -2 SD (%)	2 (15.4)	2 (15.4)	2 (15.4)	-
> 2 SD (%)	0 (0)	0 (0)	0 (0)	-
Height (m)	1.3 (0.05)[a]	1.4 (0.05)[b]	1.40 (0.05)	< 0.0001
Height z-score	-0.19 (0.30)	-0.15 (0.28)	-0.20 (0.28)	0.917
< -2 SD (%)	2 (15.4)	1 (7.7)	1 (7.7)	-
> 2 SD (%)	0 (0)	0 (0)	0 (0)	-
BMI (Kg/m^2)	16.5 (0.67)[c]	16.6 (0.71)	17.14 (0.79)	0.014
BMI z-score	-0.40 (0.34)	-0.45 (0.33)	-0.31 (0.33)	0.334
< -2 SD (%)	2 (15.4)	2 (15.4)	2 (15.4)	-
< -2 SD (%)	0 (0)	0 (0)	0 (0)	-

Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p<0.0001 compared to GFD-6 months & GFD-12 months; [b]: p< 0.0001 compared to GFD-12 months; [c]: p=0.018 compared to GFD-12 months

4.3.1 Dietary intake

Dietary intake of carbohydrates, proteins and fat (g, %TDEI), along with TDEI as a percentage to EAR did not change significantly during follow-up (*p-overall*> 0.05, Table 4.18). However, the dietary intake of PUFA (%TDEI), NMES (%TDEI), NSP (g, %DRV), starch (%TDEI) and dietary fibre (g) changed significantly (*p-overall*= 0.027, 0.01, < 0.0001, 0.001, < 0.0001 and 0.006 respectively, Table 4.18). Compared to baseline values, consumption of PUFA and dietary fibre was significantly different at six months on GFD; mean (SEM) PUFA intake (%TDEI) decreased from 4.4 (0.44) to 3.1 (0.11) (p= 0.036), whereas mean (SEM) dietary fibre (gr) intake increased from 11.6 (2.3) to 17.9 (1.9) (p= 0.005). Accordingly, compared to the diagnosis values, mean (SEM) dietary intake of NMES (%TDEI) (16.0 (1.1)) and NSP [9.8 (1.1) g, 58.3 (7.7) %DRV] was significantly lower both at six [NMES (%TDEI): 10.4 (1.3), p= 0.001; NSP (g): 4.9 (0.62), p< 0.0001; NSP (%DRV): 29.4 (4.3), p= 0.001] and 12 months on GFD [NMES (%TDEI): 12.2 (1.5), p= 0.030; NSP (g): 5.2 (0.74), p< 0.0001; NSP (%DRV): 27.3 (4.3), p= 0.001, Table 4.18]. Compared with the baseline value point (22.16 (3.16)), mean (SEM) starch intake

(%TDEI) was significantly decreased at six (5.05 (1.45)) and 12 months on GFD (5.15 (0.86), $p<$ 0.0001) as well (Table 4.18).

Table 4.18: Dietary intake of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD-6 months (13)	GFD-12 months (13)	*p-overall*
Energy (%EAR)	93.8 (6.9)	97.1 (6.7)	89.2 (6.3)	0.639
Fat (g)	59.8 (5.5)	64.0 (4.9)	63.5 (4.7)	0.724
Fat (%TDEI)	30.9 (1.2)	30.5 (1.0)	32.0 (0.92)	0.523
SFA (%TDEI)	13.4 (0.60)	13.9 (0.73)	13.4 (0.61)	0.756
PUFA (%TDEI) [A]	4.4 (0.44) [a]	3.1 (0.31)	4.2 (0.44)	0.027
MUFA (%TDEI)	7.2 (0.58)	6.5 (0.49)	7.2 (0.60)	0.502
Protein (g)	55.7 (4.7)	58.5 (4.5)	57.8 (4.3)	0.830
Protein (%TDEI)	13.2 (0.41)	12.9 (0.52)	13.4 (0.36)	0.321
Protein (%RNI)	195.1 (16.7)	180.2 (13.6)	168.6 (14.5)	0.427
Carbohydrates (g)	235.4 (18.9)	250.1 (16.0)	231.2 (14.8)	0.563
Carbohydrates (%TDEI)	53.0 (1.4)	52.4 (1.4)	50.8 (1.3)	0.400
Sugars (%TDEI)	36.7 (4.2)	40.1 (3.7)	40.8 (4.8)	0.673
Starch (%TDEI) [B]	22.2 (1.36) [b]	5.1 (1.45)	5.2 (0.86)	< 0.0001
NMES (%TDEI) [C]	16.0 (1.1) [c, d]	10.4 (1.3)	12.2 (1.5)	0.001
NSP (g) [D]	9.8 (1.1) [e]	4.9 (0.62)	5.2 (0.74)	< 0.0001
NSP (%DRV) [E]	58.3 (7.7) [f]	29.4 (4.3)	27.3 (4.3)	0.001
Dietary fibre (g) [F]	11.6 (2.3) [g]	17.9 (1.9)	14.4 (1.6)	0.006

EAR: estimated average requirement, RNI: reference nutrient intake; TDEI: total daily energy intake; SFA: saturated fatty acids; PUFA: polyunsaturated fatty acids; MUFA: monounsaturated fatty acids; NMES: non-milk extrinsic sugars; NSP: non-starch polysaccharides; Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p= 0.036 compared to GFD-6 months; [b]: $p<$ 0.0001 compared to GFD-6 months & GFD-12 months; [c]: p= 0.001 compared to GFD-6 months; [d]: p= 0.030 compared to GFD-12 months; [e]: $p<$ 0.0001 compared to GFD-6 months & GFD-12 months; [f]: p= 0.001 compared to GFD-6 months & GFD-12 months; [g]: p= 0.005 compared to GFD-6 months

4.3.2 Compliance with GFD

All 13 CD patients (100%) reported strict compliance with GFD (Biagi score > 3) six months after the initiation of GFD; 10 children (76.9%) scored Biagi = 3, while three children (23.1%) scored Biagi = 4. Accordingly, at 12 months on GFD, one patient (9.1% of 11 CD children that filled in the Biagi questionnaire at 12 months on GFD) reported bad GFD adherence (Biagi score= 1), with the remaining

10 patients (90.1% of 11 CD children that filled in the Biagi questionnaire at 12 months on GFD) reporting strict GFD adherence (Biagi score = 3-4) (Table 4.19).

Faecal GIP levels changed significantly during the follow up (*p-overall*< 0.0001, Table 4.19). Compared to the mean (SEM) baseline value (2.95 (0.76)) (µg/g wet matter), GIP concentration was almost 13 times lower at six months on GFD (0.22 (0.06), *p*< 0.0001) and six times lower at 12 months on GFD (0.49 (0.23), *p*< 0.0001, Table 4.19). Although the GIP concentration was two folds higher at 12 months on GFD than at six months on GFD, that difference was not significant. Based on the LOQ of the ELISA assay, results were classed into three ranges: (a) negative (GIP < 0.156µg/g sample), (b) weak positive (GIP = 0.156 – 0.30µg/g sample) and (c) strong positive (GIP > 0.30µg/g sample) [22]. Figure 4.18 shows the distribution of the samples among these three classes throughout the observational period. Based on any positive GIP results (GIP concentration > 0.156 µg/g faeces), two of 13 (15.4%), and three of 13 (23.1%) CD patients deemed to be non-compliant at six and 12 months on GFD respectively (Table 4.19).

Table 4.19: Biagi score & GIP concentration in faecal samples of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD-6 months (13)	GFD-12 months (13)	*p-overall*
Biagi score			[2]	
1 [n (%)]	-	0 (0)	1 (9.1)	-
2 [n (%)]	-	0 (0)	0 (0)	-
3 [n (%)]	-	10 (76.9)	8 (72.7)	-
4 [n (%)]	-	3 (23.1)	2 (18.2)	-
GIP (µg/g wet matter)	2.95 (0.76) [a] [1]	0.22 (0.06)	0.49 (0.23)	< 0.0001
< 0.156 [n (%)]	0 (0)	11 (84.6)	10 (76.9)	-
0.156 – 0.30 [n (%)]	0 (0)	1 (7.7)	0 (0)	-
> 0.30 [n (%)]	12 (100)	1 (7.7)	3 (23.1)	-

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*< 0.0001 compared to GFD-6 months & GFD-12 months

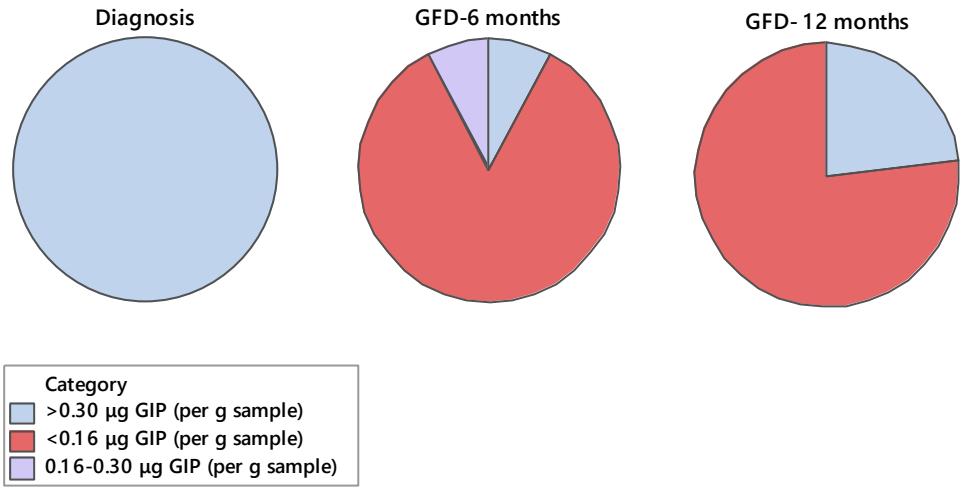


Figure 4.18: Classification of CD children into three groups, based on faecal GIP levels at diagnosis and at six and 12 months on GFD

4.3.3 Serological markers

Serum tTG concentration changed significantly during the follow-up (*p-overall*< 0.0001, Table 4.20). Compared with the mean (SEM) baseline value (68.5 (19.6)) (U/mL), there was an almost sevenfold decrease in tTG levels at six months on GFD (9.8 (3.3), *p*< 0.0001) and a respective ninefold decrease at 12 months on GFD (7.7 (2.0), *p*< 0.0001, Table 4.20). Accordingly, serum tTG concentration was significantly lower at 12 months on GFD compared to the value at six months on GFD (*p*= 0.029, Table 4.20). Although 83.3% of the UCD patients (five of six UCD children for whom serum tTG was measured) had tTG levels higher than 7 U/mL at diagnosis, at six and 12 months on GFD the respective percentage fell to 66.7% (six of nine children for whom serum tTG was measured at six months on GFD) and 33.3% (three of nine children for whom serum tTG was measured at 12 months on GFD, Table 4.20). Finally, there was no difference in IgA, IgG and IgM levels during the follow-up (*p-overall*= 0.952, 0.259 and 0.370 respectively).

Table 4.20: CD serological markers in CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD-6 months (13)	GFD-12 months (13)	*p-overall*
tTG (U/mL)	68.5 (19.6)[a] [7]	9.8 (3.3)[b] [4]	7.7 (2.0) [4]	< 0.0001
< 7 [n (%)]	1 (16.7)	3 (33.3)	6 (66.7)	-
≥ 7 [n (%)]	5 (83.3)	6 (66.7)	3 (33.3)	-
IgA (g/L)	1.16 (0.08)	1.19 (0.16) [9]	0.90 [12]	0.952
IgG (g/L)	8.72 (0.47)	9.07 (0.29) [9]	9.10 [12]	0.259
IgM (g/L)	0.92 (0.09)	1.04 (0.05) [9]	1.06 [12]	0.370

Values expressed as mean (SEM); the number of missing data is shown in brackets; at 12 months on GFD IgA, IgG and IgM were only measured in one CD child, thus no mean (SEM) is calculated; GLM accounted for paired data; Box-Cox

transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: $p < 0.0001$ compared to GFD-6 months & GFD-12 months; [b]: $p = 0.029$ compared to GFD-12 months

4.3.4 Gastrointestinal symptoms (PedsQL-GSS questionnaire)

GI symptoms were evaluated using the PedsQL-GSS questionnaire. At diagnosis six of 13 UCD children (46%) reported high levels (3-4) of abdominal pain, abdominal discomfort and passing wind, while four of 13 UCD patients (30.8%) reported high levels (3-4) of constipation and not feeling hungry (Table 4.21). During follow-up, the percentage of CD children that reported high levels of abdominal pain fell to 23.1% (three CD children) at six months on GFD and to 16.6% (two CD children) at 12 months on GFD. Compared to the diagnosis, the percentage of CD patients that reported high levels of passing wind did not decrease at six sixth months on GFD, but fell to 33.2% (four CD children) at 12 months on GFD (Table 4.21). The proportion of CD patients who reported high levels of abdominal discomfort fell from 46.2% (six CD children) to 38.5% (five CD children) at six months on GFD and to 24.9% (three CD children) at 12 months on GFD (Table 4.21).

PedsQL-GSS score was significantly different during the follow-up ($p = 0.011$). It increased (mean, SEM) significantly from 58.3 (6.2) to 73.6 (6.4) 12 months after the initiation of GFD ($p=0.009$). However, there was no significant difference at six months on GFD (Table 4.21).

Table 4.21: Frequency of self-reported GI symptoms in CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD – 6 months (13)	GFD – 12 months (12)
Abdominal pain			
0	1 (7.7)	1 (7.7)	4 (33.2)
1	3 (23.1)	3 (23.1)	3 (24.9)
2	3 (23.1)	6 (46.2)	3 (24.9)
3	3 (23.1)	1 (7.7)	0
4	3 (23.1)	2 (15.4)	2 (16.6)
Diarrhoea			
0	4 (30.8)	6 (46.2)	6 (49.8)
1	3 (23.1)	5 (38.5)	4 (33.2)
2	4 (30.8)	2 (15.4)	2 (16.6)
3	2 (15.4)	0	0
4	0	0	0
Constipation			
0	6 (46.2)	8 (61.5)	5 (41.5)
1	2 (15.4)	2 (15.4)	3 (24.9)
2	1 (7.7)	2 (15.4)	4 (33.2)
3	2 (15.4)	1 (7.7)	0
4	2 (15.4)	0	0
Nausea			
0	3 (23.1)	4 (30.8)	5 (41.5)
1	4 (46.2)	6 (46.2)	5 (41.5)
2	1 (7.7)	2 (15.4)	1 (8.3)
3	3 (23.1)	0	0
4	0	1 (7.7)	1 (8.3)

Vomiting			
0	8 (61.6)	10 (76.9)	8 (66.4)
1	5 (38.5)	2 (15.4)	3 (24.9)
2	0	0	1 (8.3)
3	0	0	0
4	0	1 (7.7)	0
Abdominal			
discomfort	2 (15.4)	1 (7.7)	5 (41.5)
0	2 (15.4)	5 (38.5)	2 (16.6)
1	3 (23.1)	2 (15.4)	2 (16.6)
2	2 (15.4)	3 (23.1)	1 (8.3)
3	4 (30.8)	2 (15.4)	2 (16.6)
4			
Passing wind			
0	0	2 (15.4)	5 (41.5)
1	1 (7.7)	2 (15.4)	0
2	6 (46.2)	3 (23.1)	3 (24.9)
3	1 (7.7)	3 (23.1)	4 (33.2)
4	5 (38.5)	3 (23.1)	0
Not feeling hungry			
0	5 (38.5)	5 (38.5)	5 (41.5)
1	1 (7.7)	0	1 (8.3)
2	3 (23.1_	5 (38.5)	5 (41.5)
3	2 (15.4)	2 (15.4)	0
4	2 (15.4)	1 (7.7)	1 (8.3)
Bloating			
0	2 (15.4)	6 (46.2)	6 (49.8)
1	5 (38.5)	2 (15.4)	1 (8.3)
2	3 (23.1)	3 (23.1)	4 (33.2)
3	2 (15.4)	1 (7.7)	1 (8.3)
4	1 (7.7)	1 (7.7)	0
PedsQL-GSS score*	58.3 (6.2)	67.1 (5.3)	73.6 (6.4)[a]

Values expressed as Number of subjects (percentage out of the total), but PedsQL-GSS score expressed as mean (SEM): a low total score is a good outcome indicating fewer GI symptoms; *: GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; *p-overall*= 0.011; [a]: *p*= 0.009 compared to diagnosis

4.3.5 Faecal sample characteristics

Stool sample characteristics, such as total sample weight (i.e. faecal output), pH, water content and ammonia concentration, did not change significantly across the observational period (Table 4.22). Only ammonia tended to be different throughout the one-year follow-up (*p–overall*= 0.067 < 0.10).

Table 4.22: Stool samples' characteristics of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
Total sample weight (gr)	58.0 (14.5)	34.4 (9.2)	54.2 (10.8)	0.278
Faecal pH	6.4 (0.46)	7.2 (0.18) [1]	6.8 (0.16)	0.113
Faecal water content (%)	67.6 (0.46)	65.8 (1.1)	69.1 (2.0)	0.506
Ammonia ($*10^{-4}$ mg/g wet matter)	11.2 (0.96)	8.0 (1.2)	11.1 (1.6) (2)	0.067

Values expressed as mean (SEM); the number of missing data is shown in brackets; GLM accounted for paired data; Box-Cox transformation with optimal λ

4.3.6 Sulphide

Free faecal sulphide measured in dry matter changed significantly during follow up (p= 0.050, Table 4.23). Compared with the baseline mean (SEM) value (0.10 (0.03)), there was a threefold increase in free sulphide concentration in dry matter at 12 months on GFD (0.32 (0.062), p= 0.046). Faecal free sulphide concentration in wet matter tended to differ during follow up, but that difference did not reach significance (p= 0.074). Free sulphide per total output had no significant difference during follow up (p= 0.150). Similarly, there was no difference in faecal concentration of faecal bound and total sulphide at six and 12 months GFD regardless of the way of measurement (p> 0.05, Table 4.23).

Table 4.23: Faecal sulphide concentration (µmol/g) of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (12)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
Free sulphide (µmol/g)				
dry matter	0.10 (0.03)[a]	0.15 (0.04)	0.32 (0.06)	0.050
wet matter	0.03 (0.01)	0.05 (0.01)	0.09 (0.02)	0.074
Total output	1.73 (0.42)	1.6 (0.54)	5.8 (2.2)	0.150
Bound sulphide (µmol/g)				
dry matter	2.4 (0.48)	1.9 (0.35)	1.5 (0.32)	0.258
wet matter	0.83 (0.17)	0.67 (0.12)	0.48 (0.11)	0.171
Total output	57.6 (17.6)	23.0 (7.8)	25.9 (4.9)	0.187
Total sulphide (µmol/g)				
dry matter	2.5 (0.48)	2.1 (0.37)	1.8 (0.3)	0.346
wet matter	0.87 (0.17)	0.72 (0.13)	0.57 (0.11)	0.211
Total output	59.3 (17.8)	24.7 (8.3)	31.7 (5.9)	0.206

Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p= 0.046 compared to GFD -12 months

4.3.7 Lactate

Faecal lactate did not change significantly during follow up (p> 0.05, Table 4.24). There was no significant change in the concentration of D-lactate throughout the treatment period. L-lactate when measured per dry and wet matter tended to differ, but that difference did not reach significance (p-*overall*= 0.085 and 0.087 respectively). Likewise, L-lactate per total output had no difference throughout the treatment on GFD (p-*overall*= 0.217, Table 4.24).

Table 4.24: Faecal D-, L-lactate concentration (g/ 100g matter) of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (12)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
D-lactate (g / 100g matter)				
dry matter	0.038 (0.01)	0.035 (0.005)	0.042 (0.006)	0.233
wet matter	0.013 (0.003)	0.012 (0.002)	0.012 (0.002)	0.812
Total output	0.007 (0.002)	0.004 (0.001)	0.006 (0.001)	0.304
L-lactate (g / 100g matter)				
dry matter	0.018 (0.006)	0.021 (0.004)	0.030 (0.007)	0.085
wet matter	0.006 (0.002)	0.007 (0.001)	0.006 (0.001)	0.087
Total output	0.003 (0.001)	0.003 (0.001)	0.004 (0.001)	0.217

Values expressed as mean (SEM); GLM accounted for pared data; Box – Cox transformation with optimal λ; pairwise comparison, Bonferroni method

4.3.8 SCFA

The absolute concentrations of SCFA did not change significantly during the follow up, except for the concentration (mean, SEM) of butyric acid (μmol/g) in dry matter that was significantly decreased at six months on GFD (58.0 (15.6)) compared to the diagnosis value (87.3 (17.1), p= 0.028) (Table 4.25, Figure 4.19c). In contrast, concentration of BCFA isobutyric and isovaleric in dry matter, as well as the concentration of isocaproic acid per total faecal output changed significantly during follow up (*p-overall*= 0.033, 0.029 and 0.030, respectively, Table 4.26). The mean (SEM) concentration (μmol/g) of isobutyric acid in dry matter tended to be lower at six (9.3 (1.1)) and 12 months (9.3 (1.2)) on GFD compared to the diagnosis (12.6 (1.8), p= 0.092 and 0.053 < 0.10, respectively). The mean (SEM) concentration (μmol/g) of isovaleric acid in dry matter was significantly lower at 12 months on GFD (8.9 (1.1)) than at diagnosis (12.7 (1.8), p= 0.040), whereas isocaproic (mean, SEM) expressed per total faecal output was significantly higher at 12 months on GFD (11.9 (7.0)) than at six months on GFD (2.7 (0.78), p= 0.041, Table 4.26).

Table 4.25: Faecal SCFA concentration (μmol/g) of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
Acetic acid (μmol/g)				
dry matter	387.7 (47.4)	369.5 (40.0)	424.9 (67.7)	0.898
wet matter	121.9 (13.8)	124.0 (12.6)	116.6 (11.1)	0.807
Total output	7570 (1846)	4313 (1126)	6940 (1866)	0.273
Propionic acid (μmol/g)				

dry matter	85.3 (11.9)	65.7 (9.4)	87.6 (16.8)	0.331
wet matter	26.7 (3.47)	22.0 (2.9)	23.63(3.0)	0.438
Total output	1864 (555)	792 (231)	1353 (365)	0.171
Butyric acid (µmol/g)				
dry matter	87.3 (17.1) [a]	58.0 (15.6)	75.7 (14.4)	0.026
wet matter	26.7 (4.4)	19.8 (5.4)	20.5 (2.5)	0.053
Total output	1769 (512)	750 (250)	1196 (317)	0.076
Valeric acid (µmol/g)				
dry matter	10.2 (1.6)	6.9 (1.0)	7.9 (1.2)	0.053
wet matter	3.2 (0.51)	2.3 (0.37)	2.4 (0.33)	0.196
Total output	226 (66.9)	79.9 (19.3)	133.4 (34.5)	0.211
Caproic acid (µmol/g)				
dry matter	1.6 (0.47)	1.5 (0.39)	2.3 (0.63)	0.320
wet matter	0.46 (0.14)	0.49 (0.13)	0.71 (0.19)	0.820
Total output	28.0 (11.4)	16.8 (5.3)	41.5 (17.6)	0.327
Heptanoic acid (µmol/g)				
dry matter	0.21 (0.09)	0.09 (0.03)	0.16 (0.08)	0.580
wet matter	0.06 (0.03)	0.03 (0.01)	0.05 (0.02)	0.608
Total output	3.9 (2.0)	0.62 (0.22)	2.5 (2.0)	0.498
Octanoic acid (µmol/g)				
dry matter	0.37 (0.14)	0.23 (0.15)	0.16 (0.08)	0.306
wet matter	0.11 (0.04)	0.08 (0.05)	0.82 (0.54)	0.298
Total output	702 (3.5)	4.6 (3.9)	63.7 (58.6)	0.320
Total SCFA (µmol/g)				
dry matter	598.3 (76.6)	520.7 (65.7)	619.6 (95.8)	0.535
wet matter	187.3 (21.8)	175.1 (21.3)	170.2 (14.8)	0.628
Total output	11993 (3061)	6153 (1645)	10031 (2615)	0.237

Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: p= 0.028 compared to GFD-6 months

Table 4.26: Faecal BCFA concentration (µmol/g) of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
Isobutyric acid (µmol/g)				
dry matter	12.6 (1.8)	9.3 (1.1)	9.3 (1.2)	0.033
wet matter	4.1 (0.64)	3.1 (0.37)	2.7 (0.35)	0.063
Total output	259.1 (71.2)	97.2 (21.1)	150.6 (39.0)	0.291

Isovaleric acid (μmol/g)				
dry matter	12.7 (1.8) [a]	9.2 (1.1)	8.9 (1.1)	0.029
wet matter	4.1 (0.66)	3.1 (0.37)	2.6 (0.33)	0.052
Total output	258.7 (70.7)	97.1 (21.3)	138.6 (32.7)	0.265
Isocaproic acid (μmol/g)				
dry matter	0.34 (0.05) [a]	0.29 (0.07)	0.56 (0.19)	0.060
wet matter	0.11 (0.02)	0.10 (0.03)	0.16 (0.05)	0.149
Total output	6.5 (1.6)	2.7 (0.78)	11.9 (7.0)	0.030

Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*= 0.040 compared to GFD-12 months

With regard to the relative abundance of SCFA, acetic, butyric, isobutyric and isovaleric acid changed significantly during the follow up (*p-overall*= 0.004, 0.015, 0.009 and 0.016 respectively) (Table 4.27). Compared with the baseline value (mean, SEM) [(%), acetic acid: 65.5 (1.3); butyric acid: 13.6 (1.04)], the relative abundance of acetic acid significantly increased at six months on GFD (72.5 (1.4), *p*= 0.003, Figure 4.19a), whereas the relative abundance of butyric acid was significantly decreased at six months on GFD (9.7 (1.4), *p*= 0.013, Table 4.27, Figure 4.19b). Accordingly, the mean (SEM) relative abundance of isobutyric and isovaleric acids were significantly decreased at 12 months on GFD (1.6 (0.16) and 1.6 (0.18)) compared to the diagnosis (2.2 (0.20), *p*= 0.008 and 2.3 (0.23), *p*= 0.014 respectively, Table 4.27, Figure 4.19d).

Table 4.27: Faecal SCFA and BCFA relative abundance (%) of CD children at diagnosis and at six and 12 months on GFD

	Diagnosis (13)	GFD – 6 months (13)	GFD – 12 months (13)	*p-overall*
Acetic acid (%)	65.5 (1.3)	72.5 (1.4) [a]	68.4 (2.0)	0.004
Propionic acid (%)	14.3 (0.93)	12.4 (0.52)	13.8 (1.53)	0.493
Butyric acid (%)	13.6 (1.04)	9.7 (1.4) [b]	11.9 (1.03)	0.015
Isobutyric acid (%)	2.2 (0.20)	1.9 (0.13)	1.6 (0.16) [c]	0.009
Valeric acid (%)	1.7 (0.20)	1.3 (0.16)	1.5 (0.21)	0.203
Isovaleric acid (%)	2.3 (0.23)	1.9 (0.16)	1.6 (0.18) [d]	0.016
Caproic acid (%)	0.3 (0.08)	0.3 (0.10)	0.5 (0.13)	0.524
Isocaproic acid (%)	0.06 (0.01)	0.05 (0.01)	0.09 (0.02)	0.111
Heptanoic acid (%)	0.04 (0.02)	0.02 (0.01)	0.03 (0.01)	0.572
Octanoic acid (%)	0.07 (0.03)	0.04 (0.03)	0.6 (0.4)	0.308

Values expressed as mean (SEM); GLM accounted for paired data; Box-Cox transformation with optimal λ; pairwise comparison, Bonferroni method; [a]: *p*= 0.003 compared to diagnosis, [b]: *p*= 0.013 compared to diagnosis; [c]: *p*= 0.008 compared to diagnosis; [d]: *p*= 0.014 compared to diagnosis

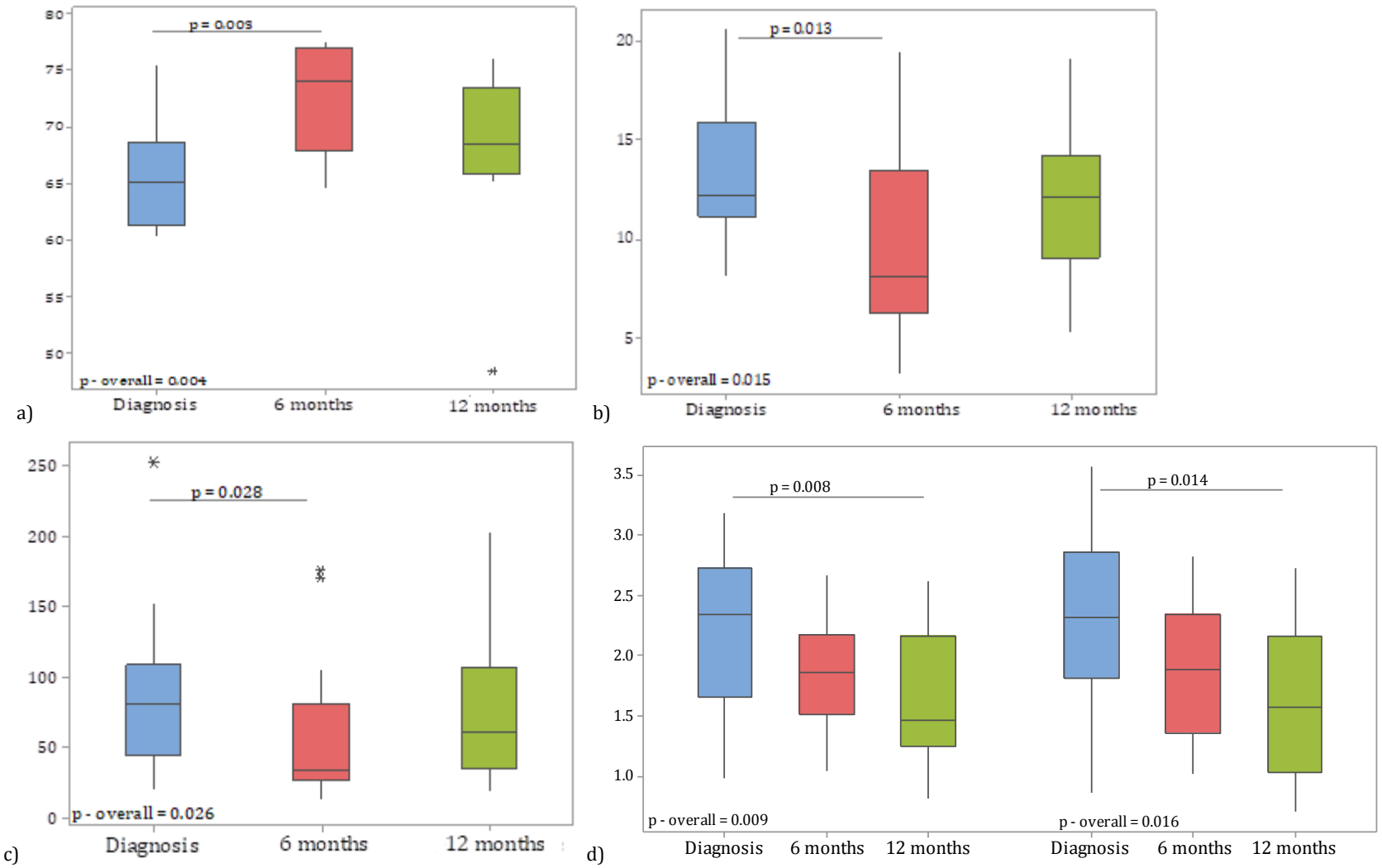


Figure 4.19: SCFA and BCFA in 13 CD children at diagnosis and at six and 12 months on GFD; (a) relative abundance (%) of acetic acid, (b) relative abundance (%) of butyric acid, (c) concentration of butyric acid in dry matter (µmol/g), (d) relative abundance (%) of isobutyric (left) and isovaleric (right) acid

4.3.9 Faecal microbiota

4.3.9.1 Faecal microbiota community structure

Faecal microbial community structure of 13 CD patients did not change during follow-up. Differences in microbiota diversity richness and evenness (Shannon H and Simpson index) did not reach significance, neither at OTU (Figure 4.20a) nor at genus (Figure 4.21a) level (p> 0.05, Table 4.28). There was no significant difference in the community structure (β-diversity) of the samples at diagnosis compared with six and 12 months on GFD (*p*> 0.05). That was evident from (a) the NMDS plots at OTU level for both the Bray-Curtis dissimilarity index and unweighted Unifrac distance analysis (Figure 4.22), (b) the NMDS plot for the Bray-Curtis dissimilarity index at genus level (Figure 4.23) and (c) the LCBD analysis at both OTU and genus level (*p*> 0.05, Figure 4.24). Similarly, there was no significant difference in the variation of the community structure at OTU and genus level among samples of the same group (β-dispersion).

These results were confirmed when DADA2 algorithm was used to infer the sample sequences in ASV. Alpha diversity did not change during the follow up, neither at ASV (Figure 4.20b) nor at genus (Figure 4.21b) level (*p*> 0.05, Table 4.28). Similarly, the microbial community structure (β-diversity) was not different in the NMDS plots for the Bray – Curtis dissimilarity index and unweighted Unifrac distance analysis, between the three time points, neither at ASV nor at genus level (*p*> 0.05).

Table 4.28: Alpha diversity indices in faecal samples of CD children at diagnosis, six and 12 months on GFD

Microbial diversity	*p-value* (USEARCH)	*p-value* (DADA2)
	OTU level	**ASV level**
Richness	0.99	0.73
Shannon H	0.73	0.39
Simpson index	0.45	0.27
	Genus level	
Richness	0.67	0.6
Shannon H	0.19	0.65
Simpson index	0.12	0.57

OTU: operational taxonomic unit; ASV: amplicon sequencing variants

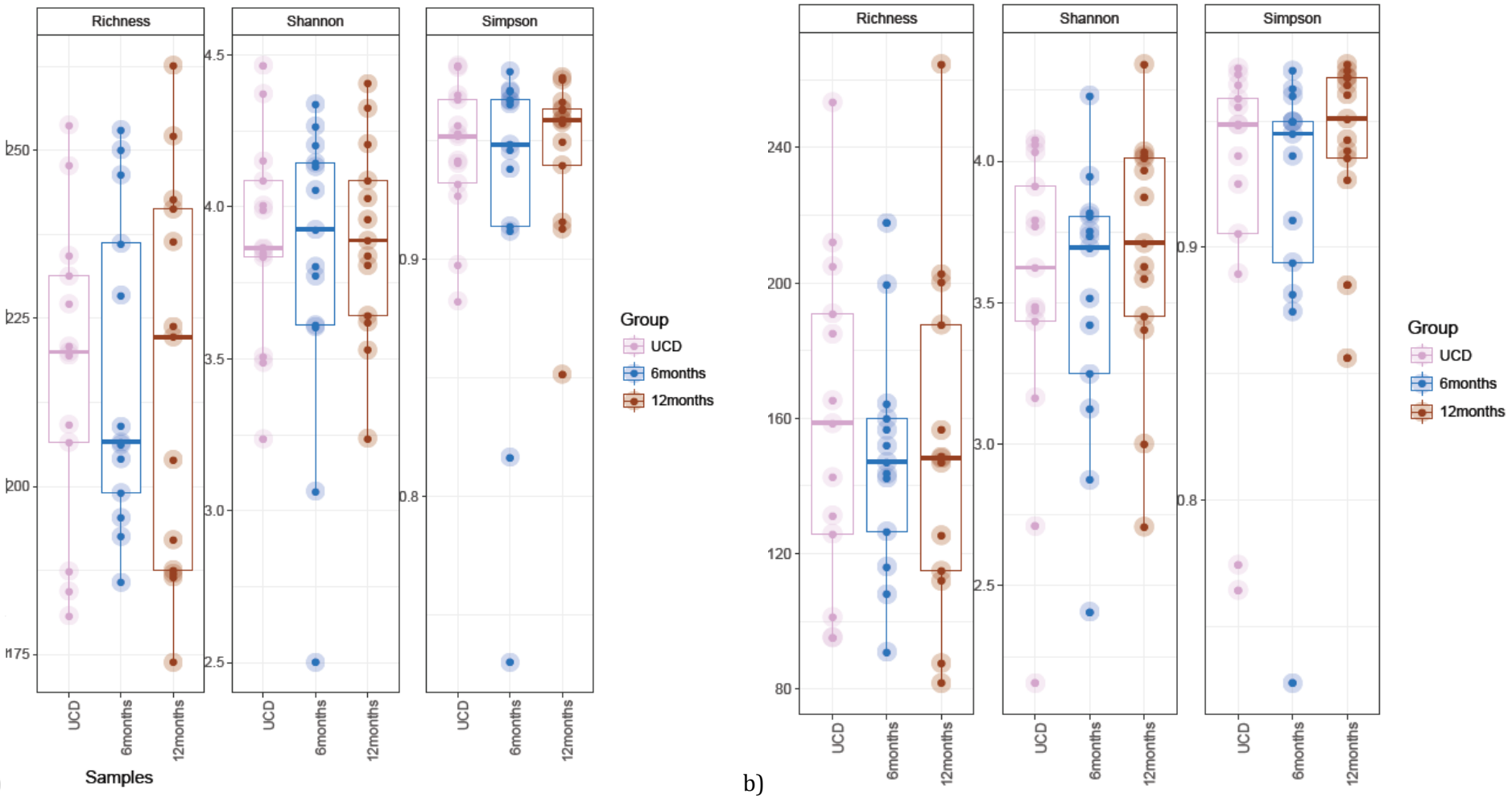


Figure 4.20: Alpha diversity measures at OTU and ASV level using (a) USEARCH and (b) DADA2 algorithms respectively, for CD patients at diagnosis, six and 12 months on GFD

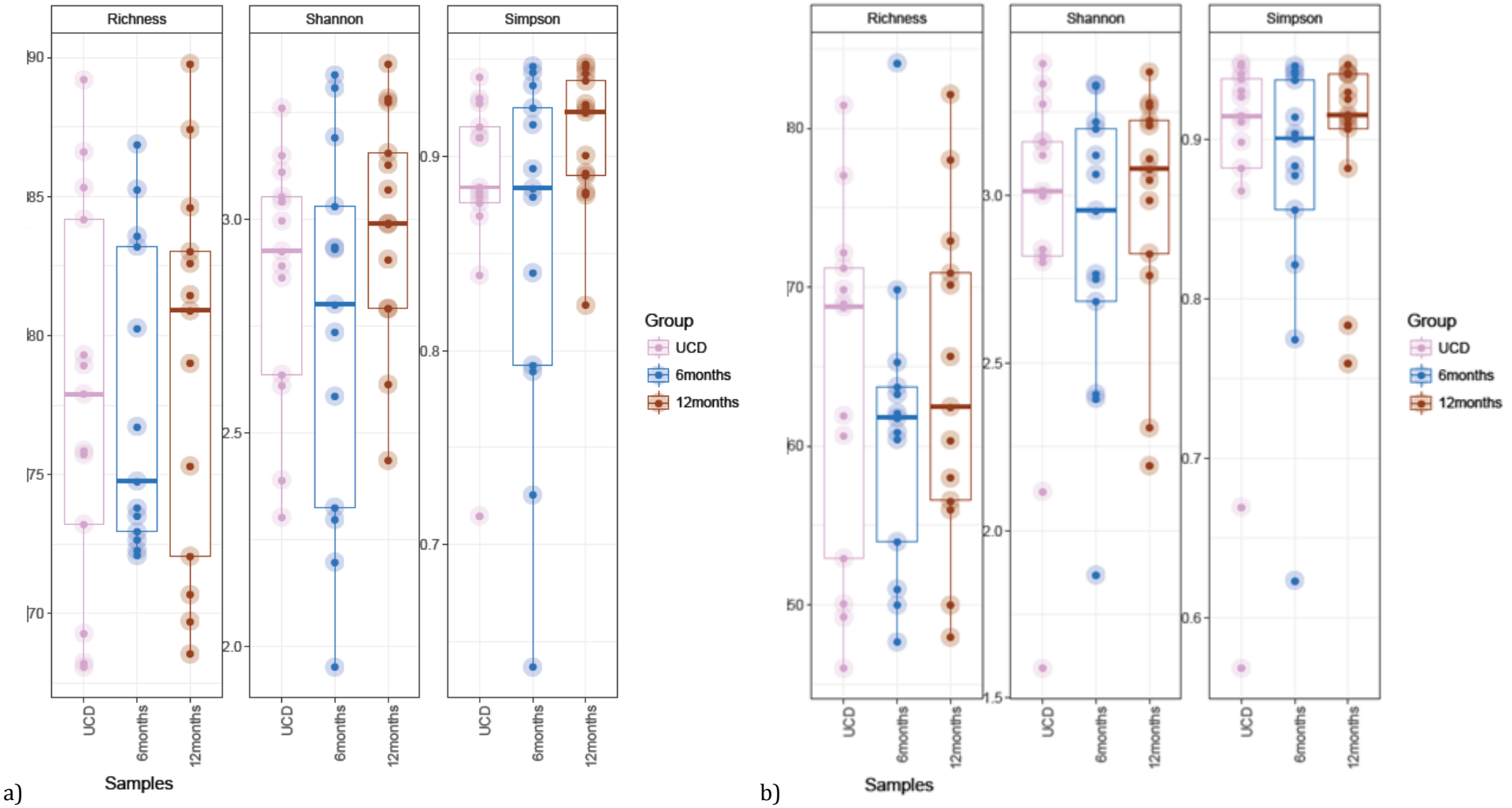


Figure 4.21: Alpha diversity measures at genus level using (a) USEARCH and (b) DADA2 algorithms for CD patients at diagnosis, six and 12 months on GFD

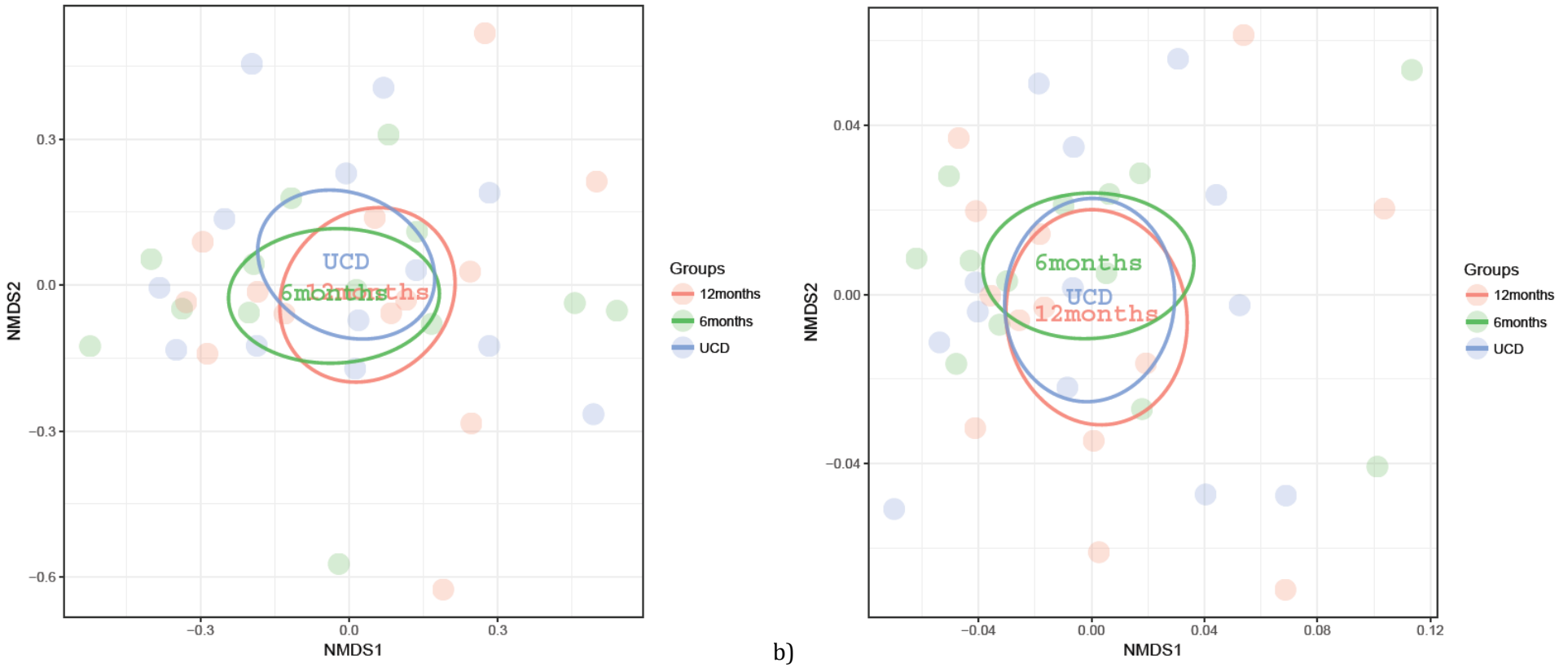


Figure 4.22: Non-metric multidimensional scaling (NMDS) plot at OTU level, using (a) Bray-Curtis dissimilarity index and (b) unweighted Unifrac for CD patients at diagnosis, six and 12 on GFD (OTU clustering using USEARCH)

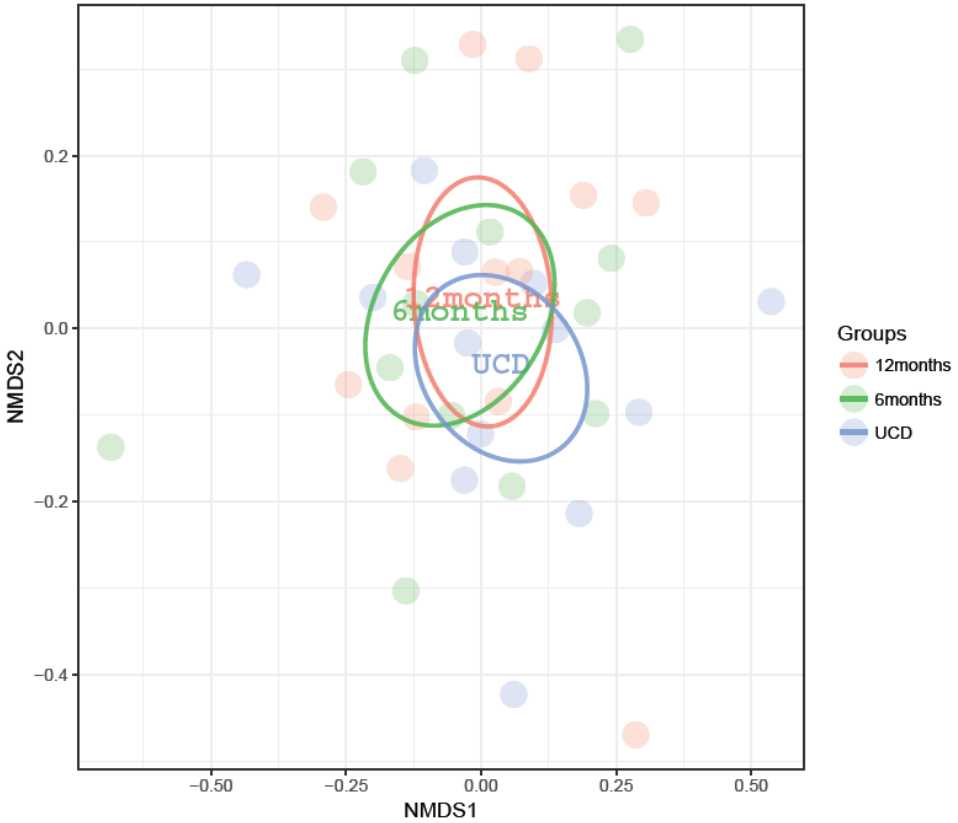
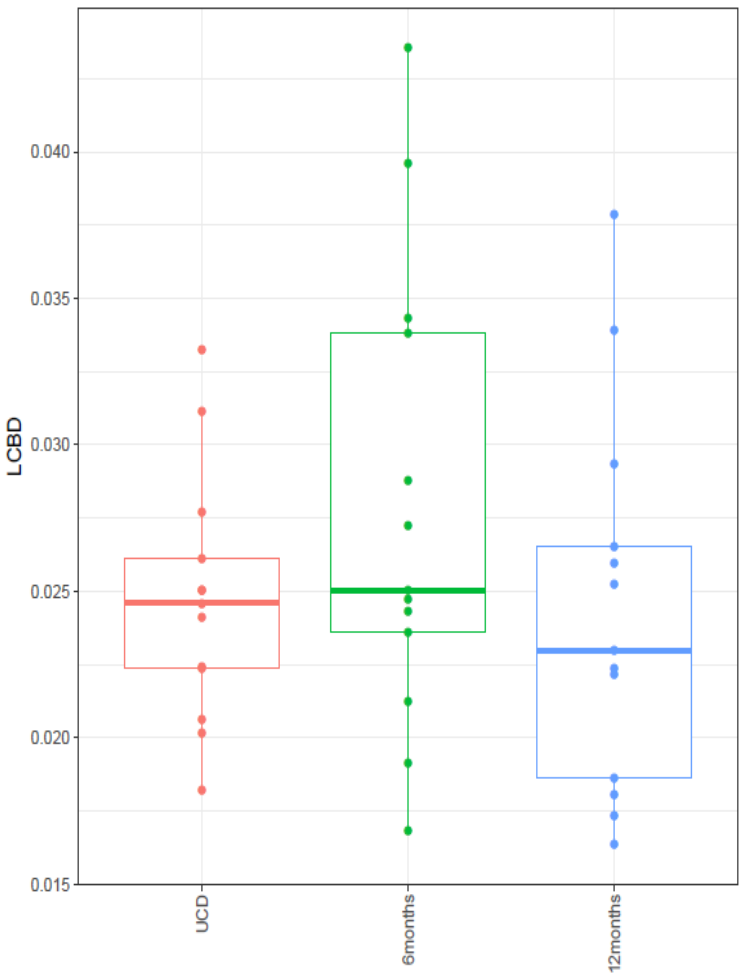
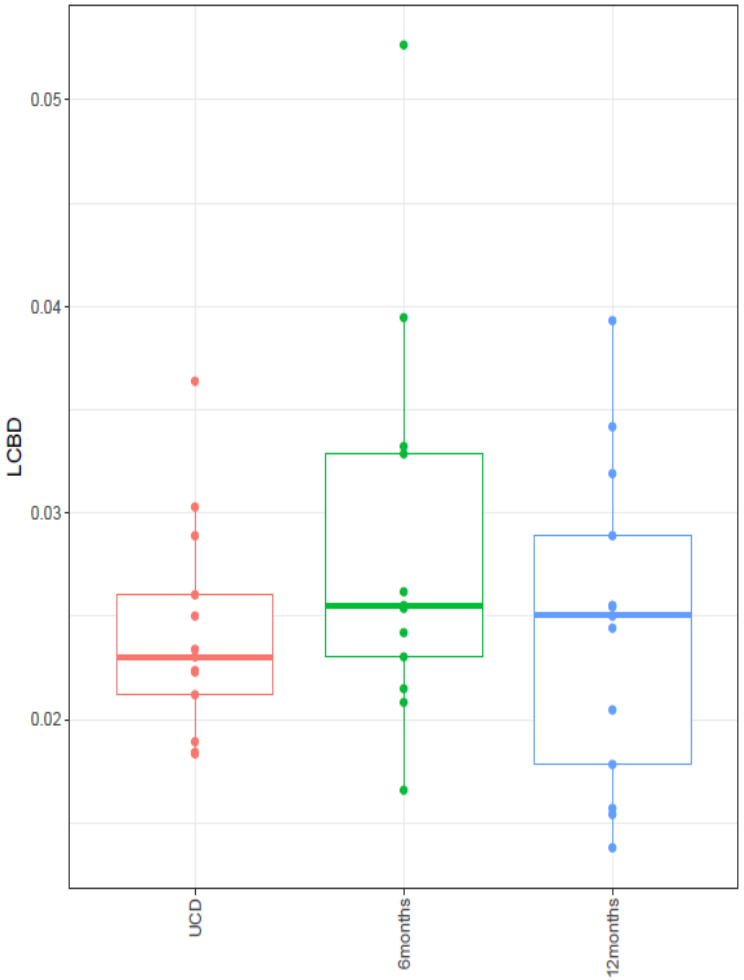


Figure 4.23: Non-metric multidimensional scaling (NMDS) plot at genus level, using Bray-Curtis dissimilarity index for CD patients at diagnosis, six and 12 months on GFD (OTU clustering using USEARCH)



a)



b)

Figure 4.24: Local contribution of β-diversity (LCBD) analysis at (a) OTU and (b) genus level for CD patients at diagnosis, six and 12 months on GFD (% of total community dispersion)

4.3.9.2 Bacterial taxon relative abundance

The faecal samples of prospective cohort of 13 CD children at diagnosis, at six and 12 months on GFD were characterised by 835 distinct OTU.

Six months on GFD

Forty out of 835 OTU (4.8% of all) had significantly different relative abundance six months after the initiation of GFD compared to diagnosis (Table 4.29, Figure 4.25). Twenty-two of these 40 OTU had significantly lower relative abundance at six months on GFD, while the remaining 18 OTU had significantly higher relative abundance at six months on GFD than at diagnosis (*p-value (adjusted)*< 0.05, Table 4.29). Collectively, OTU belonging to *Phascolarctobacterium, Cronobacter, Morganella, Erysipelatoclostridium, Tyzzerella 4, Clostridium 9, Flavonifractor, Eggerthella, Intestinimonas* and [Eubacterium] *hallii* group were significantly less abundant at six months on GFD than at diagnosis, while OTU belonging to *Dialister, Ruminococcaceae UCG-005 / UCG-002, Ruminococcaceae NK4A214* group*, Ruminococcus 2, Subdoligranulum, Veillonella, Clostridium sensu stricto 1, Dorea, Parasutterella, Coprococcus 2, Terrisporobacter* were significantly more abundant at six months on GFD than at diagnosis (*p*< 0.05, Table 4.30, Figure 4.26). Compared to the baseline, a respective decrease in relative abundance of OTU belonging to *Acidaminococcaceae*, along with an increase of OTU belonging to *Veillonellaceae* and *Clostridiaceae 1* was observed at six months on GFD (*p-value (adjusted)* = 1.74E-09, 2.46E-04 and 1.22E-02 respectively).

The relative abundance of the 40 discriminatory OTU was tested against faecal GIP levels, serum tTG concentration, dietary intake of PUFA (%TDEI), NMES (%TDEI), NSP (g, %TDEI), starch (%TDEI) and dietary fibre (g), concentration in dry matter of butyric and isobutyric acids, as well as relative the abundance of acetic and butyric acid. Kendall rank correlation revealed no significant difference between the relative abundance of the 40 OTU and the faecal, subjects' characteristics that differentiated CD children at diagnosis and at six months on GFD.

Twelve months on GFD

Twelve out of 835 OTU (1.4% of all) had significantly different relative abundance at 12 months on GFD compared to the diagnosis. All but four OTU had significantly higher relative abundance at 12 months on GFD than at diagnosis. OTU_576 *Akkermansia,* OTU_5 *Phascolarctobacterium,* OTU_18 *Phascolarctobacterium* and OTU_50 *Tyzzerella* 4 had significantly lower relative abundance at 12 on GFD than at diagnosis (Table 4.29, Figure 4.27a). Likewise, OTU belonging to *Phascolarctobacterium* and *Tyzzerella 4* had significantly lower relative abundance at 12 months on GFD, but OTU belonging to *Dialister, Ruminococcus 2, Veillonella, Clostridium sensu strict 1* and *Terrisporobacter* had significantly higher relative abundance at 12 months on GFD compared to the diagnosis (Table 4.30, Figure 4.27b). As a whole, compared to baseline, the relative abundance of OTU belonging widely to *Acidaminococcaceae* decreased at 12 months on GFD (*p-value (adjusted)*= 7.62E-06), while the relative

abundance of OTU belonging to *Veillonellaceae* and *Clostridiaceae 1* increased (*p-value (adjusted*= 3.71E -03 and 3.92E-02 respectively).

Relative abundance of 12 OTU that differentiated CD children at diagnosis and at 12 months on GFD was associated with the following variables, which were significantly different between the two groups; faecal GIP levels, serum tTG concentration, PedsQL-GSS score, dietary intake of NMES (%TDEI), NSP (g, %TDEI) and starch (%TDEI), concentration in dry matter of isobutyric, isovaleric and isocaproic acids, as well as relative abundance of isobutyric and isovaleric acid. Correlation between relative abundance of 12 OTU and these variables was examined using Kendall correlation. Similar to the analysis between CD children at diagnosis and at six months on GFD, no significant associations were found.

Table 4.29: OTU with significantly different relative abundance in faecal samples of (a) paired data from 13 CD children at diagnosis and at six and 12 months after the initiation of GFD (b) independent data from UCD patients (n=20) and TCD patients (n=45)

(OTU clustering using USEARCH)

	Group comparison	BaseMean	log2Fold Change	*p-value*	*p-value (adjusted)*
OTU_576 *Akkermansia*[1]	UCD Vs GFD 6 mos	212.2	-8.53	1.79E-05	1.27E-03
	UCD Vs GFD 12 mos	109.1	-7.00	3.20E-06	1.31E-03
	UCD Vs TCD	124.9	1.19	1.27E-01	2.38E-01
OTU_6 *Dialister*[2]	UCD Vs GFD 6 mos	1,428.3	6.82	1.99E-06	2.83E-04
	UCD Vs GFD 12 mos	1,283.0	6.88	1.48E-05	1.52E-03
	UCD Vs TCD	1,540.8	-0.21	7.67E-01	8.48E-01
OTU_5 *Phascolarctobacterium*[1]	UCD Vs GFD 6 mos	1,522.3	-9.80	5.98E-11	1.70E-08
	UCD Vs GFD 12 mos	1,256.2	-7.66	8.71E-06	1.52E-03
	UCD Vs TCD	2,385.3	0.83	3.26E-01	4.69E-01
OTU_537 *Dialister*[2]	UCD Vs GFD 6 mos	25.8	6.82	2.51E-05	1.42E-03
	UCD Vs GFD 12 mos	25.9	7.69	1.15E-05	1.52E-03
	UCD Vs TCD	4.68	0.23	3.11E-01	7.29E-01
OTU_8 *Ruminococcus 2*[2]	UCD Vs GFD 6 mos	2,208.9	4.79	6.18E-03	4.46E-02
	UCD Vs GFD 12 mos	3,458.3	6.24	1.71E-04	1.40E-02
	UCD Vs TCD	2,692.5	0.21	7.48E-01	8.33E-01
OTU_43 *Veillonella*[2]	UCD Vs GFD 6 mos	343.6	4.80	6.14E-04	1.43E-02
	UCD Vs GFD 12 mos	1,222.8	5.26	3.85E-04	2.26E-02
	UCD Vs TCD	367.3	-0.42	4.66E-01	6.00E-01
OTU_55 [*Eubacterium*] *coprostanoligenes group*[2]	UCD Vs GFD 6 mos	119.2	6.46	1.39E-03	2.28E-02
	UCD Vs GFD 12 mos	129.2	6.76	3.69E-04	2.26E-02
	UCD Vs TCD	86.6	-0.38	5.89E-01	7.09E-01

OTU_1045 *Roseburia*	UCD Vs GFD 6 mos	98.6	0.85	5.22E-01	8.06E-01
	UCD Vs GFD 12 mos	140.7	5.49	5.53E-04	2.52E-02
	UCD Vs TCD	276.2	-0.27	6.85E-01	7.87E-01
OTU_50 *Tyzzerella 4*[1]	UCD Vs GFD 6 mos	59.9	-6.20	1.22E-04	3.84E-03
	UCD Vs GFD 12 mos	62.6	-5.39	5.08E-04	2.52E-02
	UCD Vs TCD	55.6	1.74	5.06E-03	3.02E-02
OTU_191 *Christensenellaceae R-7 group*[2]	UCD Vs GFD 6 mos	16.9	6.09	1.01E-04	3.57E-03
	UCD Vs GFD 12 mos	17.4	6.00	6.65E-04	2.73E-02
	UCD Vs TCD	14.4	-0.74	2.07E-01	3.39E-01
OTU_22 *Ruminococcaceae UCG-002*[2]	UCD Vs GFD 6 mos	748.0	3.92	1.29E-03	2.28E-02
	UCD Vs GFD 12 mos	772.1	6.01	8.30E-04	2.84E-02
	UCD Vs TCD	785.6	-0.22	7.35E-01	8.24E-01
OTU_18 *Phascolarctobacterium*	UCD Vs GFD 6 mos	732.0	-0.48	7.88E-01	9.85E-01
	UCD Vs GFD 12 mos	109.8	-5.38	7.91E-04	2.84E-02
	UCD Vs TCD	9.4	0.96	1.36E-01	2.46E-01
OTU_166 *Cronobacter*	UCD Vs GFD 6 mos	25.3	-7.90	5.84E-06	5.53E-04
	UCD Vs GFD 12 mos	3.57	-0.02	9.91E-01	9.98E-01
	UCD Vs TCD	3.4	0.65	1.54E-01	2.73E-01
OTU_113 *Clostridium sensu stricto 1*	UCD Vs GFD 6 mos	268.6	5.73	7.91E-05	3.21E-03
	UCD Vs GFD 12 mos	297.8	4.92	2.03E-03	6.43E-02
	UCD Vs TCD	148.6	0.67	2.78E-01	4.25E-01
OTU_259 *Erysipelatoclostridium*[3]	UCD Vs GFD 6 mos	26.8	-6.95	7.37E-05	3.21E-03
	UCD Vs GFD 12 mos	10.5	-1.98	2.78E-01	9.98E-01
	UCD Vs TCD	8.6	2.61	4.31E-07	1.64E-05
OTU_24 *Lachnoclostridium*	UCD Vs GFD 6 mos	1,315.9	5.22	2.79E-04	7.91E-03
	UCD Vs GFD 12 mos	1,471.1	2.51	9.34E-02	6.32E-01
	UCD Vs TCD	939.0	-0.70	2.28E-01	2.83E-02
OTU_14 *Subdoligranulum*	UCD Vs GFD 6 mos	1,006.0	4.16	6.55E-04	1.43E-02
	UCD Vs GFD 12 mos	1,258.2	2.21	1.30E-01	7.30E-01
	UCD Vs TCD	1,446.3	1.24	2.59E-02	9.09E-02
OTU_83 *Alistipes*	UCD Vs GFD 6 mos	93.9	5.52	6.50E-04	1.43E-02
	UCD Vs GFD 12 mos	113.2	5.40	9.00E-03	1.83E-01
	UCD Vs TCD	132.2	0.65	3.06E-01	4.49E-01
OTU_66 *Ruminococcaceae UCG-005*	UCD Vs GFD 6 mos	558.9	4.49	7.20E-04	1.46E-02
	UCD Vs GFD 12 mos	505.6	2.38	1.64E-01	7.98E-01
	UCD Vs TCD	304.6	0.77	1.55E-01	2.73E-01
OTU_98 *Ruminococcaceae*[3]	UCD Vs GFD 6 mos	57.0	-3.66	8.64E-04	1.63E-02
	UCD Vs GFD 12 mos	53.1	-1.81	2.57E-01	9.98E-01
	UCD Vs TCD	139.2	2.06	1.39E-04	1.80E-03
OTU_3 *Akkermansia*	UCD Vs GFD 6 mos	1,033.2	5.40	1.45E-03	2.28E-02
	UCD Vs GFD 12 mos	1,582.8	2.19	2.35E-01	9.95E-01

	UCD Vs TCD	2,466.0	1.45	5.12E-02	1.43E-01
OTU_129 *Eggerthella*	UCD Vs GFD 6 mos	73.1	-3.76	2.99E-03	3.14E-02
	UCD Vs GFD 12 mos	72.6	-1.69	2.12E-01	9.49E-01
	UCD Vs TCD	70.16	0.53	2.89E-01	4.36E-01
OTU_893 [*Eubacterium*] *hallii*	UCD Vs GFD 6 mos	82.7	-2.95	2.60E-03	3.14E-02
Group	UCD Vs GFD 12 mos	85.3	-1.93	8.97E-02	6.32E-01
	UCD Vs TCD	50.2	0.48	3.50E-01	4.91E-01
OTU_303 *Lachnospiraceae*	UCD Vs GFD 6 mos	46.0	-3.69	2.90E-03	3.14E-02
UCG-008	UCD Vs GFD 12 mos	62.5	-0.42	7.67E-01	9.98E-01
	UCD Vs TCD	114.8	1.47	3.77E-03	2.47E-02
OTU_115 *Parasutterella*	UCD Vs GFD 6 mos	56.8	4.42	2.33E-03	3.14E-02
	UCD Vs GFD 12 mos	73.4	3.26	4.43E-02	4.54E-01
	UCD Vs TCD	36.5	1.62	5.14E-03	3.03E-02
OTU_1005 *Fusicatenibacteri*	UCD Vs GFD 6 mos	9.71	-4.70	2.31E-03	3.14E-02
	UCD Vs GFD 12 mos	8.7	-2.47	1.28E-01	7.30E-01
	UCD Vs TCD	4.25	0.11	8.07E-01	8.76E-01
OTU_133 *Flavonifractor*	UCD Vs GFD 6 mos	57.0	-3.64	4.45E-03	4.36E-02
	UCD Vs GFD 12 mos	59.1	-2.48	1.07E-01	6.58E-01
	UCD Vs TCD	99.5	1.02	5.73E-02	1.55E-01
OTU_1044 [*Eubacterium*]	UCD Vs GFD 6 mos	150.3	-3.14	5.18E-03	4.37E-02
hallii group	UCD Vs GFD 12 mos	149.9	-2.11	1.40E-01	7.58E-01
	UCD Vs TCD	152.3	0.04	9.38E-01	9.61E-01
OTU_185 *Dorea*	UCD Vs GFD 6 mos	19.4	4.18	5.41E-03	4.37E-02
	UCD Vs GFD 12 mos	20.0	3.10	7.03E-02	5.77E-01
	UCD Vs TCD	6.5	-0.05	9.29E-01	9.56E-01
OTU_90 *Ruminococcaceae*	UCD Vs GFD 6 mos	103.6	4.71	5.05E-03	4.37E-02
UCG-002	UCD Vs GFD 12 mos	92.9	3.83	4.53E-02	4.54E-01
	UCD Vs TCD	83.8	1.83	1.89E-03	1.34E-02
OTU_122 *Ruminiclostridium 9*	UCD Vs GFD 6 mos	25.3	-4.32	5.09E-03	4.37E-02
	UCD Vs GFD 12 mos	83.1	-4.10	1.70E-02	2.79E-01
	UCD Vs TCD	16.9	0.20	7.33E-01	8.23E-01
OTU_15 *Dorea*	UCD Vs GFD 6 mos	275.7	3.58	6.28E-03	4.46E-02
	UCD Vs GFD 12 mos	446.8	2.15	1.67E-01	7.98E-01
	UCD Vs TCD	538.8	0.82	1.62E-01	2.81E-01
OTU_68 *Ruminococcaceae*	UCD Vs GFD 6 mos	152.6	5.45	6.19E-03	4.46E-02
UCG-005	UCD Vs GFD 12 mos	77.5	3.89	4.99E-02	4.88E-01
	UCD Vs TCD	150.1	1.36	4.10E-02	1.24E-01

Table 4.30: Genera with significantly different relative abundance in faecal samples of (a) paired data from 13 CD children at diagnosis and at six and 12 months after the initiation of GFD (b) independent data from UCD patients (n=20) and TCD patients (n=45) (OTU clustering using USEARCH)

	Group comparison	BaseMean	log2Fold Change	*p-value*	*p-value (adjusted)*
Phascolactobacterium[1]	UCD Vs GFD 6 mos	1,676.12	-9.8	8.08E-12	1.06E-09
	UCD Vs GFD 12 mos	1,174.29	-7.68	7.02E-08	7.09E-06
	UCD Vs TCD	2,706.7	0.93	2.42E-01	4.81E-01
Dialister[2]	UCD Vs GFD 6 mos	1,090.6	6.93	8.86E	5.8E-06
	UCD Vs GFD 12 mos	925.78	7.4	5.24E-07	2.65E-05
	UCD Vs TCD	1,501.3	-0.23	7.32E-01	8.90E-01
Ruminococcaceae UCG -005	UCD Vs GFD 6 mos	534.9	5.14	3.25E-06	1.42E-04
	UCD Vs GFD 12 mos	483.0	2.91	4.57E-02	2.88E-01
	UCD Vs TCD	913.1	0.25	6.43E-01	8.29E-01
Cronobacter	UCD Vs GFD 6 mos	30.7	-7.99	4.69E-06	1.54E-04
	UCD Vs GFD 12 mos	3.8	-0.20	9.10E-01	9.45E-01
	UCD Vs TCD	3.4	0.59	1.99E-01	4.42E-01
Erysipelatoclostridium	UCD Vs GFD 6 mos	29.08	-6.59	9.43E-06	2.47E-04
	UCD Vs GFD 12 mos	11.3	-1.85	1.81E-01	5.79E-01
	UCD Vs TCD	16.6	0.12	8.16E-01	9.40E-01
Tyzzerella 4[1]	UCD Vs GFD 6 mos	58.71	-6.22	4.68E-05	1.02E-03
	UCD Vs GFD 12 mos	62.47	-5.57	3.01E-04	7.58E-03
	UCD Vs TCD	41.4	1.19	5.21E-02	1.92E-01
Subdoligranulum	UCD Vs GFD 6 mos	990.88	3.79	1.74E-04	3.26E-03
	UCD Vs GFD 12 mos	1,336.3	2.37	8.56E-02	4.12E-01
	UCD Vs TCD	2,306.2	0.24	6.38E-01	8.29E-01
Veillonella[2]	UCD Vs GFD 6 mos	283.98	4.45	3.08E-04	5.05E-03
	UCD Vs GFD 12 mos	933.6	4.82	4.73E-04	9.55E-03
	UCD Vs TCD	374.3	-0.53	3.39E-01	5.84E-01
Clostridium sensu stricto 1[2]	UCD Vs GFD 6 mos	256.18	5.05	4.78E-04	6.26E-03
	UCD Vs GFD 12 mos	275.2	4.44	1.15E-03	1.94E-02
	UCD Vs TCD	227.8	-0.09	8.68E-01	9.62E-01
Clostridium 9	UCD Vs GFD 6 mos	25.05	-4.54	4.53E-04	6.26E-03
	UCD Vs GFD 12 mos	n/a	n/a	n/a	n/a
	UCD Vs TCD	44.2	0.55	2.73E-01	5.13E-01
Flavonifractor	UCD Vs GFD 6 mos	44.60	-3.71	7.31E-04	7.97E-03
	UCD Vs GFD 12 mos	39.2	-2.37	5.93E-02	3.15E-01
	UCD Vs TCD	89.1	0.86	9.60E-02	3.29E-01
Dorea	UCD Vs GFD 6 mos	253.97	3.77	6.76E-04	7.97E-03
	UCD Vs GFD 12 mos	419.2	2.09	1.26E-01	4.77E-01

	UCD Vs TCD	639.9	0.70	2.02E-01	4.42E-01
Parasutterella	UCD Vs GFD 6 mos	35.47	4.42	9.21E-04	9.28E-03
	UCD Vs GFD 12 mos	48.8	3.05	4.54E-02	2.88E-01
	UCD Vs TCD	36.0	1.53	7.55E-03	5.10E-02
Morganella	UCD Vs GFD 6 mos	5.31	-5.05	1.21E-03	9.94E-03
	UCD Vs GFD 12 mos	n/a	n/a	n/a	n/a
	UCD Vs TCD	n/a	n/a	n/a	n/a
Ruminococcaceae	UCD Vs GFD 6 mos	974.68	3.22	1.14E-03	9.94E-03
UCG-002	UCD Vs GFD 12 mos	1,077.5	1.77	1.87E-01	5.79E-01
	UCD Vs TCD	1,823.9	0.22	6.20E-01	8.18E-01
Eggerthella	UCD Vs GFD 6 mos	66.40	-3.75	1.53E-03	1.18E-02
	UCD Vs GFD 12 mos	57.6	-1.83	1.42E-01	4.94E-01
	UCD Vs TCD	70.83	0.66	1.61E-01	4.25E-01
Ruminococcus 2[2]	UCD Vs GFD 6 mos	1,357.6	4.72	2.57E-03	1.87E-02
	UCD Vs GFD 12 mos	2,421.7	5.93	4.69E-05	1.58E-03
	UCD Vs TCD	2,372.6	0.06	9.20E-01	9.73E-01
Coprococcus 2	UCD Vs GFD 6 mos	202.57	4.95	4.36E-03	3.01E-02
	UCD Vs GFD 12 mos	342.3	2.66	1.19E-01	4.77E-01
	UCD Vs TCD	279.9	0.49	4.73E-01	7.33E-01
Intestinimonas	UCD Vs GFD 6 mos	8.83	-3.32	6.17E-03	4.04E-02
	UCD Vs GFD 12 mos	4.32	-1.14	3.51E-01	7.53E-01
	UCD Vs TCD	7.43	0.52	2.08E-01	4.45E-01
Terrisporobacter[2]	UCD Vs GFD 6 mos	42.64	4.01	6.97E-03	4.35E-02
	UCD Vs GFD 12 mos	28.06	4.77	3.42E-03	4.93E-02
	UCD Vs TCD	59.5	0.79	1.77E-01	4.31E-01
Ruminicoccaceae	UCD Vs GFD 6 mos	137.07	4.53	8.28E-03	4.85E-02
NK4A214 group	UCD Vs GFD 12 mos	98.6	3.63	4.51E-02	2.88E-01
	UCD Vs TCD	124.8	-0.82	2.04E-01	4.42E-01
[Eubacterium]	UCD Vs GFD 6 mos	790.4	-2.33	8.52E-03	4.85E-02
hallii group	UCD Vs GFD 12 mos	794.5	-1.38	1.89E-01	5.79E-01
	UCD Vs TCD	1,039.2	-0.34	4.68E-01	7.33E-01

n/a: not applicable

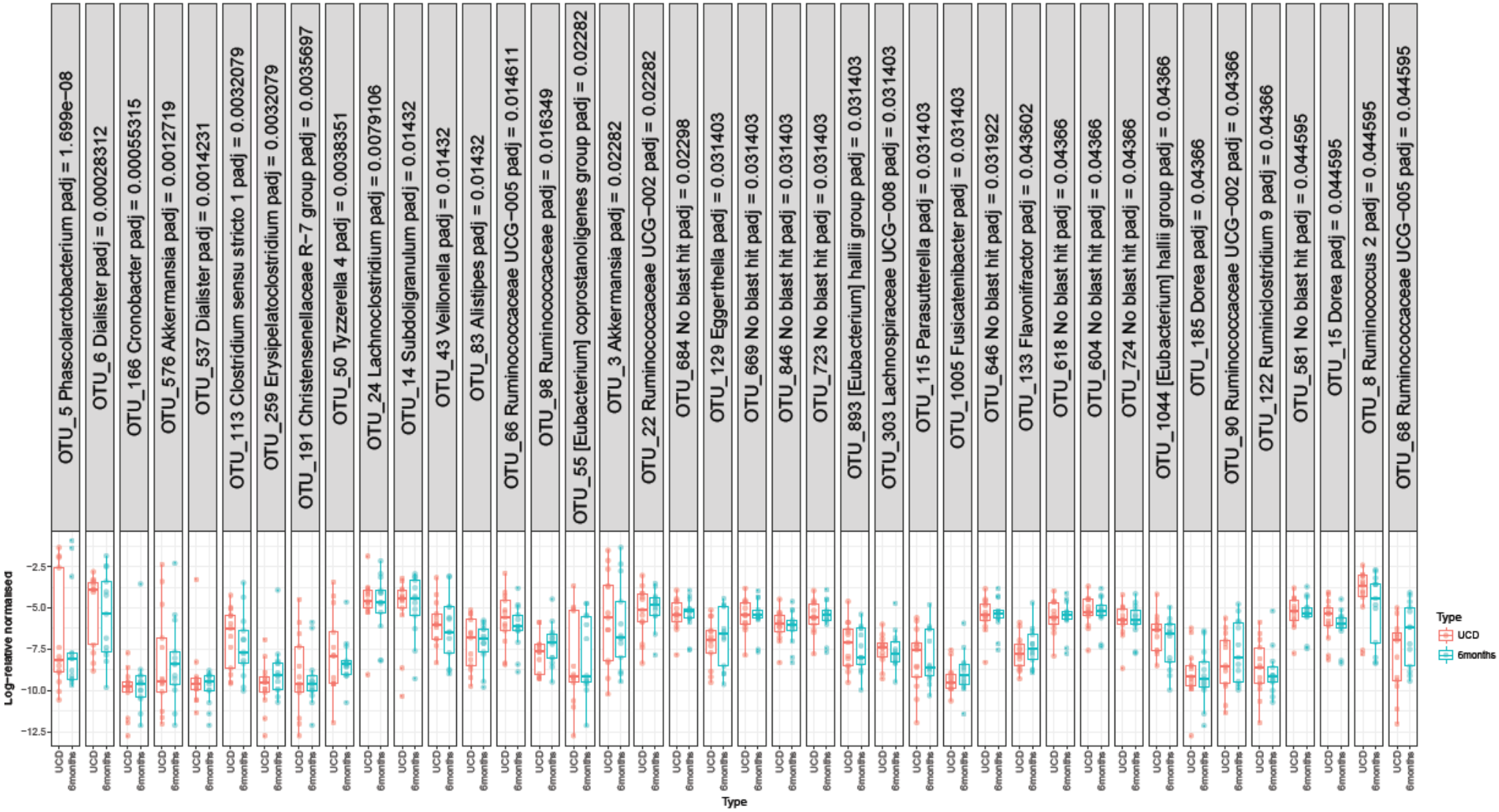


Figure 4.25: Log-relative abundance of OTU that changed significantly at six months on GFD compared to the diagnosis

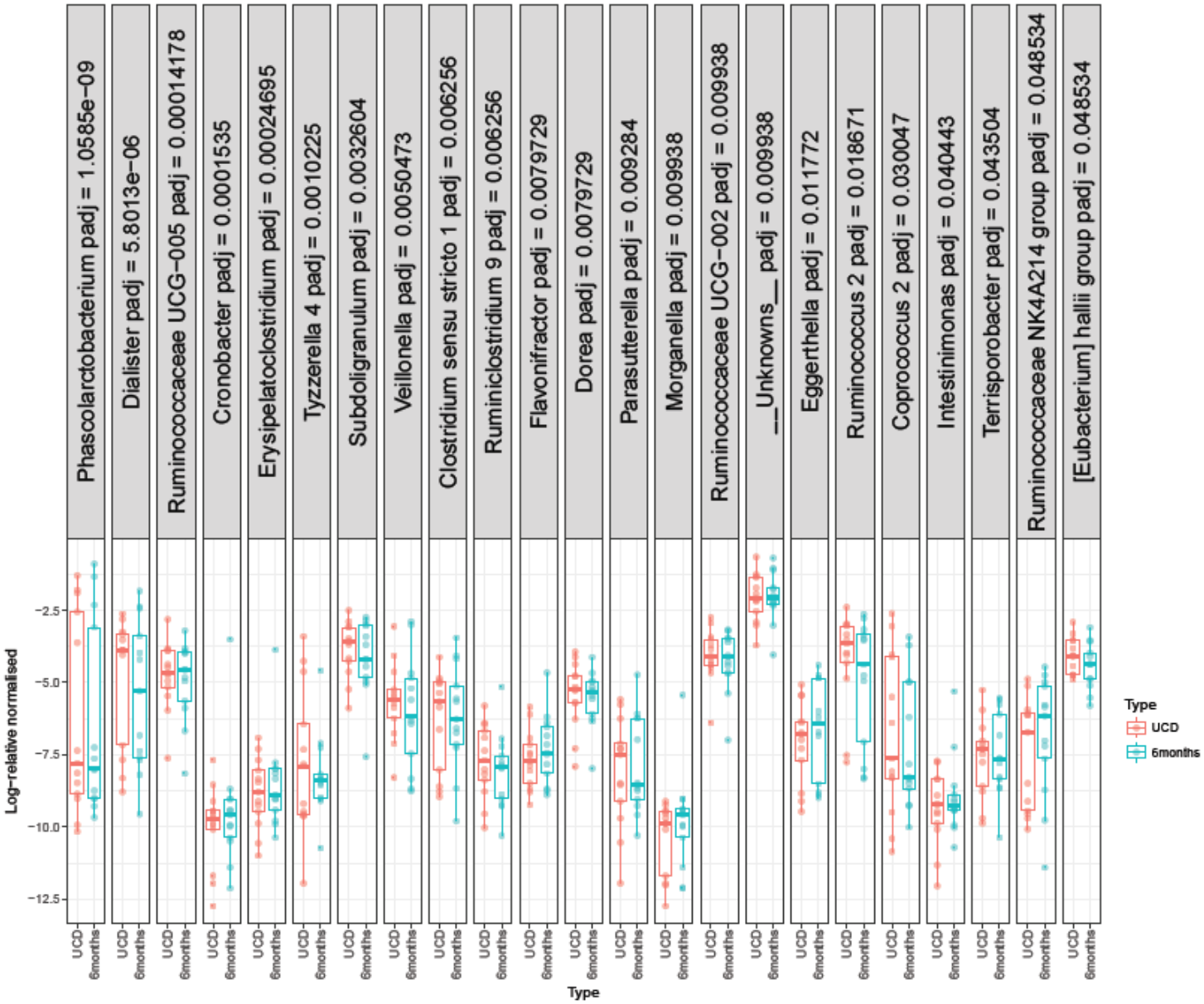


Figure 4.26: Log-relative abundance of genera that changed significantly at six months on GFD compared to the diagnosis

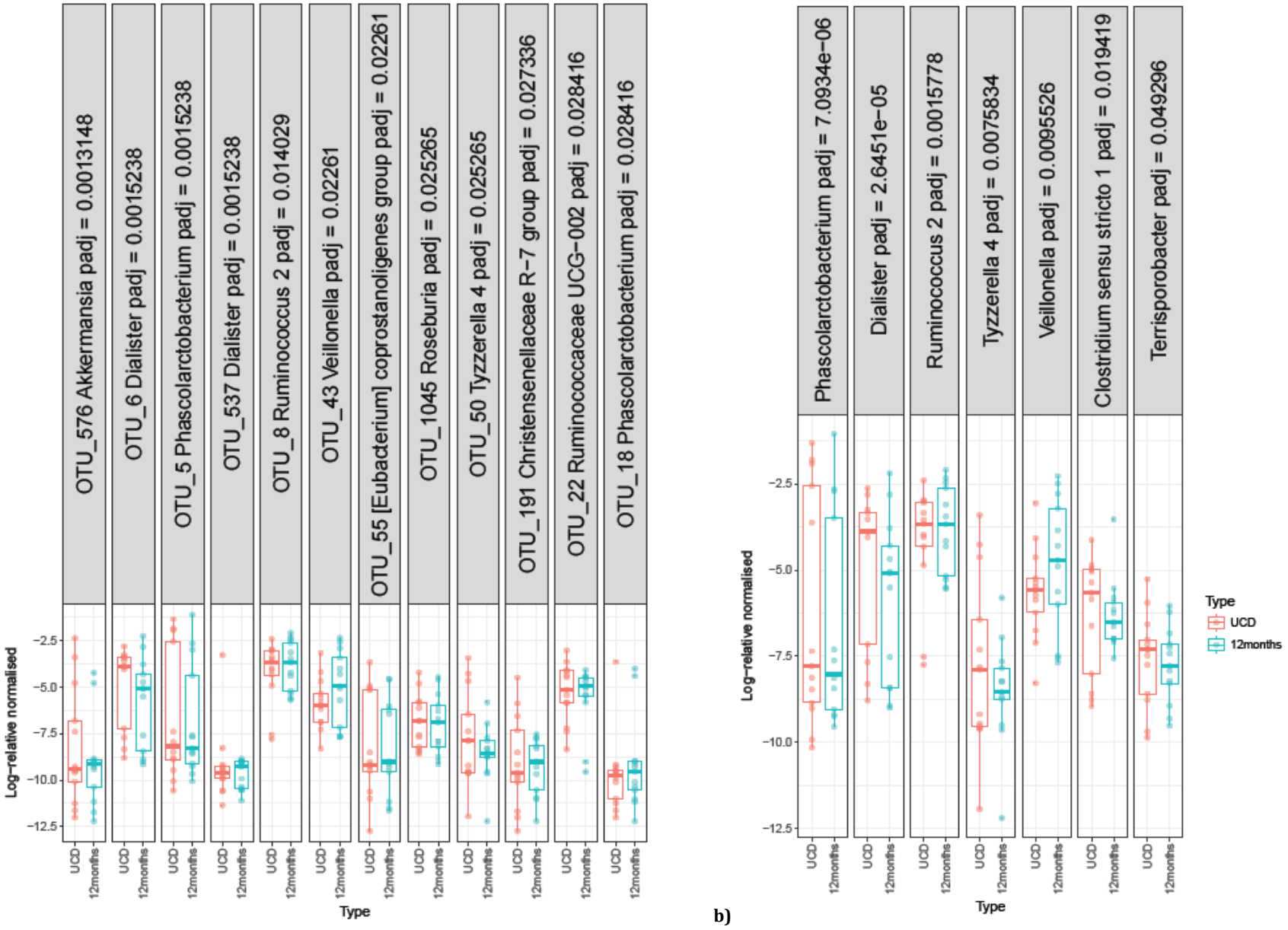


Figure 4.27: Log-relative abundances of (a) OTU and (b) genera that changed significantly at 12 months on GFD compared to the diagnosis

GFD dependent OTU

OTU that had significantly different relative abundance between the independent UCD and TCD groups, as well as within a CD cohort at six and 12 months after the initiation of GFD were considered to be GFD dependent.

Compared to baseline values, ten of 835 OTU, which characterised the faecal samples of CD prospective cohort, were found to have significantly different relative abundance at six and 12 months on GFD. All but three OTU had higher relative abundance at six and 12 months on GFD than at diagnosis (shown with superscript No2 ([2]) in Table 4.29). OTU_576 *Akkermansia,* OTU_5 *Phascolarctobacterium* and OTU_50 *Tyzzerella 4* had significantly lower relative abundance at six and 12 months on GFD than at diagnosis (shown with superscript No1 ([1]) in Table 4.29). Collectively, compared to baseline values, relative abundance of OTU belonging to *Phascolarctobacterium* and *Tyzzerella 4* was significantly decreased at six and 12 months on GFD (shown with superscript No1 ([1]) in Table 4.30), while relative abundance of OTU belonging to *Dialister, Veillonella, Clostridium sensu stricto 1, Ruminococcus 2* and *Terrisporobacter* was significantly increased (shown with superscript No2 ([2]) in Table 4.30).

Although none of the 10 OTU described above had significantly different relative abundance in the comparison of the independent UCD and TCD groups, OTU_259 *Erysipelatoclostridium* and OTU_98 *Ruminococcaceae* had significantly different relative abundance both at six months on GFD compared to the diagnosis, and at TCD group compared to UCD (shown with superscript No3 ([3]) in Table 4.29). However, this difference was of opposite direction. OTU_259 and OTU_98 had significantly lower relative abundance in faeces of CD children who were on GFD for six months compared to the diagnosis, but significantly higher relative abundance in independent TCD children who were on GFD for at least one year compared to UCD. In contrast, OTU_185 *Dorea* and OTU_90 *Ruminococcaceae* UCG-002 had significantly higher relative abundance in faecal samples of both TCD patients on GFD for at least one year and CD cohort on GFD for six months than UCD [OTU_185 *Dorea*: UCD Vs TCD: log2fold change= 2.06, *p-value (adjusted)*= 1.98E-02; UCD Vs GFD – 6 months: log2fold change= 4.22, *p-value (adjusted)*= 4.28E-02; OTU_90 *Ruminococcaceae* UCG-002: UCD Vs TCD: log2fold change= 2.04, *p-value (adjusted)*= 3.93E-03; UCD Vs GFD – 6 months: log2fold change= 4.67, *p-value (adjusted)*= 4.73E-02].

4.4 Association between microbiota SCFA and metadata

SCFA and BCFA of CD children on GFD at least one year (n= 60) were plotted against selected metadata. These data were age, gender, BMI z-score, faecal GIP, serum tTG, PedsQL-GSS score, diet composition characteristics [energy (%EAR), fat (g, %TDEI), protein (g, %TDEI, %RNI), carbohydrates (g, %TDEI), sugars (%TDEI), starch (%TDEI), NMES (%TDEI), NSP (g, %DRV), dietary fibre (g)], faecal pH and faecal water content to investigate potential correlations.

Regarding patients' demographic characteristics, age was positively associated with the concentration of heptanoic acid in dry matter (Rho= 0.311, *p*= 0.016) and its relative abundance (Rho= 0.283, *p*= 0.028, Table 4.31). When we looked for differences according to gender, the concentration of acetic, propionic, butyric, isocaproic acid and total SCFA in dry matter was significantly lower in female than in male patients (*p*= 0.003, 0.003, 0.016, 0.001 and 0.002 respectively, Table 4.34, Figure 4.29). Similarly, the concentration in wet matter of all the SCFA but butyric was significantly lower in female than in male CD children on GFD for at least one year (acetic; *p*= 0.005, propionic; *p*= 0.006, isocaproic; *p*= 0.004 and total SCFA; *p*= 0.006, Table 4.34). With regards to the percentage representation of SCFA, the relative abundance of propionic and isocaproic acid was significantly decreased in female compared with male CD children (*p*= 0.021 and 0.026 respectively), whereas the relative abundance of isobutyric, isovaleric and caproic acid was significantly increased in female compared with male patients (*p*= 0.046, 0.019 and 0.003 respectively, Table 4.34).

BMI z-score (SD) was positively associated with the concentration of isocaproic acid in wet matter (Rho= 0.266, *p*= 0.046, Table 4.31). Faecal GIP concentration (μg/g wet matter) had no significant association with faecal SCFA. When CD patients were classified into compliant and non-compliant based on their faecal GIP levels (compliant: GIP < 0.156 μg/g wet matter, non-compliant: GIP ≥ 0.156 μg/g wet matter), there was no difference among the two groups neither in SCFA concentration in dry, wet matter nor in their relative abundance (*p*> 0.05, Table 4.34). PedsQL-GSS score, where higher values indicate lower GI symptoms, was positively associated with the concentration in dry matter of acetic (Rho= 0.399, *p*= 0.002, Figure 4.30a), propionic (Rho= 0.317, *p*= 0.014), butyric acid (Rho= 0.271, *p*= 0.038) and total SCFA (Rho= 0.377, *p*= 0.003, Figure 4.30b, Table 4.31), showing that as GI symptoms increase, the concentrations in dry matter of faecal acetic, propionic, butyric and total SCFA decrease.

With regard to patients' dietary intake, fat consumption (g) was negatively associated with the concentration of isocaproic acid in dry matter (Rho= -0.276, *p*= 0.036, Table 4.31). Percentage of fat to TDEI was negatively associated with the concentration of caproic acid in dry matter (Rho= -0.266, *p*= 0.043, Table 4.32) and the relative abundance of isocaproic acid (Rho= -0.263, *p*= 0.046, Table 4.32). Total daily energy intake as a percentage to EAR was negatively associated with the concentration in dry matter of acetic (Rho= -0.314, *p*= 0.016, Figure 4.30c), butyric (Rho= -0.222, *p*= 0.026), isocaproic

(Rho= -0.351, *p*= 0.007) and total SCFA (Rho= -0.317, *p*= 0.015, Figure 4.30d) (Table 4.31). Total daily energy intake as a percentage to EAR was negatively associated with the concentration of isocaproic acid in wet matter (Rho= -0.293, *p*= 0.029, Table 4.31) as well. However, it was positively associated with the relative abundance of isovaleric acid (Rho= 0.284, *p*= 0.031, Table 4.31). Dietary consumption of protein (g) tended to be negatively associated with the concentration of isocaproic acid in dry matter (Rho= -0.233, *p*= 0.079 < 0.10, Table 4.32). Similarly, protein's percentage to RNI tended to be negatively associated with isocaproic acid's concentration in dry matter (Rho= -0.244, *p*= 0.064 < 0.10) and its relative abundance (Rho= -0.231, *p*= 0.082 < 0.10, Table 4.32).

Consumption of carbohydrates (g) was negatively associated with butyric acid (concentration in dry matter; Rho= -0.285, *p*= 0.030; concentration in wet matter; Rho= -0.283, *p*= 0.034; relative abundance; Rho= - 0.282, *p*= 0.032, Table 4.32). It was negatively associated with the concentration of total SCFA in dry matter as well (Rho= -0.269, *p*= 0.041). The percentage of carbohydrates to TDEI was positively associated with caproic acid's concentration in dry matter (Rho= 0.289, *p*= 0.028, Table 4.32). However, the percentage of starch, sugars and NMES to the TDEI was not associated with faecal SCFA. Similarly, the consumption of dietary fibre and NSP (g) was not significantly associated with any faecal SCFA.

Faecal pH was negatively and strongly associated with propionic, butyric and total SCFA [concentration in dry matter; propionic: Rho= -0.395, *p*= 0.020, butyric: Rho= -0.531, *p*< 0.0001, total SCFA: Rho= -0.310, *p*= 0.019; concentration in wet matter; propionic: Rho= -0.471, *p*< 0.0001, Figure 4.31c, butyric: Rho= -0.595, *p*< 0.0001, total SCFA: Rho= -0.408, *p*= 0.002, Figure 4.31a; relative abundance; propionic: Rho= -0.338, *p*= 0.010, butyric: Rho= -0.683, *p*< 0.0001, Figure 4.31d, Table 4.33]. Faecal pH was negatively associated with the concentration of isocaproic acid in dry (Rho= -0.332, *p*= 0.012) and wet matter (Rho= -0.340, *p*= 0.010, Table 4.33) as well. In contrast, faecal pH was positively associated with the relative abundance of acetic (Rho= 0.613, *p*< 0.0001, Figure 4.31b) and caproic acid (Rho= 0.274, *p*= 0.039, Table 4.33), with the former being strong correlation (Rho= 0.613, Figure 4.31b). Faecal water content was positively associated with the concentration in dry matter of acetic (Rho= 0.626, *p*< 0.0001, Figure 4.32a), propionic (Rho= 0.509, *p*< 0.0001, Figure 4.32b), butyric (Rho= 0.447, *p*= 0.001), isocaproic (Rho= 0.311, *p*= 0.017) and total SCFA (Rho= 0.593, *p*< 0.0001, Table 4.33). Finally, it was negatively associated with both the concentration in wet matter and the relative abundance of isobutyric (Rho= -0.285, *p*= 0.030 and Rho= -0.478, *p*< 0.0001, Figure 4.32c respectively), isovaleric (Rho= -0.414, *p*= 0.001 and Rho= -0.582, *p*< 0.0001, Figure 4.32d respectively) and valeric acid (Rho= -0.279, *p*= 0.034 and Rho= -0.479, *p*< 0.0001 respectively, Table 4.33).

Table 4.31: Spearman rank correlation between the absolute concentration (µmol/g dry matter) / relative abundance (%) of faecal SCFA and the age (years), BMI z-score (SD), faecal GIP, serum tTG, PedsQL-GSS score, energy intake and dietary intake of fat of CD patients on GFD at least one year

	Age (years)		BMI z-score (SD)		GIP (µg/g wet matter)		tTG (U/ mL)		PedsQL-GSS score		Energy (%EAR)		Fat (g)	
Absolute concentration (µmol/g dry matter)	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*
Acetic acid	0.093	0.482	0.152	0.251	-0.152	0.247	0.107	0.446	0.399	0.002	-0.314	0.016	-0.195	0.143
Propionic acid	0.033	0.801	0.145	0.272	-0.007	0.960	0.092	0.513	0.317	0.014	-0.227	0.086	-0.121	0.367
Isobutyric acid	0.088	0.504	0.027	0.840	0.067	0.612	0.055	0.694	0.192	0.144	-0.152	0.256	-0.104	0.439
Butyric acid	0.063	0.632	0.197	0.134	0.011	0.933	0.159	0.256	0.271	0.038	-0.292	0.026	-0.163	0.222
Isovaleric acid	0.141	0.284	0.020	0.879	0.123	0.347	0.033	0.813	0.115	0.385	-0.114	0.393	-0.063	0.638
Valeric acid	0.116	0.379	0.118	0.374	-0.004	0.977	-0.037	0.793	0.160	0.225	-0.209	0.115	-0.160	0.229
Isocaproic acid	0.163	0.214	0.238	0.069	-0.123	0.350	0.119	0.398	0.229	0.081	-0.351	0.007	-0.276	0.036
Caproic acid	0.218	0.094	-0.034	0.797	0.021	0.874	0.184	0.186	0.128	0.332	0.002	0.986	-0.049	0.714
Heptanoic acid	0.311	0.016	0.157	0.234	-0.023	0.859	0.122	0.383	0.020	0.881	0.039	0.769	0.049	0.715
Octanoic acid	0.121	0.356	0.093	0.482	0.105	0.427	0.203	0.145	-0.122	0.359	0.030	0.823	0.052	0.699
Total SCFA	0.065	0.624	0.145	0.273	-0.083	0.528	0.133	0.341	0.377	0.003	-0.317	0.015	-0.197	0.138
Absolute concentration (µmol/g wet matter)														
Acetic acid	0.024	0.860	0.217	0.105	-0.102	0.445	0.000	0.997	0.213	0.111	-0.178	0.189	-0.059	0.667
Propionic acid	0.060	0.654	0.155	0.250	0.058	0.666	0.070	0.623	0.215	0.108	-0.155	0.255	-0.006	0.967
Isobutyric acid	0.070	0.601	0.116	0.389	0.103	0.444	-0.114	0.425	0.069	0.612	-0.048	0.724	-0.007	0.958
Butyric acid	0.086	0.521	0.213	0.112	0.064	0.633	0.144	0.314	0.185	0.169	-0.256	0.057	-0.105	0.443
Isovaleric acid	0.066	0.622	0.139	0.304	0.136	0.307	-0.150	0.292	-0.011	0.937	-0.020	0.886	-0.001	0.995
Valeric acid	0.11	0.408	0.184	0.170	0.030	0.826	-0.074	0.607	0.010	0.943	-0.129	0.343	-0.066	0.630
Isocaproic acid	0.151	0.257	0.266	0.046	-0.094	0.484	0.091	0.523	0.101	0.454	-0.293	0.029	-0.229	0.090

Caproic acid	0.142	0.287	0.059	0.665	0.052	0.699	0.063	0.662	-0.009	0.950	0.085	0.534	0.021	0.879
Heptanoic acid	0.241	0.069	0.219	0.102	-0.036	0.786	0.006	0.967	-0.022	0.870	0.130	0.340	0.083	0.543
Octanoic acid	0.063	0.638	0.130	0.335	0.085	0.526	0.157	0.270	-0.124	0.359	0.047	0.734	0.040	0.768
Total SCFA	0.062	0.643	0.189	0.159	-0.023	0.863	0.060	0.676	0.204	0.128	-0.197	0.146	-0.048	0.726
Relative abundance (%)														
Acetic acid	-0.058	0.658	-0.166	0.209	-0.210	0.107	-0.045	0.750	-0.061	0.644	0.142	0.287	0.059	0.661
Propionic acid	0.050	0.703	0.101	0.444	0.049	0.709	-0.059	0.673	0.159	0.229	-0.084	0.529	0.006	0.965
Isobutyric acid	-0.017	0.900	-0.064	0.628	0.160	0.222	-0.058	0.682	-0.165	0.211	0.255	0.053	0.128	0.340
Butyric acid	0.025	0.850	0.216	0.100	0.112	0.392	0.163	0.245	0.155	0.384	-0.218	0.100	-0.131	0.327
Isovaleric acid	-0.018	0.894	-0.049	0.710	0.163	0.212	-0.093	0.509	-0.205	0.119	0.284	0.031	0.151	0.256
Valeric acid	0.037	0.780	0.058	0.663	0.127	0.334	0.017	0.902	-0.185	0.160	0.161	0.227	0.063	0.637
Isocaproic acid	0.155	0.238	0.224	0.088	-0.099	0.450	0.098	0.484	0.042	0.753	-0.257	0.052	-0.256	0.052
Caproic acid	0.165	0.208	-0.153	0.249	0.038	0.772	0.047	0.738	-0.129	0.329	0.192	0.148	0.092	0.490
Heptanoic acid	0.283	0.028	0.082	0.537	-0.046	0.726	0.060	0.669	-0.047	0.721	0.156	0.244	0.132	0.322
Octanoic acid	0.030	0.821	0.097	0.467	0.088	0.504	0.152	0.278	-0.219	0.096	0.096	0.475	0.077	0.566

Table 4.32: Spearman rank correlation between the absolute concentration (µmol/g dry matter) / relative abundance (%) of faecal SCFA and the dietary intake of fat (%TDEI), protein (g, %TDEI, %RNI), carbohydrates (g, %TDEI) and sugars (%TDEI) of CD patients on GFD at least one year

	Fat (%TDEI)		Protein (g)		Protein (%TDEI)		Protein (%RNI)		Carbohydrates (g)		Carbohydrates (%TDEI)		Sugars (%TDEI)	
Absolute concentration (µmol/g dry matter)	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*
Acetic acid	0.029	0.829	-0.146	0.274	0.143	0.284	-0.153	0.252	-0.249	0.059	0.033	0.806	-0.014	0.916
Propionic acid	0.042	0.755	-0.085	0.525	0.076	0.570	-0.106	0.427	-0.163	0.222	0.065	0.629	-0.150	0.262
Isobutyric acid	-0.151	0.257	0.002	0.988	0.036	0.787	-0.076	0.572	-0.067	0.616	0.160	0.232	-0.017	0.898
Butyric acid	0.062	0.642	-0.182	0.171	0.040	0.766	-0.181	0.174	-0.285	0.030	0.022	0.868	-0.126	0.344
Isovaleric acid	-0.157	0.238	0.028	0.836	0.039	0.769	-0.075	0.576	-0.034	0.802	0.143	0.283	-0.044	0.742
Valeric acid	-0.098	0.466	-0.096	0.472	-0.017	0.898	-0.143	0.286	-0.149	0.263	0.096	0.474	-0.026	0.847
Isocaproic acid	-0.127	0.342	-0.233	0.079	0.041	0.759	-0.244	0.064	-0.215	0.105	0.170	0.203	0.069	0.607
Caproic acid	-0.266	0.043	-0.001	0.996	-0.162	0.223	-0.186	0.162	0.140	0.294	0.289	0.028	0.138	0.302
Heptanoic acid	-0.152	0.254	0.115	0.388	0.036	0.786	-0.163	0.221	0.136	0.307	0.071	0.595	0.088	0.512
Octanoic acid	-0.027	0.841	0.117	0.383	0.118	0.377	0.026	0.847	-0.001	0.992	-0.030	0.824	-0.012	0.928
Total SCFA	0.043	0.746	-0.161	0.228	0.119	0.374	-0.147	0.271	-0.269	0.041	0.034	0.803	-0.067	0.619
Absolute concentration (µmol/g wet matter)														
Acetic acid	0.113	0.407	-0.081	0.555	0.148	0.277	0.020	0.882	-0.175	0.196	-0.123	0.368	-0.089	0.512
Propionic acid	0.114	0.403	-0.007	0.957	0.102	0.455	-0.039	0.777	-0.107	0.432	-0.031	0.819	-0.147	0.279
Isobutyric acid	-0.057	0.678	0.028	0.839	0.055	0.688	0.048	0.727	-0.024	0.860	-0.021	0.879	-0.041	0.764
Butyric acid	0.108	0.427	-0.170	0.211	0.051	0.708	-0.154	0.257	-0.283	0.034	-0.077	0.574	-0.105	0.441
Isovaleric acid	-0.051	0.710	0.034	0.804	0.060	0.663	0.062	0.647	-0.012	0.930	-0.026	0.848	-0.054	0.693
Valeric acid	-0.036	0.790	-0.051	0.708	0.033	0.811	-0.048	0.724	-0.100	0.462	-0.042	0.756	-0.034	0.804
Isocaproic acid	-0.136	0.317	-0.187	0.167	0.036	0.791	-0.191	0.158	-0.168	0.216	0.122	0.371	0.0997	0.477

Caproic acid	-0.163	0.231	0.023	0.864	-0.161	0.235	-0.064	0.640	0.141	0.300	0.128	0.348	0.063	0.646
Heptanoic acid	-0.161	0.237	0.112	0.410	-0.012	0.928	-0.060	0.663	0.164	0.227	0.033	0.808	0.012	0.932
Octanoic acid	0.009	0.946	0.091	0.503	0.120	0.378	0.094	0.492	-0.030	0.829	-0.075	0.585	-0.039	0.776
Total SCFA	0.122	0.371	-0.079	0.565	0.120	0.378	-0.028	0.836	-0.190	0.160	-0.113	0.407	-0.116	0.394
Relative abundance (%)														
Acetic acid	-0.080	0.549	0.065	0.628	0.004	0.976	0.115	0.388	0.177	0.183	0.007	0.956	0.127	0.344
Propionic acid	0.031	0.819	0.053	0.692	0.089	0.505	-0.018	0.894	-0.003	0.980	0.051	0.702	-0.163	0.221
Isobutyric acid	-0.181	0.173	0.176	0.187	-0.023	0.865	0.147	0.272	0.202	0.128	0.068	0.615	-0.065	0.626
Butyric acid	0.068	0.613	-0.184	0.166	-0.027	0.841	-0.150	0.261	-0.282	0.032	-0.016	0.905	-0.143	0.284
Isovaleric acid	-0.150	0.261	0.175	0.190	-0.045	0.735	0.170	0.201	0.216	0.103	0.050	0.711	-0.063	0.636
Valeric acid	-0.128	0.338	0.063	0.636	-0.067	0.617	0.065	0.627	0.118	0.379	0.046	0.733	-0.035	0.793
Isocaproic acid	-0.263	0.046	-0.195	0.142	-0.029	0.829	-0.231	0.082	-0.093	0.489	0.219	0.099	0.146	0.273
Caproic acid	-0.167	0.210	0.076	0.569	-0.217	0.103	-0.107	0.424	0.232	0.079	0.145	0.279	0.134	0.317
Heptanoic acid	-0.133	0.321	0.147	0.272	-0.049	0.712	-0.104	0.436	0.212	0.110	0.052	0.701	0.066	0.624
Octanoic acid	-0.013	0.922	0.120	0.368	0.094	0.482	0.095	0.478	0.034	0.800	-0.042	0.754	-0.027	0.843

Table 4.33: Spearman rank correlation between the absolute concentration (µmol/g dry & wet matter)/ relative abundance (%) of SCFA and the dietary intake of starch (%TDEI), NMES (%TDEI), NSP (g, %TDEI), dietary fibre (g), the faecal pH and water content (%) of CD patients on GFD at least one year

	Starch (%TDEI)		NMES (%TDEI)		NSP (g)		NSP (%DRV)		Dietary fibre (g)		Faecal pH		Faecal water content (%)	
Absolute concentration (µmol/g dry matter)	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*	Rho	*p-value*
Acetic acid	0.116	0.388	-0.014	0.915	-0.112	0.404	0.016	0.912	-0.124	0.363	-0.190	0.156	0.626	< 0.0001
Propionic acid	-0.008	0.954	0.034	0.797	-0.206	0.122	-0.025	0.863	-0.104	0.448	-0.395	0.002	0.509	< 0.0001
Isobutyric acid	0.039	0.772	-0.051	0.703	0.013	0.922	-0.047	0.752	0.097	0.479	-0.048	0.723	0.216	0.104
Butyric acid	0.064	0.633	-0.013	0.921	-0.157	0.240	-0.089	0.546	-0.162	0.234	-0.531	< 0.0001	0.447	< 0.0001
Isovaleric acid	0.063	0.640	-0.087	0.514	0.112	0.403	-0.047	0.752	0.128	0.346	-0.028	0.836	0.063	0.638
Valeric acid	0.038	0.777	0.001	0.994	-0.090	0.501	-0.013	0.930	-0.047	0.732	-0.138	0.305	0.122	0.362
Isocaproic acid	-0.136	0.310	-0.050	0.708	-0.168	0.207	-0.030	0.837	-0.162	0.233	-0.332	0.012	0.311	0.017
Caproic acid	0.198	0.136	-0.061	0.649	0.221	0.096	0.229	0.117	0.156	0.252	0.110	0.415	0.137	0.305
Heptanoic acid	0.222	0.093	-0.145	0.277	0.242	0.068	0.226	0.122	0.154	0.257	0.149	0.268	-0.029	0.827
Octanoic acid	0.244	0.065	-0.145	0.278	0.191	0.151	0.020	0.892	0.056	0.684	0.100	0.458	-0.101	0.451
Total SCFA	0.110	0.413	-0.035	0.792	-0.137	0.304	-0.030	0.842	-0.139	0.306	-0.310	0.019	0.593	< 0.0001
Absolute concentration (µmol/g wet matter)														
Acetic acid	0.103	0.450	-0.129	0.345	0.015	0.911	-0.121	0.417	-0.023	0.868	-0.290	0.030	0.140	0.294
Propionic acid	-0.016	0.906	-0.043	0.754	-0.164	0.227	-0.103	0.491	-0.053	0.702	-0.471	< 0.0001	0.222	0.093
Isobutyric acid	0.008	0.955	-0.133	0.328	0.074	0.586	-0.202	0.174	0.145	0.294	-0.095	0.487	-0.285	0.030
Butyric acid	0.074	0.587	-0.099	0.468	-0.124	0.361	-0.204	0.170	-0.128	0.356	-0.595	< 0.0001	0.145	0.276
Isovaleric acid	0.003	0.981	-0.136	0.317	0.133	0.327	-0.157	0.291	0.146	0.292	-0.099	0.470	-0.414	0.001
Valeric acid	0.029	0.832	-0.076	0.579	-0.023	0.869	-0.118	0.430	-0.016	0.906	-0.175	0.198	-0.279	0.034
Isocaproic acid	-0.172	0.206	-0.096	0.479	-0.132	0.331	-0.092	0.537	-0.110	0.428	-0.340	0.010	0.042	0.753

Caproic acid	0.077	0.573	-0.063	0.647	0.193	0.153	0.191	0.198	0.153	0.270	0.034	0.804	-0.244	0.065
Heptanoic acid	0.170	0.209	-0.139	0.308	0.262	0.051	0.204	0.170	0.211	0.126	0.167	0.218	-0.199	0.135
Octanoic acid	0.203	0.133	-0.133	0.329	0.172	0.204	0.003	0.982	0.034	0.806	0.096	0.482	-0.178	0.181
Total SCFA	0.077	0.572	-0.109	0.425	-0.039	0.777	-0.127	0.393	-0.037	0.790	-0.408	0.002	0.144	0.279
Relative abundance (%)														
Acetic acid	0.051	0.705	0.000	0.999	0.153	0.250	0.116	0.433	0.155	0.253	0.613	< 0.0001	-0.026	0.845
Propionic acid	-0.194	0.144	0.100	0.456	-0.223	0.093	-0.034	0.818	-0.030	0.824	-0.338	0.010	0.227	0.086
Isobutyric acid	0.020	0.880	-0.020	0.883	0.178	0.181	-0.060	0.686	0.245	0.068	0.231	0.084	-0.478	0.0001
Butyric acid	0.032	0.810	-0.030	0.825	-0.118	0.380	-0.167	0.255	-0.184	0.175	-0.683	< 0.0001	0.151	0.256
Isovaleric acid	0.012	0.929	-0.062	0.644	0.250	0.059	-0.024	0.870	0.246	0.068	0.184	0.171	-0.582	< 0.0001
Valeric acid	-0.027	0.842	0.022	0.872	0.015	0.913	-0.105	0.478	0.083	0.542	0.038	0.780	-0.479	< 0.0001
Isocaproic acid	-0.249	0.059	-0.015	0.914	-0.155	0.245	-0.062	0.673	-0.113	0.408	-0.211	0.116	-0.001	0.995
Caproic acid	0.073	0.586	-0.048	0.720	0.194	0.144	0.224	0.126	0.192	0.157	0.274	0.039	-0.241	0.069
Heptanoic acid	0.173	0.194	-0.123	0.356	0.239	0.070	0.209	0.154	0.212	0.118	0.252	0.059	-0.180	0.177
Octanoic acid	0.234	0.076	-0.116	0.384	0.216	0.103	0.020	0.893	0.045	0.742	0.154	0.254	-0.216	0.103

Table 4.34: Association and Univariate regression analysis between the absolute concentration (µmol/g dry & wet matter)/ relative abundance (%) of faecal SCFA and the categorical variables (a) gender and (b) compliance with GFD based on faecal GIP levels of CD patients on GFD at least one year

	Compliant (GIP < 0.156 µg/g wet matter) / Non-compliant (GIP > 0.156 µg/g wet matter)			Male / Female		
	Mann - Whitney	Univariate regression analysis		Mann - Whitney	Univariate regression analysis	
Absolute concentration (µmol/g dry matter)	*p-value*	R^2	*p-value*	*p-value*	R^2	*p-value*
Acetic acid	0.223	0.00	0.340	0.003	12.83	0.003
Propionic acid	0.851	0.00	0.957	0.003	13.35	0.002
Isobutyric acid	0.670	0.00	0.848	0.052	1.79	0.155
Butyric acid	0.929	0.00	0.681	0.016	2.94	0.100
Isovaleric acid	0.388	0.00	0.470	0.223	0.00	0.443
Valeric acid	0.897	0.00	0.836	0.172	0.00	0.329
Isocaproic acid	0.298	0.00	0.373	0.001	8.79	0.012
Caproic acid	0.945	0.00	0.905	0.298	0.00	0.350
Heptanoic acid	0.929	0.00	0.685	0.485	0.00	0.971
Octanoic acid	0.528	5.11	0.046	0.360	0.38	0.273
Total SCFA	0.457	0.00	0.504	0.002	11.80	0.004
Absolute concentration (µmol/g wet matter)						
Acetic acid	0.146	0.00	0.377	0.005	10.75	0.007
Propionic acid	0.797	0.00	0.963	0.006	8.23	0.016
Isobutyric acid	0.452	0.00	0.530	0.171	0.00	0.408
Butyric acid	0.781	0.00	0.678	0.066	0.25	0.290
Isovaleric acid	0.299	0.00	0.362	0.467	0.00	0.799

Valeric acid	0.910	0.00	0.665	0.344	0.00	0.709
Isocaproic acid	0.428	0.13	0.305	0.004	6.95	0.026
Caproic acid	0.734	0.00	0.857	0.313	2.06	0.144
Heptanoic acid	0.893	0.00	0.778	0.462	0.00	0.796
Octanoic acid	0.640	5.59	0.041	0.394	0.32	0.281
Total SCFA	0.765	0.00	0.575	0.006	8.07	0.017
Relative abundance (%)						
Acetic acid	0.145	3.02	0.098	0.342	0.95	0.216
Propionic acid	0.774	1.85	0.152	0.021	8.52	0.013
Isobutyric acid	0.194	2.72	0.109	0.046	1.89	0.149
Butyric acid	0.506	0.00	0.649	0.240	0.00	0.362
Isovaleric acid	0.181	3.01	0.098	0.019	4.14	0.065
Valeric acid	0.356	0.92	0.218	0.115	2.31	0.127
Isocaproic acid	0.433	0.00	0.333	0.026	3.54	0.080
Caproic acid	0.774	0.00	0.445	0.003	9.08	0.011
Heptanoic acid	0.842	0.00	0.909	0.159	0.09	0.309
Octanoic acid	0.592	5.73	0.036	0.188	0.81	0.229

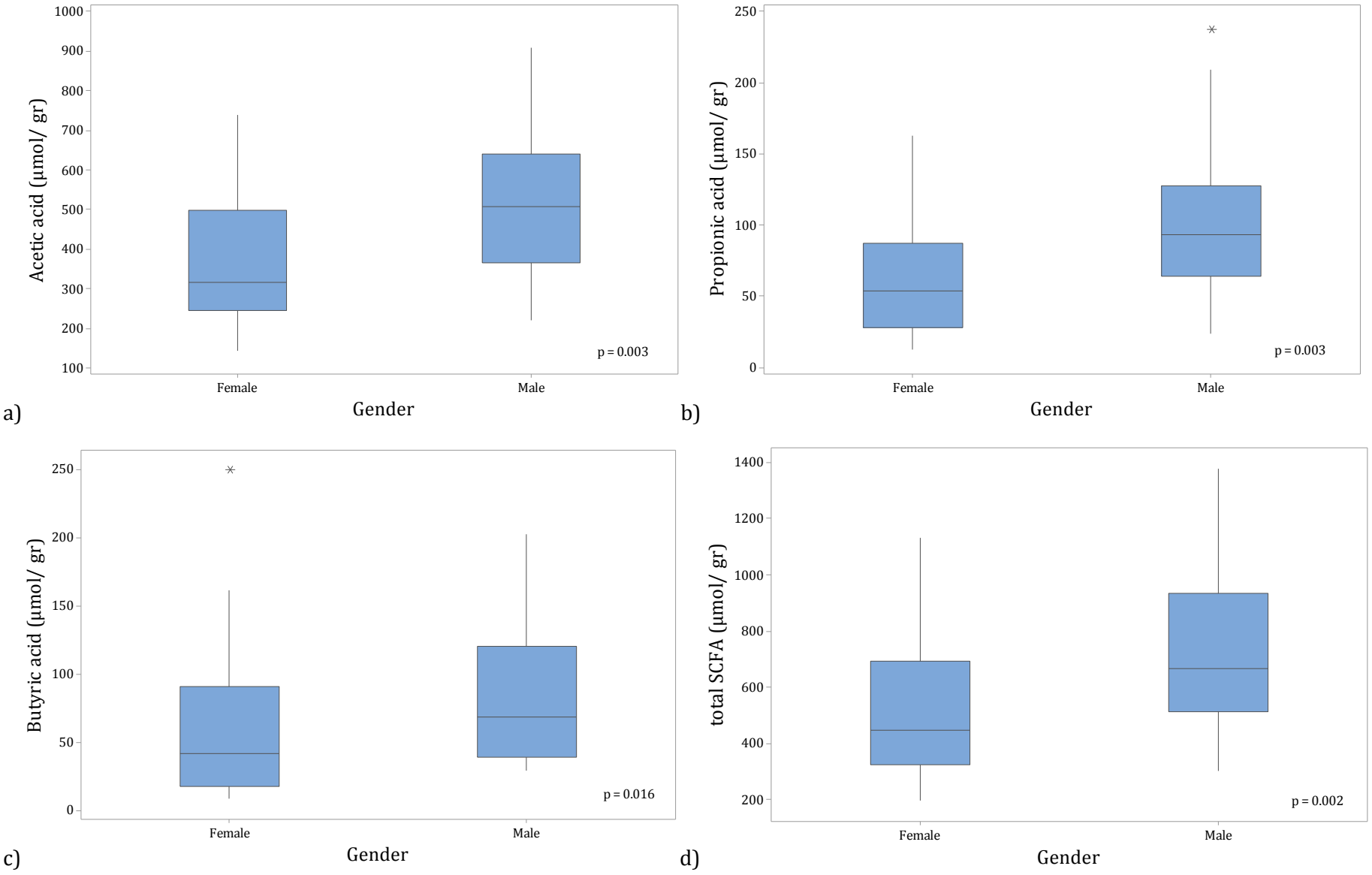


Figure 4.28: Concentration in dry matter (µmol/g) of (a) acetic, (b) propionic, (c) butyric acid, (d) total SCFA in female and male CD children on GFD at least one year

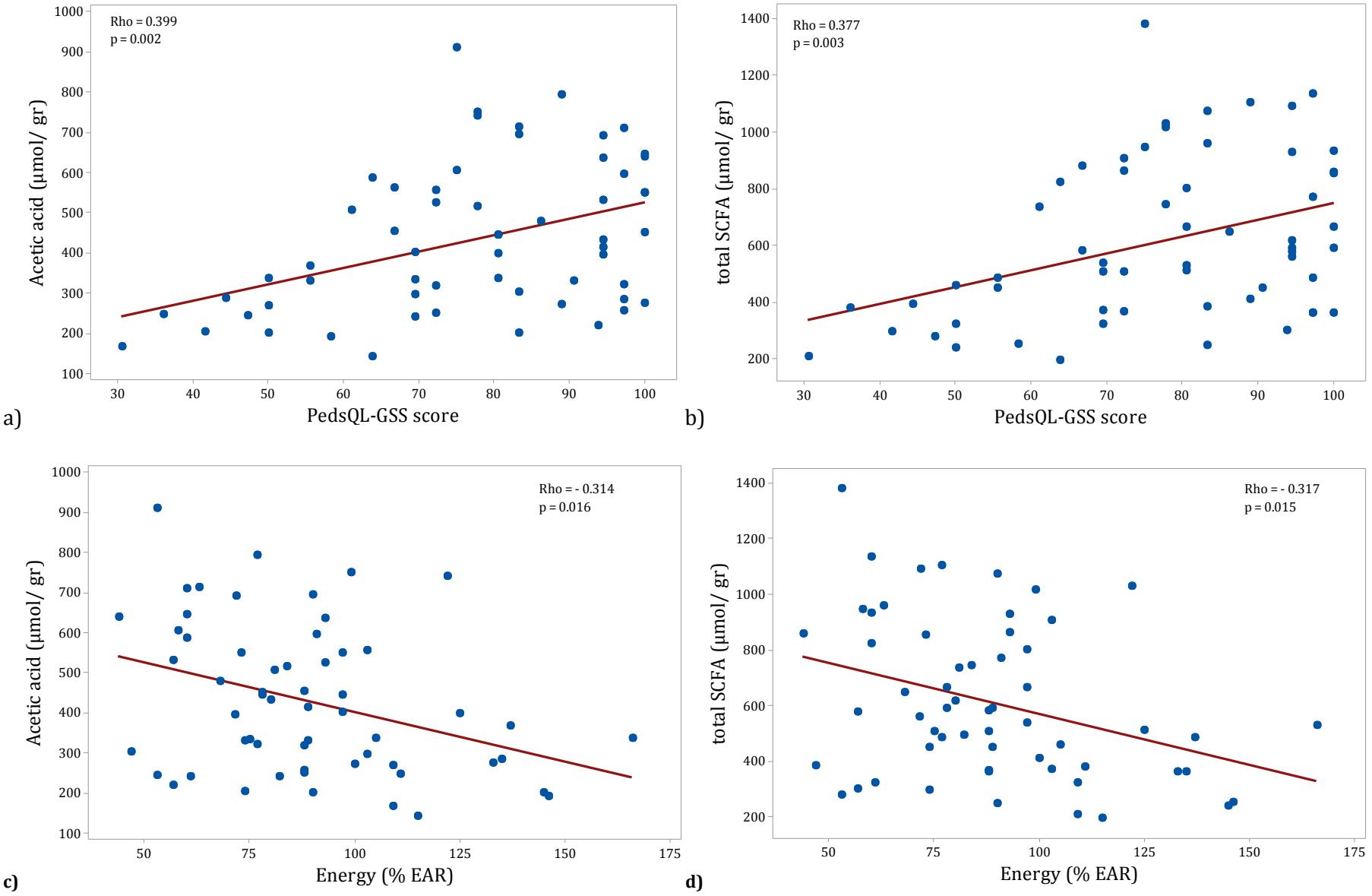


Figure 4.29: Spearman rank correlation between the PedsQL-GSS score and the concentration in dry matter of faecal (a) acetic acid, (b) total SCFA, as well as between total daily energy intake (%EAR) and the concentration in dry matter of (a) acetic acid, (d) total SCFA in CD children (n=60) on GFD at least one year

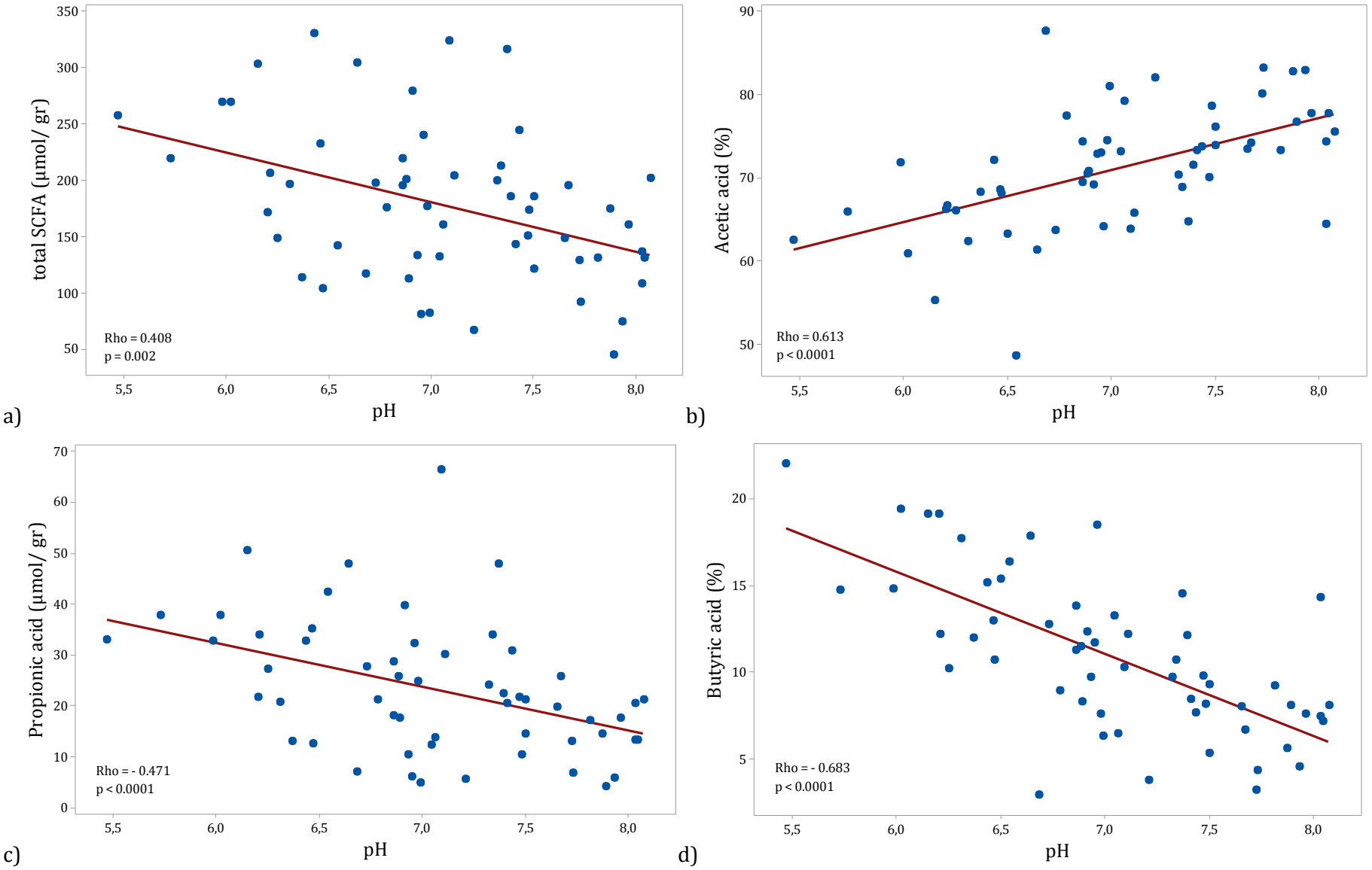


Figure 4.30: Spearman rank correlation between faecal pH and the (a) concentration of total SCFA (µmol/g) in wet matter, (b) relative abundance of acetic acid, (c) concentration of propionic acid (µmol/g) in wet matter, (d) relative abundance of butyric acid of CD children (n=60) on GFD at least one year

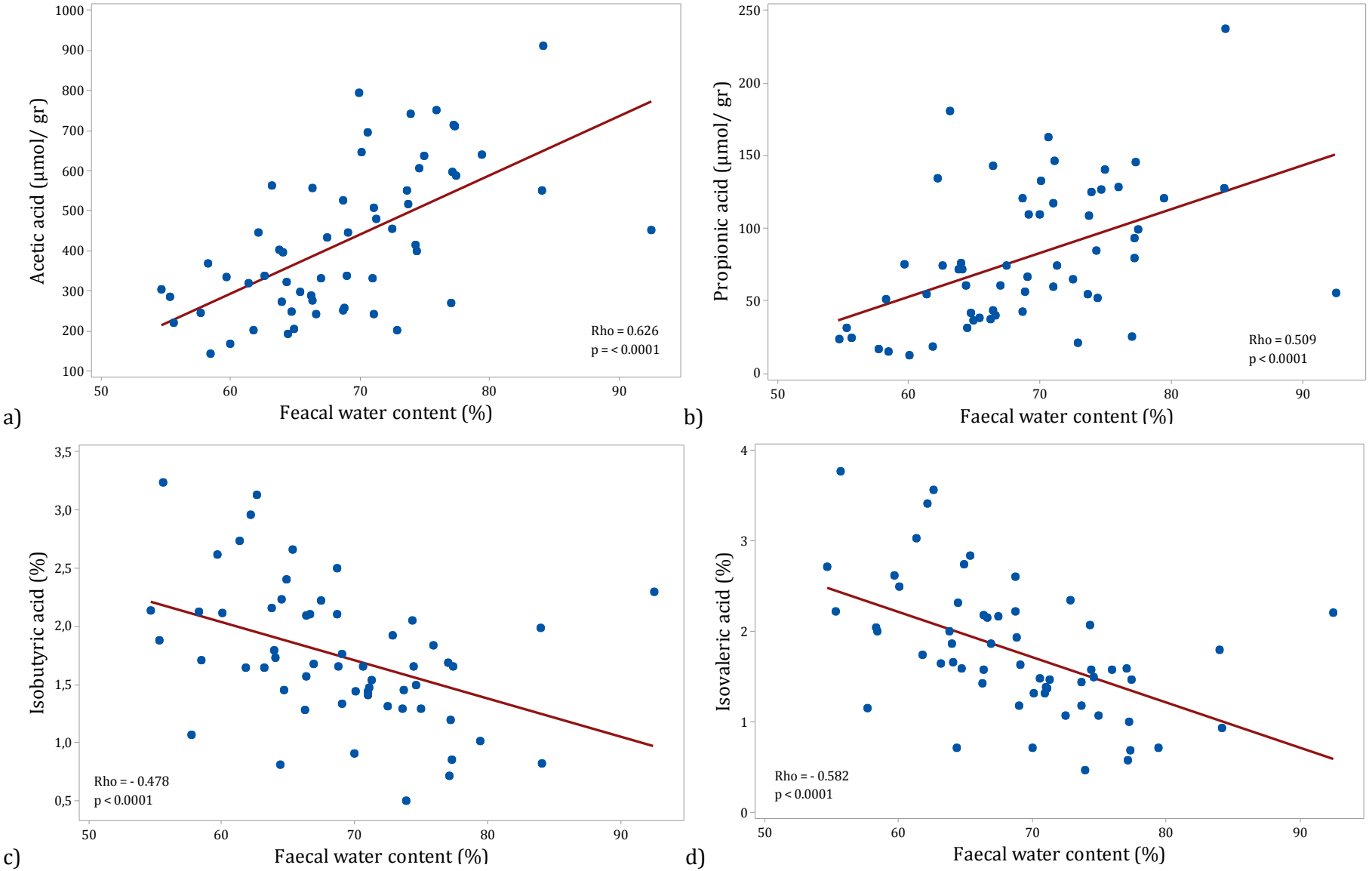


Figure 4.31: Spearman rank correlation between the faecal water content (%) and the concentration of (a) acetic acid in dry matter (µmol/g), (b) propionic acid in dry matter (µmol/g), relative abundance of (c) isobutyric acid (%), and (d) isovaleric acid (%) of CD children (n=60) on GFD at least one year

4.4.1.1 Predictors of faecal SCFA levels

Potential predictors of faecal SCFA levels in CD children who are on GFD for at least one year (n= 60) were defined a priori. Subjects and faecal samples characteristics including age, gender, BMI z-score, diet components [total daily energy intake (%EAR), fat (g, %TDEI), protein (g, %TDEI, %RNI), carbohydrates (g, %TDEI), sugars (%TDEI), starch (%TDEI), NMES (%TDEI), NSP (g, %DRV) and dietary fibre (g)], faecal GIP concentration, serum tTG levels, PedsQL-GSS score, and faecal water content were tested using linear regression analysis with backward stepwise method to assess their association with SCFA relative abundance and concentration in dry and wet matter.

In the univariate regression analysis, independent associations showed that increased concentrations of serum tTG, low TDEI (%EAR), increased PedsQL-GSS score, thus decreased GI symptoms, and increased faecal water content were strongly associated with increased acetic acid concentration in dry matter (tTG concentration: R^2= 7.3%, p= 0.029, TDEI (%EAR): R^2= 15.3%, p= 0.001, PedsQL-GSS score: R^2= 10.9%, p= 0.007, Table 4.35, and faecal water content: R^2= 35.8%, p< 0.0001, Table 4.37). Moreover, males were more likely to have increased acetic acid concentration in dry matter than females (p= 0.003, Table 4.34). Accordingly, increased serum tTG concentration, increased PedsQL-GSS score, thus fewer GI symptoms, along with high BMI z-score and elevated proportion of starch to TDEI were significantly associated with increased concentration of acetic acid in wet matter (tTG concentration: R^2= 6.5%, p= 0.039, PedsQL-GSS score: R^2= 5.3%, p= 0.047, BMI z-score: R^2= 5.2%, p= 0.048, Table 4.35, and starch intake (%TDEI): R^2= 8.7%, p= 0.015, Table 4.37). Similarly, to acetic acid concentration in dry matter, males were more likely to have increased acetic acid concentration in wet matter than females (p= 0.007, Table 4.34).

In the multivariate regression analysis, where predictors with *p-value* < 0.10 were used, faecal water content, percentage of starch to TDEI, as well as the child's gender were significant predictors, explaining 45.6% of the variance in the concentration of acetic acid in dry matter (p< 0.0001, Table 4.38). Likewise, serum tTG concentration and child's gender explained 17.8% of the variation in the concentration of acetic acid in wet matter (p= 0.005, Table 4.39). Finally, variance in the relative abundance of acetic acid was only explained by faecal GIP concentration, with low GIP levels in faeces, so increased compliance with GFD, being strongly associated with increased relative abundance of acetic acid (R^2= 13.8%, p= 0.002, Table 4.35).

All significant predictors of the concentration of acetic acid in dry matter except for serum tTG were significant predictors of the concentration of propionic acid in dry matter as well. Specifically, 13.4% of the variance in the concentration of propionic acid in dry matter was explained independently by patient's gender (p= 0.002, Table 4.34), with boys being more likely to have increased propionic acid concentration in dry matter than girls. Moreover, low TDEI (%EAR), increased PedsQL-GSS score, and increased faecal water content were significantly associated with increased concentration of faecal propionic acid in dry matter (TDEI (%EAR): R^2= 6.4%, p= 0.031,

PedsQL-GSS score R^2= 7.6%, *p*= 0.019, Table 4.35, and faecal water content: R^2= 21.3%, *p*< 0.0001, Table 4.37). In the multivariate analysis, only the faecal water content remained significant predictor of the concentration of propionic in dry matter, explaining 26.8% of the variation in propionic levels (*p*< 0.0001, Table 4.38). Males were more likely to have increased levels of propionic acid expressed per wet matter compared with females, as well as increased relative abundance of propionic (R^2= 8.23%, *p*= 0.016, Table 4.34). Variance in the relative abundance of propionic acid was also independently explained by faecal GIP concentration, and child's gender (R^2= 12.90%, *p*= 0.003, and R^2= 8.23%, *p*= 0.016 respectively). Increased levels of GIP in faeces, which indicate decreased compliance with GFD, were significantly associated with increased relative abundance of propionic. In the multivariate analysis both the faecal GIP concentration and child's gender remained significant predictors (*p*= 0.001 and 0.029 respectively, Table 4.40), and combined with the dietary intake of NSP (g) that tended to predict relative abundance of propionic in the univariate analysis (*p*= 0.086 < 0.010, Table 4.37) explained 25.7% of the variance in the relative abundance of propionic acid (*p*< 0.0001, Table 4.40).

Although faecal GIP concentration was, as mentioned above, significantly associated with the relative abundance of acetic and propionic acid, the classification of CD children into compliant and non-compliant with GFD according to their faecal GIP levels was not a significant categorical predictor of the relative abundance of acetic and propionic acid in the univariate regression analysis (*p*= 0.098 and 0.152 respectively, Table 4.34), and hence it was not included in the multivariate regression analysis.

Variance in the concentration of butyric acid in dry matter was independently explained by the following predictors; PedsQL-GSS score (R^2= 6.29%, *p*= 0.031, Table 4.35), TDEI (%EAR) (R^2= 7.09%, *p*= 0.024, Table 4.35), and faecal water content (R^2= 14.50%, *p*= 0.002, Table 4.37). Increased PedsQL-GSS score and faecal water content, along with decreased TDEI (%EAR) were significantly associated with increased acetic acid levels in dry matter. However, in the multivariate analysis, where all predictors with *p-value* < 0.10 were used, only the faecal water content and dietary intake of starch (%TDEI) remained significant, explaining 19.5% of the variance in the concentration of butyric acid in dry matter (*p*= 0.003, Table 4.38). With regard to the concentration pf butyric acid expressed per wet matter, it was found that although there was no significant predictor in the univariate analysis regression, in the multivariate analysis were predictors with *p-value* < 0.10 (BMI z- score, serum tTG concentration, and starch dietary intake (%TDEI) were also accounted, the serum tTG concentration explained 7.1% of the variance (*p*= 0.037, Table 4.39).

In the univariate regression analysis, faecal water content explained 7.6% of the variance in the concentration of valeric acid in wet matter and 16.2% of the variance in its relative abundance (*p*= 0.021 and 0.001 respectively, Table 4.37). In the multivariate regression analysis, faecal water content remained significant in both cases, explaining 7.6% and 16.3% of the variance in the valeric acid

concentration (expressed per wet matter) and relative abundance respectively (p= 0.021 and 0.001 respectively).

Patient's gender was the only significant predictor of the relative abundance of caproic acid (R^2= 9.1%, p= 0.011, Table 4.34), with females being more likely to have increased levels of caproic acid relative abundance than boys. Accordingly, patient's age was the only significant predictor of the concentration of heptanoic acid in dry matter (R^2= 5.8%, p= 0.035, Table 4.35), with older children being more likely to have increased heptanoic acid concentration in dry matter than younger children. There was no significant predictor of the concentration of heptanoic acid in wet matter and its relative abundance.

Regarding octanoic acid's predictors, it was found that the classification to compliant and non-compliant with GFD patients based on the faecal GIP levels explained 5.1% (p= 0.046) and 8.1% (p= 0.017) of the variance in its concentration expressed per dry and wet matter respectively, and 5.7% of the variance in its relative abundance (p= 0.036) (Table 4.34). In all three cases non-compliant children were more likely to have increased octanoic acid levels compared to compliant. Decreased PedsQL-GSS score, so more GI symptoms, was significantly associated with octanoic acid's concentration in wet matter and its relative abundance (R^2= 5.2%, p= 0.049 and R^2= 6.7%, p= 0.027 respectively). However, in the multivariate regression analysis, only PedsQL-GSS score remained significant and explained 5.2% (p= 0.049, Table 4.39) and 6.7% (p= 0.027, Table 4.40) of the variance in the concentration per wet matter and relative abundance of octanoic acid respectively.

Faecal water content was the only independent predictor of isobutyric and isovaleric acids concentration in wet matter (R^2= 9.4%, p= 0.011 and R^2= 16.2%, p= 0.001 respectively, Table 4.37). Accordingly, there was no significant predictor when isobutyric and isovaleric acid was measured in dry matter. When predictors were tested against the relative abundance of isobutyric and isovaleric acids, not only faecal water content but also the TDEI (%EAR), and dietary intake of carbohydrates (g) and dietary fibre (g) were independent predictors (p< 0.05, Tables 4.35, 4.36 and 4.37 respectively). However, in the multivariate analysis, only the faecal water content remained significant and a model consisting of this and the percentage intake of fat (%TDEI), which tended to predict the relative abundance of isobutyric and isovaleric acid (p< 0.10), explained 22.2% and 29.5% of the variance in the relative abundance of isobutyric and isovaleric acid respectively (p= 0.001 and <0.0001 respectively, Table 4.40).

Patient's gender was a significant independent predictor of isocaproic acid's concentration in dry and wet matter (R^2= 8.8%, p= 0.012 and R^2= 6.3%, p= 0.034 respectively), with males being more likely to have increased levels compared with females (Table 4.34.) Increased levels of serum tTG were also significantly associated with increased concentration of isocaproic acid in wet matter, but in the multivariate analysis only the patients' gender remained significant predictor explaining 8.4% (p=

0.021, Table 4.38) and 6.7% (p= 0.037, Table 4.39) of the variance in isocaproic acid's concentration expressed per dry and wet matter respectively.

Finally, increased serum tTG levels, PedsQL-GSS score and faecal water content, along with decreased TDEI (%EAR) were significantly associated with increased total SCFA concentration in dry matter [tTG: R^2=6.1%, p= 0.041, PedsQL-GSS score: R^2=13.4%, p= 0.003, faecal water content: R^2=31.0%, p< 0.0001 and TDEI (%EAR): R^2=10.3%, p= 0.008]. Increased levels of starch intake (%TDEI) were significantly associated with increased levels of total SCFA in wet matter (R^2=5.4%, p= 0.048, Table 4.37). Moreover, boys were more likely to have increased total SCFA concentration expressed per both dry and wet matter (p= 0.004 and 0.017 respectively, Table 4.34). However, in the multivariate analysis only the faecal water content remained significant predictor explaining 20.9% (p< 0.001) of the variance in the concentration of total SCFA in dry matter, and the serum tTG that tended to predict the total SCFA concentration in wet matter in the univariate analysis (p= 0.073 < 0.10, 4.35) but in the multivariate explained significantly 7.2% of this variance (p= 0.038, Table 4.39).

Table 4.35: Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and age (years), BMI z-score, GIP (μg/g wet matter), tTG (U/ mL), PedsQL-GSS score, energy intake (%EAR) and fat intake (gr) of CD patients on GFD at least one year

	Age (years)		BMI z-score		GIP (μg/g wet matter)		tTG (U/ mL)		PedsQL-GSS score		Energy (% EAR)		Fat (g)	
Absolute concentration (μmol/g dry matter)	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*
Acetic acid	0.00	0.768	0.00	0.351	0.00	0.543	7.25	0.029	15.46	0.001	10.92	0.007	3.41	0.088
Propionic acid	0.00	0.949	0.07	0.312	0.72	0.237	0.24	0.294	7.63	0.019	6.37	0.031	1.01	0.213
Isobutyric acid	0.00	0.498	0.00	0.900	0.00	0.733	0.00	0.374	4.31	0.062	0.00	0.533	0.00	0.605
Butyric acid	0.00	0.996	2.42	0.124	0.00	0.552	5.27	0.054	6.29	0.031	7.09	0.024	2.89	0.106
Isovaleric acid	0.00	0.467	0.00	0.769	0.00	0.487	0.00	0.741	1.02	0.211	0.00	0.907	0.00	0.790
Valeric acid	0.00	0.444	0.00	0.618	0.00	0.471	1.83	0.167	0.95	0.217	0.15	0.302	0.00	0.522
Isocaproic acid	0.00	0.832	0.00	0.481	0.00	0.531	5.20	0.055	3.44	0.085	1.31	0.191	0.76	0.236
Caproic acid	3.91	0.070	0.00	0.335	0.00	0.790	0.00	0.824	0.00	0.458	0.00	0.664	0.00	0.998
Heptanoic acid	5.83	0.035	0.00	0.334	0.00	0.487	0.00	0.517	0.51	0.260	0.00	0.614	0.00	0.809
Octanoic acid	0.00	0.332	0.00	0.772	0.00	0.400	0.00	0.570	3.72	0.077	0.00	0.526	0.00	0.366
Total SCFA	0.00	0.795	0.38	0.274	0.00	0.972	6.10	0.041	13.35	0.003	10.31	0.008	3.26	0.093
Absolute concentration (μmol/g wet matter)														
Acetic acid	0.00	0.997	5.22	0.048	0.00	0.475	6.51	0.039	5.26	0.047	1.31	0.194	0.00	0.556
Propionic acid	0.00	0.770	1.24	0.197	0.47	0.265	0.00	0.470	2.72	0.115	0.74	0.240	0.00	0.690
Isobutyric acid	0.00	0.955	0.36	0.278	0.00	0.537	0.00	0.968	0.00	0.440	0.00	0.683	0.00	0.906
Butyric acid	0.00	0.990	3.82	0.078	0.00	0.523	4.59	0.071	2.98	0.105	2.43	0.130	0.41	0.273
Isovaleric acid	0.00	0.982	0.73	0.240	0.00	0.435	0.00	0.636	0.00	0.692	0.00	0.488	0.00	0.853
Valeric acid	0.00	0.596	0.00	0.335	0.93	0.221	0.00	0.577	0.00	0.858	0.00	0.967	0.00	0.979
Isocaproic acid	0.00	0.828	0.92	0.223	0.00	0.414	6.37	0.041	2.16	0.141	0.25	0.291	0.45	0.269

Caproic acid	2.15	0.139	0.00	0.570	0.00	0.933	0.00	0.594	0.00	0.776	0.00	0.927	0.00	0.933
Heptanoic acid	4.71	0.056	0.00	0.599	0.00	0.488	0.00	0.525	0.27	0.288	0.00	0.803	0.00	0.826
Octanoic acid	0.00	0.370	0.00	0.698	0.00	0.434	0.00	0.915	5.19	0.049	0.00	0.444	0.11	0.308
Total SCFA	0.00	0.921	4.86	0.055	0.00	0.972	4.51	0.073	4.57	0.060	1.46	0.183	0.00	0.506
Relative abundance (%)														
Acetic acid	0.00	0.666	0.05	0.314	13.75	0.002	0.00	0.552	0.00	0.379	0.00	0.537	0.00	0.654
Propionic acid	0.00	0.822	0.00	0.722	12.90	0.003	0.00	0.977	2.13	0.138	0.00	0.633	0.00	0.991
Isobutyric acid	0.00	0.682	0.00	0.800	0.00	0.433	0.00	0.351	0.05	0.315	5.98	0.036	0.68	0.244
Butyric acid	0.00	0.928	2.63	0.115	3.95	0.069	2.16	0.149	1.07	0.207	2.80	0.110	1.76	0.161
Isovaleric acid	0.00	0.698	0.00	0.961	0.00	0.415	0.95	0.227	1.31	0.189	6.08	0.035	0.47	0.264
Valeric acid	0.00	0.455	0.00	0.930	3.47	0.083	0.00	0.649	4.57	0.057	0.86	0.227	0.00	0.479
Isocaproic acid	0.00	0.495	0.00	0.386	0.00	0.367	0.00	0.368	0.00	0.505	0.38	0.274	1.10	0.206
Caproic acid	0.16	0.299	0.00	0.571	0.00	0.968	0.00	0.489	0.45	0.266	0.00	0.408	0.00	0.606
Heptanoic acid	4.06	0.067	0.00	0.927	0.00	0.416	0.00	0.445	0.00	0.566	0.00	0.924	0.00	0.805
Octanoic acid	0.00	0.393	0.00	0.674	0.00	0.433	0.00	0.915	6.69	0.027	0.00	0.347	0.46	0.266

Table 4.36: Univariate regression analysis between the absolute concentration (μmol/g dry & wet matter)/ relative abundance (%) of SCFA and dietary intake of fat (%TDEI), protein (g, %TDEI, % RNI), carbohydrates (g, %TDEI) and sugars (%TDEI) of CD patients on GFD at least one year

	Fat (%TDEI)		Protein (g)		Protein (%TDEI)		Protein (% RNI)		Carbohydrates (g)		Carbohydrates (% TDEI)		Sugars (% TDEI)	
Absolute concentration (μmol/g dry matter)	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*
Acetic acid	0.00	0.481	1.22	0.197	0.00	0.902	1.09	0.207	3.27	0.093	0.00	0.888	0.00	0.861
Propionic acid	0.00	0.360	0.00	0.452	0.00	0.585	0.00	0.331	1.04	0.211	0.00	0.783	0.60	0.251
Isobutyric acid	0.00	0.366	0.00	0.725	0.00	0.655	0.00	0.688	0.00	0.583	0.23	0.292	0.00	0.825
Butyric acid	0.00	0.389	1.04	0.211	0.00	0.688	0.71	0.240	4.03	0.071	0.00	0.810	0.00	0.356
Isovaleric acid	1.00	0.215	0.00	0.619	0.00	0.787	0.00	0.837	0.00	0.344	0.42	0.270	0.00	0.719
Valeric acid	0.00	0.918	0.00	0.731	0.00	0.621	0.43	0.269	0.00	0.736	0.00	0.695	0.00	0.752
Isocaproic acid	0.00	0.418	0.00	0.837	0.00	0.713	0.00	0.604	0.00	0.853	0.00	0.362	0.00	0.655
Caproic acid	0.00	0.855	0.00	0.869	0.00	0.674	3.21	0.095	0.00	0.932	0.00	0.770	0.00	0.393
Heptanoic acid	0.00	0.966	0.00	0.889	0.00	0.664	2.49	0.123	0.00	0.940	0.00	0.960	0.00	0.396
Octanoic acid	0.00	0.479	0.00	0.362	0.00	0.757	0.00	0.959	0.00	0.759	0.00	0.678	0.00	0.979
Total SCFA	0.00	0.443	0.89	0.224	0.00	0.778	1.15	0.202	3.18	0.096	0.00	0.896	0.00	0.611
Absolute concentration (μmol/g wet matter)														
Acetic acid	0.00	0.545	0.00	0.528	0.00	0.453	0.00	0.916	0.50	0.264	0.00	0.409	0.00	0.538
Propionic acid	0.00	0.393	0.00	0.957	0.00	0.773	0.00	0.667	0.00	0.630	0.00	0.935	0.67	0.247
Isobutyric acid	0.00	0.363	0.00	0.595	0.00	0.865	0.00	0.613	0.00	0.426	0.00	0.768	0.00	0.794
Butyric acid	0.00	0.496	0.00	0.330	0.00	0.895	0.00	0.418	1.44	0.185	0.00	0.691	0.00	0.374
Isovaleric acid	0.79	0.236	0.00	0.632	0.00	0.737	0.00	0.583	0.00	0.333	0.00	0.744	0.00	0.736
Valeric acid	0.00	0.784	0.00	0.999	0.00	0.923	0.00	0.730	0.00	0.891	0.00	0.894	0.00	0.629
Isocaproic acid	0.00	0.337	0.00	0.883	0.00	0.776	0.00	0.680	0.00	0.978	0.00	0.371	0.00	0.760

Caproic acid	0.00	0.793	0.00	0.768	0.00	0.961	1.00	0.218	0.00	0.991	0.00	0.965	0.00	0.503
Heptanoic acid	0.00	0.894	0.00	0.951	0.00	0.736	1.07	0.212	0.00	0.896	0.00	0.881	0.00	0.373
Octanoic acid	0.00	0.418	0.00	0.345	0.00	0.785	0.00	0.911	0.00	0.753	0.00	0.592	0.00	0.983
Total SCFA	0.00	0.524	0.00	0.555	0.00	0.686	0.00	0.858	0.17	0.300	0.00	0.554	0.00	0.418
Relative abundance (%)														
Acetic acid	0.00	0.819	0.00	0.961	0.00	0.546	0.00	0.475	0.00	0.908	0.00	0.647	0.76	0.235
Propionic acid	0.00	0.542	0.00	0.695	0.00	0.430	0.00	0.650	0.00	0.759	0.00	0.447	1.73	0.162
Isobutyric acid	3.96	0.080	3.29	0.092	0.00	0.864	0.00	0.298	6.85	0.026	0.00	0.394	0.00	0.801
Butyric acid	0.00	0.682	0.69	0.242	0.00	0.829	0.44	0.268	1.50	0.177	0.00	0.916	0.00	0.393
Isovaleric acid	3.89	0.074	1.85	0.155	0.00	0.937	0.07	0.312	5.76	0.039	0.00	0.490	0.00	0.894
Valeric acid	0.00	0.439	0.00	0.481	0.00	0.811	0.00	0.917	0.00	0.325	0.00	0.835	0.00	0.898
Isocaproic acid	1.53	0.175	0.00	0.880	0.00	0.672	0.00	0.405	0.00	0.883	1.34	0.189	0.00	0.682
Caproic acid	0.00	0.557	0.00	0.721	0.00	0.856	0.00	0.596	0.00	0.512	0.00	0.856	0.00	0.427
Heptanoic acid	0.00	0.545	0.00	0.755	0.00	0.638	0.95	0.219	0.00	0.597	0.00	0.555	1.07	0.209
Octanoic acid	0.00	0.450	0.20	0.295	0.00	0.780	0.00	0.827	0.00	0.648	0.00	0.610	0.00	0.968

Table 4.37: Univariate regression analysis between the absolute concentration (µmol/g dry & wet matter)/ relative abundance (%) of SCFA and dietary intake of starch (% TDEI), NMES (% TDEI), NSP (g, % TDEI) and dietary fibre (g), and faecal water content (%) of CD patients on GFD at least one year

	Starch (%TDEI)		NMES (%TDEI)		NSP (g)		NSP (% DRV)		Dietary fibre (g)		Faecal water content (%)	
Absolute concentration (µmol/g dry matter)	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*
Acetic acid	4.83	0.053	0.00	0.919	0.00	0.632	0.23	0.298	1.53	0.179	35.81	< 0.0001
Propionic acid	0.46	0.612	0.00	0.875	0.97	0.217	0.00	0.557	1.49	0.182	21.31	< 0.0001
Isobutyric acid	0.00	0.967	0.00	0.912	0.00	0.971	0.00	0.601	0.00	0.358	2.64	0.116
Butyric acid	3.09	0.099	0.00	0.605	0.00	0.628	0.00	0.863	3.75	0.082	14.50	0.002
Isovaleric acid	0.00	0.852	0.00	0.799	0.00	0.643	0.00	0.437	1.02	0.216	0.00	0.683
Valeric acid	0.00	0.752	0.00	0.984	0.00	0.587	0.00	0.648	0.00	0.881	0.00	0.461
Isocaproic acid	0.00	0.825	0.00	0.503	0.00	0.801	0.00	0.677	0.00	0.915	2.77	0.111
Caproic acid	0.00	0.826	0.00	0.526	0.00	0.838	1.25	0.213	0.00	0.715	0.00	0.371
Heptanoic acid	0.00	0.748	0.00	0.523	0.00	0.738	2.02	0.167	0.00	0.621	0.07	0.313
Octanoic acid	0.00	0.820	0.00	0.887	0.00	0.769	0.00	0.713	0.00	0.905	0.00	0.847
Total SCFA	3.29	0.092	0.00	0.988	0.00	0.526	0.00	0.414	1.99	0.152	31.02	< 0.0001
Absolute concentration (µmol/g wet matter)												
Acetic acid	8.72	0.015	0.00	0.944	0.00	0.457	0.00	0.845	0.00	0.773	0.00	0.810
Propionic acid	0.00	0.573	0.00	0.954	0.00	0.404	0.00	0.775	0.00	0.587	0.00	0.439
Isobutyric acid	0.00	0.785	0.00	0.919	0.00	0.635	1.15	0.221	1.97	0.156	9.39	0.011
Butyric acid	3.59	0.087	0.00	0.371	0.00	0.906	0.00	0.693	0.84	0.234	0.00	0.448
Isovaleric acid	0.00	0.603	0.00	0.674	0.00	0.475	1.37	0.207	2.60	0.126	16.23	0.001
Valeric acid	0.00	0.869	0.00	0.827	0.00	0.978	0.00	0.688	0.00	0.654	7.56	0.021
Isocaproic acid	0.00	0.805	0.00	0.472	0.00	0.897	0.00	0.875	0.00	0.838	0.00	0.664

Caproic acid	0.00	0.696	0.00	0.403	0.00	0.889	0.00	0.587	0.00	0.664	0.00	0.496
Heptanoic acid	0.00	0.823	0.00	0.512	0.00	0.987	0.00	0.341	0.00	0.530	0.00	0.787
Octanoic acid	0.00	0.791	0.00	0.972	0.00	0.787	0.00	0.718	0.00	0.947	0.00	0.532
Total SCFA	5.36	0.048	0.00	0.780	0.00	0.745	0.00	0.932	0.00	0.647	0.00	0.791
Relative abundance (%)												
Acetic acid	0.00	0.810	0.00	0.942	0.45	0.267	0.00	0.385	0.00	0.422	0.00	0.654
Propionic acid	0.00	0.395	0.00	0.393	3.48	0.086	0.00	0.847	0.00	0.627	2.47	0.123
Isobutyric acid	0.00	0.326	0.00	0.881	0.98	0.216	0.00	0.402	6.62	0.031	16.13	0.001
Butyric acid	0.00	0.424	0.00	0.573	0.00	0.422	0.04	0.318	3.10	0.103	0.20	0.296
Isovaleric acid	0.00	0.323	0.00	0.725	2.34	0.130	0.00	0.411	5.46	0.046	24.89	< 0.0001
Valeric acid	0.00	0.555	0.00	0.908	0.00	0.762	0.00	0.881	0.00	0.652	16.20	0.001
Isocaproic acid	0.00	0.417	0.00	0.569	0.00	0.591	0.00	0.985	0.00	0.987	0.00	0.850
Caproic acid	0.00	0.619	0.00	0.592	0.00	0.720	0.00	0.480	0.00	0.606	1.38	0.186
Heptanoic acid	0.00	0.850	0.00	0.612	0.00	0.591	0.28	0.293	0.00	0.515	0.00	0.798
Octanoic acid	0.00	0.790	0.00	0.993	0.00	0.720	0.00	0.766	0.00	0.969	0.00	0.471

Table 4.38: Predictors of SCFA concentration (µmol/g) in dry matter of CD patients on GFD at least one year (n=60), using multiple linear regression analysis (Beta Coefficient, *p-value*)

	Acetic acid (µmol/g)	Propionic acid (µmol/g)	Butyric acid (µmol/g)	Isocaproic acid (µmol/g)	Heptanoic acid (µmol/g)	Total SCFA (µmol/g)
Constant coefficient (B_o)	B_o= -523	B_o= -142.1	B_o= -145.1	B_o= 0.47	B_o= -0.05	B_o= -847
Age	-	-	-	-	B= 0.02 *p*= 0.035	-
Gender = Female	B= -82.7 *p*= 0.045	-	-	B= -0.26 *p*= 0.21	-	-
Starch (%TDEI)	B= 9.6 *p*= 0.049		B= 3.97 *p*= 0.049			-
Faecal water content (%)	B = 13.5 *p* < 0.0001	B = 3.18 *p* < 0.0001	B= 2.79 *p* = 0.008	-	-	B= 20.9 *p* < 0.0001
***p* – model**	< 0.0001	< 0.0001	0.003	0.021	0.035	< 0.0001
R^2 – model	45.6 %	26.8 %	19.5 %	8.4 %	5.8 %	30.8 %

Table 4.39: Predictors of SCFA concentration (µmol/g) in wet matter of CD patients on GFD at least one year (n=58), using multiple linear regression analysis (Beta Coefficient, *p-value*)

	Acetic acid (µmol/g)	Propionic acid (µmol/g)	Isobutyric acid (µmol/g)	Butyric acid (µmol/g)	Isovaleric acid (µmol/g)	Valeric acid (µmol/g)	Isocaproic acid (µmol/g)	Octanoic acid (µmol/g)	Total SCFA (µmol/g)
Constant coefficient	B_o= 128.7	B_o= 27.5	B_o= 7.6	B_o= 17.3	B_o= 9.50	B_o= 6.6	B_o= 0.13	B_o= 1.29	B_o= 160.6
Gender = Female	B= -25.9 *p*= 0.029	B= -8.1 *p*= 0.016	-	-	-	-	B= -0.06 *p*= 0.037	-	-
tTG (U/ mL)	B= 0.703 *p*= 0.029	-	-	B= 0.239 *p*= 0.037	-	-	-	-	B= 1.13 *p* = 0.038
PedsQL-GSS score	-	-	-	-	-	-	-	B= -0.014 *p*= 0.049	-
Faecal water content (%)	-	-	B= -0.07 *p*= 0.011	-	B= - 0.10 *p*= 0.001	B = -0.06 *p* = 0.021	-	-	-
***p* – model**	0.005	0.016	0.011	0.037	0.001	0.021	0.037	0.049	0.038
R^2 – model	17.8 %	8.2 %	9.4 %	7.1 %	16.2 %	7.6%	6.7 %	5.2 %	7.2 %

Table 4.40: Predictors of the relative abundance of SCFA (%) of CD patients on GFD at least one year (n=58), using multiple linear regression analysis (Beta Coefficient, *p-value*)

	Acetic acid (%)	Propionic acid (%)	Isobutyric acid (%)	Isovaleric acid (%)	Valeric acid (%)	Caproic acid (%)	Octanoic acid (%)
Constant coefficient	B_o= 72.4	B_o= 15.3	B_o = 5.16	B_o= 6.51	B_o= 3.72	B_o = 0.20	B_o= 1.04
Gender = Female	-	B= -2.11 *p* = 0.029	-	-	-	B = 0.19 *p* = 0.011	-
Fat (%TDEI)	-	-	B= -0.027 *p*= 0.037	B= - 0.03 *p*= 0.032	-	-	-
NSP (g)	-	B = - 0.40 *p* = 0.028	-	-	-	-	-
GIP (µg/g wet matter)	B = -5.47 *p* = 0.002	B = 2.85 *p* = 0.001	-	-	-	-	-
PedsQL-GSS score	-	-	-	-	-	-	B= -0.011 *p*= 0.027
Faecal water content (%)	-	-	B= -0.037 *p*< 0.0001	B= -0.05 *p*< 0.0001	B= -0.033 *p*= 0.001	-	-
***p* – model**	0.002	< 0.0001	0.001	< 0.0001	0.001	0.011	0.027
R^2 – model	13.8 %	25.7 %	22.2 %	29.5 %	16.3 %	9.1 %	6.7%

CHAPTER 5 General Discussion and Conclusions

This study characterised the faecal microbiota composition and metabolic activity of children with CD before and during treatment with GFD and compared this with healthy controls and siblings of the patients. This is the first study to combine both observational and prospective designs and to use NGS in faeces of such patients. The prospective cohort of CD children was followed from diagnosis (UCD) to six and 12 months on GFD. Clinical, dietary and microbiological outcomes were evaluated, as well as patients' compliance with GFD. Compliance was examined both directly, measuring GIP in faeces, and indirectly, evaluating GI symptoms, serology tTG and self-reported Biagi score. Gut microbiota composition was characterised using 16S rRNA sequencing, while targeted faecal metabolites, SCFA, were measured using gas chromatography.

CD is an immune-based enteropathy elicited by gluten ingestion in genetically susceptible individuals. Although the pathologic process involved in the development of CD has yet to be fully elucidated, we know that once the patient is exposed to gluten, an inflammatory cascade is induced in the small intestinal mucosa that will lead to villous atrophy, crypt hyperplasia and an increase in the number of lymphocytes in the lamina propria [51]. Although the HLA-DQ genes explain approximately 35% of the risk for CD, only a small percentage of genetically susceptible individuals develop the disease. This in conjunction with the fact that introduction of gluten into the diet occurs in early childhood, but CD may develop at any point during lifetime implies that other environmental and genetic factors involved in the onset of CD have yet to be found. Indeed, Kalliomaki et al (2012) demonstrated a potential role for a disturbed host-microbiota interaction in the pathogenesis of CD, based on a significant decrease in the Toll intracellular protein mRNA levels observed in duodenal tissues of UCD compared with HC [112]. The Toll intracellular protein inhibits the Toll-like receptor bacteria-induced signalling, and dysregulation of the later has been implicated in chronic epithelial responses (i.e. IBD) [112]. Moreover, De Palma et al (2010) suggested a possible relationship between the gut microbiota and the host defences in CD, based on a significant decrease in the levels of IgA-coated bacteria observed in CD patients compared with HC [43]. However, it is still unclear whether the microbiota "dysbiosis" is involved in the development of CD or is a secondary effect of the inflammatory active phase of CD and the impact of treatment with GFD.

During the active phase of CD, ingestion of gluten results in activation of an inflammatory cascade characterised, as described above, by villous atrophy and crypt hyperplasia [51]. In this process, epithelial cell turnover is increased, with a greater load of endogenous mucus and sloughed epithelial cells reaching the colon in UCD than HC. Moreover, because of the villous atrophy, the small intestine of UCD has impaired epithelial barrier function that leads to nutrient malabsorption. As a result, more macro- and micronutrients escape absorption in the upper GI tract of UCD than HC, and reach the large intestine where they are available for bacterial fermentation. Hence, substrate (endogenous and exogenous) availability to the gut microbiota is expected to be higher in UCD than

HC. On the other hand, UCD is usually characterised by a faster GI transit time that along with severe diarrhoea may limit bacterial fermentation of the available substrate in the proximal and distal colon [58]. Therefore, research is needed to explore the true effect of mucosal damage, GI transit time and diarrhoea on the UCD microbiota, especially when these are accompanied by modifications in the diet prior to initiation of GFD because of severe GI symptoms that are often associated with dietary exclusion.

Patients who suffer from persistent GI symptoms, such as diarrhoea and abdominal pain, may modify their diet eliminating gluten prior to diagnosis, although it is strictly recommended not to remove gluten before diagnosis. Once the patient is diagnosed with CD, then lifelong adherence to GFD is currently the only proven effective treatment [113]. Transition to GFD is associated with a decrease in the dietary intake of a large group of non-digestible carbohydrates that are fermented in the large intestine by the gut microbiota to yield energy for bacterial growth [51]. SCFA are the main products of this fermentation, which may act as an energy source and inflammation modulator [114]. Transition to GFD creates social and economic burden and is quite challenging, considering the presence of gluten, even in trace amounts, in most food products. The median time required to achieve normal villous height on GFD is 3.8 years [50]. However, normalization of serum tTG levels usually takes up to one year, suggesting that resolution of inflammation is achieved by the first year of GFD [51]. Thus, decreased dietary non-digestible polysaccharides seen with GFD, resolution of gut inflammation, and changes in gut motility and improvement in GI symptoms may all modify the TCD microbiota.

The evidence of studies looking at the gut microbiota of CD so far is controversial and inconclusive. Collado et al (2008) demonstrated that TCD are characterised by an increase in *B. adolescentis* compared with HC and UCD [41]. Golfetto et al (2014) reported that faecal samples of TCD children are characterised by an increase in *Bifidobacterium*, however no UCD children were recruited in this study to evaluate the effect of GFD [48]. On the other hand, Nistal et al (2012) observed a decrease in the diversity of *Bifidobacteria* in TCD compared with HC, along with a decrease in the levels of *B. catenulatum* in TCD compared with UCD [37]. Similar to the evidence of studies looking at the gut microbiota community, evidence of studies looking at the microbiota metabolic activity is inconsistent. Most of studies have suggested an increase in SCFA of CD patients regardless of disease activity [37, 83, 84], but not all studies support this [39].

In our study, the relative abundance of SCFA in CD patients on GFD was significantly different compared with HC. The relative abundance of acetic acid was significantly higher, while propionic, butyric and valeric acid levels were significantly lower in TCD than HC. These findings, in conjunction with the lack of difference in the concentration of SCFA between UCD and HC, suggest that this difference in faecal SCFA levels is a secondary effect of GFD adherence. The prospective study with paired sample collection confirmed the observations from the cross-sectional study. Compared with baseline, the relative abundance of acetic was significantly higher, while the relative abundance of butyric was significantly lower at six months on GFD. Similarly, when we looked at the faecal microbial

community structure, there was no difference between UCD and HC groups. TCD patients clustered separately from HC in the NMDS plot for both the Bray-Curtis dissimilarity index, and unweighted Unifrac distance analysis, while they tended to cluster separately from UCD. Therefore, it can be concluded that it is treatment with GFD that drives faecal microbiota community structure changes and similarly the changes seen in faecal metabolites.

Changes in the metabolic activity of the gut microbiota are broadly reflected by changes in faecal pH. In accordance with previous findings by Golfetto et al (2014) [48] and Kopecny et al (2008)[88], in the current study, there was no significant difference in the faecal pH of UCD, TCD and HC children. Although changes were seen in SCFA pattern, these did not affect luminal pH, as relative abundance of some SCFA decreased (e.g. butyric) and other increased (e.g. acetic), counterbalancing the effect on pH.

L-lactate was significantly higher in UCD than HC and TCD groups, an effect that was independent of faecal water content. In contrast, D-lactate had significantly lower concentration in wet matter of faeces of UCD children than HC. Lactate is an intermediate bacterial product that is further metabolised to SCFA. In the current study, no significant differences in dietary intake of UCD and HC were observed. However, the estimated intake of NSP was significantly higher in UCD than TCD children. This may explain the difference in faecal L-lactate concentration between UCD and TCD, but the difference between UCD and HC remains to be explained. Therefore, the fast GI transit time, and diarrhoea that differentiate UCD children from HC need to be considered; 20% of UCD reported frequent diarrhoea in the PedsQL-GSS questionnaire, while no HC reported such symptoms. Fast GI transit time may limit SCFA and lactate production and conversion to other SCFA in UCD children compared to HC. In the current study, D-lactate, which is mainly produced via bacterial fermentation in the colon, was significantly lower in UCD than HC. In contrast, L-lactate was significantly higher in UCD than HC. This means that a fast GI transit time may explain differences in the mainly bacterially produced D-lactate between UCD and HC, but it does not explain differences in L-lactate. Instead, an increased L-lactate production in mammalian cells, counterbalancing the decrease in bacterial L-lactate production could potentially explain the significantly increased levels of L-lactate in UCD than HC. However, there has been no research on the mammalian production of L-lactate to support such a hypothesis. Therefore, further research is needed to examine the increase in L-lactate levels of faeces of UCD patients and then explore the effect of mucosal damage, GI transit time, diarrhoea, and respective diet modifications, along with production in mammalian cells.

Unlike faecal lactate, there was no significant difference in faecal SCFA between UCD and TCD children, nor between UCD and HC. Acetic, isobutyric and isovaleric tended to differ between UCD and TCD children, but this difference did not reach significance. However, SCFA concentration was significantly different between TCD and HC. Propionic and butyric acid concentrations per total faecal output were significantly lower in TCD (approximately -22%) than HC. There was also significantly lower molar proportion of butyric acid in TCD (-16%) than HC. The prospective study confirmed that faecal butyric decreased after compliance with GFD. Compared to the baseline values, both the concentration per dry matter and the relative abundance of butyric acid significantly decreased at six

months on GFD (-34% and -29% respectively). These findings are consistent with previous research regarding faecal SCFA in children with CD by Di Cagno et al (2011) where TCD had significantly lower concentrations of propionic and butyric acid than HC [39]. However, they disagree with findings by Tjellstrom et al (2005) who found higher propionic acid concentration in faecal samples of TCD children than HC [84]. The difference in treatment duration of CD children may explain differences in the levels of SCFA; in the present study, TCD were on GFD for at least one year, whereas in the study by Tjellstrom et al (2005) TCD were on GFD for at least three months. Our findings disagree with findings by Nistal et al (2012) as well, who found higher concentrations of both propionic and butyric acid in TCD than HC [37], and this difference may be due to difference in the age of patients. We measured SCFA in samples of children with CD, whereas Nistal et al (2012) measured SCFA in samples of adults with CD.

The decreased dietary intake of NSP of TCD compared with HC may explain the decrease in butyric and propionic levels of the former. However, there was no significant association between the intake of NSP (g and %DRV) and the concentrations of butyric (p= 0,774 and 0,407 respectively) and propionic acid per total faecal output (p= 0,716 and 0,490 respectively), nor between the intake of NSP (g and %DRV) and the relative abundance of butyric acid (p= 0,380 and 0,255 respectively). Therefore, if the lower concentration of butyric acid observed in TCD children was to be explained by alterations in the composition of gut microbiota, a decrease in the relative abundance of butyrate-producing bacteria, such as *Faecalibacterium prausnitzii, Anaerostipes, Eubacterium* and *Roseburia* species [115], would be expected in TCD children. Indeed, in the current study, the faeces of TCD had significantly lower relative abundance of OTU_908 *Anaerostipes* than HC. However, there was no significant correlation between the relative abundance of OTU_908 *Anaerostipes* and butyric levels, which might be due to lack of statistical power (i.e. not enough faecal samples of TCD children with non-zero relative abundance of the OTU_908 to correlate significantly with butyric levels), but it is still not clear and further research is needed to clarify this.

Moreover, the relative abundances of OTU_893 *Eubacterium hallii* group, OTU_1044 *Eubacterium hallii* group and all OTU belonging widely to *Eubacterium hallii*, an important butyrate-producing genus, were significantly lower at six months on GFD than at diagnosis, but there was no difference between the independent TCD and UCD groups. This may be due to the inter-individual variability in microbiota composition seen in distinct subjects in our cross-sectional study, a problem that we addressed in the prospective study by collecting serial samples from the same CD children at diagnosis, and at six and 12 months after the initiation of GFD. A recent study by Engels et al (2016) showed that the role of *Eubacterium hallii* is very important in the total microbiota metabolic activity not only because of its involvement in glucose fermentation to butyric acid and hydrogen, but also because of its ability to convert glycerol to 3-hydroxypropionaldehyde and produce propionic [116]. Although the relative abundance of *Eubacterium hallii* was not significantly associated with the faecal concentration of butyric, nor with the faecal concentration of GIP, at six months on GFD both *Eubacterium hallii* and butyric levels were significantly decreased from diagnosis. Further research is

needed to evaluate the effect of GFD adherence, examined by measuring GIP in faeces, on the butyrate-producing bacteria and butyric acid levels.

The relative abundance of acetic acid was significantly higher in TCD than HC. This is in agreement with findings by Nistal et al (2012) [37], Di Cagno et al (2011) [39] and Tjellstrom et al (2005)[84], but in disagreement with findings by Tjellstrom et al (2013) who reported no difference in acetic between TCD on GFD for more than one year and HC [83]. Differences in the findings may be due to differences in the characteristics of the responding populations, compliance with GFD and overall dietary habits of the TCD individuals studied. Tjellstrom et al (2013) did not provide any data regarding their diet or treatment adherence, and hence we cannot compare the TCD groups apart from their mean age. Tjellstrom et al (2013) measured microbiota metabolites in faeces of adults that were on GFD for more than one year (median: 4 years), whereas in the current study only samples of children were collected and analysed. Likewise, results may disagree because of differences in the microbiota structure and metabolic activity seen during aging, regardless of the onset of CD and treatment with GFD.

Nevertheless, it should be noted that, unlike findings by Tjellstrom et al (2013), in the current study, faecal SCFA of TCD patients did not return to HC levels. Instead, acetic was significantly higher in TCD than HC, and the GFD adherence, examined by faecal GIP concentration, was a significant predictor of the relative molar concentration of acetic of faeces of CD children who were on GFD for at least one year. Decreased GIP levels, hence increased compliance GFD, were strongly associated with increased relative abundance of acetic acid, suggesting that an increase in acetic acid levels is a secondary consequence of treatment with GFD, rather than an effect of the onset of CD. However, the exact mechanisms through which GFD drives acetic acid increase, and butyric and propionic acid decrease have yet to be found. These findings could be used to unravel the mechanisms through which GFD exhibits its beneficial effect except for excluding the environmental trigger of the inflammatory response of CD.

Unlike previous findings [39, 84], no significant difference in faecal SCFA levels was observed between UCD and HC groups. This is in accordance with our results of microbiota structure, where no difference between UCD children and HC was observed. However, they do not support the hypothesis for an altered UCD microbiota structure and functionality, suggested because of the villous atrophy, consequent malabsorption and increased delivery of fermentable substrate into the colon of UCD patients. Further research is needed to clarify this and determine whether differences in the microbiota of CD patients are a secondary consequence of treatment with GFD, as seen in the present study, or a genuine phenomenon of CD not affected by either the treatment with GFD, or the disease activity as proposed by Tjellstrom et al (2005) [84].

Another interesting finding of our study was the significantly higher concentration of free faecal H_2S (deemed to be a toxic compound) in HC than UCD children. Considering the demonstrated beneficial effects of H_2S on the GI tract, liver [79], and circulatory system homeostasis [73], and the lack of studies looking at faecal sulphide of CD, our results raise the question of whether increased H_2S

levels seen in HC protect against the development of CD, or the onset of CD diminishes the production of H_2S (observed in UCD). Our data suggest the latter. The relative abundance of OTU belonging widely to *Euryarcheota*, one of the seven main phylogenetic lineages of SRB [117], was decreased in faeces of UCD children compared with HC, potentially explaining the respective decrease seen in H_2S concentration. However, cross-sectional analysis does not allow establishment of causality, and hence further research of the role of H_2S in CD is needed.

Overall, our findings demonstrate that the gut microbiota community structure and functional capacity changes after treatment with GFD, with no difference between UCD and HC. The relative abundance of 31 identified OTU was significantly different between the UCD and HC groups, with all 31 OTU having significantly lower relative abundance in UCD than HC. OTU belonging widely to nine distinct genera (*Senegalimassilia, Methanobrevibacter, Sutterella, Alloprevotella, Slackia, Ruminococcaceae UCG-014, Parabacteroides, Holdemanella* and *Prevotella* 9) had significantly lower relative abundance in faeces of UCD than HC. Previous research has not reported alterations in these specific taxa, which may be explained by limitations in the methods used to capture differences in different levels of the microbial hierarchy. Only Collado et al (2007) found a significant difference in *Prevotella* between UCD and HC, but the microbial composition was characterised using the FISH method, with UCD having more log cells per g of faecal sample [44]. Multiple factors, (i.e. fast GI transit time, modifications in the diet because of severe GI symptoms, malabsorption) may explain differences in the relative abundance of the 31 OTU, but a distinct microbiota "dysbiosis" similar to IBD has not been identified in the present study. However, the use of NGS to characterise the gut microbiome of patients with CD, and healthy subjects, in the future, could examine whether differences in these 31 OTU are replicated in a larger CD cohort, explore the metabolic pathways regulated by these OTU and potentially increase our understanding of the role of these bacteria in the development of CD.

The effect of GFD on the microbial composition of CD children was evaluated comparing the 51 OTU, which had significantly different relative abundance between UCD and TCD, with the OTU that had significantly different relative abundance in faeces of the CD prospective cohort at six and 12 months on GFD compared to the baseline. OTU_259 *Erysipelatoclostridium* and OTU_98 *Ruminococcaceae* had significantly lower relative abundance at six months on GFD than at diagnosis, but both OTU had significantly higher relative abundance in faeces of TCD than UCD independent groups. This may be due to differences in treatment duration or compliance with GFD. Based on the findings of our prospective study, CD children tend to comply with the diet more strictly at six months after the initiation of GFD than at 12 months. However, the level of adherence was not associated with the relative abundance of the OTU_259 and OTU_98, as the reported relative abundance of both OTU was replicated when those patients who did not comply strictly to a GFD and had faecal GIP levels higher than 0.156 μg per g wet matter were excluded from the analysis. Therefore, it is either the different treatment duration, along with the consequent disease improvement, between the CD cohort (at six months) and the TCD group (on GFD for at least one year), or probable inter-individual

variability in the microbiota complexity of distinct UCD and TCD subjects of the cross-sectional study that causes this difference in the findings of our prospective and cross-sectional studies.

In terms of treatment duration, it should be noted that resolution of inflammation is achieved only after one year of GFD. As ongoing disease may influence the microbiota community, the microbiota of CD children at six months on GFD is not the same as those on the diet for one year. Further prospective cohort studies are needed, to clarify the effect of treatment with GFD and disease improvement on the microbiota community using NGS technologies in a long-term follow-up. For instance, it would be interesting to follow the CD cohort of the present study at two or three years on GFD, examine the adherence to treatment with GFD by measuring GIP in faeces, compare their microbiota characteristics and explore the correlation between the long-term adherence to GFD and the microbiota community.

Finally, although it was not a primary outcome, this study evaluated the agreement between different methods used to assess compliance with GFD, including the self-reported Biagi score, clinicians' reports, serum tTG levels and a novel faecal biomarker, termed GIP. The GIP is a 33-mer peptide in the molecule of α-gliadin. It is stable against breakdown by all gastric, pancreatic and intestinal brush border membrane endoproteases [112], and hence GIP detection in a faecal sample confirms gluten consumption and poor compliance with GFD. The Kappa agreement between the methods was poor regardless of the chosen comparison. Although most of the CD children identified as compliant by GIP were also confirmed by all three traditional methods, only 16.7% of the non-compliant CD children by faecal GIP method, were correctly identified as non-compliant by the Biagi score, clinicians' evaluation and serum tTG analysis, revealing important limitations of the traditional methods compared to faecal GIP. Measuring GIP in faeces is an easy, precise, non-invasive method which would require only a faecal sample, and so could be used to facilitate disease improvement.

5.1 Study strengths and limitations

The double design of the present study with sample collection at baseline, as well as at six and 12 months on GFD was its major strength. It is the first study of this kind. The decrease in levels of butyric acid after treatment with GFD was evident not only when we compared the independent UCD and TCD groups, but also when we compared the CD cohort at baseline and at six months on GFD. However, it was mostly the results regarding the gut microbiota metabolic activity that were reproduced in both the cross-sectional and prospective study. Based on the clear separation of the TCD microbiota from UCD and HC, we would anticipate a gradual clustering of the CD microbiota at six and 12 months on GFD compared to the diagnosis. Instead, there was no separation, regardless of the examined taxonomic level or the index used. This could be explained by the smaller sample size of the prospective study. More prospective studies are needed, with bigger cohort size to examine the microbiota structure of patients with CD during treatment with GFD.

A strength of this study was the use of the novel faecal biomarker GIP and the NGS technique 16S rRNA sequencing. NGS technologies provide the most powerful data to characterise the whole

microbiome, allowing for phylogenetic identification and quantification of both known and unknown species, while the faecal biomarker GIP assesses accurately the adherence to treatment with GFD, and was an important aspect of this study. One limitation of this study was the lack of age matching in the cross-sectional sub-study, with HC being (mean) 2.3 years younger than UCD and 1.5 years younger than TCD. It may be that assessment only twice in a year, given the great variability in the dietary intake and the volume of the faecal samples is not enough to characterise the microbiome. Tjellstrom et al. (2014) suggested recently that based on their experience the relative proportion of SCFA varies slowly, so less frequent sampling is probably adequate [86]. However, this is just an opinion, and the need for actual evidence of studies assessing more often the microbiota of subjects under specific exclusive dietary treatment is emerge. Moreover, it could be questioned whether the gut microbial composition and functionality observed in faecal samples of CD children is representative of the gut microbiota in the colon. Collado et al (2009) explored the gut microbiota of CD patients in both faecal and duodenal biopsies, and suggested that since similar bacterial taxa are related to CD in both the duodenal and faecal tissues, the latter could be an acceptable biological index in the case of CD [42]. However, one study is not enough to confirm such a hypothesis, and further research is needed to clarify this.

Other limitations may raise during the DNA extraction process considering that increasing evidence has been reporting bacterial DNA contamination of the extraction reagents [118]. If the isolation of the genomic DNA is not done early enough, you may lose some bacteria and get a disturbed view of the microbiome, which may be further disturbed by an inappropriate bioinformatic approach. The bioinformatic analysis of the present thesis required an intense training. Comparing the microbiota community of the paired data was a challenge that the researcher had to account for the fact that these samples were collected from the same patient. Moreover, when correlating the relative abundance of distinct species with all metadata that differentiated TCD to UCD or HC children, correction for multiple statistic testing was necessary making the interpretation of the 16S rRNA sequencing a challenging task with increased practicalities.

5.2 Conclusion

In conclusion, the findings of this study supported our hypothesis that the altered gut microbiota in CD is an epiphenomenon of the dietary treatment with GFD. Even though no UCD microbiota "dysbiosis" was observed, faeces of UCD children had lower relative abundance of 31 distinct OTU compared to HC. This finding could be used to increase our understanding of the role of the gut microbiota in the development of CD and potentially our knowledge of the multifactorial aetiology of the disease, but replication of these alterations in larger cohorts of CD patients is necessary. The altered microbiota metabolic activity of children with CD after the treatment with GFD could be used to broaden our understanding of the mechanisms through which GFD exhibits its beneficial effects apart from excluding the environmental trigger of CD, and affects the health of patients with CD.

Considering that after the initiation of GFD the levels of both the beneficial butyric acid and butyrogenic bacteria decreased, concerns about the effects of GFD on the gut microbiota, GI tract and overall health of patients with CD are increased. Therefore, further studies to replicate our findings are needed. Based on our findings, an administration of butyrogenic probiotics could potentially improve in a synergistic manner with the GFD the levels of the beneficial butyric acid, and then the GI symptoms and health of patients with CD. Therefore, further studies to compare the microbiota composition of UCD to HC, and the microbiota structure and functionality of patients with CD after the initiation of GFD combined with butyrogenic probiotics are recommended.

References

1. D'Argenio, V., et al., *Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N. flavescens Strain in Duodenum of Adult Celiac Patients.* Am J Gastroenterol, 2016. **111**(6): p. 879-90.
2. Husby, S., et al., *European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(1): p. 136-60.
3. Serena, G., et al., *The Role of Gluten in Celiac Disease and Type 1 Diabetes.* Nutrients, 2015. **7**(9): p. 7143-62.
4. Koning, F., *Celiac disease: caught between a rock and a hard place.* Gastroenterology, 2005. **129**(4): p. 1294-301.
5. Moron, B., et al., *Sensitive detection of cereal fractions that are toxic to celiac disease patients by using monoclonal antibodies to a main immunogenic wheat peptide.* Am J Clin Nutr, 2008. **87**(2): p. 405-14.
6. Garampazzi, A., et al., *Clinical pattern of celiac disease is still changing.* J Pediatr Gastroenterol Nutr, 2007. **45**(5): p. 611-4.
7. Admou, B., et al., *Atypical Celiac Disease: From Recognizing to Managing.* Gastroenterol Res Pract, 2012. **2012**.
8. van Rijn, J.C., et al., *Short stature and the probability of coeliac disease, in the absence of gastrointestinal symptoms.* Arch Dis Child, 2004. **89**(9): p. 882-3.
9. Rostom, A., et al., *Celiac disease.* Evid Rep Technol Assess (Summ), 2004(104): p. 1-6.
10. Anderson, R.P., et al., *A novel serogenetic approach determines the community prevalence of celiac disease and informs improved diagnostic pathways.* BMC Med, 2013. **11**: p. 188.
11. Sarno, M., et al., *Risk factors for celiac disease.* Ital J Pediatr, 2015. **41**.
12. Mearin, M.L., et al., *HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease.* Gut, 1983. **24**(6): p. 532-7.
13. Dubois, P.C., et al., *Multiple common variants for celiac disease influencing immune gene expression.* Nat Genet, 2010. **42**(4): p. 295-302.
14. Trynka, G., C. Wijmenga, and D.A. van Heel, *A genetic perspective on coeliac disease.* Trends Mol Med, 2010. **16**(11): p. 537-50.
15. Tye-Din, J.A., et al., *Appropriate clinical use of human leukocyte antigen typing for coeliac disease: an Australasian perspective.* Intern Med J, 2015. **45**(4): p. 441-50.
16. Akobeng, A.K., et al., *Effect of breast feeding on risk of coeliac disease: a systematic review and meta-analysis of observational studies.* Arch Dis Child, 2006. **91**(1): p. 39-43.
17. Norris, J.M., et al., *Risk of celiac disease autoimmunity and timing of gluten introduction in the diet of infants at increased risk of disease.* Jama, 2005. **293**(19): p. 2343-51.
18. Vriezinga, S.L., et al., *Randomized feeding intervention in infants at high risk for celiac disease.* N Engl J Med, 2014. **371**(14): p. 1304-15.
19. Myleus, A., et al., *Early infections are associated with increased risk for celiac disease: an incident case-referent study.* BMC Pediatr, 2012. **12**: p. 194.
20. Collin, P., et al., *The safe threshold for gluten contamination in gluten-free products. Can trace amounts be accepted in the treatment of coeliac disease?* Aliment Pharmacol Ther, 2004. **19**(12): p. 1277-83.
21. Valdes, I., et al., *Innovative approach to low-level gluten determination in foods using a novel sandwich enzyme-linked immunosorbent assay protocol.* Eur J Gastroenterol Hepatol, 2003. **15**(5): p. 465-74.
22. Comino, I., et al., *Fecal Gluten Peptides Reveal Limitations of Serological Tests and Food Questionnaires for Monitoring Gluten-Free Diet in Celiac Disease Patients.* Am J Gastroenterol, 2016. **111**(10): p. 1456-1465.
23. Candela, M., et al., *Dynamic efficiency of the human intestinal microbiota.* Crit Rev Microbiol, 2015. **41**(2): p. 165-71.
24. Turnbaugh, P.J., et al., *The human microbiome project.* Nature, 2007. **449**(7164): p. 804-10.
25. Guarner, F. and J.R. Malagelada, *Gut flora in health and disease.* Lancet, 2003. **361**(9356): p. 512-9.
26. Tremaroli, V. and F. Backhed, *Functional interactions between the gut microbiota and host metabolism.* Nature, 2012. **489**(7415): p. 242-9.

27. Lupp, C., et al., *Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae.* Cell Host Microbe, 2007. **2**(2): p. 119-29.
28. Haller, D. and C. Jobin, *Interaction between resident luminal bacteria and the host: can a healthy relationship turn sour?* J Pediatr Gastroenterol Nutr, 2004. **38**(2): p. 123-36.
29. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.
30. Gill, S.R., et al., *Metagenomic analysis of the human distal gut microbiome.* Science, 2006. **312**(5778): p. 1355-9.
31. Sekirov, I., et al., *Gut microbiota in health and disease.* Physiol Rev, 2010. **90**(3): p. 859-904.
32. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* Science, 2005. **308**(5728): p. 1635-8.
33. O'Hara, A.M. and F. Shanahan, *The gut flora as a forgotten organ.* EMBO Rep, 2006. **7**(7): p. 688-93.
34. Houghteling, P.D. and W.A. Walker, *Why is initial bacterial colonization of the intestine important to the infant's and child's health?* J Pediatr Gastroenterol Nutr, 2015. **60**(3): p. 294-307.
35. Round, J.L. and S.K. Mazmanian, *The gut microbiota shapes intestinal immune responses during health and disease.* Nat Rev Immunol, 2009. **9**(5): p. 313-23.
36. Medina, M., et al., *Bifidobacterium strains suppress in vitro the pro-inflammatory milieu triggered by the large intestinal microbiota of coeliac patients.* J Inflamm (Lond), 2008. **5**: p. 19.
37. Nistal, E., et al., *Differences in faecal bacteria populations and faecal bacteria metabolism in healthy adults and celiac disease patients.* Biochimie, 2012. **94**(8): p. 1724-9.
38. Di Cagno, R., et al., *Different fecal microbiotas and volatile organic compounds in treated and untreated children with celiac disease.* Appl Environ Microbiol, 2009. **75**(12): p. 3963-71.
39. Di Cagno, R., et al., *Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization.* BMC Microbiol, 2011. **11**: p. 219.
40. Sanz, Y., et al., *Differences in faecal bacterial communities in coeliac and healthy children as detected by PCR and denaturing gradient gel electrophoresis.* FEMS Immunol Med Microbiol, 2007. **51**(3): p. 562-8.
41. Collado, M.C., et al., *Imbalances in faecal and duodenal Bifidobacterium species composition in active and non-active coeliac disease.* BMC Microbiol, 2008. **8**: p. 232.
42. Collado, M.C., et al., *Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease.* J Clin Pathol, 2009. **62**(3): p. 264-9.
43. De Palma, G., et al., *Intestinal dysbiosis and reduced immunoglobulin-coated bacteria associated with coeliac disease in children.* BMC Microbiol, 2010. **10**: p. 63.
44. Collado, M.C., M. Calabuig, and Y. Sanz, *Differences between the fecal microbiota of coeliac infants and healthy controls.* Curr Issues Intest Microbiol, 2007. **8**(1): p. 9-14.
45. Sanchez, E., et al., *Reduced diversity and increased virulence-gene carriage in intestinal enterobacteria of coeliac children.* BMC Gastroenterol, 2008. **8**: p. 50.
46. Sanchez, E., et al., *Intestinal Staphylococcus spp. and virulent features associated with coeliac disease.* J Clin Pathol, 2012. **65**(9): p. 830-4.
47. Lorenzo Pisarello, M.J., et al., *Decrease in lactobacilli in the intestinal microbiota of celiac children with a gluten-free diet, and selection of potentially probiotic strains.* Can J Microbiol, 2015. **61**(1): p. 32-7.
48. Golfetto, L., et al., *Lower Bifidobacterium counts in adult patients with celiac disease on a gluten-free diet.* Arq Gastroenterol, 2014. **51**(2): p. 139-43.
49. Olivares, M., et al., *Double-blind, randomised, placebo-controlled intervention trial to evaluate the effects of Bifidobacterium longum CECT 7347 in children with newly diagnosed coeliac disease.* Br J Nutr, 2014. **112**(1): p. 30-40.
50. Rubio-Tapia, A., et al., *Mucosal recovery and mortality in adults with celiac disease after treatment with a gluten-free diet.* Am J Gastroenterol, 2010. **105**(6): p. 1412-20.
51. Nylund, L., K. Kaukinen, and K. Lindfors, *The microbiota as a component of the celiac disease and non-celiac gluten sensitivity.* Clinical Nutrition Experimental, 2016. **6**(Supplement C): p. 17-24.
52. Fraher, M.H., P.W. O'Toole, and E.M. Quigley, *Techniques used to characterize the gut microbiota: a guide for the clinician.* Nat Rev Gastroenterol Hepatol, 2012. **9**(6): p. 312-22.
53. Chen, K. and L. Pachter, *Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities.* PLoS Comput Biol, 2005. **1**(2).

54. Naluai, A.T., et al., *Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11.* Eur J Hum Genet, 2001. **9**(12): p. 938-44.
55. Cook, S.I. and J.H. Sellin, *Review article: short chain fatty acids in health and disease.* Aliment Pharmacol Ther, 1998. **12**(6): p. 499-507.
56. Gibson, G.R., G.T. Macfarlane, and J.H. Cummings, *Sulphate reducing bacteria and hydrogen metabolism in the human large intestine.* Gut, 1993. **34**(4): p. 437-9.
57. Cummings, J.H. and G.T. Macfarlane, *The control and consequences of bacterial fermentation in the human colon.* J Appl Bacteriol, 1991. **70**(6): p. 443-59.
58. Macfarlane, S. and G.T. Macfarlane, *Regulation of short-chain fatty acid production.* Proc Nutr Soc, 2003. **62**(1): p. 67-72.
59. den Besten, G., et al., *The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism.* J Lipid Res, 2013. **54**(9): p. 2325-40.
60. Bergman, E.N., *Energy contributions of volatile fatty acids from the gastrointestinal tract in various species.* Physiol Rev, 1990. **70**(2): p. 567-90.
61. Hague, A., et al., *Apoptosis in colorectal tumour cells: induction by the short chain fatty acids butyrate, propionate and acetate and by the bile salt deoxycholate.* Int J Cancer, 1995. **60**(3): p. 400-6.
62. Gamet, L., et al., *Effects of short-chain fatty acids on growth and differentiation of the human colon-cancer cell line HT29.* Int J Cancer, 1992. **52**(2): p. 286-9.
63. Samuel, B.S., et al., *Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16767-72.
64. Kimura, I., et al., *The SCFA Receptor GPR43 and Energy Metabolism.* Front Endocrinol (Lausanne), 2014. **5**.
65. Campbell, J.M., G.C. Fahey, Jr., and B.W. Wolf, *Selected indigestible oligosaccharides affect large bowel mass, cecal and fecal short-chain fatty acids, pH and microflora in rats.* J Nutr, 1997. **127**(1): p. 130-6.
66. Duncan, S.H., et al., *The role of pH in determining the species composition of the human colonic microbiota.* Environ Microbiol, 2009. **11**(8): p. 2112-22.
67. Cherrington, C.A., et al., *Short-chain organic acids at ph 5.0 kill Escherichia coli and Salmonella spp. without causing membrane perturbation.* J Appl Bacteriol, 1991. **70**(2): p. 161-5.
68. Ewaschuk, J.B., J.M. Naylor, and G.A. Zello, *D-lactate in human and ruminant metabolism.* J Nutr, 2005. **135**(7): p. 1619-25.
69. Mortensen, P.B., et al., *Fermentation to short-chain fatty acids and lactate in human faecal batch cultures. Intra- and inter-individual variations versus variations caused by changes in fermented saccharides.* Scand J Gastroenterol, 1991. **26**(12): p. 1285-94.
70. Bustos, D., et al., *Fecal lactate and short bowel syndrome.* Dig Dis Sci, 1994. **39**(11): p. 2315-9.
71. Tanner, F.W., *Studies on the Bacterial Metabolism of Sulfur.* J Bacteriol., 1917. **2**(5): p. 585–593.
72. Levine, J., et al., *Fecal hydrogen sulfide production in ulcerative colitis.* Am J Gastroenterol, 1998. **93**(1): p. 83-7.
73. Tomasova, L., P. Konopelski, and M. Ufnal, *Gut Bacteria and Hydrogen Sulfide: The New Old Players in Circulatory System Homeostasis.* Molecules, 2016. **21**(11).
74. M Pitcher, E.B., and J Cummings, *The contribution of sulphate reducing bacteria and 5-aminosalicylic acid to faecal sulphide in patients with ulcerative colitis.* Gut, 2000. **46**(1): p. 64–72.
75. Levitt, M.D., et al., *Detoxification of hydrogen sulfide and methanethiol in the cecal mucosa.* J Clin Invest, 1999. **104**(8): p. 1107-14.
76. Pitcher, M.C. and J.H. Cummings, *Hydrogen sulphide: a bacterial toxin in ulcerative colitis?* Gut, 1996. **39**(1): p. 1-4.
77. Christl, S.U., et al., *Antagonistic effects of sulfide and butyrate on proliferation of colonic mucosa: a potential role for these agents in the pathogenesis of ulcerative colitis.* Dig Dis Sci, 1996. **41**(12): p. 2477-81.
78. Verbeke, K.A., et al., *Towards microbial fermentation metabolites as markers for health benefits of prebiotics.* Nutr Res Rev, 2015. **28**(1): p. 42-66.
79. Fiorucci, S., et al., *The emerging roles of hydrogen sulfide in the gastrointestinal tract and liver.* Gastroenterology, 2006. **131**(1): p. 259-71.

80. Richardson, A.J., N. McKain, and R.J. Wallace, *Ammonia production by human faecal bacteria, and the enumeration, isolation and characterization of bacteria capable of growth on peptides and amino acids.* BMC Microbiol, 2013. **13**: p. 6.
81. Vince, A.J. and S.M. Burridge, *Ammonia production by intestinal bacteria: the effects of lactose, lactulose and glucose.* J Med Microbiol, 1980. **13**(2): p. 177-91.
82. Down, P.F., et al., *The interrelations of faecal ammonia, pH and bicarbonate: evidence of colonic absorption of ammonia by non-ionic diffusion.* Clin Sci, 1972. **43**(1): p. 101-14.
83. Tjellstrom, B., et al., *Faecal short-chain fatty acid pattern in childhood coeliac disease is normalised after more than one year's gluten-free diet.* Microb Ecol Health Dis, 2013. **24**.
84. Tjellstrom, B., et al., *Gut microflora associated characteristics in children with celiac disease.* Am J Gastroenterol, 2005. **100**(12): p. 2784-8.
85. Jakobsdottir, G., et al., *Fasting serum concentration of short-chain fatty acids in subjects with microscopic colitis and celiac disease: no difference compared with controls, but between genders.* Scand J Gastroenterol, 2013. **48**(6): p. 696-701.
86. Tjellstrom, B., et al., *The effects of oats on the function of gut microflora in children with coeliac disease.* Aliment Pharmacol Ther, 2014. **39**(10): p. 1156-60.
87. Tjellstrom, B., et al., *Screening-detected and symptomatic untreated celiac children show similar gut microflora-associated characteristics.* Scand J Gastroenterol, 2010. **45**(9): p. 1059-62.
88. Kopecny, J., et al., *The intestinal microflora of childhood patients with indicated celiac disease.* Folia Microbiol (Praha), 2008. **53**(3): p. 214-6.
89. Tsiountsioura, M., et al., *Detailed assessment of nutritional status and eating patterns in children with gastrointestinal diseases attending an outpatients clinic and contemporary healthy controls.* Eur J Clin Nutr, 2014. **68**(6): p. 700-6.
90. Biagi, F., et al., *A gluten-free diet score to evaluate dietary compliance in patients with coeliac disease.* Br J Nutr, 2009. **102**(6): p. 882-7.
91. Biagi, F., et al., *A score that verifies adherence to a gluten-free diet: a cross-sectional, multicentre validation in real clinical life.* Br J Nutr, 2012. **108**(10): p. 1884-8.
92. Strocchi, A., J.K. Furne, and M.D. Levitt, *A modification of the methylene blue method to measure bacterial sulfide production in feces.* Journal of Microbiological Methods, 1992. **15**(2): p. 75-82.
93. Cline, J., *Spectrophotometric determination of hydrogen sulfide in natural waters.* Limnology and Oceanography, 1969.
94. Fleckenstein, B., et al., *Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process.* J Biol Chem, 2002. **277**(37): p. 34109-16.
95. Kumar, V., et al., *Celiac disease and immunoglobulin a deficiency: how effective are the serological methods of diagnosis?* Clin Diagn Lab Immunol, 2002. **9**(6): p. 1295-300.
96. Gerasimidis, K., et al., *Decline in presumptively protective gut bacterial species and metabolites are paradoxically associated with disease improvement in pediatric Crohn's disease during enteral nutrition.* Inflamm Bowel Dis, 2014. **20**(5): p. 861-71.
97. Schloss, P.D., *The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies.* PLoS Comput Biol, 2010. **6**(7): p. e1000844.
98. Joshi NA, F.J., *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*

(Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011.
99. Masella, A.P., et al., *PANDAseq: paired-end assembler for illumina sequences.* BMC Bioinformatics, 2012. **13**: p. 31.
100. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nat Meth, 2016. **13**(7): p. 581-583.
101. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.
102. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.* PLOS ONE, 2010. **5**(3): p. e9490.
103. Douglas, A.G., *Practical Statistics for Medical Research.* 1991.
104. Ciacci, C., et al., *Long-term follow-up of celiac adults on gluten-free diet: prevalence and correlates of intestinal damage.* Digestion, 2002. **66**(3): p. 178-85.

105. Moreno, M.L., et al., *Biomarkers to Monitor Gluten-Free Diet Compliance in Celiac Patients.* Nutrients, 2017. **9**(1).
106. Tursi, A., G. Brandimarte, and G.M. Giorgetti, *Lack of usefulness of anti-transglutaminase antibodies in assessing histologic recovery after gluten-free diet in celiac disease.* J Clin Gastroenterol, 2003. **37**(5): p. 387-91.
107. Dickey, W., D.F. Hughes, and S.A. McMillan, *Disappearance of endomysial antibodies in treated celiac disease does not indicate histological recovery.* Am J Gastroenterol, 2000. **95**(3): p. 712-4.
108. Sharkey, L.M., et al., *Optimising delivery of care in coeliac disease - comparison of the benefits of repeat biopsy and serological follow-up.* Aliment Pharmacol Ther, 2013. **38**(10): p. 1278-91.
109. Comino, I., et al., *Monitoring of gluten-free diet compliance in celiac patients by assessment of gliadin 33-mer equivalent epitopes in feces.* Am J Clin Nutr, 2012. **95**(3): p. 670-7.
110. Shan, L., et al., *Structural basis for gluten intolerance in celiac sprue.* Science, 2002. **297**(5590): p. 2275-9.
111. Skovbjerg, H., et al., *Deamidation and cross-linking of gliadin peptides by transglutaminases and the relation to celiac disease.* Biochim Biophys Acta, 2004. **1690**(3): p. 220-30.
112. Kalliomaki, M., et al., *Expression of microbiota, Toll-like receptors, and their regulators in the small intestinal mucosa in celiac disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(6): p. 727-32.
113. Green , P.H.R. and C. Cellier *Celiac Disease.* New England Journal of Medicine, 2007. **357**(17): p. 1731-1743.
114. Corrêa-Oliveira, R., et al., *Regulation of immune cell function by short-chain fatty acids.* Clin Transl Immunology, 2016. **5**(4): p. e73-.
115. Rivière, A., et al., *Bifidobacterium and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut.* Front Microbiol, 2016. **7**.
116. Engels, C., et al., *The Common Gut Microbe Eubacterium hallii also Contributes to Intestinal Propionate Formation.* Front Microbiol, 2016. **7**.
117. Muyzer, G. and A.J.M. Stams, *The ecology and biotechnology of sulphate-reducing bacteria.* 2008. **6**: p. 441.
118. Glassing, A., et al., *Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples.* Gut Pathog, 2016. **8**.

Appendix: PedsQL-GSS questionnaire

The PARENT proxy report **PedsQL™** Gastrointestinal Symptom Scale is composed of 9 items comprising 1 dimension.

DESCRIPTION OF THE GASTROINTESTINAL SYMPTOM SCALE:

Dimensions	Number of Items	Cluster of Items	Reversed Scoring	Direction of Dimensions
Gastrointestinal Symptoms	9	1-9	1-9	Higher scores indicate lower problems.

SCORING OF DIMENSIONS:

Item Scaling	5-point Likert scale from 0 (Never) to 4 (Almost always)
Weighting of Items	No
Extension of the Scoring Scale	Scores are transformed on a scale from 0 to 100.
Scoring Procedure	**Step 1: Transform Score** Items are reversed scored and linearly transformed to a 0-100 scale as follows: 0=100, 1=75, 2=50, 3=25, 4=0. **Step 2: Calculate Total Score** Sum of all the items over the number of items answered on all the Scales.
Interpretation and Analysis of Missing Data	If more than 50% of the items in the scale are missing, the Scale Scores should not be computed. If 50% or more items are completed: Impute the mean of the completed items in a scale.