Comparative Biology of γδ T Cells and the Gut Microbiota: Studies with Avian and Murine Systems.



Robert J Dixon

Department of Zoology
Hertford College

Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Hilary Term 2021

Declaration

I, Robert James Dixon, declare that this thesis has never been submitted, either in the same or a different form to this or any other university, for a degree or any other qualification. All the work presented in this thesis is the result of my own work unless otherwise stated and was carried out while I was registered as a graduate student at the Department of Zoology, University of Oxford, under the supervision of Dr. Adrian Smith



Oxford, Hilary Term 2021

Abstract

Interactions between the mucosal immune system, the microbiota and pathogens in the gut have far reaching implications for the host in terms of disease. The enteric mucosal immune system is known to develop in tandem with the microbiota during maturation and they drive changes in each other over time. γδ T cells are an enigmatic T cell subset which are greatly enriched at the intestinal mucosa and despite being in close contact with the microbiota, interactions between the two are not well described. This thesis aims to elucidate these interactions using Next Generation Sequencing methods to probe the TCR gamma repertoire and microbiota of both chicken and mouse, which fundamentally differ in terms of number of circulating γδ T cells.

This thesis first establishes a TCR gamma repertoire sequencing strategy in chicken. This strategy required an updated annotation of the gamma locus, indicating its evolution by tandem duplication. This enabled the first TCR gamma repertoire analysis in chicken, revealing that a single V gene segment, TRGV3.3, with a highly diverse, private CDR3 repertoire, was dominant in all tissues.

This annotated chicken gamma locus is then used to investigate the gut TCR gamma repertoire of Germ Free (GF) chickens, which lack an enteric microbiota. Surprisingly, the differences observed in the TCR gamma repertoire of these animals were relatively minor, suggesting that avian γδ T cells do not depend on the microbiota for development. To examine the reverse relationship, the gut microbial community of a recently constructed TCRγ -/- chicken strain, which lacked functional γδ T cells, was examined. The microbiota of these animals was again mostly unaffected, though reductions in potentially beneficial genera, such as *Bifidobacterium,* were observed in caecal contents.

Mixed *Campylobacter* strain infections are then utilised as an initiator of inflammation and dysbiosis in the chicken, resulting in a loss of homeostasis, which allows the examination of the γδ T cell-microbiota relationship in that context. Large differences in microbiota composition and changes

to the TCR gamma repertoire were observed. Interestingly these changes were dependant on the composition of the mixed strain *Campylobacter* infection.

Finally, a mouse model is used to examine whether the finding that the microbiota is largely unaffected in the absence of γδ T cells is specific to chicken or whether it has a general applicability. Studies in mouse also allowed the investigation of a lack of gut tropic TRGV7 γδ T cells and thymic development on both the gut TCR gamma repertoire and the microbiota. In contrast to what was observed in chicken, all of these γδ T cell perturbations in mice lead to large scale microbiota changes when compared with Wild Type.

The findings presented in this thesis provide the first detailed insight into the interactions between γδ T cells and the enteric microbiota of chickens and further explore the generality of these findings through similar experiments in mice.

Acknowledgements

I was fortunate enough to have been given the opportunity to study a genuinely fascinating project, which would not have been possible without the invaluable help of a number of people I would like to acknowledge.

I would firstly like to give thanks to my supervisor Adrian Smith, without whom this work would not have been possible. I will be forever grateful for and his unique and interesting insights and will never forget how much he helped me along the way.

I would like to thank the Oxford Doctoral Training Centre for initially providing me with this opportunity and the BBSRC for providing the funding to make my project possible.

Furthermore, I wish to express my gratitude to Patrik Flammer, Stefan Dascalu, Steven Fiddaman, Stephen Preston and Frances Colles for providing assistance and reassurance both experimental and emotional, and without whom I wouldn't have made it to the end.

I would like to thank Ivan Rychlik and Benjamin Schusser and their respective research groups, who facilitated the animal work used in this thesis and made me feel comfortable and welcome when visiting their facilities.

Shout out to my coursemates and housemates who have made my time at Oxford memorable and among whom are individuals I would consider lifelong friends.

I would finally like to thank my family, for supporting me, believing in me and knowing not to ask how much longer it was going to take me to finish.

List of Abbreviations

Gut Associated Lymphoid Tissue (GALT)
Peyer's patches (PP)
Tight Junction (TJ)
Intraepithelial lymphocytes (IEL)
Immunoglobulin A (IgA)
polymeric Ig receptor (pIgR)
T Cell Receptor (TCR)
Variable (V)
Joining (J)
Constant (C)
diversity (D)
palindromic (P)
Recombination-Activating Gene (RAG)
non-templated (N)
Terminal deoxynucleotidyl Transferase (TdT)
Butyrophilin-like molecules (Btnl)
T Receptor Gamma (TRG)
Feed Conversion Ratio (FCR)
Germ Free (GF)
Specific Pathogen Free (SPF)
Brain Heart Infusion (BHI)
Colony Forming Units (CFUs)
Rapid Amplification of cDNA Ends (RACE)
Amplicon Sequence Variant (ASV)
Nuclease Free (NF)
Restriction Signal-Sequence (RSS)
Recombination Information Content (RIC)
Next Generation Sequencing (NGS)
Complementarity-Determining Region (CDR)
Non-Metric Multidimensional Scaling (NMDS)
Passing Filter (PF)
Linear Mixed Model (LMM)
Jensen-Shannon divergence (JSD)
center log ratio (CLR)
Akaike Information Criterion (AIC)
Maximum Likelihood (ML)
Principal Coordinates Analysis (PCA)
Redundancy Analysis (RDA)
Permutational multivariate analysis of variance (PERMANOVA)
Dendritic Epidermal T Cells (DETCs)
Single Molecule Real Time sequencing (SMRT)
Complementarity Determining Region 3 Nucleotide (CDR3NT)
Dextran Sodium Sulfate (DSS)
similarity percentages (SIMPER)
Linear Model (LM)
False Discovery Rate (FDR)
Ribosomal Database Project (RDP)

Kyoto Encyclopedia of Genes and Genomes (KEGG)
Whole Genome Sequences (WGS)
galactooligosacharides (GOS)
Inflammatory Bowel Disease (IBD)
Toll-like Receptor (TLR)
Mono-Infected Group (MIG)
Co-Infection Group 1 (CIG1)
Co-Infection Group 2 (CIG2)
control (CTRL)
Mueller Hinton (MH)
Operational Taxonomic Units (OTUs)
Short Chain Fatty Acids (SCFA)
Type 6 Secretion System (T6SS)
Residual Feed Intake (RFI)
double negative (DN)
international ImMunoGeneTics (IMGT)
butyrophilin (Btn)
Keratinocyte Growth Factor (KGF)
antimicrobial peptides (AMPs)
Regenerating islet-derived protein 3 gamma (RegIIIγ)
Inflammatory Bowel Disease (IBD)

Table of Contents

Declaration II
Abstract III
Acknowledgements V
List of Abbreviations VI
Table of Contents VIII
List of Figures XIII

Chapter 1: General Introduction

1.1 The Enteric Immune System of Birds and Mammals – Common Features 2
1.2 The Enteric Immune System of Birds and Mammals – Differences 3
1.3 γδ T cells – A Non-Classical Lymphocyte 4
1.3.1 T Cell Subsets - γδ vs αβ TCRs 4
1.3.2 VDJ Recombination 5
1.3.3 γδ T Cell High vs Low Animals 6
1.3.4 γδ T Cell Subsets, Localisation to Different Tissues 6
1.4 The Gut Microbial Community 7
1.4.1 The Enteric Microbiota - Chickens 7
1.4.2 The Enteric Microbiota - Mice 9
1.5 Interactions between the Gut Microbiota and the Immune System – Focus on γδ T Cells 9
1.5.1 The Microbiota is Involved in Immune Development of the Intestine 9
1.5.2 The Immune System Monitors and Regulates the Gut Microbiota 10
1.6 Campylobacter – A Zoonotic, Dysbiosis Causing Pathogen 11
1.7 Thesis Structure 13
1.8 References 15

Chapter 2: Materials and Methods

2.1 General Sample Processing Workflow 22
2.2 Shared Methods 24
2.2.1 Tissue RNA Extraction 24
2.2.2 Faecal/Ileal Contents/Caecal Contents DNA Extraction 24
2.2.3 Qubit Nucleic Acid Quantification 25

2.2.4 cDNA Synthesis 25
2.2.5 Primers/barcoding 26
2.2.6 PCR 27
2.2.7 Agarose Gel 29
2.2.8 Library Mixing 30
2.2.9 Gel Extraction 30
2.2.10 Library Preparation 30
2.2.11 Library Quantification 30
2.2.12 Tapestation 31
2.2.13 Library Blending 31
2.2.14 MiSeq Loading - Denature and Dilute 31
2.3 Chapter 4 Specific Methods 31
2.3.1 Gamma Locus Annotation 31
2.4 Chapter 5 Specific Methods 32
2.4.1 Experimental Chickens 32
2.5 Chapter 6 Specific Methods 33
2.5.1 Culturing and Mixing of *Campylobacter* Strains 33
2.5.2 Experimental Chickens and *Campylobacter* Gavage 34
2.5.3 Sanger Sequencing 35
2.6 Chapter 7 Specific Methods 36
2.6.1 Experimental Mice 36
2.6.2 RNA/DNA Simultaneous Extraction from Tissue 37
2.7 References 37

Chapter 3: Method Development

3.1 Introduction 40
3.2 Experimental Results and Discussion 44
3.2.1 PCR 44
3.3 Analytical Results and Discussion 48
3.3.1 Repertoire Analytical Development 48
3.3.2 Repertoire and 16S Analytical Development 56
3.4 References 60

Chapter 4: The Chicken TCR Gamma Locus and Repertoire

4.1 Introduction 65

4.2 Chapter Outline 66
4.3 Results 66
4.3.1 The Chicken Gamma Locus shows a Repeating Motif suggesting Tandem Duplication based Expansion 67
4.3.2 The Chicken Gamma Repertoire Features a Dominant TCR gamma Rearrangement in every Tissue and Novel Public TCRs 72
4.3.3 Private and Public Compartments within the Chicken TCR Gamma Repertoire 76
4.3.4 Diversity of the TCR Gamma Repertoire 79
4.3.5 The Recombination Signal Sequence provides a Partial Explanation for Differential TRGV Gene Usage 82
4.4 Discussion 83
4.5 Conclusion 87
4.6 References 88

Chapter 5: Interactions between γδ T Cells and the Enteric Microbiota in the Chicken

5.1 Introduction 93
5.2 Chapter Outline 95
5.3 Results 95
5.3.1 Study Design 95
5.3.2 Tissue Sample Preparation and Processing for TCR Gamma Repertoire Sequencing 96
5.3.3 Beta-Diversity plots reveal small Differences in the Germ Free group 100
5.3.4 TRGV Gene Segment Usage is mostly Similar in the Germ Free Birds with some Small Differences 106
5.3.5 The Chicken TCR Gamma Repertoire contains highly Public TRGV Gene Segments 111
5.3.6 Diversity Plots reveal massive Increases in Diversity over Time but no Differences in the Germ Free Birds when compared to Wild Types 114
5.3.7 TCR gamma repertoire sequencing reveals differences between the Germ Free and Complex Birds at the level of individual clones 117
5.3.8 Chicken Sample Collection and Processing for 16S rRNA Sequencing 120
5.3.9 Beta-Diversity Plots reveal Differences in the Caecal Microbiota of TCRγ -/- Birds 124
5.3.10 Taxonomic Plots reveal Small Scale Differences in the Caecal Contents Microbiota of the TCRγ -/- Birds 130
5.3.11 Plotting Individual Genera reveals Sources of the Differences in the TCRγ -/- Caecal Contents Microbiota 136
5.3.12 Pseudometagenomic Plots Generated using Tax4Fun Predict Changes to Metabolic Genes in the Caecal Contents of TCRγ -/- Birds due to Reduced *Bifidobacterium* 139
5.4 Discussion 141
5.5 Conclusion 149

5.6 References 150

Chapter 6: Effect of *Campylobacter* Colonisation on the Microbiota and γδ T Cells in the Chicken

6.1 Introduction 157

6.2 Chapter Outline 158

6.3 Results 159

6.3.1 Experimental Design – PorA Type Confirmation and *Campylobacter* Strain Determination 159

6.3.2 Experimental Termination and PorA Sequencing from Caecal Sweeps 163

6.3.3 16S PCR, Sequencing and Data Processing using the DADA2 Pipeline 166

6.3.4 Beta-Diversity using Bray-Curtis, UniFrac, Aitchison's and PhILR Distances Reveals Microbiome Differences between the *Campylobacter* Infection Groups 170

6.3.5 Taxonomic Plots Reveal Significant Differences between the *Campylobacter* Infected Groups, Mainly at the Order Level 182

6.3.6 Genus Plots Reveal Further Differences between the *Campylobacter* Infection Groups 187

6.3.7 Pseudometagenomics Analysis with Tax4Fun Predicts Metabolic Changes in the *Campylobacter* Infection Groups 191

6.3.8 Chicken Tissue Sample Processing and TCR Gamma Repertoire PCRs 194

6.3.9 Illumina MiSeq Run and Processing Reads 196

6.3.10 Beta-Diversity Plots shows Small Differences in Repertoire Depending on *Campylobacter* Infection Group. 197

6.3.11 TRGV Gene Segment Usage Plots show Differences in the Proportion of some Minority Rearrangements according to *Campylobacter* Infection Group 202

6.3.12 Publicity in some TRGV Gene Segments is Enhanced by *Campylobacter* Colonisation .. 206

6.3.13 TCR Gamma Diversity in the Jejunum and Ileum is Altered by *Campylobacter* Colonisation 208

6.3.14 *Campylobacter* Colonisation Increases the Expression of Selected Largely Public TCR Gamma Clones 209

6.4 Discussion 213

6.5 Conclusion 222

6.6 References 223

Chapter 7: Interactions between γδ T Cells and the Enteric Microbiota in the Mouse

7.1 Introduction 230

7.2 Chapter Outline 232

7.3 Results 233

7.3.1 Mouse Tissue Extraction and PCR Amplification of TCR Gamma Repertoire using Universal Primers 233
7.3.2 The Butyrophilin -/- Mice have a Lower Proportion of γδ T cells Expressing TRGV7 in the Intestinal Tissues. 236
7.3.3 Beta-Diversity Plots reveal Differences in Overall TCR Gamma Repertoire between Mouse Strains 238
7.3.4 The TCR Gamma Repertoire is Highly Public in all Mouse Strains 242
7.3.5 TCR Gamma Repertoire Diversity is Greatly Reduced in the Nude Mice 243
7.3.6 PCR Amplification of the TRGV7 Repertoire using Specific Primers 247
7.3.7 Beta-Diversity Plots show Differences in the Overall TRGV7 Repertoire of both Butyrophilin -/- and Nude Mice 250
7.3.8 The TRGV7 Repertoire Diversity is Decreased in the Nude Mice 253
7.3.9 The Most Common, Canonical Clones Appear to be Most Affected in the Butyrophilin -/- Mice 254
7.3.10 16S Sequencing from Caecal DNA 257
7.3.11 MiSeq 16S Data Processing using the DADA2 Pipeline 258
7.3.12 Beta-Diversity Plots Reveal Microbiota Differences between all of the Mouse Strains. 260
7.3.13 Microbiome Taxonomic Composition Plots Reveal Differences between the Mouse Strains at the Phylum and Order Level 265
7.3.14 Individual Genus Plots Reveal Further Differences between the Mouse Strains, Particularly the Decrease in *Helicobacter* in all but the Wild Type mice 268
7.4 Discussion 271
7.5 Conclusions 277
7.6 References 278

Chapter 8: General Discussion and Future Work

8.1 General Discussion 285
8.1.1 Methods for the Processing and Analysis of 16S and Immune Repertoire Sequence Data 286
8.1.2 TCR Gamma Repertoire of Complex vs Germ Free Chickens showed only Minor Changes, however, this does not Rule Out Microbiome Interactions 287
8.1.3 Mouse vs Chicken TCR Gamma Repertoire 288
8.1.4 Mouse vs Chicken γδ T Cell Influence over the Microbiome 290
8.2 Future Work 293
8.2.1 Common to all Chapters 293
8.2.2 Chapter 5 294
8.2.3 Chapter 6 296
8.2.4 Chapter 7 296

8.3 Closing Remarks 298
8.4 References 299
8.5 Appendices 304

List of Figures

Chapter 1: General Introduction

Figure 1.1 - Structure of the Gut Mucosal Immune System 3
Figure 1.2 – Rearrangement of the Mouse TCR Gamma Locus 6

Chapter 2: Materials and Methods

Figure 2.1 - Workflow for 16S rRNA and Repertoire Sequencing 23
Figure 2.2 – SMARTer Reverse Transcription and Universal 5' RACE Repertoire PCRs 26

Chapter 3: Method Development

Figure 3.1 - Schematic Showing the 16S Ribosomal RNA Gene 42
Figure 3.2 – Bray-Curtis NMDS plot of Triplicate PCRs and No Template Controls 45
Figure 3.3 – TCR Gamma Chain RNA after Recombination, Transcription and Splicing 46
Figure 3.4 - Diluted 16S PCR Samples Highlight the need to Dilute Faecal PCR Inhibitors prior to Amplification 47
Figure 3.5 - Plots to Confirm Data fits the Assumptions of Linear Mixed Models 51
Figure 3.6 - Example BoxCox Plot for Linear Model Transformations 56

Chapter 4: The Chicken TCR Gamma Locus and Repertoire

Figure 4.1 - PCR-Restriction Digest Strategy to Resolve the TCR Gamma Locus 68
Figure 4.2 - Schematic Representation of the Chicken Gamma Locus 70
Figure 4.3 - Dotplot reveals that the Chicken TCR Gamma Locus Evolved via a set of Duplication Events 71
Figure 4.4 - Phylogenetic Tree of TRGV Gene Ig-like Domains 72
Figure 4.5 - The Proportion of each TRGV Gene Represented in TCR Gamma Chains within Different Tissues 74
Figure 4.6 - TRGV Usage in Tissues of two Distinct lines of Chicken reveals Conserved Patterns of Expression 75
Figure 4.7 - The Proportion of Public TCR Gamma CDR3 Sequences Varies According to TRGV Gene Usage 77
Figure 4.8 - Comparison of Publicity Profiles and Shared Public Clones between ISA Brown and PA12 White Leghorn Chickens 78
Figure 4.9 - Variation in the patterns of TCR gamma diversity in different tissues of ISA Brown chickens 79
Figure 4.10 - Diversity of TRGV rearrangements in different tissues 82

Figure 4.11 - Weak Positive Correlation between RSS Score and TRGV Gene Expression in Thymic TCR Gamma Repertoire 83
Table 4.1 - The Number of TRGV, J and C Genes in Various Animals 84

Chapter 5: Interactions between γδ T Cells and the Enteric Microbiota in the Chicken

Figure 5.1 - Example Library of TCR Gamma PCRs from 5' RACE Generated Chicken cDNA shows a Range of Amplicon Concentrations 97
Figure 5.2 - DNA Tapestation of NEBNext prepared MiSeq Amplicon Libraries 98
Figure 5.3 – Plot of Numbers of Reads at each Stage of Repertoire Sequence Processing 100
Figure 5.4 – β-Diversity Plots of Bray-Curtis, JSD and Aitchison's Distance Metrics Displaying Differences between Groups Split by Tissue and Age 106
Figure 5.5 - Proportion of TCR Gamma Rearrangements Originating from each TRGV Gene Segment 110
Figure 5.6 - Proportional Publicity for each TRGV Gene Segment at each Time Point 113
Figure 5.7 – TCR Gamma Repertoire Hill Number Diversity Indices Calculated for each Tissue at each Time Point 116
Figure 5.8 - Clone Plots of the Biggest Contributors to Bray-Curtis Differences between Bursa, Spleen and Jejunum as Determined by SIMPER Analysis 119
Figure 5.9 - Example Library of 16S PCR Products Generated from Faecal Samples 122
Figure 5.10 - Read Quality Profiles Highlight the need to Truncate Reads as Sequence Quality Declines with Length 122
Figure 5.11 – Numbers of Reads at each Stage of the 16S Processing Pipeline 124
Figure 5.12 – β-Diversity Ordination Plots showing Comparisons between the Faecal Microbiota of TCR γ-/-, TCRγ +/- and Wild Type Birds 128
Figure 5.13 – β Diversity Ordination Plots showing Comparisons between the Ileal and Caecal Microbiota of TCRγ -/-, TCRγ +/- and Wild Type Birds 129
Figure 5.14 - Faecal Microbiota 16S Taxonomic Stacked Barplots at the Phylum and Order Level 133
Figure 5.15 - Caecal and Ileal contents 16S taxonomic stacked barplots at the Phylum and Order level 135
Figure 5.16 - Individual Genera Plots from Caecal and Ileal Contents 138
Figure 5.17 - Tax4Fun Individual Gene Plots 141

Chapter 6: Effect of Campylobacter Colonisation on the Microbiota and γδ T Cells in the Chicken

Figure 6.1 – Experimental Design 161
Figure 6.2 - PorA and *Campylobacter jejuni* vs *coli* PCRs 163
Figure 6.3 - PorA Sequencing from Caecal Sweeps 165
Figure 6.4 - Example of 16S PCRs before Pooling and Sequencing 167
Figure 6.5 - Read Quality Profiles Highlight the need to Truncate as Sequence Quality Declines with Length 169
Figure 6.6 - The Majority of 16S Reads are Removed due to their Lack of Desired Barcodes/Primers 170
Figure 6.7 - Bray-Curtis, UniFrac, Aitchison and PhILR plots show Differences between CTRL and Campylobacter Infected Groups. 181
Figure 6.8 - Compositional Plots for each Campylobacter Infected Group show Increased Lactobacillus in CTRL 186
Figure 6.9 - Bar plots showing Genus/Species Level Differences uncovered by SIMPER Analysis 190

Figure 6.10 - Pseudometagenomic Analysis reveals Increases in Certain Genes Associated with Microbiota Differences 193
Figure 6.11 – TCR Gamma Repertoire PCRs from Multi Strain *Campylobacter* Infected Chickens 196
Figure 6.12 - Numbers of Reads at each Stage of Repertoire Data Processing 197
Figure 6.13 – TCR Gamma Repertoire β-Diversity Plots show Differences between *Campylobacter* Infection Groups 201
Figure 6.14 - TRGV Gene Usage Plots show Subtle TCR Gamma Rearrangement Differences in *Campylobacter* Infection Groups 205
Figure 6.15 - Campylobacter infected birds show slight increases in TCR gamma repertoire publicity relative to CTRL 207
Figure 6.16 - Hill Number Diversity Indices for each Tissue split by *Campylobacter* Infection Group. 209
Figure 6.17 - Individual TCR Gamma Clone Plots show Further Differences between *Campylobacter* Infected Groups 213

Chapter 7: Interactions between γδ T Cells and the Enteric Microbiota in the Mouse

Figure 7.1 – Agarose Gel of TCR Gamma Repertoire PCRs. 234
Figure 7.2 – Numbers of Reads at Various Stages of TCR Gamma Sequence Data Processing 236
Figure 7.3 – TRGV Gene Usage Profiles for each Strain in Duodenum, Jejunum, Ileum and Caecum 237
Figure 7.4 – β Diversity Plots of TCR Gamma Repertoire Reveal Differences between Mouse Strains 242
Figure 7.5 – Publicity Plots by TRGV Gene Segment 243
Figure 7.6 – TCR Gamma Hill Number Diversity Plots Reveal Differences in the Mouse Strains 247
Figure 7.7 – Agarose Gel of TRGV7 Specific PCRs 248
Figure 7.8 – Numbers of Reads after each Stage of TRGV7 Specific Repertoire Processing 249
Figure 7.9 - β-Diversity Plots of TRGV7 Specific Repertoire Data shows that the TRGV7+ Clonal Populations are Altered in Butyrophilin -/- and Nude Mice 252
Figure 7.10 – Hill Number Diversity Plots for TRGV7 Specific Repertoire 253
Figure 7.11 – Individual TRGV7 Clone Plots reveal Clones Reduced in Butyrophilin -/- Mice 256
Figure 7.12 – Agarose Gel showing 16S PCR Products from Caecal Tissues 257
Figure 7.13 – Illumina Sequence read FASTQ Quality Score Plots 259
Figure 7.14 – Read Tracking Plots for 16S Sequence Processing 260
Figure 7.15 – 16S β-Diversity Plots by Strain and Tissue reveal clear Differences between the Caecal Floral of each of the Mouse Strains 264
Figure 7.16 – 16S Microbiome Composition Plots reveal a large Decrease in *Campylobacterales* in all Mouse Strains except Wild Type 267
Figure 7.17 – Individual Genera Plots reveal Fine Scale Microbiota Changes in each of the Mouse Strains 270

Appendices

Figure A1 – Sequencing Primer Barcodes 304
Figure A2 - Levels of Publicity in each TRGV gene Segment in PA-12 White Leghorn Chickens 305
Figure A3 - Rarefaction Plots Generated using iNEXT 306
Figure A4 - Variation in the Patterns of TCR Gamma Diversity in different Tissues of PA12 White Leghorn Chickens. 307
Figure A5 - Diversity of TRGV Gene Segment Rearrangements in different Tissues 308

Figure A6 – Libraries used in GF TCR Gamma Repertoire Experiment (See Chapter 5) 311
Figure A7 – Antibody Testing for Chicken IgA-Seq..... 311

Department of Zoology, University of Oxford

Chapter 1

General Introduction

1.1 The Enteric Immune System of Birds and Mammals – Common Features

The enteric immune systems of mammals and birds contain various differences when compared to their circulatory immune systems. These differences reflect the unique nature of the gut environment, characterised by a high microbial load and constant turnover of the epithelial cells that line it[1]. The gut mucosal immune system carries out the important functions of identifying and responding to antigens of the microbiota, eliminating potential food-borne pathogens and maintaining the crucial separation between microbes and the rest of the body. In order to carry out these roles, the mucosal immune system maintains a tightly regulated structure known as the Gut Associated Lymphoid Tissue (GALT).

At the tissue level, the gut is lined with alternating villi and crypts, which increase the surface area available for nutrient absorption[2]. These gut undulations are interspersed with lymphoid structures called Peyer's patches (PP) (Fig 1.1A). PP are distributed across the intestine and contain high concentrations of M cells, which reside in the epithelium and specialise in phagocytosis and transfer of microbial antigens for presentation to lymphoid follicles[3]. These follicles are rich in developing lymphocytes of both the T and B cell lineages.

The intestinal epithelium itself consists predominantly of enterocytes, which are joined by Tight Junction (TJ) proteins[4] and interspersed with goblet cells, which secrete mucins, forming a mucosal layer[5] (Figure 1B). Intraepithelial lymphocytes (IELs) are a group of specialised immune cells found dispersed across the epithelium, where they are resident. This population consists mainly of T cells which carry out important functions, such as maintaining and protecting the epithelial barrier[6]. The lamina propria is a cell rich layer of connective tissue beneath the epithelium and hosts large numbers of lymphocytes[7] including both CD8+ve and CD4+ve T cells, as well as B cells, plasma cells, dendritic cells, neutrophils, granulocytes and macrophages[8,9]. Plasma cells in the lamina propria are

biased towards Immunoglobulin A (IgA) production, which is dependent upon microbiota presence[10], secreted IgA is exported into the intestinal lumen through transcytosis via the polymeric Ig receptor (pIgR)[11].

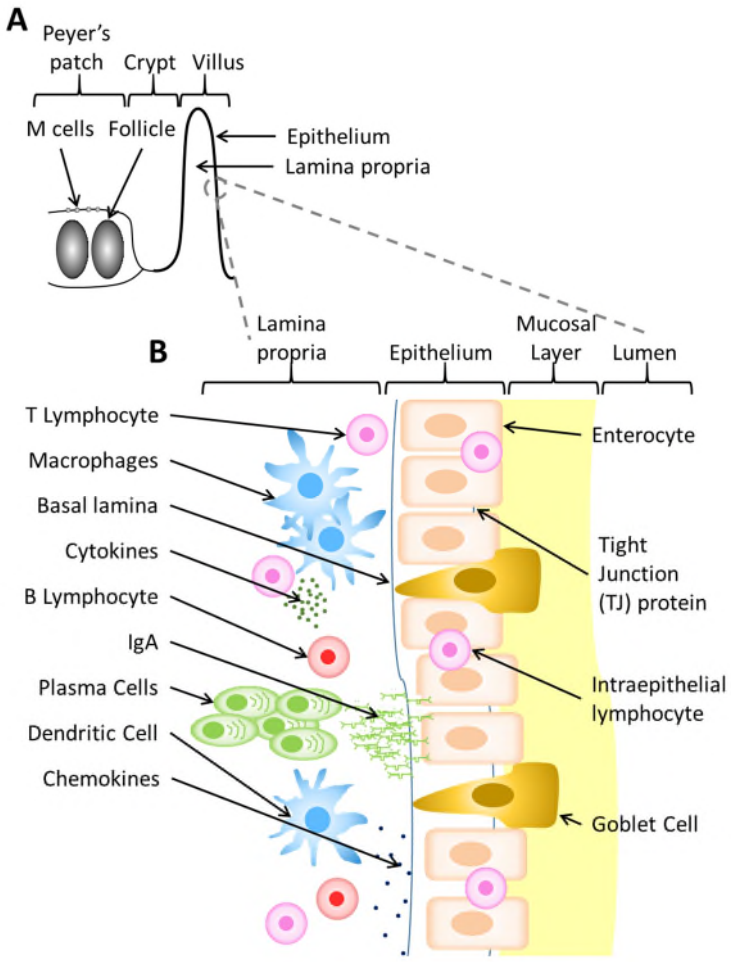


Figure 1.1 - Structure of the Gut Mucosal Immune System
Schematic of the organisation of the gut immune system shown at both **A)** the tissue level **B)** and the cell level.

1.2 The Enteric Immune System of Birds and Mammals – Differences

The avian gut immune system is mostly similar to that of mice and other mammals, but there are a number of differences that should be highlighted. The most obvious difference between the overall structure of the mouse and chicken gut is the presence of paired blind caeca in the

chicken. Though mice do possess caeca that are relatively large in comparison to those found in humans, the anatomy of avian caeca is very different. In chicken, the caeca are 2 intestinal protrusions which extend out to roughly the same length and terminate in blind ends, an anatomy totally distinct to the single kidney-shaped caecum found in mouse[12]. The caeca in chicken are the largest reservoir of microbes in the gut and periodically empty into the intestine at the ileal-caecal junction[13]. The microbiota of the caecum is distinct from that of the small intestine and colon and this should be considered when conducting microbiome studies on chicken faecal samples[14]. Additionally, the villi in the caeca display a more complex, ruffled structure leading to crypts both at the bases of villi, as in mouse, but also between ruffles[15]. The GALT of chicken and mouse both feature PP, made up of M cells overlaying lymphoid follicles, however one major difference between the two is the lack of defined lymph nodes in chicken. In mouse these PP drain into mesenteric lymph nodes via afferent lymph vessels, however these structures are entirely absent in the chicken. Despite this, the function of the M cells remains the same, sampling of the gut microbiota and delivery of antigens to underlying macrophages and dendritic cells[16–18].

The makeup of the avian GALT at the cellular level is very similar to that observed in mammals, with a few minor differences. Avian neutrophils are instead known as heterophils and some differences exist in cytokine and chemokine profiles, though clear homology in terms of family and function is present. Avian γδ T cells are predominantly CD8+ and are able to express a wide range of different cytokines[15,19].

1.3 γδ T cells – A Non-Classical Lymphocyte

1.3.1 T Cell Subsets - γδ vs αβ TCRs

Immune surveillance in jawed vertebrates, including mice and chickens, is carried out by 2 lineages of T lymphocytes, which are present at the gut as part of the IEL, as well as in other tissues within the body. These lymphocytes are named according to the somatically recombined surface receptor they display, which can consist of a heterodimer made up of either α and β subunits as in

αβ T cells, or γ and δ subunits as in γδ T cells. αβ T cells are the most abundant in blood and therefore the best characterized. The αβ T Cell Receptor (TCR) recognizes peptides from degraded proteins, as well as other molecules, only when they are bound to Major Histocompatibility protein (MHC). Conversely γδ T cells are less abundant in circulation and can bypass MHC restriction, enabling them to recognize molecules in their native context[20]. These cells are known for their non-classical behaviour, unlike the more commonly studied αβ T cells, which are said to have more conventional behaviour. The IELs are greatly enriched in γδ T cells, which can make up at much as 60% of the T cells of this population[21,22]. Whereas αβ T cells recognise antigen only when it is bound to MHC proteins, γδ T cells are able to bypass this restriction and thus can recognise molecules in their native context[23]. Additionally, γδ T cells are able to display both innate-like behaviour, characterised by fast responses and a lack of TCR diversity, and adaptive-like behaviour, characterised by slower responses by cells expressing clonally distinct TCRs[24].

1.3.2 VDJ Recombination

Despite these differences, the development of γδ and αβ TCRs is similar in many ways and both develop their receptor diversity through VDJ recombination. In order to recognize a vast array of foreign molecules using a limited set of gene segments, both T cell subsets must generate the diversity they require by a process of gene rearrangement (Fig 1.2). During this process, in the TCRγ locus, which is presented as it is the focus of the repertoire section of this thesis, Variable (V) and Joining (J) segments recombine with each other, becoming spatially adjacent and within close proximity of a Constant (C) region. This brings the V promoter close to the enhancer regions, enabling transcription to occur[11]. Rearrangement in the TCRβ and TCRδ locus occurs in a similar manner, however an additional segment, diversity (D) is also present. Additional diversity is generated at the VDJ junctions, both through palindromic (P) nucleotides, generated by random hairpin cleavage activity of the Recombination-Activating Gene (RAG) complex, and through additional random non-templated (N) nucleotides, added by Terminal deoxynucleotidyl Transferase (TdT),which increases in expression during early life, resulting in mature γδ TCRs[20]. RNA processing

then occurs, removing the intervening sequences, leaving the mature mRNA, which is then translated into the TCR protein. The diversity of these TCRs varies between individuals and the number of different sub-types which an organism produces is known as the TCR repertoire.

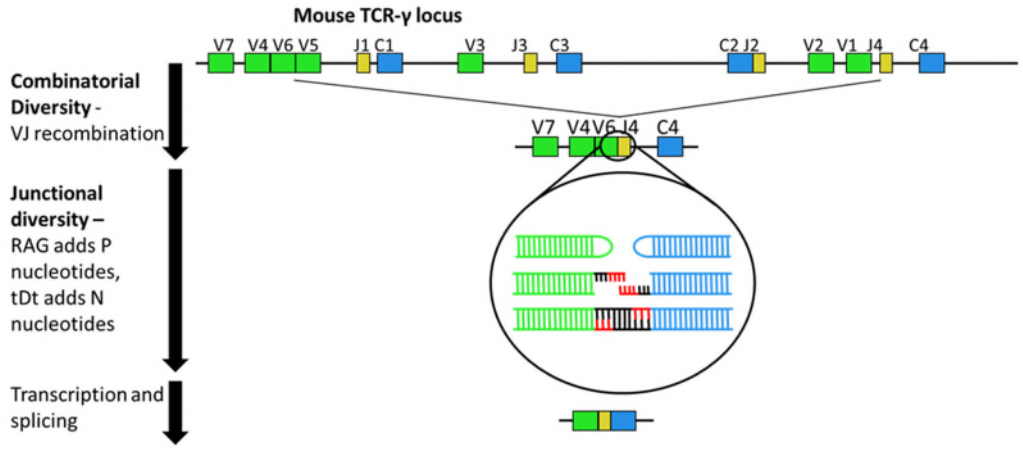


Figure 1.2 – Rearrangement of the Mouse TCR Gamma Locus
Diagram showing the process of recombination of the mouse TCR Gamma locus. Black bases are P nucleotides generated by RAG, red bases are N nucleotides added by TdT

1.3.3 γδ T Cell High vs Low Animals

Though they are mostly similar, γδ T cells of mice and chickens differ in a few crucial ways. In terms of their circulating γδ T cell populations, animals can be broadly separated into 2 groups, 'γδ high animals' and 'γδ low animals'[25]. These groups differ in their proportions of circulating T cells; γδ lows have <10% circulating γδ T cells, wheras γδ highs have >10%. Mice and humans fit into the γδ low category, whereas chickens and domestic animals and ruminants are γδ high[25]. This suggests a conserved fundamental difference in the emphasis different species place on γδ T cells. Thus it is important that γδ T cell behaviour is studied in representatives of both of these groups.

1.3.4 γδ T Cell Subsets, Localisation to Different Tissues

One aspect of γδ T cell biology which is conserved across species is their enrichment at mucosal sites, including the gut. In mouse it has been shown that different subsets, named according

to the V gene segment they acquire during VDJ recombination, localise to different tissues[26]. This process in mice is thought to be mediated by a family of proteins known as the Butyrophilin-like molecules (Btnl)[27]. In chickens a similar family exists, the BG genes, however their activity is yet to be characterised[28]. A recent paper showed that the migration of T Receptor Gamma (TRG) V7, the dominant gamma segment in the mouse gut, is mediated by Btnl1[26]. Despite showing that the numbers of gut TRGV7+ γδ T cells are reduced in the Btnl1 -/- mouse strain, the more precise nature of the impact of this knockout on the gamma repertoire is not known.

1.4 The Gut Microbial Community

The presence of the gut microbiota, the most diverse and numerous microbial community of any tissue is one of the reasons the enteric immune system must be so specialised. Most, but not all, organisms possessing a long, complex gut have a complex resident microbial community. In both mouse and chicken this microbiota is initially inherited by maternal transfer and develops rapidly as the animals mature[29,30]. In both of these animals, the mature microbiota consists mainly of the two phyla *Firmicutes* and *Bacteroidetes,* though their relative proportions only reach stability in adulthood.

1.4.1 The Enteric Microbiota - Chickens

Gaining an understanding of chicken gut microbiota composition is of interest to the poultry industry because it affects bird health, welfare and productivity. Chicken colonisation by the microbiota begins at hatch and prior to this, the developing embryo is sterile. In the wild, much of the early colonising microbiota would originate from maternal faeces, through bacteria present on the egg surface and during rearing[31,32], or from the environment. Once hatch and initial colonisation has occurred, the microbiota develops rapidly over the first few weeks of life, with successive colonisation of strains transferred from the environment in the form of food and water. This initial development period is crucial for obtaining a healthy adult microbiota and its disruption can lead to

a loss of productivity or increased susceptibility to disease[33]. It is the importance of this period that has fuelled the development of a range of pre and probiotics for industrial farm use[34].

A study into the composition of the caecal microbiota in egg laying hens was recently conducted revealing large scale changes, which occur at different stages of life, with particularly rapid development occurring in birds just days post hatch[29]. This study showed that the birds were initially dominated by the phyla *Proteobacteria,* which was mostly made up of *Escherichia* at the genus level. However, the amount of *Firmicutes* rapidly increased to supplant the *Proteobacteria* and become the dominant phyla at around day 10. Following this, at around 3 weeks of age, the *Bacteroidetes* are first observed in the microbiota. By 2 months of age, this population has increased in proportion to occupy around 50% of the microbiota at the expense of the previously dominant *Firmicutes*. In addition, though the chicken gut microbiota has been shown to vary considerably between individuals, it has been shown that this inter-individual variation tends to decrease as birds age[35]. As well as being highly variable between individuals, there is also considerable microbiome variation between different breeds of chicken, with broilers showing considerable differences, particularly during the earliest stages of life[33,36]. Though environmental factors, such as feed, may account for some of these differences, they are at least partly attributable to host genetics.

The chicken gut microbiota impacts on productivity in terms of total body weight and Feed Conversion Ratio (FCR)[37]. Though the composition of the flora in other gut sections impacts on these performance measures, changes to the microbial composition of the caeca have been shown to have the greatest impact. The genera *Lactobacillus, Clostridium, Bacteroides, Akkermansia,* and *Faecalibacterium* were all shown to be associated with good performance, whereas *Eschericia/Shigella* and *Campylobacter* were shown to have negative effects in terms of FCR and digestion[38–40]. Despite this, the evidence for microbial associations with productivity is often contradictory and should be considered cautiously and with this in mind.

1.4.2 The Enteric Microbiota - Mice

Studies on the mouse and its microbiota are mainly interested in using it as a model organism and a proxy for the human enteric flora[41]. Perhaps because developmental studies using faecal samples can be performed on humans, the literature on the development of the mouse microbiota is surprisingly limited. Despite this, and similar to chickens, it appears that as mice mature, the amount of the phyla *Bacteroidetes,* which seems to make up the majority of the microbiota in younger animals, decreases and is replaced by *Firmicutes,* which becomes the dominant phyla in the adult mouse[30,42]. However, some studies have in fact seen the opposite relationship over shorter periods of time[43].

At the genera level, the combination of multiple studies[42–47] estimate the core mature mouse gut microbiota to consist of around 352 genera[41]. Interestingly, overlap was poor when comparing these studies and only 4 genera, were found in every study, though some studies overlapped more than others. This suggests that though a core microbial community exists, the mouse gut microbiota is also highly variable across different colonies and individuals.

1.5 Interactions between the Gut Microbiota and the Immune System – Focus on γδ T Cells

1.5.1 The Microbiota is Involved in Immune Development of the Intestine

A healthy gut microbiota can be thought of as a part of the enteric immune system and competes with incoming pathogens, both directly, through interfering with their growth and adhesion, and indirectly, by competing for resources[48]. Additionally, the gut microbiome plays a crucial role in development of the enteric immune system and the absence of this community has profound and wide reaching impacts on many gut lymphocytes[49]. Interestingly, these changes are generally not permanent and can be reversed if they animals are re-conventionalised later in life.

This reconstitution ability tends to be host specific however, and some immune changes cannot be rescued, even with floras from closely related species[50].

Little is known about the influence the microbiota has on γδ T cell development specifically. It has been shown that in re-conventionalization studies of previously Germ Free (GF) rats, an increase in numbers of αβ but not γδ IELs can be observed following microbial exposure[51,52]. Similarly, early studies in GF mice showed no difference in numbers of IELs, or numbers of γδ T cells in the absence of a microbiota[53,54]. This may be due to the early importance of γδ T cells in immunity and by the time young animals become colonised, the majority of their development has already occurred. On the other hand, αβ T cells develop later and thus may use microbial cues for the timing of their development[51]. More recently, studies have shown that certain aspects of γδ T cell effector functions are impaired in GF mice[55]. However, γδ T cell development regarding the microbiota remains little studied and these interactions in γδ high animals, such as chicken, are even more neglected in the literature.

1.5.2 The Immune System Monitors and Regulates the Gut Microbiota

As many members of the gut flora are themselves opportunistic pathogens and can cause disease in the context of injury, inflammation or dysbiosis[56] they are constantly monitored by the resident lymphocytes. In order to prevent the harmful outgrowth of a single species, it is often necessary for various players within the enteric immune system to take punitive action against certain microbes, thus regulating the composition of the microbial community and maintaining homeostasis[57].

Some initial studies have begun to examine the influence γδ T cells can exert over the microbiota. Studies in mouse have shown that γδ T cells are involved in maintaining microbial homeostasis by releasing antimicrobial factors and in the absence of γδ cells the gut was more prone to microbial penetration into the mucosa following injury[55,58,59]. TCRδ -/- animals, which are unable to make functional γδ T cells offer a valuable opportunity to examine the impacts the loss of these

lymphocytes has on the microbiota. Recent construction of a TCRγ -/- chicken also allows us to examine this relationship in a γδ high animal and the Btnl -/- mouse allows us to look at the effect of removing the Vγ7+ γδ T cell subset specifically.

1.6 Campylobacter – A Zoonotic, Dysbiosis Causing Pathogen

Campylobacter is a zoonotic, bacterial pathogen, which causes serious financial harm to global economies. In the UK alone there are around 700,000 cases per year of diseases caused by *Campylobacter*, costing the national economy around £2 billion pa[60]. *C. jejuni* is the cause of around 85% of cases of campylobacteriosis from poultry, the other 15% can be attributed to *C. coli*[61]. From this point onwards, *Campylobacter* can be taken to refer to *C. jejuni* and *coli* only, as these are the most significant species within the genus. *Campylobacter* is from the epsilon group of *proteobacteria*. It is spiral shaped, but can become coccoid under stress and most strains are motile via polar flagella. *Campylobacter* is a microaerophilic facultative intracellular pathogen and grows best at 37 or 42°C[62]. Food borne campylobactereriosis from poultry products is extremely common and is difficult to eliminate due to the ability of *Campylobacter* to colonise the GI tract with as few as 500 Colony Forming Units (CFUs), combined with its resilience to antibacterial agents when contaminating poultry, which is in contrast with its notorious fastidiousness in culture[63,64]. Infection can lead to extra-intestinal spread in immune deficient individuals[65] resulting in Guillain Barre syndrome[66,67] and reactive arthritis[68].

In developed countries, 50-70% of *Campylobacter* infection is attributable to the consumption of poultry products[69]. It was previously thought to be a commensal in chickens[70,71] but new studies have shown that it triggers innate and adaptive immune responses and can affect bird welfare[72]. *Campylobacter* in chickens is certainly not highly pathogenic, in most circumstances being restricted to the gut lumen, but due to the strong immune responses and potential for causing mild pathologies may be considered at the cusp of the pathogen-commensal continuum. Chicken co-colonization with several strains of *Campylobacter* is common[73–77], particularly in breeder flocks. It

has been calculated that in the commercial chicken rearing environment where thousands of birds are co-housed, *Campylobacter* can go from colonization of 5% of individuals to 95% prevalence in just 6 days[78]. Additionally, different strains are capable of interaction and have been shown to interfere with one another's colonization[79].

It has been shown that *Campylobacter* is exceptionally difficult to eliminate from chicken flocks and colonization can only be prevented by extremely rigorous biosecurity measures unfeasible for most farm settings[80]. As a result, despite intensive research interest, the case rate has remained consistent. As eradication and prevention of colonization have thus far been unsuccessful, there is increasing interest in the field in the idea of control and management approaches. Such approaches involve attempting to prevent outgrowth of *Campylobacter* rather than eliminating it altogether. Recent evidence using modern molecular detection methods shows that *Campylobacter* was detectable at low levels in every flock tested in birds under 8 days old. Not only this, but these birds were already colonized by multiple different *Campylobacter* strains as determined using the PorA gene typing system[81]. Therefore, efforts should instead be focused on preventing outgrowth of a single strain, which leads to carcass contamination during slaughter[81,82].

Some experimental *Campylobacter* co-colonization studies have been performed on chickens. These studies agree that during experimental infection with multiple *Campylobacter* strains, though they are able to stably cohabit a host, a single strain tends to become dominant and outnumber the others, both at the individual and at the flock level[83]. Additionally, some strains appear to have different colonization dynamics, and preferentially infect certain tissues at different rates[84]. Moreover, multiple studies have shown that the dominant strain does not remain dominant over long periods of time, and is often supplanted by a previously less common strain[83,85–87]. These switches seem to occur independent of growth rate and microbiota differences and it has instead been speculated that they are likely to be down to host immune effects or even the effects of bacteriophage[85,87].

Initially it was thought that *Campylobacter* was a harmless commensal of chickens and a normal part of the gut flora[88]. However, more recent studies have shown that colonization has impacts on bird welfare and are associated with a Th17 mediated inflammatory response[72,89]. Additionally, during infection of the chicken gut, very high *Campylobacter* loads have been reported in the caeca[90,91], leading to dysbiosis, and differences in a wide range of genera[90,92]. As *Campylobacter* is prevalent in chickens and is known to causes perturbations to the gut flora and initiate an immune response, in the context of this thesis it was utilized as a 'natural stressor' and used to study microbiota-γδ T cell interactions in the context of a loss of gut homeostasis.

1.7 Thesis Structure

This thesis will consider the interactions between the microbiota, the immune system and pathogens in the gut. The immunological studies will focus on γδ T cells, as a unique and unconventional subset of lymphocytes, which are enriched at the intestinal epithelia and studied here using NGS repertoire sequencing. We perform our analyses on both chicken and mouse as representatives of both γδ high and low animals respectively to test for the generality of any interactions. Additionally, studies with *Campylobacter* infection were chosen to provide biotic stress to the chicken gut immune system and microbiota.

Chapter 3 will consider the development and refinement of both repertoire and 16S microbiome profiling approaches that will be used throughout the thesis. The technical and analytical pipelines will be presented and discussed in terms of their development during the course of this work.

The first data chapter, Chapter 4, represents the first gamma repertoire study of chicken gut γδ T cells. To fully establish repertoire analysis in the chicken it was important to begin with a well-annotated, accurate version of the TCR gamma locus. This chapter confirmed the structure and updated the annotation of this locus, identifying four families of TRGV gene segments and a characteristic duplication pattern of expansion. This framework allowed formulation of a strong TCR

gamma repertoire analysis pipeline that was applied to tissue samples from chickens of two different breeds. The repertoire was remarkably diverse and included both public and private CDR3 sequences. This dataset assisted and confirmed annotation of the TCR gamma locus and also revealed a very wide level of variation in the usage of different TRGV segments, which is discussed in detail in this chapter.

Chapter 5 examines the relationship between γδ T cells and the microbiota from the perspective of both. We first examine the TCR gamma repertoire of GF birds and compare it to that of those with a complex flora to look at the effects of γδ T cell development in the absence of a microbiota. Furthermore, the collection of tissue samples from birds of different ages allows us to examine how the TCR gamma repertoire develops over time. The very recent availability of a TCRγ -/- chicken allowed examination of the reciprocal relationship to test the hypothesis that γδ T cells play a role in regulating the microbiota and therefore the loss of these lymphocytes would lead to changes in the 16S microbiome profile.

Chapter 6 examines the impact of *Campylobacter* on the γδ T cell microbiota relationship. This study was performed using a realistic co-infection made up of strains isolated from farm birds and included an assessment of the impact of a single *Campylobacter* strain versus multiple strains. This chapter aimed to test the hypothesis that γδ T cell microbiota interactions would show a differing phenotype under conditions of enteric stress.

The final data chapter, Chapter 7, will consider the generality of the findings from chicken models, by testing them in mice. This chapter will firstly consider the impact on the TCR gamma repertoire of loss of the gut γδ localization agent Btnl1, which is responsible for enteric homing of TRGV7+ γδ T cells in mouse, as well as the loss of thymic development in a Nude mouse strain. Though this work has been previously done using lower resolution flow cytometry methods, TCR gamma repertoire sequencing will provide a deeper insight into the impact of these γδ T cell perturbations. This chapter will then consider the effects of these mutations, as well as a TCRδ -/-

strain, which lacks all functional γδ T cells, on the microbiota, using a similar set up to that used in the parallel chicken experiment.

Finally, in the general discussion, the data from all chapters will be integrated to explore chicken γδ T cells and their microbiota interactions under conditions of both microbiota homeostasis and dysbiosis as well as assessing the generality of these findings and how they compare to mouse. To conclude the thesis, possible follow up analyses will be discussed.

1.8 References

1. Williams JM, Duckworth CA, Burkitt MD, Watson AJM, Campbell BJ, Pritchard DM. Epithelial Cell Shedding and Barrier Function: A Matter of Life and Death at the Small Intestinal Villus Tip. Vet Pathol. 2015;52(3):445–55.
2. Martini E, Krug SM, Siegmund B, Neurath MF, Becker C. Mend Your Fences: The Epithelial Barrier and its Relationship With Mucosal Immunity in Inflammatory Bowel Disease. Cmgh. 2017;4(1):33–46.
3. Ohno H. Intestinal M cells. J Biochem. 2015;159(2):151–60.
4. Gonzalezmariscal L. Tight junction proteins. Prog Biophys Mol Biol. 2003;81(1):1–44.
5. McCauley HA, Guasch G. Three cheers for the goblet cell: Maintaining homeostasis in mucosal epithelia. Trends Mol Med. 2015;21(8):492–503.
6. Cheroutre H, Lambolez F, Mucida D. The light and dark sides of intestinal intraepithelial lymphocytes. Nat Rev Immunol. 2011;11(7):445–56.
7. Burkitt, H. George; Young, Barbara; Heath JW. Wheater's Functional Histology. 3rd ed. 1993.
8. Mowat AM, Agace WW. Regional specialization within the intestinal immune system. Nat Rev Immunol. 2014;14(10):667–85.
9. Brandtzaeg P, Pabst R. Let's go mucosal: Communication on slippery ground. Trends Immunol. 2004;25(11):570–7.
10. Brandtzaeg P, Johansen FE. Mucosal B cells: Phenotypic characteristics, transcriptional regulation, and homing properties. Immunol Rev. 2005;206(1):32–63.
11. Murphy K. Janeway's Immunobiology. 8th ed. 2011.
12. Snipes RL. Anatomy of the cecum of the laboratory mouse and rat. Anat Embryol (Berl). 1981;162(4):455–74.
13. Clench M, Mathias J. The avian cecum: a review. Wilson Bull. 1995;107(1):93–121.
14. Glendinning L, Watson KA, Watson M. Development of the duodenal, ileal, jejunal and caecal microbiota in chickens. bioRxiv. 2019;1–11.
15. Schat KA, Kaspers B, Kaiser P. Avian Immunology. 2014.

16. Burns RB, Maxwell MH. Ultrastructure of Peyer's patches in the domestic fowl and turkey. J Anat. 1986;147:235–43.

17. Gallego M, del Cacho E, Bascuas JA. Antigen-binding cells in the cecal tonsil and Peyer's patches of the chicken after bovine serum albumin administration. Poult Sci. 1995;74(3):472–9.

18. Jeurissen SHM, Wagenaar F, Janse EM. Further characterization of M cells in gut-associated lymphoid tissues of the chicken. Poult Sci. 1999;78(7):965–72.

19. Pieper J, Methner U, Berndt A. Heterogeneity of avian γδ T cells. Vet Immunol Immunopathol. 2008;124(3–4):241–52.

20. Attaf M, Legut M, Cole DK, Sewell AK. The T cell antigen receptor: The Swiss army knife of the immune system. Clin Exp Immunol. 2015;181(1):1–18.

21. Guy-Grand D, Cerf-Bensussan N, Malissen B, Malassis-Seris M, Briottet C, Vassalli P. Two gut intraepithelial CD8+ lymphocyte populations with different T cell receptors: a role for the gut epithelium in T cell differentiation. J Exp Med. 1991;173(2):471–81.

22. Goodman T, Lefrançois L. Expression of the γ-δ T-cell receptor on intestinal CD8+ intraepithelial lymphocytes. Nature. 1988;333(6176):855–8.

23. Godfrey DI, Uldrich AP, Mccluskey J, Rossjohn J, Moody DB. The burgeoning family of unconventional T cellsGodfrey, D.I. et al., 2015. The burgeoning family of unconventional T cells. , 16(11). Nat Immunol. 2015;16(11):1114–24.

24. Hayday AC. γδ T Cell Update: Adaptate Orchestrators of Immune Surveillance. J Immunol. 2019;203(2):311–20.

25. Holderness J, Hedges JF, Ramstead A, Jutila MA. Comparative biology of γδ T cell function in humans, mice, and domestic animals. Vol. 1, Annual Review of Animal Biosciences. 2013. p. 99–124.

26. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.

27. Jandke A, Melandri D, Monin L, Ushakov DS, Laing AG, Vantourout P, et al. Butyrophilin-like proteins display combinatorial diversity in selecting and maintaining signature intraepithelial γδ T cell compartments. Nat Commun. 2020;11(1).

28. Chen L, Fakiola M, Staines K, Butter C, Kaufman J. Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. Front Immunol. 2018;9(MAY):1–17.

29. Videnska P, Sedlar K, Lukac M, Faldynova M, Gerzova L, Cejkova D, et al. Succession and replacement of bacterial populations in the caecum of egg laying hens over their whole life. PLoS One. 2014;9(12):1–14.

30. Sun BL, Li WW, Wang J, Xu YL, Sun HL, Tian DiY, et al. Gut Microbiota Alteration and Its Time Course in a Tauopathy Mouse Model. J Alzheimer's Dis. 2019;70(2):397–410.

31. Donaldson EE, Stanley D, Hughes RJ, Moore RJ. The time-course of broiler intestinal microbiota development after administration of cecal contents to incubating eggs. PeerJ. 2017;2017(7):1–19.

32. Kubasova T, Kollarcikova M, Crhanova M, Karasova D, Cejkova D, Sebkova A, et al. Contact

with adult hen affects development of caecal microbiota in newly hatched chicks. PLoS One. 2019;14(3):1–13.

33. Kers JG, Velkers FC, Fischer EAJ, Hermes GDA, Stegeman JA, Smidt H. Host and environmental factors affecting the intestinal microbiota in chickens. Front Microbiol. 2018;9(FEB):1–14.

34. Dizaji BR, Hejazi S, Zakeri A. Effects of dietary supplementations of prebiotics, probiotics, synbiotics and acidifiers on growth performance and organs weights of broiler chicken. Eur J Exp Biol. 2012;2(6):2125–9.

35. Crhanova M, Hradecka H, Faldynova M, Matulova M, Havlickova H, Sisak F, et al. Immune response of chicken gut to natural colonization by gut microflora and to Salmonella enterica serovar enteritidis infection. Infect Immun. 2011;79(7):2755–63.

36. Danzeisen JL, Kim HB, Isaacson RE, Tu ZJ, Johnson TJ. Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. PLoS One. 2011;6(11).

37. Carrasco JMD, Casanova NA, Miyakawa MEF. Microbiota, gut health and chicken productivity: What is the connection? Microorganisms. 2019;7(10):1–15.

38. Yan W, Sun C, Yuan J, Yang N. Gut metagenomic analysis reveals prominent roles of Lactobacillus and cecal microbiota in chicken feed efficiency. Sci Rep. 2017;7(October 2016):1–11.

39. Stanley D, Denman SE, Hughes RJ, Geier MS, Crowley TM, Chen H, et al. Intestinal microbiota associated with differential feed conversion efficiency in chickens. Appl Microbiol Biotechnol. 2012;96(5):1361–9.

40. Torok VA, Ophel-Keller K, Loo M, Hughes RJ. Application of methods for identifying broiler chicken gut bacterial species linked with increased energy metabolism. Appl Environ Microbiol. 2008;74(3):783–91.

41. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? DMM Dis Model Mech. 2015;8(1):1–16.

42. Riboulet-Bisson E, Sturme MHJ, Jeffery IB, O'Donnell MM, Neville BA, Forde BM, et al. Effect of lactobacillus salivarius bacteriocin ABP118 on the mouse and pig intestinal microbiota. PLoS One. 2012;7(2).

43. Ward NL, Pieretti A, Dowd SE, Cox SB, Goldstein AM. Intestinal aganglionosis is associated with early and sustained disruption of the colonic microbiome. Neurogastroenterol Motil. 2012;24(9).

44. Cho I, Yamanishi S, Cox L, Methé BA, Zavadil J, Li K, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. Nature. 2012;488(7413):621–6.

45. Nagy-Szakal D, Ross MC, Dowd SE, Mir SAV, Schaible TD, Petrosino JF, et al. Maternal micronutrients can modify colonic mucosal microbiota maturation in murine offspring. Gut Microbes. 2012;3(5).

46. Ubeda C, Bucci V, Caballero S, Djukovic A, Toussaint NC, Equinda M, et al. Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. Infect Immun. 2013;81(3):965–73.

47. Zenewicz LA, Yin X, Wang G, Elinav E, Hao L, Zhao L, et al. IL-22 Deficiency Alters Colonic Microbiota To Be Transmissible and Colitogenic. J Immunol. 2013;190(10):5306–12.

48. Pan D, Yu Z. Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes. 2014;5(1):108–19.

49. Smith K, McCoy KD, Macpherson AJ. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin Immunol. 2007;19(2):59–69.

50. Chervonsky A V. Microbiota and autoimmunity. Cold Spring Harb Perspect Biol. 2013;5(3):1–11.

51. Helgeland L, Vaage JT, Rolstad B, Halstensen TS, Midtvedt T, Brandtzaeg P. Regional phenotypic specialization of intraepithelial lymphocytes in the rat intestine does not depend microbial colonization. Scand J Immunol. 1997;46(4):349–57.

52. Helgeland L, Vaage JT, Rolstad B, Midtvedt T, Brandtzaeg P. Microbial colonization influences composition and T-cell receptor Vβ repertoire of intraepithelial lymphocytes in rat intestine. Immunology. 1996;89(4):494–501.

53. Umesaki Y, Setoyama H, Matsumoto S, Okada Y. Expansion of αβ T-cell receptor-bearing intestinal intraepithelial lymphocytes after microbial colonization in germ-free mice and its independence from thymus. Immunology. 1993;79:32–7.

54. Bandeira A, Mota-Santos T, Itohara S, Degermann S, Hausser C, Tonegawa S, et al. Localization of γ/δ T cells to the intestinal epithelium is independent of normal microbial colonization. J Exp Med. 1990;172(1):239–44.

55. Ismail AS, Behrendt CL, Hooper L V. Reciprocal Interactions between Commensal Bacteria and γδ Intraepithelial Lymphocytes during Mucosal Injury. J Immunol. 2009;182(5):3047–54.

56. Rath S, Rud T, Karch A, Pieper DH, Vital M. Pathogenic functions of host microbiota. Microbiome. 2018;6(1):1–13.

57. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492–506.

58. Ismail AS, Severson KM, Vaishnava S, Behrendt CL, Yu X, Benjamin JL, et al. Γδ Intraepithelial Lymphocytes Are Essential Mediators of Host-Microbial Homeostasis At the Intestinal Mucosal Surface. Proc Natl Acad Sci U S A. 2011;108(21):8743–8.

59. Hoytema van Konijnenburg DP, Reis BS, Pedicord VA, Farache J, Victora GD, Mucida D. Intestinal Epithelial and Intraepithelial T Cell Crosstalk Mediates a Dynamic Response to Infection. Cell. 2017;171(4):783-788.e13.

60. DEFRA. Zoonoses Report 2012. 2013;(September):31–2.

61. Wegener HC. Epidemiology of Campylobacter jejuni infections in the United States and other industrializd nations. 2nd ed. Washington, USA: ASM International; 2000. 121–138 p.

62. Silva J, Leite D, Fernandes M, Mena C, Gibbs PA, Teixeira P. Campylobacter spp . as a foodborne pathogen : a review. 2011;2(September):1–12.

63. Robinson AD. Infective dose of Campylobacter jejuni in milk. 1981;282(May).

64. Solomon EB, Hoover DG. Campylobacter jejuni: A bacterial paradox. J Food Saf. 1999;19(2):121–36.

65. Man SM. The clinical importance of emerging Campylobacter species. Nat Rev Gastroenterol Hepatol. 2011;8(12):669–85.

66. Nachamkin I, Allos BANM. Campylobacter Species and Guillain-Barre ´ Syndrome.

1998;11(3):555–67.

67. Nyati KK, Nyati R. Role of Campylobacter jejuni Infection in the Pathogenesis of Guillain-Barré Syndrome : An Update. 2013;2013.

68. Pope JE. Campylobacter Reactive Arthritis : A Systematic Review. 2010;37(1):48–55.

69. Allos BM. Campylobacter jejuni Infections : Update on Emerging Issues and Trends. 2001;32(August):1201–6.

70. Shane SM, Shane SM. The significance of campylobacter jejuni infection in poultry : A review The significance of Campylobacter jejuni infection in poultry : a review. 1992;9457.

71. Stern NJ, Stan Baily J, Blankenship LC, Cox NA, McHan F. Colonization Characteristics of Campylobacter jejuni in Chick Ceca. 1988;32(2):330–4.

72. Humphrey S, Chaloner G, Kemmett K, Davidson N, Williams N, Kipar A, et al. Campylobacter jejuni is not merely a commensal in commercial broiler chickens and affects bird welfare. MBio. 2014;5(4):1–7.

73. Hiett KL, Stern NJ, Cox NA, Musgrove MT, Ladely S. Molecular Subtype Analyses of Campylobacter spp . from Arkansas and California Poultry Operations. 2002;68(12):6220–36.

74. Thomas LM, Long KA, Good RT, Panaccio M, Widders PR. Genotypic Diversity among Campylobacter jejuni Isolates in a Commercial Broiler Flock. 1997;63(5):1874–7.

75. Emanuelson U, Engvall A. A l-year epidemiological study of campylobacters in 18 Swedish chicken farms. 1995;26(95):167–85.

76. Jacobs-Reitsma WF, Giessen AWVD, Bolder NM, Mulder RWAW. Epidemiology of Campylobacter spp . at two Dutch broiler farms. 1995;413–21.

77. Conlan AJK, Coward C, Grant AJ, Maskell DJ, Gog JR. Campylobacter jejuni colonization and transmission in broiler chickens : a modelling perspective. 2007;(May):819–29.

78. Van Gerwe TJWM, Bouma A, Jacobs-Reitsma WF, Van Den Broek J, Klinkenberg D, Stegeman JA, et al. Quantifying transmission of Campylobacter spp. among broilers. Appl Environ Microbiol. 2005;71(10):5765–70.

79. Smith CK, AbuOun M, Cawthraw SA, Humphrey TJ, Rothwell L, Kaiser P, et al. Campylobacter colonization of the chicken induces a proinflammatory response in mucosal tissues. FEMS Immunol Med Microbiol. 2008;54(1):114–21.

80. Hermans D, Van Deun K, Messens W, Martel A, Van Immerseel F, Haesebrouck F, et al. Campylobacter control in poultry by current intervention measures ineffective: Urgent need for intensified fundamental research. Vet Microbiol. 2011;152(3–4):219–28.

81. Colles FM, Hedges SJ, Dixon R, Preston SG, Thornhill P, Barfod KK, et al. Deep sequencing reveals Campylobacter in commercial meat chickens less than 8 days old. bioRxiv. 2020;1–14.

82. Rawson T, Paton RS, Colles FM, Maiden MCJ, Dawkins MS, Bonsall MB. A Mathematical Modeling Approach to Uncover Factors Influencing the Spread of Campylobacter in a Flock of Broiler-Breeder Chickens. Front Microbiol. 2020;11(October):1–17.

83. Höök H, Fattah MA, Ericsson H, Vågsholm I, Danielsson-Tham ML. Genotype dynamics of Campylobacter jejuni in a broiler flock. Vet Microbiol. 2005;106(1–2):109–17.

84. Chaloner G, Wigley P, Humphrey S, Kemmett K, Lacharme-lora L, Humphrey T, et al. Dynamics of Dual Infection with Campylobacter jejuni Strains in Chickens Reveals Distinct Strain-to-

Strain Variation in Infection Ecology. 2014;80(20):6366–72.

85. El-Shibiny A, Connerton PL, Connerton IF. Campylobacter succession in broiler chickens. Vet Microbiol. 2007;125(3–4):323–32.

86. Korolik V, Alderton MR, Smith SC, Chang J, Coloe PJ. Isolation and molecular analysis of colonising and non-colonising strains of Campylobacter jejuni and Campylobacter coli following experimental infection of young chickens. Vet Microbiol. 1998;60(2–4):239–49.

87. Skånseng B, Trosvik P, Zimonja M, Johnsen G, Bjerrum L, Pedersen K, et al. Co-infection dynamics of a major food-borne zoonotic pathogen in chicken. PLoS Pathog. 2007;3(11):1790–7.

88. Hendrixson DR, DiRita VJ. Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. Mol Microbiol. 2004;52(2):471–84.

89. Lacharme-Lora L, Chaloner G, Gilroy R, Humphrey S, Gibbs K, Jopson S, et al. B lymphocytes play a limited role in clearance of Campylobacter jejuni from the chicken intestinal tract. Sci Rep. 2017;7.

90. Thibodeau A, Fravalo P, Yergeau E, Arsenault J, Lahaye L, Letellier A. Chicken Caecal Microbiome modifications induced by campylobacter Jejuni colonization and by a non-antibiotic feed additive. PLoS One. 2015;10(7):1–14.

91. Johansen CH, Bjerrum L, Finster K, Pedersen K. Effects of a Campylobacter jejuni infection on the development of the intestinal microflora of broiler chickens. Poult Sci. 2006;85(4):579–87.

92. Kaakoush NO, Sodhi N, Chenu JW, Cox JM, Riordan SM, Mitchell HM. The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. Gut Pathog. 2014;6(1):1–10.

Department of Zoology, University of Oxford

Chapter 2

Materials and Methods

2.1 General Sample Processing Workflow

In order to facilitate the TCR repertoire and 16S rRNA sequencing protocols used throughout this project, samples, including tissue, faecal, ileal contents and caecal contents, were processed using a standard pipeline developed as part of this thesis (Fig 2.1). DNA for 16S sequencing was extracted from tissue, faecal and ileal/caecal contents samples and RNA for repertoire sequencing was extracted from tissue samples only. RNA was converted to cDNA using 5' Rapid Amplification of cDNA Ends (RACE) 16S and repertoire PCRs were then performed in triplicate. PCRs were pooled and prepared for sequencing using a library preparation protocol followed by qPCR and Tapestation analysis, which act as quality control steps. The amplicons were then sequenced and Illumina reads first processed by demultiplexing which involves splitting the sequencing reads by primer and barcode into separate files for processing. 16S samples were then processed using DADA2, which involved denoising, read merging, Amplicon Sequence Variant (ASV) table construction and chimera removal before hand off to the analysis packages Phyloseq and Vegan among others. Additionally, any 16S sequences belonging to genera found in the controls are removed, generally *Halomonas* and various green algae. Repertoire sequences have their T receptor gamma (TRG) V, J and C genes assigned by BLAST searching against a database followed by denoising and filtering of non-coding CDR3s. The experimental methods involved in this processing pipeline are described below, whereas the data processing techniques are described in greater detail in the Method Development chapter (See Chapter 3). This is due to the fact that there is no 'standard' analysis for repertoire and 16S data and therefore many possible ways in which it can be processed. As a result, it is important to discuss and justify any data processing steps as they have possible implications for downstream interpretation of data and may inform future work.

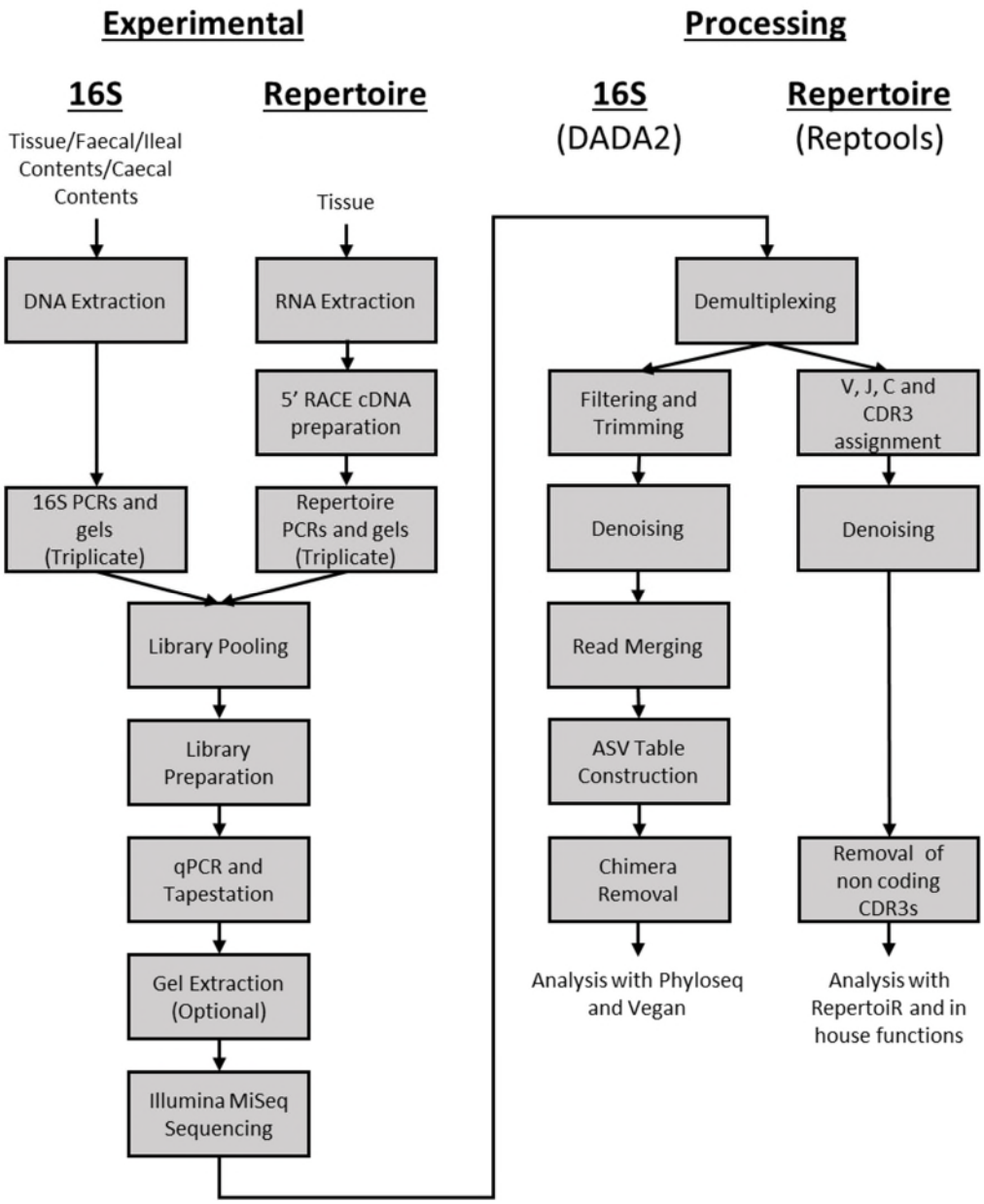


Figure 2.1 - Workflow for 16S rRNA and Repertoire Sequencing
Schematic shows workflow for 16S rRNA and repertoire sequencing studies. On the experimental side of the 16S rRNA work, DNA is first extracted from tissue, faecal and ileal/ceacal contents, followed by amplification using 16S PCR primers. Repertoire amplification is performed on tissue samples only and includes RNA extraction and 5' RACE cDNA preparation, enabling downstream universal repertoire amplification using UPA primers. The PCRs from both 16S and repertoire are then pooled into their respective libraries before being processed for sequencing on the Illumina MiSeq platform (in house). The resulting reads are then demultiplexed, which involves splitting the MiSeq files by primer and barcode. The 16S data is then processed by DADA2, which filters, trims, denoises and merges reads then constructs an ASV table and removes chimeric sequences before handing off the data to Phyloseq, Vegan and other R packages which are used for analysis. The repertoire side of data processing begins with BLAST assignment of V, J and C gene segments and CDR3 assignment. Reads are then denoised using VDJtools, non-coding CDR3s removed and the resulting reads passed on to our in house R package RepertoiR, among other packages, for analysis.

2.2 Shared Methods

Note: All animal handling work described in this chapter was performed by our collaborators at INRAE – France, Veterinary Research Institute – Czech Republic, TUM – Munich and The Francis Crick Institute – UK.

2.2.1 Tissue RNA Extraction

Tissue samples in RNAlater were first defrosted on ice. Pieces from each sample were then excised and weighed, using a fresh sterile scalpel for each. The removed pieces were then added to 350µl RLT plus with 7µl 2M DTT and 1.75µl DX reagent, and ~50µl 0.1mm glass beads (Thistle Scientific). The sample was then homogenised in a bead beater on full power for 8x 30 seconds with 30 seconds on ice between bursts. The lysate was then spun at ~16,000xg at room temperature and the RNeasy spin column protocol used according to manufacturer's instructions. DNase I digestion was performed on column. The RNA elution step was repeated twice using 40µl RNase free water (Qiagen). RNA concentration and quality, based on 28S and 18S ratio, was then measured using Agilent Tapestation RNA ScreenTape with an electronic ladder according to manufacturer's instructions.

2.2.2 Faecal/Ileal Contents/Caecal Contents DNA Extraction

150mg of material, 150µl Nuclease Free (NF) H_2O, 200µl 0.1mm glass beads (Thistle Scientific) and 600µl Lysis buffer (Maxwell Blood DNA Purification Kit) was added to a screw cap tube and disrupted for 2 mins using a bead beater at 3800 strokes/min and placed on ice for 1 min. The samples were then centrifuged at 13,000rpm for 2 mins to remove bubbles and pellet beads and debris. 400µl of supernatant was then added to a fresh tube and incubated at 56°C for 20 min. The resulting suspension was then spun at 13,000rpm for 2 mins and 300µl added to well 1 of the Maxwell Blood DNA Purification Kit strip tube which was then run using the 'Blood Protocol'. Dilution was then performed in NF H_2O according to previous PCR success to surmount the issue of PCR inhibitors from faecal contaminants (See Chapter 3)[1,2].

2.2.3 Qubit Nucleic Acid Quantification

Nucleic acid quantification using Qubit 2.0 fluorimeter (ThermoFisher) was performed according to manufacturer's instructions. Briefly, Qubit working solution was made up from reagents belonging to the relevant Assay Kit (DNA BR/HS or RNA BR/HS). 10µl (or less depending on estimated concentration and available volume) sample was added to the working solution alongside the Qubit standards. These solutions were then vortexed for 2-3 seconds and incubated at room temperature for 2 minutes. Standards were then used to calibrate the device. Samples were then measured and final sample concentration calculated using the working solution dilution factor.

2.2.4 cDNA Synthesis

5' RACE cDNA was synthesised from RNA using the Takara Bio SMARTer reverse transcriptase kit. This method of converting RNA to cDNA has the crucial advantage that it enables universal amplification of all cDNA molecules via the addition of a common 5' sequence (Fig. 2.2). This enables unbiased amplification of the entire immune repertoire in a single PCR reaction.

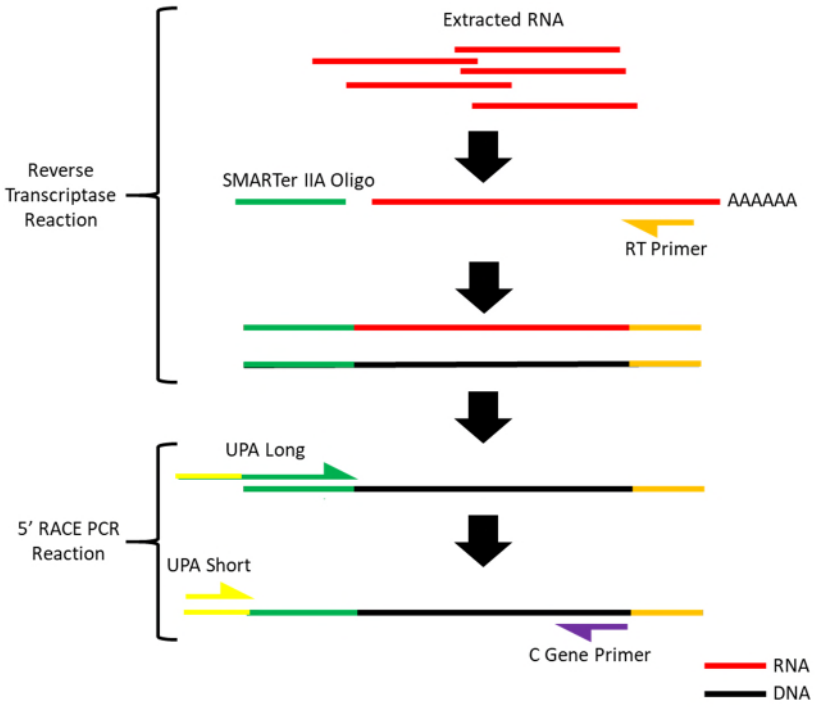


Figure 2.2 – SMARTer Reverse Transcription and Universal 5' RACE Repertoire PCRs
Universal amplification of all TCR gamma rearrangements, including every TRGV gene, was achieved using a standard 5' RACE strategy. The template switching activity of SMARTScribe Reverse Transcriptase (Takara Bio) is utilised to anneal the SMARTer IIA Oligo onto the 5' end of the cDNA molecules produced during the reverse transcription step. A 2 stage step-down PCR reaction is then used to first amplify cDNA molecules using the UPA long primer only, which produces products with the UPA long sequence at the 5' end. During the second stage, a UPA short primer anneals to a region of the UPA long sequence and enables targeted amplification of the TCR gamma repertoire when used in combination with a primer complementary to the TCR gamma C gene. As the UPA long sequence is at the 5' end of every cDNA molecule regardless of TRGV gene, use of UPA short enables unbiased amplification of the entire TCR gamma repertoire in a single reaction.

Master mix was made up as follows:

Component	Volume (µl)
5x First Strand Buffer	2
DTT 20mM	1
dNTP Mix (10mM)	1
SMARTer IIA oligo (12µM)	1
RNAse Inhibitor (20U/µl)	0.5
-For RT +ve: SMARTScribe Reverse Transcriptase (100U)	1
-For RT –ve: Nuclease Free H_2O	1

1µl 5'-CDS Primer A was then added to 2.75µl of each sample RNA and incubated at 72°C for 3 min, then cooled to 4°C for 2 min. The RNA, CDS primer mix was then added to the master mix and incubated at 42°C for 90 mins then 70°C for 10 mins in a Thermocycler. 20µl Tris-EDTA was then added to the resulting mix and the samples stored at -20°C.

2.2.5 Primers/barcoding

Primer	Sequence (5'-3')
5'CDS	TTTTTTTTTTTTTTTTTTTTTTTTVN
SMARTer oligo	AAGCAGTGGTATCAACGCAGAGTACATGGG
UPA short	CTAATACGACTCACTATAGGGC
UPA long	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
MuMGV7 Specific	TATCATGTTTATGAAGGCCCG
MuMGC*	CTTATGGAGATTTGTTTCAGCA
ChGC	GGTAGAATTTCTCAATAAGACAAACATACACCACTT
E338*	ACTCCTACGGGAGGCAGCAGT
R806	GGACTACHVGGGTWTCTAAT
MompB	CCACAATTATGGTTAGCTTA
Momp2R	TGAGAAGTTAAGTTTTGGAGAG
hipO-F - (Persson05)**	GACTTCGTGCAGATATGGATGCTT
hipO-R - (Persson05)**	GCTATAACTATCCGAAGAAGCCATCA
CC18F - (Persson05)**	GGTATGATTTCTACAAAGCGAG
CC519R - (Persson05)**	ATAAAAGACTATCGTCGCGTG
16S_cjcc-F - (Persson05)**	GGAGGCAGCAGTAGGGAATA

16S_cjcc-R - (Persson05)**	TGACGGGCGGTGAGTACAAG

*Primer barcodes were designed based on Hamming codes according to the recommendations set out previously in the literature[3]. Hamming codes are a principle originally used in telecommunications for transmission of information through an error prone medium. Repurposed for use as DNA barcodes, Hamming codes consist of 7 bases in total with 4 making up the barcode itself and the other 3 making up the quality control component (Fig. A1). The quality control part of the barcode enables correction of single base pair changes drastically reducing the risk of misidentification in the event of incorrect base calls made during sequencing.

**Primers were combined according to Persson 2005 and used as multiplex primers[4]

2.2.6 PCR

2.2.6.1 Master Mixes

PCR master mix was made up as below:

2.2.6.1.1 UPA (MuMGC or ChGC)

Component	**Volume (µl)**
Nuclease Free H_2O	16.75
5x High Fidelity Buffer	5
dNTP	0.5
Phusion	0.25
10µM UPA S	0.5
2µM UPA L	0.5
10µM MuMGC/ChGC (Barcoded)*	0.5
cDNA	1

2.2.6.1.2 MuMGV7 Specific

Component	**Volume (µl)**
Nuclease Free H_2O	17.25
5x High Fidelity Buffer	5
dNTP	0.5
Phusion	0.25
10µM V7	0.5
10 µM MuMGC (Barcoded)*	0.5
cDNA	1

2.2.6.1.3 16S

Component	**Volume (µl)**
Nuclease Free H_2O	17.25
5x High Fidelity Buffer	5
dNTP	0.5

Phusion	0.25
10µM E338*	0.5
10µM R806	0.5
cDNA	1

2.2.6.1.4 Momp (PorA)

Component	**Volume (µl)**
Nuclease Free H_2O	11.75
5x High Fidelity Buffer	5
dNTP	0.5
Phusion	0.25
10µM Momp B	1.25
10µM Momp 2R	1.25
DNA Template	5

2.2.6.1.5 Persson

Component	**Volume (µl)**
Nuclease Free H_2O	15.875
10x Taq Buffer	2.5
dNTP	0.5
Taq Polymerase	0.125
Persson05 multiplex mix	1
DNA Template	5

*A unique barcode was used for each PCR within a library. Different barcodes were used for each sample between triplicate repeats/libraries.

2.2.6.2 Running Conditions

PCR was run on a thermocycler using the following conditions.

2.2.6.2.1 UPA (MuMGC or ChGC)

Stage	**Cycles**	**Temperature (°C)**	**Time (m:s)**
1	1	98	1:00
2	5	98	0:15
		72	0:30
3	5	98	0:15
		70	0:15
		72	0:30
4	30	98	0:15
		62	0:15
		72	0:30
5	1	72	2:00

		4	∞

2.2.6.2.2 MuMGV7 Specific

Stage	Cycles	Temp (°C)	Time (m:s)
1	1	98	1:00
2	35	98	0:15
		60	0:15
		72	0:30
3	1	72	2:00
		4	∞

2.2.6.2.3 16S

Stage	Cycles	Temperature (°C)	Time (m:s)
1	1	98	1:00
2	35	98	0:15
		60	0:15
		72	0:30
3	1	72	2:00
		4	∞

PCR products were then stored at -20°C

2.2.6.2.4 Momp

Stage	Cycles	Temperature (°C)	Time (m:s)
1	1	98	0:30
2	35	98	0:10
		64	0:30
		72	0:30
3	1	72	10:00
		4	∞

2.2.6.2.5 Persson

Stage	Cycles	Temperature (°C)	Time (m:s)
1	1	94	6:00
2	35	94	0:50
		57	0:40
		72	0:50
3	1	72	3:00
		4	∞

2.2.7 Agarose Gel

Gels were run at 120V for 25 minutes using 1.5% agarose with 1:25000 Ethidium Bromide alongside a 100bp NEB ladder.

2.2.8 Library Mixing

The concentration of each PCR product was roughly estimated by comparison to the ladder bands. The volume of product required to add an equal amount of DNA from each amplicon was then calculated and the products pooled accordingly.

2.2.9 Gel Extraction

Where necessary, a gel extraction was then performed in order to excise the band of interest and remove unwanted contaminants. This was done by running a gel at 120V for 45 minutes using 1.5% agarose with 1:10000 SYBR safe alongside a 100bp NEB ladder. A UV light box was then used to visualise the gel whilst the bands of interest were excised using a scalpel. Exposure time of the gel to UV light was minimised and the gel slice was cut as small as possible. Gel extraction was then performed using the Qiagen QIAquick Gel Extraction Kit according to manufacturer's instructions.

2.2.10 Library Preparation

Library preps were then performed using the NEB NEBNext Ultra DNA Library Prep Kit for Illumina according to manufacturer's instructions. AMPure XP beads were used for size selection. Generally as few cycles as possible were used for the PCR amplification stage in order to reduce the chances of introducing PCR artefacts into the library.

2.2.11 Library Quantification

To obtain accurate library concentrations, which are necessary for mixing libraries for the sequencing, library quantification was performed using the NEBNext Library Quant Kit for Illumina according to manufacturer's instructions. The concentration of each library was then calculated using the NEB online NEBioCalculator. In order to ensure these calculations were accurate, it was ensured that amplification efficiency be over 90% and R^2 >0.98.

2.2.12 Tapestation

To confirm the libraries were free of contaminants and contained products of the size expected, the Agilent D1000 ScreenTape System was used according to manufacturer's instructions. Particular care was taken to exclude libraries with high amounts of adapter dimer (~125bp) as it competes for slots on the flow cell, reducing the amount of sequencing available for actual libraries.

2.2.13 Library Blending

Libraries were then blended according to Illumina recommendations. The desired run percentage attributed to each library was decided based on the estimated number of reads required to achieve a good level of coverage. The final concentration of the library was then decided based on the concentrations of the libraries to be mixed. This concentration was kept as high as possible and potential values ranged from 4-0.5nM. The desired volume of each library to be added to the final pool was then calculated. Any library that had a volume to be added less than 2µl was subsequently diluted, allowing for increased pipetting volume and thus increased accuracy.

2.2.14 MiSeq Loading - Denature and Dilute

In order to prepare libraries for the sequencing run, the Illumina Denature and Dilute Libraries guide was followed, for 4nM libraries, the MiSeq protocol was used, for libraries 2-0.5nM the NextSeq protocol was used. Libraries were loaded on the MiSeq at 10pM and 10% PhiX used as a sequencing control.

2.3 Chapter 4 Specific Methods

2.3.1 Gamma Locus Annotation

V genes were identified by performing BLAST searches of known sequences already present in the galGal6a genome. These prospective V genes were then confirmed by a number of factors. Existence of a plausible Restriction Signal-Sequence (RSS), assessed by Recombination Information Content (RIC) score using RSSsite[5]. Additionally, Interpro's protein prediction tool was used to

ensure the sequence corresponded to an Ig-like like motif[6,7]. Finally, to ensure no V genes had been missed, the sequences obtained from UPA sequencing were realigned to the gamma locus using Bowtie2[8] and the results visualised using SeqMonk[9]. V gene families were assigned using the Ig-like domain predicted by InterPro. Using the IMGT subgroup concept, which requires 75% sequence identity at the nucleotide level, we assigned V gene families[10]. V genes within these families were numbered according to their position in the locus, distal to proximal with respect to the C gene. Sequence identity was determined by the following formula

$$100 \times \frac{Identical\ Positions}{Length\ Shorter\ Sequence}$$

2.4 Chapter 5 Specific Methods

2.4.1 Experimental Chickens

2.4.1.1 Germ Free, Tetra, Complex and Conventional TCR gamma repertoire birds

The birds used in this experiment were of the INRAE PA-12 line, an outbred white Leghorn line with the heterozygous MHC haplotype B2/1. Germ Free (GF) birds were generated by sterilisation of eggs immediately following laying using 1.5% Divosan Plus VT53 (Johnson Diversey, France), eggs were then transferred into a sterile HEPA-filtered incubator and after 18 days were again surface sterilised as previously described. Hatched chicks were fed an X-ray sterilised standard diet and had access to sterile water ad libitum. GF status was confirmed by collection of faecal samples every 7 days, which was then cultured in brain-heart infusion, a rich growth medium. A lack microbial growth was taken to be confirmation that the birds had no microflora and therefore remained GF. Upon experimental termination, 16S qPCR was performed on faecal and caecal samples to check for the presence of non-culturable microbial contaminants. Complex birds were generated by re-conventionalisation of GF birds by gavage of caecal contents collected from a single healthy SPF hen. Tetra birds were generated by inoculation with a defined mixture of *Eschericia coli* Nissle 1917, *Enterococcus faecium* DSM 7134, *Lactobacillus rhamnosus* DSM 7133 and *Clostridium*

butyricum DSM 10702. Conventional birds were raised in a Specific Pathogen Free (SPF) facility. Complex, Tetra and Conventional birds were all fed on a standard diet and had access to water ad libitum. Each group contained 5 birds, except the Tetra, which contained 3. Tetra and Conventional birds were raised in the same facility at different times to Complex and GF birds.

2.4.1.2 TCRγ-/- 16S birds

Knockout chickens were produced from TCRγ +/- heterozygous Brown and White Leghorn parents generated using homologous recombination of germ cells[11]. In order to ensure a normalised starting microbiota, three days prior to hatch, the mother hens were transferred to the aviary which was to house the experimental birds. These hens were removed the day of hatch and the litter mixed to allow colonisation from maternal faecal microbiota. All chicks were then housed together and genotyped within the first three days of life. After 1 week the Wild Type, TCRγ +/- and TCRγ-/- birds were separated to be co-housed according to group. Each of these genetic groups contained 5 males and 5 females. All birds were fed on a standard diet and had access to water ad libitum. Chickens were tagged and faecal samples collected such that each faecal sample could be traced back to the bird of origin. The birds were then sacrificed by cervical dislocation on Day 65 post-hatch and caecal and ileal contents collected.

2.5 Chapter 6 Specific Methods

2.5.1 Culturing and Mixing of *Campylobacter* Strains

Campylobacter strains were first isolated from farm chickens by plating on charcoal agar and frozen down as glycerol stocks at -80°C. These isolates then had their genomic DNA extracted using the Maxwell 16 Cell DNA purification kit according to manufacturer's instructions. The PorA gene of each of these gDNA samples was then amplified and Sanger sequenced. The resulting Sanger trace was then used to determine the PorA type of each isolate. The glycerol stocks were then partially defrosted to allow the top layer to be scraped off and subbed onto several blood agar plates. The

plates were then incubated at 42°C in a microaerophilic environment generated using Genbox Microaer Gas pouches (biomerieux) in an airtight box.

After 2 days of growth, these plates were scraped into BHI broth and diluted to an OD600 of 0.6 using a spectrophotometer. The cultures had their genomic DNA extracted as previous, then were serially diluted and plated onto charcoal agar plates which were incubated as before. The genomic DNA was used for PorA PCRs and Sanger sequencing as previous to confirm their identity. Additionally, PCRs designed to discriminate between *C. jejuni* and *C. coli* were performed. The number of colonies was then counted from the serially diluted plates and CFU/ml calculated. These CFU calculations were used to mix the strains evenly according to the table below

CIG1		**CIG2**	
Isolate	**PorA**	**Isolate**	**PorA**
784.11	PorA 11	784.18	PorA 400
784.12	PorA 44	5643.1	PorA 108
784.18	PorA 400	5814.1	PorA 36
784.20	PorA 33	6235.10	PorA 25

2.5.2 Experimental Chickens and *Campylobacter* Gavage

Male ISA Brown chickens were obtained on the first day of hatch from a commercial hatchery. Birds were treated with Broilact (Nimrod Veterinary Products Ltd.), a chicken start of life probiotic, to ensure a consistent starting microbiota. On Day 7, the chickens were split into the following *Campylobacter* infection groups: Mono-Infected Group (MIG), which was infected by a single strain, Co-Infection Group 1 (CIG1), which was infected with a mix of 4 different strains all isolated from the same farm bird, Co-Infection Group 2 (CIG2), which was infected with a mix of 4 different strains all isolated from different birds and Control (Ctrl) which was not infected with *Campylobacter* (Chapter 6, Fig 6.1A). Each of these groups was gavaged with $5x10^5$ CFU total of the appropriate *Campylobacter* mix in 100µl Brain Heart Infusion (BHI) broth. Control birds were given just BHI broth.

Successful *Campylobacter* colonisation was confirmed by qPCR and 16S sequencing. The chickens then had individual faecal samples taken on days 8, 10, 15, 17, 21, 28, 35, 42 and 49 allowing the microbiota to be followed over time in any given bird. The birds were then sacrificed on Day 49 and their thymus, spleen, duodenum, jejunum, ileum, caeca and colon sampled for downstream repertoire sequencing. 8 birds were used per group and were fed on a standard diet with access to water ad libitum.

2.5.3 Sanger Sequencing

Using a multi-channel pipette, 60µl of 20% PEG-8000/2.5M NaCl was added to each well of the PCR plate containing the samples to be sequenced. The contents were then mixed using a vortex mixer and briefly centrifuged at 500g. The plate was then incubated at room temperature for 30 minutes. Samples to be sequenced were then pelleted by centrifugation for 1 hour at 2750g (4°C). Supernatants were then discarded by gently inverting the plate, and spinning upside down on a folded piece of blue roll at 500×g for 1 minute. The DNA pellet was washed twice by adding 150µl of cold 70% ethanol and spinning at 2750g for a further 5 minutes, supernatants were then discarded as previous (spinning inverted on to blue roll).

PCR pellets were resuspended by adding 10-30µl Milli-Q water to each of the wells using a multi-channel pipette, re-sealed and vortexed for 10 – 20 seconds. Plates were then briefly centrifuged at 500g to ensure all the liquid was at the bottom of the wells.

Sequencing master mix was then prepared containing the reagents as below:

Reagent	**Volume (µl)**
BigDye	0.25
5× dilution buffer	2.0
Milli-Q water	1.75
10µM primer (diluted 1:15)	4

8µl of the sequencing master mix was then pipetted into the bottom of the wells of a 96-well PCR plate and 2µl of DNA template from the resuspended PCR plate added to the top wall of each of the wells. The plate was then vortexed briefly and the run on a thermocycler as below.

Cycles	Temperature (°C)	Time (m:s)
30	96	0:10
	50	0:05
	60	2:00

The plate was removed from the thermocycler and briefly centrifuged, 10µl of Milli-Q water was then added to each well.

7.0ml absolute ethanol was mixed with 280µl of 3M sodium acetate, pH 4.6 in a small reagent reservoir and 52µl of the mix added to each of the wells of the PCR plate. The plate was then resealed, vortexed briefly to mix and spun briefly at 500g before being incubated at room temperature for 30-45 min. The plate was then spun at 2750g for 1 hour (4°C) and the supernatant discarded as previous. The pellets were then washed by adding 150µl of cold 70% ethanol and spinning at 2750g for a further 5 minutes. Supernatants were discard as before. Pellets were then Sanger sequenced on the 3750xl DNA analyser (Applied Biosciences).

2.6 Chapter 7 Specific Methods

2.6.1 Experimental Mice

C57/BL6 mice were purchased from Charles River, Btnl1 -/- and TCRδ-/- mice were obtained from the international mouse phenotyping consortium (IMPC) and Nude mice were purchased from the Jackson Laboratory. These mice were raised in the Francis Crick institute and fed a standard diet with access to water ad libitium and caged according to genotype. Mice were sacrificed and tissues harvested once they reached 10 weeks of age.

2.6.2 RNA/DNA Simultaneous Extraction from Tissue

RNA and DNA was simultaneously extracted using Qiagen AllPrep micro columns according to manufacturer's instructions. ~5mg of tissue was taken and added to 350µl RLT plus with 7µl 2M DTT and 1.75µl DX reagent in a screwcap 2ml tube with ~50µl 0.1mm glass beads. The sample was then homogenised in a bead beater on full power for 8x 30 seconds with 30 seconds on ice between bursts. DNA/RNA extraction was then performed according the manufacturer's instructions using the minimum elution volume in order to maximise nucleic acid concentration for subsequent steps.

RNA and DNA concentrations where then measured using their relevant Qubit BR kit. RNA was subsequently assayed for quality, based on 28S and 18S ratio, using gel electrophoresis or Agilent Tapestation RNA ScreenTape.

2.7 References

1. Lantz PG, Matsson M, Wadström T, Rådström P. Removal of PCR inhibitors from human faecal samples through the use of an aqueous two-phase system for sample preparation prior to PCR. J Microbiol Methods. 1997;28(3):159–67.
2. Monteiro L, Bonnemaison D, Vekris A, Petry KG, Bonnet J, Vidal R, et al. Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. J Clin Microbiol. 1997;35(4):995–8.
3. Bystrykh L V. Generalized DNA barcode design based on Hamming codes. PLoS One. 2012;7(5).
4. Persson S, Olsen KEP. Multiplex PCR for identification of Campylobacter coli and Campylobacter jejuni from pure cultures and directly on stool samples. 2005;(54):1043–7.
5. Merelli I, Guffanti A, Fabbri M, Cocito A, Furia L, Grazini U, et al. RSSsite: A reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. 2010;38:262–7.
6. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 2001;29(1):37–40.
7. Smith DK, Xue H. Sequence profiles of immunoglobulin and immunoglobulin-like domains. J Mol Biol. 1997;274(4):530–45.
8. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
9. Andrews S. SeqMonk: A tool to visualise and analyse high throughput mapped sequence data [Internet]. Babraham Bioinformatics. 2007.

10. Lefranc M. Ontology for immunogenetics : the IMGT-ONTOLOGY. Bioinformatics. 1999;15(12):1047–54.

11. Schusser B, Collarini EJ, Yi H, Izquierdo SM, Fesler J, Pedersen D, et al. Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells. Proc Natl Acad Sci U S A. 2013;110(50):20170–5.

Department of Zoology, University of Oxford

Chapter 3

Method Development

3.1 Introduction

Much of the analysis of Next Generation Sequencing (NGS) data from the T cell repertoire (TCR) and 16S is not standard and is still subject to debate in the literature. Therefore, any given analysis should be treated with some level of caution and its use justified to ensure it best suits the data in question. As much of the analyses used throughout this thesis are very similar and discussion of them in each chapter would cause repetition, the thought processes behind them are described here for the readers convenience. This chapter is presented with integrated results and discussion as this format makes it easier to provide a clear narrative describing our reasoning. As it is described before the data chapters, the method development presented here may appear out of context and the reader may instead wish to read these first and refer back to this chapter as necessary.

Repertoire sequencing allows the examination of T cell subsets at extremely high resolution. The rapidly decreasing price of sequencing allows researchers to achieve reasonable levels of coverage at relatively low cost. The use of NGS sequencing means that millions of receptors can be sequenced from a single sample allowing multiple analyses to be conducted, providing a population level insight into the T cell receptor, a key determinant of T cell biology.

We focus our repertoire analysis exclusively on the gamma chain of the γδ receptor. This is because different gamma chains have been shown to localize to different tissues in mouse and therefore we concentrate our efforts on the gamma in order to investigate this phenomenon[1]. Furthermore, during gamma delta T cell development the delta chain rearranges first, followed by the gamma[2]. Therefore the chances of a clone displaying the same delta chain but a different gamma chain is higher than the chances of the reverse event and thus the gamma chain is better able to represent the γδ TCR as a whole. Additionally, in the case of the αβ TCR, the beta chain is most often studied because of its increased potential for combinatorial diversity due to the presence of the D segments as well as V, J and C. However, despite the presence of D segments in the TCR delta locus, only 2 exist, and therefore the combinatorial diversity of the delta locus is 31 (TRDV) x 2

(TRDJ) x 2 (TRDD) x 1 (TRDC)[3] = 134 compared to 40 (TRGV) x 3 (TRGJ) x 1 (TRGC) (see Chapter 4) = 120 for the gamma locus. Therefore, the chicken TCR gamma locus has only slightly less potential diversity than the TCR delta locus and is therefore just as beneficial in terms of capturing the diversity of the γδ TCR.

We focus our analysis on the Complementarity-Determining Region (CDR) 3, the highly variable region resulting from V and J splicing, as it shows the greatest sequence diversity and thus has the greatest power to distinguish clones from one another. The hyper variability of the CDR3 is partly down to recombination derived diversity, due to its location between the V and J gene segments, and partly down to junctional diversity, generated by addition and deletion of P and N nucleotides, mediated by Recombination-activating gene (RAG) and terminal deoxynucleotidyl transferase (TdT), respectively (see Chapter 1). During the recombination of two TCR gene segments and subsequent generation of junctional diversity, the RAG enzyme complex cleaves the DNA hairpin, and adds P nucleotides, followed by the addition of random N nucleotides by TdT. The remaining gaps between the strands which remain at this point are filled by DNA repair machinery producing the mature recombined region with additional junctional diversity. The CDR3 region is identified using the IMGT definition which states that it is located between the 2nd conserved cysteine at the end of the V gene and a phenylalanine residue in a conserved motif at the start of the J gene[4,5].

16S rRNA is a technique involving amplification of a variable but universally present sequence of DNA encoding a section of ribosomal RNA, found in all prokaryotic cells[6]. The 16S rRNA is a part of the 30S subunit of the bacterial ribosome, a cytoplasmic nucleoprotein, which serves as the site of protein synthesis. The 16S section of the ribosome is an effective tool for the identification of bacteria and interrogation of complex microbial communities for a number of reasons. Importantly, it consists of 10 common conserved regions to which primers can be designed, interspersed with 9 hypervariable regions (V1-9), allowing different species to be distinguished from

one another (Fig. 3.1). The relatively short distance between these conserved regions means that they offer convenient primer sites for PCRs spanning the variable regions. The products can then be thoroughly sampled using an amplicon sequencing approach on a short read NGS platform such as the Illumina MiSeq, which is capable of producing millions of reads in a single run. Additionally, as 16S becomes more widely adopted, the number of sequences deposited in publically available databases rapidly increases. As reference to these databases allows 16S sequences to be attributed to a particular species, this allows the accurate identification of thousands of microbes in a complex community. Though other targets for this type of microbial identification are available, such as gyrB, a gene which encodes the B subunit of DNA gyrase, the widescale use of 16S is one of its greatest strengths and has driven the development of multiple reference databases and data processing packages and pipelines. In addition, with gyrB specifically, various studies have shown 16S to have higher resolution[7].

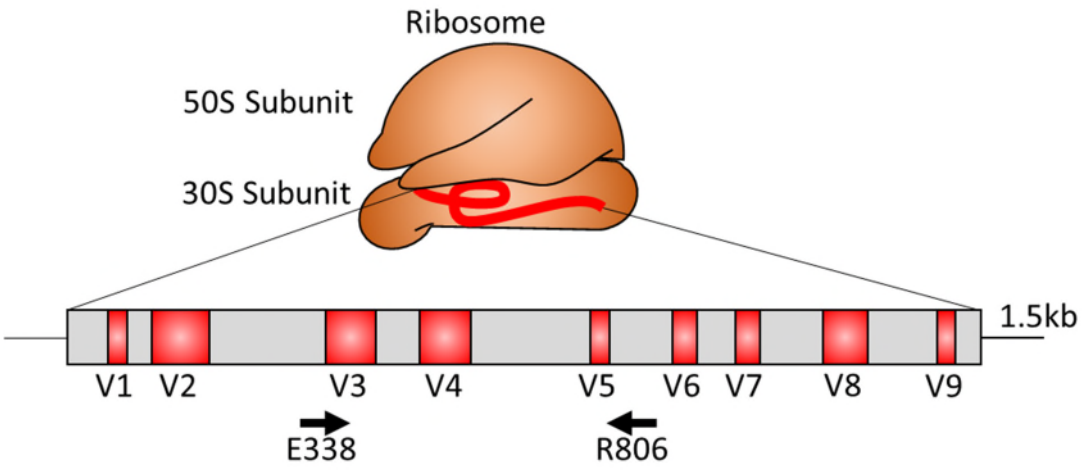


Figure 3.1 - Schematic Showing the 16S Ribosomal RNA Gene
16S ribosomal RNA is located on the 30S subunit of the bacterial ribosome and is represented by a red strand. The gene encoding the 16S rRNA is made up of 9 hypervariable regions (V1-V9) with high levels of sequence diversity, here coloured red. Interspersed between these hypervariable regions are constant regions, shown here in grey, with low levels of sequence diversity. Locations of the binding sites for the forward, E338, and reverse, R806, primers used in this thesis are shown by black arrows.

In some ways, NGS has become a victim of its own success and though error rates are relatively low, when millions of reads are produced by a single run, the dataset will inevitably become contaminated with these sequencing errors. Commonly used 16S pipelines, such as QIIME[8], MOTHUR[9] and USEARCH-UPARSE, originally solved this problem by coining the term Operational Taxonomic Unit (OTU), which is simply a cluster of sequences within a certain sequence identity threshold, normally 97%, and analysing these OTUs as a single group. The obvious problem with this simple approach is that it substantially reduces taxonomic resolution, which may lead to an analysis missing important information. Even within closely related species, bacterial variation in terms of behaviour, lifestyle and gene expression remains high[10–14]. In order to solve this problem, more recent, sophisticated denoising approaches, such as DADA2[15], Qiime2-Deblur[16] and USEARCH-UNOISE3 have been developed. Rather than pooling data together, these approaches attempt to correct the sequencing errors in various ways, in order to provide a more high resolution unit of 16S analysis, the Amplicon Sequence Variant (ASV)[15]. Recent studies show that the DADA2 pipeline provides the greatest sensitivity and specificity and is therefore the package used in this thesis[17,18].

Briefly, DADA2 functions by first inspecting the read quality profiles, and trimming the reads according to these profiles. The trimming length represents a compromise between retaining enough length for later read merging and removing lower quality data, which improves the packages ability to pick out rare sequence variants. DADA2 then uses a model to learn the error rates, which differ with each amplicon dataset. The error rate refers to the rate at which any one sequence is erroneously interpreted as another as a function of read composition and quality. This function uses a machine learning approach, starting with an estimate of the expected error rates and alternating between this and sample composition inference until a convergent solution is reached. Once the error rate model has been estimated, sample inference and calculation of unique sequence variants can begin. This stage involves grouping together all unique sequences and sequentially asking what the probability is of each unique partition arising from another as a result of sequencing errors. If the probability of this occurring is determined to be above a certain threshold, these 2 partitions are

combined and their read counts summed. This process continues until all of the probabilities that one partition arose from another are below the threshold. Importantly, DADA2 does not infer singletons and thus any analysis relying on their presence, eg. diversity, cannot be conducted on DADA2 processed data. Forward and reverse reads are then merged and an ASV table is produced. Finally, chimeric reads are then identified and removed by examining each of the merged reads and asking if any could have resulted from the combination of 2 more abundant parent reads. DADA2 also facilitates an easy transition to ecological packages featuring a variety of useful analysis and graphing functions such as phyloseq[19] and vegan[20].

This chapter aims to justify the data processing methods and analyses used throughout this thesis and act as a common reference for chapters using similar methods. The experimental method development will first be discussed, followed by the computational method development. The methods are described in the order they appear in the following data chapters.

3.2 Experimental Results and Discussion

3.2.1 PCR

3.2.1.1 Repertoire and 16S

3.2.1.1.1 Triplicate PCRs

The use of triplicate PCRs as a technique to mitigate the effects of PCR jackpotting, the potential for sequences to be overrepresented due to random unequal amplification in the early rounds of a reaction, has recently been called into question in the microbiome field[21]. The idea behind repeat PCRs is to reduce this bias with multiple different random starts. However, even if this is the case, the triplicates remain extremely useful for eliminating the possibility that contamination has occurred due to pipetting error at the PCR set up stage. By plotting the PCR replicates on a Bray Curtis plot, the degree to which they cluster can be examined. Any replicates clustering with the incorrect sample would be indicative of a pipetting error, which can often subsequently be

corrected. For both 16S and repertoire sequencing, the triplicate PCRs are visualised alongside no template controls using the Bray-Curtis distance metric and plotted using Non-Metric Multidimensional Scaling (NMDS). The example triplicate PCRs come from 7 Day old Wild Type chicken faecal DNA samples amplified with 16S primers (See Chapter 5) and show good clustering of each PCR replicate within a given sample (Fig 3.2). Furthermore, the No Template Controls cluster randomly around the faecal samples and have very low numbers of reads associated with them. There are some examples of faecal PCR replicates clustering away from the rest of the samples, though it should be noted that these have very low read numbers associated with them and will therefore be filtered out during processing. Once it has been confirmed that the sample triplicate PCRs cluster close to one another and the No Template Controls do not resemble them, the triplicates are combined by summing read counts.

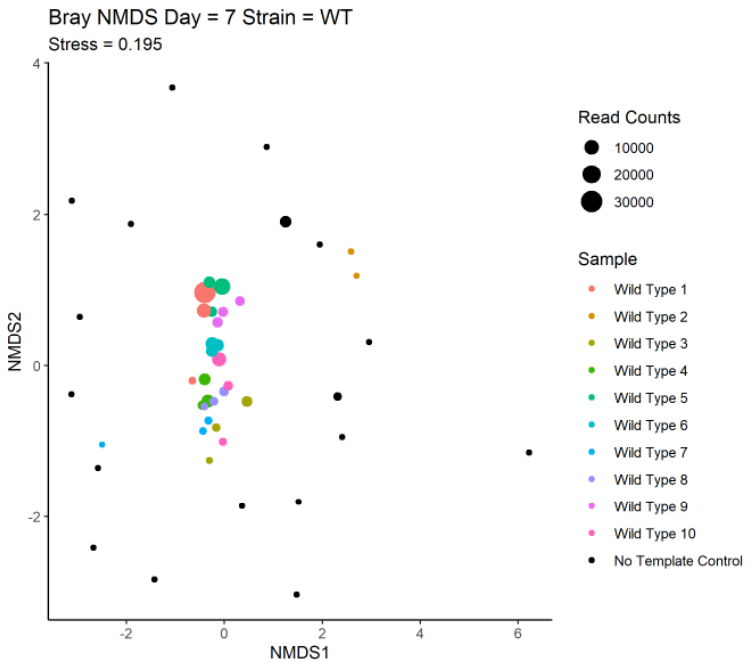


Figure 3.2 – Bray-Curtis NMDS plot of Triplicate PCRs and No Template Controls
β-diversity plot shows Bray-Curtis distances between triplicate PCRs and No Template Controls ordinated using NMDS. Bray-Curtis is calculated based on proportional data by treating every 16S sequence as an individual species. NMDS plot was generated with 20 random starts and each point represents a PCR replicate. Wild Type chicken faecal samples are shown in colour and No Template Controls are shown in black. Stress values are indicated below the title.

3.2.1.2 Repertoire PCRs

The length of a rearranged TCR gamma chain, for both chicken and mouse, is approximately 610bp depending on the V or J used and the degree of addition or deletion of nucleotides at the recombination sites (Fig 3.3). With the incorporation of the UPA long primer, the expected product length is around 650bp. Due to the incorporation of the poly dT adapter (5' CDS primer A) into every cDNA molecule, methods using the universal UPA primers often result in 'smeary' products due to mispriming as can be seen on the UPA gels[22]. Additionally, an unidentified smaller product around 200bp is often observed in these PCRs, however this and the smears are removed during downstream gel extraction.

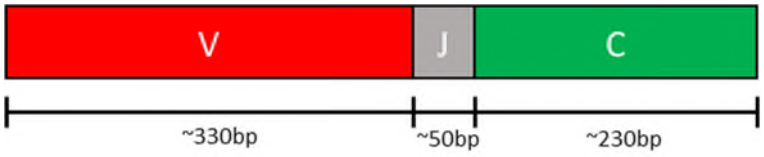


Figure 3.3 – TCR Gamma Chain RNA after Recombination, Transcription and Splicing
TRGV length shown is a rounded average of each of the TRGV gene lengths. Only the partial length of the TRGC gene is shown as the TRGC primers bind approximately half way along it.

3.2.1.3 16S

3.2.1.3.1 Chicken Faecal DNA Extraction and PCRs

DNA was extracted from chicken faecal samples using the Maxwell Blood DNA Purification Kit. After initial 16S PCRs resulted in a lack of amplification, diluting the DNA prior to PCR was tested as it has been shown previously that DNA extracted from chicken faecal samples is often contaminated with PCR inhibitors[23] (Fig 3.4). To test the impact of dilution on this issue, we set up 16S PCRs with faecal DNA (4.21ng/µl) which was serially diluted by a factor of 10 from 1:1 down to 1:1000 in Nuclease Free (NF) H_2O. Using this method, we showed that undiluted faecal DNA was contaminated with PCR inhibitors and failed to yield 16S bands, but 1:1000 diluted DNA has too low a template concentration to produce a satisfactorily strong band under the PCR conditions (Fig

3.4A). However, DNA diluted 1:10 or 1:100 had a low enough concentration of PCR inhibitor to allow amplification but a high enough concentration of DNA to produce strong bands. In order to confirm these PCR issues were indeed caused by PCR inhibitor presence in the extracted faecal DNAs, control PCRs were set up using a pTarget plasmid with a TLR insert from another project (Fig 3.4B). This test showed that amplification of the TLR insert is inhibited in the presence of the extracted chicken faecal DNA and that this inhibition diminishes with dilution. DNA concentration suggests this was unlikely to be inhibition due to DNA overload and that a secondary inhibiting agent must therefore be present. All chicken faecal DNA samples were subsequently diluted 1:100 for amplification with barcoded 16S primers.

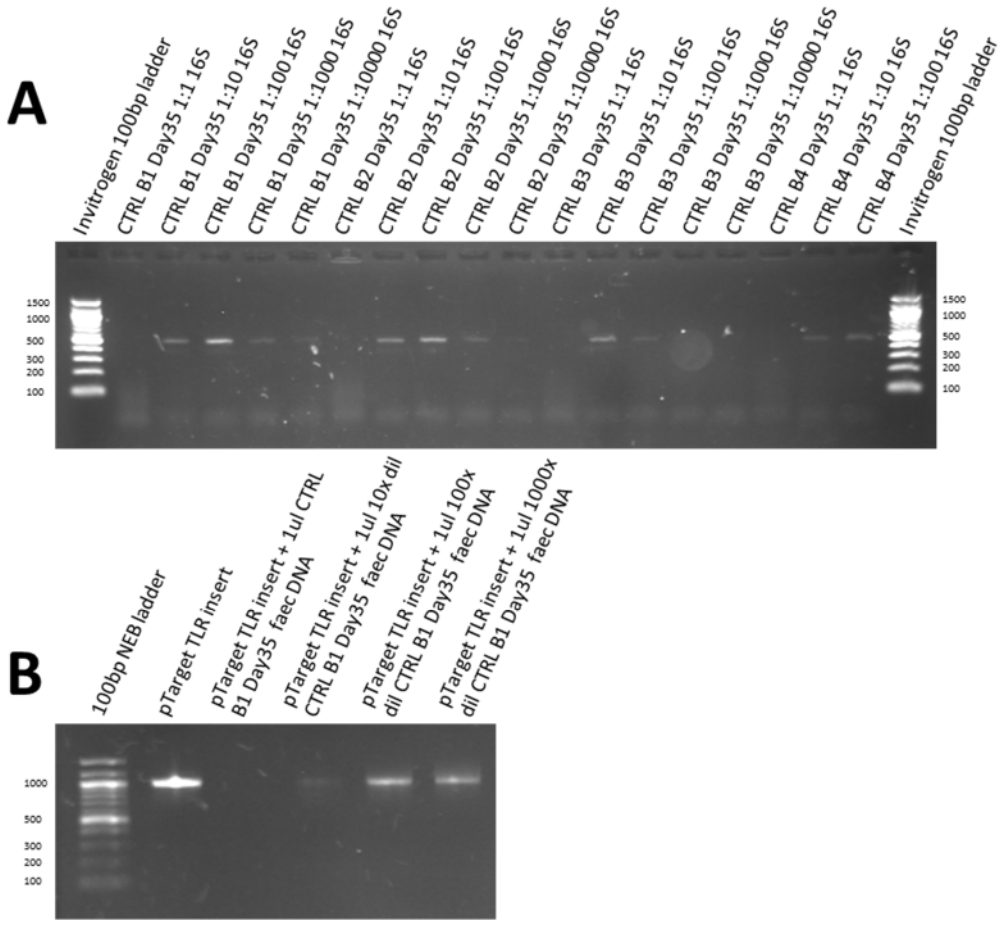


Figure 3.4 - Diluted 16S PCR Samples Highlight the need to Dilute Faecal PCR Inhibitors prior to Amplification
A) Extracted faecal DNAs were serially diluted 10x down to 1:1000 using NF H_2O. 16S PCRs were then performed on these samples. **B)** To confirm the presence of PCR inhibitors in the extracted faecal DNAs, a PCR designed to confirm that a pTarget plasmid contained a TLR insert during a cloning experiment was used. The extracted chicken faecal DNA was added at various dilutions to demonstrate inhibition. Gels were run at 120V for 30 minutes alongside a 100bp Invitrogen/NEB ladder. Ladder band sizes are along either side of the gel in bp.

3.3 Analytical Results and Discussion

3.3.1 Repertoire Analytical Development

3.3.1.1 Initial Processing

Reads Passing Filter (PF) generated by the Illumina MiSeq run, are first passed on to our in house Python demultiplexing code, which takes each MiSeq library and separates the reads into folders by searching for the primers and barcodes used in the amplicon generation. To eliminate erroneous CDR3 calls introduced via sequencing error, the reads are then denoised using the VDJtools correct function[24]. This function examines the data and identifies sequences with low read numbers. It then searches the rest of the data looking for a higher read number sequence which differs from the low read count sequence by a set number of base mismatches, in this case 1. If the higher read number sequence also has 20 times more reads than the lower one then there is a good chance that a sequencing error has been occurred and the lower read number sequence count has its count added onto the higher one and is discarded.

Following this, VJC assignment code was run on the reads using another in house package, reptools. The reads are then searched against a dictionary of possible V, J and C genes using BLAST with a percentage sequence identity cut off. The CDR3s are identified and extracted based on the presence of a conserved Cysteine residue, representing the end of the V gene segment, and a conserved Phenylalanine residue, representing the start of the J gene segment. Reads may be excluded at this stage due to incomplete recombination, intergenic sequences being retained and concatenated J genes[25]. Each clone is assigned a number based on unique V gene, J gene, C gene and CDR3 nucleotide sequence combinations. Though these numbers are essentially arbitrary, lower clone numbers are those which have been encountered earlier by the naming function and therefore tend to be the most common. Reads from unexpected barcodes, presumably present due to misallocation or contamination are then removed from the dataset, generally resulting in minor

data loss. Finally, the translated CDR3 nucleotide sequences are examined and any clones containing incomplete codons, or stop codons are removed.

3.3.1.2 Chicken Gene Assignment

In mice, differences between V gene segments are relatively large and therefore V gene assignment does not produce ambiguous designations. However, chickens present the problem of having very low germline sequence diversity within each of their T Receptor Gamma (TRG) V gene families, and therefore ambiguously assigned sequences occur during TRGV gene designation. As other animals typically have higher sequence diversity[26] this is not normally an issue in repertoire studies and raises the issue of how ambiguously assigned sequences should be addressed. The most obvious cases are TRGV genes which are identical and can therefore be treated as a single entity for the purposes of analysis. Genes which differ from one another by 1-2bps are also treated in this manner, as this level of similarity makes them difficult to reliably tell apart. Additionally, it is very likely that these genes will be almost identical in overall protein structure and therefore it probably makes little biological difference if one or the other is used. The ambiguous V genes which were grouped were “TRGV2.7/2.18” and “TRGV2.4/2.8/2.14/2.19”. The majority of ambiguous gene groups account for a negligible amount of the sequence data and are simply excluded from analysis to enable more straightforward interpretation of the data.

3.3.1.3 Linear Mixed Models for TRGV Expression Profiles

In statistical analysis of proportions it is generally accepted that normality can be assumed for a binomial distribution once a certain number of trials has been reached, normally said to be around 20. In the case of our data, we never have this many animals in a group, mainly due to the complexity of the experiments, thus the data does not meet this requirement. Standard deviation can be calculated directly from binomial distribution, however this gave unacceptably large error bars.

Therefore it was decided to analyse the TRGV gene proportion data using a Linear Mixed Model (LMM), as it would subvert the problem of large error bars and allow calculation of P values, whilst taking into account the nested nature of the data. Data was logit transformed and fixed and random effects chosen from the metadata. Models of this type come with their own assumptions however, and it must be confirmed that residuals are normally distributed and residual variance is approximately constant. Example figures for confirmation of these assumptions are shown as calculated from the chicken TRGV gamma repertoire LMM calculated for Chapter 5 (Fig. 3.5). Plots show that the variation in the residuals is even across the entire dataset (Fig 3.5A) and that the residuals are normally distributed (Fig 3.5B-C), therefore the data fits the assumptions of the LMM. These assumption tests are performed for each individual usage of LMMs in TRGV proportional usage plots. Without the Logit transformation of the proportions prior to LMM calculation, these assumptions are not met, and the residual variance shows a cone-shaped profile, characteristic of uneven distribution. Confidence intervals can then be calculated using bootstraps and are far less variable that those calculated from Binomial Distribution and P values can be calculated directly from the LMM using the LmerTest package using the Satterthwaite approximation of degrees of freedom[27].

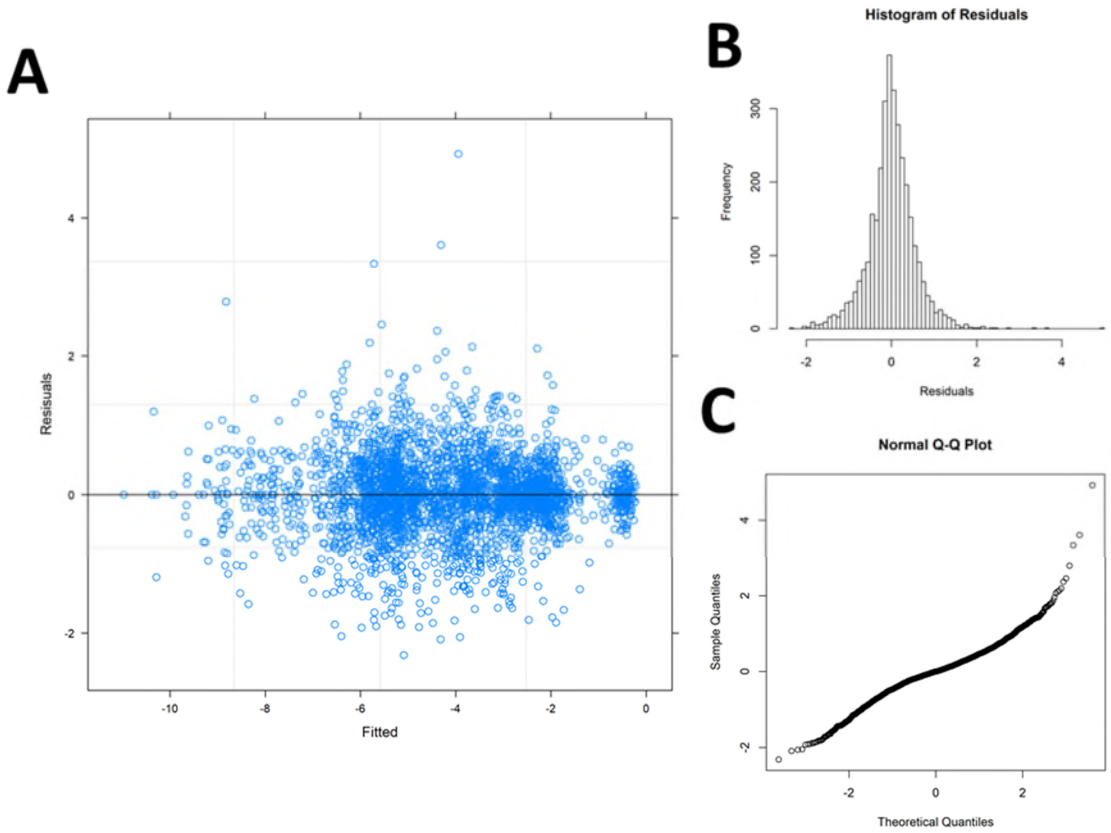


Figure 3.5 - Plots to Confirm Data fits the Assumptions of Linear Mixed Models
A) Plot of model residuals, Y axis is actual values, X axis is fitted/expected values. Even variation across entire datasets shows data fits model assumptions
B) Histogram showing normally distributed data with characteristic bell curve
C) Quantile –quantile plot. Points follow a roughly straight line so data is normally distributed, uptick at end is likely due to long tail present in panel B.

3.3.1.4 Diversity

3.3.1.4.1 How Diversity is Calculated

Diversity metrics are values traditionally used in ecology and are assigned to a sampling site to give a measure of how 'diverse' it is. Diversity at a site is measured based on 2 parameters, how many species are present at a given site, richness, and how evenly distributed individuals are between those species, evenness. The most diverse site based on these 2 parameters would be one with a high number of species with individuals distributed equally amongst them, whereas the least diverse would be made up of individuals all of a single species[28].

Diversity metrics differ based on the emphasis they put on these parameters, represented by their Hill number, q. Species richness (q = 0), the simplest diversity metric, measures only the number of species present at a site, Shannon diversity (q = 1) also considers evenness and Simpsons diversity (q = 2) weights evenness the most heavily[29] of the three. Additionally, these metrics are influenced by sampling effort, in this case represented by the number of reads. As numbers of reads vary widely for each sample this must be accounted for before diversity metrics can be compared.

Traditionally this problem has been addressed by rarefying, in which data is essentially discarded down to the level of the least sampled site. However, this approach has been criticized harshly as discarding data reduces statistical power and wastes sampling effort[30]. A wide range of other approaches have been suggested in the literature, including the use of proportions, various transformations and rarefying repeatedly[30–34].

Proportions have been criticised for their inability to account for unequal variances between samples, referred to as heteroscedasticity. Despite its flaws, rarefying does normalise variances between samples, but only by reducing read numbers down to the least sampled site and therefore inflating variance to the level of the worst case site[30]. However, proportions have been recommended for their use in β diversity metrics due to their ability to provide accurate comparisons across whole communities[33]. Log transformations have also been criticized for their inability to fully standardise variance and tendency to distort data by emphasising the contributions of less common species[33]. The issues of discarding data can theoretically be overcome by simply rarefying multiple times as, assuming sufficient subsampling repeats are performed, all data should be included in one of the many replicates[31]. In practise however, this method leads to a very large margin of error, particularly when rarefying down to lower read numbers in an effort to include as many sites as possible.

The majority of these issues however, can be overcome using a related technique known as rarefaction. Whereas rarefying works by random sub-sampling without replacement, rarefaction

operates via non-parametric resampling. Furthermore, to avoid discarding data sampling curves can be calculated and used to extrapolate lower sample-size sites with an increasing degree of uncertainty as the extrapolation factor increases[35]. Additionally, rarefaction and extrapolation based on sample completeness rather than sample size has been shown to be a better indicator of differences in diversity between communities and also often reduces the degree to which data must be rarefied[36]. Sample completeness is indicated by the proportion of individuals that fit within a species which has already been observed, thus if all species present have been observed, sample completeness or coverage would be 1. In order to implement this method, iNEXT was used to facilitate rarefaction based on 2x extrapolation of the sample with the lowest sample coverage and subsequent calculation of Species Richness, Shannon diversity and Simpsons diversity[37].

Rarefaction via non-parametric resampling is a common technique in ecology and has become the gold standard for performing diversity comparisons with samples which differ in sampling depth. This method has been incorporated into several different analysis packages including iNEXT[37], EstimateS[38] and Vegan, though iNEXT is the most comprehensive of the three. iNEXT based diversity calculations are absolutely routine in the ecological literature and despite the existence of various other methods and tools remains extremely widely used and cited, 1389 times at the time of writing. Though this type of down-sampling remains an imperfect solution to a difficult problem, it is accepted in the ecological literature and has been developed the Chao lab, a highly reputable and reliable mathematical biology group and can therefore be considered trustworthy and robust.

Species richness is calculated by combining the number of species seen in a site with an estimate of the unseen species. The number of unseen species is calculated from singletons, the greater the number of singletons, the greater the chance that there are other species yet to be observed. With NGS sequencing, the potential is high for spurious single sequences to be introduced

into the data through base calling error, chimeras or contaminants. The Species Richness should therefore be interpreted with scepticism when it comes to NGS data.

3.3.1.4.2 Hill Numbers

We express the diversity indices calculated as above as their Hill number variants. There are a number of advantages to expressing diversity in these terms which have been discussed extensively in the literature[39]. However, briefly there are a few main advantages to this. Diversity indices as Hill numbers are expressed as effective numbers of species, regardless of what metric is actually being shown. Effective number of species is the number of evenly distributed species which would be required to achieve the same diversity score. This has the advantage of ensuring that the indices obey the doubling property, which is to say that if 2 assemblages of equal diversity but with no common species are combined, then the diversity metric is doubled. Additionally, having all metrics expressed in the same 'units' makes it easier to compare between them and allows diversity to be more intuitively interpreted. Furthermore, the algebraic conversions between Hill number diversity indices and their more traditional forms are straightfoward[40]. In the case of Shannon's diversity, the conversion is simply an exponential conversion. For Simpson's diversity, normally expressed as a number from 0-1, it is the inverse of the complement of this number.

3.3.1.4.3 Why Diversity is Calculated only on Repertoire and not Microbiome Data

We do not calculate diversity metrics for our 16S data because of the way it has been processed. It is well accepted in the literature that ASV based methods such as DADA2 have difficulty in accounting for sequence errors due to base changes, chimeras and other sources. In the case of the DADA2 algorithm specifically, a sequence observed only a small number of times which is suspected of arising from an error in a more common sequence, has its count combined with that sequence if the bespoke sequencing run error model supports this[15]. However, this is still an imperfect process and some singletons from real rare sequences may still remain. This is an impossible problem in that at this stage, real rare sequences are impossible to tell apart from a singleton sequence arising from some error. DADA2 solves this problem by simply removing all

singletons after the error correction stage. One way to limit the chances of real sequences being removed is to pool all of the sequences before this error correction process to maximise the total read depth and therefore increase the power to confirm rare sequences as real, whilst the genuine errors should not be affected. While this mitigates the problem to a certain extent, it still does not provide a resolution, and error correction of all sequences in a combined manner massively increases the computational power needed and is therefore unfeasible for large datasets.

Species Richness estimation methods not based on singletons do exist, however they produce highly variable results and are therefore their use is inadvisable, even by their authors[41]. While Species Richness is heavily affected by the presence of such sequencing errors and cannot be used for DADA2 processed data, diversity indices that also take into account evenness, such as Shannon and Simpsons diversity indices are more robust. However, as described above we have concluded from the literature that iNEXT diversity estimation using coverage based rarefaction and extrapolation, though still flawed, is the best way to calculate these indices. In order to perform coverage based rarefaction one has to be able to calculate coverage, which is itself based on singletons and doubletons. Though methods do exist to calculate coverage not based on singletons, they are subject to similar issues present in the Species Richness estimation methods[41]. Therefore, we do not calculate alpha-diversity (within-sample diversity, calculated from a single population) metrics for DADA2 processed 16S data.

In the case of the repertoire data, our pipeline uses a more crude denoising process and thus does not remove singletons. It is therefore possible to calculate iNEXT estimates for diversity indices including Species Richness but these should still be treated with caution.

3.3.1.4.4 Linear Models

A linear model was calculated for each of the diversity data sets and used to calculate confidence intervals and P values. The optimal transformation to ensure normality of residuals was determined using Box-Cox plots (Fig 3.6). A cube root transformation was found to perform best in

all cases. The terms used to calculate the model were determined by examination of the P values. If no significant values were obtained for a given term, it was eliminated from the model to reduce overfitting. Normality and homoscedasticity was then confirmed as described for the LMMs.

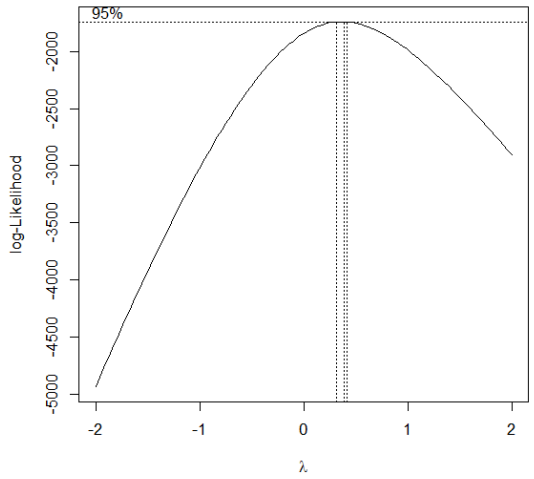


Figure 3.6 - Example BoxCox Plot for Linear Model Transformations
BoxCox plots were used to determine the optimal transformation to ensure normality of residuals. The peak of the plot was used to select the optimal root transformation.

3.3.2 Repertoire and 16S Analytical Development

3.3.2.1 Beta-Diversity Metrics

3.3.2.1.1 Proportions for Beta-Diversity

When comparing communities using beta diversity (between-sample diversity, calculated from two populations), in the case of Bray-Curtis, UniFrac and Jensen-Shannon divergence (JSD), dissimilarity indices were calculated based on proportions. It has been shown that calculating dissimilarity indices in this way is the optimum solution when seeking to preserve differences between ecological communities[33]. This is of particular importance for our data where none of the differences observed between communities are large.

3.3.2.1.2 Dissimilarity Indices

3.3.2.1.2.1 Common

3.3.2.1.2.1.1 Bray-Curtis

Bray-Curtis dissimilarity is a metric favoured in the literature because it is simple, quick to calculate and has few assumptions. Bray-Curtis dissimilarity can take on a value from 0 to 1, where a value of 0 means the samples are identical and a value of 1 means they are totally different and share no common species/sequences. However, Bray-Curtis distance does suffer some drawbacks. In terms of 16S and particularly repertoire sequencing, datasets often have high sparsity, meaning that samples being compared have a high number of 0 counts where any one sequence is often not observed at all in another sample. Additionally, Bray-Curtis does not take into account phylogenetic relatedness and therefore a difference in the abundance of highly related sequences gives the same dissimilarity as less related sequences. This is important because differences in related sequences may not result from an actual biological difference. This is of special importance to 16S data as many of the sequences, and therefore the microbes present, will be closely related. In addition, processing errors in the DADA2 pipeline may lead to a single species being corrected to 2 difference sequences in different samples. Thus, accounting for relatedness is also of value when addressing these problems.

3.3.2.1.2.1.2 Aitchison's Distance

Despite non-compositional distance metrics, such as Bray-Curtis and UniFrac, continuing to be standard practise in the literature, it has recently been argued that, due to the nature of 16S Illumina sequencing, the data it produces should be treated as compositional, as the run can only generate reads up to a certain numerical capacity. Therefore, the reads produced can be thought of as occupying a limited number of 'sequencing slots' in the run and an increase in the number of these slots assigned to any one read is associated with a decrease in slots available to other reads[42]. To address this issue, we also include Aitchison's dissimilarity metric. In order to calculate Aitchison's distance, a small pseudocount is added onto each sample and the counts are transformed using

centre log ratio (CLR). Following this transformation, Aitchison's distance is simply calculated from the transformed counts using Euclidian distance[43]. One drawback when using Aitchison's distance for repertoire data is the low amounts of variation explained by the X and Y axes. This tends to be far higher for 16S data, and may be due to the increased multidimensionality of the immune repertoire compared to the microbiome.

3.3.2.1.2.2 Repertoire Specific

3.3.2.1.2.2.1 JSD

Due to the high level of CDR3 diversity and the inevitable sparsity this introduces into the data, the usefulness of Bray-Curtis for repertoire analysis has been questioned, as this metric typically performs better with overlapping datasets[44]. As a result of this, ordination plots using the JSD index were also produced, as JSD uses random sampling from the combined datasets of all samples to calculate the probability of each distribution occurring from the others and is thus more robust to sparse data.

3.3.2.1.2.3 Microbiota Specific

3.3.2.1.2.3.1 UniFrac

UniFrac is a dissimilarity metric which is highly similar to Bray-Curtis except it does account for relatedness between individuals[45,46]. In order to input the relatedness component required to compute UniFrac metrics, a phylogenetic tree must first be constructed. The nucleotide evolution model which best firsts the data is first selected based on highest log Likelihood and lowest Akaike Information Criterion (AIC) as calculated by modeltest[47]. A distance matrix is then calculated based on this nucleotide evolution model and a phylogenetic tree computed using the neighbour joining method[48]. This tree is then used as a starting point for Maximum Likelihood (ML), which is used to produce a final tree.

3.3.2.1.2.3.2 PhILR

Similarly, to Bray-Curtis, Aitchison's distance does not take into account relatedness between samples. PhILR is simply an adaptation of Aitchison's distance which also incorporates a phylogenetic tree, similar to UniFrac. PhILR stands for Phylogenetic Isometric Log Ratio and as the name suggests, it involves data transformation using isometric log ratio taking into account information about relatedness between sequences inputted via a phylogenetic tree[49]. The tree is calculated as with UniFrac distance.

3.3.2.2 Beta-Diversity Ordination

3.3.2.2.1 NMDS

NMDS is a technique whereby highly multidimensional data which could be compared on multiple different individual axes is collapsed into 2 or 3 dimensions for visualisation. It has several advantages over other commonly used ordination methods such as Principal Coordinates Analysis (PCA), including flexibility and ability to represent all data on a single plot, not just a percentage of the variation. Rather than the percentage of variation represented on an axis, NMDS plots produce a stress value, which conveys how well the plot represents the displayed data, >0.1 is a good representation, >0.2 is an acceptable representation and >0.3 is a poor representation. Calculation of NMDS plots occurs iteratively as repeat configurations are compared in terms of their stress values to find the optimal arrangement. In NMDS, the axes scales and orientation are essentially arbitrary.

3.3.2.2.2 RDA

Redundancy Analysis (RDA) is another technique used for ordination of Aitchison's and PhILR distance metrics. RDA is simply a constrained version of PCA where axes are constructed from combinations of variables selected to explain the maximum variation of the data possible. The percentage of variation explained by each axis is displayed next to the axis title. The variation

explained tends to be low for the repertoire plots, suggesting the multidimensionality present in these datasets overwhelms this type of ordination.

3.3.2.3 Statistical Analysis

3.3.2.3.1 PERMANOVA

Permutational multivariate analysis of variance (PERMANOVA) is then used to calculate P and R^2 values for the β-Diversity plots. PERMANOVA testing utilises multiple random permutations, 999 were used in our analyses, to calculate the probability that the observed distribution could be produced by chance. The null hypothesis for PERMANOVA testing is that there is no difference between the centroids of the clusters being tested. This type of testing can be performed both globally, on multiple categories, or used to test differences between just 2 groups. One assumption for this type of test is homogeneity of variance and therefore prior to PERMANOVA testing, the vegan betadisper function, which calculates multivariate homogeneity of group dispersions, in conjunction with ANOVA, is used to test this assumption.

3.3.2.3.2 Multiple Testing Correction

P values were adjusted using a multiple testing correction for each individual analysis as testing was *post-hoc*. As we were mostly concerned about avoiding rejecting a true hypothesis, or Type II error, the FDR correction was selected. We defined our family as the set of statistical tests in any one analysis. This was because though data is arguably dependant between analyses, each effectively tests a different hypothesis and therefore can be corrected separately[50,51].

3.4 References

1. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.
2. Murphy K. Janeway's Immunobiology. 8th ed. 2011.
3. Liu F, Li J, Lin IYC, Yang X, Ma J, Chen Y, et al. The Genome Resequencing of TCR Loci in Gallus gallus Revealed Their Distinct Evolutionary Features in Avians . ImmunoHorizons. 2020;4(1):33–46.

4. Lefranc M. IMGT [Internet]. 2021.

5. Lefranc M. Ontology for immunogenetics : the IMGT-ONTOLOGY. Bioinformatics. 1999;15(12):1047–54.

6. Fukuda K, Ogawa M, Taniguchi H, Saito M. Molecular approaches to studying microbial communities: Targeting the 16S ribosomal RNA gene. J UOEH. 2016;38(3):223–32.

7. Wang LT, Lee FL, Tai CJ, Kasai H. Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group. Int J Syst Evol Microbiol. 2007;57(8):1846–50.

8. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat Publ Gr. 2010;7(5):335–6.

9. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.

10. Guarner F. The gut microbiota era marches on. Nat Rev Gastroenterol Hepatol. 2014;11(11):647–9.

11. Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data. BMC Bioinformatics. 2012;13(1).

12. Tikhonov M, Leach RW, Wingreen NS. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. ISME J. 2015;9(1):68–80.

13. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 2015;9:968–79.

14. Mcelroy K, Zagordi O, Bull R, Luciani F, Beerenwinkel N. Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering. 2013;

15. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2 : High resolution sample inference from amplicon data. Nat Methods. 2015;13(7):0–14.

16. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

17. Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One. 2020;15(1):1–19.

18. Nearing JT, Douglas GM, Comeau AM, Langille MGI. Denoising the Denoisers: An independent evaluation of microbiome sequence error- correction approaches. PeerJ. 2018;(8):1–22.

19. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4).

20. Dixon P. Computer program review VEGAN , a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.

21. Marotz C, Sharma A, Humphrey G, Gottel N, Daum C, Gilbert JA, et al. Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. Biotechniques. 2019;67(1):29–32.

22. Bower NI, Johnston IA. Targeted rapid amplification of cDNA ends (T-RACE) - An improved RACE reaction through degradation of non-target sequences. Nucleic Acids Res. 2010;38(21).

23. Monteiro L, Bonnemaison D, Vekris A, Petry KG, Bonnet J, Vidal R, et al. Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. J Clin Microbiol. 1997;35(4):995–8.

24. Shugay M, Bagaev D V., Turchaninova MA, Bolotin DA, Britanova O V., Putintseva E V., et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. PLoS Comput Biol. 2015;11(11):1–16.

25. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2017;19(November 2016):bbw138.

26. Su C, Jakobsen I, Gu X, Nei M. Diversity and evolution of T-cell receptor variable region genes in mammals and birds. Immunogenetics. 1999;50(5–6):301–8.

27. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models . J Stat Softw. 2017;82(13).

28. Schleuter D, Daufresne M, Massol F, Argillier C. A user's guide to functional diversity indices. Ecol Monogr. 2010;80(3):469–84.

29. Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, et al. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. Ecol Monogr. 2014;84(1):45–67.

30. McMurdie PJ, Holmes S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol. 2014;10(4).

31. Cameron ES, Schmidt PJ, Tremblay BJM, Emelko MB, Muller KM. To rarefy or not to rarefy: Enhancing microbial community analysis through next-generation sequencing. bioRxiv. 2020;

32. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2013;14(6):671–83.

33. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. Methods for normalizing microbiome data: An ecological perspective. Methods Ecol Evol. 2019;10(3):389–400.

34. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5(1):1–18.

35. Colwell RK, Chao A, Gotelli NJ, Lin SY, Mao CX, Chazdon RL, et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. J Plant Ecol. 2012;5(1):3–21.

36. Chao A, Jost L. Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. Ecology. 2012;93(12):2533–47.

37. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol Evol. 2016;7(12):1451–6.

38. Colwell RK. EstimateS: Statistical estimation of species richness and shared species from samples. [Internet]. 2013.

39. Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, et al. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. Ecol Monogr. 2014;84(1):45–67.

40. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, et al. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. Ecol Evol. 2014;4(18):3514–24.

41. Willis A. Species richness estimation with high diversity but spurious singletons. 2016;1–18.

42. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8(NOV):1–6.

43. Aitchison J. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B. 1982;44(2):139–60.

44. Yokota R, Kaminaga Y, Kobayashi TJ. Quantification of inter-sample differences in T-cell receptor repertoires using sequence-based information. Front Immunol. 2017;8(NOV):1–15.

45. Lozupone C, Hamady M, Knight R. UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics. 2006;7:1–14.

46. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5(2):169–72.

47. Posada D, Crandall AC. MODELTEST: testing the model of DNA substitution David. Chaos. 1998;14(9):817–8.

48. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

49. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. Elife. 2017;6:1–20.

50. Jafari M, Ansari-pour N. Why , When and How to Adjust Your P Values ? 2019;20(4):604–7.

51. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc. 1995;57(1):289–300.

Department of Zoology, University of Oxford

Chapter 4

The Chicken TCR Gamma Locus and Repertoire

4.1 Introduction

The adaptive immune system of jawed vertebrates functions through the production of antibodies by B lymphocytes, combined with the cytotoxic, helper and regulatory functions of T lymphocytes. T lymphocytes can be divided into two lineages according to their expression of either the αβ or γδ form of the T cell receptor (TCR), and the appearance of these lineages coincides with the emergence of the jawed vertebrates[1,2]. The function of TCRαβ+ T cells is well defined in many animal species, with most αβ T cells recognizing peptides presented on the surfaces of cells in the context of MHC class I or II. In contrast, TCRγδ+ T cells are much less well understood, although a range of stimulatory ligands have been described[3], and they have been shown to have an array of cytotoxic and cytokine based effector functions[4]. At least some populations of γδ T cells are considered part of the innate immune system, operating in the early phases of the host's response. These γδ T cell subsets have very restricted TCR diversity, which is characteristic of an innate-like response[5,6]. Other γδ T cell populations have diverse CDR3 repertoires, as are seen in classical αβ T cells, more characteristic of the adaptive immune system. Further characterization of the TCR repertoire of γδ T cell populations will be important in exploring whether these populations function in innate immunity, adaptive immunity or both.

Most information on the biology of γδ T cells derives from research on human blood or rodent tissues. Whilst providing a solid foundation for understanding the role of γδ T cells in immune responses, it is noteworthy that some of the aspects of their biology are not shared between humans and rodents, such as the population of Dendritic Epidermal T Cells (DETCs) which express an invariant TCR and are located in the skin of mice[7]. Moreover, humans and rodents are both considered γδ low animals whereas many others, including chickens, are considered γδ high animals[8]. γδ low animals are loosely characterised as having a peripheral blood T cell compartment which consists of less than 10% γδ T cells, the remainder being αβ T cells. This group includes mice (0.5-10%), humans (0.5-10%)[9] and dogs (2.5%)[10]. Conversely, γδ high animals have a circulatory T cell

compartment comprised of greater than 10% γδ T cells. This group includes various domesticated species such as chickens (15-50%)[11], cattle (20-40%)[12], pigs (30%)[13] and goats (5-20%)[14,15]. In both γδ high and γδ low animals, the proportions of γδ T cells are greatly enriched in some tissues, particularly at epithelial sites such as the intestine. To fully appreciate the biology of γδ T cells it is important to consider these lymphocytes in a range of different animals.

Typical γδ T cell functions include; rapid response to pathogens[16,17], regulation of enterocyte turnover[18], protection against epithelial tumours[19], and limitation of inflammatory damage[20]. γδ T cells are thought to be particularly important for immunity in early life[17,21]. In birds and mammals, γδ T cells are the first T cells to develop in the thymus[22,23], and in mice these early γδ T cells migrate from the thymus in discrete waves, each expressing different T Receptor Gamma Variable (TRGV) genes and homing to particular tissues to perform discrete functions[5,24]. The chicken is a γδ high species in which our knowledge of the response and function of these cells is limited. What is clear is that chicken γδ T cells are heterogeneous and can produce a variety of cytokines and interferons, as well as displaying cytotoxicity[25,26], and responding to infectious challenge[27,28] or changes in hormone levels[29]. As T cell specificity is determined by the TCR, advances in high throughput next generation sequencing (NGS) have enabled their biology to be studied in previously unattainable detail using TCR repertoire sequencing.

4.2 Chapter Outline

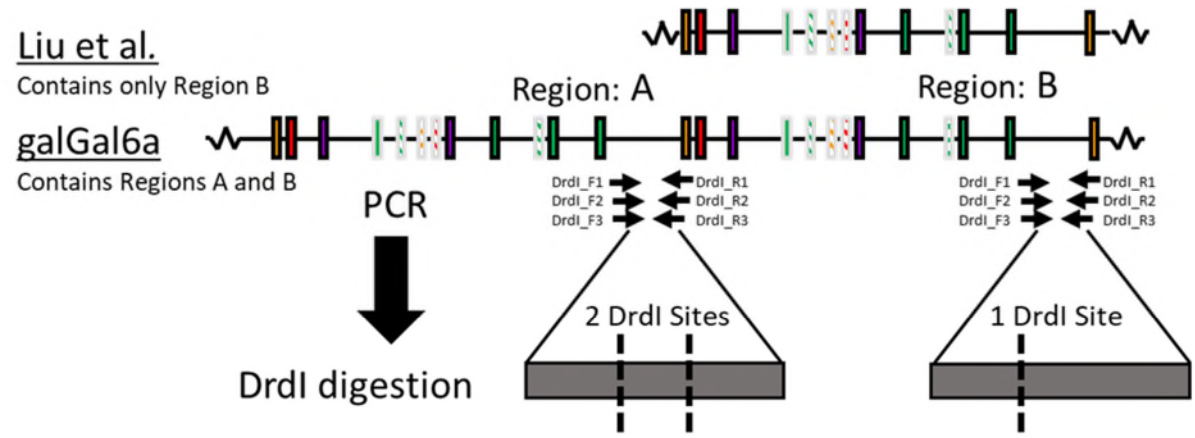
Accurate description of the clonal repertoire of TCR gamma requires a correctly assembled and annotated genomic locus. Early iterations of the chicken genome did not accurately depict the structure or content of the gamma locus, presumably due to the repetitive nature of the genes and the difficulties this poses to assembly. Although the most recent genome build (galGal6a) is a vast improvement, the sequence and structure of the TCR gamma locus is different to that proposed recently by Liu et al. 2020[30], based upon BAC insert resequencing. This chapter aims to resolve this discrepancy and add additional annotation that highlights the duplication-based evolution of this

locus. Using a 5'RACE based parallel sequencing approach for assessment of repertoire, this chapter will show that most TRGV genes are utilised, but that some TRGVs are used much more commonly than others. The proportional usage of each TCRV gene is partially predicted by the sequence of the recombination signal sequence (RSS) but other factors are clearly involved. Furthermore, the chicken TCR gamma repertoire includes substantial portions of highly public (shared between clonotypes) CDR3 sequences as well as a large private (individual-specific) repertoire.

4.3 Results

4.3.1 The Chicken Gamma Locus shows a Repeating Motif suggesting Tandem Duplication based Expansion

The most recent version of the chicken genome assembly, galGal6a, was supported by PacBio Single Molecule Real Time sequencing (SMRT), which produces reads with an average length of 8.5 – 14kb. The assembly has 82x coverage, which combined with the long SMRT reads should make it a high quality and reliable whole genome sequence. However, the TCR gamma locus of this assembly was recently contradicted by Liu et al. 2020[30] in a paper that used a BAC resequencing approach to correct an older genome assembly, galGal4. galGal6a features a 15kb region not present in the BAC derived gamma locus[30]. To address this discrepancy, we employed a PCR and restriction endonuclease digest approach (Fig. 4.1). Digestion with DrdI yielded fragment sizes consistent with the 15kb sequence being absent from the genomic DNA of PA12 White Leghorn chickens, confirming the genomic organisation proposed by Liu et al 2020.



Primer pair	Expected band sizes for A	Expected band sizes for B
Drd_F1 x R1	(910) 340,267,303	(907) 341,566
Drd_F2 x R2	(782) 293,267,222	(779) 294,485
Drd_F3 x R3	(673) 238,267,168	(670) 239, 431

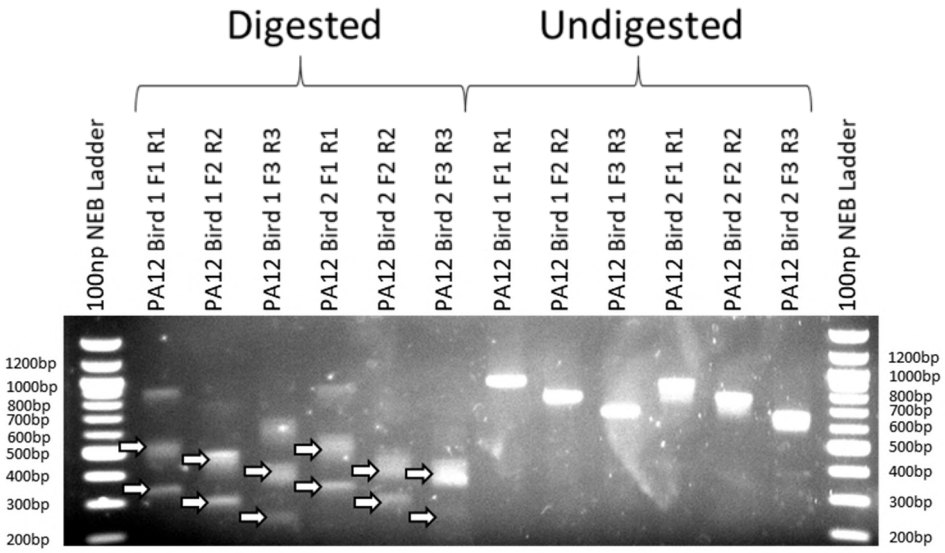


Figure 4.1 - PCR-Restriction Digest Strategy to Resolve the TCR Gamma Locus
A schematic showing the TCR gamma locus according to Liu et al. 2020 and as depicted from galGal6a, with the latter appearing to have a tandem duplicated region. The PCR-restriction digest strategy used to resolve this discrepancy is depicted including primer locations, DrdI restriction sites and expected product sizes (undigested product size in brackets). Regions of genomic DNA extracted from the spleens of 2 PA12 birds were amplified using three different primer sets, digested with the restriction enzyme DrdI and separated on a 1% agarose gel, visualised with ethidium bromide. Lanes 2-7 are with DrdI digestion, Lanes 8-13 are with no digestion. Product length was determined using a 100bp NEB ladder. Ladder band lengths are indicated at either side of the gels in base pairs. Digested PCR products of expected sizes are indicated (white arrow) and support the Liu et al. 2020[30] proposed structure for the TCR gamma locus.

Annotation of the chicken TCR gamma locus was performed using BLAST to identify gamma-like sequences based on previously known TRGV, TRGJ and TRGC genes (Fig. 4.2). To provide further

experimental validation of the annotation and to ensure no genes had been missed, immune repertoire derived sequence data was also mapped to the genome and BAC sequences. This revealed four TRGV genes not present in the annotation proposed by Liu *et al.* 2020[30], TRGV4.1,TRGV4.3 TRGV2.12 and TRGV4.5. The TRGV gene families were classified based on their predicted Ig like domains and members within families numbered according to location 5'-3' as proposed by IMGT. The TRGV gene families were named according to Six et al*.* 1996[31] with the addition of a TRGV4 family. Forty TRGV genes were identified and grouped into four families with three TRGJs and one TRGC gene as previously reported[31]. Multiple members were identified within each TRGV family and three of the four families contained pseudogenes. The TRGV1 family contains six members (with two pseudogenes), the TRGV2 family contains twenty-two members (with eight pseudogenes), the TRGV3 family contains seven members (with no pseudogenes), and the TRGV4 family contains five members (with three pseudogenes). In the case of TRGV1.4 and TRGV2.10, the sequences of these pseudogenes have deletions compared to other family members, which could have resulted in inappropriate classification. Therefore, they were classified according to the sequence upstream of the deletion site only. Most of the intact TRGV genes were found to be expressed at the mRNA level in thymus, spleen and intestinal tissues, with the exceptions of TRGV1.2 and TRGV2.9, which were not detected.

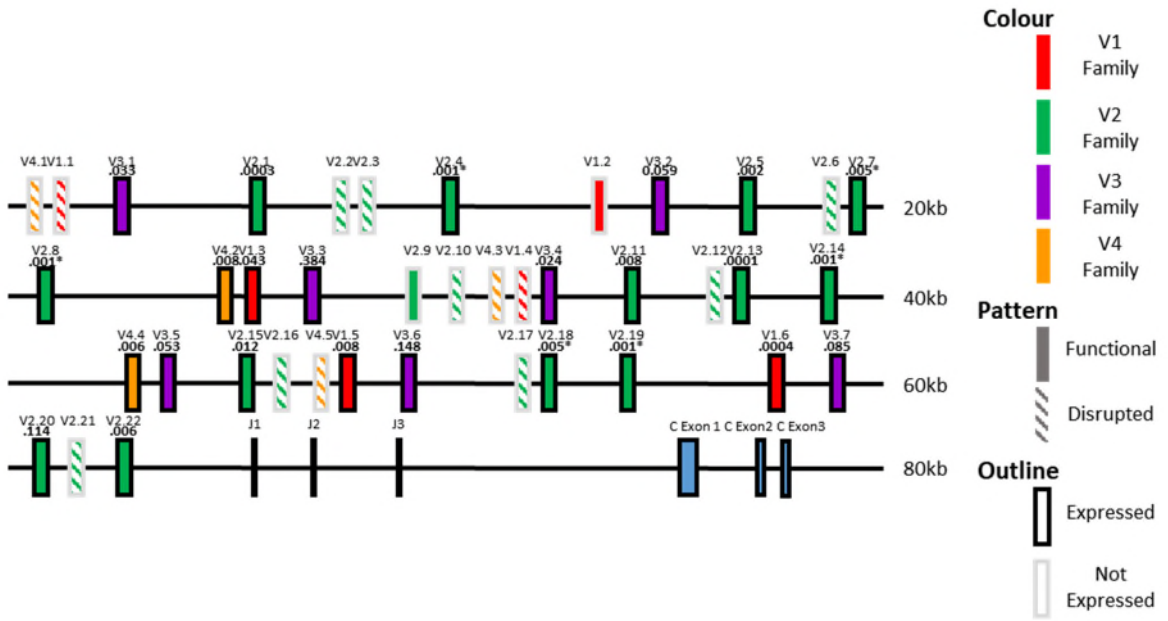


Figure 4.2 - Schematic Representation of the Chicken Gamma Locus
The schematic displays the location of the coding exons for the TRGV, TRGJ and TRGC genes of TCR gamma. Four distinct families of TRGV genes (different coloured boxes) are indicated with subfamily numbering based upon genomic location. RSS sequences were identified downstream of all indicated, non-pseudogenised TRGV genes. TRGV genes with a full length ORF are indicated with a solid fill and those with a disrupted ORF by a striped fill. Those found to be expressed and rearranged by RTPCR and sequencing are bounded by a black outline and those not identified as expressed are bounded by a grey outline. The proportion of rearrangements attributed to each TRGV gene are displayed below the label. These were calculated from the averages of every tissue across every bird in this study. Genes which are too similar to be reliably distinguished from one another during the TRGV gene assignment stage of processing are represented by an asterisk.

The addition of these new TRGV genes and their re-assignment into families offers a picture of a consistent repeating motif comprised of a TRGV4, TRGV1, TRGV3 and four TRGV2s interspersed by non-coding regions (Fig. 4.2). The lengths of these non-coding regions remain fairly consistent, though different iterations of this motif are subject to small differences presumably due to insertion/deletion mutations. Some of these motifs are truncated but the regions which are present still show this same characteristic structure.

A locus-versus-locus sequence comparison revealed that the TRGV motif also showed high sequence similarity to other repeated motifs in the non-coding as well as coding regions (Fig. 4.3). The regions of sequence similarity between motifs are largely unbroken and often span 10-20kbp, thus strongly supporting expansion of the region by tandem duplication events. The grouping of TRGV genes into four families was also supported by a phylogenetic tree (Fig. 4.4). Interestingly, the

TRGV2 genes clustered into four subgroups based on their position within the repeated motif (labelled A-D). The core TRGV motif therefore comprises TRGV4, TRGV1, TRGV3 followed by a TRGV2A, TRGV2B, TRGV2C and TRGV2D.

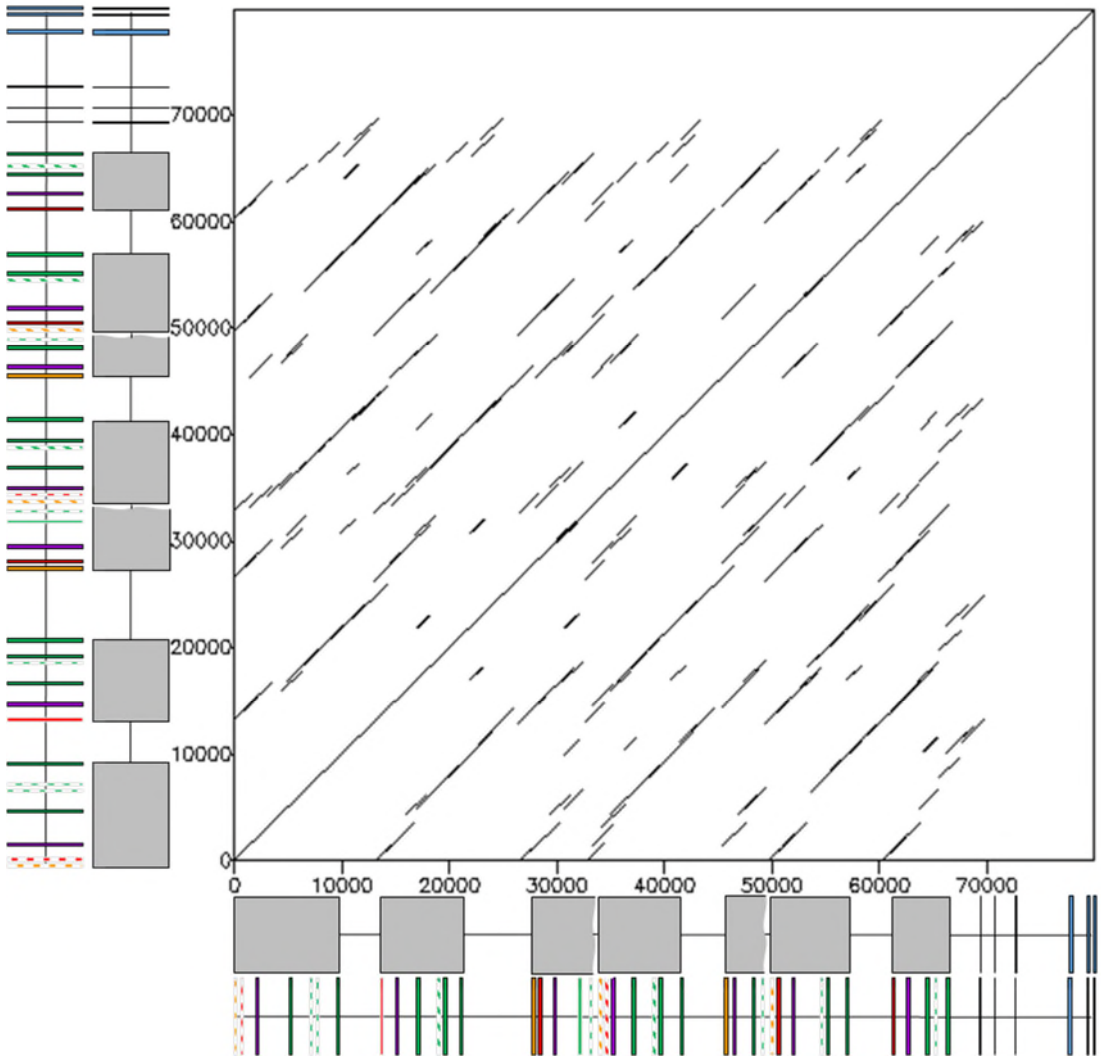


Figure 4.3 - Dotplot reveals that the Chicken TCR Gamma Locus Evolved via a set of Duplication Events
A comparison of sequence similarity across the TCR gamma locus using a dot plot generated with Dotmatcher, window size of 1000 and threshold of 500. The default EMBOSS matrix was used (with match =+5 and a mismatch =-4) and a dot plotted where the value was above the threshold. The X and Y axis include schematic representations of the chicken TCR gamma locus and an indication of potentially duplicated blocks of TRGV genes. Homologous comparison is represented by the diagonal passing through the origin and the parallel diagonal indications represent regions with high similarity.

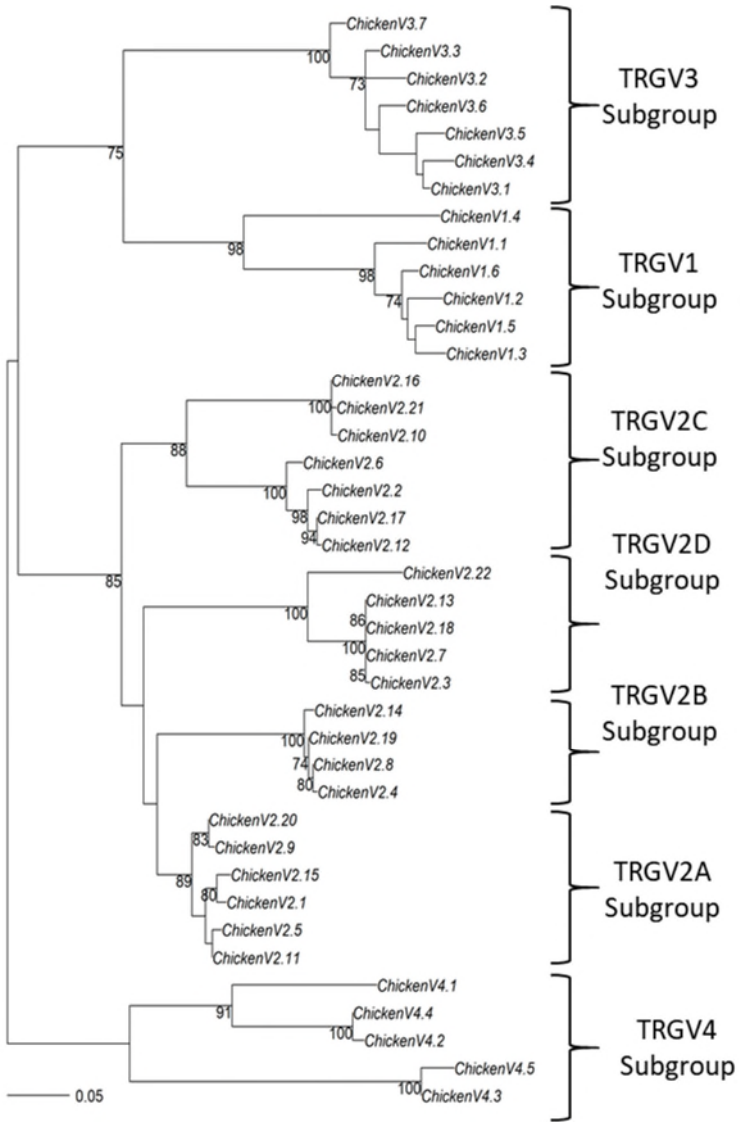


Figure 4.4 - Phylogenetic Tree of TRGV Gene Ig-like Domains
TRGV gene Ig-like domains were aligned using MUSCLE and a phylogenetic tree calculated using maximum likelihood with a neighbour joining tree as the start point. Bootstrap values were calculated using 1000 replicates. The tree shows TRGV families clustering together in branches, which are well supported by the bootstrap values (bootstrap values >70 are displayed). The TRGV2 genes also cluster by their position in the repeating motif, consistent with evolution by tandem duplication. Tree scale is indicated on the bottom left of the figure and shows the branch length which represents 5% of the variation.

4.3.2 The Chicken Gamma Repertoire Features a Dominant TCR gamma Rearrangement in every Tissue and Novel Public TCRs

A 5' RACE approach was used with multiple tissues from two lines of chicken, ISA Brown (49 days old) and PA12 White Leghorn (55 days old), to generate a broad representation of TRGV gene usage and to explore the repertoire of TCR gamma rearrangements. After amplicon sequencing and pre-processing, this yielded a total of 425,870 in frame CDR3 sequences for the ISA Browns, with an

average of 53,233.75 per bird and a total of 51,534 in frame CDR3 sequences for the PA12 White Leghorns, with an average of 10,306.8 per bird. To facilitate TRGV gene usage analysis, some TRGV2 genes were grouped together due to very high levels of sequence similarity.

ISA Brown birds expressed all potentially productive TRGV genes as productive rearrangements, with the exception of TRGV1.2 and TRGV2.9, although some genes were represented only as a small fraction of the sequences. The proportion of sequences attributable to each productively rearranged TRGV gene varied considerably, but the pattern was remarkably consistent between different tissues (Fig. 4.5). Interestingly, the profile of TRGV usage in the thymus was very close to that seen in spleen and intestinal tissues, except for a small but significant increase in the proportion of TRGV1.3 in thymus compared with jejunal and ileal tissues ($p = 0.04$ and 0.041 respectively). The mean proportion of TRGV2.20 appeared elevated in the thymic tissue, but this was variable between birds and not statistically significant ($p>0.05$) compared with other tissues.

TRGV3.3 was the most highly expressed TRGV gene, comprising 30-40% of the entire repertoire, and this dominance persisted across all tissues (Fig. 4.5). TRGV3.3 was at least twice as common as the second most prevalent TRGV gene in all of the tissues, and several times the proportion of the third most prevalent. In both the TRGV1 and TRGV2 families, a single family member is expressed at much higher levels than any other, TRGV1.3 and TRGV2.20 respectively. This is not the case with the TRGV3 family, as although TRGV3.3 was more highly represented than other TRGV3 family members, all other TRGV3 family genes were well represented in sequence data derived from all tissues. The two intact members of the TRGV4 family, TRGV4.2 and TRGV4.4, were expressed at relatively low proportions and without a strong bias in expression of one over the other. It is noteworthy that the more highly expressed TRGV genes are distributed across the entire genomic region and not located in any one of the duplicated TRGV motifs.

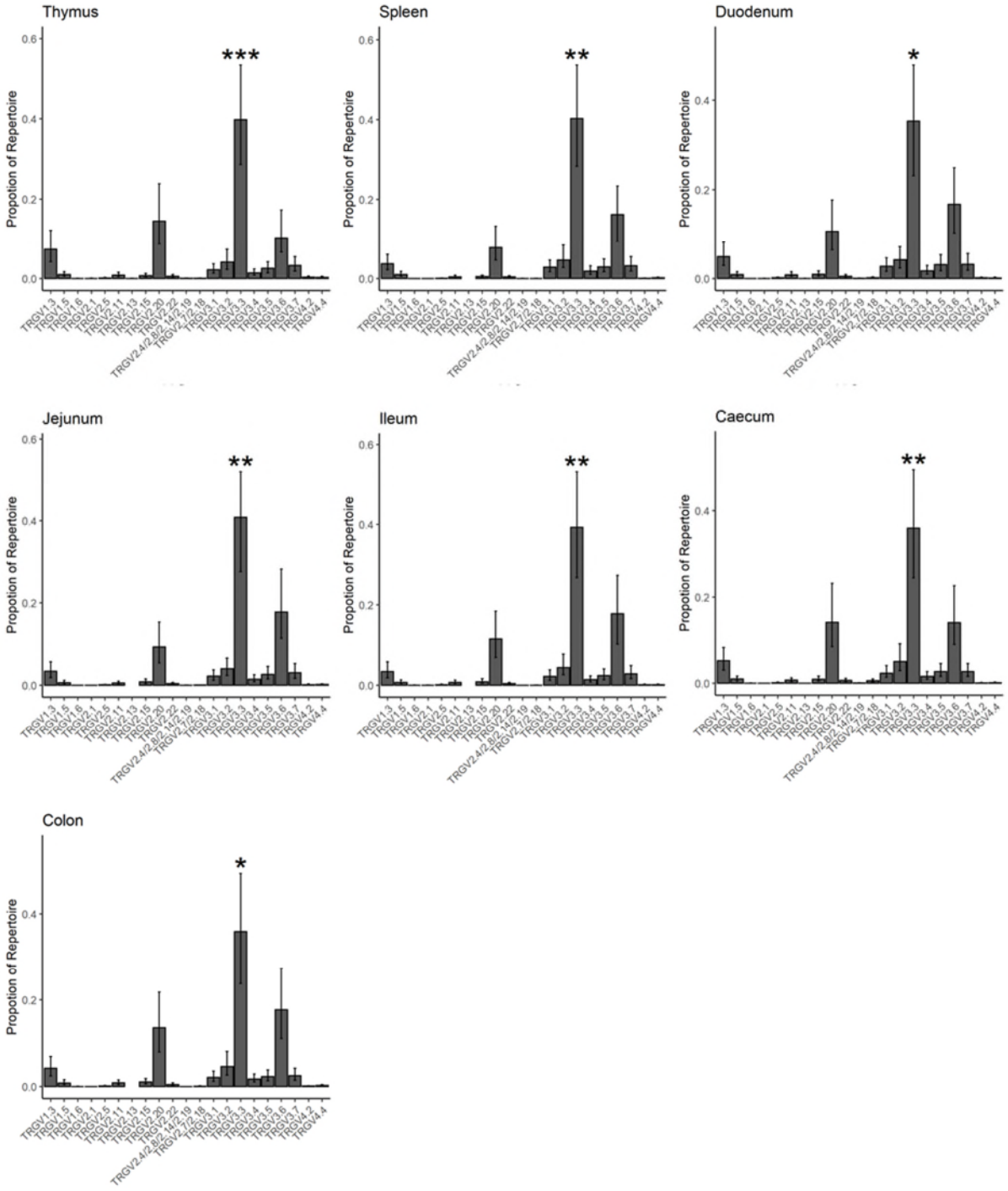


Figure 4.5 - The Proportion of each TRGV Gene Represented in TCR Gamma Chains within Different Tissues
The proportional usage of each TRGV gene in productively (in-frame) rearranged TCR gamma is displayed in thymus, spleen and various intestinal tissues extracted from ISA Brown chickens. Due to a high level of identity, some TRGV2 genes have been grouped. The TRGV genes on the X axis were all expressed in at least one tissue, pseudogenes and non-expressed genes are omitted. Bars represent the mean proportion for each TRGV gene. Error bars are 95% confidence intervals derived from 500 bootstrap replicates of an LMM using Bird number as a random variable and Tissue and TRGV gene as fixed variables. Proportions were logit transformed before model calculation. TRGV3.3 was expressed at greater levels than all other TRGV in all tissues. P values are comparisons between TRGV3.3 and the next most common TRGV gene in that tissue and were derived from the LMM using lmerTest with the Satterthwaite approximation of degrees of freedom. (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

Tissue samples from PA12 White Leghorn chickens were used for comparison of interline TRGV usage in spleen, jejunum, caecum and colon with the ISA Brown birds. The profiles between lines were very similar, and each included the expression of all potentially productive genes, with the exception of TRGV1.2 and TRGV2.9, although there were some small differences. TRGV1.4, TRGV3.5 and TRGV4.2 were present at a slightly higher proportion in all tissues of the PA12 White Leghorn birds, whereas TRGV1.3, TRGV3.1 and TRGV3.6 were more prominent in the ISA Browns (Fig. 4.6). The major difference between the two genetic backgrounds was in TRGV3.7, which represented a higher proportion of the TRGV profile in PA12 White Leghorn chickens compared with the ISA Brown birds. TRGV3.7 represented the third most prevalent TRGV gene in all tissues examined in the PA12 White Leghorn chickens, occupying 20% of all rearrangements compared with just 3% of rearrangements in the ISA Brown chickens. Both TRGV4.2 and TRGV4.4 were at higher levels in the spleen of PA12 White Leghorn chickens compared with other tissues in the PA12 background or any tissue in ISA Brown chickens.

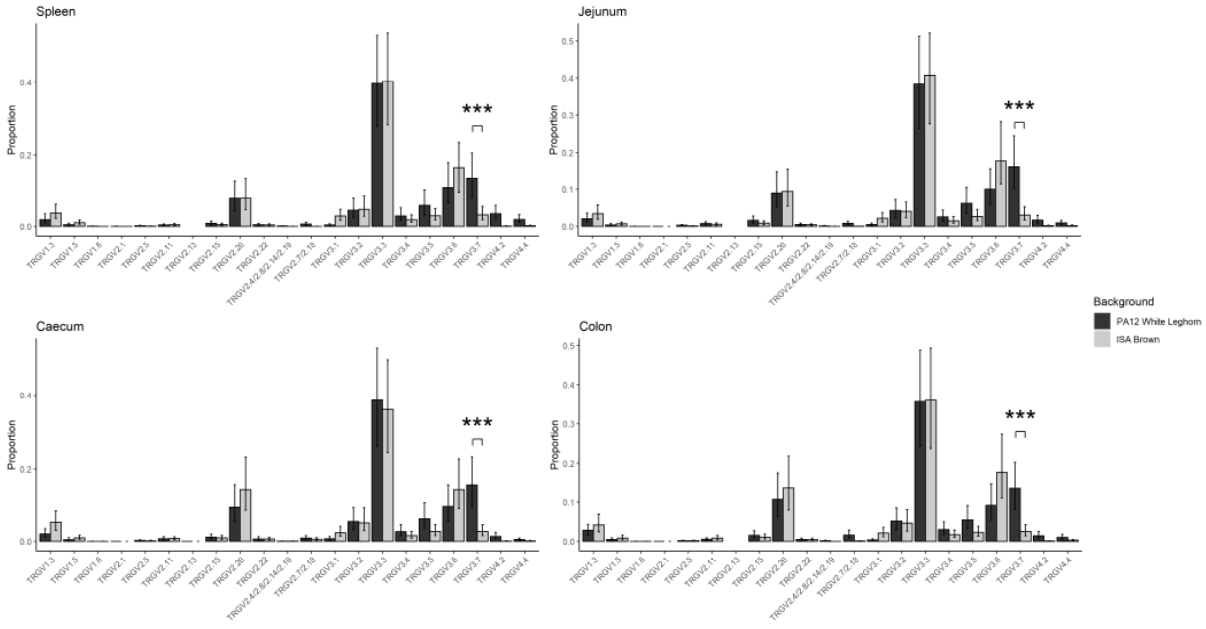


Figure 4.6 - TRGV Usage in Tissues of two Distinct lines of Chicken reveals Conserved Patterns of Expression
The proportional usage of each TRGV gene in productively (in-frame) rearranged TCR gamma is displayed in spleen and various intestinal tissues from ISA Brown and PA12 White Leghorn lines of chicken. Due to a high level of identity, some TRGV2 genes have been grouped. The TRGV genes on the X axis were all expressed in at least one tissue, pseudogenes and non-expressed genes are omitted. Bars represent the mean proportion for each TRGV gene displayed, with 95% confidence intervals derived from 500 bootstrap replicates of an LMM using Bird number as a random variable and Tissue and TRGV gene as fixed variables. Proportions were logit

transformed before model calculation. P values were derived from the LMM using lmerTest with the Satterthwaite approximation of degrees of freedom. (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

4.3.3 Private and Public Compartments within the Chicken TCR Gamma Repertoire

All TCR repertoires contain a mixture of CDR3 sequences that are unique to each individual, known as the private repertoire, and CDR3 sequences that are shared between individuals, known as the public repertoire. Patterns of publicity may be important in understanding the nature and function of the T cells carrying these receptors and, in the context of γδ T cells, whether these lymphocytes might have patterns of ligand recognition consistent with a role in innate or adaptive immunity. Although definitions of publicity vary, we use it here to refer to CDR3 nucleotide sequences that are found in multiple birds in any tissue and at any proportion. These public CDR3s can be shared between several individuals or the entire population.

For most of the chicken TRGV genes, the CDR3 repertoire was dominated by private sequences, with 80-100% of all CDR3s only found in single individuals (Fig. 4.7). In the case of public CDR3 sequences, most were found in all individuals with a lower proportion exhibiting 60% or 80% penetrance in the population. The public CDR3s were represented by TCR gamma rearrangements utilising a range of TRGV genes, but publicity was much more common in some genes than others, and three genes, TRGV1.6, 2.1 and 2.13, had no public sequences at all, at a 60% publicity cut-off. By far the highest level of publicity was found in TRGV2.7/2.18 containing sequences, 75% of which were public, most often to 100% penetrance (Fig. 4.7). The very highly represented TRVG3.3 rearrangement was dominated by private CDR3 sequences, with public CDR3s representing ~10% of the entire TRVG3.3+ve repertoire. A mostly similar pattern of public/private CDR3s was seen with PA12 White Leghorn chickens (Fig. A2).

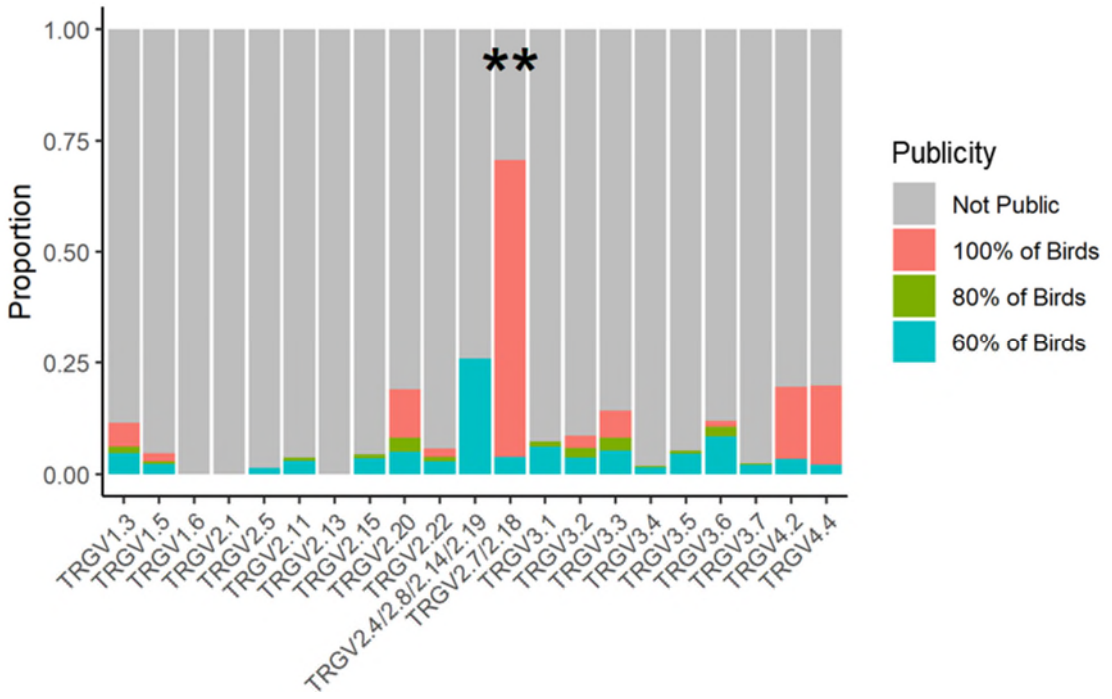


Figure 4.7 - The Proportion of Public TCR Gamma CDR3 Sequences Varies According to TRGV Gene Usage
Publicity is defined as a CDR3 nucleotide sequence found in 60%, 80% or 100% of ISA Brown birds (n=8) in any tissue at any frequency. The proportions of public clones in a given TRGV were calculated and displayed as a proportion of total sequences in the respective gene. Data is presented as stacked bars representing the proportion of the total repertoire for the particular TRGV. Due to a high level of identity, some TRGV genes have been grouped. P values are comparisons between the '100% of Birds' category of TRGV2.7/2.18 and the next most common TRGV gene, in this case TRGV4.4, and were calculated using an unpaired Wilcoxon signed rank test. (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

Having established that public CDR3 sequences comprise a significant proportion of the within-line repertoire in both PA12 White Leghorn and ISA Brown chickens, it was important to consider the similarities and differences between these chicken lines. The proportions of public sequences within each TRGV were broadly similar between both lines although overall a slightly higher fraction of the overall repertoire was totally public in the PA12 White Leghorn birds (Mean 0.111, SD 0.021) compared to the ISA Browns (Mean 0.044, SD 0.010), p=0.008. Within individual TRGVs these differences were significant for TRGV2.11, 2.15, 2.4/2.8/2.14/2.19, 3.4, 3.5 and 3.7. In contrast, the TRGV specific public repertoire was larger in ISA Brown birds for TRGV1.3 and 1.5, though these differences were not statistically significant. Within TRGV2.7/2.18, the public clones dominated the repertoire in both lines of chicken. Interestingly, the public repertoire of both PA12 and ISA Brown lines of chicken was largely shared (Fig 4.8B, 4.8C). For example, 70 to 80 percent of

the public repertoire comprised identical Complementarity Determining Region 3 Nucleotide (CDR3NT) sequences at the level of complete publicity (i.e. found in all birds in both lines). When considering the common public clones at an equal weighting, the proportion of shared clones (Fig 4.8C) was lower than when accounting for their relative contributions to the repertoire (Fig 4.8B). This pattern indicates that a significant proportion of the overall public repertoire in both lines was composed of “expanded” or “over-represented” clones.

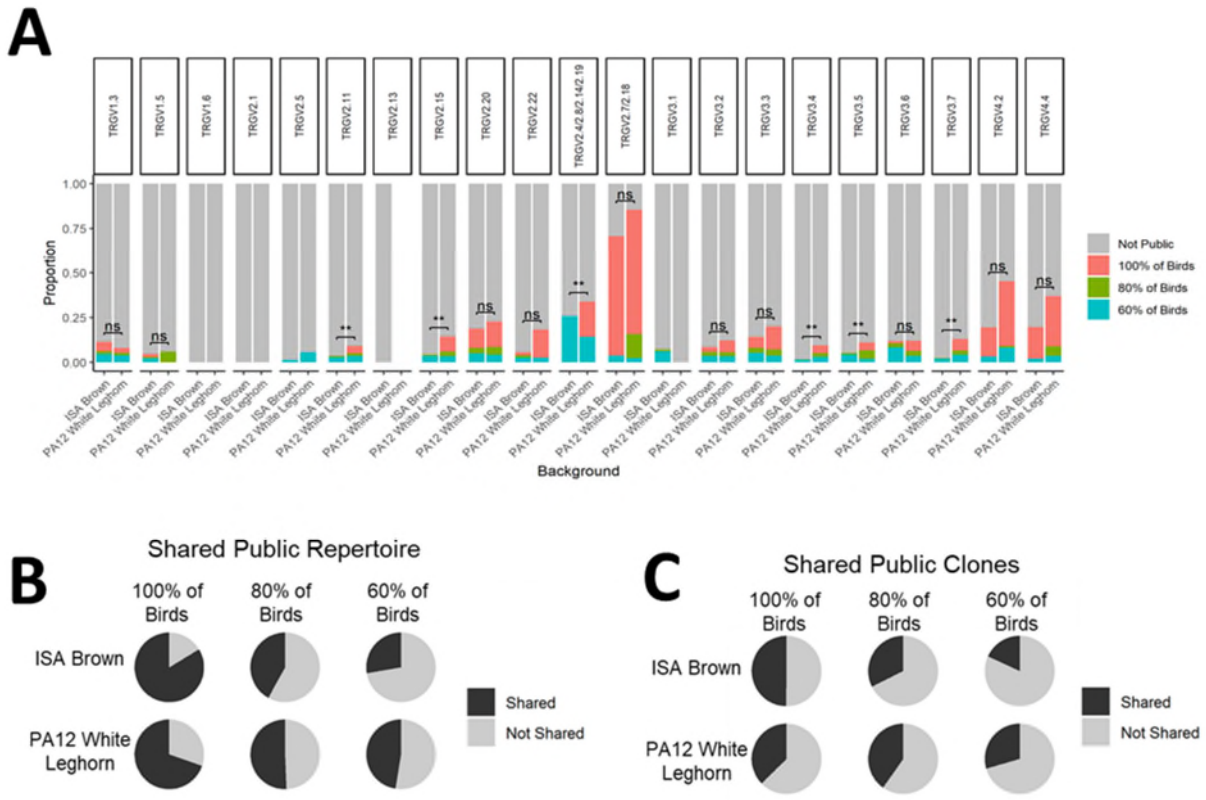


Figure 4.8 - Comparison of Publicity Profiles and Shared Public Clones between ISA Brown and PA12 White Leghorn Chickens

A) Publicity was defined as a CDR3 nucleotide sequence present in 100%, 80% and 60% of birds in any tissue at any frequency in ISA Brown and PA12 chicken lines (n=5/group). Publicity values were compared within TRGV genes and displayed as a proportion of all the clones within the population of TCRs defined by each TRGV. Due to a high level of identity, some TRGV genes have been grouped. P values are comparisons between the ‘100% of Birds’ category of the two lines, and were calculated using an FDR corrected unpaired Wilcoxon signed rank test. *** p value < 0.001, ** p value < 0.01, * p value < 0.05. TRGV2.13 was only detected in the repertoire of the ISA Brown birds.

B) The proportion of the overall public repertoire shared between the two lines was calculated. A public clone was considered to be shared between the two lines if it was present at the same publicity level or higher. These proportions are depicted for the 100%, 80% and 60% publicity categories for each line.

C) The number of public clones that were shared were identified using the same definition as in B. The proportion of shared clones are depicted for the 100%, 80% and 60% publicity categories for each line.

4.3.4 Diversity of the TCR Gamma Repertoire

Diversity indices are metrics that can be used to compare the structure of populations, in this case clones within the TCR gamma repertoire. Since CDR3 rarefaction curves for each sample make it clear that the novel CDR3 discovery rate has not reached a plateau for any of the samples (Fig. A3), diversity indices were calculated using rarefaction/extrapolation to a constant coverage of 0.1[32]. Repertoire profiles were compared using Hill numbers 0, 1 and 2, which correspond to species richness, the exponential of Shannon's entropy and the inverse of Simpson's concentration index respectively.

We first considered the diversity of the global TCR gamma repertoire of ISA Brown birds according to tissue. As expected, as a site of T cell development, the thymus samples contained the most diverse repertoire, as measured by all three Hill numbers, followed by the spleen, jejunum and ileum, and then by the duodenum and large intestine (Fig. 4.9). In PA12 White Leghorn birds, thymus samples were not available, and lower read depth reduced the power of the analysis to the extent that it is impossible to make concrete comparisons between the other tissues, but the spleen appears to have greater diversity than the gut tissues (Fig. A4).

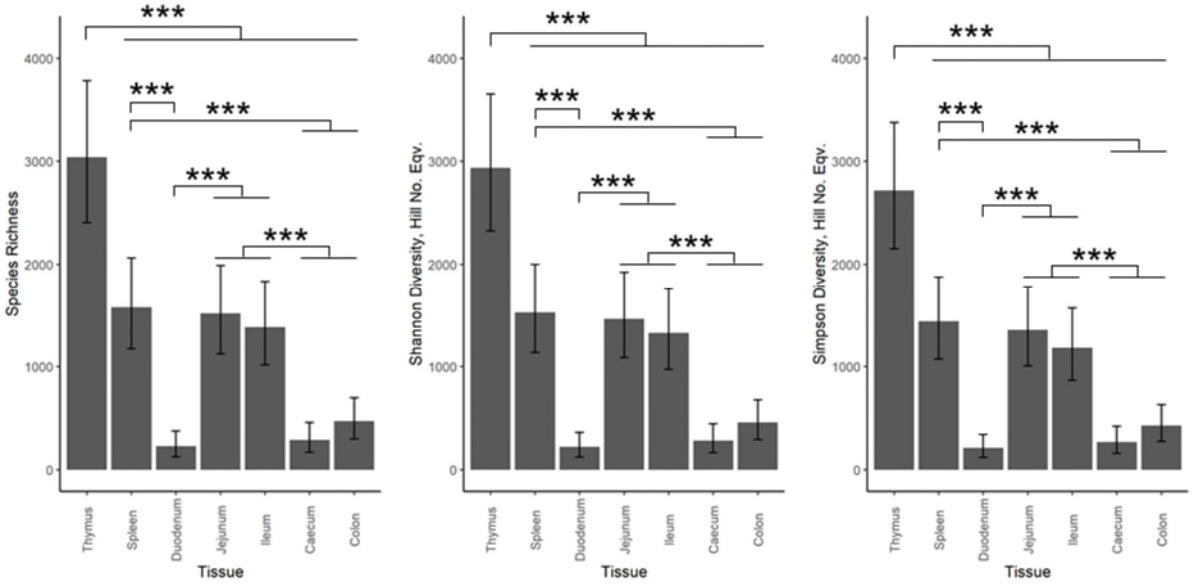


Figure 4.9 - Variation in the patterns of TCR gamma diversity in different tissues of ISA Brown chickens
Diversity indices were calculated from sample coverage rarefied abundance data for all unique clones in a tissue using iNEXT. Diversity indices are displayed as Hill number variants and thus represent effective species

counts - the number of evenly distributed species required to achieve the same diversity score. A linear model was constructed from the diversity index data using a cube root transform. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. Indications of significance are shown only for comparisons which exceeded these thresholds.

Since the overall level of expression and the proportion of public clones differed for each TRGV gene, it was important to explore whether the patterns of diversity also differed in each of these genes. These analyses were bound by the number of sequencing reads available for each gene, and therefore there was insufficient depth to meaningfully calculate diversity for the rarer genes. Of the more common genes, the clonal diversity of TRGV3.3 was higher than any other at all three Hill numbers in both ISA Brown (Fig. 4.10) and PA12 chickens (Fig. A5).

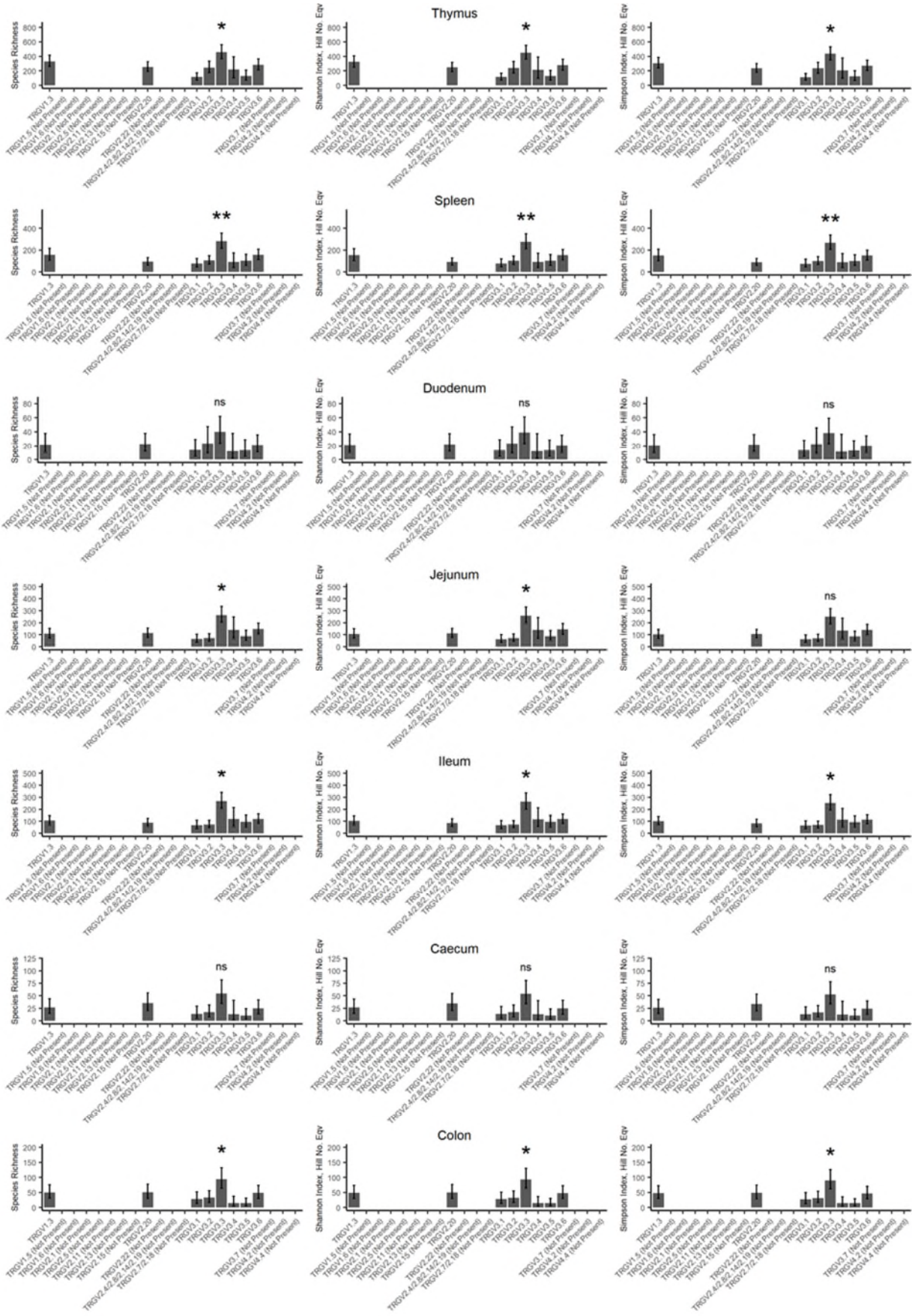

Thymus
Spleen
Duodenum
Jejunum
Ileum
Caecum
Colon
Species Richness
Shannon Index, Hill No. Eqv
Simpson Index, Hill No. Eqv
*
**
ns

Figure 4.10 - Diversity of TRGV rearrangements in different tissues
Diversity indices were calculated as for the aggregate ISA Brown chicken tissue diversity plots using all of the clones in a TRGV gene. Due to their widely varying expression levels, genes below a read count of 100 were omitted. This overcomes the issue of rarefying down to extremely low sample coverage and therefore discarding an unacceptable amount of data. 95% confidence intervals were calculated using iNEXT. A linear model was constructed from the diversity index data using a cube root transform. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values. P values are comparisons between TRGV3.3 and the next most diverse TRGV gene in that tissue Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

4.3.5 The Recombination Signal Sequence provides a Partial Explanation for Differential TRGV Gene Usage

Since the most highly expressed TRGVs were dispersed throughout the locus and therefore gene position was a poor indicator of TRGV gene usage, we considered whether other features, such as the Recombination Signal Sequence (RSS), might influence preferential TRGV expression. The RSS has the potential to influence the relative usage of TRGV genes and can be calculated as a Recombination Information Content (RIC) score[33]. RIC scores were calculated for each of the chicken TRGV RSSs using Bayesian models and expressed as natural logs, ranging from -1000 to 0 where 0 would be a perfect RSS and -1000 would represent a very poor RSS[33]. To avoid any complications associated with local peripheral tissue expansion of specific TRGVs, the proportional expression levels for each TRGV were taken from the thymus as the main site for developing T cells. The RIC scores for the TRGVs lay between -70 and -40 with the proportion of each TRGV ranging between <1% and 40% of the thymic TCR gamma repertoire. Plotting the RIC scores against the proportional expression of each TRGV gene in the thymus gives a weak positive correlation with an R^2 value of 0.1166. The P value from regression analysis does not quite reach a significance threshold of 0.05 (Fig. 4.11). The weakness of the correlation is in part due to the very high expression of TRGV3.3.

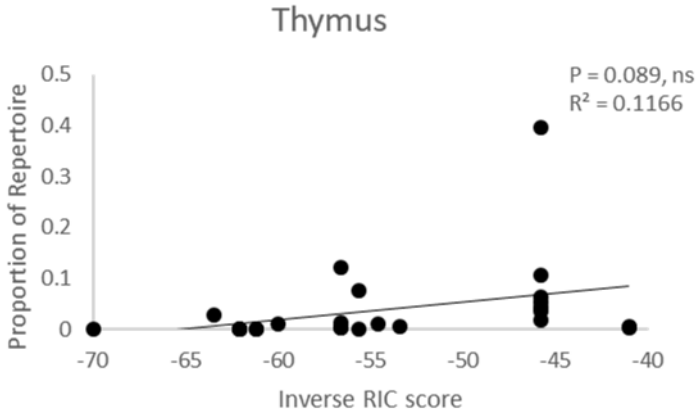


Figure 4.11 - Weak Positive Correlation between RSS Score and TRGV Gene Expression in Thymic TCR Gamma Repertoire
RIC scores were calculated using RSSsite and plotted against the proportion of the repertoire rearrangements attributable to each TRGV gene in the thymus of ISA Brown chickens. As there is currently no chicken RSS consensus, mouse RSS was used as a reference. These plots reveal a weak positive correlation but this only accounts for a minority of the variation we observe in the data. RIC scores are expressed as natural logs where 0 would be a perfect RSS and -1000 would be the worst. P and R^2 values were generated using regression analysis (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

4.4 Discussion

γδ T cells represent an enigmatic part of the immune system of all jawed vertebrates, and their retention over a prolonged period of evolution indicates that these cells play essential roles in the function of the immune system. To understand the role of γδ T cells, it is critical that functions are studied in diverse animal groups with both high and low circulating levels of γδ T cells. The differences in repertoire biology are likely to relate to the potential for γδ T cells to display a spectrum of different response types, ranging from adaptive to innate[34]. To enable repertoire studies, it is important that the TCR loci in question are accurate and well annotated. Although the most recent galGal6a genome assembly has high coverage (~82 fold) and includes long PacBio SMRT sequencing reads, the locus organisation was different to the recent BAC resequencing based TCR gamma locus reported by Liu et al. 2020[30]. The discrepancy was the presence of a 15kb potential duplication in the galGal6a assembly. Our PCR restriction-digest based strategy showed that the most likely organisation of this region of the genome was indeed that proposed by Liu et al. 2020[30]. Adding information from the repertoire sequencing revealed four further TRGV genes and showed

that most intact genes were productively rearranged and expressed. The final locus map now contains 40 TRGV genes, of which 14 are pseudogenes, divided into 4 families (by 75% identity), 3 TRGJ genes and a single TRGC gene (Table 4.1). There was also strong phylogenetic support for the TRGV2 family to be subdivided into four sub-groups (A, B, C and D). The chicken TCR gamma locus contains a larger number of TRGV genes than any other species (Table 4.1)[8], although there is very little within family diversification compared with mammals[35]. It is also noteworthy that chicken TRGV gene sequences are divergent from those found in mammals and do not fall into any of the conserved mammalian groups[36]. Moreover, all of the TRGVs are associated with a single TRGC as in duck[37], the only other bird that has been examined to date, and different to all of the currently resolved mammalian TCR gamma loci[36].

Species	***TRGV Gene Families***	***TRGV Genes***	***TRGJ Genes***	***TRGC Genes***
Chicken	4	40(27)	3	1
Duck	6	13(8)	5	1
Cow	10	17(17)	9	7
Sheep	11	13(11)	13	6
Human	6	15(6)	5	2
Mouse	5	7(7)	4	4

Table 4.12 - The Number of TRGV, J and C Genes in Various Animals
Numbers of TRGV, TRGJ and TRGC genes in the chicken compared with those of duck[37], cow, sheep, human, and mouse[53]. TRGV gene families were grouped according to IMGT subgroup concept of 75% sequence identity at the nucleotide level. Total numbers of TRGV genes are shown, with the number of functional TRGV genes, not including ORFs, shown in brackets. The mouse TCR gamma locus is based on multiple cassettes and one mouse TRGC is a pseudogene and therefore one TRGV gene is not used.

Confirmation of the TCR gamma locus sequence, the discovery of four additional TRGVs, and the assignment of gene families and the TRGV2 subfamilies collectively, revealed a pattern that suggests a block-based expansion of the locus by tandem duplication. Further evidence for this hypothesis is provided by the dot plot which shows multiple regions of high sequence similarity

dispersed throughout the locus. The patterns of similarity include both the TRGV motif patterns and extensive noncoding regions, providing strong evidence for tandem duplication. Duplication based expansion of TCR loci is widely reported with other animals[35] and the very high sequence similarity in non-coding and coding regions supports the premise that these represent recent events. The lengths of most intergenic regions are consistent and many of the differences can be identified as partial block duplication or deletion events. This extensive duplication suggests recent strong selective pressure for diversification of the TCR gamma region as reported for the TCR and other loci[35,38,39]. The level of identity within chicken TRGV family members is much higher than can be seen with other animals[35] which also supports the hypothesis of a recent duplication event. One driver for duplication might be to expand the repertoire of the γδ T cell population, however, the very high identity of some TRGV and the pseudogenisation of a number of others are not consistent with this hypothesis, though this may simply be due to a lack of time, which would allow diversification of function to develop. The motif most distal from the TRGC gene is likely to be the ancestral region as it possesses the longest array of duplicated TRGV genes with the most highly conserved gap sizes, though this is speculative. Moreover, the repertoire analysis shows that a single TRGV dominates the repertoire within each family with other family members being represented at very low levels. This potentially suggests a complex array of selective pressures expanding the region then limiting the use of the duplicated TRGVs. However, considering the genetic plasticity of other immune loci it is also possible that these duplication events are merely due to chance.

In terms of explaining the differential usage of TRGV genes, the RSS RIC score has some predictive value but only explains a small amount of the variation in thymic TRGV usage. Other features such as gene order can influence TRGV expression, particularly during foetal development[40]. However, this did not correlate with proportional expression of TRGV in the chicken thymus or other tissues. A more likely explanation is that usage may also be affected by promoter sequence or modifications such as methylation[40–42].

The similarity of TRGV usage profiles between tissues was surprising since in both humans and mice, there is a significant bias in TRGV usage in different tissues. In particular, with the foetal-thymus derived epithelial-associated subsets of γδ T cells in the mouse, TRGV5, 6 and 7, home preferentially to different sites within the body[43,44]. Similarly, TRGV9 expressing γδ T cells predominate in the blood of humans whereas TRGV4 is most common in the intestine[45]. At least some of these patterns of TRGV bias are influenced by epithelial expression of selective agents such as the butyrophilin-like molecules[45]. The influence of these types of molecule on TRGV repertoire remains undefined in the chicken, or any other γδ high animal, although the lack of a strong tissue-associated bias in chicken TRGV expression would need to be considered in such studies.

In terms of the repertoire of chicken γδ T cells, this population contains a high diversity of CDR3s, especially within the dominantly expressed TRGV genes such as TRGV3.3. A large proportion of this diversity was represented as private CDR3 sequences. That said, there were also substantial fractions of the overall repertoire that can be considered highly public and were found in a high proportion, often 100%, of the birds in this study. Moreover, the proportion of these hyper-public CDR3s varied between TCR gamma rearrangements that utilised different TRGV genes and the patterns of publicity were remarkably similar between PA12 and ISA Brown lines of chicken. The biology of these public versus private repertoires could be critical in defining the function of chicken γδ T cells. Indeed, the skin and uterine/tongue focussed murine γδ T cells represent hyper-public canonically rearranged versions of TRGV5 and TRGV6 respectively[16]. The Vγ5Vδ1+ DETCs in mice have been shown to play important roles in tumour surveillance, inflammation and epithelial cell regulation[19,20]. The Vγ9+Vδ2+ T cells found in the blood of humans are a semi-invariant population of γδ T cells where approximately 80% of clones contain public CDR3s[46,47]. Interestingly, these Vγ9+Vδ2+ T cells have been shown to respond to pyrophosphate antigens and are activated during infection with pathogens such as *M. tuberculosis*[48]. Highly public CDR3s were detected in chicken TCR gamma, and the clones which were present in 100% of birds made up over 70% of the rearrangements expressing TRGV2.7/2.18, which suggests a specialised function for cells bearing this

receptor. Moreover, many of the public clones, and their proportions within the repertoire, were shared between two genetically distinct lines of chicken, indicating the ubiquity of the public TRGV repertoire. The sub-division of γδ T cells according to publicity could be important in understanding the spectrum of roles fulfilled by γδ T cells in different species.

The diversity of TCR gamma in the chicken is highest in the thymus, which is in line with expectations, as this is the primary site where γδ T cell maturation occurs[49], and would therefore potentially contain all the TCR gamma rearrangements that exist in the chicken, though some debate remains about the existence and function of extrathymically derived γδ T cells in some species[49–51]. On the other hand, the large intestinal tissues, caecum and colon, being less diverse than jejunum and ileum is contrary to our expectations. Considering the role of γδ T cells in the early response to pathogens at epithelial sites[16], we might expect to see a greater receptor diversity in the caecum and colon as these tissues host the major microbial population in the gut and therefore the greatest "antigenic" diversity. This result potentially implies that γδ T cells may be more important in responses to microbes in the small intestine, or that they are carrying out some other function in this part of the gut that requires greater receptor diversity. Furthermore, we showed that TRGV gene diversity correlated with proportional expression, therefore we conclude that the TRGV genes with the highest diversity are rearranged preferentially. The results presented here represent the state of the "mature" TCR gamma repertoire, as the birds examined were all 7-8 weeks of age, and therefore in a period of stable γδ T cell export from the thymus, about 4 weeks after the last of the three "waves of thymic γδ T cell emigration"[52] . It will be of interest to consider the dynamics of TRGV repertoire patterns in younger birds during the pre-hatch and early post hatch periods.

4.5 Conclusion

The regional differences in repertoire may reflect the numbers of γδ T cells in these different tissues and could influence how different tissue resident populations of γδ T cells interact with the outside world, including pathogens and the microbiota. The TRGV repertoire analyses used here

represent an important tool for use in future studies and have been critical in understanding the TCR gamma locus in chickens. This work is a starting point for understanding the role for γδ T cells in chickens and more broadly in γδ high animals that may be more reliant on these cells than the more commonly studied γδ low animals. A broad understanding of γδ T cell biology and the driving forces that have underpinned their retention across the jawed vertebrates will help us understand the evolution of immune capabilities and may reveal new ways for us to exploit these cells to protect humans and livestock species.

4.6 References

1. Flajnik MF. Re-evaluation of the immunological big bang. Curr Biol. 2014;24(21):R1060–5.
2. Cooper MD, Alder MN. The evolution of adaptive immune systems. Cell. 2006;124(4):815–22.
3. Attaf M, Legut M, Cole DK, Sewell AK. The T cell antigen receptor: The Swiss army knife of the immune system. Clin Exp Immunol. 2015;181(1):1–18.
4. Paul S, Lal G. Regulatory and effector functions of gamma-delta (γδ) T cells and their therapeutic potential in adoptive cellular therapy for cancer. Int J Cancer. 2016;139(5):976–85.
5. Prinz I, Silva-Santos B, Pennington DJ. Functional development of γδ T cells. Eur J Immunol. 2013;43(8):1988–94.
6. Bonneville M, O'Brien RL, Born WK. γ δ T cell effector functions: A blend of innate programming and acquired plasticity. Nat Rev Immunol. 2010;10(7):467–78.
7. Sumaria N, Roediger B, Ng LG, Qin J, Pinto R, Cavanagh LL, et al. Cutaneous immunosurveillance by self-renewing dermal γδ T cells. J Exp Med. 2011;208(3):505–18.
8. Holderness J, Hedges JF, Ramstead A, Jutila MA. Comparative biology of γδ T cell function in humans, mice, and domestic animals. Vol. 1, Annual Review of Animal Biosciences. 2013. p. 99–124.
9. Hayday AC. γδ Cells: A right time and a right place for a conserved third way of protection. Annu Rev Immunol. 2000;18:975–1026.
10. Faldyna M, Levá L, Knötigová P, Toman M. Lymphocyte subsets in peripheral blood of dogs - A flow cytometric study. Vet Immunol Immunopathol. 2001;82(1–2):23–37.
11. Sowder JT, Chen CLH, Lanier Ager L, Chan MM, Cooper MD. Large subpopulation of avian T cells express a homologue of the mammalian Tγ/δ receptor. J Exp Med. 1988;167(2):315–22.
12. Mackay CR, Hein WR. A large proportion of bovine T cells express the γδ T cell receptor and show a distinct tissue distribution and surface phenotype. Int Immunol. 1989;1(5):540–5.
13. Piriou-Guzylack L, Salmon H. Membrane markers of the immune cells in swine: an update. Vet Res. 2008;39(54).

14. Caro MR, Gallego MC, Buendía AJ, Navarro E, Navarro JA. Postnatal evolution of lymphocyte subpopulations in peripheral blood and lymphoid organs in the goat. Res Vet Sci. 1998;65(2):145–8.

15. Yirsaw A, Baldwin CL. Goat γδ T cells. Dev Comp Immunol. 2021;114(July 2020):103809.

16. Khairallah C, Chu TH, Sheridan BS. Tissue Adaptations of Memory and Tissue-Resident Gamma Delta T Cells. Front Immunol. 2018;9:1–22.

17. Ramsburg E, Tigelaar R, Craft J, Hayday A. Age-dependent Requirement for γδ T Cells in the Primary but Not Secondary Protective Immune Response against an Intestinal Parasite. J Exp Med. 2003;198(9):1403–14.

18. Komano H, Fujiura Y, Kawaguchi M, Matsumoto S, Hashimoto Y, Obana S, et al. Homeostatic regulation of intestinal epithelia by intraepithelial gamma delta T cells. Proc Natl Acad Sci U S A. 1995;92(13):6147–51.

19. Girardi M, Glusac E, Filler RB, Roberts SJ, Propperova I, Lewis J, et al. The distinct contributions of murine T cell receptor TCRγδ+ and TCRαβ+ T cells to different stages of chemically induced skin cancer. J Exp Med. 2003;198(5):747–55.

20. Roberts SJ, Smith AL, West BA, Wen L, Findly CR, Owen MJ, et al. T-cell αβ+ and γδ+ deficient mice display abnormal but distinct phenotypes toward a natural, widespread infection of the intestinal epithelium. Proc Natl Acad Sci. 1996;93:11774–9.

21. Smith AL, Hayday AC. An αβ T-cell-independent immunoprotective response towards gut coccidia is supported by γδ cells. Immunology. 2000;101(3):325–32.

22. Chien Y, Meyer C, Bonneville M. γδ T Cells: First Line of Defense and Beyond . Annu Rev Immunol. 2014;32(1):121–55.

23. Schat KA, Kaspers B, Kaiser P. Avian Immunology. 2014.

24. Carding SR, Egan PJ, Walter T, Hall E. γδ T Cells : Functional Plasticity and Heterogeneity. Immunology. 2002;2(May):336–45.

25. Pieper J, Methner U, Berndt A. Heterogeneity of avian γδ T cells. Vet Immunol Immunopathol. 2008;124(3–4):241–52.

26. Fenzl L, Göbel TW, Neulen ML. γδ T cells represent a major spontaneously cytotoxic cell population in the chicken. Dev Comp Immunol. 2017;73:175–83.

27. Berndt A, Methner U. Gamma/delta T cell response of chickens after oral administration of attenuated and non-attenuated Salmonella typhimurium strains. Vet Immunol Immunopathol. 2001;78(2):143–61.

28. Choi KD, Lillehoj HS. Role of chicken IL-2 on γδ T-cells and Eimeria acervulina-induced changes in intestinal IL-2 mRNA expression and γδ T-cells. Vet Immunol Immunopathol. 2000;73(3–4):309–21.

29. Arstila TP, Lassila O. Androgen-induced expression of the peripheral blood gamma delta T cell population in the chicken. J Immunol. 1993;151(12):6627–33.

30. Liu F, Li J, Lin IYC, Yang X, Ma J, Chen Y, et al. The Genome Resequencing of TCR Loci in Gallus gallus Revealed Their Distinct Evolutionary Features in Avians . ImmunoHorizons. 2020;4(1):33–46.

31. Six A, Rast JP, McCormack WT, Dunon D, Courtois D, Li Y, et al. Characterization of avian T-cell

receptor γ genes. 1996;93:15329–34.

32. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol Evol. 2016;7(12):1451–6.

33. Merelli I, Guffanti A, Fabbri M, Cocito A, Furia L, Grazini U, et al. RSSsite: A reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. 2010;38:262–7.

34. Hayday AC. γδ T Cell Update: Adaptate Orchestrators of Immune Surveillance. J Immunol. 2019;203(2):311–20.

35. Su C, Jakobsen I, Gu X, Nei M. Diversity and evolution of T-cell receptor variable region genes in mammals and birds. Immunogenetics. 1999;50(5–6):301–8.

36. Antonacci R, Massari S, Linguiti G, Jambrenghi AC, Giannico F, Lefranc MP, et al. Evolution of the T-cell receptor (TR) Loci in the adaptive immune response: The tale of the TRG locus in mammals. Genes (Basel). 2020;11(6).

37. Yang Z, Sun Y, Ma Y, Li Z, Zhao Y, Ren L, et al. A comprehensive analysis of the germline and expressed TCR repertoire in White Peking duck. Sci Rep. 2017;7(September 2016):1–13.

38. Kiemnec-Tyburczy KM, Richmond JQ, Savage AE, Lips KR, Zamudio KR. Genetic diversity of MHC class i loci in six non-model frogs is shaped by positive selection and gene duplication. Heredity (Edinb). 2012;109(3):146–55.

39. Peatman E, Liu Z. Evolution of CC chemokines in teleost fish: A case study in gene duplication and implications for immune diversity. Immunogenetics. 2007;59(8):613–23.

40. Xiong N, Baker JE, Kang C, Raulet DH. The genomic arrangement of T cell receptor variable genes is a determinant of the developmental rearrangement pattern. Proc Natl Acad Sci U S A. 2004;101(1):260–5.

41. Migone N, Padovan S, Zappador C, Giachino C, Bottaro M, Matullo G, et al. Restriction of the T-cell receptor V delta gene repertoire is due to preferential rearrangement and is independent of antigen selection. Immunogenetics. 1995;42(5):323–32.

42. Tunyaplin C, Knight KL. Fetal VDJ gene repertoire in rabbit: evidence for preferential rearrangement of VH1. Eur J Immunol. 1995;25(9):2583–7.

43. Guy-Grand D, Vassalli P, Eberl G, Pereira P, Burlen-Defranoux O, Lemaitre F, et al. Origin, trafficking, and intraepithelial fate of gut-tropic T cells. J Exp Med. 2013;210(9):1839–54.

44. Vantourout P, Hayday A. Six-of-the-best: Unique contributions of γδ T cells to immunology. Nat Rev Immunol. 2013;13(2):88–100.

45. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.

46. Davey MS, Willcox CR, Hunter S, Kasatskaya SA, Remmerswaal EBM, Salim M, et al. The human Vδ2+ T-cell compartment comprises distinct innate-like Vγ9+ and adaptive Vγ9- subsets. Nat Commun. 2018;9(1):1–14.

47. Delfau M -H, Hance AJ, Lecossier D, Vilmer E, Grandchamp B. Restricted diversity of Vγ9-JP rearrangements in unstimulated human γ/δ T lymphocytes. Eur J Immunol. 1992;22(9):2437–43.

48. Willcox CR, Davey MS, Willcox BE. Development and selection of the human Vγ9Vδ2+ T-Cell Repertoire. Front Immunol. 2018;9:1–7.

49. Dunon D, Cooper MD, Imhof BA. Thymic origin of embryonic intestinal γ/δ T cells. J Exp Med. 1993;177(2):257–63.

50. Chen CH, Göbel TW, Kubota T, Cooper MD. T cell development in the chicken. Poult Sci. 1994;73(7):1012–8.

51. Cihak J, Losch U, Hoffmann-Fezer G, Chen CH, Cooper MD, Ziegler-Heitbrock HWL. In Vivo Depletion of Chicken T-Cell Subsets. Scand J Immunol. 1993;38(2):123–9.

52. Dunon D, Courtois D, Vainio O, Six A, Chen CH, Cooper MD, et al. Ontogeny of the immune system: γ/δ and α/β T cells migrate from thymus to the periphery in alternating waves. J Exp Med. 1997;186(7):977–88.

53. Lefranc M. IMGT [Internet]. 2021.

Chapter 5

Interactions between γδ T cells and the Enteric Microbiota in the Chicken

5.1 Introduction

The gut microbiota in chickens as in other animals can be consider a part of the enteric immune system and is able to prevent pathogens from colonising by simple competition, reduction of adherence and production of antagonistic metabolites and bacteriocins[1]. In mouse and other mammals, it has also been shown that the microbiota plays a role in immune development and function[2,3]. Studies in germ free (GF) birds confirm a similar relationship in the chicken. In addition to exhibiting a range of differences in enteric tissue, such as a reduction in vascularity, enzyme activity and smooth muscle thickness, GF birds have been shown to display reduced expression of gut immune factors such as immunoglobulins and cytokines[4], as well as immature goblet cells and mucin profiles[5,6], reduced B cells and a lack of caecal tonsil germinal centres[7,8]. Additionally, these GF animals display a reduction in the size of Peyer's patches (PP) and intra-epithelial lymphocyte numbers (IEL). More specifically, members of the 2 of the phyla which dominate the chicken gut at different stages of life, *Firmicutes* and *Bacteroidetes*, have been shown to modulate cytokine expression in different ways at different ages[9]. For example, the commensal genera *Faecalibacterium,* of the phyla *Bacteroidetes,* is sometimes given as a probiotic, and was shown to reduce inflammation by inducing the downregulation of pro-inflammatory cytokines such as IL-1β, IL-18 and IL-6, whilst upregulating the anti-inflammatory cytokine TGF-β4.

γδ T cells present in the gut carry out a range of important functions including rapid response to pathogens[10], regulation of enterocyte turnover[11], protection against epithelial tumours[12], and limitation of inflammatory damage[13]. As γδ T cells are enriched in the epithelial layer, where they are present as part of a population of IELs, they are in constant intimate contact with various members of the microbiota. Therefore, as has been shown for other immune cells, we would expect the microbiota to play a role in γδ T cell development. However, despite the increasing interest in this population, their development relative to the microbiota has been little studied in γδ high animals, such as chickens. These γδ high species possess a high proportion of circulating T cells,

which express a γδ T cell receptor (>10% T cells by proportion). γδ low species include mice and humans and therefore γδ T cells in γδ high animals are relatively poorly studied.

Despite early studies showing no difference in numbers of γδ T cells in GF mice[14,15], more recent experiments show that in a Dextran Sodium Sulfate (DSS) gut inflammatory model, stimulation of tissue healing by γδ T cells was impaired in the absence of a microbiota[16]. Furthermore, γδ T cell mediated healing of epithelial tissue was shown to be restored in mice born GF and reconventionalised during adulthood by microbial colonisation. In humans, a recent study showed that the huge increase in Vγ9Vδ2+ T cells in blood after birth was due to polyclonal expansion and the authors suggest this is due to the transition from a sterile to a non-sterile environment during birth and the microbial colonisation that occurs at this time[17].

γδ T cells have also been shown to exert an influence over the microbiota and it was shown that in mice lacking γδ T cells, mucosal penetration by commensal microbes was increased following intestinal injury[16]. This is unsurprising considering that expression of antimicrobial factors such as REGIIIγ and chemotactic cytokines by γδ T cells was also shown to be impaired in GF animals. Curiously, this wasn't the case for all antimicrobials and upregulation of lysozyme was shown to be independent of the microbiota. This γδ T cell activation and subsequent release of antimicrobial factors to control the microbiota and maintain homeostasis was shown to be mediated by epithelial cells which seem to act as something of a surveillance system[18]. In addition, gut epithelial cells were also shown to be involved in facilitating γδ T cell motility and chemotaxis to sites of infection[19].

However, chicken has a far higher proportion of circulating γδ T cells, suggesting the increased importance of these cells. Despite this, studies looking at the influence the microbiota can have over gut γδ T cells in avians is lacking. A recent study investigating the impact of an adult vs immature microbiota on broiler chickens showed no difference between these groups in terms of percentages of γδ T cells overall, activation or CD8αα+ or CD8αβ subsets[20].

5.2 Chapter Outline

This chapter seeks to begin to address the lack of studies examining γδ T cell microbiota interactions in γδ high animals. We first use our repertoire sequencing method to compare the TCR gamma repertoires of conventional and complex flora vs GF birds. We examine the gamma repertoire in detail at various different scales, from population structure, down to the level of individual clones. We then examine the inverse relationship and use 16S sequencing to look at the impacts on the microbiota in the absence of γδ T cells, using a recently constructed TCRγ -/- bird line. Using these techniques, we are able to examine γδ T cell-microbiota interactions from the perspective of the influence of microbiota on γδ T cells and the reverse relationship. As chickens are, to our knowledge, the first γδ high animal these analyses have been performed on, this also has implications for various other species within this group.

5.3 Results

5.3.1 Study Design

The different groups of chickens used in this study are Complex, Germ Free, Conventional and Tetra. All birds are PA12 white leghorns. In order to somewhat reduce the variation in the microbiota, and therefore reduce the variation in the TCRG repertoire if the microbial status of a bird has a substantial effect on it, Complex birds were generated by re-conventionalising birds born Germ Free, using an oral gavage of caecal contents collected from a single healthy SPF hen. 5 Germ Free and 5 Complex birds were sacrificed at Days 14, 28 and 55 for a total of 15 birds per group. Additionally, 5 Conventional birds were also included in the study and sampled at Day 55 to provide a reference for the effect of a naturally occurring microbiota. Finally, 3 Tetra colonised birds, born Germ Free and reconstituted with a defined culture of *E. coli* Nissle, *Enterococcus faecium, Clostridium butyricum, Lactobacillus rhamnosus,* were also included to study the effects of a more limited flora, these birds were again sampled at Day 55 only. It should be noted that rearing of the Conventional and Tetra birds occurred in parallel at a separate time to the Complex and Germ Free

birds. Bursa, Spleen, Jejunum, Caeca and Colon tissues were harvested from sacrificed birds and stored in RNAlater at -80°C.

5.3.2 Tissue Sample Preparation and Processing for TCR Gamma Repertoire Sequencing

Chicken tissue samples were processed, RNA extracted, cDNA preparation performed using 5’ Rapid Amplification of cDNA Ends (RACE) and PCRs run as detailed in the methods section. Though the majority of samples result in a band at the expected size (650bp), some do not display this product, but are represented in another of the triplicate PCRs (Fig 5.1). The concentrations of each of these products are estimated by comparison to the 100bp NEB ladder, which informs the later pooling steps.

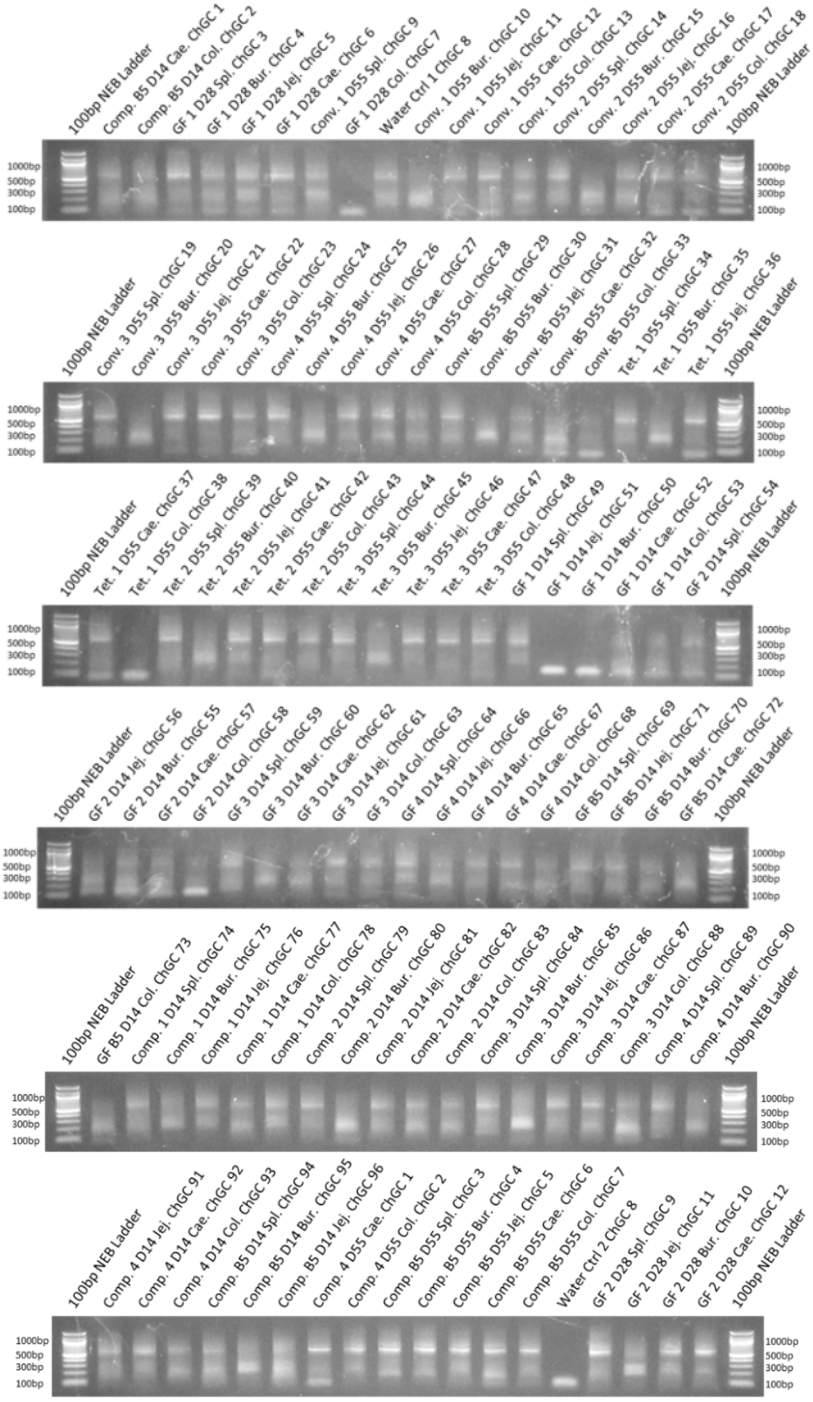


Figure 5.1 - Example Library of TCR Gamma PCRs from 5' RACE Generated Chicken cDNA shows a Range of Amplicon Concentrations

Agarose gel image showing amplicons generated from repertoire PCRs performed on cDNAs produced using 5'RACE on RNA extracted from the Bursa, Spleen, Jejunum, Caecum and Colon of Complex, Germ Free, Conventional and Tetra colonised birds. These cDNAs were amplified using a primer annealing to the single chicken TRGC region and a UPA long and short primer combination, resulting in universal amplification of all TRGV gene rearrangements in a single reaction. Sizes of the 100bp NEB ladder bands are indicated in bp at either side of the gel image and the desired band length is ~600bp. Bird group names and tissue names were shortened for labelling as follows; Bird Groups - Conventional (Conv.), GermFree (GF), Complex. (Comp.), Tetra (Tet.), Tissues – Bursa (Bur.), Spleen (Spl.), Jejunum (Jej.), Caecum (Cae.), Colon (Col.).

After pooling and library preparation was performed as detailed in the methods section (See Chapter 2), the Tapestation plot revealed that Library 5 had an unacceptably high concentration of a product at ~125bp, assumed to be adaptor dimer, relative to the desired 700bp band (Fig 5.2). Adapter dimers, due to their role in the sequencing process are able to bind to the flow cell and generate sequence data, competing with the desired amplicons. This is a particular problem as due to their small size, the number of molecules will often outnumber those of the desired product, despite them appearing at a similar intensity on the Tapestation gel. This library was therefore omitted and not taken forward to the sequencing stage.

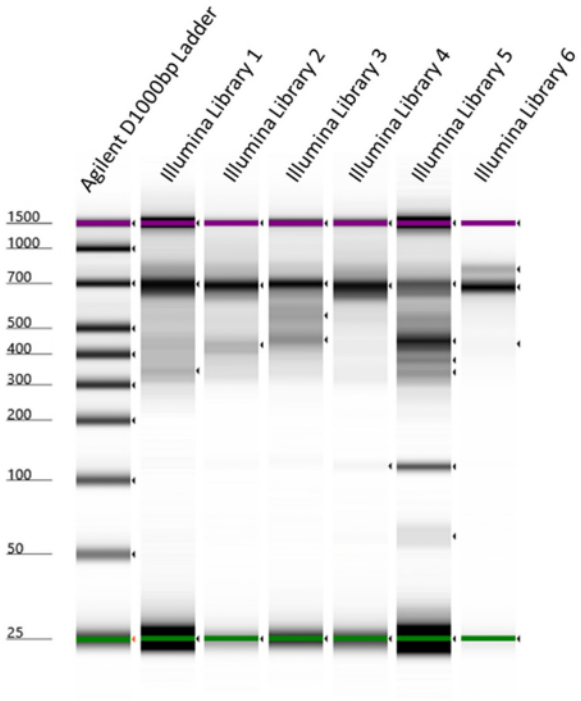


Figure 5.2 - DNA Tapestation of NEBNext prepared MiSeq Amplicon Libraries
Agilent TapeStation image shows the libraries generated using the NEBNext Library Prep Kit for Illumina which were visualised on a DNA ScreenTape (Agilent) to confirm they were of the expected size and free of any contaminants. All of the Libraries show the expected product at 700bp, and library 5 also shows an adaptor dimer band at ~125bp. Ladder band sizes are indicated in bp along the side of the gel image. The upper purple and lower green band are automatically generated by the Agilent software and are manually checked to ensure it has detected the largest and smallest ladder bands correctly.

The initial MiSeq run produced 48,848,718 paired Reads Passing filter (PF) with an average Q30 of 80.3%. However, due to an order of magnitude loading error resulting in Library 6 accounting for 10x more of the library concentration than intended, a subsequent run was conducted with the aim of increasing the read numbers of the other libraries. This run produced 38,692,024 paired PF reads with a Q30 of 88.2% and increased the read counts of the other libraries such that Library 6 had ~5x as many reads as any of the other libraries. For reference, Fig. A6 contains details of which samples were included in each library, though the omission of Library 5 and over-representation of Library 6 should not affect any downstream results.

The reads were then processing using our in house python package reptools (Fig 5.3). The desired barcoded primer stage involves searching the MiSeq reads for one of the expected primer and barcode combinations which passes a FASTQ quality filter. The majority of the read loss occurs at this stage and is around 2/3 of all reads on average. This is mainly due to reads which have a barcoded ChGC primer but no UPA primer. The next processing stage involves BLAST searching the reads against a known database of T Receptor Gamma (TRG) V, J and C regions, created from our annotated chicken TCR gamma locus annotation (See Chapter 4). The CDR3 region is then extracted from these reads. This region is located by identification of the conserved conserved cysteine and phenylalanine residues which mark the junction between the TRGV and J.This stage represents another substantial loss of reads, which are accounted for by incomplete recombination, intergenic sequences or concatenated C gene segments[21]. Following this, any sequences which are associated with unexpected barcodes, presumably present due to sequencing error, are removed. Finally, any sequences which are non-functional are removed from the analysis. These are reads which contain frame shift mutations or incomplete codons. The processed reads are then passed on to our in house repertoire analysis package RepertoiR for analysis.

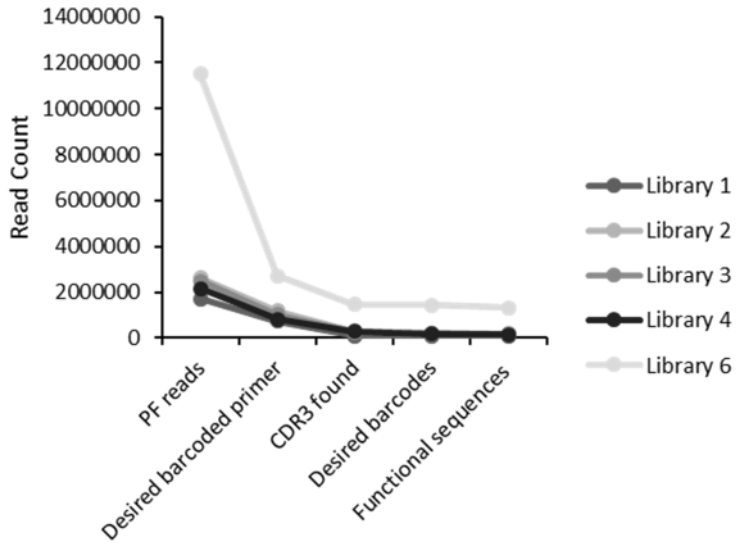


Figure 5.3 – Plot of Numbers of Reads at each Stage of Repertoire Sequence Processing
Plot shows the number of reads at each stage of processing following Illumina MiSeq sequencing. The PF reads stage refers to sequences which pass the internal quality filter within the MiSeq instrument, Desired barcoded primer refers to reads for which an expected primer and barcode combination was found and both pass their respective quality filtering. At the CDR3 found stage, the reads are searched against a BLAST database of possible TRGV, J and C sequences and the Complementarity Determining Region 3 Nucleotide (CDR3NT) region extracted by identification of conserved residues. The Desired barcode stage removes unexpected barcodes, which may be present due to contamination or sequencing error. Finally, the Functional sequences stage removes any CDR3 which is out of frame or includes an incomplete codon. Samples used in this study were split across 2 libraries and 3 replicates were performed for a total of 6 libraries. However, Library 5 failed the Tapestation quality control stage and was omitted from sequencing.

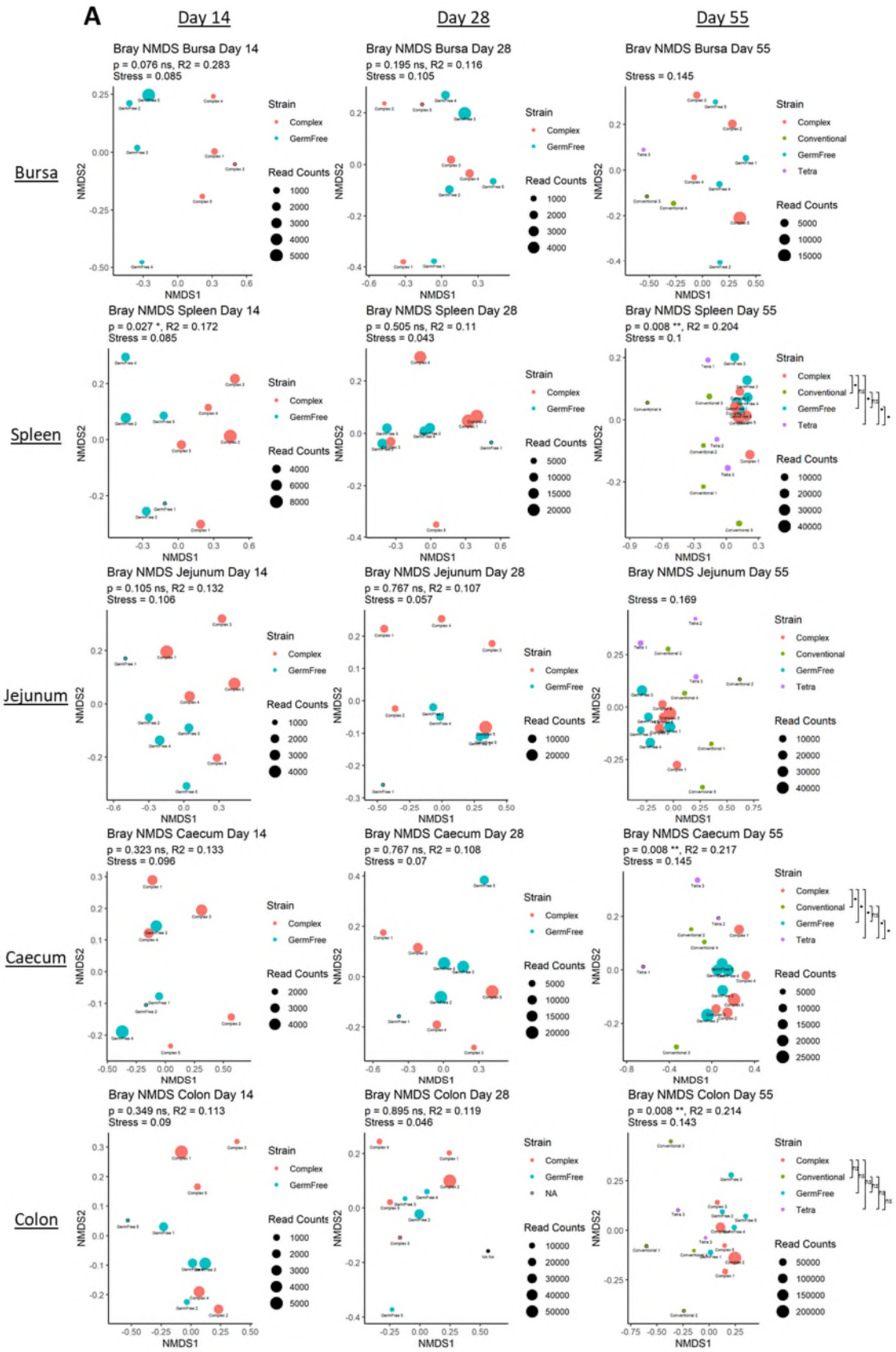
5.3.3 Beta-Diversity plots reveal small Differences in the Germ Free group

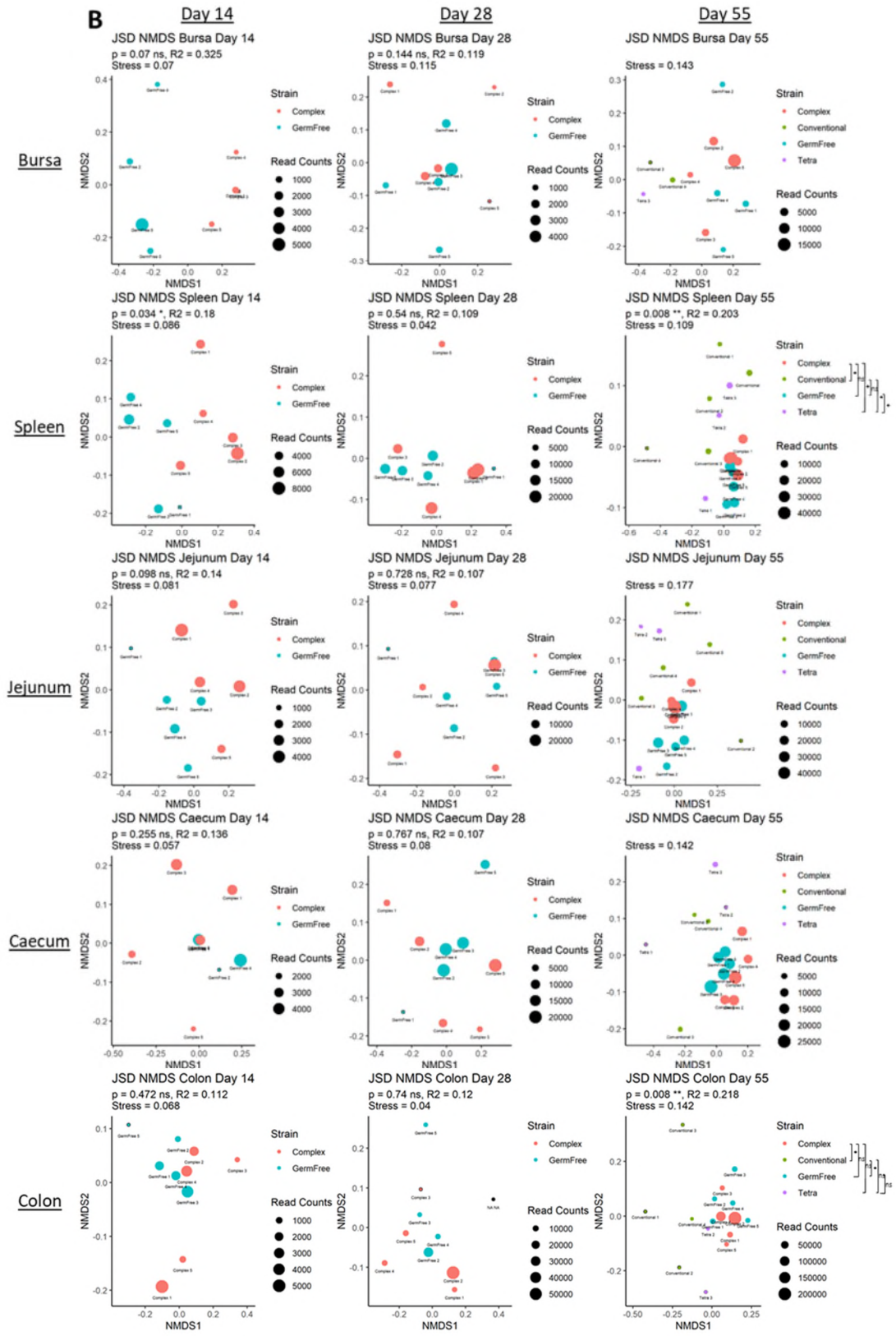
Bray-Curtis distance is a common metric for immune repertoire studies since it is easy to calculate and interpret[22–24]. It also allows more in depth downstream analysis to be conducted through assessment of the relative contribution of each clone through similarity percentages (SIMPER) analysis. The Bray-Curtis distances are visualised using Non-Metric Multidimensional Scaling (NMDS). At day 14 in the Bursa, Spleen and Jejunum, there is good visible separation between the Complex and Germ Free groups (Fig 5.4A). However, PERMANOVA tests are significant only in the Spleen, and the low R^2 values in each of these tissues suggests there remains much variation in the TCR gamma repertoire not explained by microbial status alone. At Day 28 however, there are no clear visible differences in clustering in any of these tissues and all of the P values are high and non-significant. Additionally, there are no visible differences at Day 14 or 28 in the Caecum and Colon

plots. At Day 55, there are differences in the caecum between the 2 groups, but not in any other tissues. These differences are significant in the Caecum for both the overall PERMANOVA testing and the paired test for Complex vs Germ Free, though with a low R^2 value of 0.22. The inclusion of the Conventional and Tetra groups in these Day 55 Bray-Curtis plots reveals that both of these groups are more heterogeneous than Complex and Germ Free in terms of dispersion in the all of the tissues. Due to the increased variance and therefore greater dispersion of the Conventional and Tetra groups, it is more difficult to judge visual differences and the PERMANOVA test is no longer valid in the Bursa and Jejunum. The Conventional and Tetra groups appear to cluster together and both are significant in the Spleen and Caecum when compared with both Complex and Germ Free. However, as both the Conventional and Tetra groups were present in the animal facility at different times to the Complex and Germ Free birds, which were reared in parallel, and clustered together on these plots despite their overwhelmingly different microbiota compositions, this difference in timing was thought to be the reason for these differences rather than any microbial effect.

JSD reveals similar patterns to those observed using Bray-Curtis, with visual separation at day 14 in Bursa, Spleen and Jejunum (Fig 5.4B). Again however, these differences are statistically significant only in the case of Spleen. At day 28 these differences disappear and the groups become interspersed, indicating the lack of any compositional repertoire differences at this age. At Day 55 however, there are again visual differences between the Complex and Germ Free groups in the Caecum, however due to a significant betadisper test and therefore heterogeneous variance, PERMANOVA testing is not valid. When considering the other tissues at day 55, the plots are once again similar to those obtained using Bray-Curtis, with the increase in variation between samples in the Conventional and Tetra groups making it more difficult to make comparisons at these time points, invalidating PERMANOVA testing in the Bursa, Jejunum and Caecum. However, upon visual inspection, the Conventional and Tetra groups do not seem to be clustering separately from the Complex and Germ Free groups to the same degree as in the Bray-Curtis plots, though the same general pattern of separation is still present.

As a result of recent arguments that Illumina NGS data should be treated as compositional[25], Aitchison's distance was also calculated and plots generated using Redundancy Analysis (RDA) (Fig 5.4C). Aitchison distance is simply Euclidean distance calculated from centre log ratio (CLR) transformed count data. The Aitchison's distance plots show relatively similar patterns to the Bray-Curtis and JSD plots with visual differences in the Bursa, Spleen and Jejunum at day 14. However, these differences are not significant in any of these tissues. Despite the differences in distribution in the Aitchison's distance plots, which seem to show a tendency for individual points to cluster distal from the main group, betadisper tests maintain that there are no significant differences in distribution and thus PERMANOVA remains a valid test in these cases. This pattern of one or 2 points breaking away from the main cluster continues in the Day 28 plots, but similar to the Bray-Curtis and JSD plots there is a lack of obvious visual differences at this age and all of the P values are high and therefore non-significant. Interestingly, the Day 55 plots do not show the variation in the Conventional and Tetra groups that was present in the Bray-Curtis and JSD plots and tend to cluster more closely than the Complex or Germ Free birds on these plots. Due to this difference in variation, PERMANOVA is not valid in any of the day 55 tissues. Despite this, there are no obvious differences and overall each of the groups seems to cluster relatively closely in these plots, again with the exception of a few breakaway points.


A
Day 14
Day 28
Day 55
Bursa
Spleen
Jejunum
Caecum
Colon
Bray NMDS Bursa Day 14
p = 0.076 ns, R2 = 0.283
Stress = 0.085
Bray NMDS Bursa Day 28
p = 0.195 ns, R2 = 0.116
Stress = 0.105
Brav NMDS Bursa Day 55
Stress = 0.145
Bray NMDS Spleen Day 14
p = 0.027 *, R2 = 0.172
Stress = 0.085
Bray NMDS Spleen Day 28
p = 0.505 ns, R2 = 0.11
Stress = 0.043
Bray NMDS Spleen Day 55
p = 0.008 **, R2 = 0.204
Stress = 0.1
Bray NMDS Jejunum Day 14
p = 0.105 ns, R2 = 0.132
Stress = 0.106
Bray NMDS Jejunum Day 28
p = 0.767 ns, R2 = 0.107
Stress = 0.057
Bray NMDS Jejunum Day 55
Stress = 0.169
Bray NMDS Caecum Day 14
p = 0.323 ns, R2 = 0.133
Stress = 0.096
Bray NMDS Caecum Day 28
p = 0.767 ns, R2 = 0.108
Stress = 0.07
Bray NMDS Caecum Day 55
p = 0.008 **, R2 = 0.217
Stress = 0.145
Bray NMDS Colon Day 14
p = 0.349 ns, R2 = 0.113
Stress = 0.09
Bray NMDS Colon Day 28
p = 0.895 ns, R2 = 0.119
Stress = 0.046
Bray NMDS Colon Day 55
p = 0.008 **, R2 = 0.214
Stress = 0.143
NMDS1
NMDS2
Strain
Complex
GermFree
Conventional
Tetra
NA
Read Counts


B
Day 14
Day 28
Day 55
Bursa
JSD NMDS Bursa Day 14
p = 0.07 ns, R2 = 0.325
Stress = 0.07
JSD NMDS Bursa Day 28
p = 0.144 ns, R2 = 0.119
Stress = 0.115
JSD NMDS Bursa Day 55
Stress = 0.143
Spleen
JSD NMDS Spleen Day 14
p = 0.034 *, R2 = 0.18
Stress = 0.086
JSD NMDS Spleen Day 28
p = 0.54 ns, R2 = 0.109
Stress = 0.042
JSD NMDS Spleen Day 55
p = 0.008 **, R2 = 0.203
Stress = 0.109
Jejunum
JSD NMDS Jejunum Day 14
p = 0.098 ns, R2 = 0.14
Stress = 0.081
JSD NMDS Jejunum Day 28
p = 0.728 ns, R2 = 0.107
Stress = 0.077
JSD NMDS Jejunum Day 55
Stress = 0.177
Caecum
JSD NMDS Caecum Day 14
p = 0.255 ns, R2 = 0.136
Stress = 0.057
JSD NMDS Caecum Day 28
p = 0.767 ns, R2 = 0.107
Stress = 0.08
JSD NMDS Caecum Day 55
Stress = 0.142
Colon
JSD NMDS Colon Day 14
p = 0.472 ns, R2 = 0.112
Stress = 0.068
JSD NMDS Colon Day 28
p = 0.74 ns, R2 = 0.12
Stress = 0.04
JSD NMDS Colon Day 55
p = 0.008 **, R2 = 0.218
Stress = 0.142
NMDS1
NMDS2
Strain
Complex
Conventional
GermFree
Tetra
Read Counts

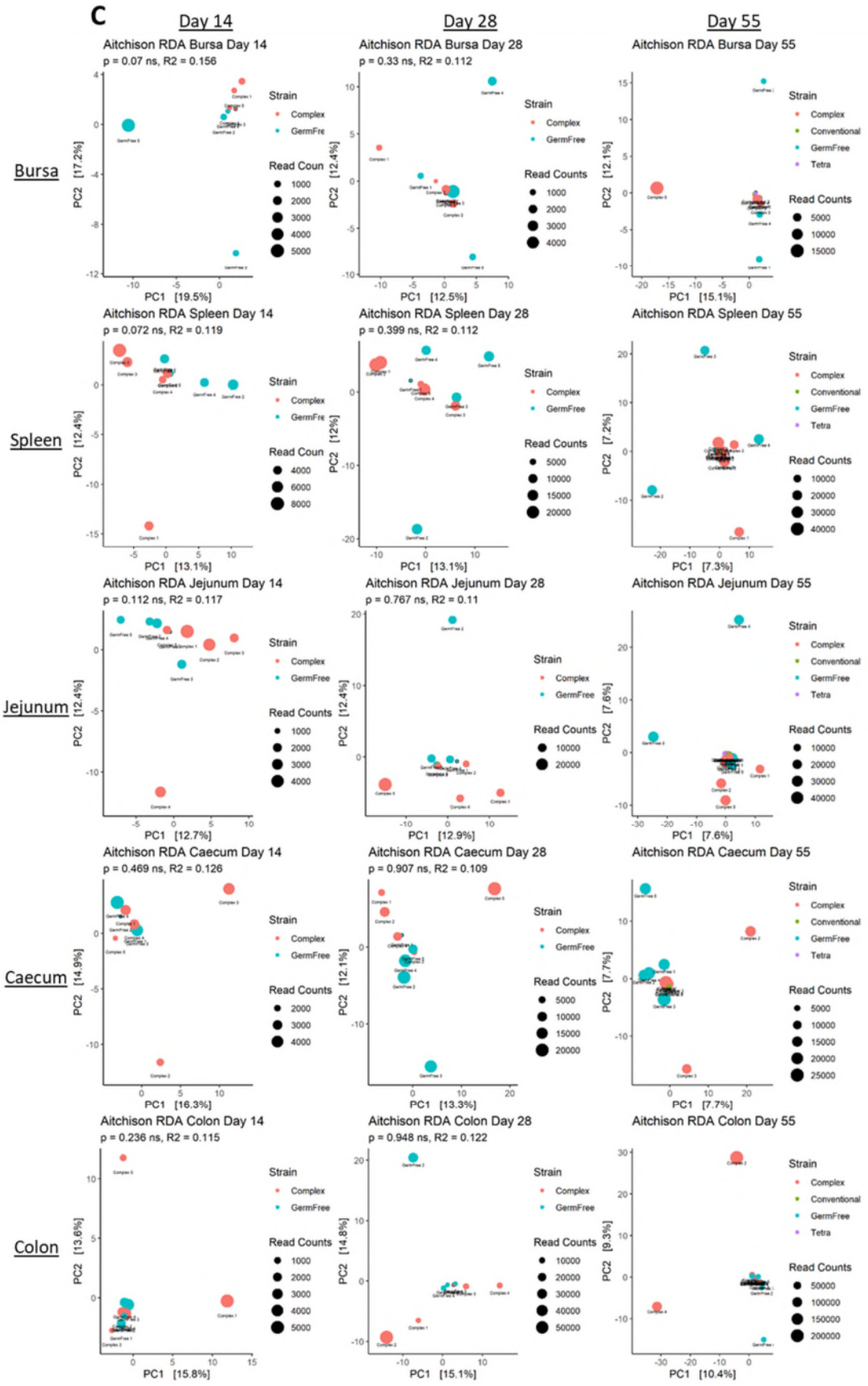

C
Day 14
Day 28
Day 55
Bursa
Spleen
Jejunum
Caecum
Colon
Aitchison RDA Bursa Day 14
p = 0.07 ns, R2 = 0.156
Aitchison RDA Bursa Day 28
p = 0.33 ns, R2 = 0.112
Aitchison RDA Bursa Day 55
Aitchison RDA Spleen Day 14
p = 0.072 ns, R2 = 0.119
Aitchison RDA Spleen Day 28
p = 0.399 ns, R2 = 0.112
Aitchison RDA Spleen Day 55
Aitchison RDA Jejunum Day 14
p = 0.112 ns, R2 = 0.117
Aitchison RDA Jejunum Day 28
p = 0.767 ns, R2 = 0.11
Aitchison RDA Jejunum Day 55
Aitchison RDA Caecum Day 14
p = 0.469 ns, R2 = 0.126
Aitchison RDA Caecum Day 28
p = 0.907 ns, R2 = 0.109
Aitchison RDA Caecum Day 55
Aitchison RDA Colon Day 14
p = 0.236 ns, R2 = 0.115
Aitchison RDA Colon Day 28
p = 0.948 ns, R2 = 0.122
Aitchison RDA Colon Day 55
Strain
Complex
Conventional
GermFree
Tetra
Read Counts
PC1 [19.5%]
PC2 [17.2%]
PC1 [12.5%]
PC2 [12.4%]
PC1 [15.1%]
PC2 [12.1%]
PC1 [13.1%]
PC2 [12.4%]
PC1 [13.1%]
PC2 [12%]
PC1 [7.3%]
PC2 [7.2%]
PC1 [12.7%]
PC2 [12.4%]
PC1 [12.9%]
PC2 [12.4%]
PC1 [7.6%]
PC2 [7.6%]
PC1 [16.3%]
PC2 [14.9%]
PC1 [13.3%]
PC2 [12.1%]
PC1 [7.7%]
PC2 [7.7%]
PC1 [15.8%]
PC2 [13.6%]
PC1 [15.1%]
PC2 [14.8%]
PC1 [10.4%]
PC2 [9.3%]

Figure 5.4 – β-Diversity Plots of Bray-Curtis, JSD and Aitchison's Distance Metrics Displaying Differences between Groups Split by Tissue and Age
Ordination plots show a visual summary of dissimilarity matrices, which were generated for the samples using **A)** Bray-Curtis, **B)** JSD and **C)** Aitchison's distance indices using unique clones, defined by their combination of TRGV, TRGJ and CDR3NT, as species. Matrices were then subsetted for each tissue and age combination and ordination plots generated for each β-diversity metric. NMDS plots were generated for Bray-Curtis and JSD indices using 20 random start iterations and RDA plots were generated for Aitchison distance. Each point represents an individual bird, strain is indicated by colour and read count is indicated by point size. P values and R2 values are indicated beneath the subtitle, as well as next to the legend for testing pairs of groups at day 55 post-hatch. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. P values were FDR corrected. Stress values are indicated below the title for NMDS and percentage of variation explained in the axes for RDA

5.3.4 TRGV Gene Segment Usage is mostly Similar in the Germ Free Birds with some Small Differences

As the chicken TCR gamma repertoire was sequenced and subsequently characterised using universal UPA repertoire PCRs that allow us to sample all of the TRGV gene rearrangements in a single reaction, we were able to accurately compare the relative proportions of each gene segment and thus their contribution to the overall TCR gamma repertoire. As a complex flora was used to re-conventionalise the microbiota of the Complex birds which were born germ free, we sought to satisfy ourselves that these birds displayed a TRGV gene usage profile similar to that of conventionally reared birds, especially as no conventional samples were collected at days 14 and 28. In order to confirm this, we first compared the TRGV gene usage profiles of Conventional and Complex birds at Day 55 in each tissue (Fig 5.5A-E). The Complex group look mostly similar in every tissue examined, however there were some slight differences in TRGV2.20 proportion, with it being slightly higher than that of the Conventional group in the spleen and to a lesser extent caecum and colon. There were also increases in TRGV4.2 in the Conventional strain in the Bursa. The absence of any large scale differences between the Complex and Conventional birds allows us to confidently make comparisons based on microbial status using the Complex birds.

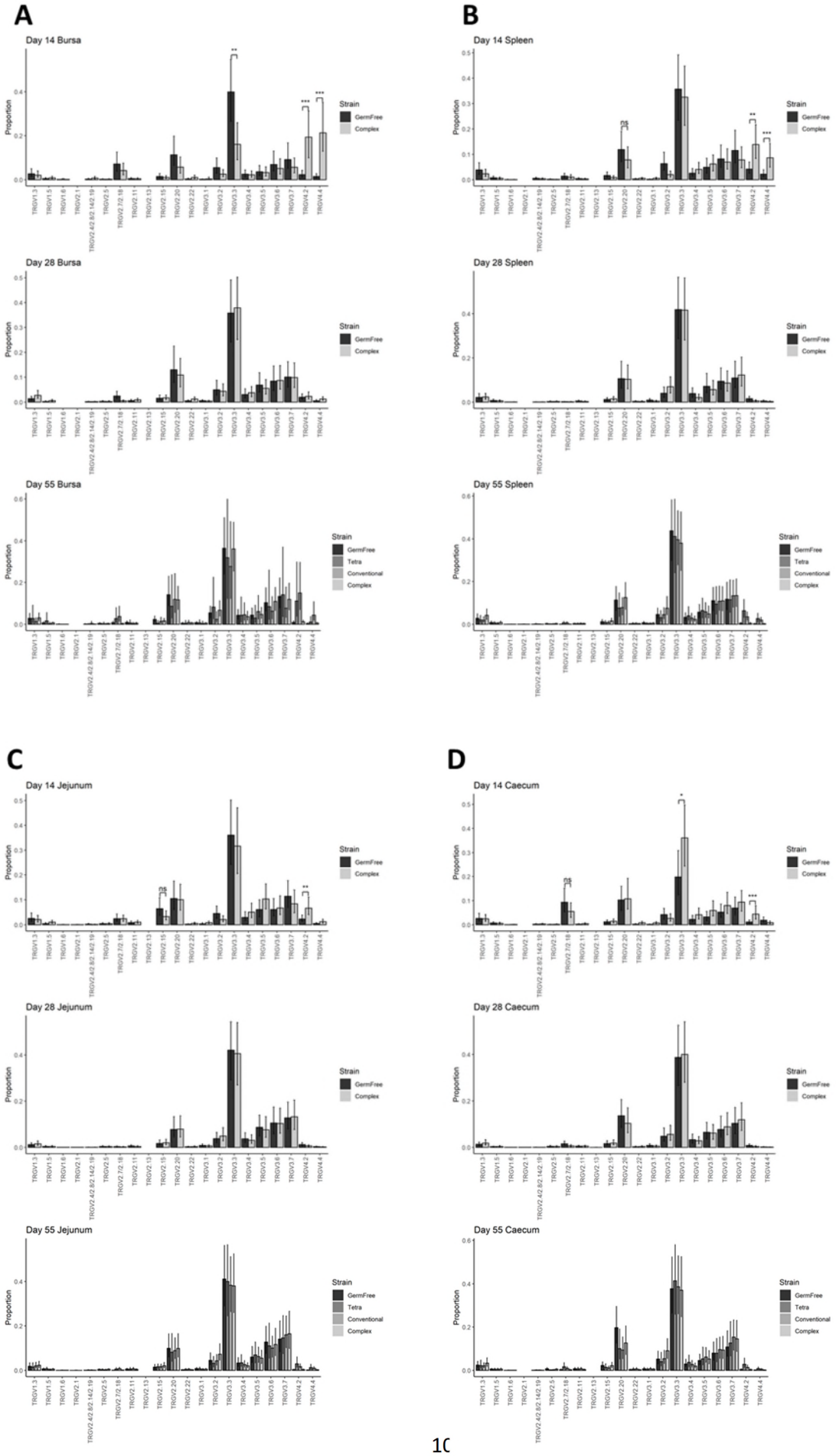
Though there were no large differences in these groups, the Conventional and Tetra groups did show some slight differences from the Complex birds. Both the Conventional and Tetra birds showed increases in the proportion of TRGV4.2 and TRGV4.4 particularly in the Bursa. As mentioned

for the β-diversity plots (Fig 5.4), due to the fact that both the Conventional and Tetra groups were present in the animal facility at different times to the Complex and Germ Free birds, this difference in timing was thought to be the reason for these differences rather than their microbiota differences.

Examination of the Complex group over time reveals some interesting differences about how the TCR gamma repertoire appears to change with age. The Bursa undergoes a fairly dramatic change from day 14 to day 28. At day 14, the Bursa is the only tissue where we observed TRGV3.3 not being the most dominant rearrangement and instead both TRGV4.2 and TRGV4.4 occupy a higher proportion of the repertoire (Fig 5.5A). However, at day 28 the proportions of TRGV4.2 and TRGV4.4 are dramatically reduced, with TRGV3.3 increasing in proportion to 39% and remaining around the same at day 55. In comparison, the other tissues remain more consistent over time. In Spleen, from day 14 to day 28 there is a substantial decrease in TRGV4.2 and TRGV4.4, which are reduced down to almost negligible levels at day 28 and remain at this low level at day 55 (Fig 5.5B). This reduction is associated with a slight increase in the usage of the dominant TRGV3.3, which increases from 35% to 45%. The Jejunum remains mostly consistent across all of the time points, but similarly to Spleen and Bursa, shows a decrease in TRGV4.2 from day 14 to day 28, though in this case no accompanying drop in TRGV4.4, which is already at very low proportion on day 14 (Fig 5.5C). In the large intestine, there are some more notable TRGV gene expression differences worth highlighting. Looking first to the caecum, there is a substantial decrease in the proportion of TRGV2.7/2.18 from 6% at day 14 down to a negligible level at days 28 and 55 (Fig 5.5D). A similar, more pronounced difference is present in the colon where TRGV2.7/2.18 decreases from high proportions of 20% at Day 14 down to almost negligible amounts at Day 28 (Fig 5.5E).

By comparing the Complex and Germ Free strain TRGV gene proportions, we are able to gain an insight into the role of the microbiota in gamma repertoire development in terms of TRGV expression. While some differences do exist, particularly at the earlier time points, they are for the

most part relatively small. The one exception to this is the Bursa at day 14, in which some fairly obvious differences do exist. At this time point, Complex birds have significantly less TRGV3.3 and more TRGV4.2 and TRGV4.4. Spleen on the other hand, has far more subtle but still observable differences, with a higher proportion of TRGV2.20 (non-significant) and significantly lower TRGV4.2 and TRGV4.4 at Day 14 in Germ Free compared to Complex. Looking to the jejunum, the TRGV gene profiles are largely very similar with a slight increase in TRGV2.15 (non-significant) and accompanying significant decrease in TRGV4.2 in Germ Free relative to Complex. Caecum has some substantial and noteworthy differences, particularly at Day 14 with TRGV2.7/2.18 reduced by almost half from 15% in Germ Free to 8% in Complex (non-significant) and TRGV3.3 significantly increased from 23% in Germ Free to 35% in Complex. Additionally, as in Bursa, Spleen and Jejunum at day 14 there is a significant decrease in TRGV4.2 in the Germ Free birds relatively to Complex. The colon however, is the exception to this and there are only very slight variations in some of the proportions at Day 14.


A
Day 14 Bursa
Day 28 Bursa
Day 55 Bursa
B
Day 14 Spleen
Day 28 Spleen
Day 55 Spleen
C
Day 14 Jejunum
Day 28 Jejunum
Day 55 Jejunum
D
Day 14 Caecum
Day 28 Caecum
Day 55 Caecum
Proportion
Strain
GermFree
Tetra
Conventional
Complex

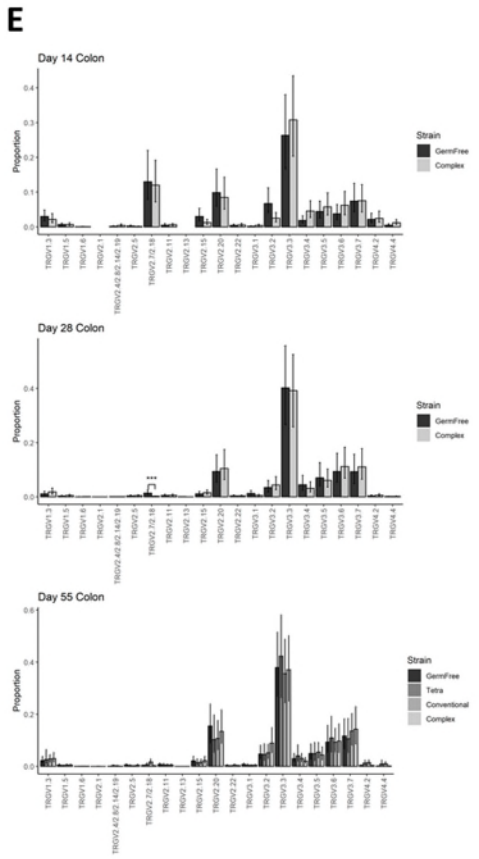


Figure 5.5 - Proportion of TCR Gamma Rearrangements Originating from each TRGV Gene Segment
Plots show average proportional TRGV gene usage and are displayed in A) Bursa, B) Spleen, C) Jejunum, D) Caecum and E) Colon at days 14, 28 and 55 post-hatch. TCR gamma rearrangements were blasted against a repository of TRGV genes, enabling the initial gene segment that produced the rearrangement to be identified. The proportion of rearrangements corresponding to each TRGV gene were then calculated. Error bars are 95% confidence intervals derived from 500 bootstrap replicates of a linear mixed model using bird number as a random variable and tissue, group, day and TRGV gene as fixed variables. P values were generated using the LmerTest package using the Satterthwaite approximation of degrees of freedom (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

5.3.5 The Chicken TCR Gamma Repertoire contains highly Public TRGV Gene Segments

Understanding the structure of public (shared between individuals) and private (unique to an individual) rearrangements, gives an insight into the nature and function of the γδ T cells expressing them. Publicity was defined here as CDR3 sequences at the nucleotide level which were shared between individuals at any proportion in any tissue. Percentage publicity was calculated for each group at different levels and plotted for each TRGV gene segment, revealing interesting changes over time (Fig 5.6A-C). In the Complex strain at day 14, many of the TRGV2s and all of the TRGV4s were highly public, namely TRGV2.4/2.8/2.14/2.19, TRGV2.7/2.18, TRGV2.15, TRGV2.20, TRGV2.22, TRGV4.2 and TRGV4.4 (Fig 5.6A). However, at day 28 many of these genes are reduced in publicity, most strikingly TRGV2.4/2.8/2.14/2.19 and TRGV4.4, which both go from being ~75% to ~40% overall publicity (Fig 5.6B). The "100% of Birds" category was most affected and this difference was significant in both of these genes. Additionally, many other TRGV genes are reduced in publicity to a lesser degree, including TRGV2.7/2.18, TRGV2.15, TRGV2.20 and TRGV4.2. Similarly, from day 28 to 55, there is another decrease in publicity in a number of V genes with reductions in overall publicity in TRGV4.4 from ~50% to ~25%, however this does not reach significance for the "100% of Birds" category for TRGV4.4 or any other V gene (Fig 5.6C).

The publicity profile of the Germ Free group is relatively similar to the Complex, however, some differences can be observed. At day 14, the Germ Frees have higher levels of publicity in TRGV3.2 and lower publicity in TRGV2.22. Both of these differences are significant when comparing the "100% of Birds" category. Additionally, the Germ Free birds also have higher levels of public clones expressing rearrangements with the TRGV genes TRGV4.4 and TRGV3.3. At day 28, there are also differences in TRGV2.22 and TRGV4.4, which are both higher in the Complex group relative to the Germ Free birds. These differences however are not significant in both cases. At day 55, the Germ Free group features a publicity profile more similar to that of Complex, but still with small

differences in some V genes, including higher levels of publicity in TRGV2.7/2.18 and lower levels in TRGV1.5, TRGV2.11, TRGV2.4/2.8/2.14/2.19, TRGV2.15, TRGV2.22, TRGV3.2 and TRGV3.3 (all non-significant).

Interestingly, there are fairly widespread differences between the Complex and Conventional publicity profiles at Day 55. Conventional has a far less public repertoire, featuring decreased publicity in TRGV2.5, TRGV2.20, TRGV2.11, TRGV2.15, TRGV2.7/2.18, TRGV3.1, TRGV3.2, TRGV3.3, TRGV3.4, TRGV3.5, TRGV3.6, TRGV3.7. Interestingly however, this is accompanied by increases in TRGV4.2 and TRGV4.4 in Conventional relative to complex. The Tetra group contains fewer individuals so publicity is harder to interpret relative to the others, however, it does show a profile more like that of Complex and Germ Free rather than Conventional.

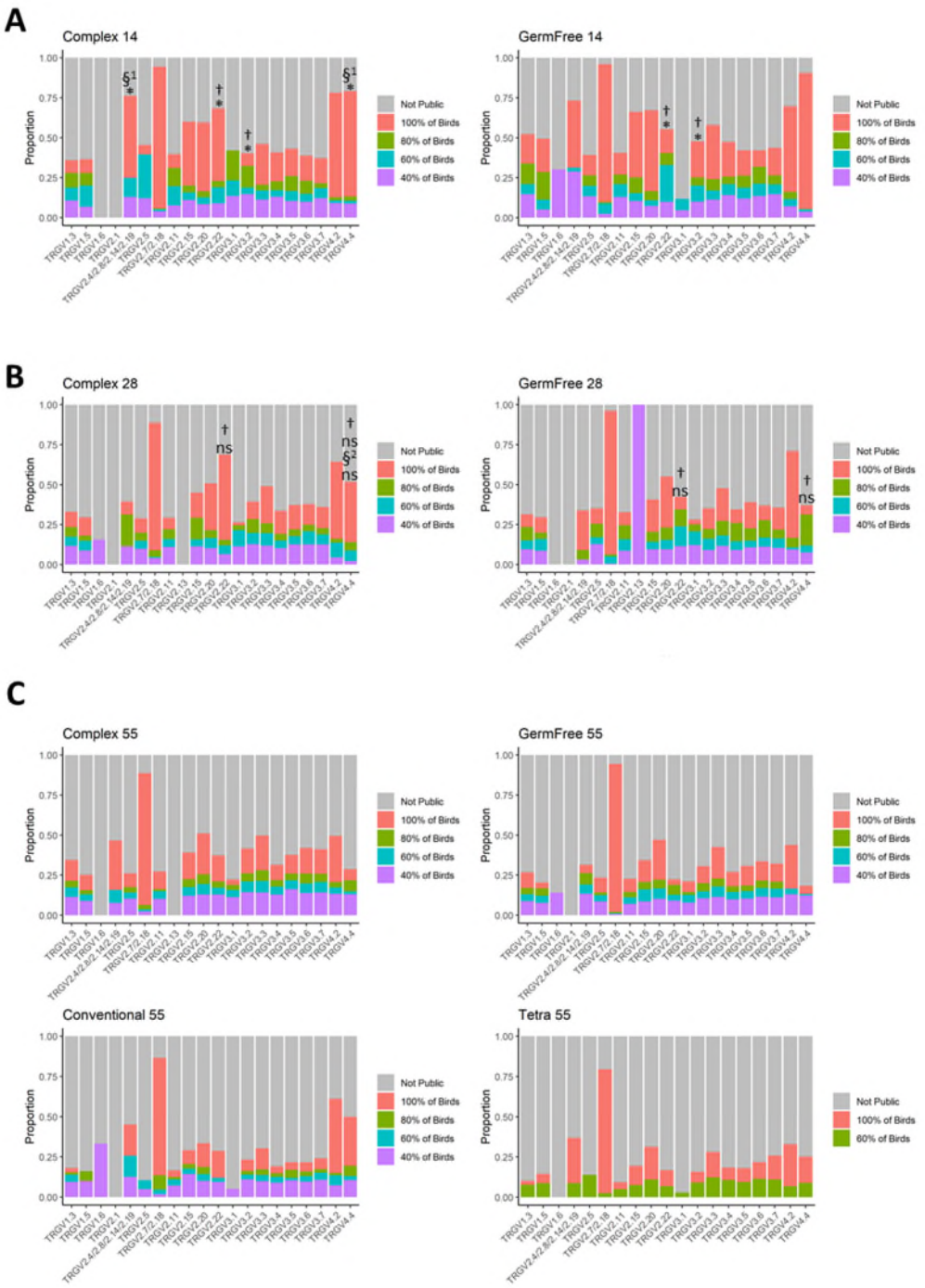


Figure 5.6 - Proportional Publicity for each TRGV Gene Segment at each Time Point
Plots show the proportions of public clones in a given TRGV gene segment calculated at different levels at days **A)** 14, **B)** 28 and **C)** 55 post-hatch. Publicity is defined by the presence of a clone in more than one individual within a group in any tissue at any proportion. The number of individuals a clone is found in defines its percentage publicity, which is represented by colour in the stacked bars. P values are calculated using an unpaired Wilcoxon rank test performed on the 100% of birds category only (*** p value < 0.001, ** p value < 0.01, * p value < 0.05). † represents comparisons between groups with P value indications above both bars being compared. §1 is comparisons between age 14-28, §2 is comparisons between age 28-55 with P value indications only on the younger of the ages being compared. P values were FDR corrected.

5.3.6 Diversity Plots reveal massive Increases in Diversity over Time but no Differences in the Germ Free Birds when compared to Wild Types

Diversity indices were calculated from rarefied data and reveal the tissue level clonal structure of the global chicken TCR gamma repertoire. (Fig 5.7A-C) Before calculation of diversity indices, the repertoire of every bird for each tissue and time point was rarefied down to the coverage level of the lowest, and though this does mean discarding more data it allows comparisons to be made across time as well as tissues. A Linear Model (LM) was calculated for the diversity data using a cube root transform allowing generation of 95% confidence intervals from 500 bootstrap replicates and P values.

The diversity plots show an increase in diversity as the Complex and Germ Free birds age, which matches our observations from the publicity plots (Fig 5.6). Unsurprisingly, this age-dependent increase in diversity occurred in every tissue, at every time point and all diversity indices. These differences were also significant for every age related comparison with the sole exceptions of Day 28 Germ Free Bursa vs Day 55 Germ Free Bursa and Day 28 Germ Free Caecum vs Day 55 Germ Free Caecum.

Comparing diversity across tissues in the Complex strain reveals that as we would expect, Bursa is the least diverse at every time point. However, unexpectedly at Day 14, Caecum is the more diverse than Spleen. This however, is not the case at Days 28 and 55 where Spleen becomes the most diverse tissue, with Jejunum, Caecum and Colon at roughly the same level.

There are some interesting differences between the Complex and Germ Free groups. At Days 14 and 28, the Germ Free Spleen TCR gamma repertoires are around twice as diverse as Complex Spleen by every diversity index, however, these differences do not reach the 0.05 significance threshold. Despite this, the p values for Day 28 Spleen Germ Free vs Complex are very low and close to 0.05, with respective values of 0.052, 0.055 and 0.067 for Hill numbers 0, 1 and 2

respectively and therefore this difference is still worth considering. Perhaps a larger group size or greater sequencing depth may have resulted in significance. Additionally, the Day 14 Complex Caeca were around 3 times as diverse as that of the Germ Free group, however again due to high inter-bird variation, this difference did not reach statistical significance.

These differences between the Complex and Germ Free groups seem to disappear at day 55. Perplexingly however, both the Conventional and Tetra groups appear less diverse than both Complex and Germ Free by every metric. As outlined previously though, these birds were reared at a different time to the Complex and Germ Free groups. As diversity appeared to correlate with age it seems that the TCR gamma repertoire of these birds has developed at a reduced rate, however the reason for this is unclear. Again, it seems highly unlikely that these differences are down to gut flora composition and are most likely down to some difference in conditions or experimental artefact.

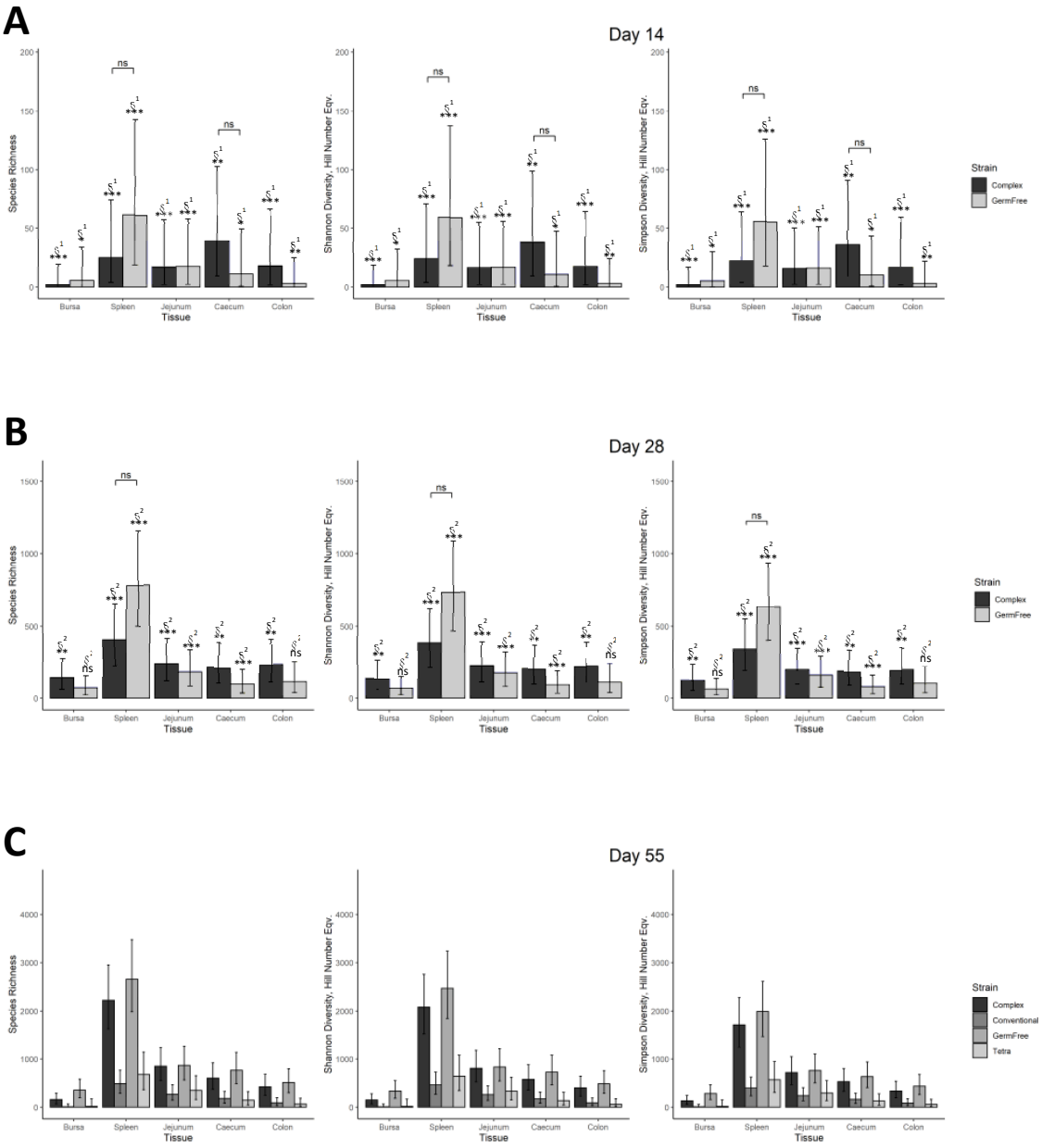


Figure 5.7 – TCR Gamma Repertoire Hill Number Diversity Indices Calculated for each Tissue at each Time Point

Plots show diversity indices calculated from iNEXT rarefied count data separated by each time point **A)** Day 14, **B)** Day 28 and **C)** Day 55 post-hatch. Diversity indices are displayed as their Hill number variants. A linear model was generated for each diversity index using cube root transformed data. Cube root was determined to be the optimal transform for normality of residuals using Box Cox plots. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values (*** p value < 0.001, ** p value < 0.01, * p value < 0.05). §1 is comparisons between age 14-28, §2 is comparisons between age 28-55 with P value indications displayed only on the younger of the ages being compared.

5.3.7 TCR gamma repertoire sequencing reveals differences between the Germ Free and Complex Birds at the level of individual clones

Applying SIMPER analysis to the Bray-Curtis dissimilarity indices from day 14, Bursa, Jejunum and Spleen, all of which showed significant group differences when tested by PERMANOVA (Fig 5.4A), allows the relative contribution of each clone in the dataset to be assessed. The largest contributing clones can then be isolated and assessed individually. Picking out clones in this way gives a higher resolution picture of the differences between the Complex and Germ Free groups (Fig 8). Though there are many clones which are contributors to Bray-Curtis differences between Complex and Germ free birds, one issue with analysing individual clones is that proportions tend to vary widely between individuals, resulting in very high standard deviation. Unfortunately, this makes analysis on many other potentially interesting clones untenable and limits it to those with more consistent proportions.

Clone 15 is the top contributing clone to differences in the Bursa and the second from the top in Spleen, it uses a TRGV4.4 TRGV gene, with a germline CDR3 and is present at extremely high levels in the Complex repertoire (Fig 5.8A). Interestingly, Clone 15 is almost solely responsible for the differences observed in TRGV4.4 proportion between the Complex and Germ Free groups at day 14 in both the Bursa and Spleen. At day 14, Clone 15 comprises almost ~20% of the repertoire in Complex Bursa and ~6% in Complex Spleen but is present at negligible levels in both of these tissues in Germ Free birds. Both of these differences are significant when tested using an unpaired Wilcoxon Signed Rank test. At day 28, the proportions of clone 15 fall sharply in Bursa and Spleen in both groups, and it is present in the Complex birds at just ~2.5% and ~0.1% respectively, and at almost undetectably low levels in both of these tissues in Germ Free birds. This trend continues with clone 15 being present at similarly low levels in both of these groups, as well as conventional and tetra, at Day 55. Levels of clone 15 at Day 55 are still very low in Conventional birds but notably higher than in the other groups.

Clone 7 expresses a TRGV3.3 and is another considerable contributor to Bray-Curtis differences between Complex and Germ Free birds at Day 14 in Spleen, Jejunum and Bursa (Fig 5.8B). Clone 7 is present at far higher levels in the Germ Free birds and at 14 day, it makes up ~2.5%, ~1.5% and ~1.8% of the respective TCR gamma repertoire in Bursa, Spleen and Jejunum, compared with ~0.3%, ~0.7% and ~0.2% in Complex birds. However, these differences are significant only in the Jejunum, although the P value for Spleen is 0.056 and therefore very close to the 0.05 significance threshold. In the Day 28 birds, the proportions of Germ Free Clone 7 tend to decrease to roughly the levels present in 14 Day old Complex birds. They remain at this proportion in Day 55 birds, consistent with what is seen in Conventional and Tetra at the same time point.

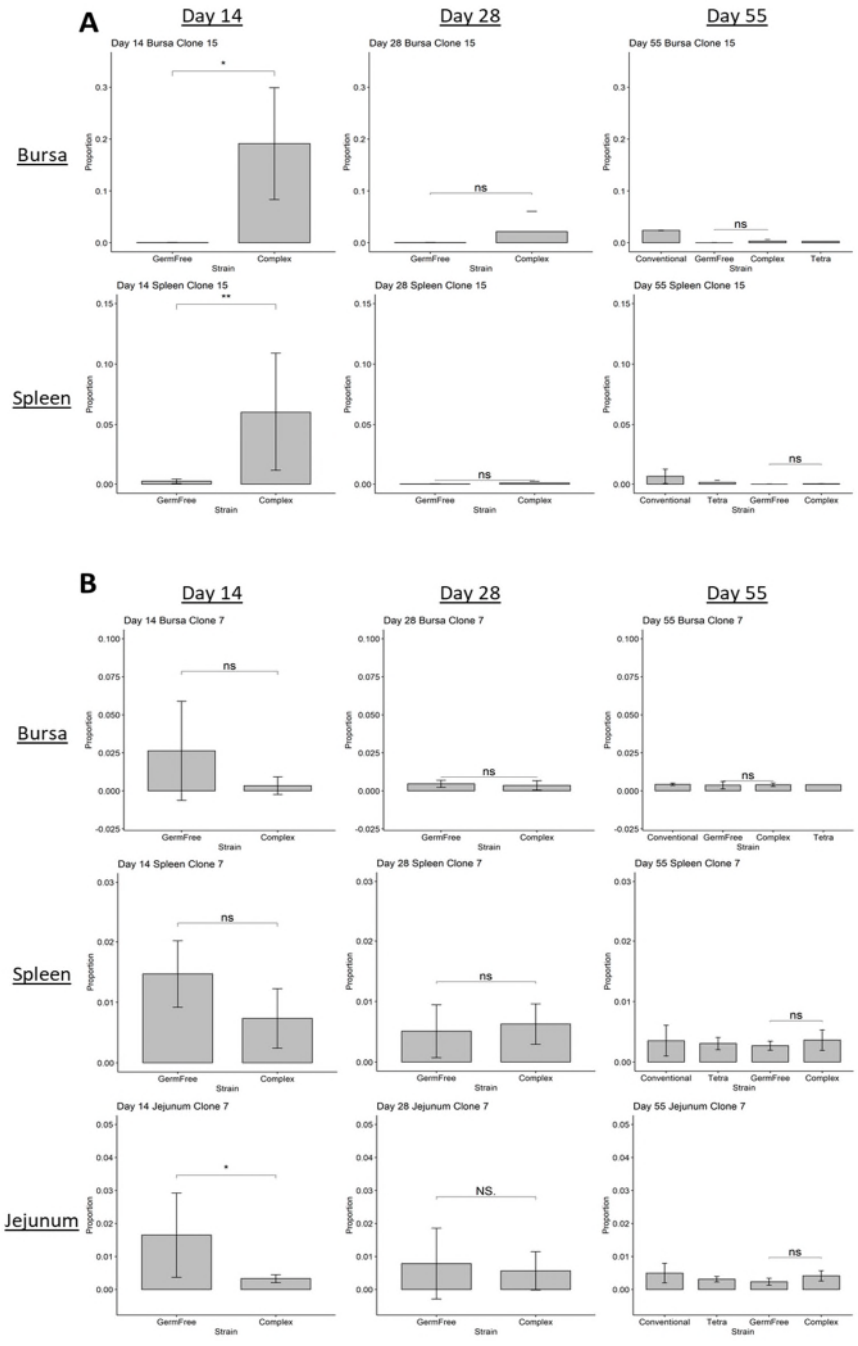


Figure 5.8 - Clone Plots of the Biggest Contributors to Bray-Curtis Differences between Bursa, Spleen and Jejunum as Determined by SIMPER Analysis

Plots show the relative contribution of individual clones to the overall repertoire of each group of birds split by tissue and day post-hatch. The clones highlighted by SIMPER analysis were **A)** Clone 15 and **B)** Clone 7 (numbers are arbitrary). SIMPER was used to calculate the contribution of each clone to the significant differences observed between the Complex and Germ Free Groups in the day 14 post-hatch Bursa, Spleen and Jejunum ordination plots. Unfortunately, the majority of clones highlighted by this analysis were extremely variable among individuals, as tends to be the case when examining repertoire at this fine scale, and thus were not plotted. Only clones which were present at consistent proportions and appeared in the top 20 for multiple tissues were selected for plotting. Error bars are standard deviation and proportions are of whole repertoire for each tissue and age. P values were calculated using an unpaired Wilcoxon Signed Rank test (*** p value < 0.001, ** p value < 0.01, * p value < 0.05).

5.3.8 Chicken Sample Collection and Processing for 16S rRNA Sequencing

Chicken faecal samples were collected from Wild Type, TCRγ +/- and TCRγ -/- birds at day 3, 7, 21, 35 and 49 days post hatch. The chickens were tagged and samples collected such that each faecal sample could be traced back to the bird of origin. Additionally, upon sacrifice at day 65, caecal and ileal contents were collected. The DNA of these samples was extracted and triplicate 16S PCR reactions performed and visualised on agarose gels (Fig 5.9). The desired products are visible at 500bp. These products were blended, library prepped, quantified and sequenced on an Illumina MiSeq.

The Illumina MiSeq run produced a total of 29,574,824 reads with a Q30 of 86.9% and %PF of 92.68%. Of those reads, 10,088,819 were attributable to the 16S amplicons for used in this experiment. The Illumina reads were demultiplexed and separated according to barcode using reptools and processed using DADA2. Read quality profiles were generated and inspected to inform the DADA2 read trimming stage (Fig 5.10). The reverse reads were worse quality than the forward, though this is a common facet of Illumina sequencing. As a result, it was decided that the forward reads be trimmed to 260bp and the reverse reads to 210bp as a trade-off between trimming the lowest quality base calls and retaining enough length that the reads overlap and allow downstream read merging.

100bp NEB ladder
No Template Control 1 16S #93
No Template Control 2 16S #94
No Template Control 3 16S #95
No Template Control 4 16S #96
No Template Control 5 16S #1
No Template Control 6 16S #2
No Template Control 7 16S #3
No Template Control 8 16S #4
100bp NEB ladder
100bp NEB ladder
TCRγ+/- 5 M D49 Faeces 16S #57
TCRγ+/- 6 F D49 Faeces 16S #58
TCRγ+/- 7 F D49 Faeces 16S #59
TCRγ+/- 8 F D49 Faeces 16S #60
TCRγ+/- 9 F D49 Faeces 16S #61
TCRγ+/- 10 F D49 Faeces 16S #62
Wild Type 1 M D65 Ileal Contents 16S #63
Wild Type 2 M D65 Ileal Contents 16S #64
Wild Type 3 M D65 Ileal Contents 16S #65
Wild Type 4 M D65 Ileal Contents 16S #66
Wild Type 5 M D65 Ileal Contents 16S #67
TCRγ-/- 1 M D65 Ileal Contents 16S #68
TCRγ-/- 2 M D65 Ileal Contents 16S #69
TCRγ-/- 3 M D65 Ileal Contents 16S #70
TCRγ-/- 4 M D65 Ileal Contents 16S #71
TCRγ-/- 5 M D65 Ileal Contents 16S #72
TCRγ+/- 1 M D65 Ileal Contents 16S #73
TCRγ+/- 2 M D65 Ileal Contents 16S #74
100bp NEB ladder
100bp NEB ladder
TCRγ-/- 9 F D35 Faeces 16S #21
TCRγ-/- 10 F D35 Faeces 16S #22
TCRγ+/- 1 M D35 Faeces 16S #23
TCRγ+/- 2 M D35 Faeces 16S #24
TCRγ+/- 3 M D35 Faeces 16S #25
TCRγ+/- 4 M D35 Faeces 16S #26
TCRγ+/- 5 M D35 Faeces 16S #27
TCRγ+/- 6 F D35 Faeces 16S #28
TCRγ+/- 7 F D35 Faeces 16S #29
TCRγ+/- 8 F D35 Faeces 16S #30
TCRγ+/- 9 F D35 Faeces 16S #31
TCRγ+/- 10 F D35 Faeces 16S #32
Wild Type 1 M D49 Faeces 16S #33
Wild Type 2 M D49 Faeces 16S #34
Wild Type 3 M D49 Faeces 16S #35
Wild Type 4 M D49 Faeces 16S #36
Wild Type 5 M D49 Faeces 16S #37
Wild Type 6 F D49 Faeces 16S #38
100bp NEB ladder
100bp NEB ladder
TCRγ-/- 3 M D21 Faeces 16S #81
TCRγ-/- 4 M D21 Faeces 16S #82
TCRγ-/- 5 M D21 Faeces 16S #83
TCRγ-/- 6 F D21 Faeces 16S #84
TCRγ-/- 7 F D21 Faeces 16S #85
TCRγ-/- 8 F D21 Faeces 16S #86
TCRγ-/- 9 F D21 Faeces 16S #87
TCRγ-/- 10 F D21 Faeces 16S #88
TCRγ+/- 1 M D21 Faeces 16S #89
TCRγ+/- 2 M D21 Faeces 16S #90
TCRγ+/- 3 M D21 Faeces 16S #91
TCRγ+/- 4 M D21 Faeces 16S #92
TCRγ+/- 5 M D21 Faeces 16S #93
TCRγ+/- 6 F D21 Faeces 16S #94
TCRγ+/- 7 F D21 Faeces 16S #95
TCRγ+/- 8 F D21 Faeces 16S #96
TCRγ+/- 9 F D21 Faeces 16S #1
TCRγ+/- 10 F D21 Faeces 16S #2
100bp NEB ladder
100bp NEB ladder
Wild Type 7 F D49 Faeces 16S #39
Wild Type 8 F D49 Faeces 16S #40
Wild Type 9 F D49 Faeces 16S #41
Wild Type 10 F D49 Faeces 16S #42
TCRγ-/- 1 M D49 Faeces 16S #43
TCRγ-/- 2 M D49 Faeces 16S #44
TCRγ-/- 3 M D49 Faeces 16S #45
TCRγ-/- 4 M D49 Faeces 16S #46
TCRγ-/- 5 M D49 Faeces 16S #47
TCRγ-/- 6 F D49 Faeces 16S #48
TCRγ-/- 7 F D49 Faeces 16S #49
TCRγ-/- 8 F D49 Faeces 16S #50
TCRγ-/- 9 F D49 Faeces 16S #51
TCRγ-/- 10 F D49 Faeces 16S #52
TCRγ+/- 1 M D49 Faeces 16S #53
TCRγ+/- 2 M D49 Faeces 16S #54
TCRγ+/- 3 M D49 Faeces 16S #55
TCRγ+/- 4 M D49 Faeces 16S #56
100bp NEB ladder
100bp NEB ladder
TCRγ+/- 3 M D65 Ileal Contents 16S #75
TCRγ+/- 4 M D65 Ileal Contents 16S #76
TCRγ+/- 5 M D65 Ileal Contents 16S #77
Wild Type 1 M D65 Caecal Contents 16S #78
Wild Type 2 M D65 Caecal Contents 16S #79
Wild Type 3 M D65 Caecal Contents 16S #80
Wild Type 4 M D65 Caecal Contents 16S #81
Wild Type 5 M D65 Caecal Contents 16S #82
TCRγ-/- 1 M D65 Caecal Contents 16S #83
TCRγ-/- 2 M D65 Caecal Contents 16S #84
TCRγ-/- 3 M D65 Caecal Contents 16S #85
TCRγ-/- 4 M D65 Caecal Contents 16S #86
TCRγ-/- 5 M D65 Caecal Contents 16S #87
TCRγ+/- 1 M D65 Caecal Contents 16S #88
TCRγ+/- 2 M D65 Caecal Contents 16S #89
TCRγ+/- 3 M D65 Caecal Contents 16S #90
TCRγ+/- 4 M D65 Caecal Contents 16S #91
TCRγ+/- 5 M D65 Caecal Contents 16S #92
100bp NEB ladder
100bp NEB ladder
Wild Type 1 M D35 Faeces 16S #3
Wild Type 2 M D35 Faeces 16S #4
Wild Type 3 M D35 Faeces 16S #5
Wild Type 4 M D35 Faeces 16S #6
Wild Type 5 M D35 Faeces 16S #7
Wild Type 6 F D35 Faeces 16S #8
Wild Type 7 F D35 Faeces 16S #9
Wild Type 8 F D35 Faeces 16S #10
Wild Type 9 F D35 Faeces 16S #11
Wild Type 10 F D35 Faeces 16S #12
TCRγ-/- 1 M D35 Faeces 16S #13
TCRγ-/- 2 M D35 Faeces 16S #14
TCRγ-/- 3 M D35 Faeces 16S #15
TCRγ-/- 4 M D35 Faeces 16S #16
TCRγ-/- 5 M D35 Faeces 16S #17
TCRγ-/- 6 F D35 Faeces 16S #18
TCRγ-/- 7 F D35 Faeces 16S #19
TCRγ-/- 8 F D35 Faeces 16S #20
100bp NEB ladder
100bp NEB ladder
TCRγ+/- 5 M D7 Faeces 16S #63
TCRγ+/- 6 F D7 Faeces 16S #64
TCRγ+/- 7 F D7 Faeces 16S #65
TCRγ+/- 8 F D7 Faeces 16S #66
TCRγ+/- 9 F D7 Faeces 16S #67
TCRγ+/- 10 F D7 Faeces 16S #68
Wild Type 1 M D21 Faeces 16S #69
Wild Type 2 M D21 Faeces 16S #70
Wild Type 3 M D21 Faeces 16S #71
Wild Type 4 M D21 Faeces 16S #72
Wild Type 5 M D21 Faeces 16S #73
Wild Type 6 F D21 Faeces 16S #74
Wild Type 7 F D21 Faeces 16S #75
Wild Type 8 F D21 Faeces 16S #76
Wild Type 9 F D21 Faeces 16S #77
Wild Type 10 F D21 Faeces 16S #78
TCRγ-/- 1 M D21 Faeces 16S #79
TCRγ-/- 2 M D21 Faeces 16S #80
100bp NEB ladder
100bp NEB ladder
Wild Type 1 M D3 Faeces 16S #9
Wild Type 2 M D3 Faeces 16S #10
Wild Type 3 M D3 Faeces 16S #11
Wild Type 4 M D3 Faeces 16S #12
Wild Type 5 M D3 Faeces 16S #13
Wild Type 6 F D3 Faeces 16S #14
Wild Type 7 F D3 Faeces 16S #15
Wild Type 8 F D3 Faeces 16S #16
Wild Type 9 F D3 Faeces 16S #17
Wild Type 10 F D3 Faeces 16S #18
TCRγ-/- 1 M D3 Faeces 16S #19
TCRγ-/- 2 M D3 Faeces 16S #20
TCRγ-/- 3 M D3 Faeces 16S #21
TCRγ-/- 4 M D3 Faeces 16S #22
TCRγ-/- 5 M D3 Faeces 16S #23
TCRγ-/- 6 F D3 Faeces 16S #24
TCRγ-/- 7 F D3 Faeces 16S #25
TCRγ-/- 8 F D3 Faeces 16S #26
100bp NEB ladder
100bp NEB ladder
TCRγ-/- 9 F D3 Faeces 16S #27
TCRγ-/- 10 F D3 Faeces 16S #28
TCRγ+/- 1 M D3 Faeces 16S #29
TCRγ+/- 2 M D3 Faeces 16S #30
TCRγ+/- 3 M D3 Faeces 16S #31
TCRγ+/- 4 M D3 Faeces 16S #32
TCRγ+/- 5 M D3 Faeces 16S #33
TCRγ+/- 6 F D3 Faeces 16S #34
TCRγ+/- 7 F D3 Faeces 16S #35
TCRγ+/- 8 F D3 Faeces 16S #36
TCRγ+/- 9 F D3 Faeces 16S #37
TCRγ+/- 10 F D3 Faeces 16S #38
Wild Type 1 M D7 Faeces 16S #39
Wild Type 2 M D7 Faeces 16S #40
Wild Type 3 M D7 Faeces 16S #41
Wild Type 4 M D7 Faeces 16S #42
Wild Type 5 M D7 Faeces 16S #43
Wild Type 6 F D7 Faeces 16S #44
100bp NEB ladder

Figure 5.9 - Example Library of 16S PCR Products Generated from Faecal Samples
Agarose gel image shows PCR products generated from faecal DNA extracted using a Maxwell Blood DNA Purification Kit. Microbial 16S DNA was amplified by PCR and run on a 1.5% agarose gel. Gel images show the expected band at around 500bp indicating amplification of the desired product, though at varying concentrations for each reaction. Sizes of the 100bp NEB ladder bands are indicated in bp on either side of the gel image.

A read tracking plot was then produced showing the numbers of reads at each stage of the 16S demultiplexing and DADA2 processing as described in detail in the Materials and Methods and Method Development Chapters (See Chapters 2 and 3). Briefly, our in house demultiplexing code detects the primers and barcodes and separates the reads into different files accordingly. Forward and reverse reads were trimmed according to the profiles in (Fig 5.10) as described above. Error rates are then learned and DADA2 sample inference is performed. Where a sufficient overlap is present, the forward and reverse reads are then merged and a high resolution Amplicon Sequence Variant table (ASV) is produced. Finally, chimeric reads are identified and eliminated by examining which sequences could have occurred from the combination of 2 more common sequences.

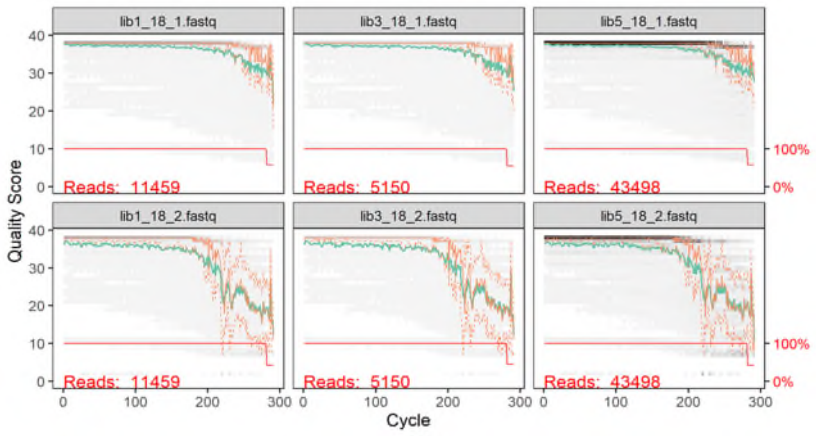


Figure 5.10 - Read Quality Profiles Highlight the need to Truncate Reads as Sequence Quality Declines with Length
Quality profile plots produced using the DADA2 function 'plotQualityProfile'. Plots provide a visual representation of the relationship between increasing read length and the decline in read quality as represented by Q score. These plots inform the length the reads are to be truncated to as a compromise between retaining enough length for merging and losing as few reads as possible. The truncation lengths decided on were 260bp for the forward and 210bp for the reverse reads. The X axis is the number of Illumina parallel sequencing cycles and is equivalent to the read length. Quality scores are represented by the greyscale heatmap as well as the mean Q score shown by the green line, the median by the orange line and the 75th and 25th quartiles by the dashed orange lines. The red line shows the percentage of sequences which extend to that position, in this case the reads are mostly 300bp.

The numbers of reads present at each of these processing stages was calculated and plotted (Fig 5.11). As we tend to see with MiSeq data processing, the majority of the reads are lost when looking for the correct barcoded primer. These lost reads tend to be those with neither the E338 nor the R806 primers, or just the R806 primer. It is implausible for so many reads to be generated without primer annealing so presumably these reads represent cases where the primer sequence becomes corrupted due to sequencing error or can't be accurately determined due to low FASTQ score. On this run, different amplicons produced using different primers were sequenced alongside the 16S products. As our demultiplexing pipeline can separate amplicons with the same Illumina indexing primer, but different barcoded primers and fewer libraries makes for simpler pooling, libraries 2, 3, 4, 5 and 6 were mixed with another amplicon library produced using the same index primer. Library 3 suffered a far higher level of read attrition during the 'Desired barcoded primer' stage than the other libraries. This was presumably due to an error either in the concentration calculations when mixing the 2 libraries with shared indexes, or a pipetting error when this pooling actually took place. The only other stage involving a substantial data loss was the DADA2 filtering stage, but this was far less severe than at the preceding stage. Almost no data was lost at any of the later stages.

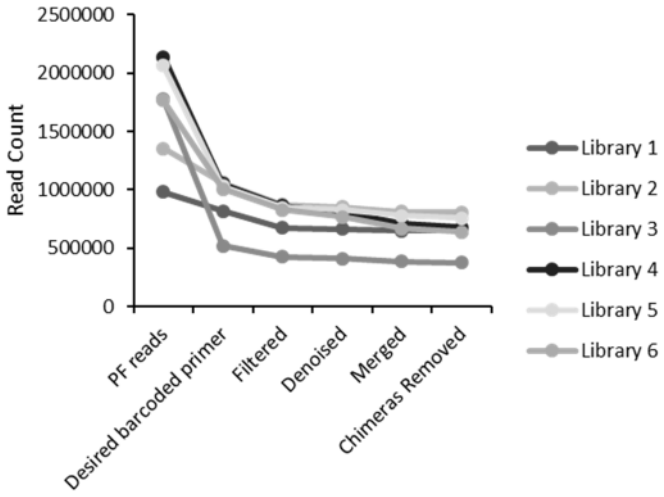


Figure 5.11 – Numbers of Reads at each Stage of the 16S Processing Pipeline
Plot shows the numbers of 16S sequencing reads at each stage of the processing. 'PF reads' is those that get through the Illumina MiSeq quality filter, 'desired barcoded primer' is those that have the expected primers and their barcodes pass quality filtering, 'Filtered' is reads which pass through the trimming phase and contain less than a certain number of predicted errors, 'Denoised' is combining rarer reads with the larger unique reads they are predicted to have arisen from mistakes in, 'Merged' is combining the forward and reverse reads, 'Chimeras removed' is removing rarer reads predicted to have arisen from a chimera of 2 more common reads. 16S PCRs were performed in triplicate and split across the 6 libraries. Y axis shows numbers of reads at each stage of the processing.

5.3.9 Beta-Diversity Plots reveal Differences in the Caecal Microbiota of TCRγ -/- Birds

Following data processing, we calculated Bray-Curtis distance metrics and visualised them using NMDS. The resulting plots did not display any obvious differences between the faecal samples of Wild Type, TCRγ +/- or TCRγ -/- birds at any of the time points (Fig 5.12A). Interestingly however, clear differences between some of the samples can be seen developing at day 21 becoming clearer at day 35, then reaching a complete segregation into 2 distinct clusters at day 49. However, there is no chicken strain association with either of these 2 clusters.

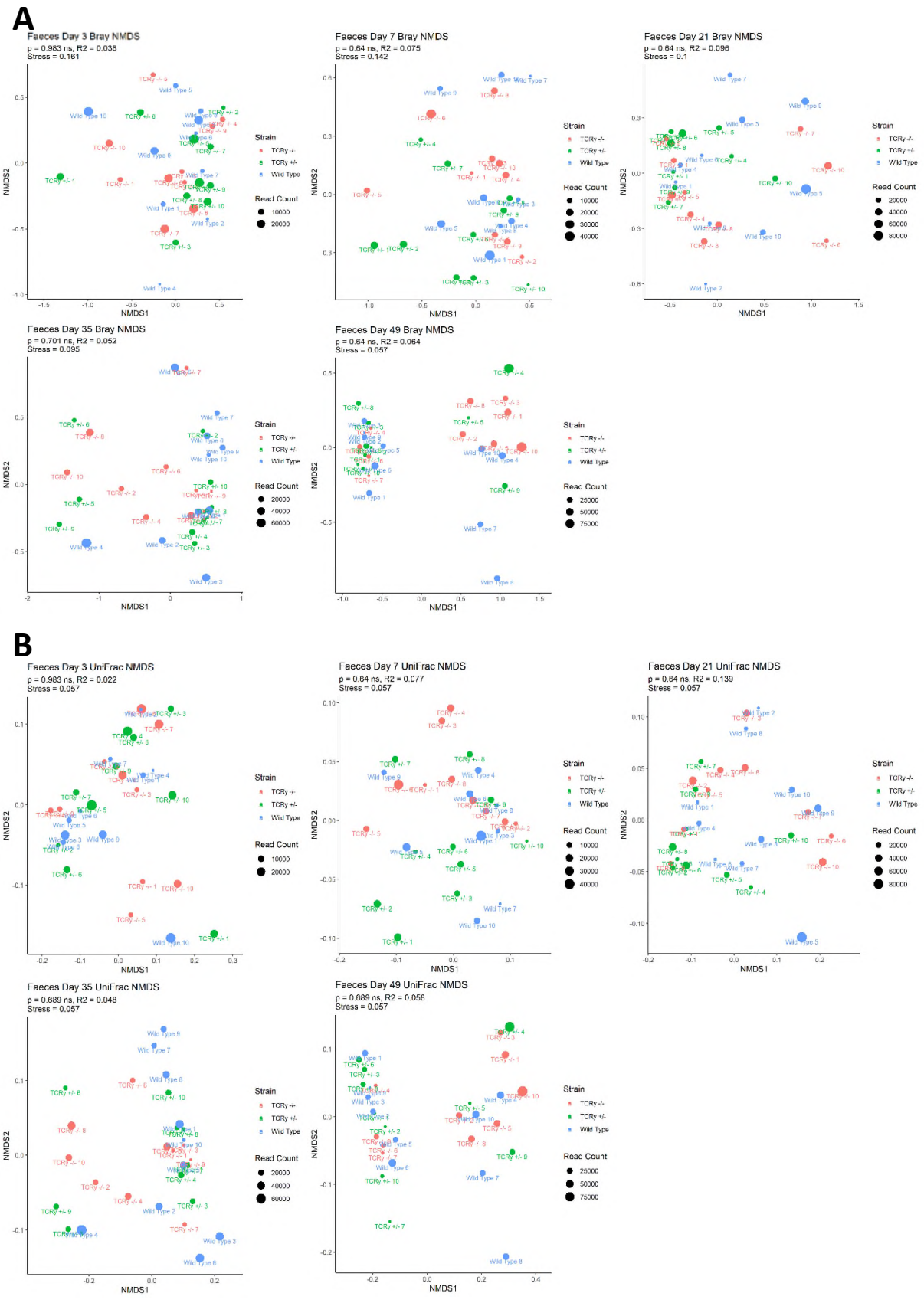
A similar pattern can be observed when comparing the samples using UniFrac, which has the advantage of incorporating a phylogenetic tree into its dissimilarity calculations and thus can account for the degree of relatedness between ASVs (Fig 5.12B). The results are almost identical to those from the Bray Curtis plots with a the same two distinct clusters developing at day 21 and

becoming more defined at days 35 and 49, but again there is a lack of chicken strain association with either of these clusters.

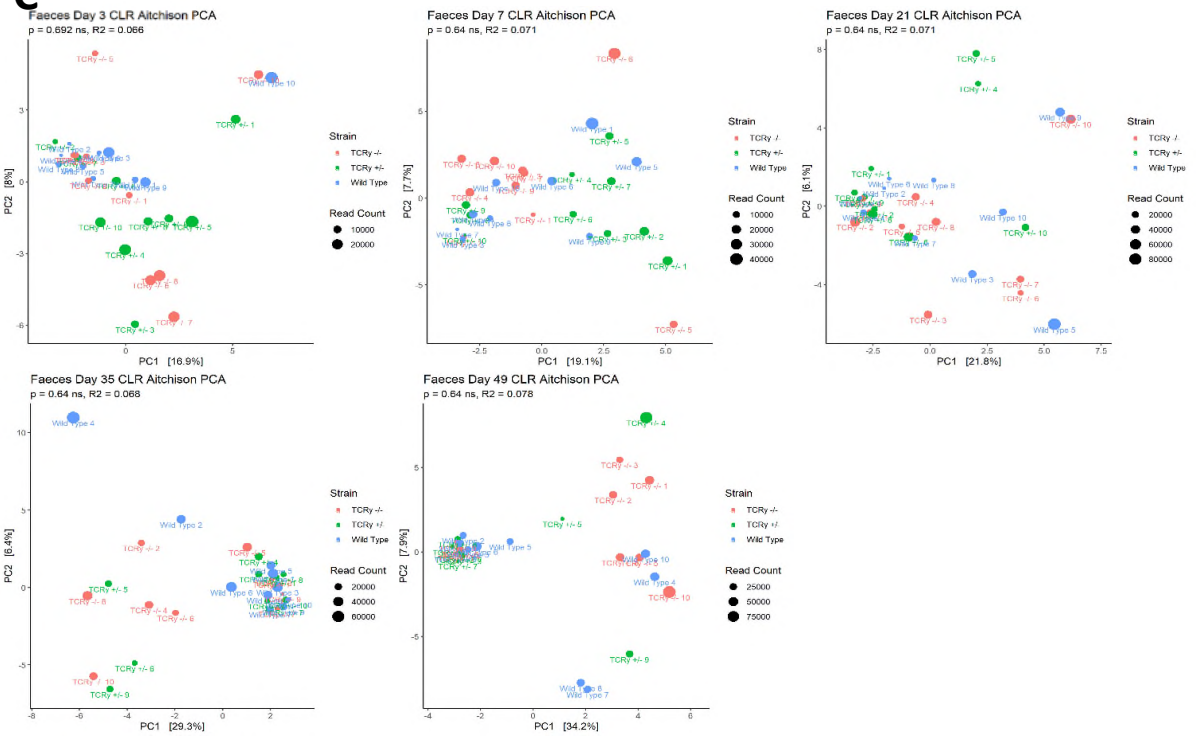
Generating plots using the Aitchison distance, thus accounting for the compositional nature of NGS sequencing data, again reveals no obvious differences between the strains (Fig 5.12C). The Aitchison plots show a similar pattern to Bray Curtis and UniFrac, however the separation into 2 groups becomes apparent slightly earlier and can be seen beginning to emerge at day 7. Interestingly, particularly at days 35 and 49, one of the 2 clusters is far tighter than the other, indicating one group is much more heterogeneous than the other.

The PhILR plots account for compositional data and similarly to UniFrac, have the additional advantage of taking into account relatedness (Fig 5.12D). These plots show an almost identical pattern to that observed in the Aitchison plots. This suggests that the differences observed using metrics not accounting for relatedness are not in closely related species and thus likely represent a genuine difference in niche, rather than just a different species occupying it.

Taken together these results show there are no large scale composition level differences between the 16S profiles of any of the strains, this is supported by a lack of any significant P values using PERMANOVA testing with False Discovery Rate (FDR) corrected P values. When examining the ordination plots, though no inter-strain differences were observed, the separation into 2 distinct clusters during the later timepoints has the potential to confound any difference that might exist between the strains and clearly shows the bird faecal samples are separating based on something. Additionally, particularly at day 49, one of these 2 separate groups is far more tightly clustered than the other, indicating it is more homogenous.


A
Faeces Day 3 Bray NMDS
p = 0.983 ns, R2 = 0.038
Stress = 0.161
Faeces Day 7 Bray NMDS
p = 0.64 ns, R2 = 0.075
Stress = 0.142
Faeces Day 21 Bray NMDS
p = 0.64 ns, R2 = 0.096
Stress = 0.1
Faeces Day 35 Bray NMDS
p = 0.701 ns, R2 = 0.052
Stress = 0.095
Faeces Day 49 Bray NMDS
p = 0.64 ns, R2 = 0.064
Stress = 0.057
B
Faeces Day 3 UniFrac NMDS
p = 0.983 ns, R2 = 0.022
Stress = 0.057
Faeces Day 7 UniFrac NMDS
p = 0.64 ns, R2 = 0.077
Stress = 0.057
Faeces Day 21 UniFrac NMDS
p = 0.64 ns, R2 = 0.139
Stress = 0.057
Faeces Day 35 UniFrac NMDS
p = 0.689 ns, R2 = 0.048
Stress = 0.057
Faeces Day 49 UniFrac NMDS
p = 0.689 ns, R2 = 0.058
Stress = 0.057
NMDS1
NMDS2
Strain
TCRγ -/-
TCRγ +/-
Wild Type
Read Count

C


Faeces Day 3 CLR Aitchison PCA
p = 0.692 ns, R2 = 0.066
PC1 [16.9%]
PC2 [8%]
Faeces Day 7 CLR Aitchison PCA
p = 0.64 ns, R2 = 0.071
PC1 [19.1%]
PC2 [7.7%]
Faeces Day 21 CLR Aitchison PCA
p = 0.64 ns, R2 = 0.071
PC1 [21.8%]
PC2 [6.1%]
Faeces Day 35 CLR Aitchison PCA
p = 0.64 ns, R2 = 0.068
PC1 [29.3%]
PC2 [6.4%]
Faeces Day 49 CLR Aitchison PCA
p = 0.64 ns, R2 = 0.078
PC1 [34.2%]
PC2 [7.9%]
Strain
TCRγ -/-
TCRγ +/-
Wild Type
Read Count

D

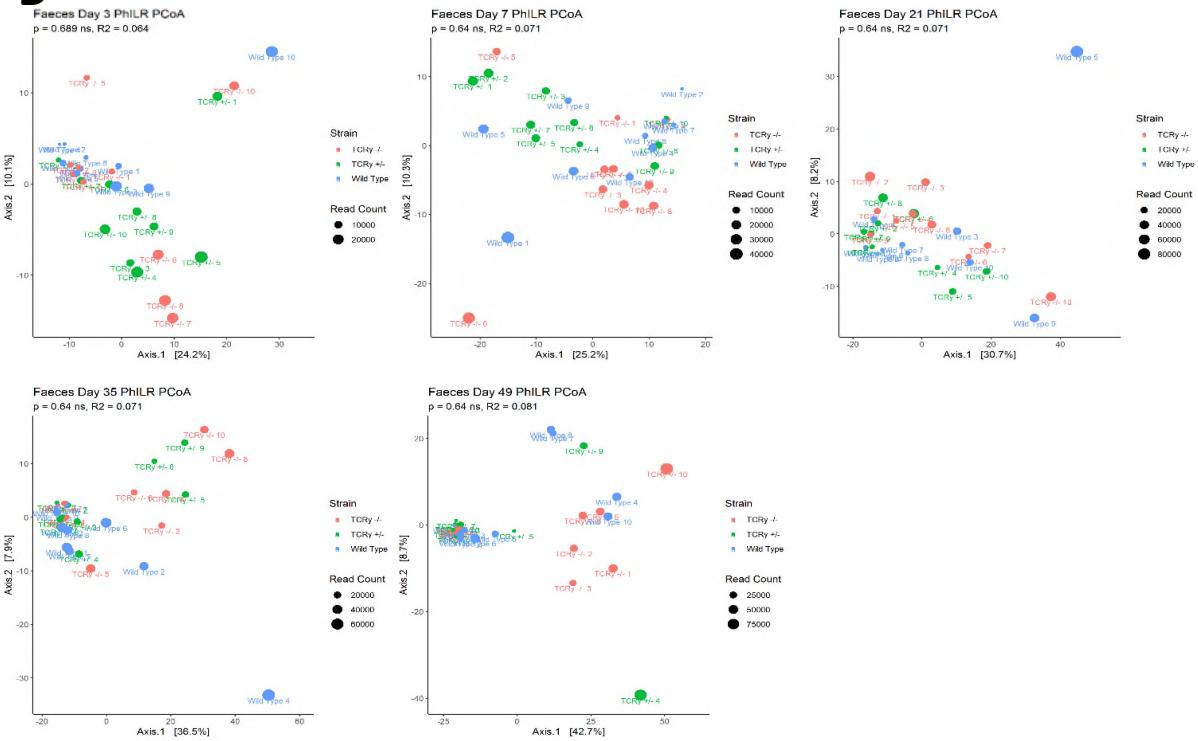

Faeces Day 3 PhILR PCoA
p = 0.689 ns, R2 = 0.064
Axis.1 [24.2%]
Axis.2 [10.1%]
Faeces Day 7 PhILR PCoA
p = 0.64 ns, R2 = 0.071
Axis.1 [25.2%]
Axis.2 [10.3%]
Faeces Day 21 PhILR PCoA
p = 0.64 ns, R2 = 0.071
Axis.1 [30.7%]
Axis.2 [8.2%]
Faeces Day 35 PhILR PCoA
p = 0.64 ns, R2 = 0.071
Axis.1 [36.5%]
Axis.2 [7.9%]
Faeces Day 49 PhILR PCoA
p = 0.64 ns, R2 = 0.081
Axis.1 [42.7%]
Axis.2 [8.7%]
Strain
TCRγ -/-
TCRγ +/-
Wild Type
Read Count

Figure 5.12 – β-Diversity Ordination Plots showing Comparisons between the Faecal Microbiota of TCR γ-/-, TCRγ +/- and Wild Type Birds
Dissimilarity matrices based on **A)** Bray-Curtis, **B)** UniFrac, **C)** Aitchison and **D)** PhILR distance indices were calculated for the faecal microbial communities sampled by 16S PCR with each unique ASV as a variable. Phylogenetic trees were calculated using MAFFT, followed by computation of pairwise distances and production of an initial tree using Neighbour joining. Optimisation of model parameters and production of the final tree was done using maximum likelihood. Each point represents an individual faecal sample. Strain is indicated by colour and read count is indicated by point size. P values and R2 values are indicated beneath the subtitle and the significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. P values were FDR corrected. Stress values are indicated below the title for NMDS and percentage of variation explained in the axes for RDA

Analysis of caecal contents using the same metrics reveals some separation when using Bray Curtis and UniFrac (Fig 5.13A). Using both of these metrics, the TCRγ -/- strain clusters distinctly from the Wild Type and only a single TCRγ +/- bird, TCRγ +/- 3, is intermixed with it. However, the FDR corrected PERMANOVA p values in both of these plots are not quite significant. *Post hoc* PERMANOVA testing of the Wild Type and TCRγ -/- strains in isolation does not quite pass the 0.05 significance threshold using Bray Curtis and UniFrac distances, however the P values of 0.065 and 0.058 respectively are extremely close. It should be considered however, that this weak separation is not robust to the distance metric used. When comparing the 3 strains using Aitchison and PhILR, no separation is present and all of the birds are totally intermixed.

Exploring the 16S data from Ileal contents using the same metrics, reveals a total lack of separation between any of the strains (Fig 5.13B). Additionally, all of the strains are homogenous in terms of distribution and the birds within them are evenly mixed. Curiously, the separation observed in the Day 49 faecal ordination plots is not present in either the caecal or ileal contents plots. This suggests that either this separation is only apparent when analysing the faecal microbiota or that the microbiota has become homogenous by day 65, at least in terms of changes over time.

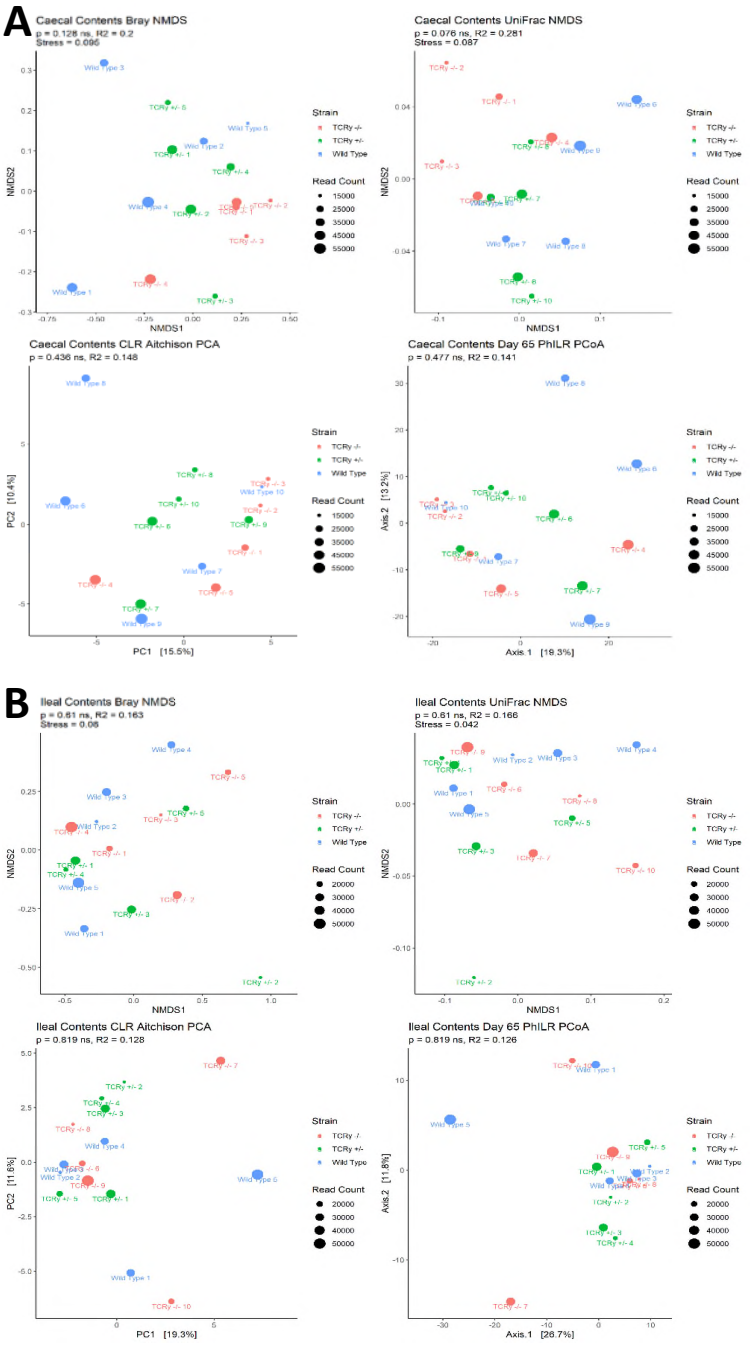


Figure 5.13 – β Diversity Ordination Plots showing Comparisons between the Ileal and Caecal Microbiota of TCRγ -/-, TCRγ +/- and Wild Type Birds

Dissimilarity matrices based on Bray-Curtis, UniFrac, Aitchison and PhILR distance indices were calculated for the **A)** caecal and **B)** ileal contents microbial communities, which were sampled by 16S PCR and each unique ASV as a variable. Phylogenetic trees were calculated using MAFFT, followed by computation of pairwise distances, production of an initial tree using Neighbour joining and optimisation of model parameters and production of the final tree using maximum likelihood. Each point represents an individual ileal/caecal sample. Strain is indicated by colour and read count is indicated by point size. PERMANOVA P values and R2 values are indicated beneath the subtitle and the significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. P values were FDR corrected. Stress values are indicated below the title for NMDS and percentage of variation explained in the axes for RDA.

5.3.10 Taxonomic Plots reveal Small Scale Differences in the Caecal Contents Microbiota of the TCRγ -/- Birds

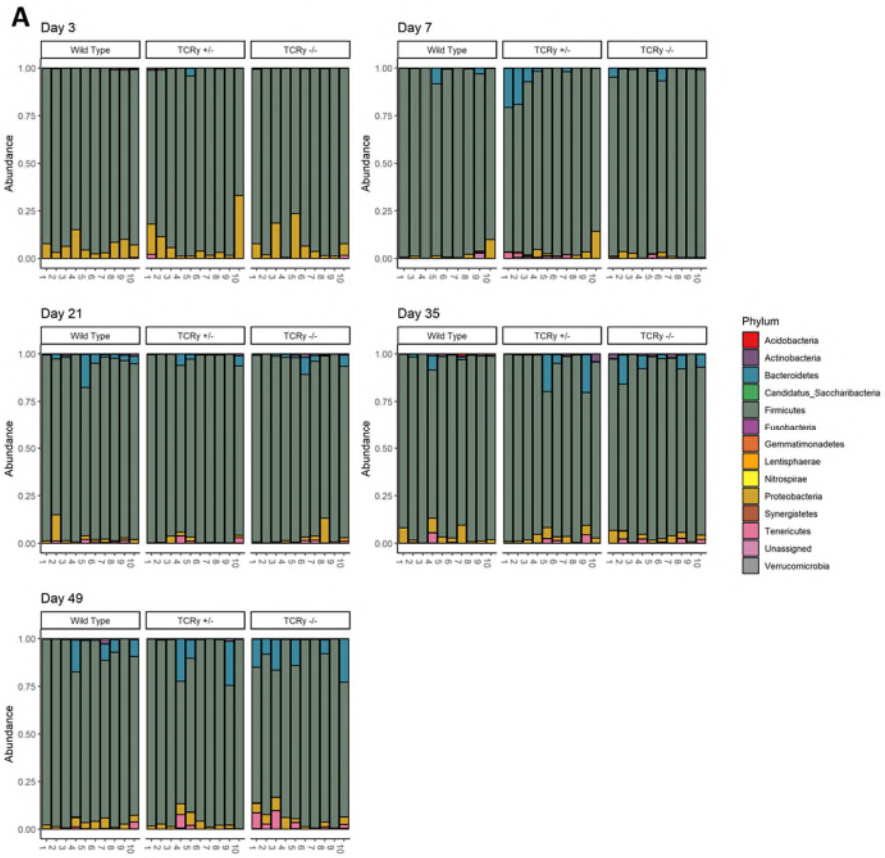
In order to gain a deeper insight into what bacteria make up the microbiota and whether any differences are present between the chicken strains, we used the Ribosomal Database Project (RDP) database to assign taxonomic classifications to the 16S sequences. This allowed the composition of the microbiota to be examined at various phylogenetic levels by generating stacked barplots for each individual bird and separating them by strain.

Examining the faecal microbiota first at the phylum level, reveals that the birds remain dominated by *Firmicutes* through the course of the study, regardless of age or strain (Fig 5.14A). Despite this, there are still some small changes in the other phyla over time, though we were unable to observe any consistent differences between the chicken strains. At day 3 there is a clear presence of *Proteobacteria,* though the proportion ranges quite widely between individuals, from ~30% to almost non-existent. Additionally, some trace amounts of *Bacteroidetes* and *Tenericutes* are present in some of the birds. At Day 7, there is a shift as the *Proteobacteria* decline and the *Bacteroidetes* increase, and interestingly these changes appear to occur in unison in the majority of the birds. Additionally, there is a slight increase in the *Tenericutes*, though they are still present at very low levels. Day 21 looks much the same as day 7, with *Bacteroidetes* becoming present in more of the individuals. Day 35 again looks similar to day 21, with low levels of *Actinobacteria* appearing in some of the birds. Finally, Day 49 sees a slight increase in *Tenericutes,* but otherwise looks similar to Day 35.

Though the microbiota was variable across individual birds as has been noted in other 16S studies, we were unable to discern any phylum level differences between the Wild Type and the TCRγ -/- and TCRγ +/- strains. However, the phylum level gives a relatively low resolution look at the microbiota and a lack of difference here does not preclude differences in lower taxonomic levels. Therefore, microbiota was next examined at the order level (Fig 5.14B). Thus, we were able to gain a

more fine scale picture of the resident flora in each of the chicken strains and over compare them over time. Examination of the microbiota using lower taxonomic levels such as Family and Genus are possible, however due to increases in the number of taxa present, it becomes impossible to view the entire microbiota composition on a single plot. Plotting the microbiotas of the chicken strains at the Order level reveals that the dominant phyla *Firmicutes* was mostly made up of *Clostridiales* and *Lactobacillales.* The levels of these orders fluctuate widely and appear to switch seemingly at random between being the majority taxa in the faecal microbiota. Though they are less common at the earlier time points and remain highly variable, as the chickens mature the *Clostridiales* generally tend to increase in proportion. As the birds age, there is also a tendency towards increases in *Bacteroidales* belonging to the phyla *Bacteroidetes* and *Anaeroplasmatales* belonging to the phyla *Tenericutes,* though the proportions of both appear to be inversely associated with the *Clostridales.*

Even when examined at a high taxonomic level, the microbiota tends to fluctuate, and levels of the same Phyla generally did not remain consistent in the same bird over time. Additionally, the microbial composition was remarkably consistent across strains, with no obvious systematic differences at any of the timepoints or in terms of development over time. This lack of a difference between strains is consistent with results from the ordination plots.


A
Day 3
Day 7
Day 21
Day 35
Day 49
Wild Type
TCRγ +/-
TCRγ -/-
Abundance
Phylum
Acidobacteria
Actinobacteria
Bacteroidetes
Candidatus_Saccharibacteria
Firmicutes
Fusobacteria
Gemmatimonadetes
Lentisphaerae
Nitrospirae
Proteobacteria
Synergistetes
Tenericutes
Unassigned
Verrucomicrobia

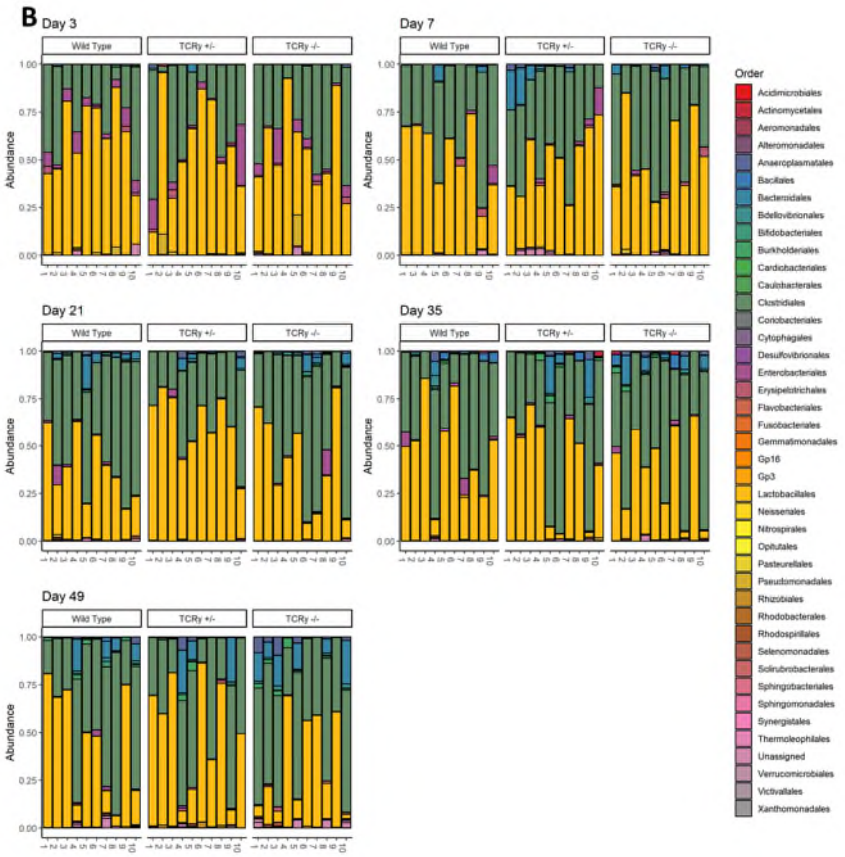

B
Day 3
Day 7
Day 21
Day 35
Day 49
Wild Type
TCRγ +/-
TCRγ -/-
Abundance
Order
Acidimicrobiales
Actinomycetales
Aeromonadales
Alteromonadales
Anaeroplasmatales
Bacillales
Bacteroidales
Bdellovibrionales
Bifidobacteriales
Burkholderiales
Cardiobacteriales
Caulobacterales
Clostridiales
Coriobacteriales
Cytophagales
Desulfovibrionales
Enterobacteriales
Erysipelotrichales
Flavobacteriales
Fusobacteriales
Gemmatimonadales
Gp16
Gp3
Lactobacillales
Neisseriales
Nitrospirales
Opitutales
Pasteurellales
Pseudomonadales
Rhizobiales
Rhodobacterales
Rhodospirillales
Selenomonadales
Solirubrobacterales
Sphingobacteriales
Sphingomonadales
Synergistales
Thermoleophilales
Unassigned
Verrucomicrobiales
Victivallales
Xanthomonadales

Figure 5.14 - Faecal Microbiota 16S Taxonomic Stacked Barplots at the Phylum and Order Level
16S sequences were assigned their relevant taxonomic classifications using the RDP database. Some sequences are ambiguous and there is therefore diminishing returns in terms of successful classification with decreasing taxonomic level. The proportions of the sequences belonging to each **A)** Phyla or **B)** Order are plotted for all the birds within each strain and the resulting values plotted at each Day post-hatch the faecal samples were taken. At the phylum level, the visible taxa from top to bottom are; Purple – Actinobacteria, Blue – Bacteroidetes, Olive Green – Firmicutes, Light Brown – Proteobacteria, Pink – Tenericutes. At the Order level the visible, more common taxa from top to bottom are; Dark Blue – Anaeroplasmatales, Blue – Bacteroidales, Olive – Clostridales, Purple – Erysipelotrichales, Yellow – Lactobacillales, Pink – Unassigned. Chicken strain is indicated in the top heading of each plot and the X axis numbers represent bird number within the chicken strain indicated at the top of the plots.

The caecal microbiota examined at the phylum level closely resembles the faecal microbiota in the later timepoints with similar levels of *Firmicutes, Tenericutes* and *Proteobacteria* (Fig 5.15A). It does however have slightly more *Proteobacteria* and the profile tends to be far more consistent across individuals. The *Firmicutes* are still the dominant phyla, though this dominance is less absolute than in the later faecal samples. Additionally, there are some hints of differences between the strains, with slightly more *Fusobacteria, Tenericutes* and *Proteobacteria* in the TCRγ -/- strain compared to Wild Type. The differences between Wild Type and TCRγ -/- were significant for *Tenericutes* and *Proteobacteria* using an FDR corrected unpaired Wilcoxon rank test and provide more information about the sources of the differences we observed in the caecal contents β-diversity plots using Bray-Curtis and UniFrac distances (Fig 5.13A). At the Order level, the *Firmicutes* again almost entirely consist of *Clostridiales,* with *Lactobacillales* making up only a small proportion of the microbiota. This profile again very much resembles that of some of the day 49 faecal profiles and has similar levels of *Anaeroplasmatales* and *Bacteroidales*. The differences that were present in the phylum level plots are far less apparent at this taxonomic level and it is extremely difficult to see any obvious differences between the strains due to the complexity of these plots.

Examining the microbiota of the ileal contents, at the Phylum level it is fairly uniformly and overwhelmingly dominated by *Firmicutes* with only trace amounts of *Actinobacteria, Bacteroidetes,* and *Proteobacteria* (Fig 5.15B). Strikingly, the profiles of the Ileal contents are totally distinct from those of the caecal contents and much more closely resemble the microbiota of the earlier faecal samples. Additionally, much like the faecal samples, there is a total lack of any difference between

the microbiotas from the different chicken strains at the phylum level. At the Order level, the *Firmicutes* mostly consist of *Lactobacillales,* however *Clostridiales* are still very prominent. The composition of the *Firmicutes* specifically also varies far more than it does in the caecal contents. In some birds, it is more dominated by *Lactobacillales* and in others *Clostridales* is more prominent. Again, at the Order level there is a lack of obvious differences between the microbiotas of the different chicken strains in the Ileal contents.

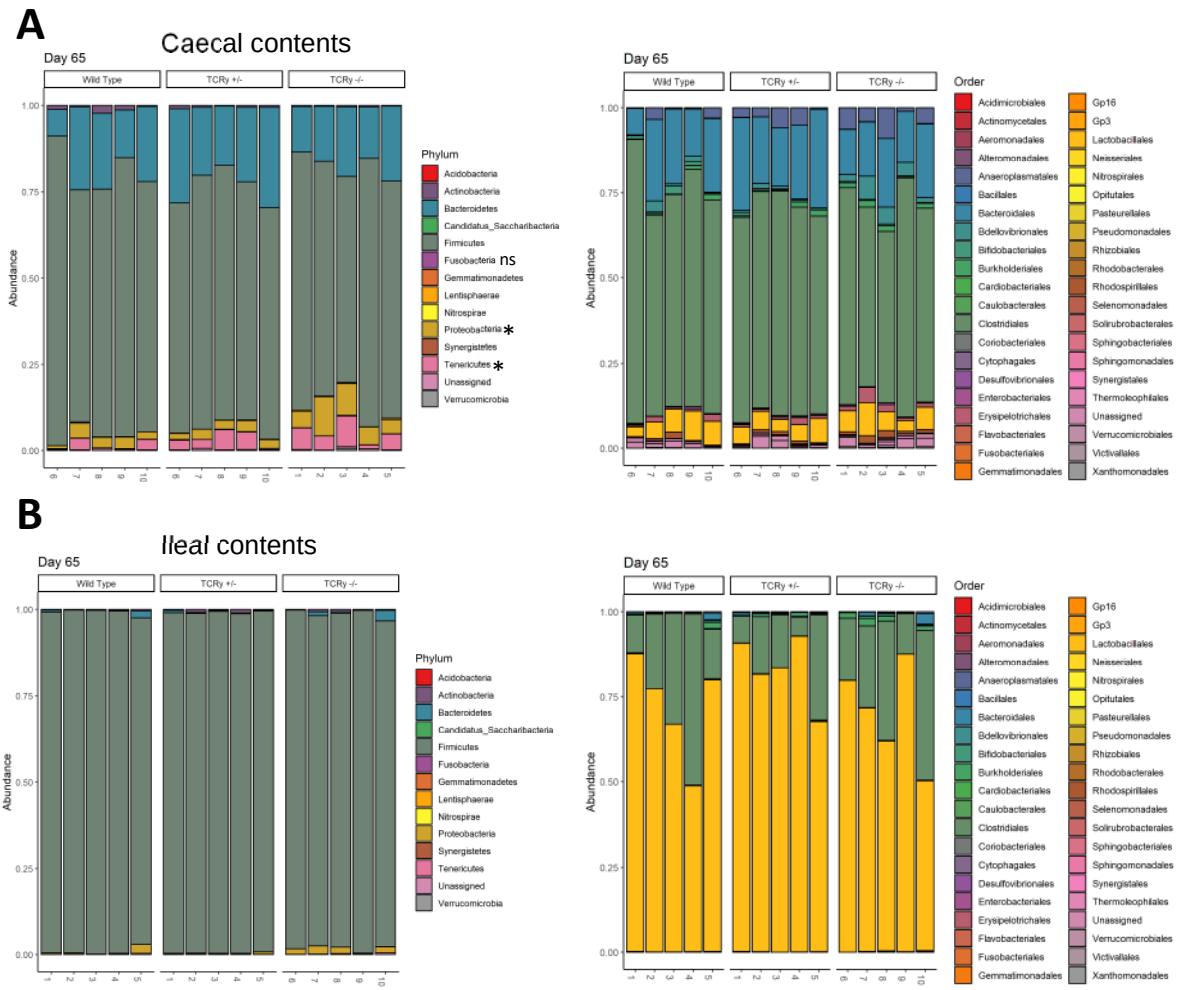


Figure 5.15 - Caecal and Ileal contents 16S taxonomic stacked barplots at the Phylum and Order level
16S sequences were assigned to their relevant taxonomic classifications using the RDP database. Some sequences are ambiguous and there is therefore diminishing returns in terms of successful classification with decreasing taxonomic level. The proportions of the sequences belonging to each Phyla or Order are plotted for all the birds within each strain and the resulting values plotted for both **A)** Caecal and **B)** Ileal Contents. At the phylum level, the visible taxa from top to bottom are; Purple – Actinobacteria, Blue – Bacteroidetes, Olive Green – Firmicutes, Light Brown – Proteobacteria, Pink – Tenericutes. At the Order level, the visible, more common taxa from top to bottom are; Dark Blue – Anaeroplasmatales, Blue – Bacteroidales, Olive – Clostridales, Purple – Erysipelotrichales, Yellow – Lactobacillales, Brown - Rhodospiralles, Pink – Unassigned. Chicken strain is indicated in the top heading of each plot and the x axis numbers represent chicken number within the strain indicated at the top of the plots, caecal contents were taken from one half of the birds and ileal contents from the other half. P values are from Wilcoxon signed rank tests conducted between Wild Type and TCRγ -/- birds only (*** p value < 0.001, ** p value < 0.01, * p value < 0.05). P values were FDR corrected.

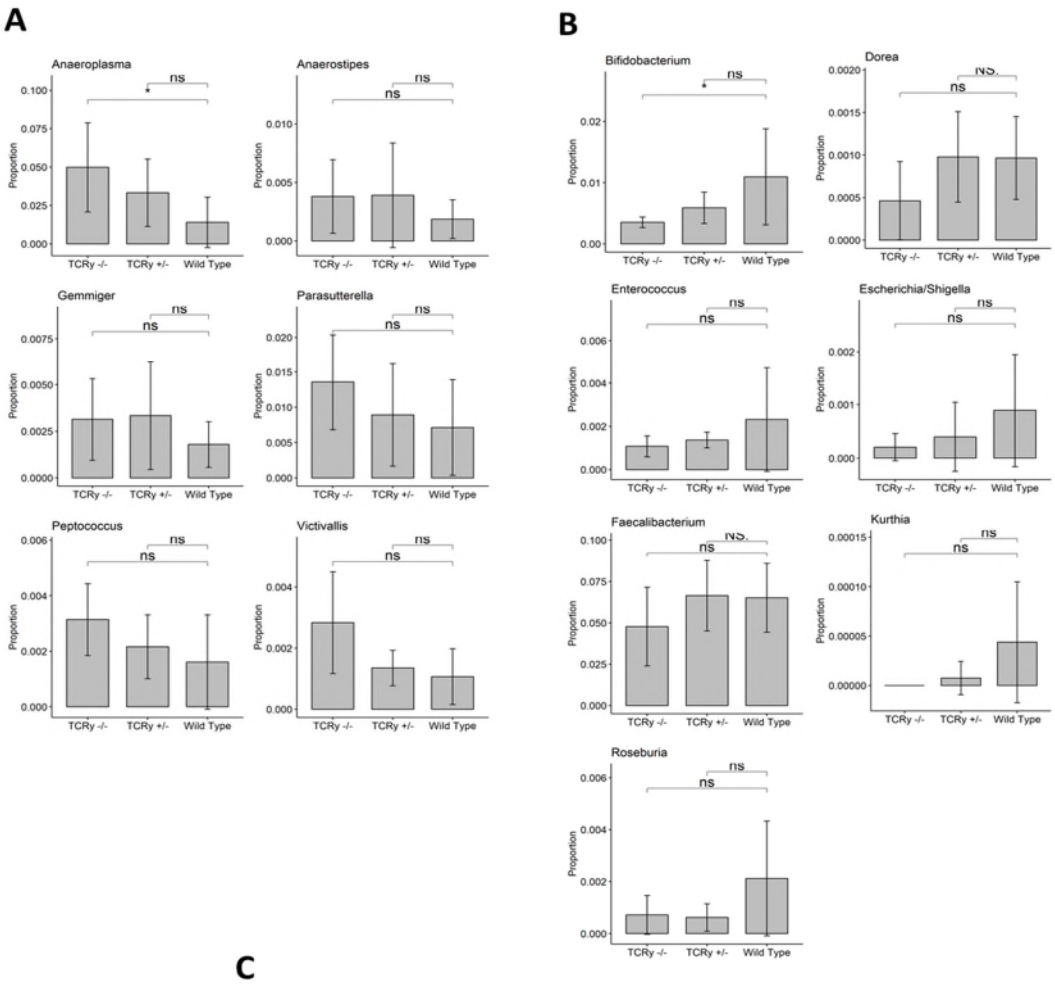
5.3.11 Plotting Individual Genera reveals Sources of the Differences in the TCRγ -/- Caecal Contents Microbiota

In order to further investigate the differences we observed between chicken strains in the caecal contents samples both using Bray-Curtis and UniFrac (Fig 5.13A) and in the phylum level composition plots (Fig 5.15A), SIMPER analysis was performed on the Bray-Curtis distance matrices to ascertain the sources of these differences at the level of individual genera. In addition to a lack of any differences between the chicken strains in the faecal and ileal samples, the variability present in their microbial profiles (Fig 5.14A,B) makes it difficult to make any meaningful comparisons at the genera level and these samples were therefore omitted from analysis at the genera level.

The SIMPER analysis produced a number of ASVs of interest, however in the majority of these cases, when examined at the genus level, the differences between chicken strains disappeared. One exception to this was the genus *Anaeroplasma,* which is made up almost entirely of the second biggest contributing 16S sequence to the differences seen in the Bray-Curtis plots. Additionally, *Anaeroplasma* makes up the majority of the *Tenericutes,* which were significantly increased in the TCRγ -/- chicken strain in the phylum level plots. (Fig 5.15A). The *Anaeroplasma* are present in the Wild Type strain at just ~1.25% on average but quadruple to 5% on average in the TCRγ -/- (Fig 5.16A). The heterozygous TCRγ +/- strain has a proportion of *Anaeroplasma* of ~3%, intermediate between the TCRγ -/- and Wild Type. This result was shown to be significant using an unpaired Wilcoxon signed rank test.

Further exploratory analysis of the caecal contents, revealed that a number of other genera differ somewhat between the strains. Genera that were increased in the TCRγ -/- strain were *Anaerostipes, Gemmiger, Parasutterella, Peptococcus* and *Victivallis* (Fig 5.15A). Genera at lower levels in the TCRγ -/- strain were *Bifidobacterium, Dorea, Enterococcus, Escherichia/Shigella, Faecalibacterium, Kurthia* and *Roseburia (*Fig 5.16B*).* Of these, only *Bifidobacterium,* which was

present at ~1% and ~0.3% in Wild Type and TCRγ -/- respectively, was significant using an unpaired Wilcoxon signed rank test.



C

Phylum	Order	Genus
Tenericutes	*Anaeroplasmatales*	*Anaeroplasma*
Firmicutes	*Clostridales*	*Anaerostripes*
Firmicutes	*Clostridiales*	*Gemmiger*
Proteobacteria	*Burkholderiales*	*Parasutterella*
Firmicutes	*Clostridales*	*Peptococcus*
Lentisphaerae	*Victivallales*	*Victivallis*
Actinobacteria	*Bifidobacteriales*	*Bifidobacterium*
Firmicutes	*Clostridales*	*Dorea*
Firmicutes	*Lactobacillales*	*Enterococcus*
Proteobacteria	*Enterobacterales*	*Escherichia/Shigella*
Firmicutes	*Clostridales*	*Faecalibacterium*
Firmicutes	*Bacillales*	*Kurthia*
Firmicutes	*Clostridales*	*Roseburia*

Figure 5.16 - Individual Genera Plots from Caecal and Ileal Contents
SIMPER analysis was performed on Bray-Curtis distances from caecal contents, in order to further investigate the contributors to the weak separation we observed between chicken strains in these samples. A number of genera were shown to be different with **A)** increases in Anaeroplasma, Anaerostripes, Gemmiger, Parasutterella, Peptococcus and Victivallis in the TCRγ -/- strain relative to wild type and **B)** decreases in Bifidobacterium, Dorea, Enterococcus, Escherichia/Shigella, Faecalibacterium, Kurthia and Roseburia in the TCRγ -/- strain relative to wild type. P values are from Wilcoxon signed rank tests (*** p value < 0.001, ** p value < 0.01, * p value < 0.05)
C) Table shows the phylum and order of each plotted genus

5.3.12 Pseudometagenomic Plots Generated using Tax4Fun Predict Changes to Metabolic Genes in the Caecal Contents of TCRγ -/- Birds due to Reduced *Bifidobacterium*

To gain an insight into the functional consequences of the microbiota differences observed in the caecal contents of the TCRγ -/- strain compared to Wild Type, the metagenomics prediction package Tax4Fun[26] was utilised. This pipeline predicts the functional profile of the microbiota, enabling pseudo-metagenomic analysis from more cost effective 16S data. This package only works with the SILVA 16S database, so taxa were first reassigned. In short Tax4Fun takes this SILVA profile as abundance data and converts it to a profile of Kyoto Encyclopaedia of Genes and Genomes (KEGG) organisms. 16S rRNA copy numbers are known to vary widely between prokaryotes and thus Tax4Fun normalises the KEGG data using a database of 16S copy numbers for each organism[27]. Finally, the KEGG profile is converted to metagenomic data using precomputed profiles for each organism.

A Bray-Curtis distance matrix was calculated for each of the birds within each chicken strain based on their predicted metagenomics profiles. This enabled the use of SIMPER analysis to extract the contribution of each gene to the differences between the Wild Type and TCRγ -/- strains and reveals a number of potential differences at the individual gene level (Fig 5.17). All of the differences come from genes which occupy a greater proportion of the predicted metagenome in Wild Type birds compared to TCRγ -/-, with TCRγ +/- birds at an intermediate level between the two. The KEGG names of the genes of interest identified were as follows, long-chain acyl-CoA synthetase, fatty acid synthase, glycogen debranching enzyme, beta-galactosidase and ATP-dependent DNA helicase RecG.

The only gene that was significant when tested using an unpaired Wilcoxon signed rank test was glycogen debranching enzyme. This gene was present at a relatively high level (~0.0006) considering that the mean proportion of all the genes in the Wild Type birds was 0.000184, and the highest proportion was 0.00739. It was reduced by ~30% when comparing the TCRγ -/- strain with

Wild Type. Fatty acid synthase was also very close to significance when tested using an unpaired Wilcoxon signed rank test with a p value of 0.056. Fatty acid synthase was reduced ~3 fold in the TCRγ -/- compared to the Wild Type, however the high variability present in the Wild Type birds prevented it from reaching significance.

To discover which change at the genera level was responsible for these differences in gene proportion, each taxa was successively removed and the gene plots re-examined. Using this method to descend down the taxonomic levels, the list of candidates was narrowed down until the responsible genera was identified. Using this technique, it was ascertained that *Bifidobacterium* was responsible for every one of these predicted metagenomics differences. The fact that *Bifidobacterium* was found to be significantly lower in the TCRγ -/- strain correlates with this finding.

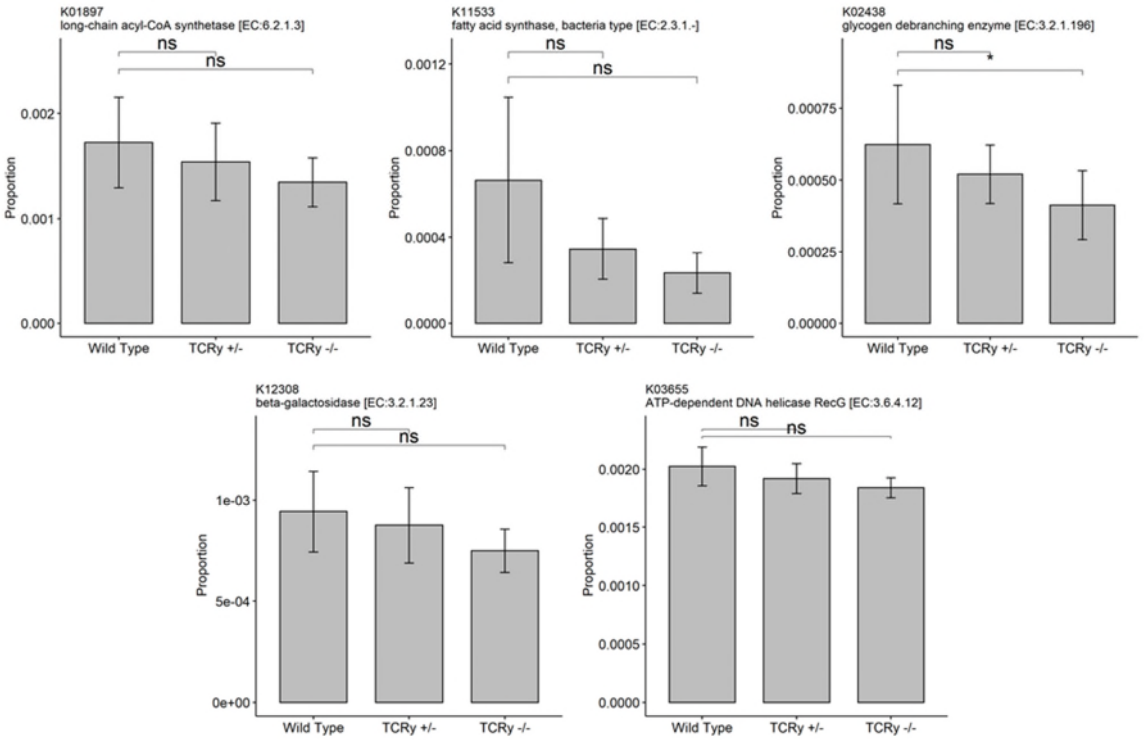


Figure 5.17 - Tax4Fun Individual Gene Plots
SIMPER analysis was again used to extract the largest contributors to Bray-Curtis differences calculated from the Caecal Tax4Fun profile of each individual bird. This analysis highlighted a range of metabolic genes which are decreased in the TCRγ -/- strain compared to the Wild Type, including long-chain acyl-CoA synthetase, fatty acid synthase, glycogen debranching enzyme, beta-galactosidase and ATP-dependent DNA helicase. By removing each genera from the data and replotting these genes it was ascertained that the reduction in Bifidobacterium in TCRγ -/- relative to Wild Type was responsible for all of these pseudo-metagenomics changes. The KEGG name and function of each of these genes is displayed above the plot, error bars are standard deviation. P values are from FDR corrected Wilcoxon signed rank tests (*** p value < 0.001, ** p value < 0.01, * p value < 0.05)

5.4 Discussion

This chapter aimed to characterise the interactions between chicken γδ T cells and the enteric microbiota. There is a clear consensus in the literature that the gut microbiota plays an important role during immune development and ensures the host immune system is healthy and functional[28]. γδ T cells are the first T cells to rearrange in the thymus and exit before αβ T cells have had time to develop[29–31]. There are 3 waves of thymic export of γδ and αβ T cells during development, however, in each of these γδ T cell export occurs first, though the first wave of αβ T cell export occurs before the second wave of γδ export and so forth[31]. Unsurprisingly then, γδ T cells have been shown to play an important role in early-life immunity and are involved in responses against pathogens in young animals, amongst a range of other functions[32,33]. Furthermore, γδ T cells have been shown to localise to mucosal sites in both chickens and mammals in a process thought to be mediated by molecules released by local epithelial cells[34]. However, one fundamental difference between chickens and more commonly studied species such as mice and humans is the prominence of circulating γδ T cells. In chickens and other γδ high animals, the proportion of circulating T cells of the γδ lineage is far higher than γδ low animals, such as mouse and human[35]. The work presented here seeks to further examine the relationship between γδ T cells and the microbiota by removing one and examining the impact on the other using NGS sequence based analyses. First, we considered the impact on the developing TCR gamma repertoire in the absence of a microbiota using Germ Free birds, with tissue samples collected at various timepoints. Second, to explore the role γδ T cells play in regulating the microbiota composition, we used 16S sequencing to observe the

microflora of TCRγ -/- birds, which lack functional γδ T cells, and compare it with heterozygote and intact birds. We used faecal samples to follow individual birds over time as well as taking end point caecal and ileal contents samples.

The results show that the whilst there were small changes in the β-diversity ordination plots and TRGV gene usage profiles during the earliest timepoints, the chicken TCR gamma repertoire appeared mostly unaffected and therefore does not seem to rely on the microbiota for development to the extent that other enteric lymphocytes do. The GF immune system has been extensively characterised in mice and the consequences of a lack of microbe mediated development is far reaching in these animals[36]. The lack of gut bacteria in these animals has a clear impact on a wide range of cell types such as αβ T cells, which have been shown to be decreased in number and B cells which are impaired in IgA production[37,38]. However, the data in this chapter suggests that the global TCR gamma repertoire is broadly unaffected by the absence of a microbiota in chickens. Interestingly the numbers of γδ T cells in the gut of GF mice and re-conventionalised rats is also reported to be independent of germ-free or conventional status[14,15,20,39,40].

Though statistically significant differences between the Complex and Germ Free birds were evident in β-diversity plots of the Day 14 Bursa, Spleen and Jejunum, the lack of obvious visual separation between points on the ordination plots and the low PERMANOVA R^2 values indicate that even where an effect of microbial absence was apparent, it was small. Furthermore, it could have been predicted that the impact of a lack of microbiota would be greatest in tissues with the highest microbial load, namely colon and caecum. According to this logic, the differences would be especially pronounced in the caeca, which have the highest concentration of microbes at 10^{11} per gram of caecal contents[41]. However, contrary to these expectations, there were no large scale differences in either of these tissues. These results provide further evidence that the chicken TCR gamma repertoire develops largely independent of the gut flora. However, some TRGV gene subsets may be more influenced by the microbiota than others. This seeming indifference to the microbiota

is perplexing in the context of studies on other aspects of the GF chicken enteric immune system, which have found reductions in mucin production, CD4+ lymphocytes and the chicken B cell marker ChB6[5,42]. The apparent insensitivity of the γδ T cells to microbial signals could be related to their role in early immunity. As chicken γδ T cells have been shown to exit the thymus prior to hatch, when the developing embryo is still sterile, it makes sense that the homing of these early T cells would not be dependent on the microbiota.

Despite a lack of any large global changes to the TCR gamma repertoire, we did observe some differences in repertoire subdivided by TRGV gene subsets. Though small differences were observed in the proportions of TRGV4.2 and TRGV4.4, when comparing the Complex and GermFree birds with the Conventional and Tetra birds, due to the radically different microbiota compositions between these groups, this difference is most likely due to a difference in rearing conditions of the 2 groups leading to some sort of drift effect. This should however, be investigated in greater detail with all of the same groups reared in parallel to ensure this effect is indeed an artefact of confounding variables introduced during rearing.

One of the notable differences between the Complex and Germ Free birds was that the proportions of 2 of the most public TRGV genes TRGV2.7/2.18 and TRGV2.20 were consistently higher at day 14 in the Germ Free birds, as well as at most of the other timepoints. As the birds aged, there was a general decrease in TCR gamma repertoire publicity and a significant increase in diversity regardless of bird group. The higher proportions of these TRGV genes in the Germ Free repertoire, particularly at the earlier timepoints suggest this process may be somewhat retarded in the Germ Free birds, though this is slightly contradicted by the higher general publicity in the Complex birds, but it should be considered that these plots do not take TRGV gene proportion into account. Additionally, the TRGV4.2 and TRGV4.4s were highly public and increased in the Complex birds in the day 14 Bursa, Spleen, Jejunum and Colon, though these V genes are distinct from

TRGV2.7/2.18 in particular because they undergo a decrease in publicity over time and do not remain dominated by super public canonical sequences.

Furthermore, as the differences between the Complex and Germ Free birds we observe at day 14 are small it may be the case that this effect is more pronounced at earlier ages and would be interesting to repeat this analysis with tissues from younger birds. Taken together, these differences suggests that the microbiota may play some limited role in facilitating the early life gut migration of the more diverse TCR gamma rearrangements that come to dominate the repertoire slightly later in life, particularly TRGV3.3. It has been shown that γδ T cells in humans and mice can respond in various different ways to different antigens, ranging from adaptive-like, featuring clonal expansions, to innate-like, featuring resident monoclonal populations[43]. These different types of responses are often governed by the TCR gamma and delta gene segments, such as the invariant human Vγ9Vδ2 T cells, which have been shown to respond to phosphoantigen, a metabolic intermediate in sterol production by mycobacteria, and are the very first T lymphocytes generated in the developing foetus[44,45]. A similar phenomena may be at work in the chicken TCR gamma repertoire whereby monoclonal γδ T cells are the first to rearrange and exit the thymus and may also respond to common pathogen associated molecules or signals of cell damage, though this would need to be tested using sorting and binding/activation assays.

Additionally, the fact that the modest impacts of the absence of a microbiota on the TCR gamma repertoire was less evident by day 28 and day 55 indicates that other signals are facilitating this migration and TCR gamma repertoire development. In mice and humans, previous studies have indicated the Butyrophilin-like molecules are important players in this process[34,43,46,47]. In chickens, there is a highly polymorphic group of Butyrophilin homologues known as the BG genes that may play a similar role[48,49], though this is yet to be experimentally determined. However, considering the small differences we observed at early time points, it does seem that the microbiota has some limited role to play in the development of a mature TCR gamma repertoire and that these

developmental signals may be mostly, but not be entirely host derived. The microbiota may perhaps be providing antigens to shape and 'train' the immune repertoire as it does in other animals and lymphocyte subsets. Unfortunately, there is a distinct paucity of information on the effect of the microbiota on developing T cells in chicken. Thus far, the only studies looking at this have found that the numbers of T cells appear to be reduced when the microbiota is absent or disturbed and that this appears to be mediated by detection of the microbiota-derived acetate[50,51]. The small impacts we observed in the early TCR gamma repertoire of Germ Free birds may be caused by the absence of similar antigens, or the impact may be indirect and mediated by other affected lymphocyte subsets, or the epithelia itself.

To determine the effects of γδ T cells on the intestinal microbiome, the 16S profiles of faeces or intestinal contents were compared between TCRγ-/-, heterozygote TCRγ+/- and conventional chickens. The faecal samples were taken from birds at regular intervals over time from 3-49 days post hatch. Although there were not any large scale faecal microbiota differences in the absence of functional γδ T cells, there were some changes in the caecal contents. The gut microbiota undergoes a rapid maturation process during the early stages of life. This development represents an ecological succession with marked points of change driven in part by the host immune system. Over time, the data showed that early microbiota development in chickens is characterised by a substantial decrease in *Proteobacteria*. A similar pattern of development has been seen in earlier studies[52,53], however our birds seem to experience this change somewhat sooner than has been previously observed. Both of these studies also saw the same increase in *Bacteroidetes* we observed at around Day 49. Additionally, Videnska *et al.* 2014 observed a similar decrease in the family *Lachnospiraceae* correlating with the increase in *Bacteroidetes,* during a long term farm experiment[53]. This developmental change of the chicken faecal microbiota was unaffected in the absence of γδ T cells, indicating that they do not play a major role in guiding gut flora composition over time. While It has been shown that γδ T cells respond to enteric pathogens such as *Salmonella*[32,54], it would seem from

our data that they are not involved in regulating microbial development or homeostasis as has been shown in other enteric immune lymphocytes, such as IgA producing plasma cells[55–58].

The microbiota of the caecal and ileal contents in this study closely resembles previous microbial communities which have been reported in the literature. In the caeca, Pandit *et al.* reported similar though slightly higher proportions of *Bacteroidales* and *Clostridiales* depending upon chicken line and location[59], whereas Glendinning *et al.* observed a similar overall profile though they found almost no *Bacteroidales*[60]. The *Bacteroidales* made up around 20% of the caecal microbiota in our samples. It should be noted however, that the caecal contents used in these studies were collected from birds younger than those described in this chapter[60]. The microbiota of the ileal contents previously reported in the literature, again very closely resembles that seen in this study, with *Lactobacillales* making up the majority of the overwhelmingly dominant *Firmicutes*[60].

Re-examination of the Day 49 faecal samples with the differing composition of the caecal and ileal contents in mind, reveals that the faecal microbiota profiles tend to resemble either caecal or ileal contents. This suggests that the faecal microbiota is highly dependent upon the periodic emptying of the caecal contents into the gut, as has been previously observed[60]. The *Clostridales* dominated profile comes from the caecal contents whereas, *Lactobacillales* as the dominant order indicates the absence of caecal emptying and more closely resembles the ileal contents. Consistent with this, the differences in ordination plots only appear during the later timepoints. In early life the caecal contents far more closely resemble the microbial profile elsewhere in the gut[60]. However, during later timepoints, the caecal contents undergo a rapid and dramatic change, diverging away from other gut sites such as the ileum[53,60]. Additionally, the clusters representing faeces containing caecal excreta clustered together far more closely, indicating their homogeneity. This is likely due to the fact that the bacterial load in the caeca is far greater than elsewhere in the gut. As a result, whenever the caeca empty into the gut, bacteria originating from them will make up the majority of the 16S DNA in the faeces. Despite the rapid changes it undergoes, the caecal microbiota has been shown to be relatively consistent at any given timepoint compared to other sites in the gut[60]. This

tendency of caecal emptying to influence the faecal microbiota and mask potential results should be considered in future studies.

Despite not observing any large scale differences in any of the samples collected, we did notice some changes to the caecal microbiota in the TCRγ -/- birds. In terms of β diversity there was some evidence of visual separation of this group when using Bray-Curtis and UniFrac distance metrics, though this was not quite significant after correction and these differences were not present in the equivalent compositional metrics, Aitchison's distance and PhILR. We did however find significant differences in the minority phyla *Proteobacteria* and *Tenericutes,* both of which were increased in the TCRγ -/- strain. Upon further examination of these differences at the genera level, we found that the increase in *Tenericutes* we observed in the TCRγ -/- strain was driven almost exclusively by the genus *Anaeroplasma,* whereas the increase in *Proteobacteria* was more multifactorial, driven by smaller differences in *Parasutterella* and other genera. Furthermore, we observed a significant decrease in the genus *Bifidobacterium* in the TCRγ -/- strain as well as a range of non-significant differences in a range of other genera.

Anaeroplasma is an obligate anaerobe and parasitic mycoplasma commonly found in the gut of chickens and ruminants but also herbivores in general[61]. Though its role in the gut is not well known, it is thought to be involved in nutrient cycling and protein turnover by degrading cell walls and enzymes[62]. Additionally, it modulates gram negative bacterial populations by microbial predation, fulfilling a function not dissimilar to that of bacteriophage[62]. In chicken, *Anaeroplasma* has been shown to be increased by heat stress[63] and a diet including plant extracts, and is associated with higher Feed Conversion Ratios (FCRs) and therefore less efficient digestion[64]. Despite this association with increased FDR, data from other animals suggests they may have some beneficial roles. They are present in swine fever resistant pigs[65], are potential anti-inflammatory probiotics enhancing mucosal IgA levels in humans[66] and increase intestinal IgA levels through TGF-β induction in mice[67].

Bifidobacterium are a probiotic genera used in chickens as well as humans and are well documented as being a bacteria which is beneficial for gut health[68,69]. In chicken their presence in the gut microbiota leads to reduced *Campylobacter*[70], a common zoonotic pathogen, as well as reductions in groups of other potential pathogens including Coliforms and *Clostridia*[69]. Additionally, the use of *Bifidobacterium* as a probiotic, significantly reduced the number of birds suffering from cellulitis, reducing meat condemnation and waste[69].

The predicted metagenomics data gives an insight into the potential consequences of these alterations in the caecal microbiota. Interestingly, the genes identified as being altered were decreased in TCRγ -/- birds and were associated with the decrease in *Bifidobacterium* in these animals. However, one issue with predicted metagenomics is its reliance on Whole Genome Sequences (WGS). Therefore, these tools will be biased in favour of more widely studied organisms, such as *Bifidobacterium.* Therefore, though this does not subtract from the value of this technique, these results should be interpreted with caution and do not eliminate the possibility of other metagenomics changes arising from some of the less well studied genera that were altered in the caecal contents of TCRγ -/- chickens. Though predictive metagenomics does allow an insight into the implications of microbial differences observed using 16S, a real metagenomics study would be needed to properly explore these differences.

Of the genes that were shown to be reduced, two of them were involved in fatty acid metabolism; long-chain acyl-CoA synthetase and fatty acid synthase. Fatty acid synthases are a family of enzymes involved in *de novo* biosynthesis of fatty acids[71]. Though the diversity of fatty acid products produced by these enzymes is incredibly broad in spectrum, the potential implications on SCFA production are of particular relevance to the microbiota as well as providing an energy source for epithelial cells and reducing inflammation through butyrate mediated Treg induction[72,73]. Indeed γδ T cells themselves may also interact with, and be influenced by, butyrate or similar metabolites[74]. Additionally, two of the other genes predicted to be reduced in the TCRγ -/- strain were involved in

sugar metabolism; glycogen debranching enzyme and beta-galactosidase. Beta-galactosidase catalyses hydrolysis of the terminal lactose of galactooligosacharides (GOS). GOS are of particular importance for *Bifidobacterium* as they rely heavily on their metabolism as a source of energy[75]. GOS can be given as prebiotic supplements to stimulate growth of beneficial bacteria such as *Bacteroides, Lactobacillus* and *Bifidobacterium* itself in chickens[76]. Glycogen debranching enzyme on the other hand is involved in glycogen metabolism. Bacterial glycogen is a bacterial energy and carbon store and is important for microbial resilience to environmental changes[77].

Therefore, the subtle changes in caecal microbiota composition in the absence of functional γδ T cells suggest a role in facilitating the maintenance of beneficial bacteria, whilst reducing the amount of potentially harmful ones. Combined with predicted metagenomics data this suggests that even though the changes we observed were small they may have implications in terms of metabolic efficiency and/or resistance to pathogens.

Initial studies in NOD1 and NOD2 KO mice have shown that the mutation is linked to colitis and dysbiosis of the gut leading to conditions such as Crohns disease and Inflammatory Bowel Disease (IBD)[78,79]. However, studies which properly controlled for maternal effects showed no change to the microbiota in these mutants, and suggested that these and other immune deficiencies may not be sufficient to induce gut dysbiosis alone[80]. Therefore, it may be the case that γδ T cells operate also in this way and have little influence over the microbiota under conditions of homeostasis. Studies looking at the effects of the absence of γδ T cells in a chicken dysbiosis model may therefore prove to be an interesting line of study, and the results shown here do not necessarily rule out a role for more large scale γδ T cell interactions with the gut flora.

5.5 Conclusion

The work in this chapter was designed to test for the effect of microbiota status on the repertoire of γδ T cells and conversely the effect of γδ T cell deficiency on microbiota composition and therefore investigate the interplay between them. The data presented here presents a picture

of a lack of large scale interactions between the γδ T cells and the microbiota. Despite the lack of a positive result when looking at the TCR gamma repertoire of GF chickens, the apparent apathy of gut γδ T cells and the microbiota towards one another, is itself an interesting phenomena, considering the parallel development and constant intimate contact between the two. There were however small differences in public clones within the TCR gamma repertoire in the absence of a microbiota and a reduction in potentially beneficial members of the microbiota in the absence of γδ T cells. Despite no clear effect on γδ T repertoire, the microbiota is known to have a strong influence on the development and function of the immune system and although the data in this chapter suggests that the effects may be quite subtle, it is not clear what their implications may be in terms of productivity and susceptibility to disease. These findings also suggests that γδ T cell expansion and gut homing may be mostly determined entirely absent of the microbiota. Though this is somewhat surprising, it makes sense for a lymphocyte which is important in early immunity, particularly in γδ high animals, as they must be functional immediately following birth/hatch and do not have time to develop in response to the microbiota in a more adaptive manner. Data in this chapter also suggests a role for γδ T cells in facilitating the survival and persistence of potentially beneficial microbes in the caeca, which may be of interest to the chicken industry and may help guide future probiotic treatments.

5.6 References

1. Pan D, Yu Z. Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes. 2014;5(1):108–19.
2. Al Nabhani Z, Eberl G. Imprinting of the immune system by the microbiota early in life. Mucosal Immunol. 2020;13(2):183–9.
3. Yu Q, Jia A, Li Y, Bi Y, Liu G. Microbiota regulate the development and function of the immune cells. Int Rev Immunol. 2018;37(2):79–89.
4. Shanahan F. The host-microbe interface within the gut. Bailliere's Best Pract Res Clin Gastroenterol. 2002;16(6):915–31.
5. Cheled-Shoval SL, Gamage NSW, Amit-Romach E, Forder R, Marshal J, Van Kessel A, et al. Differences in intestinal mucin dynamics between germ-free and conventionally reared chickens after mannan-oligosaccharide supplementation. Poult Sci. 2014;93(3):636–44.
6. Forder REA, Howarth GS, Tivey DR, Hughes RJ. Bacterial modulation of small intestinal goblet cells and mucin composition during early posthatch development of poultry. Poult Sci.

2007;86(11):2396–403.

7. Kaspers B, Lettermann S, Roell S. DEVELOPMENT OF THE GUT ASSOCIATED IMMUNE SYSTEM. In: 20th European Symposium on Poultry Nutrition. 2015. p. 12–5.

8. Broom LJ, Kogut MH. The role of the gut microbiome in shaping the immune system of chickens. Vet Immunol Immunopathol. 2018;204(October):44–51.

9. Oakley BB, Kogut MH. Spatial and temporal changes in the broiler chicken cecal and fecal microbiomes and correlations of bacterial taxa with cytokine gene expression. Front Vet Sci. 2016;3(FEB):1–12.

10. Khairallah C, Chu TH, Sheridan BS. Tissue Adaptations of Memory and Tissue-Resident Gamma Delta T Cells. Front Immunol. 2018;9:1–22.

11. Komano H, Fujiura Y, Kawaguchi M, Matsumoto S, Hashimoto Y, Obana S, et al. Homeostatic regulation of intestinal epithelia by intraepithelial gamma delta T cells. Proc Natl Acad Sci U S A. 1995;92(13):6147–51.

12. Girardi M, Glusac E, Filler RB, Roberts SJ, Propperova I, Lewis J, et al. The distinct contributions of murine T cell receptor TCRγδ+ and TCRαβ+ T cells to different stages of chemically induced skin cancer. J Exp Med. 2003;198(5):747–55.

13. Roberts SJ, Smith AL, West BA, Wen L, Findly CR, Owen MJ, et al. T-cell αβ+ and γδ+ deficient mice display abnormal but distinct phenotypes toward a natural, widespread infection of the intestinal epithelium. Proc Natl Acad Sci. 1996;93:11774–9.

14. Umesaki Y, Setoyama H, Matsumoto S, Okada Y. Expansion of αβ T-cell receptor-bearing intestinal intraepithelial lymphocytes after microbial colonization in germ-free mice and its independence from thymus. Immunology. 1993;79:32–7.

15. Bandeira A, Mota-Santos T, Itohara S, Degermann S, Hausser C, Tonegawa S, et al. Localization of γ/δ T cells to the intestinal epithelium is independent of normal microbial colonization. J Exp Med. 1990;172(1):239–44.

16. Ismail AS, Behrendt CL, Hooper L V. Reciprocal Interactions between Commensal Bacteria and γδ Intraepithelial Lymphocytes during Mucosal Injury. J Immunol. 2009;182(5):3047–54.

17. Ravens S, Fichtner AS, Willers M, Torkornoo D, Pirr S, Schöning J, et al. Microbial exposure drives polyclonal expansion of innate γδ T cells immediately after birth. Proc Natl Acad Sci U S A. 2020;117(31):18649–60.

18. Ismail AS, Severson KM, Vaishnava S, Behrendt CL, Yu X, Benjamin JL, et al. Γδ Intraepithelial Lymphocytes Are Essential Mediators of Host-Microbial Homeostasis At the Intestinal Mucosal Surface. Proc Natl Acad Sci U S A. 2011;108(21):8743–8.

19. Hoytema van Konijnenburg DP, Reis BS, Pedicord VA, Farache J, Victora GD, Mucida D. Intestinal Epithelial and Intraepithelial T Cell Crosstalk Mediates a Dynamic Response to Infection. Cell. 2017;171(4):783-788.e13.

20. Meijerink N, Kers JG, Velkers FC, van Haarlem DA, Lamot DM, de Oliveira JE, et al. Early Life Inoculation With Adult-Derived Microbiota Accelerates Maturation of Intestinal Microbiota and Enhances NK Cell Activation in Broiler Chickens. Front Vet Sci. 2020;7:1–17.

21. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2017;19(4):554–65.

22. Bischof J, Ibrahim SM. bcRep: R package for comprehensive analysis of B cell receptor

repertoire data. PLoS One. 2016;11(8):1–15.

23. Lee J, Paparoditis P, Horton AP, Frühwirth A, McDaniel JR, Jung J, et al. Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations. Cell Host Microbe. 2019;25(3):367-376.e5.

24. Clarke EL, Connell AJ, Six E, Kadry NA, Abbas AA, Hwang Y, et al. T cell dynamics and response of the microbiota after gene therapy to treat X-linked severe combined immunodeficiency. Genome Med. 2018;10(1):1–14.

25. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8:1–6.

26. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics. 2015;31(17):2882–4.

27. Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome. 2018;6(1):1–12.

28. Tanaka M, Nakayama J. Development of the gut microbiota in infancy and its impact on health in later life. Allergol Int. 2017;66(4):515–22.

29. Prinz I, Silva-Santos B, Pennington DJ. Functional development of γδ T cells. Eur J Immunol. 2013;43(8):1988–94.

30. Dunon D, Coopervn MD, Imhof B a. Migration patterns of thymus-derived. 1993;2545–50.

31. Dunon D, Courtois D, Vainio O, Six A, Chen CH, Cooper MD, et al. Ontogeny of the immune system: γ/δ and α/β T cells migrate from thymus to the periphery in alternating waves. J Exp Med. 1997;186(7):977–88.

32. Berndt A, Pieper J, Methner U. Circulating γδ T cells in response to Salmonella enterica serovar enteritidis exposure in chickens. Infect Immun. 2006;74(7):3967–78.

33. Crhanova M, Hradecka H, Faldynova M, Matulova M, Havlickova H, Sisak F, et al. Immune response of chicken gut to natural colonization by gut microflora and to Salmonella enterica serovar enteritidis infection. Infect Immun. 2011;79(7):2755–63.

34. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.

35. Holderness J, Hedges JF, Ramstead A, Jutila MA. Comparative Biology of γδ T Cell Function in Humans, Mice, and Domestic Animals. Annu Rev Anim Biosci. 2013;1(1):99–124.

36. Kennedy EA, King KY, Baldridge MT. Mouse microbiota models: Comparing germ-free mice and antibiotics treatment as tools for modifying gut bacteria. Front Physiol. 2018;9:1–16.

37. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, Spencer S, et al. Compartmentalized Control of Skin. Science. 2012;337:1115–20.

38. Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. Nature. 2014;516(729):94–8.

39. Helgeland L, Vaage JT, Rolstad B, Halstensen TS, Midtvedt T, Brandtzaeg P. Regional phenotypic specialization of intraepithelial lymphocytes in the rat intestine does not depend microbial colonization. Scand J Immunol. 1997;46(4):349–57.

40. Helgeland L, Vaage JT, Rolstad B, Midtvedt T, Brandtzaeg P. Microbial colonization influences

composition and T-cell receptor Vβ repertoire of intraepithelial lymphocytes in rat intestine. Immunology. 1996;89(4):494–501.

41. Apajalahti J, Kettunen A, Graham H. Characteristics of the gastrointestinal microbial communities, with special reference to the chicken. Worlds Poult Sci J. 2004;60(2):223–32.

42. Han Z, Willer T, Li L, Pielsticker C, Rychlik I, Velge P, et al. Influence of the gut microbiota composition on Campylobacter jejuni colonization in chicken. Infect Immun. 2017;85(11):1–15.

43. Hayday AC. γδ T Cell Update: Adaptate Orchestrators of Immune Surveillance. J Immunol. 2019;203(2):311–20.

44. Xia M, Hesser DC, De P, Sakala IG, Spencer CT, Kirkwood JS, et al. A subset of protective γ9δ2 T cells is activated by novel mycobacterial glycolipid components. Infect Immun. 2016;84(9):2449–62.

45. Papadopoulou M, Dimova T, Shey M, Briel L, Veldtsman H, Khomba N, et al. Fetal public Vγ9Vδ2 T cells expand and gain potent cytotoxic functions early after birth. Proc Natl Acad Sci U S A. 2020;117(31):18638–48.

46. Blazquez J-L, Benyamine A, Pasero C, Olive D. New Insights Into the Regulation of γδ T Cells by BTN3A and Other BTN/BTNL in Tumor Immunity. Front Immunol. 2018;9:1–11.

47. Adams EJ, Gu S, Luoma AM. Human gamma delta T cells: Evolution and ligand recognition. Cell Immunol. 2015;296(1):31–40.

48. Salomonsen J, Chattaway JA, Chan ACY, Parker A, Huguet S, Marston DA, et al. Sequence of a Complete Chicken BG Haplotype Shows Dynamic Expansion and Contraction of Two Gene Lineages with Particular Expression Patterns. PLoS Genet. 2014;10(6).

49. Chen L, Fakiola M, Staines K, Butter C, Kaufman J. Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. Front Immunol. 2018;9:1–17.

50. Honjo K, Hagiwara T, Itoh K, Takahasi E, Hirota Y. Immunohistochemical analysis of tissue distribution of B and T cells in germfree and conventional chickens. J Vetinary Med Sci. 1993;6(43):1031–4.

51. Lee IK, Gu MJ, Ko KH, Bae S, Kim G, Jin GD, et al. Regulation of CD4+CD8-CD25+ and CD4+CD8+CD25+ T cells by gut microbiota in chicken. Sci Rep. 2018;8(1):1–11.

52. Díaz-Sánchez S, Perrotta AR, Rockafellow I, Alm EJ, Okimoto R, Hawken R, et al. Using fecal microbiota as biomarkers for predictions of performance in the selective breeding process of pedigree broiler breeders. PLoS One. 2019;14(5):e0216080.

53. Videnska P, Sedlar K, Lukac M, Faldynova M, Gerzova L, Cejkova D, et al. Succession and replacement of bacterial populations in the caecum of egg laying hens over their whole life. PLoS One. 2014;9(12):1–14.

54. Pieper J, Methner U, Berndt A, Jena D-. Characterization of avian gamma / delta T-cell subsets after Salmonella Typhimurium infection of chicks. 2010;(November).

55. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature. 2017;548(7665):43–51.

56. McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. Host Selection of Microbiota via Differential Adhesion. Cell Host Microbe. 2016;19(4):550–9.

57. Melo-Gonzalez F, Kammoun H, Evren E, Dutton EE, Papadopoulou M, Bradford BM, et al. Antigen-presenting ILC3 regulate T cell-dependent IgA responses to colonic mucosal bacteria. J Exp Med. 2019;216(4):728–42.

58. Pabst O, Slack E. IgA and the intestinal microbiota: the importance of being specific. Mucosal Immunol. 2020;13(1):12–21.

59. Pandit RJ, Hinsu AT, Patel N V., Koringa PG, Jakhesara SJ, Thakkar JR, et al. Microbial diversity and community composition of caecal microbiota in commercial and indigenous Indian chickens determined using 16s rDNA amplicon sequencing. Microbiome. 2018;6(1):1–13.

60. Glendinning L, Watson KA, Watson M. Development of the duodenal, ileal, jejunal and caecal microbiota in chickens. bioRxiv. 2019;1–11.

61. Brown DR, Bradbury JM. Bergey's Manual of Systematics of Archaea and Bacteria. Bergey's Manual of Systematics of Archaea and Bacteria. 2015.

62. Joblin KN, Naylor GE. The ruminal mycoplasmas : A review. J Appl Anim Res. 2002;21(2):161–79.

63. Shi D, Bai L, Qu Q, Zhou S, Yang M, Guo S, et al. Impact of gut microbiota structure in heat-stressed broilers. Poult Sci. 2019;98(6):2405–13.

64. Zhu N, Wang J, Yu L, Zhang Q, Chen K, Liu B. Modulation of growth performance and intestinal microbiota in chickens fed plant extracts or virginiamycin. Front Microbiol. 2019;10(JUN):1–16.

65. Correa-Fiz F, Blanco-Fuertes M, Navas MJ, Lacasta A, Bishop RP, Githaka N, et al. Comparative analysis of the fecal microbiota from different species of domesticated and wild suids. Sci Rep. 2019;9(1):1–15.

66. Beller A, Kruglov A, Durek P, von Goetze V, Hoffmann U, Maier R, et al. ANAEROPLASMA, A POTENTIAL ANTI-INFLAMMATORY PROBIOTIC FOR THE TREATMENT OF CHRONIC INTESTINAL INFLAMMATION. Ann Rheum Dis. 2019;78:45–6.

67. Beller A, Kruglov A, Durek P, von Goetze V, Werner K, Heinz GA, et al. Specific microbiota enhances intestinal IgA levels by inducing TGF-β in T follicular helper cells of Peyer's patches in mice. Eur J Immunol. 2020;50(6):783–94.

68. O'Callaghan A, van Sinderen D. Bifidobacteria and their role as members of the human gut microbiota. Front Microbiol. 2016;7.

69. Estrada A, Wilkie DC, Drew M. Administration of Bifidobacterium bifidum to chicken broilers reduces the number of carcass condemnations for cellulitis at the abattoir. J Appl Poult Res. 2001;10(4):329–34.

70. Baffoni L, Gaggìa F, Di Gioia D, Santini C, Mogna L, Biavati B. A Bifidobacterium-based synbiotic product to reduce the transmission of C. jejuni along the poultry food chain. Int J Food Microbiol. 2012;157(2):156–61.

71. Schweizer E, Hofmann J. Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems. Microbiol Mol Biol Rev. 2004;68(3):501–17.

72. Singh N, Gurav A, Sivaprakasam S, Brady E, Padia R, Shi H, et al. Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. Immunity. 2014;40(1):128–39.

73. Venegas DP, De La Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short

chain fatty acids (SCFAs) mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Front Immunol. 2019;10:1–16.

74. Häselbarth L, Ouwens DM, Teichweyde N, Hochrath K, Merches K, Esser C. The small chain fatty acid butyrate antagonizes the TCR-stimulation-induced metabolic shift in murine epidermal γδ T cells. EXCLI J. 2020;19:334–50.

75. Azcarate-Peril MA, Ritter AJ, Savaiano D, Monteagudo-Mera A, Anderson C, Magness ST, et al. Impact of short-chain galactooligosaccharides on the gut microbiome of lactose-intolerant individuals. Proc Natl Acad Sci U S A. 2017;114(3):E367–75.

76. Richards PJ, Flaujac Lafontaine GM, Connerton PL, Liang L, Asiani K, Fish NM, et al. Galacto-oligosaccharides modulate the juvenile gut microbiome and innate immunity to improve broiler chicken performance. bioRxiv. 2019;5(1):1–19.

77. Wang L, Liu Q, Hu J, Asenso J, Wise MJ, Wu X, et al. Structure and evolution of glycogen branching enzyme N-termini from bacteria. Front Microbiol. 2019;10:1–12.

78. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshiftmutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001;411:603–6.

79. Natividad JMM, Petit V, Huang X, De Palma G, Jury J, Sanz Y, et al. Commensal and probiotic bacteria influence intestinal barrier function and susceptibility to colitis in Nod1-/-;Nod2-/- Mice. Inflamm Bowel Dis. 2012;18(8):1434–46.

80. Robertson SJ, Yu Zhou J, Geddes K, Rubino SJ, Ho Cho J, Girardin SE, et al. Nod1 and Nod2 signaling does not alter the composition of intestinal bacterial communities at homeostasis. Gut Microbes. 2013;4(3).

Chapter 6

Effect of Campylobacter Colonisation on the Microbiota and γδ T Cells in the Chicken

6.1 Introduction

Campylobacter jejuni and *Campylobacter coli* are found in the intestine of the majority of commercial chickens, this represents the major source of infection for zoonotic transfer into humans. It was initially thought that *Campylobacter* is a harmless commensal in the chicken gut. As commensals exist in the gut in a symbiotic relationship with the host and thus confer various benefits under homeostatic conditions, the immune system generally avoids initiating responses against them[1]. However, more recent work has shown that *Campylobacter* is not always benign in the chicken gut and can in fact impact welfare and induce an enteric immune response[2]. Studies looking at experimental colonization of birds colonized by *Campylobacter* revealed increases in various cytokines in infected birds, including CXCLi2, IL1B, IFN-γ, IL17 and IL10[3–6]. The production of IL17 is indicative of a response aimed at extracellular pathogens and is associated with inflammatory conditions.

As the major producer of IL17 in the gut, multiple studies have suggested this type of response may be mediated by γδ T cells[3,6]. Further evidence for this hypothesis comes from the finding that *Campylobacter* has been shown to activate Toll-like Receptor (TLR) 2 and TLR21 in the chicken[7]. As chicken γδ T cells express TLR2[8] and 21[9] this may provide one mechanism for recognition of Campylobacter and subsequent IL17 expression[3]. Additionally, *Campylobacter* challenge following chemical bursectomy with cyclophosphamide showed no difference in caecal load compared to control birds[10]. As the bursa is the site of B cell development in chickens, this suggests antibody production in the gut plays a very limited role in responses against *Campylobacter* and therefore emphasizes the possible role of T cells in *Campylobacter* clearance.

As *Campylobacter* colonizes the chicken gut at bacterial loads of 10^9/g caecal contents, this could lead to changes in the microbiota profile. However, though other studies have reported differences in terms of composition, these differences were surprisingly minor[11,12]. As well as differences to composition, a study using a natural colonization model also reported an increase in

microbiome diversity in young birds due to *Campylobacter* colonization[13]. Additionally, various studies have reported differences to a range of genera including *Streptococcus, Blautia, Eschericia, Alistipes, Enterococcus, Bacteroides, Shigella, Gallibacterium, Lactobacillus, Corynebacterium, Ruminococcaceae, Enterobacter*[11,14]*,* though whether these bacteria increase or decrease in proportion is often unclear and can vary from study to study, presumably partially due to differing experimental design[11].

Though differences in these genera can be observed, due to the complexity of the system, it is often difficult to elucidate the precise nature of the interplay between microbes in the microbiome. Despite this, there has been some speculation in the literature as to how *Campylobacter* might be interacting with these bacteria. For instance, a correlation between levels of various *Clostridium* species and *Campylobacter* has been reported in various studies, and it has been speculated that this is due to cross feeding of organic acids produced by *Clostridium*[14–16]. Additionally, it has been proposed that *Campylobacter* may act as a hydrogen sink, enhancing *Clostridial* fermentation. Similarly, positive associations with *Faecalibacterium* have been previously reported in the literature. Though *Faecalibacterium* is a butyrate producer and thus would normally be considered a *Campylobacter* antagonist[17], it has been speculated that *Faecalibacterium* may also increase mucin production in chicken as it does in humans, thus interfering with the ability of organic acids to act as an antimicrobial against *Campylobacter*[11,18,19]*.*

6.2 Chapter Outline

This chapter seeks to examine the effects of *Campylobacter* colonization on the chicken gamma repertoire to further investigate the γδ T cell response which has been proposed in the literature. We also used *Campylobacter* as a natural stressor of both the enteric immune system and the microbiota in order to investigate the interactions between the two under non-homeostatic conditions. The 16S gut profile was first examined to determine the effect of *Campylobacter* colonisation on the microbiota. Although studies assessing the effect of *Campylobacter* colonisation

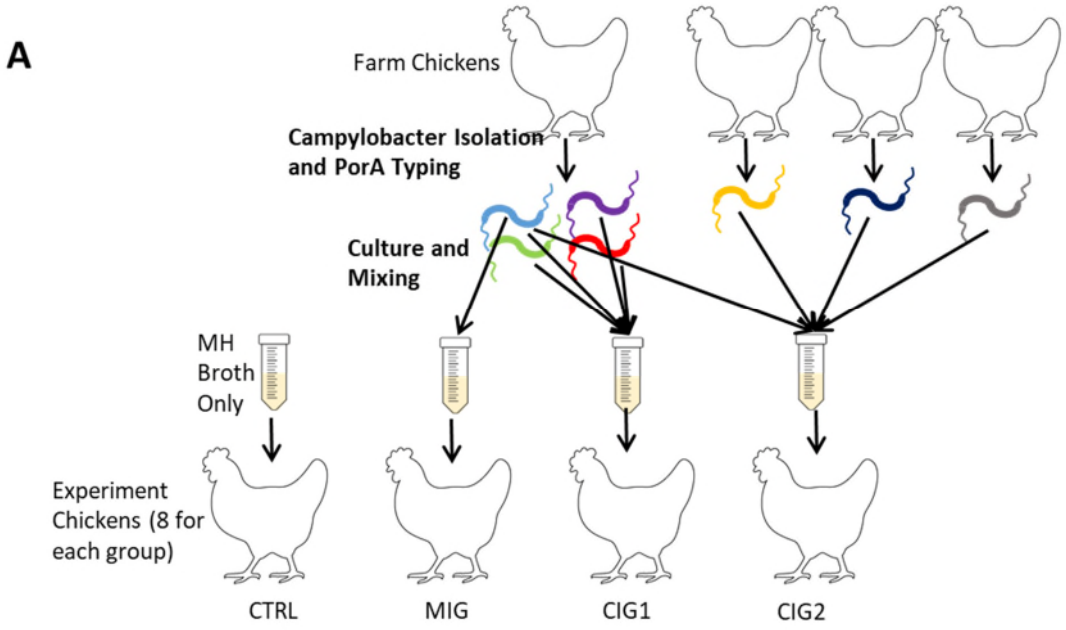
have been performed previously, the results are variable between studies and don't include multi-strain challenges, thus it was important to first determine the microbiome changes present in our birds. Crucially, we perform our experimental infections using both single and mixed strain cultures as is most realistic in a farm setting[20–24]. These *Campylobacter* strains were isolated from farm birds and therefore represent a real world infection scenario under 2 different conditions. Birds may become colonized by either strains from multiple different birds, which may not normally co-colonise, or strains from a single bird, which already co-exist. Therefore, in order to represent both of these possibilities, we perform these co-colonisation experiments using both strains that were isolated from the same bird, and strains isolated from different birds. This strategy enables us to not only examine whether or not co-infection with multiple strains makes a difference, but also examine the impact of strain composition and origin and thus fully explore the effects of *Campylobacter* status on the γδ T cell repertoire and microbiome structure.

6.3 Results

6.3.1 Experimental Design – PorA Type Confirmation and *Campylobacter* Strain Determination

The experiment consisted of four different groups of birds, one of which was an uninfected control. The other 3 groups were Mono-Infected Group (MIG), which was infected with a single strain, Co-Infection Group 1 (CIG1), which was infected with a mix of 4 different strains all isolated from the same farm bird and Co-Infection Group 2 (CIG2), which was infected with a mix of 4 different strains all isolated from different birds (Fig 6.1A). The proportions of each strain in the original donor birds is shown in Fig 6.1B as determined by colonies counted during initial isolation. These isolated strains were cultured and mixed in equal numbers for administration to the experimental groups described above. The experimental timeline is depicted in Fig 6.1C and was as follows. Shortly after hatch, chicks were dosed with Broilact, which is a chicken gut probiotic isolated from a single, healthy Specific Pathogen Free (SPF) hen. These birds were then infected with the

Campylobacter groups as above via oral gavage at Days 7 and 14 and faecal samples collected at Days 8, 10, 15, 17, 21, 28, 35, 42 and 49. At the end of the experimental period, the birds were sacrificed by cervical dislocation and had their Thymus, Spleen, Duodenum, Jejunum, Ileum and Caeca harvested and frozen down in RNAlater. The caecal contents from each bird was cultured on charcoal agar to confirm its *Campylobacter* status.



B

MIG			CIG1			CIG2		
Isolate	PorA	Percentage of Colonies in Original Bird (%)	Isolate	PorA	Percentage of Colonies in Original Bird (%)	Isolate	PorA	Percentage of Colonies in Original Bird (%)
784.18	PorA 400	52	784.11	PorA 11	33	784.18	PorA 400	52
-	-	-	784.12	PorA 44	2	5643.1	PorA 108	86
-	-	-	784.18	PorA 400	52	5814.1	PorA 36	97
-	-	-	784.20	PorA 33	5	6235.10	PorA 25	98

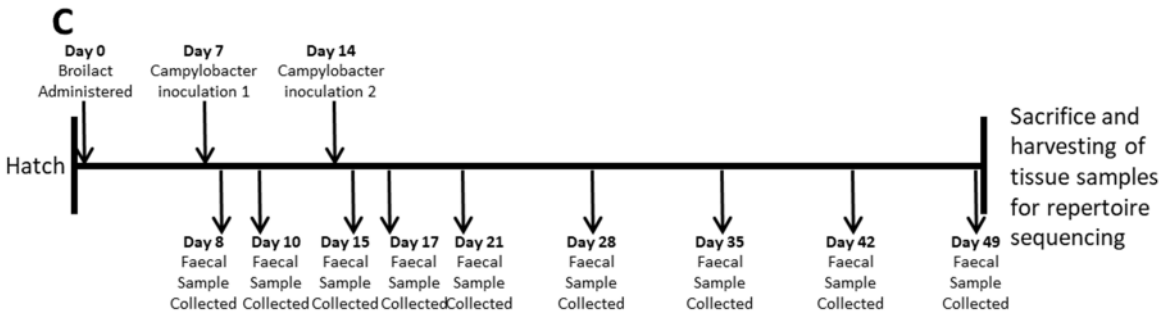


Figure 6.1 – Experimental Design

A) Schematic shows the experimental design and *Campylobacter* infection strategy. 4 Experimental groups were included in the study. For each of the infection groups, a total of $5x10^5$ CFU *Campylobacter* was used for gavage. CTRL was not infected with any *Campylobacter* and instead was given Mueller Hinton (MH) broth and acted as the control group for the experiment. The other 3 groups were infected with *Campylobacter.* The MIG group was infected with a single strain of *Campylobacter,* while CIG1 and CIG2 were both gavaged with mixed cultures of *Campylobacter*. For CIG1 all of these strains derived from a single bird and for CIG2 they were all from separate birds.

B) Table shows the Composition of the 2 Co-Infection groups. CIG1 was comprised of 4 strains all isolated from a bird referred to as 784. Conversely, CIG2 was mixed from 4 strains from different birds, including the isolate 784.18. This same isolate was also used to infect the MIG group.

C) A timeline of the experiment shows the treatments given to and samples taken from the birds. Upon hatch, the birds were treated with Broilact in an attempt to normalise the microbiota. The Birds were then infected with the previously described *Campylobacter* mixes at Day 7 post-hatch and faecal samples taken at Day 8 and 10 post-hatch. To ensure all of the birds within each group were properly colonised, the birds were infected a

second time at Day 14 post-hatch, then faecal samples taken at Days 15, 17, 21, 28, 35, 42 and 49 post-hatch. After these last samples were taken on Day 49 post-hatch, the birds were scarified and tissue samples taken for repertoire sequencing, some caecal contents was also plated to ensure the PorA types present were as expected and to gain an idea of the resident *Campylobacter* population structure.

In order to prepare the infection mixtures, we first had to ensure the isolates were assigned to the correct PorA type and determine whether they were *C. jejuni or C. coli*. After initial plating and isolation from the farm birds, each single colony was typed using PorA PCR and subsequent Sanger sequencing. However, as isolates were picked from plates often populated by hundreds or thousands of colonies, it would be easy to sample more than 1 colony, leading to a mixed culture. Additionally, initial mistyping of any of the isolates could lead to difficulties with later analysis. Therefore, the PorA type of each isolate was confirmed by an additional round of PorA PCR and Sanger sequencing after initial plating (Fig 6.2A). The Sanger traces showed that all isolates were of the expected PorA type and contained a single PorA sequence. Additional diagnostic PCRs were then performed to determine whether each isolate was *Campylobacter jejuni* or C. *coli,* since the two are indistinguishable in terms of colony appearance and by many common microbiological tests. The PCRs were designed based on the characteristic presence of hippurate and the absence of aspartokinase in *C. jejuni* and the reverse in *C. coli*[25]. Fig 6.2B depicts agarose gel images of PCRs that distinguish between *Campylobacter jejuni* and *coli* and all of the isolates chosen for this study were shown to be *C. jejuni* by the hippurate band at 344bp and lack of an aspartokinase band at 500bp. The presence of Campylobacter in the faeces of challenged birds was first confirmed by qPCR on faecal DNA and all challenged birds were shown to be PCR positive.

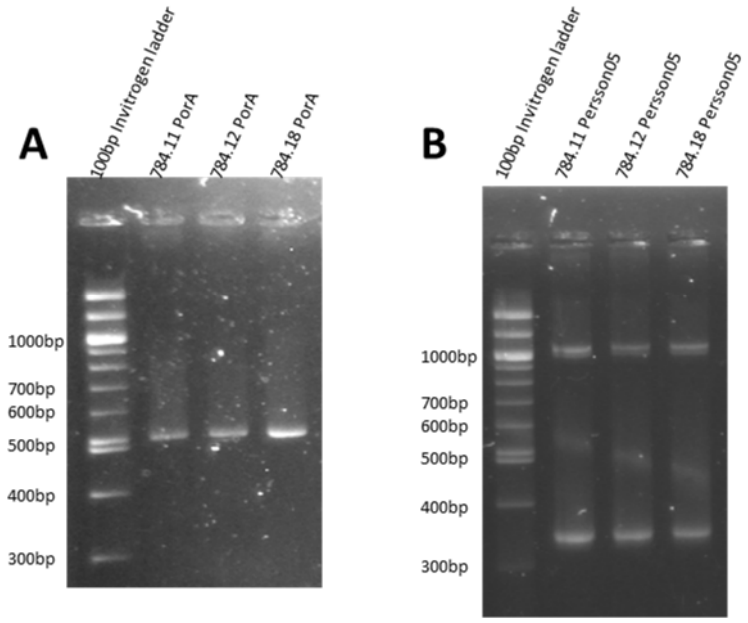


Figure 6.2 - PorA and *Campylobacter jejuni* vs *coli* PCRs
During the Campylobacter blending, genomic DNA was extracted from each of the plated strains. **A)** PorA PCR using the MompA primers and **B)** Campylobacter jejuni vs coli PCRs were performed using the Persson05 primers. The PorA PCR products were then sequenced in order to confirm their identity. The *C. jejuni* vs *C. coli* PCRs show a 16S positive control band at 1062bp and a hippurate band at 344bp showing that these strains are *Campylobacter jejuni*. This result was the same for each of the strains tested. Gels were run on 3% agarose at 130V for 1hr 45mins in an attempt to get separation of the PorA products which differ in size by a few base pairs. 100bp Invitrogen ladder was used and the sizes of the ladder bands is shown in base pairs along either size of the gel.

6.3.2 Experimental Termination and PorA Sequencing from Caecal Sweeps

As expected, the charcoal agar plates from caecal contents after termination of the experiment had multiple colonies from all of the groups, except the uncolonised CTRL birds. Sweeps of these plates were then taken using an L shaped spreader to harvest the cultured Campylobacter population, which were mixed into BHI broth and frozen in glycerols at -80°C. DNA was extracted from these samples and used as a template for PorA PCRs, which were then sequenced on an Illumina MiSeq. The reads were processed as with the 16S data described later in this chapter and analysed using Phyloseq[26]. CTRL bird 5 was removed from all downstream analysis because the sample had clearly been contaminated and it produced read counts and PorA sequencing data unlike any of the other birds from that group, despite having no colonies present on the plates.

The PorA profiles of each group were first examined using Bray-Curtis β-diversity plots showing that as expected, the four groups clustered totally distinctly (Fig 6.3A). Most of the birds in each of the four groups were tightly clustered in these plots, indicating high similarity. However 2 birds (CIG1 2 and CIG1 8) break away from the main CIG1 cluster but still do not overlap with any other group. In order to ascertain what PorA types were present in these caecal sweeps and what the proportion of each was, compositional stacked barplots were generated (Fig 6.3B).

The unchallenged CTRL group displayed a pattern of PorA types unlike the experimentally colonised groups, MIG, CIG1 and CIG2. They displayed even levels of the types Multiple PorA1, Multiple PorA3, Multiple PorA5, PorA 11, 'PorA1573, PorA1383', PorA25, PorA33, PorA400 and PorA44. The PorA types present in these sweeps appeared to originate from both the experimental innocula and from the environment, but as indicated above the plot, these read counts were an order of magnitude less than seen in any of the challenged groups. It should be noted that these birds were derived from a commercial hatchery and previous work[27] has demonstrated that all chickens in commercial environments carry very low numbers of unculturable *Campylobacter*, with diverse PorA profiles. This can therefore be considered a "natural" background level of *Campylobacter* present in the industry.

The MIG group had high levels of PorA 400, which is dominant in every bird as would be expected. However, there are also other contaminating PorA types present including, 108, 11, 33 and 44. In the CIG1 group, the majority of birds were dominated by PorA 33. There were also smaller amounts of PorAs 108, 400 and 44 present at in these animals. The exceptions to this is the birds CIG1 2 and CIG1 8, in which the dominant PorA types were 400 and 44, and 108 respectively. In both of these outlier birds however, PorA 33 is also present at a relatively high proportion (~25%). Very low levels of other PorAs were also detected that were not present in the inoculum but as noted above, these represent part of a natural profile of *Campylobacter* in commercial birds. Every bird in the CIG2 group was dominated by PorA 108 with smaller amounts (~3%) of 25, 33 and 400. Oddly,

PorA 36 is not present in any meaningful amounts. Bird CIG2 4 has relatively high levels of the contaminating PorA 1010. Additionally, PorA 33 was not in the mix given to the CIG2 birds, so there has evidently been transfer of PorA types between the infected groups.

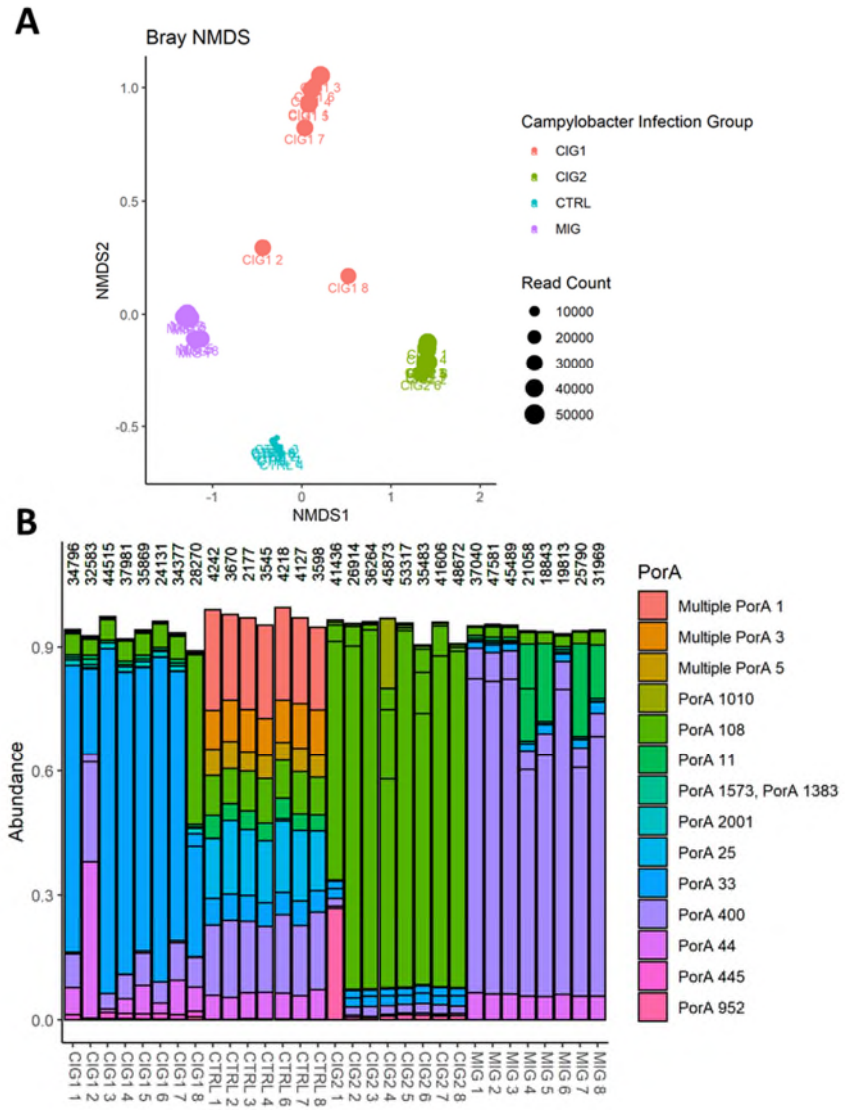


Figure 6.3 - PorA Sequencing from Caecal Sweeps.
A) After sacrifice of birds and extraction of caeca, the contents were spread on charcoal agar plates and incubated in microaerophilic conditions using microaer sachets and sealed boxes at 42°C for 48hrs. There were multiple Campylobacter colonies on each of the plates for every bird in every group except the CTRL birds. An L shaped spreader was used to mix the colonies into 500µl Mueller-Hinton broth, which was then used to make a 20% glycerol stock which was frozen down at -80°C. The stocks were later defrosted and DNA extracted from them. This DNA was amplified using the PorA primers and sequenced on an Illumina MiSeq. CTRL bird 5 was removed since it had a far higher read count and profile unlike any of the other CTRL birds. Bray-Curtis distances were then calculated and ordination plots generated using NMDS.
B) Compositional plots were then generated in order to display the proportion of each PorA type present in the caecal sweeps. The most common 20 PorA types are displayed on the compositional plot with the remaining types left out to allow easier plotting. Numbers of PorA reads from each caecal sweep is displayed along the top of the plot.

6.3.3 16S PCR, Sequencing and Data Processing using the DADA2 Pipeline

DNA was extracted from the faecal samples and diluted 100X in order to overcome the issue of faecal PCR inhibitors (See Chapter 3). Diluted faecal DNA was then amplified using barcoded 16S primers. The PCRs were performed in triplicate to mitigate the risk of PCR jackpotting and provide a way of confirming no mistakes were made during set up and potentially a means to correct them if they were (See Chapter 3) (Fig 6.4). For each PCR repeat, the barcodes were shifted 1x96 well plate column or 8 samples, such that the same sample was amplified using different barcoded primers across each of the triplicate repeats. The concentration of each PCR band was then estimated by comparison to the ladder and the volume to be added to the PCR pool adjusted accordingly, such that a roughly equal amount of all products was added. These products were then pooled, library prepped and sequenced on the Illumina MiSeq platform.

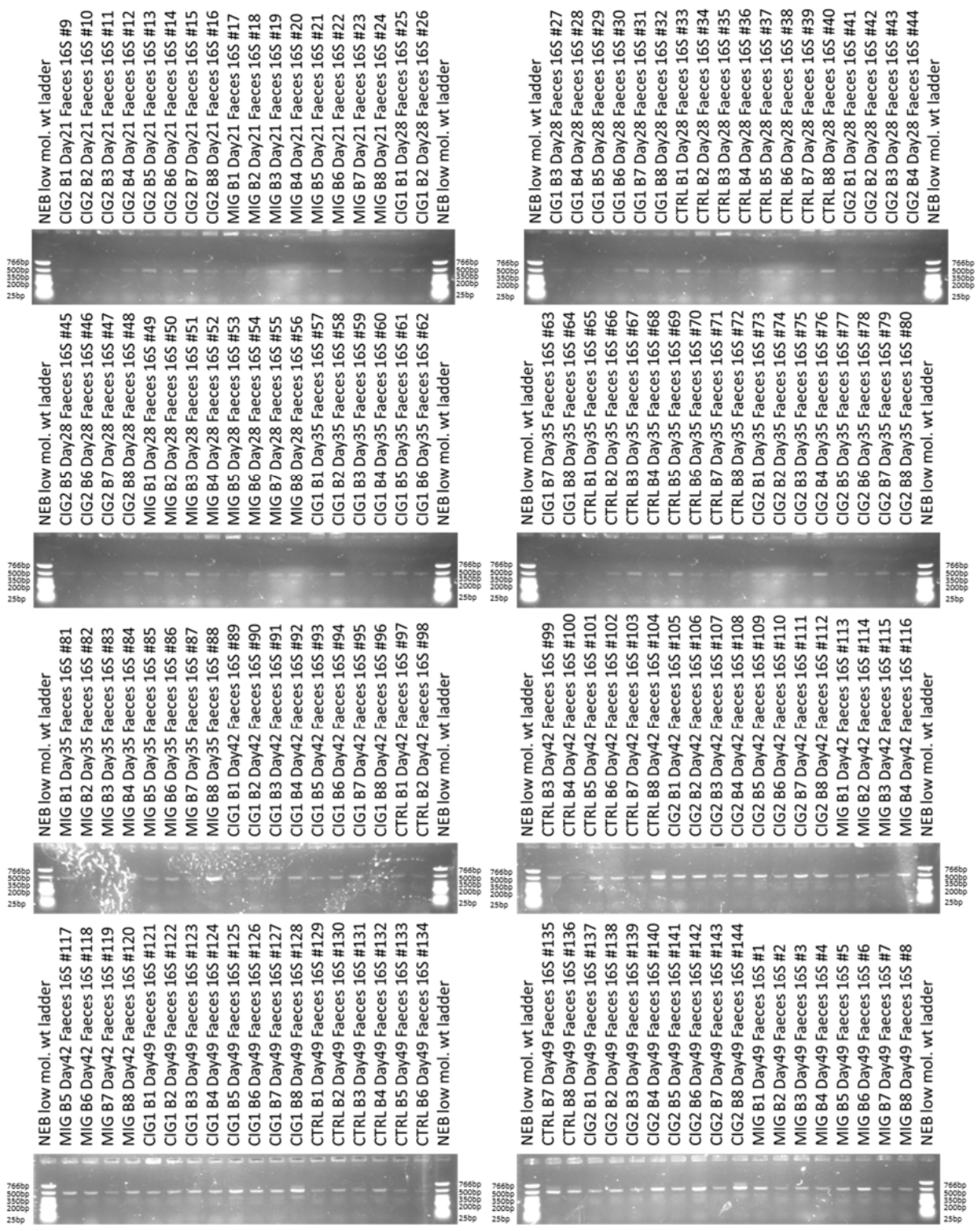


Figure 6.4 - Example of 16S PCRs before Pooling and Sequencing

DNA was extracted from chicken faecal samples and the 16S rRNA amplified using PCR. The Primers used were assigned a different barcode to allow downstream demultiplexing and attribution of sequence data to the sample of origin. PCRs show clear bands indicating amplification from the majority of samples, though some produce higher concentration bands than others. Expected products are around 500bp and product length was determined using an NEB low molecular weight ladder. Ladder band lengths are indicated at either side of the gels in base pairs.

The MiSeq run produced a total of 7,988,300 reads PF for the 16S data with an overall average Q30 of 78.49 and a %PF of 90.78%. The Illumina reads were then demultiplexed and separated according to their barcodes before being further processed in R using DADA2. DADA2 is a 16S analysis pipeline which enables processing of ribosomal RNA sequences without the construction of Operational Taxonomic Units (OTUs)[28]. OTUs are a commonly used unit of 16S studies and essentially work by grouping together sequences which are above a certain threshold similarity, generally 97%, and treating them as a single group in the analysis. This method is used as a crude way to overcome the issue of Illumina sequencing errors being interpreted as true biological variation. However, the obvious drawback of this is that when lumping data together without any phylogenetic reference to reduce the false positive rate, resolution is also lost. Even within closely related species bacterial variation can be high in terms of behaviour, lifestyle and gene expression[29–33]

Following the DADA2 pipeline, the read quality profiles were first inspected and the reads trimmed according to these profiles (Fig 6.5). The reverse reads were of lower quality than the forward, however this is a common property of Illumina sequencing and to be expected. Furthermore, it should be noted that reads can anneal to the flow cell in either orientation and hence the Illumina forward read could start from either end of the TCR gamma amplicons, therefore the drop-off in read quality does not mean reduced quality in any one part of the TCR gamma sequence. In this case, the reads were trimmed to 240 for the forward reads and 230 for the reverse reads. This trimming length represents a compromise between retaining enough length for downstream read merging and removing lower quality data, which improves the packages ability to pick out rare sequence variants.

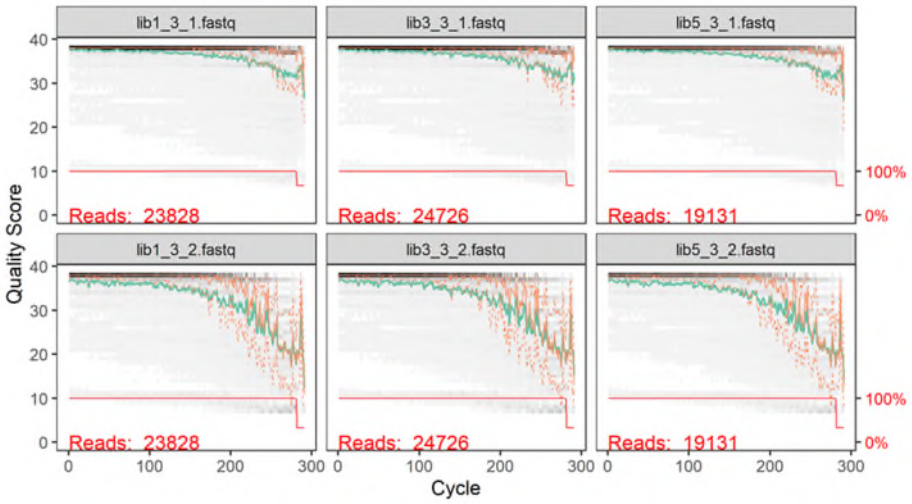


Figure 6.5 - Read Quality Profiles Highlight the need to Truncate as Sequence Quality Declines with Length
Example quality profile plots were produced using the DADA2 function 'plotQualityProfile'. These plots provide a visual representation of the relationship between increasing read length and the decline in read quality, as represented by Q score. As is a common trait of Illumina reads, this decline is accelerated in the reverse read relative to the forward. These plots inform the length the reads are to be truncated at as a compromise between retaining enough length for merging and losing as few reads as possible. The truncation lengths decided on were 240bp for the forward and 230bp for the reverse reads. The X axis is the number of Illumina parallel sequencing cycles and is equivalent to the read length. Quality scores are represented by the greyscale heatmap as well as the mean Q score by the green line, the median by the orange line and the 75th and 25th quartiles by the dashed orange lines. The red line shows the percentage of sequences which extend to that position, in our case the reads are all 300bp.

A plot was generated tracking the number of reads at each of the DADA2 processing stages (Fig 6.6). Briefly, reads are first trimmed to remove the most error prone portions. A model is then used to learn the sequencing run base call error rates, which is then used to distinguish true unique variants from sequences generated by errors. Forward and reverse reads are then merged and chimeric reads predicted and removed, resulting in the final Amplicon Sequence Variant (ASV) table. As with the repertoire data (See Chapter 4/5), the largest attrition occurs during the initial demultiplexing stage, where each read is searched for the desired 16S primers and to confirm expected E338 barcode sequence is present. This amounted to ~20% of the 16S reads per library. The second major loss of data occurred during the DADA2 filter and trimming stage where reads are trimmed to the appropriate size and filtered based on the number of expected errors, calculated based on Q score. An average of 10% of reads were lost from each library during this stage, but the

loss was 40% in Library 4, which failed during the first run and was therefore sequenced on a subsequent run, which produced lower quality reads. Almost no reads are lost at any other stage of the processing, showing that premature truncation and chimeric reads are not an issue in our dataset.

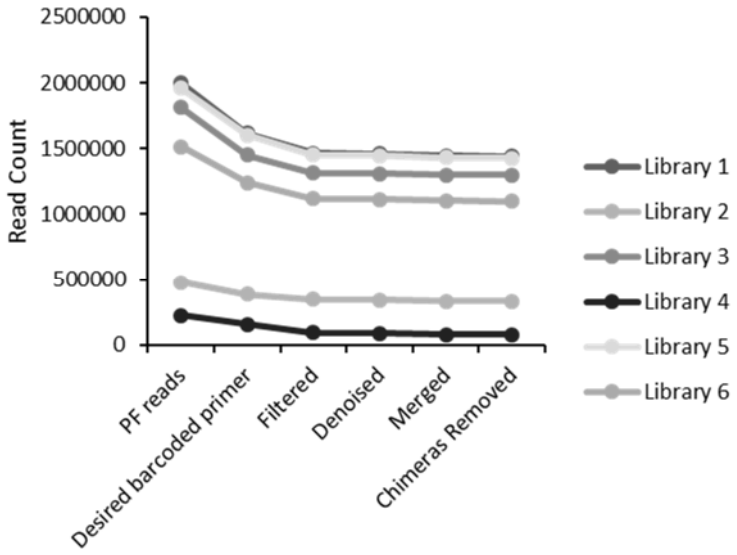


Figure 6.6 - The Majority of 16S Reads are Removed due to their Lack of Desired Barcodes/Primers
Plot shows the number of reads in each library during each stage of processing. The major data loss comes from the Desired barcoded primer stage, which involved searching the reads for the expected primers, confirming they have the desired barcodes and meet the barcode quality score threshold. The filtered stage refers to the trimming of reads to the previously determined lengths and filtering based on the number of predicted errors the reads contain. Denoised is the stage at which the error rate model is used to combine rarer sequences with the more common reads that most probably gave rise to them. Merged refers to the process of merging forward and reverse reads, thus any reads not long enough to overlap will be lost at this stage. Finally, Chimeras Removed is the elimination of any reads generated as a result of amplification errors where an erroneous sequence is generated from 2 true sequences during the PCR reaction.

6.3.4 Beta-Diversity using Bray-Curtis, UniFrac, Aitchison's and PhILR Distances Reveals Microbiome Differences between the *Campylobacter* Infection Groups

To gain a broad overview of whether any microbiota differences between the *Campylobacter* infection groups were present, they were first compared using β-diversity indices. To ensure that comparisons were robust, it was necessary to consider β-diversity using different metrics, as each has its own assumptions, strengths and weaknesses (See Chapter 3). Comparing the *Campylobacter* infection groups first by Bray-Curtis distance and plotting the resulting matrix by

Non-Metric Multidimensional Scaling (NMDS) ordination with 20 random starts, reveals relatively large differences between the groups at several of the time points (Fig 6.7A). Additionally, it should be noted that none of the observed differences were driven by *Campylobacter* itself and all of the displayed plots look virtually identical in its absence.

At Day 8 there are small hints of a visual difference, but a clear separation of the groups was not observed. At Day 10 however, these early potential differences become more apparent, with CTRL, MIG and CIG2 forming their own distinct clusters with some CIG2 birds mixed in with the other groups. Unfortunately, due to unequal dispersion, betadisper testing gives a significant result at this timepoint showing that the variances between groups are not equal and PERMANOVA testing is therefore invalid. The groups at day 15 maintain a degree of separation, however there is far more overlap between groups compared to the plots at day 10. Again MIG, CIG2 and to a certain extent CTRL are present as discrete clusters, but CIG2 seems far more variable and mixes in with the other groups. PERMANOVA testing gives a significant result but a low R^2 value of 0.2152 showing that *Campylobacter* infection group explains a small yet biologically noteworthy proportion of the variation in this plot. Despite this, none of the differences between individual pairs of groups proved to be significant, though the P value for CTRL/MIG was close with a value of 0.06. The day 17 plot features even clearer separation of the *Campylobacter* infection groups. CTRL and CIG1 in particular cluster apart from the other groups with limited overlap, whereas CIG2 and MIG are mixed in together, yet MIG is more tightly clustered and less variable. Comparing these groups using a PERMANOVA test reveals significant differences and individual paired comparisons show differences between CTRL/MIG and CTRL/CIG1. The day 21 plot shows similar levels of separation between the groups, though CTRL is more intermixed with MIG than at day 17. Unsurprisingly, betadisper reveals uneven clustering in this plot thus PERMANOVA is not valid.

At Day 28 there is good separation between CTRL, MIG and CIG2, but CIG1 is mixed in with all 4 of these groups. This plot gives a highly significant P value and a fairly high R^2 of 0.347. This

shows that there are large differences present in this plot between the groups, and that *Campylobacter* status is explaining a large portion of the variation in the plots. Paired testing of groups reveals that they were all highly significant, with the exception of MIG/CIG1. Interestingly, at day 35, while MIG remains clustered totally separate from the remaining infection groups, CTRL, CIG1 and CIG2 are far more intermixed on this plot. Despite this, PERMANOVA testing still gives a significant P value and an R^2 of 0.390. Furthermore, when conducting further PERMANOVA testing to examine the differences between paired groups, significant differences were found between CTRL/MIG, CTRL/CIG1, MIG/CIG1 and MIG/CIG2. The day 42 plot shows much more intermixing of the groups but despite this still gives a significant overall PERMANOVA P value of 0.002 and an R^2 of 0.26265. Further pairwise testing of the *Campylobacter* infection groups showed a significant difference between CIG1/CIG2 only. Similarly, day 49 shows more intermixing particularly with CIG1 and CIG2, but still some visual separation in CTRL and MIG. Despite this, differences between the samples did not reach significance.

The UniFrac plots looked very similar to those calculated using Bray-Curtis distance, showing that the previously observed differences persist when taking into account phylogenetic relatedness between 16S sequences (Fig 6.7B). Thus, the observed differences are most likely in more distantly related organisms and therefore represent real biological changes and not just stochastic variation or sequence artefacts. There are however a few notable exceptions where the Bray-Curtis and UniFrac plots do differ. The day 35 plots show similar results to those from Bray-Curtis, however CTRL 6 and CIG2 2 were such obvious outliers when compared using the UniFrac metric, they obscured the results for the rest of the birds and thus had to be removed. Subsequent PERMANOVA testing gives an R^2 of 0.655, far higher than in the Bray-Curtis plots and the highest observed in this study. Coupled with a highly significant P value of 0.001 this shows that at Day 35 *Campylobacter* infection group is in fact the major factor determining microbiota composition when the outlier birds are removed. Additionally, the UniFrac distance plots give slightly different results during paired statistical testing of the later time points. At day 35, the UniFrac PERMANOVA test for CTRL/CIG1

was not significant, though it was very close with a P value of 0.061. At day 42, CTRL/CIG1 and MIG/CIG2 were significant by UniFrac, but were not in the Bray-Curtis plots.

It has recently been argued that NGS data is compositional as there are a limited number of slots on an Illumina flow cell and an increase in one sequence will therefore lead to a decrease in another (See Chapter 3). Therefore, we also compared the groups using Aitchison's distance, which accounts for the compositional nature of NGS data (Fig 6.7C). Comparing the groups at day 8 based on Aitchison distance, CTRL, MIG and CIG1 are all relatively evenly intermixed, whereas CIG2 clusters apart from the others with some overlap. PERMANOVA testing gives a significant P value and an R^2 of 0.18255. Subsequent paired testing showed significant differences between CTRL/MIG, CTRL/CIG2 and CIG1/CIG2. At Day 10 however, there is far clearer visual separation of the groups, with MIG and CIG2 clustering together and separately from CTRL and CIG1, which also cluster together, similarly to in the Bray Curtis and UniFrac plots. However, as with Bray-Curtis and UniFrac, the differences in variation between the groups gives a significant betadisper test and thus renders PERMANOVA invalid. Day 15 also shows this developing separation, with CTRL clustering mostly separate from CIG1 and MIG, which are intermixed and CIG1, which is more independently clustered. PERMANOVA testing gives a significant result and paired testing showed significant differences in every one of the combinations of groups. Plotting by Aitchison distance at Day 17 reveals that CIG1 clusters separately from the other groups. CTRL also shows relatively good separation compared to MIG and CIG2, which are more intermixed. This is similar to the Bray-Curtis and UniFrac results but with more distinct clustering of groups. PERMANOVA shows a significant overall difference and again differences in every combination of pairs of groups. The Day 21 plot shows very good separation of the groups with very little overlap between the clusters. However, due to the tight clustering about the centroid within MIG and resulting lack of variation, the betadisper test is significant and thus PERMANOVA testing is invalid.

At day 28, again there is far more intermixing of the strains. MIG and CIG2 are indistinguishable from one another but do cluster mostly separate from CTRL. CIG1 is far more variable and has 2 outlier points. PERMANOVA testing at this time point is rendered invalid by the uneven spread, driven predominantly by CIG1. At Day 35, CIG2 and MIG remain similarly indistinguishable, whereas both CTRL and CIG1 cluster independently with virtually no overlap. Despite the presence of the outlier birds, CTRL 6 and CIG2 2, betadisper testing returns a non-significant result. Subsequent PERMANOVA testing reveals that these differences give a significant P value and subsequent paired testing of groups shows significant differences between every pair except CTRL/CIG2. The day 42 plot shows far less distinct clustering and though there is some separation between the groups it is less clear. Much like the plots at day 28, MIG and CIG2 tend to cluster together with CTRL and CIG1 being more distinct. PERMANVOVA testing gives a significant result and all of the paired tests were also significant. Day 49 shows much the same patterns of clustering as day 42 though with less overlap between CTRL and CIG1. In a marked departure from the results of Bray-Curtis and UniFrac plots, PEMANOVA testing gives a significant result and upon examination of the differences between pairs, all combinations have highly significant P values

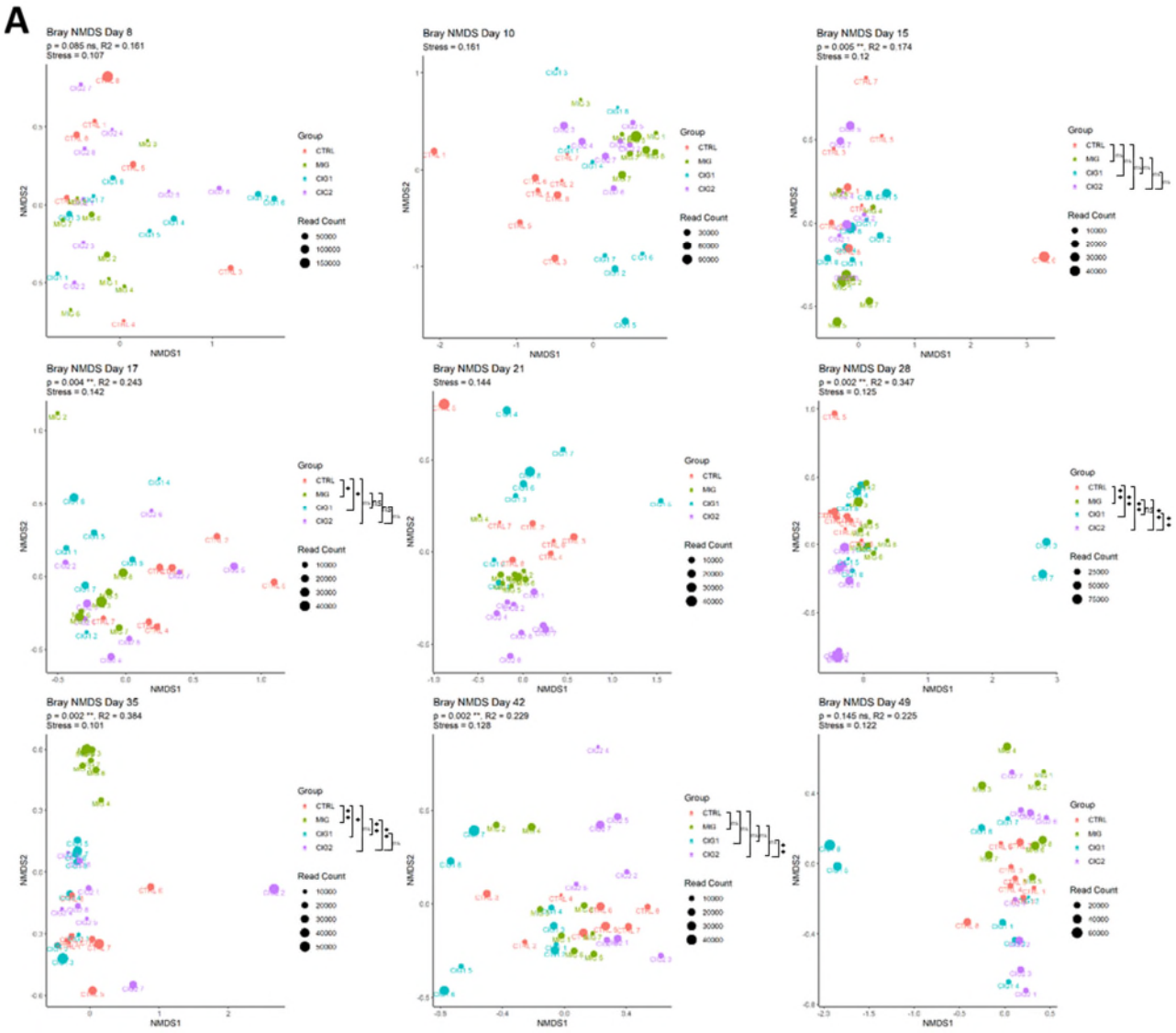
Similarly to Bray-Curtis and UniFrac, the PhILR ordination plots, which account for relatedness, closely match those for Aitchison distance, reinforcing the fact that the differences observed are not in closely related strains, and thus are more likely represent a true biological difference (Fig 6.7D). At day 8, MIG and CIG2 are far more intermixed than in the Aitchison's distance plots, however, PERMANOVA testing still gives a significant P value. Furthermore, testing of each paired combination of groups in isolation also show significant results for the same pairs as Aitchison's distance. Overlooking the orientation, the Day 10 PhILR plot looks almost identical to the Aitchison plot, however the slight reduction in variation in the CIG1 strain resulted in a non-significant betadisper test and therefore PERMANOVA testing was valid. The overall differences at day 10 were shown to be significant and paired testing of groups revealed that the PhILR distances between CTRL/MIG and CTRL/CIG2 were also significant. The Day 15 plot also shows some

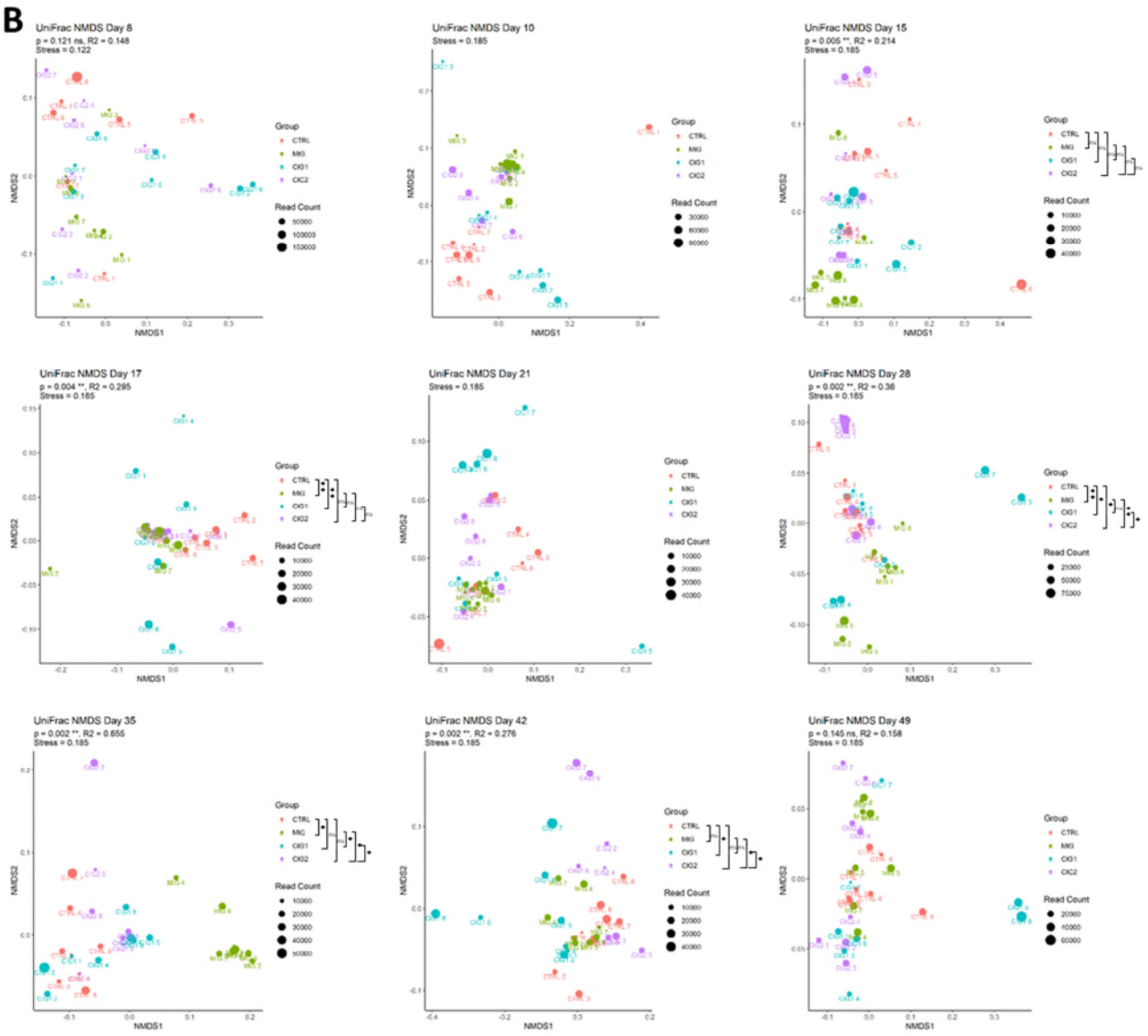
differences with far less distinct clustering between the groups. Despite the overlaps, PERMANOVA testing still gives a significant P value, and all of the paired tests were also significant, with the exception of CTRL/CIG2 and MIG/CIG2. Day 17 looks very similar to the Aitchison's distance plot when compared with PhILR and gives similar P and R^2 values when compared using PERMANOVA. Furthermore, all of the paired group comparisons were also shown to be significant. Unfortunately, the day 21 plot gives significant results with Betadisper testing and therefore cannot be tested with PERMANOVA. Despite this, the groups cluster almost totally separately indicating compositional differences exist between them.

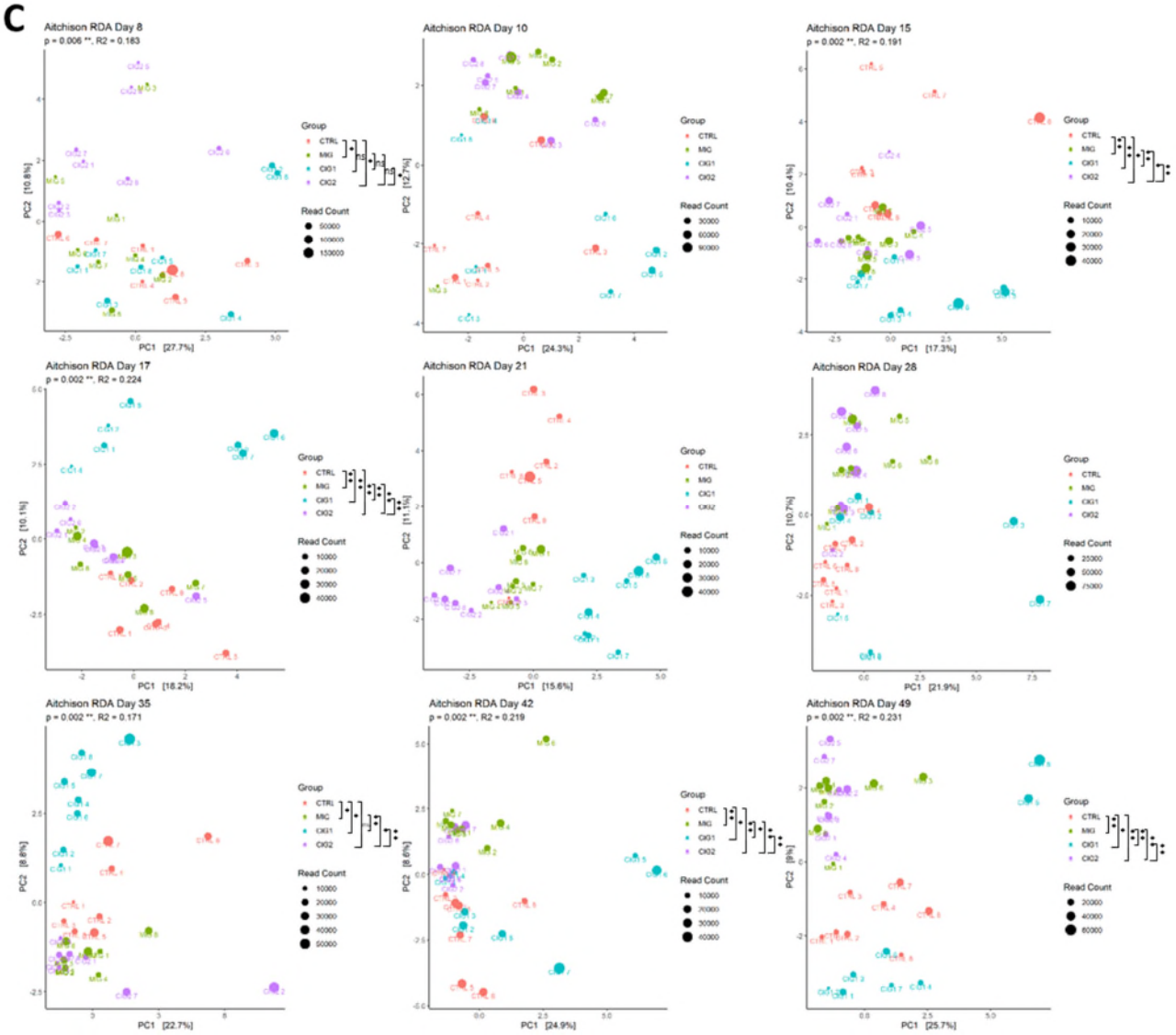
Day 28 on the other hand showed homogeneity of variance and therefore could be tested with PERMANOVA. The groups again cluster relatively distinctly and PERMANOVA testing gave a significant overall result as well as significant P values for all of the paired combinations of groups. At Day 35, the visual separation observed when comparing the groups using Aitchison's distance is greatly diminished when using PhILR. Though there is still some separate clustering, the groups are far less distinct. Nonetheless, PERMANOVA testing still gives a significant result and all of the paired comparisons were also significant with the one exception of MIG/CIG1. The Day 42 plot looks relatively similar to the Aitchison's distance plots, giving a similarly significant P value and similar R^2 with all of the paired comparisons also proving to be significant. Finally, the day 49 PhILR plot again looks mostly similar to the Aitchison's distance plot, but the CTRL strain clusters further apart from CIG1 with some intermixing with MIG and CIG2. However, this plot still gives a significant result using PERMANOVA as well as all of the paired comparisons also showing significant P values upon subsequent testing.

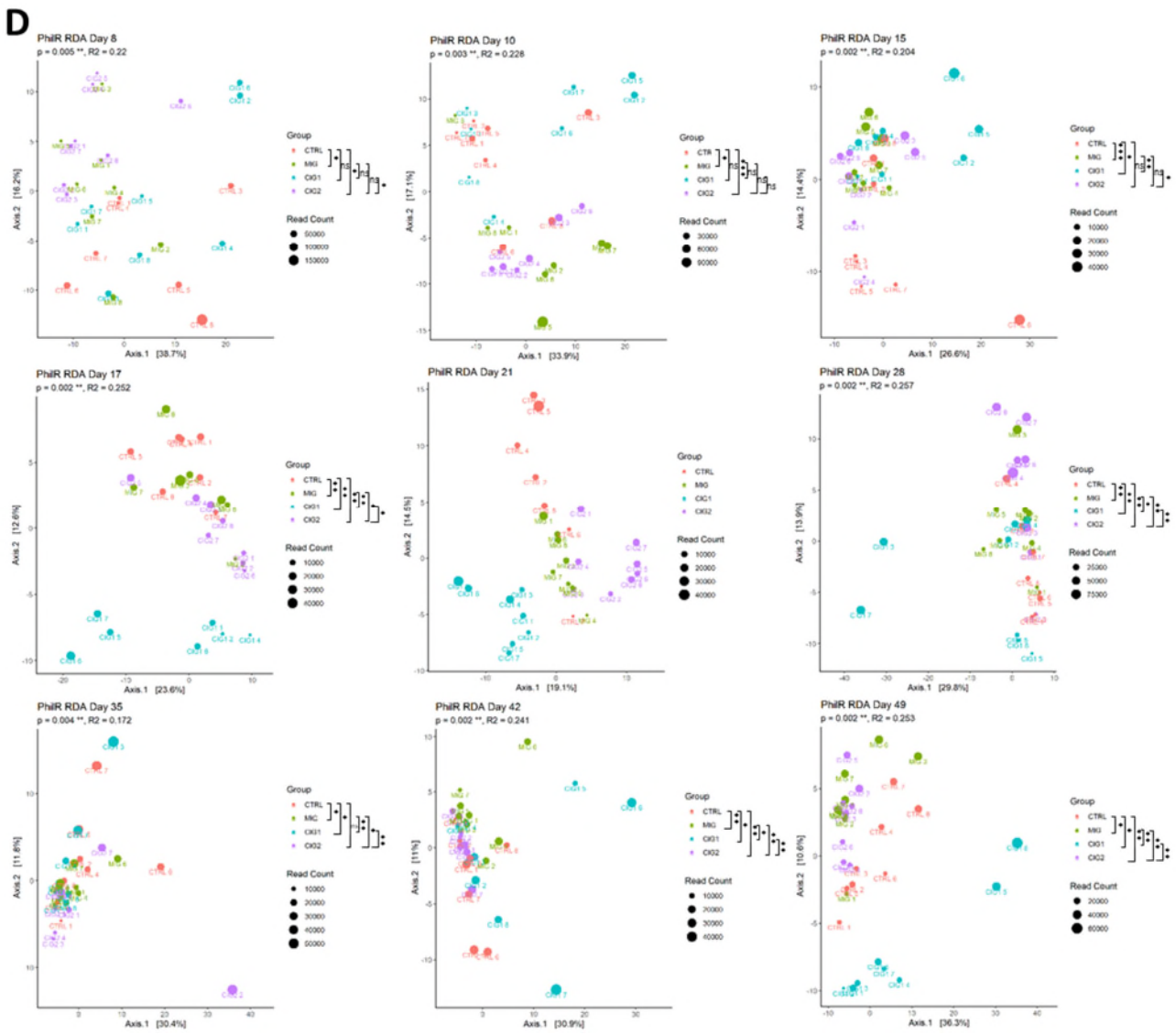
In summary then, every time point showed significant differences between *Campylobacter* infection groups in some or all of the distance metrics used, with the single exception of day 21, which showed significant differences in dispersion by every distance metric and therefore was not suitable for PERMANOVA testing. The tests between pairs of groups are fully summarised in (Fig

6.7E) and show consistent differences by the following pairs at the following days: Day 17 CTRL/MIG, CTRL/CIG1, Day 35 CTRL/MIG, MIG/CIG1, MIG/CIG2, Day 42 CIG1/CIG2. Additionally, a range of other paired comparisons were significant by one or more of the distance measures used. Therefore, we observed overall differences between the groups at nearly all of the time points, with differences between pairs of groups by one metric or another in many of the comparisons, indicating that *Campylobacter* infection and more specifically, *Campylobacter* strain composition affects the microbiota.


A
Bray NMDS Day 8
p = 0.085 ns, R2 = 0.161
Stress = 0.107
Bray NMDS Day 10
Stress = 0.161
Bray NMDS Day 15
p = 0.005 **, R2 = 0.174
Stress = 0.12
Bray NMDS Day 17
p = 0.004 **, R2 = 0.243
Stress = 0.142
Bray NMDS Day 21
Stress = 0.144
Bray NMDS Day 28
p = 0.002 **, R2 = 0.347
Stress = 0.125
Bray NMDS Day 35
p = 0.002 **, R2 = 0.384
Stress = 0.101
Bray NMDS Day 42
p = 0.002 **, R2 = 0.229
Stress = 0.128
Bray NMDS Day 49
p = 0.145 ns, R2 = 0.225
Stress = 0.122
NMDS1
NMDS2
Group
CTRL
MIG
CIG1
CIG2
Read Count


B
UniFrac NMDS Day 8
p = 0.121 ns, R2 = 0.148
Stress = 0.122
UniFrac NMDS Day 10
Stress = 0.185
UniFrac NMDS Day 15
p = 0.005 **, R2 = 0.214
Stress = 0.185
UniFrac NMDS Day 17
p = 0.004 **, R2 = 0.295
Stress = 0.185
UniFrac NMDS Day 21
Stress = 0.185
UniFrac NMDS Day 28
p = 0.002 **, R2 = 0.36
Stress = 0.185
UniFrac NMDS Day 35
p = 0.002 **, R2 = 0.655
Stress = 0.185
UniFrac NMDS Day 42
p = 0.002 **, R2 = 0.276
Stress = 0.185
UniFrac NMDS Day 49
p = 0.145 ns, R2 = 0.158
Stress = 0.185
NMDS1
NMDS2
Group
CTRL
MIG
CIG1
CIC2
Read Count


C
Aitchison RDA Day 8
p = 0.006 **, R2 = 0.183
Aitchison RDA Day 10
Aitchison RDA Day 15
p = 0.002 **, R2 = 0.191
Aitchison RDA Day 17
p = 0.002 **, R2 = 0.224
Aitchison RDA Day 21
Aitchison RDA Day 28
Aitchison RDA Day 35
p = 0.002 **, R2 = 0.171
Aitchison RDA Day 42
p = 0.002 **, R2 = 0.219
Aitchison RDA Day 49
p = 0.002 **, R2 = 0.231
Group
CTRL
MIG
CIG1
CIG2
Read Count
PC1
PC2


D
PhilR RDA Day 8
p = 0.005 **, R2 = 0.22
PhilR RDA Day 10
p = 0.003 **, R2 = 0.228
PhilR RDA Day 15
p = 0.002 **, R2 = 0.204
PhilR RDA Day 17
p = 0.002 **, R2 = 0.252
PhilR RDA Day 21
PhilR RDA Day 28
p = 0.002 **, R2 = 0.257
PhilR RDA Day 35
p = 0.004 **, R2 = 0.172
PhilR RDA Day 42
p = 0.002 **, R2 = 0.241
PhilR RDA Day 49
p = 0.002 **, R2 = 0.253
Axis.1 [38.7%]
Axis.2 [16.2%]
Axis.1 [33.9%]
Axis.2 [17.1%]
Axis.1 [26.6%]
Axis.2 [14.4%]
Axis.1 [23.6%]
Axis.2 [12.6%]
Axis.1 [19.1%]
Axis.2 [14.5%]
Axis.1 [29.8%]
Axis.2 [13.9%]
Axis.1 [30.4%]
Axis.2 [11.6%]
Axis.1 [30.9%]
Axis.2 [11%]
Axis.1 [36.3%]
Axis.2 [10.6%]
Group
CTRL
MIG
CIG1
CIG2
Read Count

E

	Bray-Curtis					UniFrac					Aitchison's Distance					PhILR				
8		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	*	ns	*	CTRL	-	*	ns	*
	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	*	-	*	ns	MIG	*	-	ns	ns
	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	ns	*	-	ns	CIG1	ns	ns	-	*
	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	*	ns	ns	-	CIG2	*	ns	*	-
10		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	*	ns	***
	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	*	-	ns	ns
	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	ns	ns	-	ns
	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	***	ns	ns	-
15		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	ns	ns	ns	CTRL	-	ns	ns	ns	CTRL	-	**	**	*	CTRL	-	***	*	ns
	MIG	ns	-	ns	ns	MIG	ns	-	ns	ns	MIG	**	-	**	*	MIG	***	-	**	ns
	CIG1	ns	ns	-	ns	CIG1	ns	ns	-	ns	CIG1	**	**	-	**	CIG1	*	**	-	*
	CIG2	ns	ns	ns	-	CIG2	ns	ns	ns	-	CIG2	*	*	**	-	CIG2	ns	ns	*	-
17		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	*	*	ns	CTRL	-	**	**	ns	CTRL	-	**	**	**	CTRL	-	**	**	**
	MIG	*	-	ns	ns	MIG	**	-	ns	ns	MIG	**	-	**	**	MIG	**	-	**	*
	CIG1	*	ns	-	ns	CIG1	**	ns	-	ns	CIG1	**	**	-	**	CIG1	**	**	-	*
	CIG2	ns	ns	ns	-	CIG2	ns	ns	ns	-	CIG2	**	**	**	-	CIG2	**	*	*	-
21		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA
	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA
	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA
	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-
28		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	**	**	**	CTRL	-	**	*	*	CTRL	-	NA	NA	NA	CTRL	-	**	**	**
	MIG	**	-	ns	**	MIG	**	-	ns	**	MIG	NA	-	NA	NA	MIG	**	-	*	**
	CIG1	**	ns	-	**	CIG1	*	ns	-	*	CIG1	NA	NA	-	NA	CIG1	**	*	-	**
	CIG2	**	**	**	-	CIG2	*	**	*	-	CIG2	NA	NA	NA	-	CIG2	**	**	**	-
35		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	*	ns	ns	CTRL	-	*	ns	ns	CTRL	-	*	*	ns	CTRL	-	*	*	ns
	MIG	*	-	**	**	MIG	*	-	*	*	MIG	*	-	**	*	MIG	*	-	**	*
	CIG1	ns	**	-	ns	CIG1	ns	*	-	*	CIG1	*	**	-	**	CIG1	*	**	-	**
	CIG2	ns	**	ns	-	CIG2	ns	*	*	-	CIG2	ns	*	**	-	CIG2	ns	*	**	-
42		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	ns	ns	ns	CTRL	-	ns	*	ns	CTRL	-	**	*	**	CTRL	-	**	*	**
	MIG	ns	-	ns	ns	MIG	ns	-	ns	*	MIG	**	-	*	**	MIG	**	-	*	**
	CIG1	ns	ns	-	**	CIG1	*	ns	-	*	CIG1	*	*	-	**	CIG1	*	*	-	**
	CIG2	ns	ns	**	-	CIG2	ns	*	*	-	CIG2	**	**	**	-	CIG2	**	**	**	-
49		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2		CTRL	MIG	CIG1	CIG2
	CTRL	-	NA	NA	NA	CTRL	-	NA	NA	NA	CTRL	-	**	*	**	CTRL	-	*	*	**
	MIG	NA	-	NA	NA	MIG	NA	-	NA	NA	MIG	**	-	**	**	MIG	*	-	**	**
	CIG1	NA	NA	-	NA	CIG1	NA	NA	-	NA	CIG1	*	**	-	**	CIG1	*	**	-	**
	CIG2	NA	NA	NA	-	CIG2	NA	NA	NA	-	CIG2	**	**	**	-	CIG2	**	**	**	-

Figure 6.7 - Bray-Curtis, UniFrac, Aitchison and PhILR plots show Differences between CTRL and Campylobacter Infected Groups.

β-Diversity plots using **A)** Bray-Curtis, **B)** UniFrac, **C)** Aitchison and **D)** PhILR distances were plotted. Bray-Curtis and UniFrac distances were calculated based on proportional data treating each sequence as an individual. Phylogenetic trees were calculated using alignments from MAFFT, followed by computation of pairwise distances, production of an initial tree using Neighbor joining, optimisation of model parameters and production of the final tree using maximum likelihood. Centre log ratio (clr) transforms of count data and Euclidean distances were used to calculate the Aitchison's distance and PhilR metrics. The number of reads associated with each sample is represented by the size of the point. PERMANOVA tests were conducted in each plot using the Adonis function from the R package vegan with 999 permutations. Where the distances tested non-significant for dispersion using betadisper testing, overall FDR corrected P and R2 values from PERMANOVA testing are displayed in the subtitle and individual paired FDR corrected P values are indicated in the figure legend. The significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. Stress values were also indicated in the subtitle for each of the plots using NMDS and percentage of variation explained was shown in the axes for RDA.

E) Table showing all of the PERMANOVA tests and their outcomes. Overall tests including all groups are represented by colour where a red fill indicates a non-specific result, and a green fill indicates p value < 0.05. Those that were not tested due to uneven dispersion are unfilled. Further tests between pairs of groups are indicated within the filled cells and the significance thresholds for these were as above.

6.3.5 Taxonomic Plots Reveal Significant Differences between the *Campylobacter* Infected Groups, Mainly at the Order Level

To investigate the differences observed in the β-diversity plots, the 16S sequences were assigned taxonomic classifications using the Ribosomal Database Project (RDP). The patterns of microbiome composition at the phyla level in the unchallenged birds indicate a pattern of ecological succession (Fig 6.8) similar to that which has been reported previously[34,35] (See Chapter 5). Though the microbiota composition shows some variation at the individual bird level, there are relatively consistent patterns of development over time. At day 8, the microbiota was mostly made up of the phylum *Firmicutes*, with a significant contribution from *Proteobacteria*, which vary but make up on average ~10% of the microbiota. Over time, the proportion of *Proteobacteria* decrease and the *Firmicutes* become yet more dominant.

At the Phylum level, the microbiota compositions between unchallenged and the three *Campylobacter* challenged groups were quite similar, though there were some note-worthy differences (Fig 6.8A). Individuals within CIG1 tended to have spikes of *Bacteroidetes* at the expense of *Firmicutes* occurring throughout the study. These spikes were present in different CIG1 birds at days 10, 17, 21, 28, 42 and 49, though they were also present in other groups to a lesser extent, such as CTRL at Days 15, 35 and 49 and CIG2 at Days 17 and 35. Interestingly, the *Bacteroidetes* spikes were never observed in any of the MIG birds. Additionally, CTRL showed an increase in the minority Phyla *Actinobacteria* at day 21. Finally, there were also some differences in terms of the proportion of *Proteobacteria*, with increases in MIG and CIG1 at Day 28 and increases in CIG2 at Days 35 and 42.

Increasing the taxonomic resolution down to the Order level revealed more profound differences between the *Campylobacter* infection groups, and helped to identify the contributors to the differences in the β-diversity plots (Fig 6.8B). As with the Phylum level plots, the microbiotas vary widely within each group and there is a lot of fluctuation between time points. Despite this, some clear differences between the groups were identified in these plots. As the majority of differences

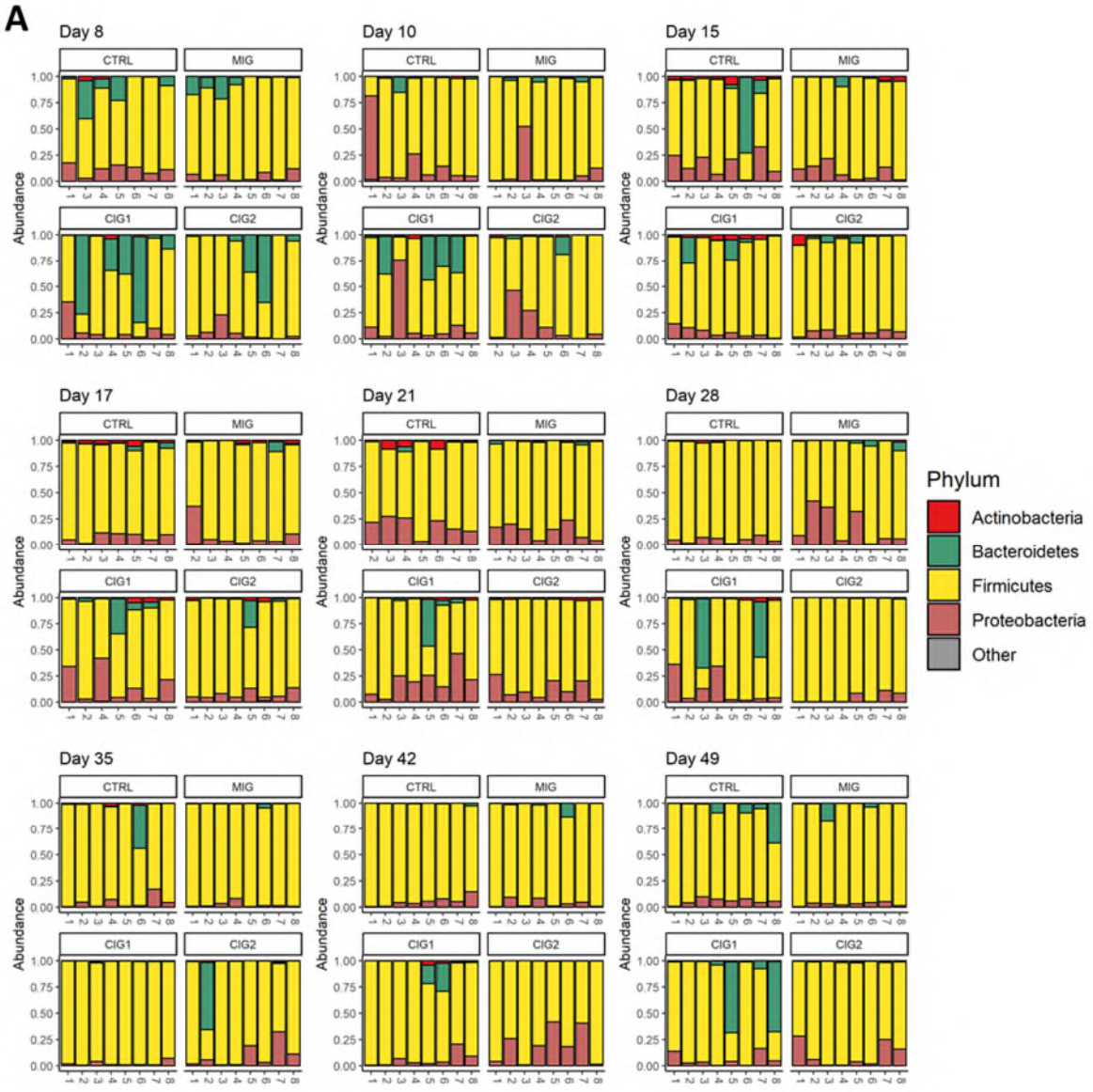
were observed in the Orders, *Lactobacillales,* and *Clostridales* and at the later timepoints *Erysipelotrichales,* unpaired Wilcoxon rank tests were performed on these taxa, comparing each paired combination of the groups at all ages at which they were present. At day 8, though there were no large differences between any of the groups, CIG1 had a higher proportion of *Bacteroidales,* which were substantially increased in some birds at the expense of *Clostridales*. At Day 10, the differences between the groups became more apparent. CTRL had a large increase in *Lactobacillales*, which are elevated in proportion in every bird except one. Comparing the proportion in CTRL to the other groups using an unpaired Wilcoxon signed rank test reveals that these differences are significant when CTRL is compared with MIG and CIG1, but not CIG2. The decrease in *Lactobacillales* was associated with an increase in *Clostridales* in each of the groups, particularly MIG, again giving significant results when comparing CTRL/MIG, as well as CTRL/CIG2 and MIG/CIG1. Additionally, some of the birds within the CIG1 group had increased proportions of *Bacteroidales,* though these increases were not consistent and appeared only in half of the birds in this group. Curiously, at Day 15, the gap between the CTRL and the *Campylobacter* infected groups narrowed again as the proportions of *Lactobacillales* were lower in the unchallenged CTRL birds. Despite this, *Lactobacillales* remained higher in CTRL than CIG1 or MIG, both of which again showed an accompanying increase in *Clostridales*. This resulted in a significant Wilcoxon test P value for proportion of *Clostridales* at this time point for CTRL/MIG only. At Day 17, the differences observed at earlier time points (day 10) appeared to return. *Lactobacillales* were substantially increased in almost every CTRL bird compared to the Day 15 levels and were present at a higher proportion than in CIG1, MIG and to a lesser extent CIG2. Again using an unpaired Wilcoxon signed rank test, this difference gave a significant P value for CTRL/MIG and CTRL/CIG1 but not CTRL/CIG2. The increased *Lactobacillaes* in CTRL birds was again associated with a decrease in *Clostridales* though there were no significant differences in this Order. Interestingly, the proportion of *Campylobacterales*, of which *Campylobacter* is the main genus (Fig 6.9B), peaked in some birds in CIG1 at this point and was present at exceptionally high levels (up to 0.4 of total microbiota). Again however, at day 21 these

differences between the groups were no longer evident as the proportion of *Lactobacillales* in CTRL birds was at levels more similar to those observed in the other groups.

The faecal samples from the day 28 CTRL birds showed higher levels of *Lactobacillales,* again giving significant P values when comparing CTRL/MIG, CTRL/CIG1 and CTRL/CIG2. Additionally at this age, the Order *Erysipelotrichales* was first observed at high proportions and was present at particularly high levels in CIG2 and to a lesser extent in some CIG1 birds. Wilcoxon signed rank testing confirmed this difference was significant when comparing CTRL/CIG1, MIG/CIG1 and CIG1/CIG2. However, it should be noted that the proportion of *Erysipelotrichales* increased over time in all strains and therefore this difference could simply be due to this occurring earlier in the CIG2 challenged group. The *Enterobacteriales* were also higher in some birds within MIG and CIG1.

At Day 35, the differences between groups due to *Lactobacillales* mostly disappeared as they decreased in proportion in the CTRL group again, though there were still significant differences when comparing MIG/CIG1. Additionally, the *Erysipelotrichales* were increased in all groups to the same levels as were present in CIG2 at day 28, with the exception of MIG. This results in significant differences when comparing CTRL/MIG and MIG/CIG1. The differences between MIG and CIG2 however, did not reach the 0.05 significance threshold. Furthermore, this increase in the *Erysipelotrichales* was also accompanied by a decrease in the *Clostridales* in every group except MIG, giving similarly significant differences for CTRL/MIG, CTRL/CIG1, CTRL/CIG2. At Day 42, the differences between the groups observed at the earlier time points once again diminish to the point where all four look relatively similar. Despite this, there was an increase in *Enterobacteriales* in the majority of birds in the CIG2 group and a corresponding decrease in *Clostridales.* Though the increase in *Enterobacteriales* was not significant for any of the compared groups, the decrease in *Clostridales* was, giving significant differences for MIG/CIG2 and CIG1/CIG2. Additionally at this age, the *Erysipelotrichales* increase in the MIG group to similar levels observed in the other groups. At

day 49, the groups looked mostly similar except for the decrease in *Clostridales* in CIG1, though this was only significant when compared with MIG.



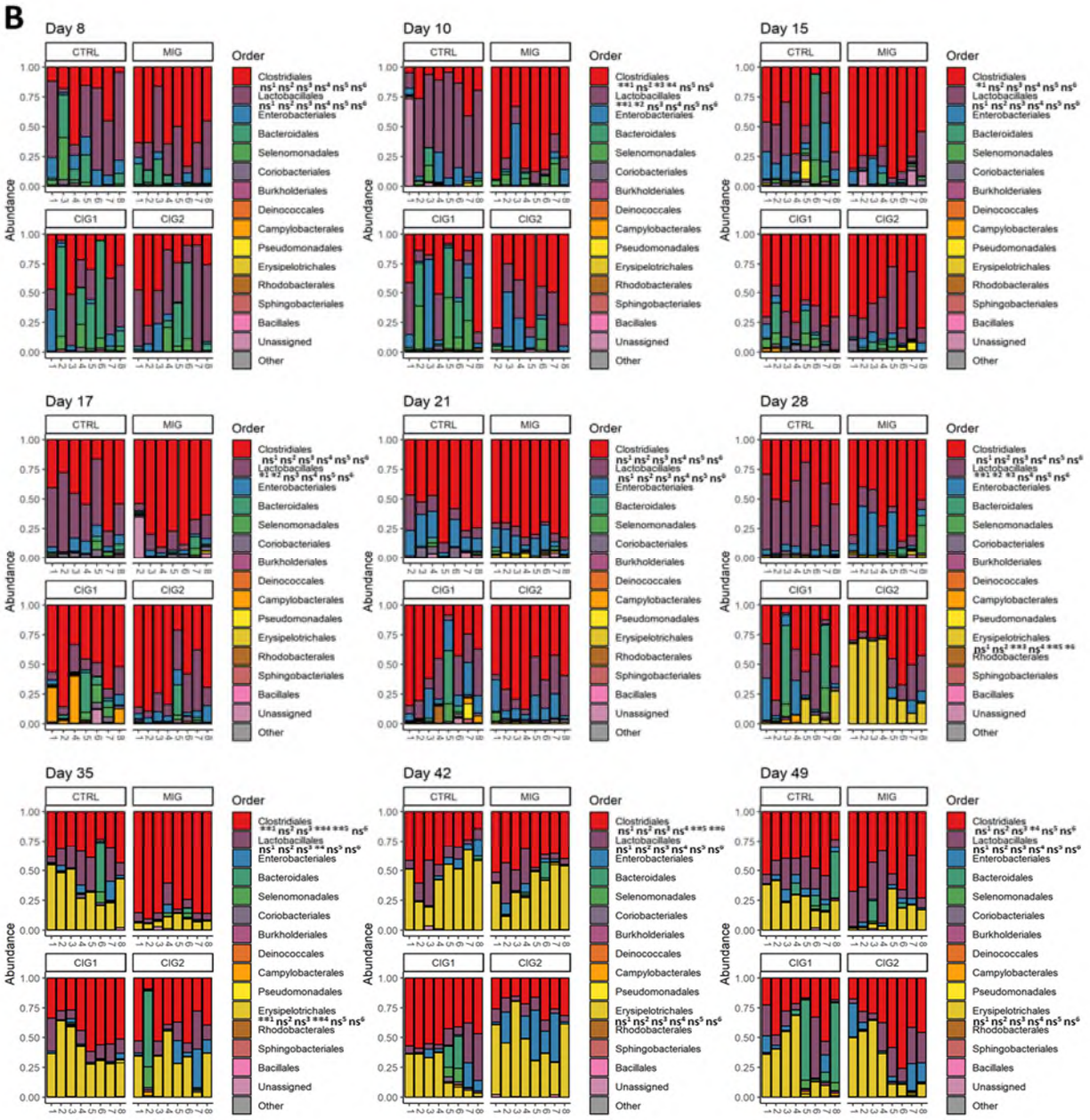


Figure 6.8 - Compositional Plots for each Campylobacter Infected Group show Increased Lactobacillus in CTRL

Each 16S sequence from the faecal samples was assigned a taxonomic group at each level using the RDP 16S database. The top 50 sequences by proportion in all samples were selected and the remaining sequences were assigned to the "Other" category. This enables easier interpretation of plots as otherwise, particularly at the Order level and below, the number of different taxonomic groups becomes overwhelming. Any sequences not assigned to a group at any given taxonomic level were assigned to the "Unassigned" category. These plots show the proportion of each taxonomic group at the **A)** Phylum and **B)** Order levels. Plots are split by day post-hatch, showing the composition of all birds in a group. Abundance on the Y axis refers to the proportion of sequences assigned to each taxonomic group. Statistical testing was performed using FDR corrected unpaired Wilcoxon signed rank tests. The significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05 and indicated below each Order being tested. The numbers correspond to which paired combination of groups is being tested as follows: 1 CTRL/MIG, 2 CTRL/CIG1, 3 CTRL/CIG2, 4 MIG/CIG1, 5 MIG/CIG2, 6 CIG1/CIG2.

6.3.6 Genus Plots Reveal Further Differences between the *Campylobacter* Infection Groups

SIMPER analysis was used to determine the biggest contributors to differences between the groups at each day and a range of genera were selected for further analysis based on these results (Fig. 6.9A). The taxonomic details of each selected genera are shown in Fig 6.9B for ease of interpretation. The main driver of differences between the CTRL group and the *Campylobacter* infected groups was the genera *Lactobacillus,* which made up the majority of the Order *Lactobacillaceae,* which represented the biggest difference between CTRL and MIG, CIG1 and CIG2. Though its levels fluctuated day to day, the biggest differences were observed at days 10, 17 and 28 where the proportion of *Lactobacillus* was hugely increased in CTRL compared to any of the other groups. Remarkably, at days 8, 10 and 17, *Lactobacillus* alone made up ~30, ~50% and ~30% of the CTRL microbiome respectively. Wilcoxon tests revealed that at these differences were significant at day 8 for CTRL/MIG, CTRL/CIG1 and MIG/CIG2, at day 17 between CTRL/MIG, CTRL/CIG1 and at day 28 between CTRL/CIG2, MIG/CIG2 and CIG1/CIG2.

It was noted that a number of 16S sequences within the *Lactobacillus* genus tended to vary more when comparing CTRL to any of the *Campylobacter* infected groups. These 16S sequences of interest were therefore searched against the EzBioCloud database to allow species level assignment. EzBioCloud has been shown to be best 16S database for genus and species assignment but has not yet been incorporated into a high throughput 16S assignment pipeline, so is only suitable for use on a small number of sequences[36,37]. The majority of the proportion of *Lactobacillus* was made up of sequences which were assigned to *Lactobacillus salivarius,* but the main driving force behind the differences between CTRL and the three *Campylobacter* infected groups was found to be *Lactobacillus johnsonii*. *L. johnsonii* levels were barely detectable at any time point in the *Campylobacter* challenged groups, except for Day 49 CIG2, where it is present at similar levels to CTRL. Despite its rarity in the *Campylobacter* challenged groups, *L. johnsonii* contributed ~12, 9 and

8% of the CTRL microbiota at Days 10, 17 and 28 and ranges between 1-3% at the other time points. These differences were significant when comparing; CTRL/CIG2 at Day 8, CTRL/MIG, CTRL/CIG1 and CTRL/CIG2 at Day 10, CTRL/CIG1 at Day 15, CTRL/MIG, CTRL/CIG1 and CTRL/CIG2 at Day 17, CTRL/MIG, CTRL/CIG1 and CTRL/CIG2 at Day 28, CTRL/MIG, CTRL/CIG1 and CTRL/CIG2 at Day 35, CTRL/MIG, CTRL/CIG1 and CTRL/CIG2 at Day 42 and CTRL/MIG, CTRL/CIG1, MIG/CIG2, and CIG1/CIG2 at Day 49.

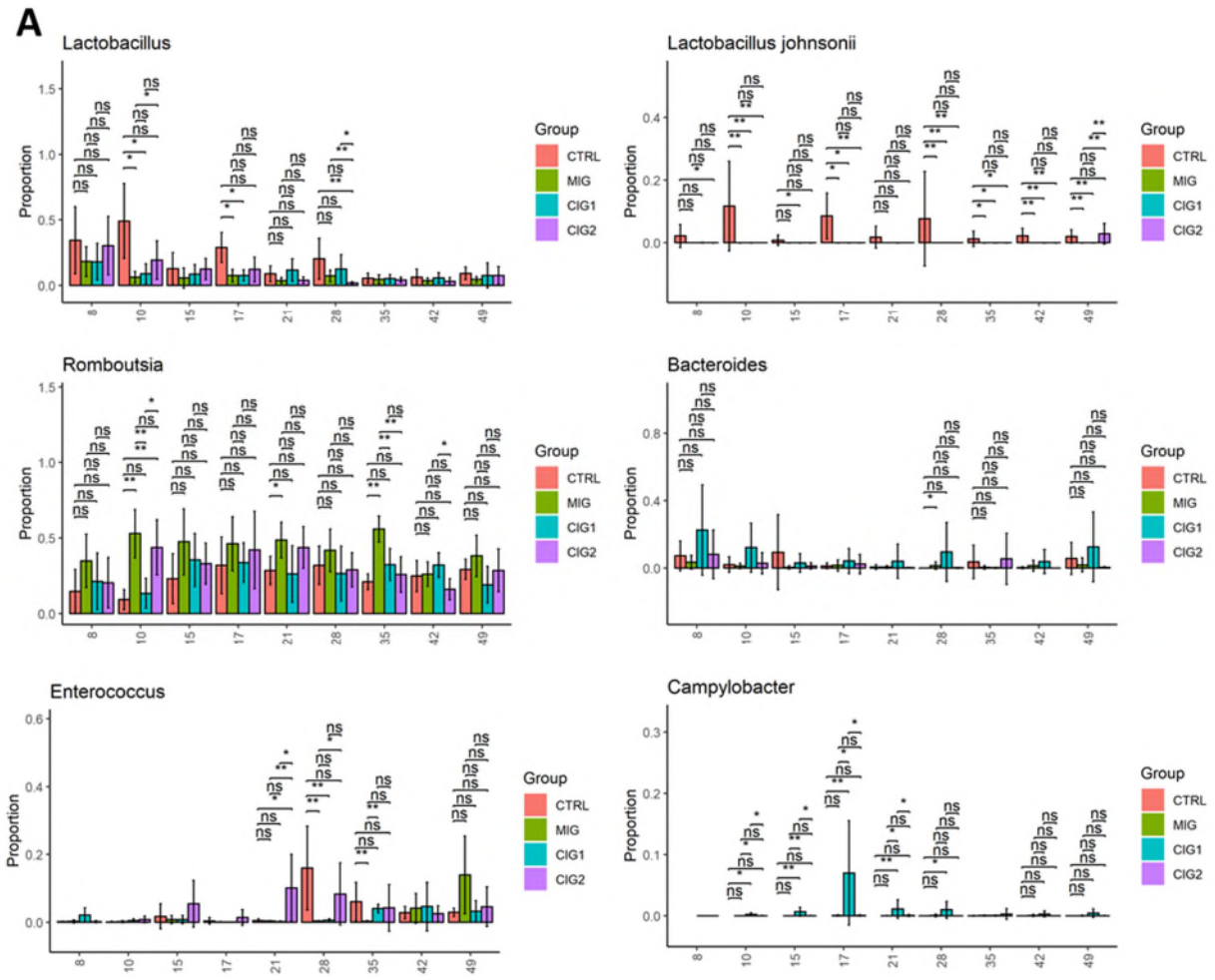
Another striking finding from the genus level microbiome analysis was the increase in *Romboutisa* in the MIG group. *Romboutsia* belongs to the Order *Clostridia* and the Phyla *Firmicutes* and is responsible for the differences we observed in the Order barplots (Fig 6.8B). Unlike *Lactobacillus*, levels of *Romboutsia* remained relatively consistent, particularly in the MIG group, which had the highest levels at every time point except day 42. CIG2 birds tended to have higher levels of *Romboutsia,* more in line with those found in MIG, at days 10, 17 and 21. These differences between MIG and the other strains are particularly profound at days 8, 10, 15, 21 and 35. FDR corrected Wilcoxon tests gave significant P values for CTRL/MIG, CTRL/CIG2, MIG/CIG1 and CIG1/CIG2 at day 10, CTRL/MIG at day 21, CTRL/MIG, MIG/CIG1 and MIG/CIG2 at day 35 and CIG1/CIG2 at day 42.

The microbiotas of the CIG1 birds contained substantially more *Bacteroides* on average than the other groups at days 8, 10, 21, 28, 42 and 49 and this genera made up ~20% of the microbiota in these birds at day 8. Though the proportion was elevated in CIG1 and thus it was highlighted by the SIMPER analysis, the levels in CIG1 were highly variable between birds. Due to this variability, unsurprisingly the differences do not reach the 0.05 significance threshold at any of the time points when compared to any other group, with the one exception of CTRL/MIG at Day 28.

Enterococcus was also highlighted by the SIMPER analysis and is present at far higher proportions in CIG2 at the earlier timepoints, particularly days 15, 21 and 28. The proportion also peaks in the unchallenged CTRL group at day 28 where it accounts for ~16% of the microbiota. This

level then drops at day 35 in both CTRL and CIG2 and continues to decline through the rest of the study. *Enterococcus* in MIG, though previously almost non-existent, increases to a ~15% peak at day 49. These differences are significant when comparing CTRL/CIG2, MIG/CIG2 and CIG1/CIG2 at Day 21, CTRL/MIG, CTRL/CIG1 and MIG/CIG2 at Day 28 and CTRL/MIG and MIG/CIG1 at Day 35.

Interestingly, one of the genera highlighted by the SIMPER analysis was *Campylobacter* itself, which was present at the highest levels by far in CIG1 at every day except days 8 and 35. Levels of *Campylobacter* peaked at day 17 where surprisingly it made up ~8% of the microbiota composition on average. It should be noted however that this mean was driven up by the exceptionally high levels found in 3 of the birds as observed in the compositional plots (Fig 6.8B). Shockingly, in one of the birds, *Campylobacter* alone made up almost 50% of the microbiota. Aside from the peak at Day 17, the levels of *Campylobacter* were relatively consistent within and remained high in every bird CIG1 giving significant P values when comparing; CTRL/CIG1, MIG/CIG1 and CIG1/CIG2 at Day 10, CTRL/CIG1, MIG/CIG1 and CIG1/CIG2 at Day 15, CTRL/CIG1, MIG/CIG1 and CIG1/CIG2 at Day 17, CTRL/CIG1, MIG/CIG1 and CIG1/CIG2 at Day 21 and CTRL/CIG1 at Day 28.



B

Phylum	Order	Genus
Firmicutes	*Lactobacillales*	*Lactobacillus*
Firmicutes	*Clostridia*	*Romboutsia*
Bacteroidetes	*Bacteroidales*	*Bacteroides*
Firmicutes	*Lactobacillales*	*Enterococcus*
Proteobacteria	*Campylobacterales*	*Campylobacter*

Figure 6.9 - Bar plots showing Genus/Species Level Differences uncovered by SIMPER Analysis
A) SIMPER analysis of Bray-Curtis differences between the groups was conducted to reveal the largest contributors to the β-diversity differences we observed. The highlighted genera were then plotted and further analysed by submitting each sequence to EzBioCloud for species level assignment. In the case of Lactobacillus, it was discovered that Lactobacillus johnsonii was largely responsible for the differences we observed between CTRL and the other Campylobacter infection groups. This analysis also revealed the exceptionally high levels of Campylobacter in CIG1 relative to the other groups as well as increases in Bacteroides. The proportion of each genus/species is shown at each time point split by Campylobacter infection group, error bars are standard deviation and statistical testing was performed using FDR corrected unpaired Wilcoxon signed rank tests. The significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05
B) Table shows taxonomic information for each of the selected genera at the Phylum and Order levels for ease of interpretation and to allow the data shown here to be related back to Fig 6.8.

6.3.7 Pseudometagenomics Analysis with Tax4Fun Predicts Metabolic Changes in the *Campylobacter* Infection Groups

In order to gain an insight into the potential implications of the microbiota changes observed using 16S sequencing, Tax4Fun was used to predict the metagenomic composition of the gut flora from pre-calculated profiles of each taxa. Tax4Fun is incompatible with phylogenetic data from RDP and therefore taxa were reassigned to the 16S sequences using the SILVA database instead. The genetic composition of the gut microbiota from the chicken samples was then estimated from precomputed metagenomic profiles for each taxa using Whole Genome Sequence (WGS) data. Using Tax4Fun to examine the predicted metagenomics in this way revealed a number of interesting changes (Fig 4.10). We then calculated Bray-Curtis distances between the predicted profiles of each bird and again ran SIMPER analysis to identify the main contributors to differences between the groups.

Examining the predicted genes highlighted by SIMPER analysis reveals that they tend to fall into 1 of 2 categories, those that are increased mostly in MIG and those that are increased mostly in CIG1. The SIMPER highlighted genes which were increased in MIG (Fig 6.10A) were K00335: NADH-quinone oxidoreductase subunit F, K00656: putative formate acetyltransferase 2, K11063: toxin A/B, K03737: flavodoxin oxidoreductase, K04072: acetaldehyde dehydrogenase. The proportions of each of these genes was the highest at every time point in MIG, with just a few exceptions. Wilcoxon testing gave relatively consistent results for all of the genes and all were significant for CTRL/MIG at Day 10 and CTRL/MIG, MIG/CIG1 and MIG/CIG2 at Day 35 as well as various other time points as shown in Fig 10A. That these genes tend to correlate with one another in terms of predicted expression levels implies they can all be attributed to changes in one specific taxa. Unsurprisingly then, systematically removing sequences from the highest taxonomic level down to the genus level revealed that all of these predicted increases in MIG were attributable to the increased proportion

of *Romboutsia* in this group and their expression closely follows the proportion of this genus as in Fig 6.9A.

A number of genes were also increased in their predicted proportions in the CIG1 group (Fig 6.10B). K01190: Beta-galactosidase, K12373: hexosaminidase and K05349: beta-glucosidase were all found to be at their highest in CIG1 at days 8, 10, 17, 21, 28, 42 and 49. However, the levels of these genes in CIG1 were highly variable and therefore only tested significant when comparing CIG1/CIG2 at day 21 and day 28 for K01190 and K12373. Using the same method as for the genes which were increased in the MIG group, we attempted to identify the microbial origin of the genes which were altered in CIG1 group. However, these changes were difficult to attribute to any specific genera, despite the match for the *Bacteroides* profile in the genus plots, and thus are most likely a result of multiple compositional changes. These differences reveal that changes to the microbiota as a result of *Campylobacter* colonisation can impact the metagenome as well as the composition of the microbiome.

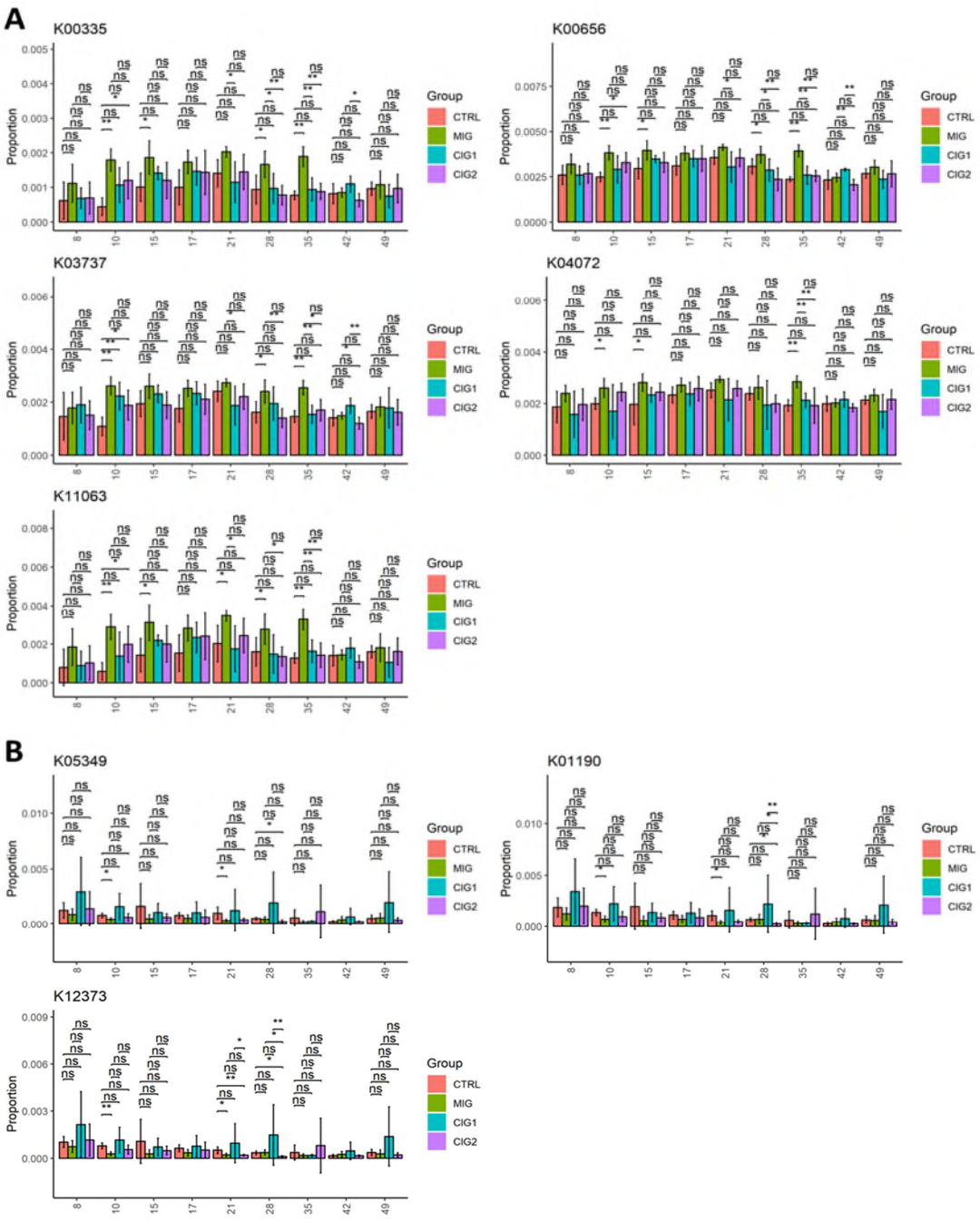


Figure 6.10 - Pseudometagenomic Analysis reveals Increases in Certain Genes Associated with Microbiota Differences

16S sequences were reassigned using SILVA to allow Tax4Fun metagenomics predictions to be made. Bray-Curtis differences based on the pseudo-metagenomic differences were then calculated and SIMPER used to find the largest contributors to differences between the chicken groups. This method highlighted a range of genes of interest, which were then plotted at each time point split by each Campylobacter infection group. This method revealed a range of interesting genes including **A)** those increased due to increased Romboutsia in MIG and **B)** a group of sugar metabolism genes increased in the CIG1 group. Following this, each taxonomic group was removed and the plots remade in order to logically identify the taxonomic differences responsible for the changes to the predicted metagenomics profile of each group. The proportion of the predicted metagenome is plotted on the Y axis and error bars are standard deviation. The significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05

6.3.8 Chicken Tissue Sample Processing and TCR Gamma Repertoire PCRs

The chicken tissue samples taken at the end of the experiment were processed, RNA extracted, cDNA preparation performed and PCRs run as detailed in the methods section (See Chapter 2). The PCR products were then run on a gel and the amplicons visualised (Fig 6.11). The length of a rearranged TCR gamma is approximately 610bp depending on the TRGV or J used and the degree of addition or deletion of nucleotides at the recombination sites. With the incorporation of the UPA primers the expected product length is around 650bp. Due to the incorporation of the poly dT adapter (5' CDS primer A) into every RNA molecule, methods using UPA primers often result in 'smeary' products as can be seen on the gels[38]. Additionally, a smaller product around 200bp is often observed in these PCRs, however this and the smears are removed during downstream gel extraction. Though the majority of samples result in a band at the expected size, many do not display this product, but often are represented in another of the triplicate PCRs. The concentrations of each of these products are estimated by comparison to the 100bp NEB ladder, which informs the later pooling steps. These products were then pooled and library prepped, then sequenced as previously described (See Earlier in this Chapter).

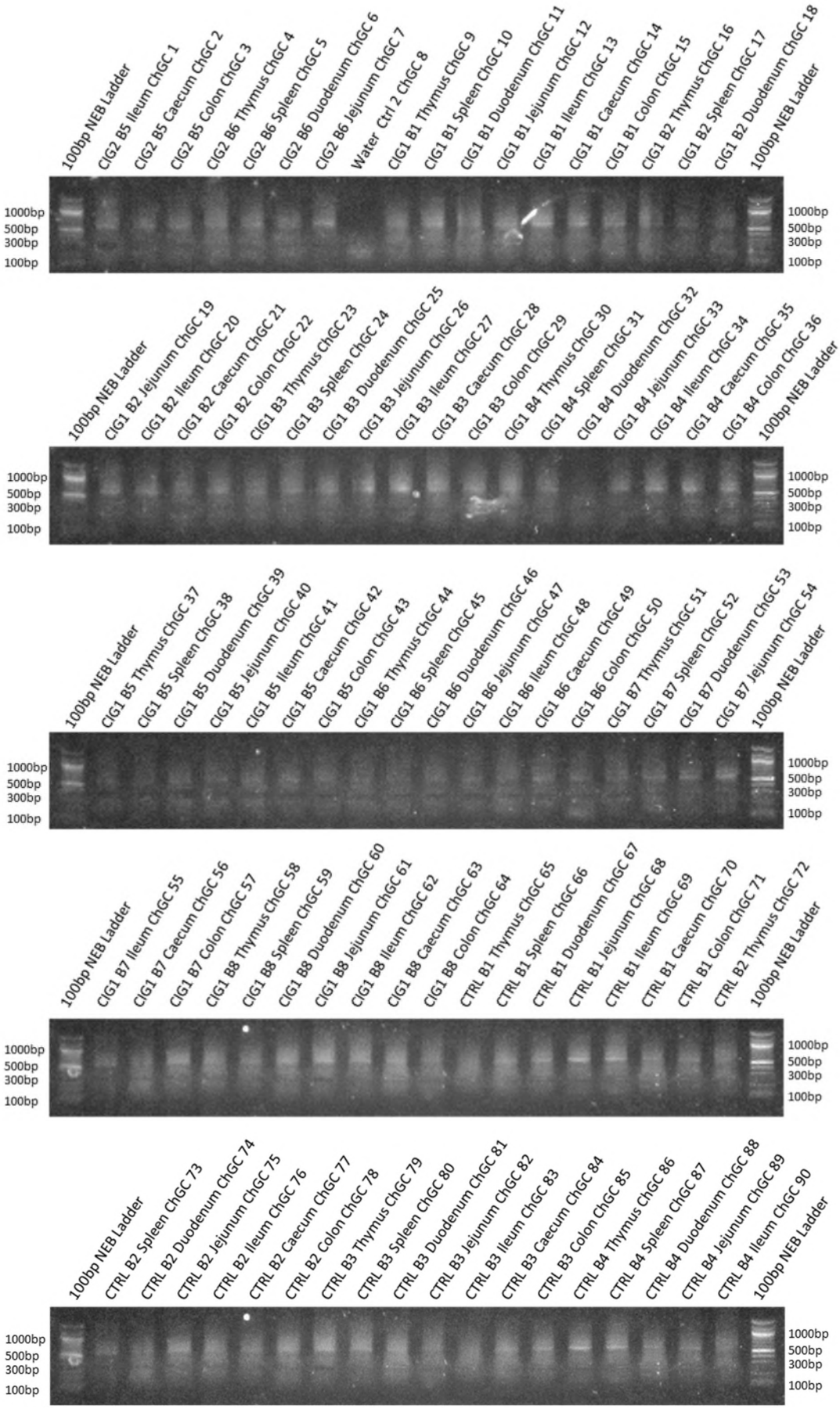

100bp NEB Ladder
CIG2 B5 Ileum ChGC 1
CIG2 B5 Caecum ChGC 2
CIG2 B5 Colon ChGC 3
CIG2 B6 Thymus ChGC 4
CIG2 B6 Spleen ChGC 5
CIG2 B6 Duodenum ChGC 6
CIG2 B6 Jejunum ChGC 7
Water Ctrl 2 ChGC 8
CIG1 B1 Thymus ChGC 9
CIG1 B1 Spleen ChGC 10
CIG1 B1 Duodenum ChGC 11
CIG1 B1 Jejunum ChGC 12
CIG1 B1 Ileum ChGC 13
CIG1 B1 Caecum ChGC 14
CIG1 B1 Colon ChGC 15
CIG1 B2 Thymus ChGC 16
CIG1 B2 Spleen ChGC 17
CIG1 B2 Duodenum ChGC 18
100bp NEB Ladder
1000bp
500bp
300bp
100bp
CIG1 B2 Jejunum ChGC 19
CIG1 B2 Ileum ChGC 20
CIG1 B2 Caecum ChGC 21
CIG1 B2 Colon ChGC 22
CIG1 B3 Thymus ChGC 23
CIG1 B3 Spleen ChGC 24
CIG1 B3 Duodenum ChGC 25
CIG1 B3 Jejunum ChGC 26
CIG1 B3 Ileum ChGC 27
CIG1 B3 Caecum ChGC 28
CIG1 B3 Colon ChGC 29
CIG1 B4 Thymus ChGC 30
CIG1 B4 Spleen ChGC 31
CIG1 B4 Duodenum ChGC 32
CIG1 B4 Jejunum ChGC 33
CIG1 B4 Ileum ChGC 34
CIG1 B4 Caecum ChGC 35
CIG1 B4 Colon ChGC 36
CIG1 B5 Thymus ChGC 37
CIG1 B5 Spleen ChGC 38
CIG1 B5 Duodenum ChGC 39
CIG1 B5 Jejunum ChGC 40
CIG1 B5 Ileum ChGC 41
CIG1 B5 Caecum ChGC 42
CIG1 B5 Colon ChGC 43
CIG1 B6 Thymus ChGC 44
CIG1 B6 Spleen ChGC 45
CIG1 B6 Duodenum ChGC 46
CIG1 B6 Jejunum ChGC 47
CIG1 B6 Ileum ChGC 48
CIG1 B6 Caecum ChGC 49
CIG1 B6 Colon ChGC 50
CIG1 B7 Thymus ChGC 51
CIG1 B7 Spleen ChGC 52
CIG1 B7 Duodenum ChGC 53
CIG1 B7 Jejunum ChGC 54
CIG1 B7 Ileum ChGC 55
CIG1 B7 Caecum ChGC 56
CIG1 B7 Colon ChGC 57
CIG1 B8 Thymus ChGC 58
CIG1 B8 Spleen ChGC 59
CIG1 B8 Duodenum ChGC 60
CIG1 B8 Jejunum ChGC 61
CIG1 B8 Ileum ChGC 62
CIG1 B8 Caecum ChGC 63
CIG1 B8 Colon ChGC 64
CTRL B1 Thymus ChGC 65
CTRL B1 Spleen ChGC 66
CTRL B1 Duodenum ChGC 67
CTRL B1 Jejunum ChGC 68
CTRL B1 Ileum ChGC 69
CTRL B1 Caecum ChGC 70
CTRL B1 Colon ChGC 71
CTRL B2 Thymus ChGC 72
CTRL B2 Spleen ChGC 73
CTRL B2 Duodenum ChGC 74
CTRL B2 Jejunum ChGC 75
CTRL B2 Ileum ChGC 76
CTRL B2 Caecum ChGC 77
CTRL B2 Colon ChGC 78
CTRL B3 Thymus ChGC 79
CTRL B3 Spleen ChGC 80
CTRL B3 Duodenum ChGC 81
CTRL B3 Jejunum ChGC 82
CTRL B3 Ileum ChGC 83
CTRL B3 Caecum ChGC 84
CTRL B3 Colon ChGC 85
CTRL B4 Thymus ChGC 86
CTRL B4 Spleen ChGC 87
CTRL B4 Duodenum ChGC 88
CTRL B4 Jejunum ChGC 89
CTRL B4 Ileum ChGC 90

Figure 6.11 – TCR Gamma Repertoire PCRs from Multi Strain *Campylobacter* Infected Chickens
Agarose gel image showing universal TCR gamma repertoire amplicons. After RNA extracted from tissues had been converted into cDNA using the RACE method, UPA and barcoded C primers were used to amplify the entire chicken gamma repertoire simultaneously, eliminating any artefacts that may otherwise be introduced into the data by PCR bias. A different barcode was used for each sample and the above is an example library. 3 PCR replicates were performed such that each sample had a different barcode in each of the repeat PCRs. The resulting products were then run on a 1.5% agarose gel at 120V for 25 minutes. Samples were run alongside a 100bp NEB ladder and ladder band sizes are shown in base pairs along either side of each gel image.

6.3.9 Illumina MiSeq Run and Processing Reads

The MiSeq run (encompassing multiple experiments) produced a total of 38,691,024 reads Passing Filer (PF) with a Q30 of 88.23. Of those reads, 6,253,614 were attributed to this TCR gamma study. During processing, almost 50% of these reads were lost at the initial demultiplexing stage, which involves quality control and barcode assignment (Fig 6.12). This stage represents the largest data attrition in the processing pipeline and the majority of this loss is attributed to reads which have a barcoded C gamma primer but no UPA, or have a UPA primer but no C gamma. After this stage, a second attrition occurs at the CDR3 assignment step where around 30% of the reads are lost. As has been reported previously, reads may be excluded at this stage due to incomplete recombination, intergenic sequences being retained or concatenated TRGJ genes[39]. Almost no reads are lost at the desired barcodes stage, which involves the removal of any read assigned to a barcode which should not be present. This is reassuring since the only way spurious barcodes could be introduced into the sequencing run is through contamination of the barcode primers or uncorrected barcode sequencing errors. The fact that very few reads are lost indicated that neither of these potential issues present a problem in this data. Finally, any CDR3 which includes incomplete codons, or stop codons is removed. In this case, this accounts for only a 10% loss of reads from the previous stage, indicating that the majority of reads are from functional γδ T cells producing productive TCR gamma rearrangements.

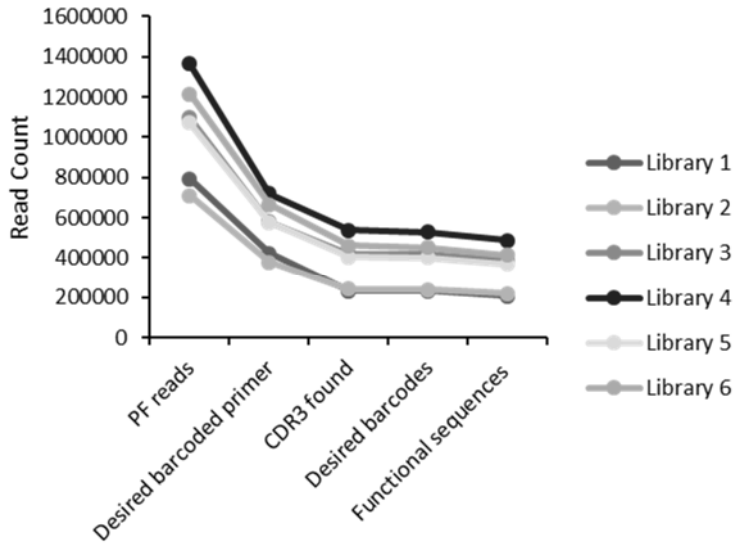


Figure 6.12 - Numbers of Reads at each Stage of Repertoire Data Processing
Plot shows the number of reads at each processing stage. After the MiSeq run was complete, the sequences were processed using reptools and the number of reads at each stage was calculated and plotted. PF reads is the number of reads after internal quality filtering was performed by the MiSeq instrument. Desired barcoded primer refers to reads which had the correct combination of primers and a barcode passing our quality filter. CDR3 found refers to reads for which a TRGV, J and C gene was found by blast searching and a Complementarity Determining Region 3 Nucleotide (CDR3NT) sequence can therefore be extracted and assigned to that read. Desired barcodes is removing any spurious barcodes we don't expect in that particular library and Functional sequences is removing any sequence which had a non-functional sequence or premature stop codon. The samples used in this study are split across 2 libraries and there are 3 replicates for each sample for a total of 6 libraries.

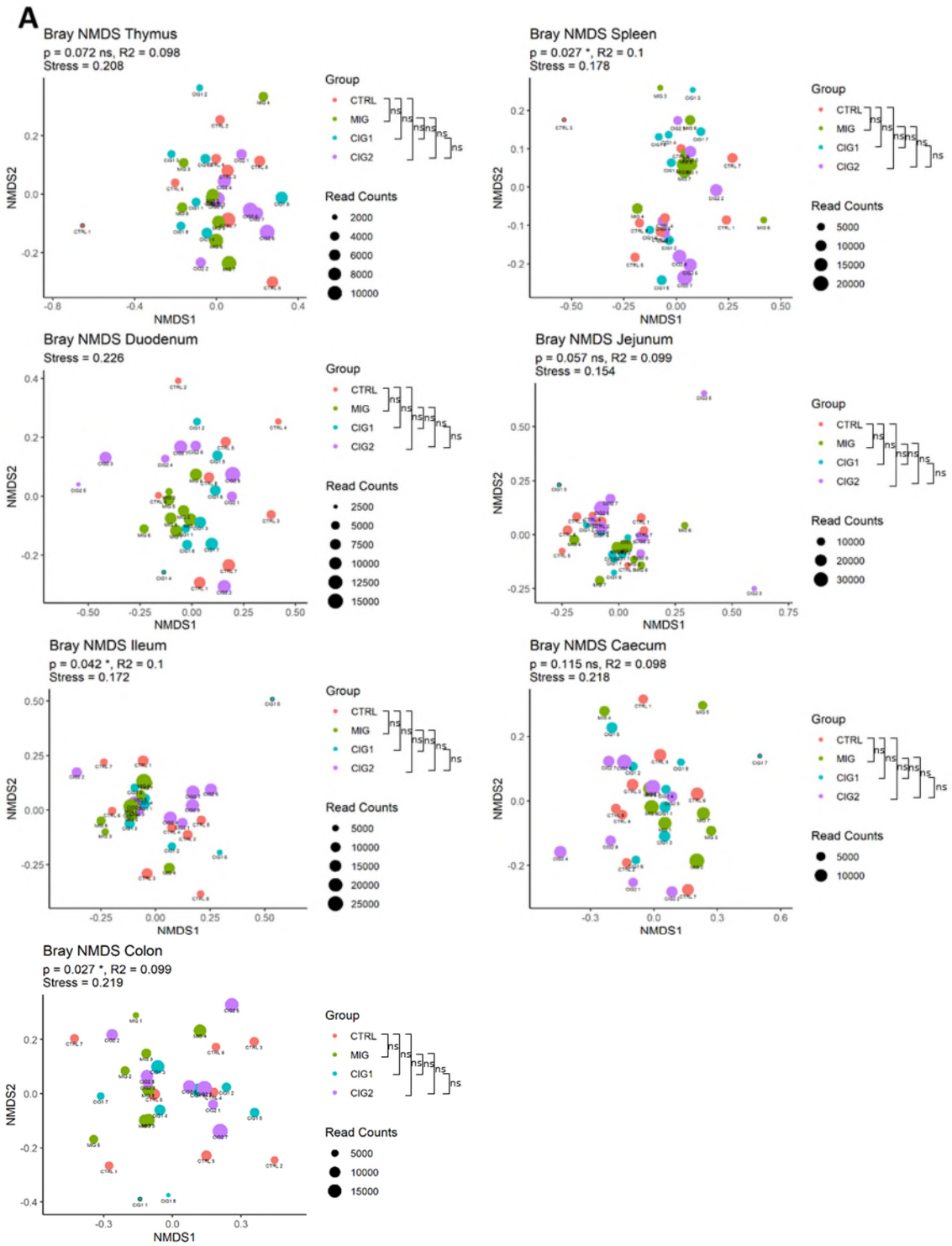
6.3.10 Beta-Diversity Plots shows Small Differences in Repertoire Depending on *Campylobacter* Infection Group.

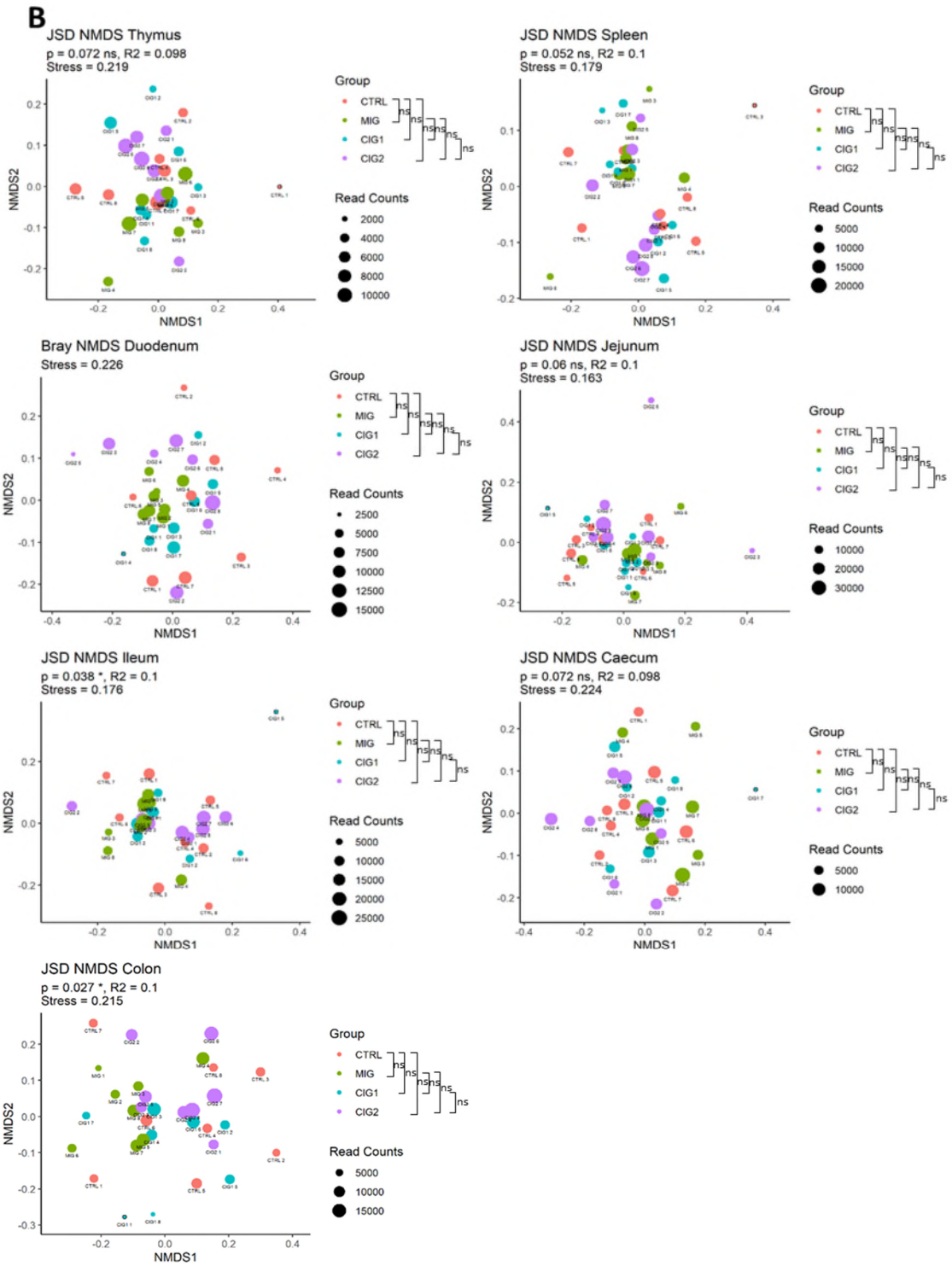
Comparing groups first by Bray-Curtis highlights a clear separation in a number of the tissues. There is an obvious visual separation in the Spleen, Duodenum, Jejunum, Ileum and Colon. It appears that CTRL and CIG2, and CIG1 and MIG birds tend to cluster together with some degree of separation between them. However, there is overlap between the 2 clusters with certain birds consistently observed clustering apart from the rest of the group to which they belong (Fig. 6.13A). All except the Duodenum show homogeneity of variance when tested using Betadisper and thus all meet the assumptions for PERMANOVA testing. Spleen, Ileum and Colon are all significant overall with FDR corrected P values and a 0.05 significance threshold, and Jejunum is very close to significance. However, all of these tissues have low R^2 values and therefore the effect sizes are small.

Testing of individual pairs of groups gave non-significant differences for every single comparison in every tissue. As we did find significant overall differences, this is likely simply due to a lack of power due to insufficient group sizes, or overcorrection by FDR.

The Jensen–Shannon divergence (JSD) results were almost identical to those obtained using Bray-Curtis (Fig 6.13B). There was similar visual separation in Spleen, Jejunum, Ileum and Colon, however from our assessment there was no separation in Duodenum, though this is open to interpretation. By JSD, only Ileum and Colon were significant, with Spleen and Jejunum on the edge of the 0.05 significance threshold. However again, the R^2 values are very low and none of the individual pairs of comparisons were significant.

Comparing the groups using Aitchison's distance showed relatively similar results to those obtained using Bray-Curtis and JSD (Fig 6.13C). The Aitchison's distance plots showed visual separation in Spleen, Duodenum, Jejunum, Ileum, Caecum and Caecum. When comparing groups using Aitchison distance, we observed more points clustering distal from the rest of the data, though other than this, the results were similar to those observed for the other 2 indices. Despite this separation, betadisper testing showed that there is homogeneity of variance in all groups and all tissues and thus PERMANOVA testing was valid in all cases. Testing with FDR corrected P values revealed significant differences in Duodenum, Ileum and Colon, with Spleen and Jejunum only narrowly above the threshold. Once again, however, this β-Diversity metric gives low R^2 values and therefore *Campylobacter* infection status is only explaining a small minority of the variation in these plots. In the case of the Aitchison's distance plots, there were significant results in between some of the pairs of groups in some of the tissues namely, Duodenum CTRL/MIG, MIG/CIG1, MIG/CIG2 and CIG1/CIG2 and Colon CTRL/MIG and CIG1/CIG2. This suggests that Aitchison's distance may have greater resolving power, and suggests that *Campylobacter* infection is causing larger differences between these groups in Duodenum and Colon.


A
Bray NMDS Thymus
p = 0.072 ns, R2 = 0.098
Stress = 0.208
Bray NMDS Spleen
p = 0.027 *, R2 = 0.1
Stress = 0.178
Bray NMDS Duodenum
Stress = 0.226
Bray NMDS Jejunum
p = 0.057 ns, R2 = 0.099
Stress = 0.154
Bray NMDS Ileum
p = 0.042 *, R2 = 0.1
Stress = 0.172
Bray NMDS Caecum
p = 0.115 ns, R2 = 0.098
Stress = 0.218
Bray NMDS Colon
p = 0.027 *, R2 = 0.099
Stress = 0.219
NMDS1
NMDS2
Group
CTRL
MIG
CIG1
CIG2
ns
Read Counts


B
JSD NMDS Thymus
p = 0.072 ns, R2 = 0.098
Stress = 0.219
JSD NMDS Spleen
p = 0.052 ns, R2 = 0.1
Stress = 0.179
Bray NMDS Duodenum
Stress = 0.226
JSD NMDS Jejunum
p = 0.06 ns, R2 = 0.1
Stress = 0.163
JSD NMDS Ileum
p = 0.038 *, R2 = 0.1
Stress = 0.176
JSD NMDS Caecum
p = 0.072 ns, R2 = 0.098
Stress = 0.224
JSD NMDS Colon
p = 0.027 *, R2 = 0.1
Stress = 0.215
NMDS1
NMDS2
Group
CTRL
MIG
CIG1
CIG2
ns
Read Counts

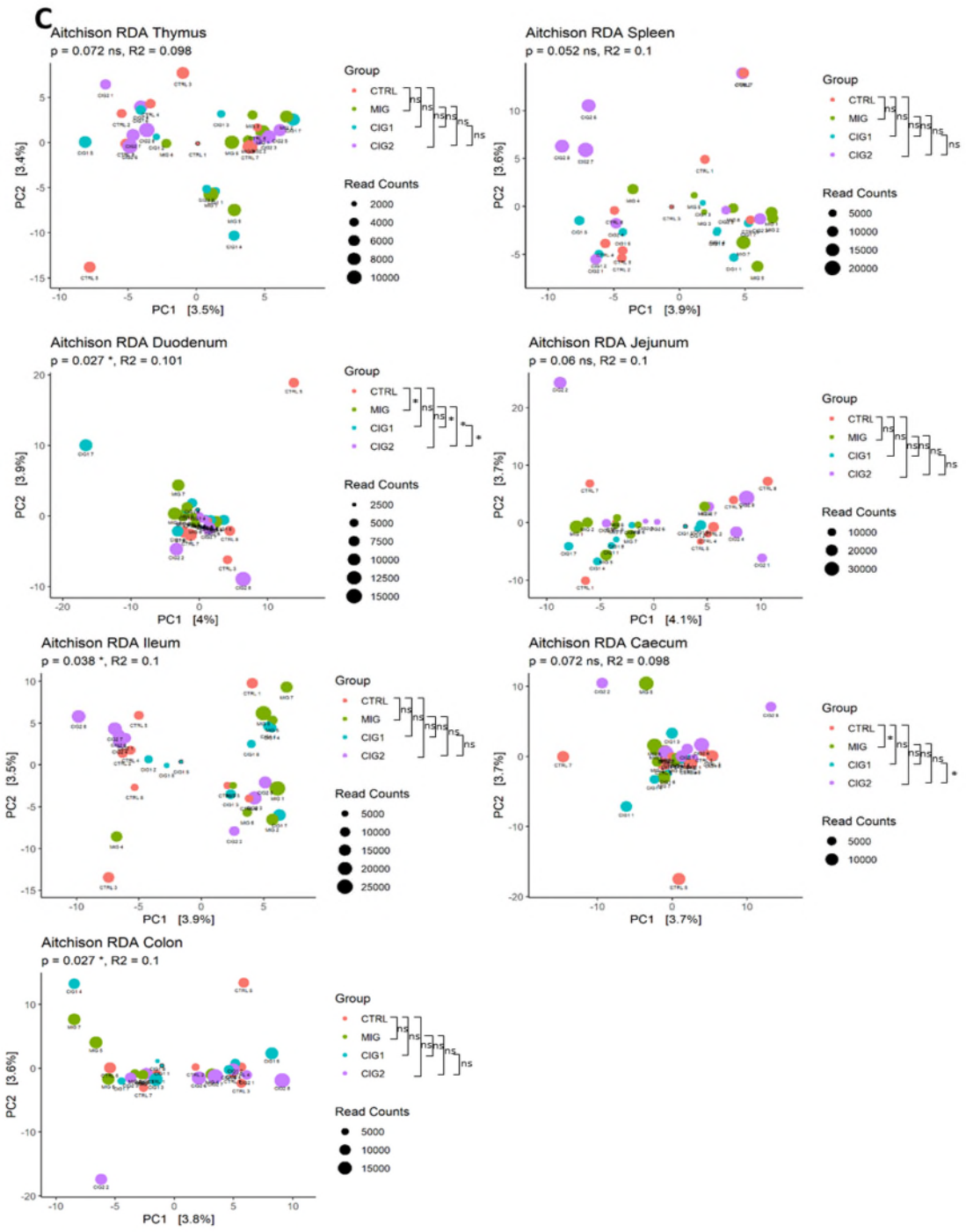


Figure 6.13 – TCR Gamma Repertoire β-Diversity Plots show Differences between *Campylobacter* Infection Groups

NMDS and RDA plots were generated by calculating **A)** Bray-Curtis, **B)** JSD and **C)** Aitchison's distances between samples treating each clone, defined by a unique combination of TRGV, TRGJ and CDR3nt, as an individual species. These distance metrics were then used to generate NMDS plots for Bray-Curtis using 20 random starts. Colours are used to represent each of the groups. Point size is determined by the number of reads associated with that sample, thus providing a basic measure of the reliability of that point. P values and R2 values are indicated beneath the subtitle and the significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. P values were FDR corrected. Stress values are indicated below the title for NMDS and percentage of variation explained in the axes for RDA.

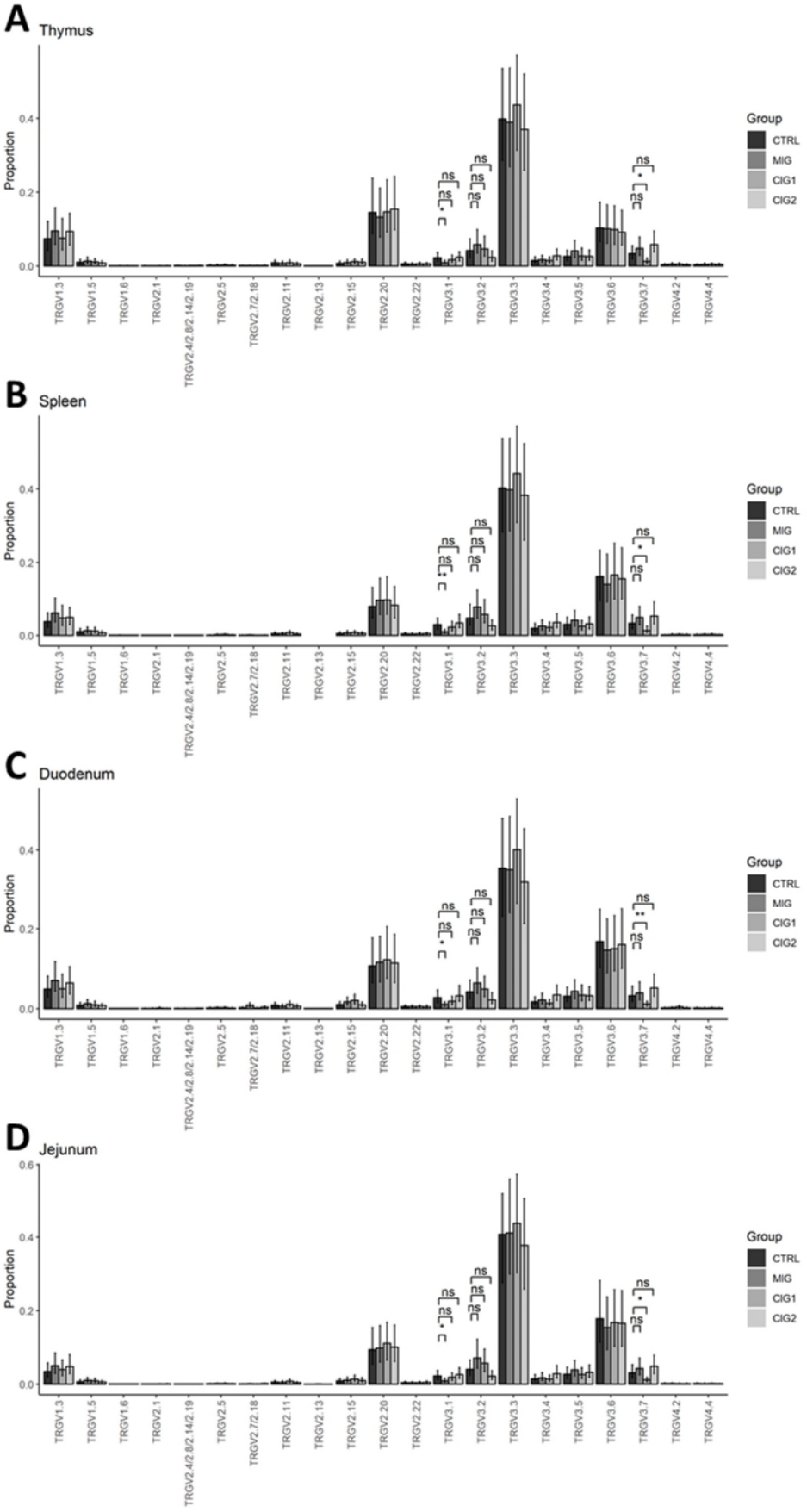
6.3.11 TRGV Gene Segment Usage Plots show Differences in the Proportion of some Minority Rearrangements according to *Campylobacter* Infection Group

Plotting TCR gamma repertoire TRGV segment usage gives an insight into how *Campylobacter* colonisation might be driving expansion of certain clones expressing specific TRGV genes (Fig 6.14A-G). A Linear Mixed Model (LMM) was generated with Bird number as a random variable and Tissue, Day and TRGV gene as fixed variables using the lme4 package with a logit transformation[40]. The TRGV gene profiles from each tissue looked mostly very similar, with TRGV3.3 as the most dominant TRGV gene by far, occupying around 40% of the repertoire in each tissue. TRGV2.20 and TRGV3.6 were also prominent rearrangements in every tissue. The one slight exception to this is was the Thymus, which displayed increased TRGV2.20 and TRGV1.5 at the expense of TRGV3.6, which was decreased in comparison to the other tissues. Though there were no differences in the proportions of the most commonly used TRGV genes, differences between the *Campylobacter* infection groups do exist in the less common ones. In every tissue compared to the CTRL group, MIG has reduced TRGV3.1, CIG2 has reduced TRGV3.2 and CIG1 has reduced TRGV3.7.

TRGV3.1 occupies around ~0.02 of total proportional V gene usage in CIG1, CTRL and CIG2, in the Thymus, Jejunum, Ileum, Caecum and Colon, and slightly higher levels in the Spleen and Duodenum of CTRL and CIG2 at around ~0.03. In all of these tissues however, the levels of TRGV3.1 were reduced ~2-4 fold in the MIG group, resulting in significant differences between CTRL and MIG in all of the tissues using P values generated from the LMM.

TRGV3.2 represents 0.04-0.07 of the TCR gamma repertoire in the Thymus, Spleen, Duodenum, Jejunum, Ileum Caecum and Colon of CIG1, CTRL and MIG birds. In each of these tissues, the levels of TRGV3.2 in CIG2 dropped down to around 0.02. However, when compared to CTRL, these differences were significant only in the Colon. In the Colon, there was also a significant increase in TRGV3.2 in the MIG strain, which rose to account for 0.095 of the repertoire.

TRGV3.7 accounts for 0.03-0.055 of the TCR gamma repertoire of CTRL, CIG2 and MIG in all of the tissues examined. In each of these tissues, the proportion of TRGV 3.7 in CIG1 was reduced down to 0.01, resulting in significant P values when compared to the CTRL birds in every tissue.


A
Thymus
B
Spleen
C
Duodenum
D
Jejunum
Proportion
Group
CTRL
MIG
CIG1
CIG2
ns
TRGV1.3
TRGV1.5
TRGV1.6
TRGV2.1
TRGV2.4/2.8/2.14/2.19
TRGV2.5
TRGV2.7/2.18
TRGV2.11
TRGV2.13
TRGV2.15
TRGV2.20
TRGV2.22
TRGV3.1
TRGV3.2
TRGV3.3
TRGV3.4
TRGV3.5
TRGV3.6
TRGV3.7
TRGV4.2
TRGV4.4

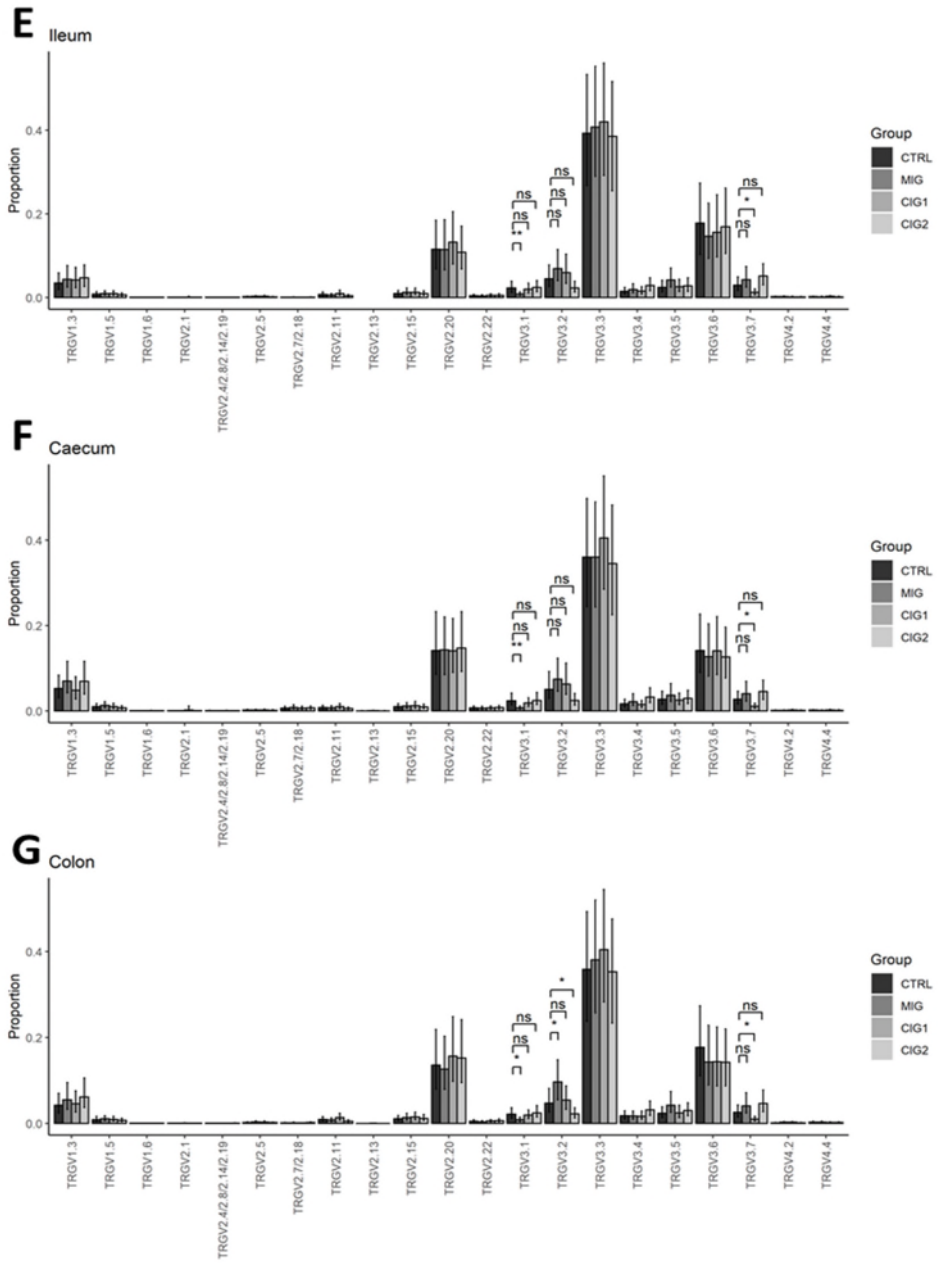


Figure 6.14 - TRGV Gene Usage Plots show Subtle TCR Gamma Rearrangement Differences in *Campylobacter* Infection Groups

Plots show TRGV gene segment usage by Campylobacter infection group in **A)** Thymus, **B)** Spleen, **C)** Duodenum, **D)** Jejunum, **E)** Ileum, **F)** Caecum and **G)** Colon. TRGV gene segments were assigned to each sequencing read by blasting them against a database of known sequences. Chicken TRGV genes have very low sequence diversity between families at the genomic level and thus some have to be combined into a single category, as they could not be reliably distinguished from one another. A linear mixed model with random intercept was calculated using Bird number as a random variable and Tissue, Day post-hatch and TRGV gene as fixed variables. Proportions were logit transformed before model calculation. This linear mixed model was used to calculate P values and 95% confidence intervals. Error bars are 95% confidence intervals calculated from 500 bootstrap replicates generated using the model. P values were generated using the LmerTest package using the Satterthwaite approximation of degrees of freedom. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05

6.3.12 Publicity in some TRGV Gene Segments is Enhanced by *Campylobacter* Colonisation

Publicity is the degree to which a TCR is shared between individuals and is an important factor in determining γδ T cell functions and response types, which range from innate to adaptive-like. Looking at publicity of the TCR gamma repertoire provides information about the clones which are shared within each group and how they are affected by *Campylobacter* colonisation (Fig 6.15). The degree of publicity was examined in each of the groups, and plotted split by TRGV gene, by calculating the percentage publicity of each clone and the contribution of each of these categories to the overall proportion of that TRGV segment. In this case, the publicity of each clone was defined as presence in two or more birds in any tissue at any proportion. The percentage publicity was calculated as the number of birds a given clone was present in, divided by the total number of birds in that group. Each of the TRGV genes was adjusted such that their proportions added up to one by gene rather than by bird for easier interpretation, though it should be noted that some are still extremely rare as shown in Fig 6.14.

In each of the *Campylobacter* infected groups, there were a number of increases in publicity relative to the CTRL group. In MIG, CIG1 and CIG2 there were significant increases in publicity in TRGV2.15, TRGV3.3 and TRGV3.6, when compared to CTRL, as well as some differences specific to each infection group. Additionally, in the MIG group, TRGV1.3, TRGV2.20, TRGV2.22, TRGV3.2 were also significantly increased. CIG1 and 2 have the same profile of differences and in addition to the common increases, both also had significantly increased TRGV1.5.

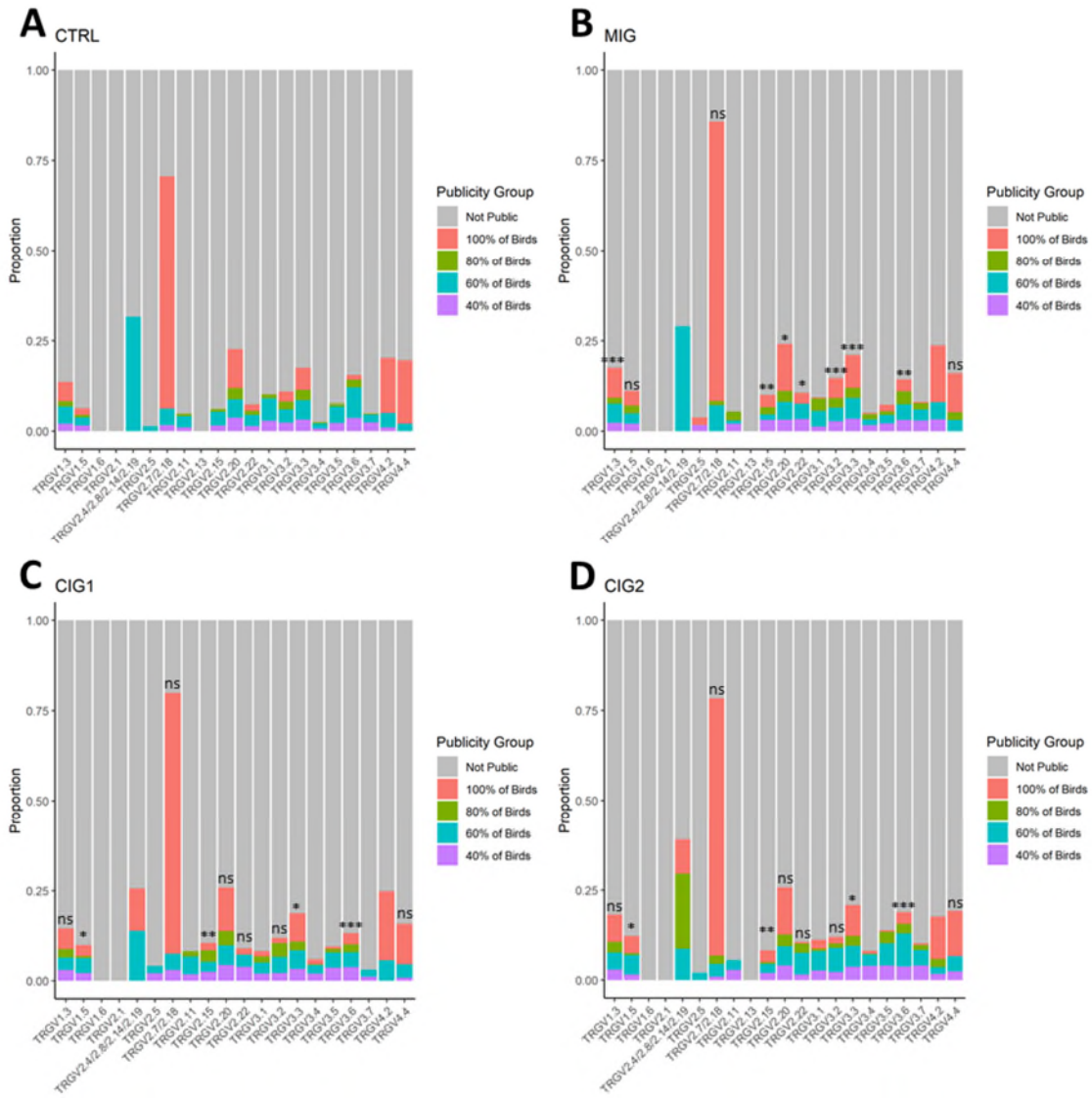


Figure 6.15 - Campylobacter infected birds show slight increases in TCR gamma repertoire publicity relative to CTRL

Plots show levels of publicity by TRGV gene in each Campylobacter infection group; **A)** CTRL, **B)** MIG, **C)** CIG1 and **D)** CIG2. The publicity of each clone was determined by the percentage of birds in which it is present. Presence includes clones observed at any proportion and in any tissue. Publicity in each group was split by each TRGV gene. Percentage publicity is represented by colour and is shown in the legend alongside each of the plots. The proportion of each TRGV gene has been adjusted to add up to 1 in order to make the plots more easily interpretable, though the actual proportions are as in Fig 6.14. Statistical comparisons were made by comparing each Campylobacter infection group with the CTRL group and are indicated above the bars of the relevant group. P values were calculated using an unpaired Wilcoxon rank test performed on the 100% of birds category only. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05 and are indicated above the TRGV gene which was compared.

6.3.13 TCR Gamma Diversity in the Jejunum and Ileum is Altered by *Campylobacter* Colonisation

To further examine the overall clonal population structure of the TCR gamma repertoire, we calculated diversity indices for each tissue in each bird. We then constructed a linear model (LM) using cube root transformed data and used this to extract confidence intervals and P values. We plotted this data split by group to examine the impact of *Campylobacter* mono and co-infection.

The diversity plots showed substantial differences across tissues with Thymus being the most diverse by a fair margin (Fig 6.16). Spleen was the next most diverse tissue and showed a substantial drop in diversity, around half that seen in the Thymus in all groups. The small intestinal tissues Jejunum and Ileum were only slightly less diverse again. There was then a large decrease in diversity down to the level of the Colon, followed closely by Caecum, then Duodenum, which was the least diverse tissue.

Comparing the diversity in each of the infection groups revealed that some differences do exist between them. Interestingly, though the differences were small, MIG was the least diverse in the Thymus and Spleen, with CIG1 being the most diverse, followed by CIG2 and CTRL, which both had very similar diversity. The diversity of each of the groups in the Duodenum was again relatively similar with only minor differences between the groups. The Jejunum and Ileum however, featured larger differences in comparison with the other tissues. While CTRL and CIG1 had fairly similar levels of diversity in the jejunum, both CIG2 and MIG had around half the diversity by every metric and this difference was significant by P values extracted from the LM using LmerTest. The Ileum showed more or less the opposite pattern with the MIG group showing the greatest diversity. However, none of the P values from the differences in the Ileum reach a significance threshold of 0.05. Contrary our expectations, the Caecum and Colon showed no differences between groups whatsoever in terms of TCR gamma repertoire diversity.

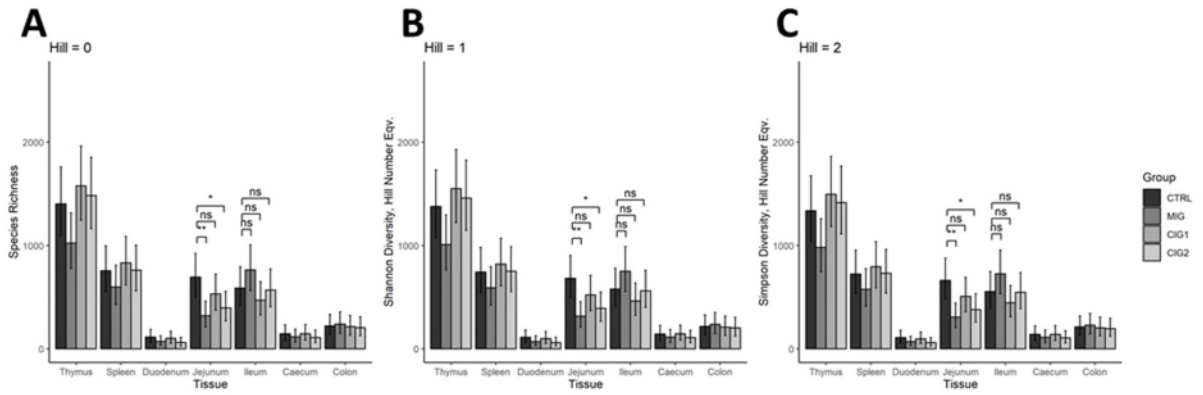


Figure 6.16 - Hill Number Diversity Indices for each Tissue split by *Campylobacter* Infection Group. Diversity plots show **A)** Species Richness, **B)** Shannon Diversity and **C)** Simpson diversity plots from iNEXT rarefied data. The diversity indices are displayed as their Hill number variants for ease of comparison and interpretation. A linear model was constructed from the diversity index data using a cube root transform, which was confirmed to best approximate normality using Box Cox plots. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

6.3.14 *Campylobacter* Colonisation Increases the Expression of Selected Largely Public TCR Gamma Clones

In order to identify any clones which might be increased in proportion indicating possible expansions, and therefore potential involvement in an immune response, SIMPER was used to select the clones contributing most to the differences observed in the Bray-Curtis plots (Fig. 6.17).

Clone 1546 appeared in the top contributors to differences between CTRL and MIG in the Spleen and Duodenum, and CTRL and CIG2 in the Thymus. This clone made up the majority of the TRGV2.7/2.18 gene segment and was 100% public in every group, appearing in every single bird in this study. TRGV2.7/2.18 was the most public TRGV gene and interestingly clone 1546 made up the entirety of the '100% of birds' publicity group for this TRGV gene (Fig 6.15). Additionally, the CDR3nt sequence of Clone 1546 is fully germline. Despite the fact that TRGV2.7/2.18 is a very rare TRGV gene, 1546 made up the vast majority of it and is therefore in the top 10% of clones by proportion in the majority of birds, hence it is able to make a substantial contribution to the Bray-Curtis distance between groups. There was a lack of any difference in 1546 between groups in the Jejunum, Ileum,

Caecum and Colon, however, in the Spleen and Thymus, clone 1546 was increased in all of the *Campylobacter* infected groups compared with the CTRL group. This increase was particularly large in CIG2 and MIG, both of which give a significant result when compared with CTRL in both of these tissues. Duodenum showed similar results, again with increases in 1546 in CIG2 and MIG, however, in this instance only the difference between CTRL and MIG was significant.

Clone 19873 appeared in the top clones contributing to Bray-Curtis distance between CTRL and CIG2 in the Duodenum, and CTRL and MIG in the Spleen. 19873 expresses a TRGV3.6 and is 100% public in MIG, 87.5% public in CIG2, 62.5% public in CTRL and 50% public in CIG1. TRGV3.6 is a very highly expressed TRGV gene and was the second most common in all tissues except Thymus where it was third. Though it was highly variable in proportion, 19873 itself is not a particularly common clone, except in the Spleen where it was in the top 5% in the majority of birds in the MIG group. Concordant with this, 19873 was significantly increased in the Spleen of the MIG group compared to CTRL. This increase in 19873 in the Spleen wasn't present in any of the other *Campylobacter* infected groups. Additionally, a similarly significant difference can be observed in the Duodenum and Jejunum. The other tissues, Thymus, Jejunum, Ileum, Caecum and Colon did not show any differences in terms of 19873 between the groups.

Clone 491 was present in the top contributing clones to Bray-Curtis differences between CTRL and CIG1 in the caecum, CTRL and CIG2 in the Duodenum and CTRL and MIG in the Thymus, Duodenum and Jejunum. It expresses a TRGV3.3 V gene and was 100% public in all chicken groups. Interestingly however, and contrary to what would be expected in a public clone, clone 491 differed from the germline sequence by 2 nucleotides. There were clear differences in terms of the proportion of 491 in the Duodenum and Spleen, and it was increased in these tissues in all of the *Campylobacter* infected groups relative to the control. These differences however, were significant in the MIG group only.

Clone 17151 was present in the top contributing clones to differences between CTRL and MIG. It expresses a fully germline TRGV3.6 and was 100% public in MIG, 75% public in CIG2, 50% public in CTRL and 75% public in CIG1. Looking across the tissues sampled, MIG had a significantly increased proportion of 17151 relative to CTRL in Duodenum, Jejunum and Colon.

Clone 971 was highlighted by SIMPER analysis as being present in the top 20 clones contributing to differences between CTRL and CIG2. It expressed a TRGV3.6 that was mostly public, with 62.5% publicity in MIG, 87.5% publicity in CIG2, 75% publicity in CTRL and 50% publicity in CIG1. Despite these high levels of publicity, the CDR3NT of clone 971 has 3 nucleotide additions when compared to the germline sequence. This clone was increased in the CTRL and CIG2 birds in the Jejunum, Ileum, Caecum and Colon relative to MIG and CIG2. Comparing CTRL with CIG1 and MIG using Wilcoxon tests gave a statistically significant P in the Caecum only.

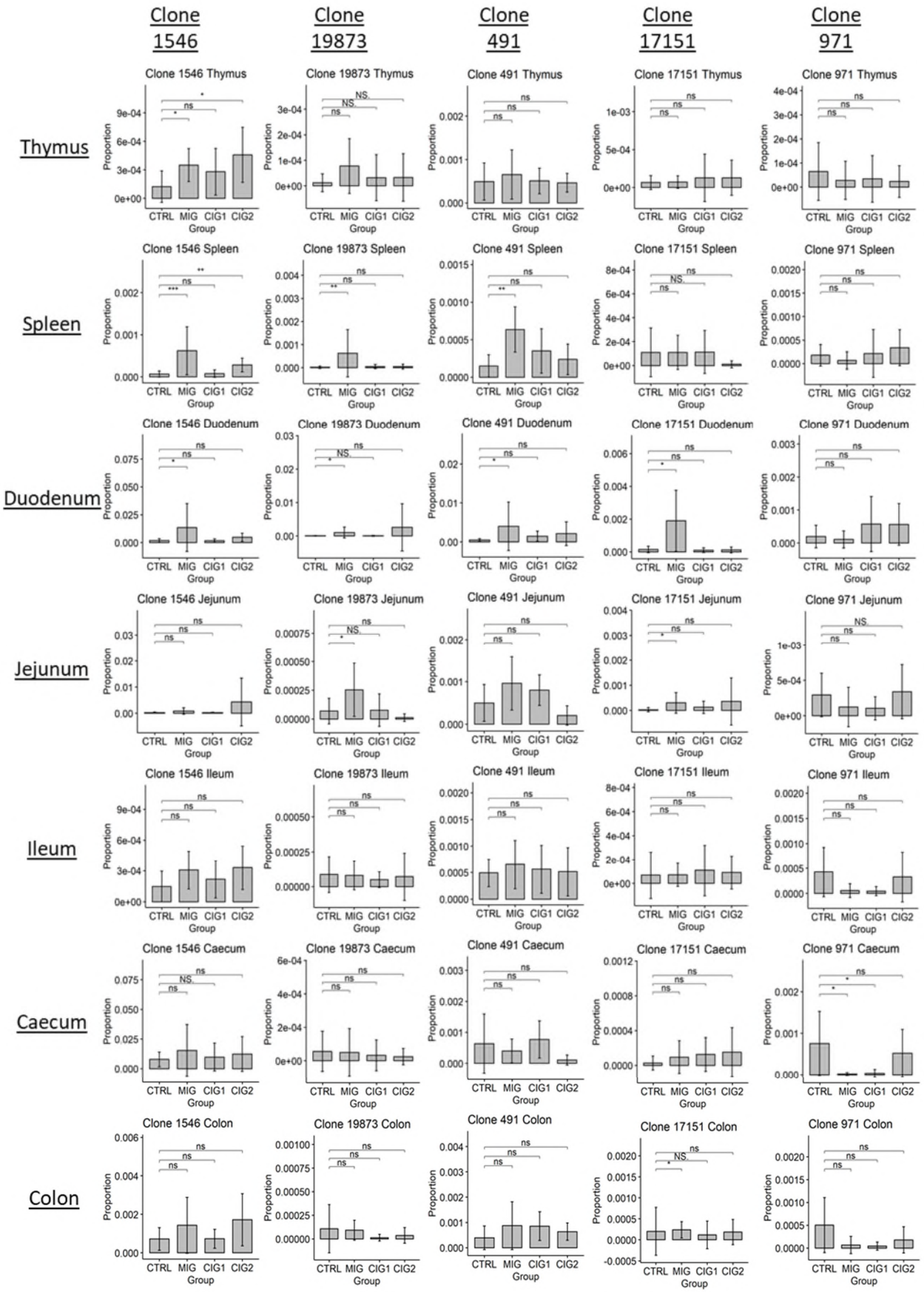

Clone 1546
Clone 19873
Clone 491
Clone 17151
Clone 971
Thymus
Spleen
Duodenum
Jejunum
Ileum
Caecum
Colon
Clone 1546 Thymus
Clone 19873 Thymus
Clone 491 Thymus
Clone 17151 Thymus
Clone 971 Thymus
Clone 1546 Spleen
Clone 19873 Spleen
Clone 491 Spleen
Clone 17151 Spleen
Clone 971 Spleen
Clone 1546 Duodenum
Clone 19873 Duodenum
Clone 491 Duodenum
Clone 17151 Duodenum
Clone 971 Duodenum
Clone 1546 Jejunum
Clone 19873 Jejunum
Clone 491 Jejunum
Clone 17151 Jejunum
Clone 971 Jejunum
Clone 1546 Ileum
Clone 19873 Ileum
Clone 491 Ileum
Clone 17151 Ileum
Clone 971 Ileum
Clone 1546 Caecum
Clone 19873 Caecum
Clone 491 Caecum
Clone 17151 Caecum
Clone 971 Caecum
Clone 1546 Colon
Clone 19873 Colon
Clone 491 Colon
Clone 17151 Colon
Clone 971 Colon
Proportion
Group
CTRL MIG CIG1 CIG2
ns
NS.
*
**
***

Figure 6.17 - Individual TCR Gamma Clone Plots show Further Differences between *Campylobacter* Infected Groups

Plots show the clones responsible for the separation between Campylobacter infection groups shown in Fig 6.13. The biggest contributing clones to the differences between the groups were identified using a SIMPER analysis. The clones that were highlighted were then manually examined and the ones showing the biggest differences were selected for display. Often SIMPER selects clones which have differing means due to outliers or clones present at high proportion but which do not show any meaningful differences when compared based on the groups and therefore such clones were excluded. This analysis revealed a number of clones which show differences between the groups in certain tissues. Error bars are standard deviation and statistical testing was conducted using an unpaired Wilcoxon signed rank test. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05

6.4 Discussion

This chapter aimed to examine the relationship between γδ T cells and the microbiota in the context of a realistic *Campylobacter* infection, where birds are exposed to multiple co-colonising strains. *Campylobacter jejuni* is a zoonotic pathogen and contaminated chicken represents the major source of infection of humans. Though initially thought to be a commensal in chickens, more recent evidence shows it does have impacts on the avian gut, inducing dysbiosis, transient intestinal inflammation and induction of strong antibody responses[2,3,41,42]. Here we employ *Campylobacter* infection as a stimulus for the enteric immune system and as a disrupter of microbiome stability. Recent literature has made it clear that real world *Campylobacter* infection of farmed chickens does not occur with a single strain[20–24]. In order to best simulate a realistic infection, we therefore used oral gavage to inoculate 3 groups of birds with different mixtures using *Campylobacter* strains isolated from commercially reared broiler chickens. These infecting mixtures were named Mono-infection group (MIG), which consisted of a single strain, Co-infection group 1 (CIG1), which consisted of 4 strains all isolated from the same bird and Co-infection group 2 (CIG2) which consisted of 4 dominant strains from 4 different birds. This experimental design of comparing single strain with multiple strain infection enabled us to fully explore the effects of *Campylobacter* status on γδ T cells and the microbiome.

It has been shown that though *Campylobacter* colonisation of chickens can be prevented with extremely rigorous biosecurity procedures[43], this is unrealistic for commercial farming. The control (CTRL) birds used in this study are a case in point for this idea as PorA sequencing showed

they were colonised with multiple strains of *Campylobacter,* of both environmental and experimental origin*,* despite not having received a defined gavage as in the other groups. Despite this, the same sequencing indicated there were far fewer reads associated with these samples, and combined with a lack of observed colonies from caecal content, in stark contrast to MIG, CIG1 and CIG2, it is safe to assume that these birds are colonised to a lesser extent and therefore still make for a useful baseline for examination of changes to the TCR gamma repertoire and microbiota. Furthermore, both modelling and *in vivo* studies have shown that though *Campylobacter* is present in almost all flocks, it only becomes a problem in the context of single strain outgrowth and resulting florid infection[27,44]. Consistent with these ideas, our data shows that the CTRL birds which have a lower *Campylobacter* load, tend to have a more even distribution of PorA types, and the experimentally inoculated groups, MIG, CIG1 and CIG2, with a higher load are dominated by a single PorA type[27]. The distinct lack of variability within groups in these plots is in itself an interesting result and suggests that the proportions of each strain are not simply due to random variation and are likely to be the result of some biological phenomena.

The repertoire results showed clear differences between the groups in the ordination plots and though this separation did not remain consistent at all time points and for all samples, we tended to observe CTRL and CIG2 clustering together, distinct from MIG and CIG1 which also seemed to group together in these plots. Furthermore, the results from the TRGV gene plots show significant decreases in the TRGV genes TRGV3.1, TRGV3.2 and TRGV3.9 for MIG, CIG2, and CIG1 respectively. Due to the absence of any cytokine expression data for TRGV subsets in chicken, it is unclear exactly what sort of immune response these changes might represent. TCR gamma repertoire sorting of CD8 sorted γδ T cell subsets would allow the linkage of this study with the available cytokine expression data in chicken[45]. Despite the fact that we found significant differences in the TRGV gene profiles, these differences are not present in any of the major TRGV gene subsets and therefore the biological impact they might have is likely to be small.

However, we also observed larger differences between the groups in terms of both publicity and diversity. We found that when compared to CTRL, all of the *Campylobacter* infected strains showed significant increases in publicity in multiple different TRGV genes, as well as significant decreases in diversity in the Jejunum. These differences were most substantial in CIG2 and MIG. The increases in publicity in the in the *Campylobacter* infected groups is an interesting finding which holds true for all of them relative to CTRL. These increases in publicity in the remaining families may represent some degree of proliferation of the public clones, recently referred to as quasi-adaptive responses[46], and is therefore somewhat reminiscent of the proliferation noted in human public Vγ9Vδ2 subsets in response to *M. tuberculosis*[47]. However, the differences we observed were smaller than those described in this study. Nonetheless, we did observe more convincing increases in various individual canonical clones in the *Campylobacter* infected groups in various tissues such as 1546 in CIG2 and MIG Spleen, 19873 in MIG Spleen and MIG and CIG2 Duodenum and 17151 in MIG Duodenum, Jejunum and Colon. Despite this, one weakness with repertoire studies is the proportional nature of the data they produce, which makes it impossible to say for certain that an increase in proportion represents a true increase in the numbers of a specific clone, as the overall population size is not known. Therefore, future repertoire studies may benefit from the incorporation of quantitative methods, such as flow cytometry or histology.

Furthermore, similar responses have been observed in studies in mice following challenge with *L. monocytogenes* and *B. subtilis,* in the gut and lung respectively, in which expansions were observed in largely invariant Vγ6Vδ1+ve T cells[48,49]. To our knowledge, this is the first evidence of such quasi-adaptive responses in chicken, and may support the existence of them in other γδ high animals, though additional studies will be needed to confirm this result. It is unclear what the role of such expanded invariant γδ T cell subsets may be, but it has been noted that they resemble the 'trained immunity' reported in previous studies, which appears to represent an innate immune memory[46,50,51]. Regarding the true innate immune system, trained immunity refers to the ability to mount a stronger response, both qualitatively and quantitatively, upon repeat infection much like

the adaptive immune system. Importantly, this mechanism is mediated by epigenetic changes, and can provide protection against pathogens in a non-specific way, for example challenge of mice by *C. albicans* was shown to provide macrophage mediated immunity against *S. aureus*[52,53]. In light of the ineffectiveness of the B-cell response against *Campylobacter,* its tendency to co-colonise and ability to undergo phase variation, quasi adaptive γδ T cell responses may represent a strategy for mounting responses against a gut pathogen with such high antigenic diversity, therefore maintaining anti-*Campylobacter* immunity[10,54,55].

This potential expansion of largely oligoclonal γδ T cells is consistent with previous *Campylobacter* challenge studies in chicken which have concluded that it is therefore likely there is a γδ response to *Campylobacter* in chickens, though an experiment looking at this directly is yet to be published. However, there is strong evidence that *Campylobacter* colonisation leads to upregulation of various pro-inflammatory mediators such as CXCLi1/2, IFN-γ and IL17A[3–6]. It has been hypothesised by multiple groups that as the major producers of IL17 it is likely that γδ T cells are involved in the immune response against *Campylobacter*[3,6]. Additionally, B lymphocytes and therefore adaptive antibody mediated immunity has been shown to play a limited role in clearance of *Campylobacter* through study of infected birds that have been subject to bursectomy[10]. Consistent with our findings that the differences in the *Campylobacter* infected groups were from invariant γδ T cells, it has been proposed that previously reported upregulation of IL17 may be driven by TLR dependant mechanisms and therefore the γδ TCR may not be important in initiating these responses[7]. However, it is currently unclear how our data fit in with these previous studies, additional work on sorted γδ subsets will allow the linkage of our TRGV expression with previous work on cytokine expression of CD8 sorted subsets[45].

To examine the impact of *Campylobacter* colonisation on the microbiota, 16S sequencing was used to follow the microbiotas of the different groups. Using faecal samples enabled us to follow the microbiota over time in a single bird. It has been shown that faecal microbiome analyses

represent 99.25% of all bacterial community members present in the caecal microbiota, indicating that the effects of conditions on the caecal microbiomes are likely to be accurately represented in faecal analyses, particularly since there was a correlation in the abundance of shared sequences[56]. Despite this, and as we have previously observed (See Chapter 5), the same study showed that despite the same microbes being present, their abundances varied drastically.

The microbiota profiles observed over time resembled the general patterns reported in previous studies[34,35] despite the initial use of Broilact, a probiotic supplement derived from the microbiota of an SPF hen, to normalise the start of life microbiota received by chicks. It should be noted that Broilact seemed to have no effect on our ability to infect the birds with *Campylobacter.* The dominance of the *Firmicutes* was evident at every time point examined, as well as the decline in *Proteobacteria* over time previously observed by ourselves (See Chapter 5) and others[34,35]. Despite this, the proportion of *Bacteroidetes* we observed was far less consistent than reported in other work[34,35] and rather than gradually increasing over time, it appeared to start high, decrease, then increase a little towards the end of the study, but was highly variable between birds. This could an effect of the Broilact treatment providing an early influx of *Bacteroidetes* which will be high in the microbiota of the mature donor hen[35]. Despite the use of Broilact, the microbiotas observed in this study still varied considerably from bird to bird.

In spite of this confounding microbiome variability, we observed differences between the groups in terms of β-diversity and composition at the Order level. Our data shows that not only does *Campylobacter* colonisation of the chicken gut cause microbiota differences, but that these changes differ in mixed vs single strain infections and remarkably also differ depending on the co-colonising strains that are present.

Considering the microbiota composition at the Order level, there was a significantly higher proportion of *Lactobacillales* in the CTRL group, but only at certain timepoints and with a decrease in proportion over time. Similar fluctuations in *Lactobacillales* associated with an overall decrease in

proportion over time have been reported in previous studies[57,58]. As the proportions of *Lactobacilalles* are far lower in caecal contents and caecal excretion into the gut occurs several times per day, it is also possible that the levels of *Lactobacillales* are decreased in the faecal samples of birds who have recently emptied their caeca. This could also explain the finding that these differences were observed even when birds are sampled just 2 days after a high proportion of *Lactobacilalles* was found, as with the compositions seen in Day 15 and Day 17 CTRL birds. However, we observed far less of the characteristic caecal microbiome profiles we reported previously (See Chapter 5) due to an attempt to avoid collecting faecal samples containing caecal excreta, identifiable by consistency, and therefore this is more likely to reflect actual variation. It has been shown that the microbiome can undergo large scale changes over these sorts of timescales particularly in response to food intake and other ingested substances[59].

Antagonistic interactions between *Lactobacillus* and *Campylobacter* are well documented in the literature, due to this and other beneficial effects, *Lactobacillus* is a common probiotic in the chicken industry[60–64]. The mechanism driving this antagonism appears to differ with *Lactobacillus* strain and may include Competitive exclusion[65,66], co-aggregation[65,67,68], production of antimicrobial bacteriocins[69,70] and lactate production[71]. It is well established in the literature that *Lactobacillus* reduces *Campylobacter* load in these and other studies. However, the results of this chapter suggest that this relationship also works in reverse as we observed reduced *Lactobacillus* in all of the *Campylobacter* infected groups.

More specifically, the finding that *Lactobacillus johnsonii* was almost totally absent from any of the *Campylobacter* inoculated groups, whereas other species, such as *Lactobacillus salivarius,* remained present in these groups, reinforces the point that different *Lactobacillus* species appear to interact with *Campylobacter* in different ways. *Lactobacillus johnsonii* has been shown to reduce *Campylobacter* load in chickens by enhancing phagocytic activity and macrophage expression of IFN-γ, the proinflammatory cytokines IL-1β, IL12p40, CXCLi2 and the regulatory cytokine IL-10[72].

However, the fact that it was undetectable in many of the *Campylobacter* infected samples suggests that *Campylobacter* may be able to employ counter measures and instead eliminate certain species of *Lactobacillus* from the microbiota. The mechanisms that may underlie this process are as yet unclear, but may be of interest in maximising the effectiveness of these probiotics.

This finding of mutual elimination from the microbiota is consistent with the competitive exclusion and co-aggregation mechanisms, where the antagonism is a two-way interaction, however this does not explain how *Campylobacter,* which was shown to be generally rare in the gut flora of infected birds, is able to overcome the numerical advantage of *Lactobacillus* species*,* which were present at a far higher proportion. Though there is a lack of evidence for widespread use of microbial weapons by *Campylobacter,* it is known to produce a Type 6 Secretion System (T6SS), a molecular injection system often used in microbial warfare[73–75]. Though this T6SS was shown to differ in structure from those of other microbes and to be involved protection against oxidative stress, it did enhance colonisation of chickens and cytotoxicity in the *Galleria mellonella* invertebrate infection model, potentially due to an enhanced ability to compete with and kill other microbes[75]. Though to our knowledge, the ability of the *Campylobacter* T6SS to deliver toxins to other microbes is yet to be shown directly, this is the primary function of the T6SS in other microbes[74] and is one potential mechanism by which a smaller population of *Campylobacter jejuni* might be able to reduce and in some cases eliminate *Lactobacillus* species.

Additionally, we showed that the MIG group had increased *Clostridiales* at the Order level, corresponding to higher proportions of the genus *Romboutsia. Romboutsia* was discovered relatively recently and all species within it are obligate anaerobes, more commonly associated with the mammalian microbiota[76]. However, since its discovery, multiple studies have reported its existence in chickens[35,77–79]. It was recently shown to be decreased during infection studies with the inflammation inducing protozoan parasite *Eimeria*[80]*.* This study described it as 'associated with the health status of the gastrointestinal tract' and though this appears to be true of mice and rats, its

role in chicken remains unclear. Despite this, these rodent studies provide the current best guess at its role in the chicken microbiota. *Romboutsia* has been associated with both amino acid and energy metabolism, producing Short Chain Fatty Acids (SCFA) in the gut such as citrate, pyruvate, creatine and lactate[81]. This is consistent with our pseudo-metagenomics data, which showed that the MIG group likely had increased levels of proteins associated with anaerobic respiration, including NADH-quinone oxidoreductase, putative formate acetyltransferase, flavodoxin oxidoreductase and acetaldehyde dehydrogenase. The increased levels of these genes may represent the metabolism of the microbiota shifting towards carbohydrate metabolism, as *Campylobacter* mostly utilises amino acids as an energy source. Paradoxically, the increase in *Romboutsia* and these genes may actually confer a benefit to the host due to increased availability of SCFAs which are an important energy source.

The main difference in CIG2, occurred at Day 28 in which an influx of the Order *Erysipelotrichales* made up around 50% of the microbiota in this group on average. Though this was a large difference compared to the other groups , they showed similar influxes at Day 35, and therefore this change most likely just occurred early in CIG2 and therefore does not constitute the clear difference it at first seems to. At the genus level, *Erysipelotrichales* is dominated by the genus *Turicibacter,* which has been associated with high productivity as determined by reduced Residual Feed Intake (RFI)[82]. However, due to its recent discovery, knowledge of the effects of this genus is limited and therefore it is difficult to say what the effects of this earlier influx of *Turicibacter* may be. These differences reveal that the *Campylobacter* strain composition of a mixed infection can in turn influence the development of the microbiota, though whether or not this is a deliberate strategy by certain strains or simply due to host responses and resulting cross reactivity remains to be shown.

One unexpected finding was that in the case of CIG1, *Campylobacter* itself was a contributor to microbiome differences observed in this group. Interestingly, this was the group made up of *Campylobacter* strains which naturally colonised together. In the bird these strains were initially

isolated from, PorA type 400 was dominant, occupying 52% of colonies picked from plated caecal contents. However, when the four strains were administered in equal quantities to experimental birds, the dominant porA type was 33 in all birds but one. This in itself was an interesting result as it suggests that the most dominant strain does not necessarily remain dominant in every infection. Mid-infection flips in *Campylobacter* co-infections have been reported in the literature where the proportion of one strain decreases rapidly and another increases concordantly[83]. This pattern of switching out the dominant infecting strain in the course of a multi strain co-infection has also been reported in *Plasmodium* and is thought to be a method of maintaining an infection despite strong immune responses against variable surface antigens[84,85]. It should be considered however, that the PorA results we observed may be due to more simple stochastic founder effects due to population bottlenecks which occur during transit through the gastrointestinal tract following inoculation by gavage[86].

As competition is often highest between related bacteria and it has been shown that *Campylobacter* strains competitively inhibit one another's binding to chicken cells *in vitro*, it is unlikely that this strain switching phenomenon reflects cooperation and as has been previously suggested it seems likely to be immune driven[87]. We take this idea one step further and show that colonisation is higher in mixed strain infections which normally colonise together in a real world setting. Additionally, we would hypothesise that the strains which are able to colonise together most frequently are more likely to be genetically diverse, thus decreasing the likelihood of cross reactivity with the dominant strain which the immune system is currently responding to. Additionally, colonisation by multiple genetically diverse strains which can swap in and out as the dominant strain may be one factor in influencing whether or not an infection goes florid, thus causing meat contamination during slaughter. Whole Genome Sequences (WGS) could provide a better picture of the diversity of these strains and may provide an initial indication of whether or not co-colonisation of genetically diverse strains confers a selective advantage in the host.

Though it was previously thought that *Campylobacter* was fully unable to utilise sugar or carbohydrates and therefore relied solely on amino acids as an energy source, more recent evidence has shown it is able to make use of a limited range of sugars[88,89]. However, it remains widely accepted that the majority energy source for *Campylobacter* is amino acids, namely serine, aspartate, asparagine, glutamate and proline, in order of metabolic preference[90]. Therefore, it comes as no surprise that in the CIG1 group, which saw by far the highest *Campylobacter* load, we also observed an increase in sugar metabolism genes such as Beta-galactosidase, hexosaminidase and beta-glucosidase as these energy sources are little utilised by *Campylobacter,* which made up a fair portion of the microbiota in this strain, particularly at certain time points. As this group had high levels of amino acid utilising *Campylobacter,* the supply of sugar would most likely increase, leading to an increase in the microbiota members expressing these sugar metabolism genes. Additionally, in a piglet infection model, C. *jejuni* was shown to only be able to utilise L-fucose in the presence of another bacterium capable of cleaving it off glycans using a fucosidase, such as *Bacteroidetes*[91]. This provides some context to the increased *Bacteroides* we observed in this group and could provide an additional driver of high *Campylobacter* loads.

6.5 Conclusion

This chapter aimed to use TCR gamma repertoire and 16S sequencing to investigate the effects of the inflammatory, dysbiosis causing pathogen *Campylobacter* and how different co-infecting strain mixtures might lead to different effects. The data presented here shows that all of the *Campylobacter* infection groups led to an increase in canonical, γδ T cell subsets, possibly representing a quasi-adaptive response, characterised by expansions in invariant γδ T cells[46]. However, each of the *Campylobacter* infection groups led to a different effect on the TRGV gene usage profiles and did not all cluster together in β-diversity plots. The impact on the microbiota was different depending on the composition of the co-infection mixture, with each leading to significant separate clustering in 16S β-diversity plots corresponding to differences at the Order and Genus

level. These results suggest that the strains which make up a mixed *Campylobacter* infection have real implications in terms of the γδ T cells and microbiota. These differences may determine the outcome of the infection and crucially whether one strain takes over, causing spread throughout a flock and carcass contamination at slaughter.

6.6 References

1. Tanoue T, Umesaki Y, Honda K. Immune responses to gut microbiota-commensals and pathogens. Gut Microbes. 2010;1(4):224–33.
2. Humphrey S, Chaloner G, Kemmett K, Davidson N, Williams N, Kipar A, et al. Campylobacter jejuni is not merely a commensal in commercial broiler chickens and affects bird welfare. MBio. 2014;5(4):1–7.
3. Reid WDK, Close AJ, Humphrey S, Chaloner G, Lacharme-Lora L, Rothwell L, et al. Cytokine responses in birds challenged with the human food-borne pathogen Campylobacter jejuni implies a Th17 response. R Soc Open Sci. 2016;3(3).
4. Connerton PL, Richards PJ, Lafontaine GM, O'Kane PM, Ghaffar N, Cummings NJ, et al. The effect of the timing of exposure to Campylobacter jejuni on the gut microbiome and inflammatory responses of broiler chickens. Microbiome. 2018;6(1):88.
5. John DA, Williams LK, Kanamarlapudi V, Humphrey TJ, Wilkinson TS. The bacterial species Campylobacter jejuni Induce diverse innate immune responses in human and avian intestinal epithelial cells. Front Microbiol. 2017;8(SEP):1–15.
6. Shaughnessy RG, Meade KG, McGivney BA, Allan B, O'Farrelly C. Global gene expression analysis of chicken caecal response to Campylobacter jejuni. Vet Immunol Immunopathol. 2011;142(1–2):64–71.
7. De Zoete MR, Keestra AM, Roszczenko P, Van Putten JPM. Activation of human and chicken toll-like receptors by Campylobacter spp. Infect Immun. 2010;78(3):1229–38.
8. Iqbal M, Philbin VJ, Smith AL. Expression patterns of chicken Toll-like receptor mRNA in tissues, immune cell subsets and cell lines. Vet Immunol Immunopathol. 2005;104(1–2):117–27.
9. Brownlie R, Zhu J, Allan B, Mutwiri GK, Babiuk LA, Potter A, et al. Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. Mol Immunol. 2009;46(15):3163–70.
10. Lacharme-Lora L, Chaloner G, Gilroy R, Humphrey S, Gibbs K, Jopson S, et al. B lymphocytes play a limited role in clearance of Campylobacter jejuni from the chicken intestinal tract. Sci Rep. 2017;7:2–11.
11. Thibodeau A, Fravalo P, Yergeau E, Arsenault J, Lahaye L, Letellier A. Chicken Caecal Microbiome modifications induced by campylobacter Jejuni colonization and by a non-antibiotic feed additive. PLoS One. 2015;10(7):1–14.
12. Johansen CH, Bjerrum L, Finster K, Pedersen K. Effects of a Campylobacter jejuni infection on the development of the intestinal microflora of broiler chickens. Poult Sci. 2006;85(4):579–

87.

13. McKenna A, Ijaz UZ, Kelly C, Linton M, Sloan WT, Green BD, et al. Impact of industrial production system parameters on chicken microbiomes: Mechanisms to improve performance and reduce Campylobacter. Microbiome. 2020;8(1):1–13.

14. Kaakoush NO, Sodhi N, Chenu JW, Cox JM, Riordan SM, Mitchell HM. The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. Gut Pathog. 2014;6(1):1–10.

15. Skånseng B, Kaldhusdal M, Rudi K. Comparison of chicken gut colonisation by the pathogens Campylobacter jejuni and Clostridium perfringens by real-time quantitative PCR. Mol Cell Probes. 2006;20(5):269–79.

16. Line JE, Hiett KL, Guard-Bouldin J, Seal BS. Differential carbon source utilization by Campylobacter jejuni 11168 in response to growth temperature variation. J Microbiol Methods. 2010;80(2):198–202.

17. van Deun K, Pasmans F, Van Immerseel F, Ducatelle R, Haesebrouck F. Butyrate protects Caco-2 cells from. Br J Nutr. 2008;100(3):480–4.

18. Hermans D, Van Deun K, Martel A, Van Immerseel F, Messens W, Heyndrickx M, et al. Colonization factors of Campylobacter jejuni in the chicken gut. Vet Res. 2011;42(1):1–14.

19. Thomas M, Wrzosek L, Miquel S, Noordine M-L, Bouet S, Chevalier-Curt MJ, et al. 264 Bacteroides Thetaiotaomicron and Faecalibacterium prausnitzii Shape the Mucus Production and Mucin O-Glycosylation in Colon Epithelium. Gastroenterology. 2013;144(5):S-59.

20. Hiett KL, Stern NJ, Cox NA, Musgrove MT, Ladely S. Molecular Subtype Analyses of Campylobacter spp . from Arkansas and California Poultry Operations. 2002;68(12):6220–36.

21. Thomas LM, Long KA, Good RT, Panaccio M, Widders PR. Genotypic Diversity among Campylobacter jejuni Isolates in a Commercial Broiler Flock. 1997;63(5):1874–7.

22. Emanuelson U, Engvall A. A l-year epidemiological study of campylobacters in 18 Swedish chicken farms. 1995;26(95):167–85.

23. Jacobs-Reitsma WF, Giessen AWVD, Bolder NM, Mulder RWAW. Epidemiology of Campylobacter spp . at two Dutch broiler farms. 1995;413–21.

24. Conlan AJK, Coward C, Grant AJ, Maskell DJ, Gog JR. Campylobacter jejuni colonization and transmission in broiler chickens : a modelling perspective. 2007;(May):819–29.

25. Persson S, Olsen KEP. Multiplex PCR for identification of Campylobacter coli and Campylobacter jejuni from pure cultures and directly on stool samples. 2005;(54):1043–7.

26. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4).

27. Colles FM, Hedges SJ, Dixon R, Preston SG, Thornhill P, Barfod KK, et al. Deep sequencing reveals Campylobacter in commercial meat chickens less than 8 days old. bioRxiv. 2020;1–14.

28. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2 : High resolution sample inference from amplicon data. Nat Methods. 2015;13(7):0–14.

29. Guarner F. The gut microbiota era marches on. Nat Rev Gastroenterol Hepatol. 2014;11(11):647–9.

30. Rosen MJ, Callahan BJ, Fisher DS, Holmes SP. Denoising PCR-amplified metagenome data.

BMC Bioinformatics. 2012;13(1).

31. Tikhonov M, Leach RW, Wingreen NS. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. ISME J. 2015;9(1):68–80.

32. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 2015;9:968–79.

33. Mcelroy K, Zagordi O, Bull R, Luciani F, Beerenwinkel N. Accurate single nucleotide variant detection in viral populations by combining probabilistic clustering. 2013;

34. Díaz-Sánchez S, Perrotta AR, Rockafellow I, Alm EJ, Okimoto R, Hawken R, et al. Using fecal microbiota as biomarkers for predictions of performance in the selective breeding process of pedigree broiler breeders. PLoS One. 2019;14(5):e0216080.

35. Videnska P, Sedlar K, Lukac M, Faldynova M, Gerzova L, Cejkova D, et al. Succession and replacement of bacterial populations in the caecum of egg laying hens over their whole life. PLoS One. 2014;9(12):1–14.

36. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613–7.

37. Park S-C, Won S. Evaluation of 16S rRNA Databases for Taxonomic Assignments Using a Mock Community. Genomics Inform. 2018;16(4):e24.

38. Bower NI, Johnston IA. Targeted rapid amplification of cDNA ends (T-RACE) - An improved RACE reaction through degradation of non-target sequences. Nucleic Acids Res. 2010;38(21).

39. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2017;19(4):554–65.

40. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1).

41. Hankel J, Jung K, Kuder H, Keller B, Keller C, Galvez E, et al. Caecal Microbiota of Experimentally Campylobacter jejuni-Infected Chickens at Different Ages. Front Microbiol. 2019;10:1–11.

42. Mortada M, Cosby DE, Akerele G, Ramadan N, Oxford J, Shanmugasundaram R, et al. Characterizing the immune response of chickens to campylobacter jejuni (strain a74c). PLoS One. 2021;16:1–21.

43. Hermans D, Van Deun K, Messens W, Martel A, Van Immerseel F, Haesebrouck F, et al. Campylobacter control in poultry by current intervention measures ineffective: Urgent need for intensified fundamental research. Vet Microbiol. 2011;152(3–4):219–28.

44. Rawson T, Paton RS, Colles FM, Maiden MCJ, Dawkins MS, Bonsall MB. A Mathematical Modeling Approach to Uncover Factors Influencing the Spread of Campylobacter in a Flock of Broiler-Breeder Chickens. Front Microbiol. 2020;11:1–17.

45. Pieper J, Methner U, Berndt A. Heterogeneity of avian γδ T cells. Vet Immunol Immunopathol. 2008;124(3–4):241–52.

46. Hayday AC. γδ T Cell Update: Adaptate Orchestrators of Immune Surveillance. J Immunol. 2019;203(2):311–20.

47. Dieli F, Sireci G, Di Sano C, Champagne E, Fourniè JJ, Ivanyi J, et al. Predominance of Vγ9/Vδ2 T lymphocytes in the cerebrospinal fluid of children with tuberculous meningitis: Reversal after chemotherapy. Mol Med. 1999;5(5):301–12.

48. Sheridan BS, Romagnoli PA, Pham QM, Fu HH, Alonzo F, Schubert WD, et al. γδ T Cells Exhibit Multifunctional and Protective Memory in Intestinal Tissues. Immunity. 2013;39(1):184–95.

49. Simonian PL, Roark CL, Diaz del Valle F, Palmer BE, Douglas IS, Ikuta K, et al. Regulatory Role of γδ T Cells in the Recruitment of CD4 + and CD8 + T Cells to Lung and Subsequent Pulmonary Fibrosis . J Immunol. 2006;177(7):4436–43.

50. Bekkering S, Arts RJW, Novakovic B, Kourtzelis I, van der Heijden CDCC, Li Y, et al. Metabolic Induction of Trained Immunity through the Mevalonate Pathway. Cell. 2018;172(1–2):135-146.e9.

51. Moorlag SJCFM, Röring RJ, Joosten LAB, Netea MG. The role of the interleukin-1 family in trained immunity. Immunol Rev. 2018;281(1):28–39.

52. Bistoni F, Vecchiarelli A, Cenci E, Puccetti P, Marconi P, Cassone A. Evidence for macrophage-mediated protection against lethal Candida albicans infection. Infect Immun. 1986;51(2):668–74.

53. Belibasakis GN, Hajishengallis G, Bostanci N, Curtis M. Oral Mucosal Immunity and Microbiome. Vol. 1197, Advances in experimental medicine and biology. 2019. 11–26 p.

54. Guerry P, Szymanski CM, Prendergast MM, Hickey TE, Ewing CP, Pattarini DL, et al. Phase variation of Campylobacter jejuni 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. Infect Immun. 2002;70(2):787–93.

55. Bayliss CD, Bidmos FA, Anjum A, Manchev VT, Richards RL, Grossier JP, et al. Phase variable genes of Campylobacter jejuni exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. Nucleic Acids Res. 2012;40(13):5876–89.

56. Stanley D, Geier MS, Chen H, Hughes RJ, Moore RJ. Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. BMC Microbiol. 2015;15(1):1–11.

57. Jurburg SD, Brouwer MSM, Ceccarelli D, van der Goot J, Jansman AJM, Bossers A. Patterns of community assembly in the developing chicken microbiome reveal rapid primary succession. Microbiologyopen. 2019;8(9):1–10.

58. Qiao H, Song Y, Shi H, Bian C. Fermented astragalus in diet altered the composition of fecal microbiota in broiler chickens. AMB Express. 2018;8(1):1–12.

59. Uhr GT, Dohnalová L, Thaiss CA. The Dimension of Time in Host-Microbiome Interactions. mSystems. 2019;4(1):1–7.

60. Koenen ME, Kramer J, Van Der Hulst R, Heres L, Jeurissen SHM, Boersma WJA. Immunomodulation by probiotic lactobacilli in layer- And meat-type chickens. Br Poult Sci. 2004;45(3):355–66.

61. Brisbin JT, Gong J, Orouji S, Esufali J, Mallick AI, Parvizi P, et al. Oral treatment of chickens with lactobacilli influences elicitation of immune responses. Clin Vaccine Immunol. 2011;18(9):1447–55.

62. Garriga M, Pascual M, Monfort JM, Hugas M. Selection of lactobacilli for chicken probiotic adjuncts. J Appl Microbiol. 1998;84(1):125–32.

63. Wang J, Ishfaq M, Guo Y, Chen C, Li J. Assessment of Probiotic Properties of Lactobacillus salivarius Isolated From Chickens as Feed Additives. Front Vet Sci. 2020;7(July):1–12.

64. Kabir SML. The role of probiotics in the poultry industry. Int J Mol Sci. 2009;10(8):3531–46.

65. Nishiyama K, Seto Y, Yoshioka K, Kakuda T, Takai S, Yamamoto Y, et al. Lactobacillus gasseri sbt2055 reduces infection by and colonization of campylobacter jejuni. PLoS One. 2014;9(9):1–9.

66. Wine E, Gareau MG, Johnson-Henry K, Sherman PM. Strain-specific probiotic (Lactobacillus helveticus) inhibition of Campylobacter jejuni invasion of human intestinal epithelial cells. FEMS Microbiol Lett. 2009;300(1):146–52.

67. Schachtsiek M, Hammes WP, Hertel C. Characterization of Lactobacillus coryniformis DSM 20001T surface protein Cpf mediating coaggregation with and aggregation among pathogens. Appl Environ Microbiol. 2004;70(12):7078–85.

68. Tareb R, Bernardeau M, Gueguen M, Vernoux JP. In vitro characterization of aggregation and adhesion properties of viable and heat-killed forms of two probiotic Lactobacillus strains and interaction with foodborne zoonotic bacteria, especially Campylobacter jejuni. J Med Microbiol. 2013;62:637–49.

69. Stern NJ, Svetoch EA, Eruslanov B V., Perelygin V V., Mitsevich E V., Mitsevich IP, et al. Isolation of a Lactobacillus salivarius strain and purification of its bacteriocin, which is inhibitory to Campylobacter jejuni in the chicken gastrointestinal system. Antimicrob Agents Chemother. 2006;50(9):3111–6.

70. Campana R, Federici S, Ciandrini E, Baffone W. Antagonistic activity of lactobacillus acidophilus ATCC 4356 on the growth and adhesion/invasion characteristics of Human Campylobacter jejuni. Curr Microbiol. 2012;64(4):371–8.

71. Neal-McKinney JM, Lu X, Duong T, Larson CL, Call DR, Shah DH, et al. Production of Organic Acids by Probiotic Lactobacilli Can Be Used to Reduce Pathogen Load in Poultry. PLoS One. 2012;7(9).

72. Taha-Abdelaziz K, Astill J, Kulkarni RR, Read LR, Najarian A, Farber JM, et al. In vitro assessment of immunomodulatory and anti-Campylobacter activities of probiotic lactobacilli. Sci Rep. 2019;9(1):1–15.

73. Granato ET, Meiller-Legrand TA, Foster KR. The Evolution and Ecology of Bacterial Warfare. Curr Biol. 2019;29(11):R521–37.

74. Navarro-Garcia F, Ruiz-Perez F, Cataldi Á, Larzábal M. Type VI secretion system in pathogenic escherichia coli: Structure, role in virulence, and acquisition. Front Microbiol. 2019;10:1–17.

75. Liaw J, Hong G, Davies C, Elmi A, Sima F, Stratakos A, et al. The Campylobacter jejuni Type VI Secretion System Enhances the Oxidative Stress Response and Host Colonization. Front Microbiol. 2019;10:1–12.

76. Gerritsen J, Fuentes S, Grievink W, van Niftrik L, Tindall BJ, Timmerman HM, et al. Characterization of Romboutsia ilealis gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus Clostridium into the genera Romboutsia gen. nov., Intestinib. Int J Syst Evol Microbiol. 2014;64:1600–16.

77. Glendinning L, Watson KA, Watson M. Development of the duodenal, ileal, jejunal and caecal microbiota in chickens. bioRxiv. 2019;1–11.

78. Qiao H, Shi H, Zhang L, Song Y, Zhang X, Bian C. Effect of Lactobacillus plantarum supplementation on production performance and fecal microbial composition in laying hens. Open Life Sci. 2019;14(1):69–79.

79. Maki JJ, Bobeck EA, Sylte MJ, Looft T. Eggshell and environmental bacteria contribute to the intestinal microbiota of growing chickens. J Anim Sci Biotechnol. 2020;11(1):1–17.

80. Chen HL, Zhao XY, Zhao GX, Huang H Bin, Li HR, Shi CW, et al. Dissection of the cecal microbial community in chickens after Eimeria tenella infection. Parasites and Vectors. 2020;13(1):1–15.

81. Gao H, Jiang Q, Ji H, Ning J, Li C, Zheng H. Type 1 diabetes induces cognitive dysfunction in rats associated with alterations of the gut microbiome and metabolomes in serum and hippocampus. Biochim Biophys Acta - Mol Basis Dis. 2019;1865(12):165541.

82. Carrasco JMD, Casanova NA, Miyakawa MEF. Microbiota, gut health and chicken productivity: What is the connection? Microorganisms. 2019;7(10):1–15.

83. Skånseng B, Trosvik P, Zimonja M, Johnsen G, Bjerrum L, Pedersen K, et al. Co-infection dynamics of a major food-borne zoonotic pathogen in chicken. PLoS Pathog. 2007;3(11):1790–7.

84. Bruce MC, Day KP. Cross-species regulation of malaria parasitaemia in the human host. Curr Opin Microbiol. 2002;5(4):431–7.

85. Bruce MC, Donnelly CA, Alpers MP, Galinski MR, Barnwell JW, Walliker D, et al. Cross-species interactions between malaria parasites in humans. Science (80-). 2000;287(5454):845–8.

86. Coward C, Van Diemen PM, Conlan AJK, Gog JR, Stevens MP, Jones MA, et al. Competing isogenic Campylobacter strains exhibit variable population structures in vivo. Appl Environ Microbiol. 2008;74(12):3857–67.

87. Konkel ME, Christensen JE, Dhillon AS, Lane AB, Hare-Sanford R, Schaberg DM, et al. Campylobacter jejuni strains compete for colonization in broiler chicks. Appl Environ Microbiol. 2007;73(7):2297–305.

88. Stahl M, Friis LM, Nothaft H, Liu X, Li J, Szymanski CM, et al. L-fucose utilization provides Campylobacter jejuni with a competitive advantage. Proc Natl Acad Sci U S A. 2011;108(17):7194–9.

89. Muraoka WT, Zhang Q. Phenotypic and genotypic evidence for L-fucose utilization by Campylobacter jejuni. J Bacteriol. 2011;193(5):1065–75.

90. Wright JA, Grant AJ, Hurd D, Harrison M, Guccione EJ, Kelly DJ, et al. Metabolite and transcriptome analysis of Campylobacter jejuni in vitro growth reveals a stationary-phase physiological switch. Microbiology. 2009;155(1):80–94.

91. Garber JM, Nothaft H, Pluvinage B, Stahl M, Bian X, Porfirio S, et al. The gastrointestinal pathogen Campylobacter jejuni metabolizes sugars with potential help from commensal Bacteroides vulgatus. Commun Biol. 2020;3(1):1–11.

Department of Zoology, University of Oxford

Chapter 7

Interactions between γδ T Cells and the Enteric Microbiota in the Mouse

7.1 Introduction

In mice, γδ T cells are present in the blood at relatively low levels, making up between 0.5-10% of the circulating T cell pool[1]. However, γδ T cells are greatly enriched at the mucosa, and are at their highest proportion in the gut, as part of the intraepithelial lymphocyte (IEL) population. Most γδ T cells in mouse develop in the thymus where they arise from a common αβ and γδ double negative (DN) T-cell progenitor. Following development, γδ T cells localise to various peripheral organs where they fulfil a wide range of functions including tumour surveillance[2], anti-pathogen effects and immune regulatory roles[3].

Multiple T Receptor Gamma (TRG) V naming conventions exist in mouse and for the sake of clarity, the international ImMunoGeneTics (IMGT) gene names will be used in this chapter. The reader should refer to the IMGT conversion table for easy interpretation and conversion of other nomenclatures[4]. Mouse IEL compartments are populated by γδ T cells expressing distinct subsets of TRGV gene segments which tend to make up a majority of TCR gamma chain rearrangements at particular anatomical sites within the body: Vγ5 in the skin, Vγ6 in the uterus and Vγ7 in the gut[5–7]. This organization is likely mediated by a range of organ specific products produced by epithelial cells. A candidate for a restricting molecule capable of mediating this type of γδ T cell compartmentalization is skint1. Skint1 was discovered in a mouse strain lacking the skin dendritic epidermal T cells, a subset of γδ T cells, and became the founding member of a larger family of 11 butyrophilin-like (Btnl) genes, which show strong similarity to the multifunctional butyrophilin (Btn) genes[8,9].

Butyrophilins are type 1 membrane proteins in the Ig superfamily[9]. The classical structure of a Butyrophilin comprises 3 domains, each encoded by a single exon; the extracellular domain, the intracellular domain and the cytoplasmic domain. This structure however, is poorly conserved, with some members lacking the extracellular or intracellular domains. The extracellular region of the Butyrophilin family has structural homology to B7 proteins, which play a role in regulation of T cell activation and tolerance, and have a genetic association with inflammatory diseases[10]. Phylogenetic

studies show that the Butyrophilin family shares a common ancestor with the B7 proteins, which was one of the early indicators of their immunological function[11]. There are 11 Btn genes in mouse, 6 of which are orthologues of human Btns; Btn1, Btn 2a2, Btnl2, Btnl9, Ermap, and Mog. The remaining 5 are only found in mouse; Btnl1, Btnl4, Btnl5, Btnl6 and Btnl7[12]. Btnl1 has been shown to play a role in selectively expanding Vγ7+ve T cells in the gut, contributing to the aforementioned tissue tropism shown by different TRGV segments[7]. However, the more precise impacts of Btnl1 on the TCR repertoire are yet to be elucidated.

Previous studies have shown that these gut γδ T cells play an important role in facilitating intestinal homeostasis. γδ T cells are involved in epithelial cell turnover and regeneration in the small intestine as shown by the reduced epithelial cell proliferation that has been observed in TCRδ -/- mice[13,14]. This regeneration process is thought to be mediated through γδ T cell production of Keratinocyte Growth Factor (KGF)-1, which differentiates them from other T cell populations[15]. Additionally, γδ T cells have a role in maintaining intestinal integrity and disruption of Tight Junction (TJ) protein complexes have been observed in TCRδ -/- mice, leading to increased permeability and susceptibility to colitis[16]. γδ T cells maintain surveillance over a large number of gut epithelial cells by migrating through them via occludin mediated cell to cell contact, however the stimulus which leads to their activation and production of growth factors is yet to be identified[17].

An additional function of enteric γδ T cells is responses to pathogens. γδ T cells have been shown to produce antimicrobial peptides (AMPs) such as Regenerating islet-derived protein 3 gamma (RegIIIγ), a C type lectin capable of lysing gram positive bacteria through formation of hexameric pores in the cell wall[18]. RegIIIγ production by γδ T cells is regulated by intestinal epithelial cells. γδ T cells also produce granzyme A and B, serine proteases involved in inducing apoptosis of infected host cells, thus contributing to antibacterial defense[19,20]. γδ T cells produce similar cytolytic proteins to αβ T cells, and have been shown to be involved in protective responses against infection by pathogens such as *Toxoplasma*[16,19]. This γδ T cell activation and subsequent release of

antimicrobial factors was shown to be mediated by epithelial cells which seem to act as something of a surveillance system and facilitate γδ T cell motility and chemotaxis[18,21]. γδ T cells in mouse have been shown to differ in terms of their effector functions on the basis of the TCR gamma chain they express. These lymphocytes can be split into two main subsets on the basis of their effector functions IFNγ (TRGV1, 5, 7) and IL17 (TRGV4, 6) producers[22,23]. IFNγ producing γδ T cells tend to be involved in anti-tumour and intracellular pathogen responses[24–26] whereas IL17 producers play a role in responses to extracellular pathogens and immunoregulation[27–30].

Despite their enrichment in the gut and presence in the enteric epithelia, coupled with their ability to induce and regulate inflammation, evidence of the ability of γδ T cells to influence the microbiota under homeostatic conditions is surprisingly limited. However, in mice lacking γδ T cells, mucosal penetration by commensal microbes was increased following intestinal injury[31]. Thus, further study into the potential involvement of γδ T cells in interactions with the microbiota and subsequent maintenance of homeostasis is needed.

7.2 Chapter Outline

In this chapter, TCR gamma repertoire approaches were employed to examine the biology of gut γδ T cells. γδ deficient mice were used to examine the effect of these cells on the microbiome composition. Firstly, I will consider the nature of the γδ T cell repertoire in the small and large intestine of mice, using both global repertoire tools and approaches more focussed on the gut enriched TRGV7 gene segment. From the perspective of repertoire, I will also consider the effects of deficiency in the Btnl1 molecule in greater detail. Furthermore, this provides confirmation of the ability of our repertoire sequencing approach to replicate findings from other immunology techniques, such as flow cytometry. Nude mice have been used to examine the requirement for the thymus in T cell development and although these animals are totally lacking in αβ T cells, they do retain some γδ T cells, which has previously been used to suggest athymic development[32]. Here I will test for the influence of the nude mutation on the repertoire of gut γδ T cells. Finally I will consider

the influence of deficiency in all γδ T cells (TCRδ -/-), reduced TRGV7+ T cells (Bntl1 -/-) and the nude mutation, on the microbial composition of the mouse gut microbiome.

7.3 Results

7.3.1 Mouse Tissue Extraction and PCR Amplification of TCR Gamma Repertoire using Universal Primers

The mouse gut tissues used in this chapter were kindly provided by the Hayday group. These tissues had been extracted from 9 week old C57BL/6 Wild Type, Btnl1 knockout (henceforth referred to as Butyrophilin -/-), TCRδ knockout (henceforth referred to at TCRδ-/-) and nude mice raised in SPF conditions. After sacrifice, tissue samples were taken from the duodenum, jejunum, ileum and caecum. These tissues were extracted from 3 mice from each of the Wild Type, Butyrophilin -/- and TCRδ -/- strains and 2 mice from the Nude strain. Nucleic acids were extracted from these tissues using the RNA/DNA AllPrep micro kit (Qiagen). The RNA was then converted into cDNA using a 5' RACE protocol and the DNA stored to act as the template for later 16S rRNA amplification. PCR was then used to amplify the gamma region from the gut tissue cDNA (Fig 7.1). The majority of samples show the expected band at ~650bp, though some only at very low concentration. This disparity was mostly attributable to RNA concentration, which was highly variable following extraction. Nevertheless, the majority of cDNAs displayed the desired band and the amplicons were subsequently pooled according to estimated concentration, library prepped and loaded on the Illumina MiSeq instrument.

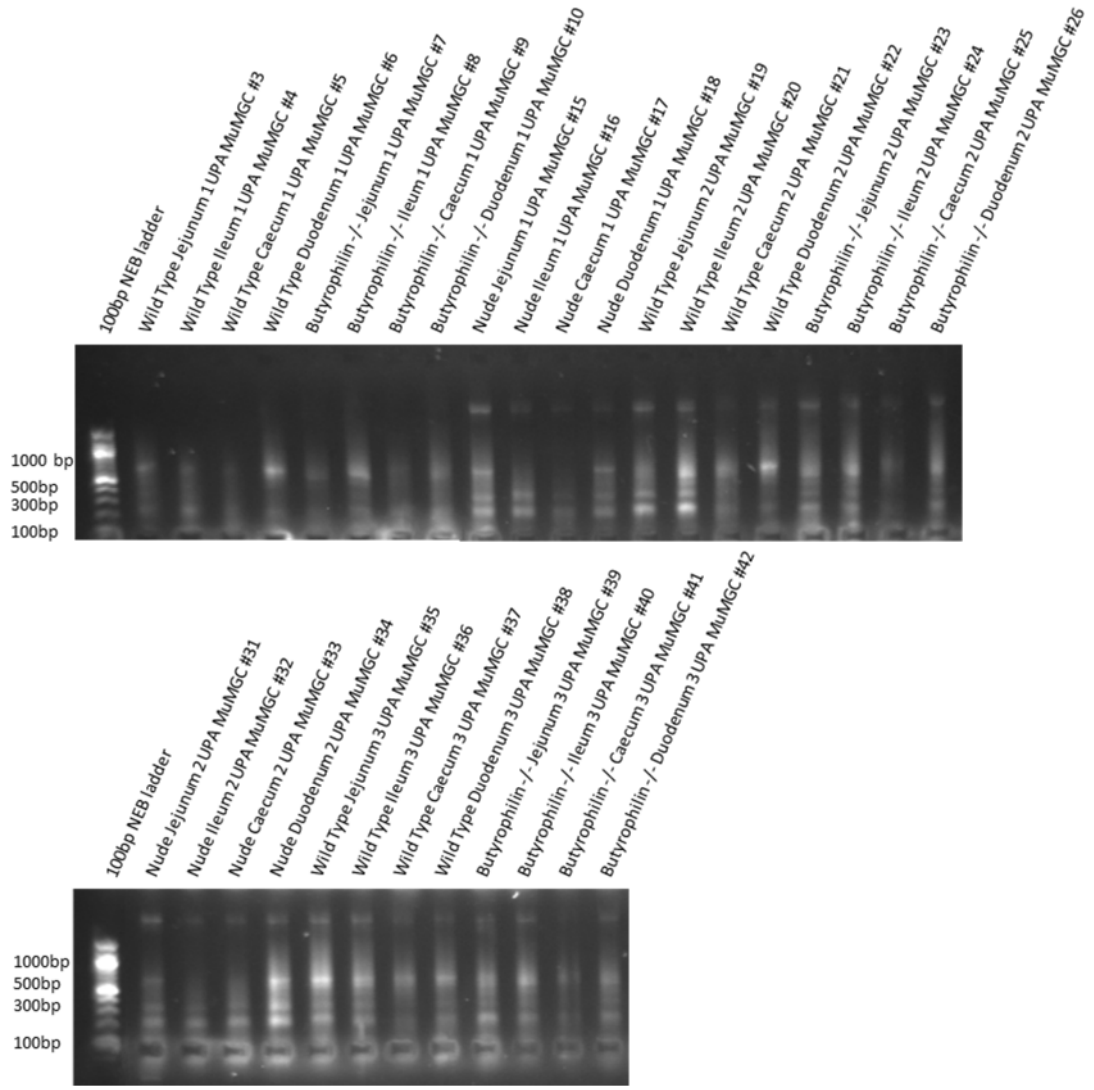


Figure 7.1 – Agarose Gel of TCR Gamma Repertoire PCRs.
Gel image shows amplicons generated using mouse TCR gamma primers with gut tissue cDNA from Wild Type, Butyrophilin -/- and Nude mice as a template. RNA was extracted from RNAlater preserved tissue samples and converted into cDNA using 5' RACE. The TCR gamma region of these cDNAs was then amplified using a universal UPA primer and a barcoded TRGC (MuMGC) primer. PCRs were performed on each sample in triplicate using different barcodes for each of the replicates. Sizes of the 100bp NEB ladder bands are indicated in bp at either side of the gel image and the desired product length is ~650bp. Products were then separated on a 1.5% agarose gel which was run at 120V for 25 minutes.

The MiSeq run produced a total of 56,081,048 reads passing filter (PF) with an average Q30 of 79.78%. Q30 reports the percentage of based with a fastq quality score of greater than 30, or an error probability of 0.001. PF refers to an internal quality control step applied by the MiSeq instrument. This filtering step consists of a calculation based on the ratio of base brightness within

each cluster to determine if it is reliable or not. Those clusters deemed unreliable are discarded from the data, though generally this loss consists of only around 5-10% of the reads. This run produced a total of 3,282,482 PF reads for the UPA PCRs and 6,196,157 for the TRGV7.

The resulting UPA reads were then processed using our in house reptools code and the numbers of reads counted at each stage (Fig 7.2). The PF reads were first searched for the expected primers and barcodes. This is the stage where we lose the most data, typically around two thirds of the reads. This loss is attributable to a range of PCR errors resulting in products which have either only the forward primer, only the reverse primer, an unexpected forward primer or neither of the forward or reverse primers. The reads are then BLAST searched against a database of known TRGV, J and C sequences and the CDR3 region extracted from each read by identification of the conserved residues that flank it. This represents the second substantial read attrition where around 50% of the reads are lost from the previous stage. Though the sources of this loss are not known, they are most likely due to incomplete recombination, leading to the absence of one of the expected gene segments. The desired barcodes stage involves removing any barcodes which are not expected to be present in the data and loss at this stage is negligible. Finally, any non-functional CDR3 sequences which are out of frame or contain stop codons are removed, this stage again involves a loss of around 40% of the remaining data. The sequence data can then be interrogated using a range of *in silico* analyses to examine the TCR gamma repertoire at various different resolutions.

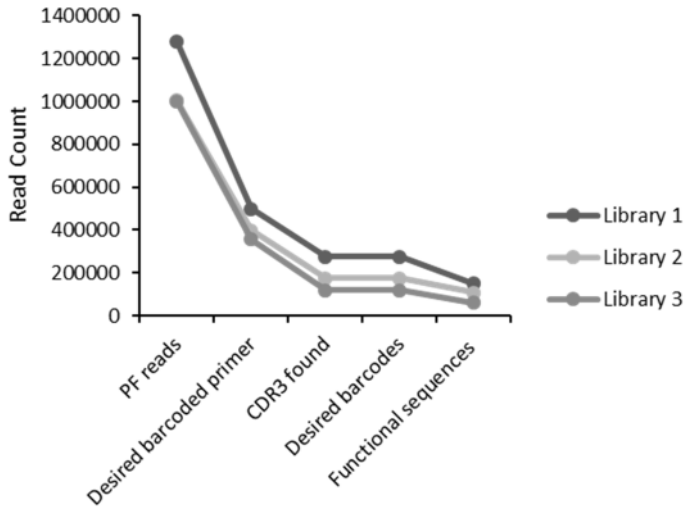


Figure 7.2 – Numbers of Reads at Various Stages of TCR Gamma Sequence Data Processing
Plot shows the numbers of TCR gamma sequences at each stage of processing, following the MiSeq run. After sequencing, the reads must be processed before analysis, using our in house processing pipeline RepertoiR. Passing Filter (PF) reads refers to the numbers of reads which pass the built in MiSeq instrument read quality cut offs. 'Desired barcoded primer' refers to reads which have both the expected primer pair, in this case UPA and MuMGC, and one of the expected barcodes which passes the quality threshold. 'CDR3 found' refers to reads for which the V, J and C regions have been identified and the CDR3 region has been found by identification of the conserved cysteine and phenylalanine residues. The 'Desired barcode' stage simply removes barcodes which were not expected to be present and 'functional' sequences removes any out of frame CDR3s. The Triplicate PCRs were pooled into separate libraries, such that replicate 1 is library 1 and so forth.

7.3.2 The Butyrophilin -/- Mice have a Lower Proportion of γδ T cells Expressing TRGV7 in the Intestinal Tissues.

The removal of Btnl1 has previously been shown to reduce the number of TRGV7+ve γδ IEL T cells in the gut, suggesting it may be mediating their homing, retention or local expansion[7]. In order to replicate this finding and validate the repertoire sequencing strategy, the TRGV gene segments used to generate the TCR gamma chain of each clone were assigned. A linear mixed model (LMM) was calculated for this data using the lme4 package with a logit transformation[33] with mouse number as a random variable and Tissue and TRGV gene as fixed variables. Plots were produced showing the proportion of rearrangements which were attributable to each TRGV gene (Fig 7.3A-D). These plots reveal that though TRGV7 is indeed the dominantly rearranged TRGV segment for small intestinal γδ T cells as shown previously[7] TRGV2 is also prominent. Furthermore, the usage of TRGV7

is lower in the caecum, with TRGV2 becoming the dominant TRGV segment detected in productively rearranged TCR gamma.

As expected, the proportion of TRGV7+ve TCRs declined substantially in the Butyrophilin -/- strain compared to Wild Type in every tissue. This decline was most evident in the Duodenum and Jejunum where it constituted a 5-6 fold loss, compared to the Ileum and Caecum where the loss was ~4 and ~2 fold respectively. These differences were highly significant using p values from the LMM in the Duodenum, Jejunum and Ileum. The decreases in the proportion of TRGV7+ TCRs in the Butyrophilin -/- strain are associated with increases in TRGV2+ TCRs, though these differences are significant in the Duodenum only. Additionally, the Nude mice appear to have a TRGV gene usage profile almost identical to that of Wild Type.

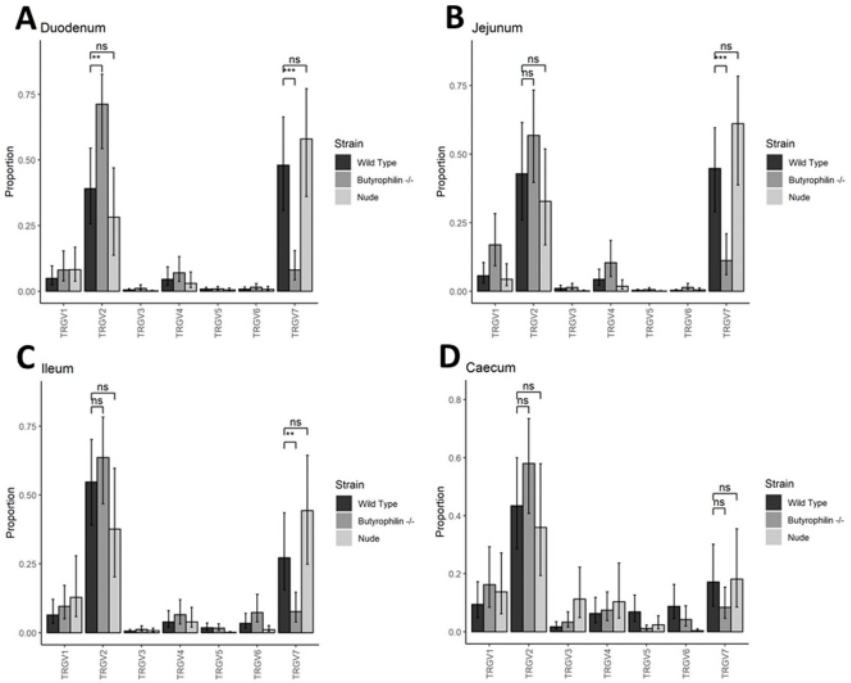


Figure 7.3 – TRGV Gene Usage Profiles for each Strain in Duodenum, Jejunum, Ileum and Caecum
Plots show the proportional TRGV gene usage profiles for the TCR gamma repertoire of **A)** Duodenum, **B)** Jejunum, **C)** Ileum and **D)** Caecum. Each sequence was assigned a TRGV gene by BLAST searching against a database of known sequences. This TRGV gene usage information was then used to generate a random intercept Linear Mixed Model (LMM) using Mouse number as a random variable and Tissue and V gene as fixed variables. Data was logit transformed before model calculation to ensure the model assumptions were met. P values were then calculated from the LMM, using the Satterthwaite approximation of degrees of freedom, and are displayed above the bars. Significance thresholds were as follows *** p value < 0.001, ** p

value < 0.01, * p value < 0.05. Error bars are 95% confidence intervals calculated from 500 bootstrap replicates generated from the LMM.

7.3.3 Beta-Diversity Plots reveal Differences in Overall TCR Gamma Repertoire between Mouse Strains

In order to explore the overall impact each mutation had on the TCR gamma repertoire, β-diversity plots were generated by treating each clone, as defined by a unique combination of Complementarity Determining Region 3 Nucleotide (CDR3NT), TCRV and TCRJ, as a separate species (Fig 7.4). Different diversity metrics were used to generate each of the plots as each comes with its own pros, cons and assumptions (See Chapter 3). These plots were generated first using Bray-Curtis distance, a common comparison metric for repertoire data (Fig 7.4A). Ordination based on Non-Metric Multidimensional Scaling (NMDS) with 20 random starts was used to visualize these plots. All of the gut tissues were plotted together for ease of interpretation since mouse strain appeared to have a far greater effect.

Unsurprisingly, the Bray-Curtis plots of total TCR gamma repertoire showed large differences between the mouse strains. The Wild Type and Butyrophilin -/- strains show complete separation with almost no overlap, with the single exception of the Caecum of Wild Type mouse 3. These plots also show separation of the Nude mice, which again cluster away from the Wild Types. Additionally, the Nude mice appear far more variable than either Wild Type or Butyrophilin -/-. As expected from visual inspection, when tested with PERMANOVA, overall P values for each tissue were significant with relatively high R^2 values of 0.33-0.41. This suggests that mouse strain makes a significant difference to the overall gut TCR gamma repertoire and explains a large portion of the variance in the data. However, when attempting to use PERMANOVA to test for differences between isolated pairs of strains, none of the P values reached a significance threshold of 0.05. This was most likely due to the low numbers of individuals and resulting lack of power, rather than the absence of any differences between the paired strains. Jensen-Shannon Divergence (JSD) plots were also generated as Bray-Curtis often has issues comparing sparse data with a high number of zeros, such as

repertoire sequencing data (Fig 7.4B). Reassuringly, the JSD plots are virtually identical and show extremely similar patterns of significance and R^2 values.

Aitchison's distance, defined as Euclidean distance calculated from centre log ratio (CLR) transformed data, was then calculated and ordination plots generated using Redundancy Analysis (RDA) (Fig 7.4C). These plots are distinct from the Bray-Curtis and JSD plots in terms of appearance, however the patterns still show separation between the 3 strains. The main difference between the Aitchison's distance plots and the Bray-Curtis and JSD plots, is the variation, with the Nude mice appearing the most tightly clustered by far, and Wild Type and Butyrophilin -/- being more spread out. Unsurprisingly, this uneven variance gives significant Betadisper tests for every tissue and therefore PERMANOVA testing is invalid. This difference in clustering is possibly due to the low percentages of variation explained by the RDA axes in this plot (See Chapter 3), which may therefore be missing the data that makes the Nude strain so much more widely distributed by Bray-Curtis and JSD.

To further explore the differences to the repertoire in the absence of Btnl1, we also generated β-diversity plots with TRGV7+ve clones removed (Fig 7.4D). Comparing the TRGV7 negative clones first by Bray-Curtis distance, with NMDS plots generated as above, shows that the Wild Type and Butyrophilin -/- strains are intermixed and indistinguishable from one another. The Nude mice were more widely dispersed across the plot, surrounding the Wild Type, Butyrophilin -/- cluster. However, this plot still gives significant PERMANOVA P values in the Duodenum, Jejunum and Caecum with R^2 values ranging from 0.33-0.34, most likely solely due to the Nude strain. Again, the JSD plots appear exactly the same as Bray-Curtis, however by JSD, the overall differences are significant for the Caecum only, with an R^2 value of 0.36 (Fig 7.4E).

On the other hand, the Aitchison's distance plots do maintain some degree of separation, with the tissues from Butyrophilin -/- Mouse 2 in particular clustering apart from the rest of the samples (Fig 7.4F). However, again this plot is not eligible for PERMANOVA testing due to a

significant Betadisper test and therefore it is unclear if these differences would be significant or not. Additionally, the variance explained by the axes is again low and all of the strains cluster more closely together than in the Aitchison's distance plot which includes the TRGV7+ve clones (Fig 7.4C).

Therefore, the reduction in TRGV7+ve clones in the Butyrophilin -/- strain clearly has a large impact on the overall proportion of TRGV7 and its influence on the global TCR gamma repertoire. Interestingly, the TRGV7-ve clones appear totally unaffected in this knockout strain and are indistinguishable from Wild Type when compared using Bray-Curtis and JSD. Furthermore, contrary to what was observed in the TRGV usage plots, the Nude strain also showed large repertoire differences compared to both Wild Type and Butyrophilin -/- when examining clonal composition.

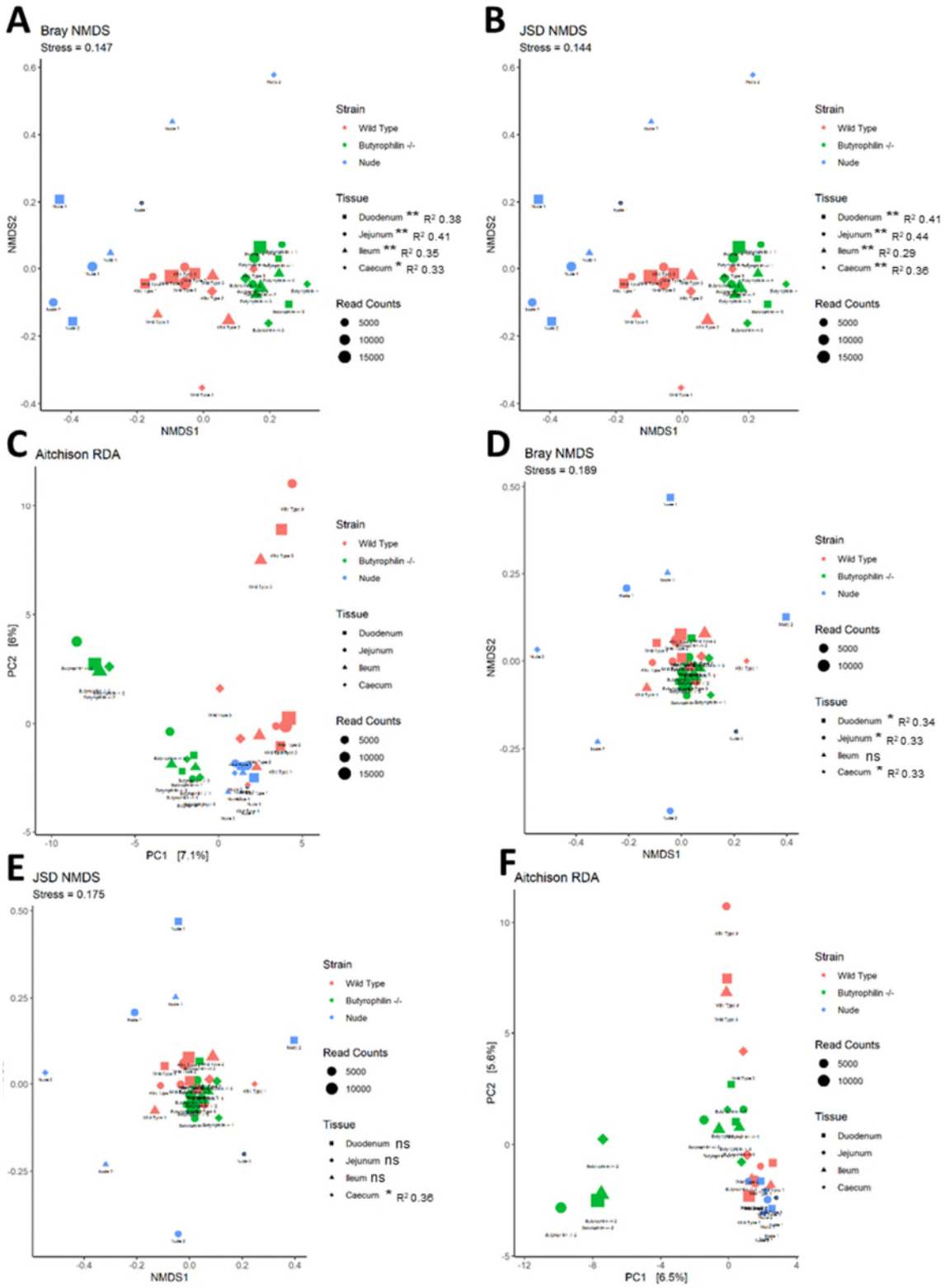

A
Bray NMDS
Stress = 0.147
NMDS1
NMDS2
Strain
Wild Type
Butyrophilin -/-
Nude
Tissue
Duodenum ** R² 0.38
Jejunum ** R² 0.41
Ileum ** R² 0.35
Caecum * R² 0.33
Read Counts
5000
10000
15000
B
JSD NMDS
Stress = 0.144
NMDS1
NMDS2
Strain
Wild Type
Butyrophilin -/-
Nude
Tissue
Duodenum ** R² 0.41
Jejunum ** R² 0.44
Ileum ** R² 0.29
Caecum ** R² 0.36
Read Counts
5000
10000
15000
C
Aitchison RDA
PC1 [7.1%]
PC2 [6%]
Strain
Wild Type
Butyrophilin -/-
Nude
Tissue
Duodenum
Jejunum
Ileum
Caecum
Read Counts
5000
10000
15000
D
Bray NMDS
Stress = 0.189
NMDS1
NMDS2
Strain
Wild Type
Butyrophilin -/-
Nude
Read Counts
5000
10000
Tissue
Duodenum * R² 0.34
Jejunum * R² 0.33
Ileum ns
Caecum * R² 0.33
E
JSD NMDS
Stress = 0.175
NMDS1
Strain
Wild Type
Butyrophilin -/-
Nude
Read Counts
5000
10000
Tissue
Duodenum ns
Jejunum ns
Ileum ns
Caecum * R² 0.36
F
Aitchison RDA
PC1 [6.5%]
PC2 [5.6%]
Strain
Wild Type
Butyrophilin -/-
Nude
Read Counts
5000
10000
Tissue
Duodenum
Jejunum
Ileum
Caecum

Figure 7.4 – β Diversity Plots of TCR Gamma Repertoire Reveal Differences between Mouse Strains
Plots were generated from the entire TCR gamma repertoire by ordination of the β-diversity indices **A)** Bray-Curtis, **B)** JSD and **C)** Aitchison's distance for each strain and tissue. Additional plots were generated from the TRGV7-ve repertoire only, using **D)** Bray-Curtis, **E)** JSD and **F)** Aitchison's distance. β-diversity indices were calculated from the proportional TCR gamma repertoire data using each unique clone, as determined by combination of TRGV, TRGJ and CDR3NT. The tissues were plotted together due to the fact that they were relatively similar within each strain and that there were very few points on each plot, so each is easily distinguishable. NMDS plots were generated from Bray-Curtis and JSD data using 20 random starts and the resulting stress values are displayed beneath the titles. Aitchison's distance plots were generated using RDA and the percentage of total variation explained by each axis is displayed in brackets. Where a negative betadisper test was obtained (in all of the Bray-Curtis and JSD plots, and none of the Aitchison's distance plots) PERMANOVA testing was conducted on each tissue in isolation using 999 permutations. PERMANOVA False Discovery Rate (FDR) corrected P and R2 values are displayed next to the tissue they correspond to in the plot key. The significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

7.3.4 The TCR Gamma Repertoire is Highly Public in all Mouse Strains

Public TCRs are those shared between multiple individuals, whereas private TCRs are those limited to a single individual. Considering the vast potential for different TCR sequences driven by recombination, it is surprising that public sequences are observed so frequently. However, as previously mentioned, this potential variability is never actually reached due to a number of factors. Particular subsets of γδ T cells in particular are known to be subject to limitations in terms of rearrangement[34]. Though the term public can have several meanings depending on interpretation, here a public sequence is defined as one found in any tissue at any proportion in a specified percentage of individuals.

Analysing the proportion of public clones in each TRGV gene segment reveals some interesting patterns (Fig 7.5). In the Wild Type mice, the TRGV5 and 6 are the most public and are made up of ~75 and ~95% public clones respectively. Conversely, TRGV3 is by far the least public and only ~10% of the clones expressing this TRGV gene are 100% public. Interestingly, the TRGV7 was also highly public, at ~65%, as well as being the dominant TRGV gene. Looking to the Butyrophilin -/- mice, the percentage publicity of each TRGV was similar to that of Wild Type, with the exception of TRGV5 and TRGV7, which both show a slight decrease in publicity. Unfortunately, as with the paired mouse strain comparisons in the β-diversity plots, there are too few individuals to get a significant P value with a Wilcoxon Signed Rank test, even if the differences were very large. These plots show

that in the absence of Btnl1, the public clones appear almost totally unaffected as a proportion of the TRGV gene they express, even those within TRGV7.

Comparing Wild Type to Nude, TRGV3 and 6 both display a lower proportion of publicity in Nude mice, with this decrease being more substantial in the TRGV3. Interestingly, TRGV5 is also increased to 97% publicity in this strain. Though interpretation is less convincing on account of the differences in sample number, this suggests some the TRGV5 public rearrangements may have an extrathymic origin.

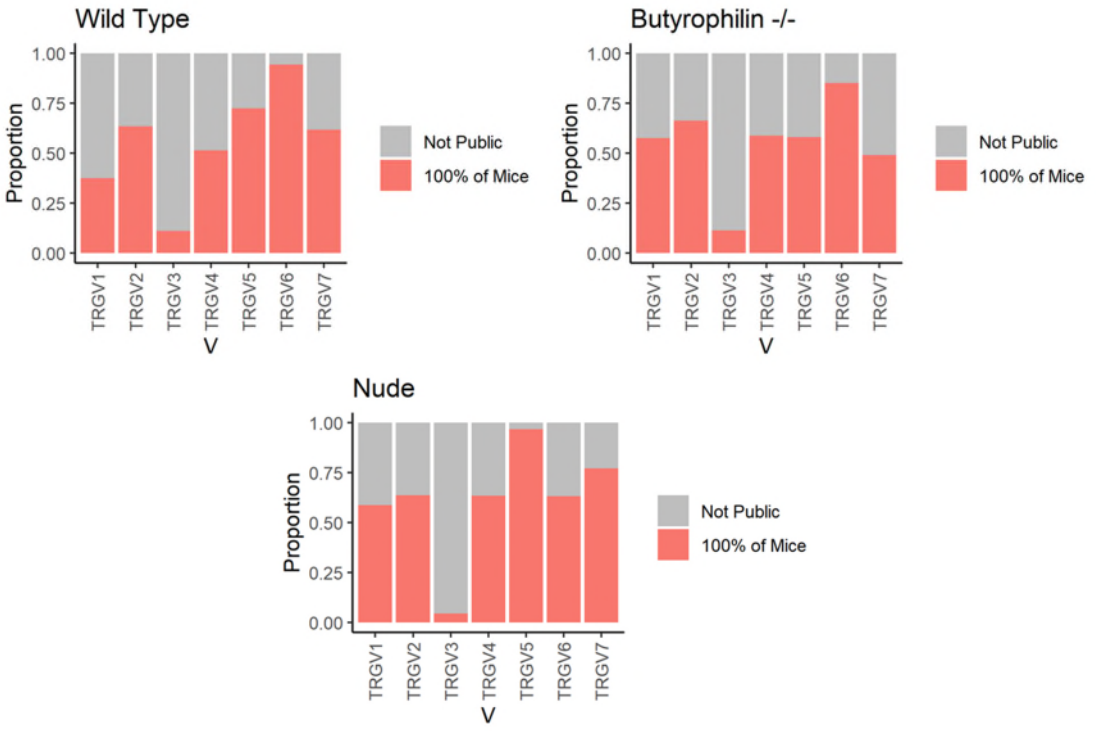


Figure 7.5 – Publicity Plots by TRGV Gene Segment
Plots show the proportion of the TRGV repertoire which is made up by public clones. In this instance, publicity is defined by presence in every mouse within a strain, in any tissue and at any proportion. The proportion of each TRGV gene has been adjusted such that each adds up to 1 for ease of interpretation, however overall proportions of each V gene are as in Fig 7.3. Wilcoxon Signed Rank testing was performed but P values are not shown as power was insufficient to obtain a significant result.

7.3.5 TCR Gamma Repertoire Diversity is Greatly Reduced in the Nude Mice

In order to further examine the clonal structure of each mouse strain, diversity indices were calculated at a common sample coverage using data rarefied with iNEXT[35] (Fig 7.6). A linear model

(LM) was then calculated for these diversity indices using a cube root transform, which was found to generate data that best approximates normality. The diversity indices show that the TCR gamma repertoire diversity decreased in every strain as the Hill number increased from Species Richness (H=0), through to Simpsons Diversity (H=2) (Fig 7.6A). As Hill number increases, the diversity score is less affected by rare species and increasingly weights the calculation towards species evenness. Therefore, these plots suggest that the mouse gamma repertoire has low evenness in every strain and every tissue, and that there is a large number of rare clones. By every diversity index, the Jejunum is the most diverse tissue in the Wild Type mice. In order of decreasing diversity, Jejunum is followed by Duodenum, Ileum and Caecum, though the differences in diversity between all of these tissues are relatively small.

Comparing the tissues by strain, the Butyrophilin -/- mice are slightly more diverse in every tissue by every metric, though none of these differences reach a significance threshold of 0.05 when compared using P values from the LM. The Nude strain on the other hand is strikingly less diverse than the Wild Type in all of the tissues. This difference leads to significant P values in Duodenum and Jejunum by Species Richness, Duodenum, Jejunum and Caecum by Shannon Diversity and Duodenum only by Simpsons Diversity. This suggests that the presence of the thymus is necessary for the gut TCR gamma repertoire to reach its full diversity.

Diversity indices were then calculated for each TRGV gene individually to gain a more detailed picture of the impact of a lack of Btnl1 and athymic development on the TCR gamma repertoire population structure (Fig 7.6B-E). Interestingly, TRGV2 was the most diverse V gene in every tissue and TRGV7 was the second most diverse. The Butyrophilin -/- strain tended to have a slight increase in TRGV2 diversity and a slight decrease in TRGV7 diversity relative to the Wild Type mice. Similar to the whole tissue diversity plots though, the differences between these two strains were mostly small and only reached significance in Duodenum TRGV2 by Shannon and Simpson's diversity and Jejunum by Simpson's diversity. Comparing the Nude mice with Wild Type revealed

significant decreases in diversity in TRGV2 and TRGV7 in the Nude Duodenum and Jejunum. Surprisingly however, the TRGV gene diversity plots showed an increase in diversity in Nude Ileum TRGV2 and Caecum TRGV7 relative to Wild Type by Species Richness, though neither of these differences reach significance and in both cases actually show a decrease in diversity by Shannon's and Simpson's diversity indices. The Nude strain also displayed a decrease in diversity in TRGV1 in all of the tissues.

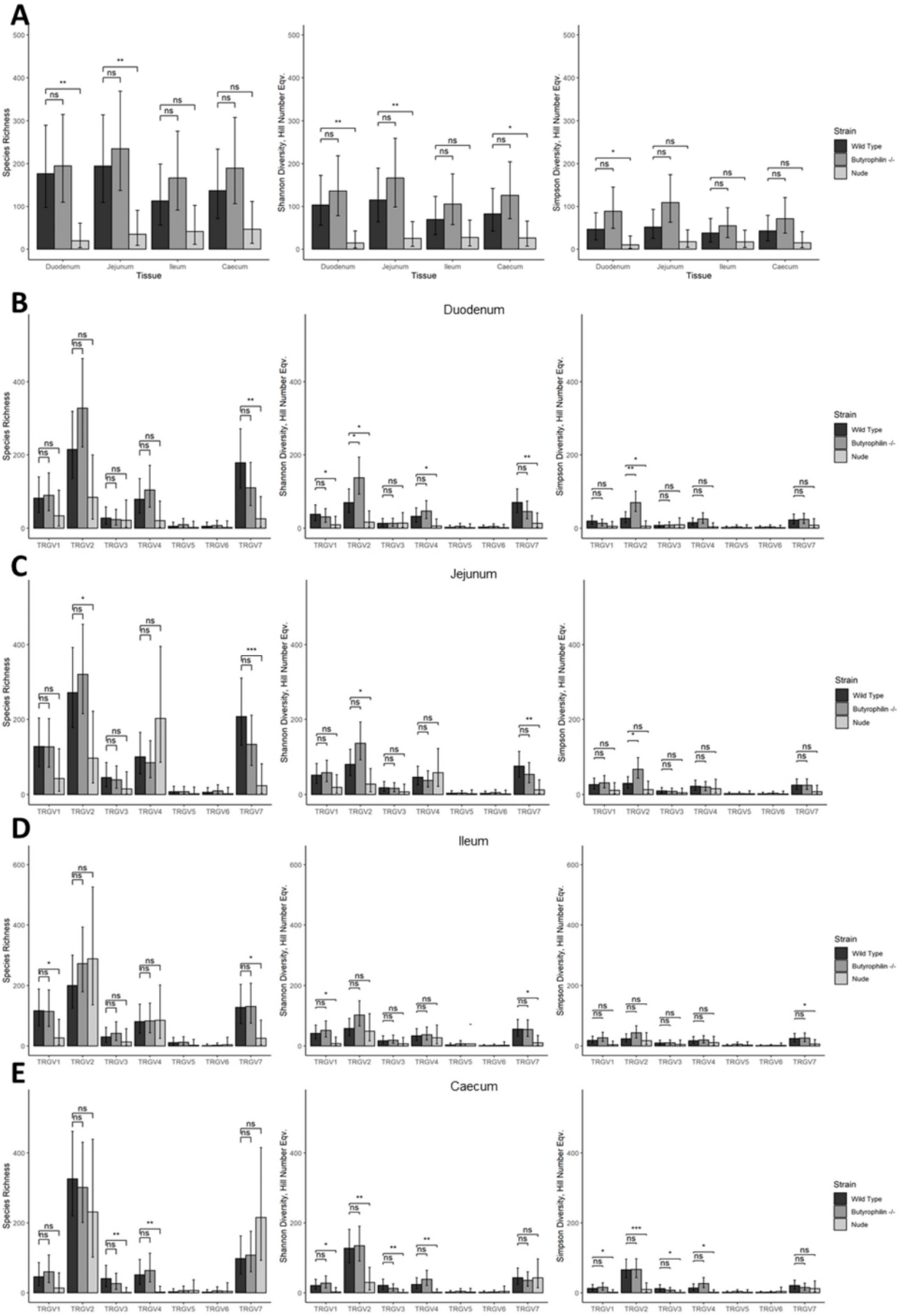

A
Species Richness
Shannon Diversity, Hill Number Eqv.
Simpson Diversity, Hill Number Eqv.
Duodenum
Jejunum
Ileum
Caecum
Tissue
B
Duodenum
C
Jejunum
D
Ileum
E
Caecum
TRGV1
TRGV2
TRGV3
TRGV4
TRGV5
TRGV6
TRGV7
Strain
Wild Type
Butyrophilin -/-
Nude

Figure 7.6 – TCR Gamma Hill Number Diversity Plots Reveal Differences in the Mouse Strains
Plots show iNEXT rarefied diversity indices for **A)** each tissue compared by strain as well as the diversity of the individual TRGV genes within **B)** Duodenum, **C)** Jejunum, **D)** Ileum and **E)** Caecum. All of the samples were first rarefied down to the lowest sample coverage using iNEXT rarefaction to allow calculation of diversity indices from comparable samples. The Hill number variants of each diversity index were then calculated and used to generate a Linear Model (LM) from cube root transformed data. Error bars are 95% confidence intervals and P values were extracted from the model. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05

7.3.6 PCR Amplification of the TRGV7 Repertoire using Specific Primers

A reduction in the proportion of TRGV7 rearrangements in the Butyrophilin -/- strain was an expected result that had been noted previously in the literature[7]. However, NGS repertoire data allows a more in depth investigation of the nature of the changes to TRGV7+ repertoire, and allows us to gain a more detailed picture of the impact Btnl1 has on γδ T cells. In order to focus more specifically on TRGV7 in isolation, repertoire PCRs were performed using a TRGV7 specific primer instead of the universal UPA primer used previously (Fig 7.7). These PCRs were then amplicon sequenced at greater depth and processed as previous.

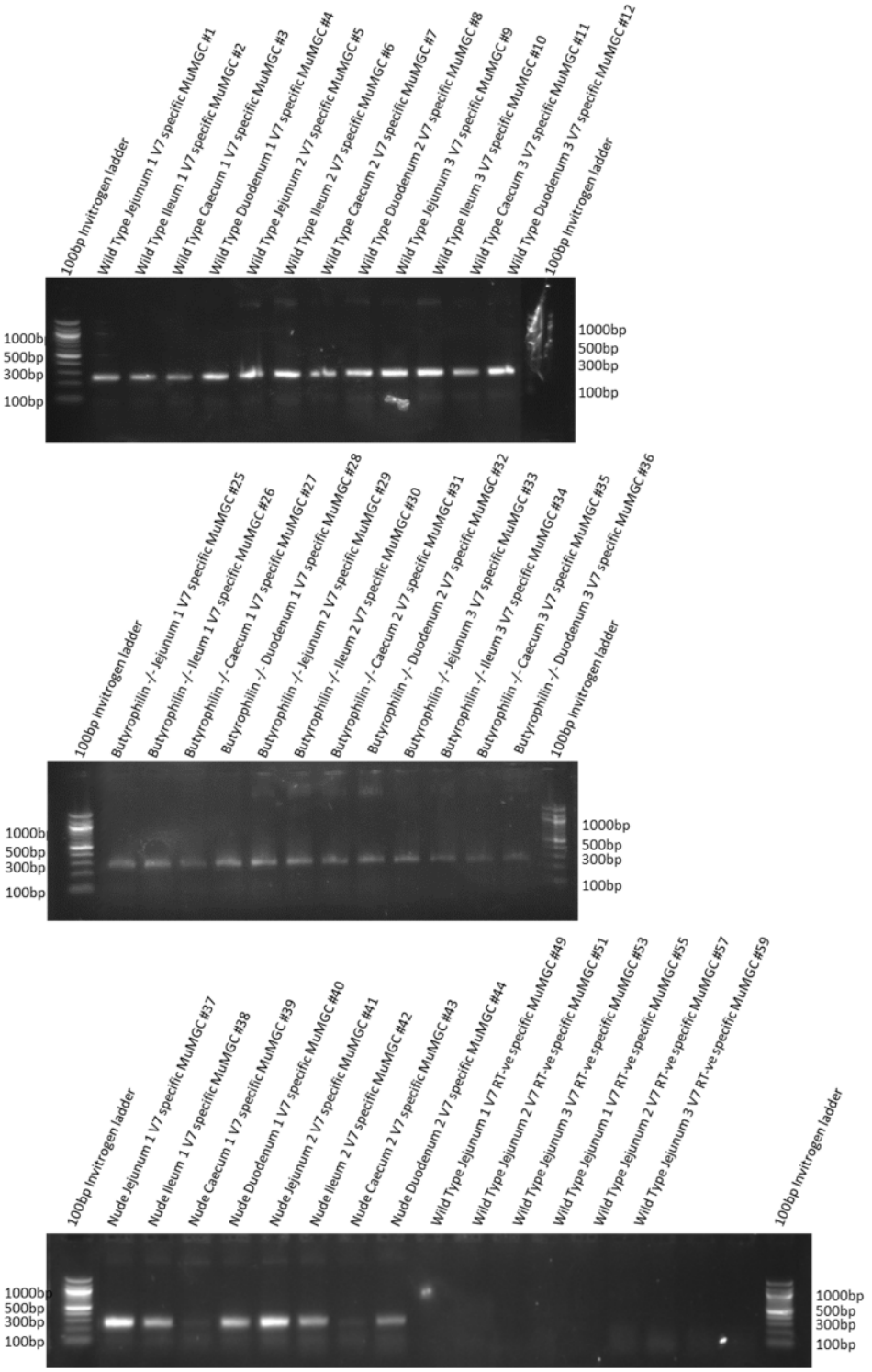


Figure 7.7 – Agarose Gel of TRGV7 Specific PCRs

Image shows an agarose gel of PCR amplicons from the TRGV7 gene segment. cDNAs were generated using 5' RACE on RNA extracted from mouse tissues using the AllPrep RNA/DNA micro kit (Qiagen). These cDNAs were used as a template for PCRs targeted at the TRGV7 gene segment using a specific forward primer and the barcoded TRGC (MuMGC) reverse primer used in the UPA PCRs (Fig 7.1). Reverse transcriptase negative (RT-ve) templates were also run alongside the cDNAs, to confirm no amplification was coming from either the RNA itself or any gDNA present in the extractions. PCRs were performed in triplicate for each sample and a different barcoded MuMGC primer was used for each of the replicates. Products were then separated on a 1.5%

agarose gel which was run at 120V for 25 minutes. A 100bp NEB ladder was run alongside these amplicons to determine their sizes and for rough estimates of concentration.

The outputs of the MiSeq run were as detailed in the 'Mouse Tissue extraction and PCR amplification of TCR gamma repertoire using universal primers' section, earlier in this chapter. A plot was generated to track the numbers of reads at each stage of the processing (Fig 7.8). As with the UPA repertoire data, the greatest loss by far occurs at the 'Desired barcoded primer' stage, which involves searching the data for the primers and barcodes used in the PCRs. With the TRGV7 specific data, the loss of reads at this stage amounts to 75%. However, following this considerable loss, very few reads are lost during any of the subsequent processing stages. The large read attrition that occurs in the UPA data at both the 'CDR3 found' and 'Functional Sequences' stages is not present in this data, presumably due to the use of a specific forward primer and the lack of spurious amplification as a result.

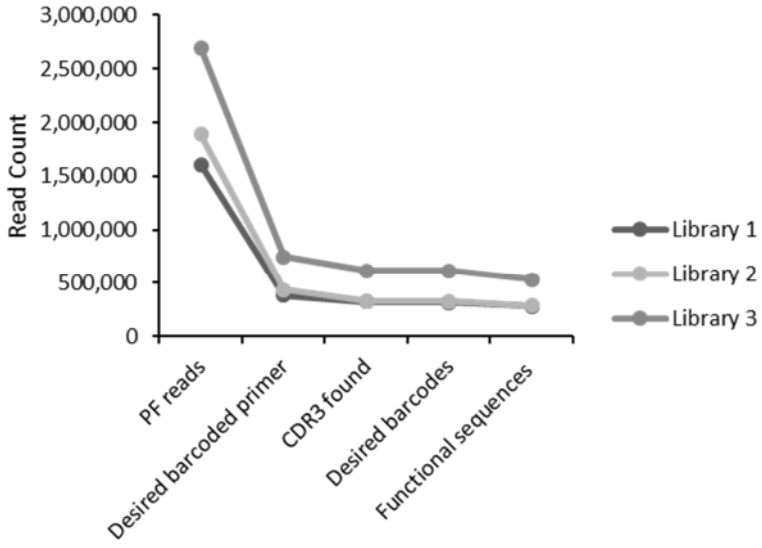


Figure 7.8 – Numbers of Reads after each Stage of TRGV7 Specific Repertoire Processing
Plot shows the number of reads at each stage as the sequence data from the TRGV7 specific PCRs was processed. Following the MiSeq run, numbers of reads at each stage in the processing were calculated and plotted. 'PF reads' refers to those that have passed the MiSeq built in quality filter. 'Desired Barcode Primer' refers to reads for which the expected primer and one of the barcodes can be found. 'CDR3 found' is the stage at which each read is BLAST searched against a database of TRGV, J and C sequences and the CDR3 region picked out by the presence of conserved cysteine and phenylalanine residues. The 'Desired barcodes' stage removes any barcode not expected to be present in the library. Finally, 'Functional Sequences' refers to the removal of any read which is not in frame.

7.3.7 Beta-Diversity Plots show Differences in the Overall TRGV7 Repertoire of both Butyrophilin -/- and Nude Mice

As with the UPA data, each of the strains was first compared using a range of β-diversity indices in order to gain an insight into potential community level differences in each of their TRGV7 repertoires. Generating β-diversity plots for the TRGV7+ repertoire in each strain first using Bray-Curtis distance, reveals large differences between them (Fig 7.9A). All of the Wild Type, Butyrophilin -/- and Nude mice cluster totally separately with virtually no overlap between them. Additionally, the TRGV7+ repertoires of all of the Wild Type mice were remarkably closely clustered and therefore very similar to one another. On the other hand, both the Butyrophilin -/- and Nude mice show far more variance in their repertoires and appear much more spread out in comparison. Unfortunately, this uneven spread does mean that in the case of the Duodenum and Jejunum, betadisper testing between the 3 strains gives a significant result and therefore violates the PERMANOVA assumption of homogeneity of variance. As a result, statistical testing is valid only in the Ileum and Caecum, both of which prove significant with R^2 values of 0.39 and 0.35 respectively, showing that strain accounts for around a third of the variation present in these plots. As with the UPA β-diversity plots, the low numbers of individuals present in this experiment results in PERMANOVA lacking the power to give significant P values for individual paired comparisons between each of the strains. Despite this, the degree of separation in each of the strain clusters suggests that differences for all of the pairs would be significant were more individuals present.

Importantly, comparing the strains using JSD reveals an identical pattern to that found in the Bray-Curtis plots, despite the differences in rotation (Fig 7.9B). In the case of the TRGV7+ repertoire specifically this is an important result. As it has been shown that the Butyrophilin -/- strain has reduced levels of TRGV7+ γδ T cells in the gut, this may lead to increased sparsity in the data and could therefore give spurious results from Bray-Curtis distance. However, the JSD plots show that this is not the case and the differences shown in the Bray-Curtis plots are likely due to real

differences in the TRGV7+ve clones. By JSD, the overall differences between the strains again prove significant in the Ileum with an R^2 value of 0.41, however PERMANOVA gives a P value just below the significance threshold in the Caecum.

The Aitchison's distance plot again shows mostly distinct clustering between the Wild Type mice and the Butyrophilin -/- and Nude mice, which oddly appear to cluster together by this metric (Fig 7.9C). Additionally, the Wild Type mice appear far more spread out in comparison to both Butyrophilin -/- and Nude which are both tightly clustered. Again, this difference in ordination compared to Bray-Curtis and JSD is most likely due to the low variation represented by the Aitchison's distance axes, just 6.3 and 6% respectively. This may be a general weakness of Aitchison's distance for comparing repertoire data and should be considered when interpreting these plots. As a result of this difference in variance, the Aitchison's distance betadisper tests are significant for all tissues and therefore PERMANOVA testing cannot be performed on these plots.

Overall, these β-Diversity plots show large differences between all of the strains, suggesting that the TRGV7+ specific repertoires are affected in both the Butyrophilin -/- and Nude mice. This is of particular relevance for the Butyrophilin -/- mice as it suggests that not only do they have reduced TRGV7+ γδ T cells in the gut, but also that the repertoire of these lymphocytes itself is altered.

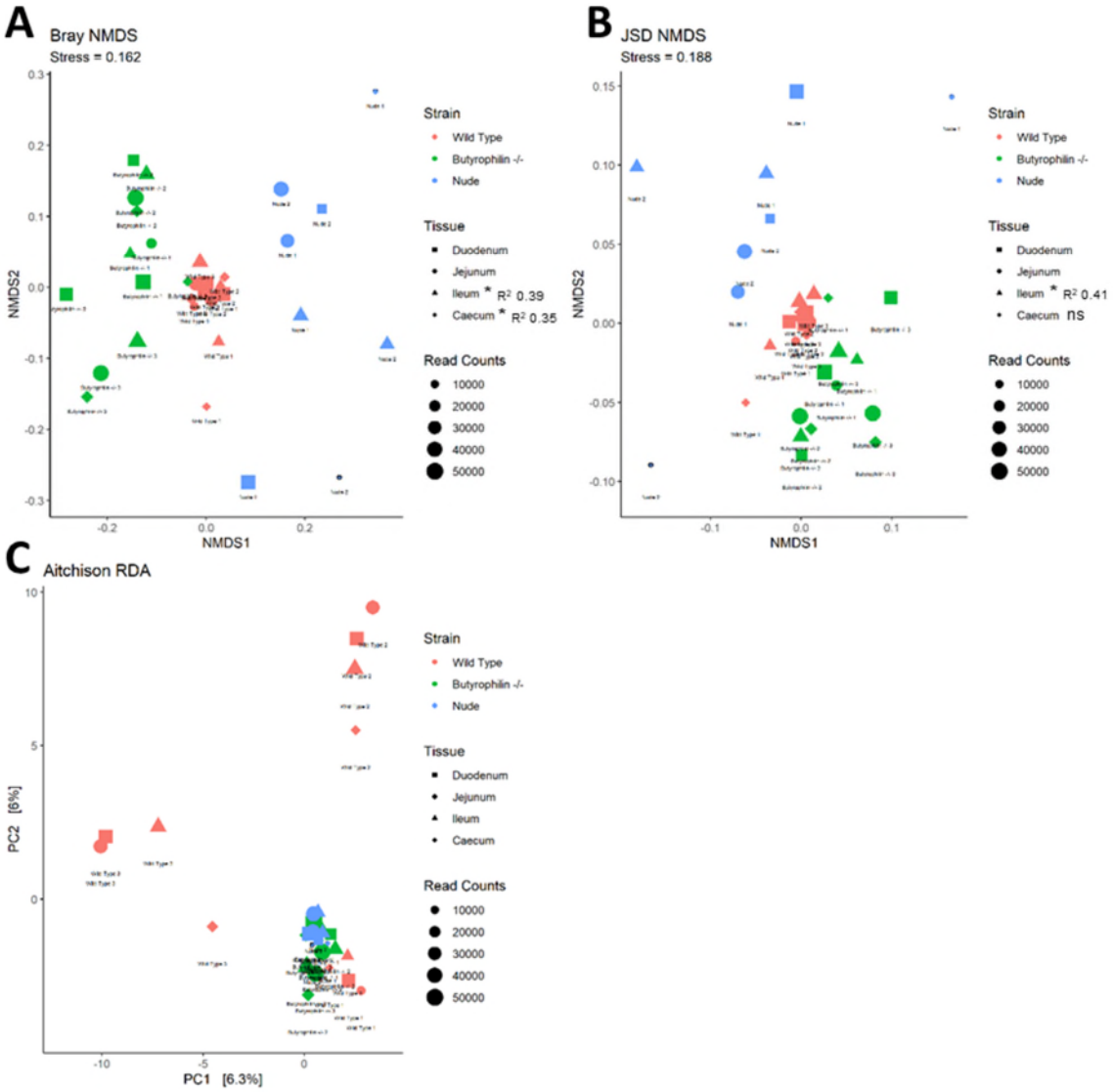


Figure 7.9 - β-Diversity Plots of TRGV7 Specific Repertoire Data shows that the TRGV7+ Clonal Populations are Altered in Butyrophilin -/- and Nude Mice

Ordination plots were generated from the β-Diversity metrics **A)** Bray-Curtis, **B)** JSD and **C)** Aitchison's distance calculated from TRGV7 specific repertoire data. Distance metrics were calculated by treating each unique clone, as defined by a unique combination of TRGV, TRGJ and CDR3NT sequence, as a species. Bray-Curtis and JSD ordination plots were generated using NMDS with 20 random starts and the resulting stress values are displayed beneath the plot titles. Aitchison's distance was calculated from centre log ratio (clr) transformed data and was displayed using RDA with the percentage of variation explained by each axis shown in brackets. Strain is indicated by point colour and Tissue by shape. All tissues were plotted together as they were relatively similar and the TRGV7 repertoires mainly differed based on strain. PERMANOVA testing was conducted for each tissue overall using 999 permutations where a non-significant betadisper test was first obtained. The lack of a P or R2 value indicates that the betadisper test was significant and therefore the assumptions of PERMANOVA were not met. P values were FDR corrected and displayed alongside the R2 values next to the tissue they are associated with in the plot keys. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

7.3.8 The TRGV7 Repertoire Diversity is Decreased in the Nude Mice

Plotting TRGV7 specific diversity allows us to gain more of an insight into the nature of the changes hinted at by the universal α-diversity plots (Fig 7.6). As in these pan TCR gamma repertoire plots, there is a slight decrease in TRGV7+ve diversity in the Btnl1 -/- Duodenum and Jejunum, and a slight increase in the Ileum and Caecum when compared to Wild Type (Fig 7.10). However, in these TRGV7 specific plots, the differences between Wild Type and Butyrophilin -/- diversity in the Duodenum do reach significance when compared by Species Richness. The diversity of the Nude mice is significantly decreased in every tissue by every metric with the two exceptions of Species Richness in the Caecum and Simpson's diversity in the Duodenum. These plots reveal that the Nude mice have a large reduction in TRGV7+ diversity and that the diversity of the TRGV7+ Butyrophilin -/- repertoire is also reduced, though to a lesser extent.

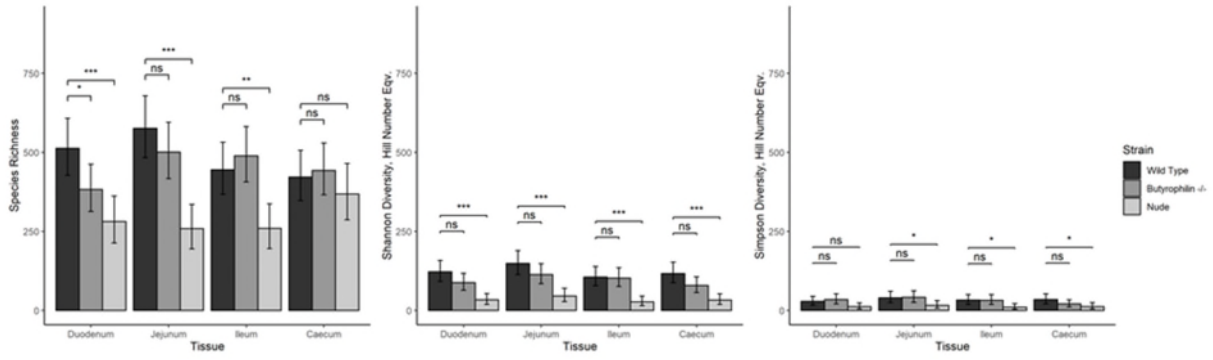


Figure 7.10 – Hill Number Diversity Plots for TRGV7 Specific Repertoire
Plots show diversity indices calculated from iNEXT rarefied TRGV7 specific repertoire data. Before calculation of diversity, to ensure that the calculated indices were not influenced by read depth, iNEXT was used to rarefy the data down to the level of the lowest coverage sample. Following this, Species Richness and Hill number variants of Shannon's and Simpson's diversity indices were calculated for each tissue and strain. An LM was then calculated based on cube root transformed diversity index data and used to generate P values and error bars, which represent 95% confidence intervals. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05

7.3.9 The Most Common, Canonical Clones Appear to be Most Affected in the Butyrophilin -/- Mice

Considering differences at the level of individual clones gives the highest resolution comparisons of the TCR gamma repertoire. Examining the TRGV7 repertoire in this way allows us to examine differences in the Butyrophilin -/- and Nude strains compared with Wild Type in extremely fine detail. Similarity percentages (SIMPER) is a method which enables assessment of the relative contribution of each clone to differences between strains assessed using a Bray-Curtis distance score. Using SIMPER, the largest contributors to dissimilarity between mouse strains in the β-diversity graphics were identified and examined on individual clone plots. The TRGV7 repertoire consists of several thousand unique CDR3s and therefore differences between strains will occur by chance, particularly when considering proportional data with low sample sizes. Therefore, though SIMPER outputs the contributions of every clone, only those with high proportions were highlighted. Unfortunately, again owing to the small sample sizes in this experiment there was insufficient power to produce a significant result when using an unpaired Wilcoxon signed rank test in any of the clone plots produced by this analysis.

Clone 1 was the largest single clone in every mouse sampled and the most common rearrangement in most individuals, as well as being entirely public and expressing a fully germline CDR3NT made up of TRGV7 and TRGJ1 (Fig 7.11A). Despite remaining the most common clone, the proportion of this clone was substantially reduced in the Bntl1 -/- strain compared to Wild Type by around 50% in every tissue except the Jejunum. Additionally, this clone is present in the Nude strain at levels mostly similar to those of the Wild Type mice, except the jejunum where it is reduced, suggesting it has an extrathymic origin.

Clone 2 is another fully public clone, it is present in every mouse and is another of the most common rearrangements in almost all of the individuals sampled (Fig 7.11B). This clone also expresses TRGV7 and TRGJ1 but differs from clone 1 and the TRGV germline sequence by 5

nucleotides, leading to an amino acid change in the CDR3. Similarly to clone 1, this clone is again reduced by around 50% in the Bntl1 -/- mice in every tissue sampled, again except in the Jejunum, where the reduction is ~25%. Additionally, clone 2 is also somewhat decreased in the Nude strain in every tissue but the Caecum. The reduction of these extremely common, public, canonical clones provides some insight into the impact of Btnl1 and suggests that it may be biased in its activity towards certain canonical γδ TCRs.

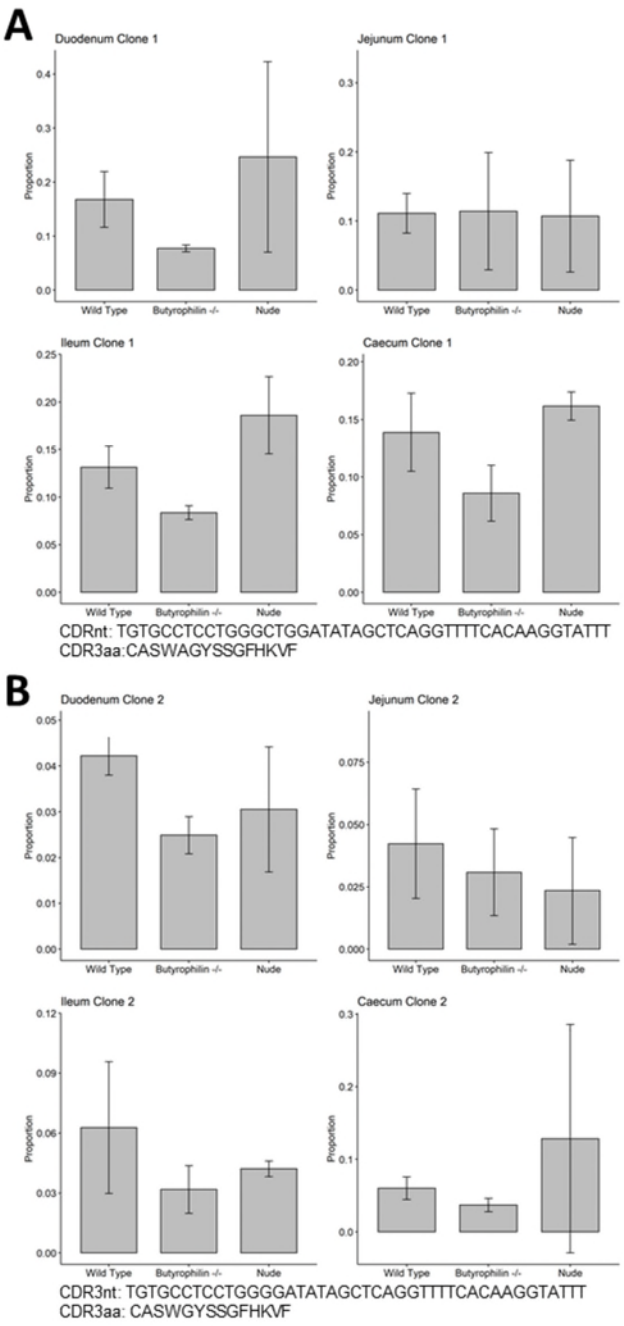


Figure 7.11 – Individual TRGV7 Clone Plots reveal Clones Reduced in Butyrophilin -/- Mice
Plots show TRGV7 clones selected by SIMPER analysis of Bray-Curtis distances between mouse strains. The mean contribution to the TRGV7 repertoire made by clones **A)** 1 and **B)** 2 was calculated for each strain in each tissue. Clones are named according to the order in which they are encountered during clone assignment, so while the numbers are arbitrary, those with lower numbers tend to be the most common clones as they are encountered earlier. CDR3NT and Complementarity Determining Region 3 Amino Acid (CDR3AA) sequences are shown below each plot. Error bars are standard deviation. Wilcoxon signed rank testing was performed but P values are not shown as power was insufficient.

7.3.10 16S Sequencing from Caecal DNA

In mice, γδ T cells are highly enriched at epithelial sites and so are in close and prolonged contact with the microbiota. To examine whether γδ T cells may play a role in influencing the composition of the microbiota in the mouse gut, 16S sequencing was used to examine the gut flora of Wild Type, TCRδ-/- mice which lack all γδ T cells, Bntl1-/- which are depleted in TRGV7+ γδ T cells and athymic Nude mice. The 16S portion of the ribosomal RNA (rRNA) was amplified using PCR (Fig 7.12). Only the PCRs from caecal tissues showed a band at the expected size and therefore the other tissues did not contain enough template bacterial DNA and were subsequently eliminated from the analysis. These PCRs were performed in triplicate using barcoded primers such that each replicate used a different barcode. The concentration of each band was then estimated by comparison to the ladder and pooled at different volumes in an attempt to obtain a solution containing equal amounts of each amplicon. These pools were then library prepped and run on an Illumina MiSeq.

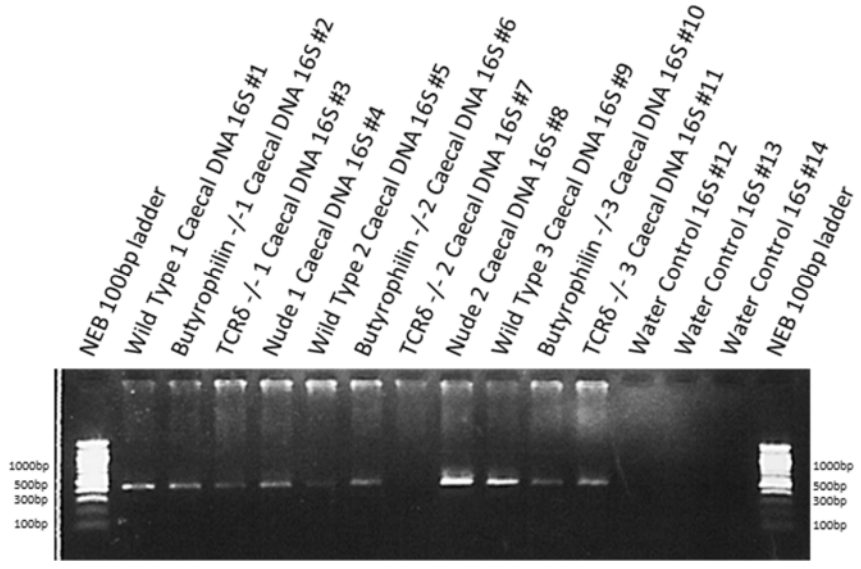


Figure 7.12 – Agarose Gel showing 16S PCR Products from Caecal Tissues
Agarose gel image showing PCR products generated from 16S PCRs. A dual RNA/DNA extraction was used to isolate nucleic acids from the tissue samples. The 16S region of the rRNA gene was then amplified from the DNA using barcoded primers in triplicate for each sample. These products were then run on an agarose gel to confirm a band was present at the expected size, around 500bp. This was done alongside water controls to allow identification of any contaminating 16S sequences introduced during sample processing. 100bp NEB ladders were run alongside the amplicons to allow estimation of concentration and size. The sizes of these bands are indicated in bp at either side of the gel image.

7.3.11 MiSeq 16S Data Processing using the DADA2 Pipeline

The MiSeq run produced a total of 4,414,180 PF reads for the 16S data, with an average Q30 of 68.83% of and a %PF of 86.51. These reads were first demultiplexed using in house code, which involves simply splitting them into separate files according to their primers and barcodes. The R package DADA2 was then used to process the 16S reads. Briefly, this package avoids the issues associated with 97% sequence identity OTUs, which are commonly used by less sophisticated analysis pipelines. DADA2 instead uses using an error correction algorithm allowing the construction of higher resolution Amplicon Sequence Variants (ASVs).

In order to facilitate DADA2 processing, the error prone ends of the forward and reverse reads must first be removed. This trimming process is a compromise between removing potential sequencing errors and retaining enough of an overlap for downstream forward and reverse read merging. The trimming process is informed by read quality plots, as the Q scores of the forward and reverse reads differ from one run to another. In order to inform our trimming strategy, read quality plots were generated using the DADA2 'plotQualityProfile' function (Fig 7.13). In this instance, the forward reads were of substantially better quality than the reverse reads, which begin to drop below Q30 at around cycle 180. As a result, a forward read cut off of 250bp and a reverse cut off of 220bp was used in an effort to remove these more error prone, lower quality reverse read ends.

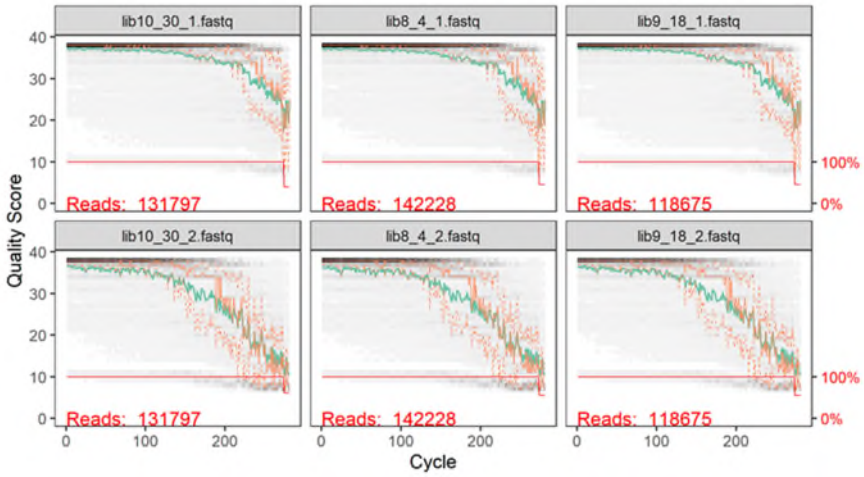


Figure 7.13 – Illumina Sequence read FASTQ Quality Score Plots
Plots show the quality profiles of the Illumina MiSeq FASTQ reads generated using the DADA2 'plotQualityProfile' function. These plots are used to guide downstream read truncation, which is a compromise between retaining as many reads as possible and limiting the predicted number of sequencing errors in the data. In this case, the chosen lengths were 250bp and 220bp for the forward and reverse reads respectively. The X axis represents the number of Illumina cycles as parallel sequencing proceeds one base at a time, and the Y axis is Q score. Q scores are represented by the greyscale heatmap as well as the solid green and orange lines which represent the mean and median Q score at each cycle respectively. The top and bottom dashed orange lines represent the upper and lower quartile respectively. The red line shows the percentage of sequences which extend to that cycle number, since all reads in these libraries are the same length, these are all 300bp in this case.

Once the remaining DADA2 processing steps had been carried out, read tracking plots showing the number of reads lost at each stage were generated (Fig 7.14). These plots provide a diagnostic look at the DADA2 processing and are helpful in identifying any issues which might have occurred. The majority of data is lost in every library at the 'Desired barcoded primer' stage, which involves the removal of any read which does not have the expected forward and reverse primers, in this case E338 and R806, and a recognised barcode associated with it. At this stage, all of the libraries lose around 1/3 of their reads. This loss was mostly due to reads which lack both primers, or reads which only have the R806 primer. These missing primers are presumably just unrecognisable due to sequencing error, since large numbers of reads could not exist without them being present during amplification. The second large attrition of reads occurs during the Filtering stage, in which reads are removed if they have more predicted errors, as determined by examination of the Q score

associated with each base, than an acceptable threshold, in this case 2 for the forward reads and 5 for the reverse.

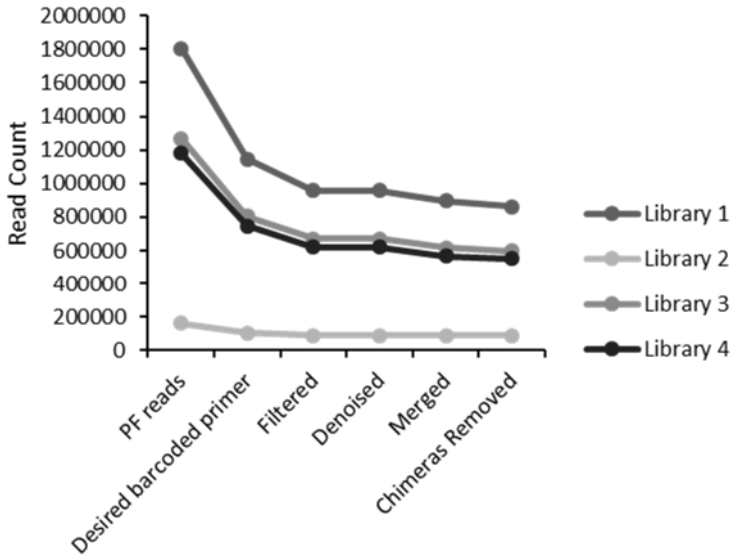


Figure 7.14 – Read Tracking Plots for 16S Sequence Processing
Plot shows the number of reads associated with each library during each stage of DADA2 processing. Libraries 1-3 contain the 16S caecal amplicons and are replicates of one another. Library 4 contains the amplicons from the water controls. The 'Passing Filter (PF) reads' stage refers to the reads which remain after filtering by the MiSeq instrument. 'Desired barcoded primer' refers to reads for which the expected primers and barcodes associated with this experiment were found, and the barcodes pass quality filtering. 'Filtered' refers to reads which pass through the trimming and filtering stage and have fewer predicted errors than the acceptable cut-off. 'Denoised' refers to the DADA2 processing stage involving construction of ASVs by combining more common reads with rarer reads that are predicted by the package to have been generated by sequencing errors. 'Merged' refers to combining forward and reverse reads and 'Chimeras removed' refers to the removal of rarer reads predicted to have originated from more common reads by chimerisation during the PCR stage.

7.3.12 Beta-Diversity Plots Reveal Microbiota Differences between all of the Mouse Strains

β-diversity plots were generated in order to gain an initial insight into the effects of each mutation, and subsequent γδ T cell perturbation, on the caecal microbiota. The strains were compared using multiple different β-diversity indices in order to ensure findings were robust and due to the fact that each has its own strengths and weaknesses (see Chapter 3). Bray-Curtis is a commonly used dissimilarity index used to quantify differences in microbial communities. Comparison of the four mouse strains using Bray-Curtis reveals that the Wild Types cluster

separately from the Nudes and the TCRδ -/- and Butyrophilin -/-, which both cluster together, revealing differences in these microbial communities (Fig 7.15A). However, ordination of this plot by NMDS ordination with 20 random starts gave a very low stress value (rounded to 0), as the Nude mice separate from the other 3 strains based on the NMDS1 axis only. Therefore, this plot provides little information about whether Wild Type, Butyrophilin -/- and TCRδ -/- separate from one another as they are only effectively being compared on a single axis. Nonetheless, PERMANOVA testing of the Bray-Curtis distances gave a highly significant overall P value with an exceptionally high R^2 value of 0.761. This shows that in these plots, mouse strain alone explains the majority of the variation in the caecal microbiota. Despite this difference in microbiota composition, due to the low number of individuals present in this experiment, it is not possible to obtain significant paired strain comparisons even with such a large effect size.

It is important to also consider microbiome composition using a metric which accounts for relatedness between species. This is of particular importance in this case, as the Bray-Curtis NMDS2 axis was likely made up of a single taxon either massively increased or decreased in the Nude strain. Therefore, incorporation of phylogenetic relatedness will likely give a clearer picture of the degree of biological difference between the mouse strains. The weighted UniFrac distance metric considers relatedness when comparing samples[36] and reveals that the Nude mice do not separate on a single axis as they did in the Bray-Curtis plot (Fig 7.15B). This suggests that the sequences responsible for this separation are phylogenetically related to those present in the other strains. Though the overwhelming difference seen in the Nude mice in the Bray-Curtis plot is diminished when compared using UniFrac, there is still good separation of these points from those of the Wild Type mice. Additionally, in this plot the TCRδ -/- and Wild Type strains do not separate at all, suggesting that the sequences that caused them to cluster separately in the Bray-Curtis plots were all closely related and that when factoring in sequence phylogeny, the two communities are relatively similar. The Butyrophilin -/- mice however, remain separated from individuals of the other 3 strains, suggesting community differences remain even when relatedness is considered. This is confirmed by

PERMANOVA testing, which gives a significant P value and high R^2 value of 0.596, showing that differences are present, and strain is still the main determinant of microbiome β-diversity when calculated using UniFrac distance.

It has recently been convincingly argued in the 16S literature that data obtained from Illumina sequencing is compositional and should be analysed with this in mind[37]. This argument was justified by the fact that during Illumina sequencing, DNA molecules within a library compete for slots on the flow cell and therefore an increase in the numbers of one read will lead to a corresponding decrease in another. In accordance with this, RDA plots were produced using Aitchison's distance in order to compare the caecal microbiotas of the four mouse strains whilst accounting for the compositional nature of the data (Fig 7.15C).These plots reveal a distinct difference in microbiota composition in each of the strains, which cluster apart from each other with no overlap. The differences between the strains are even more pronounced than in the Bray-Curtis or UniFrac plots and all mice within each strain cluster extremely close to one another. Unsurprisingly then, PERMANOVA testing again gives a significant overall P value as well as a high R^2 value of 0.632. This shows that as with the other metrics, the differences between the microbiotas of each mouse strain and very large and highly significant.

Similarly to Bray-Curtis, Aitchison's distance does not consider relatedness, therefore a compositional equivalent of UniFrac, PhILR is used, which takes the previously calculated phylogenetic tree as an additional input[37] (Fig 7.15D). These plots are almost identical to those produced using Aitchison's distance, though the strains cluster even further apart from one another when using the PhILR metric. Again, the PERMANOVA P value is significant and R^2 of 0.654 shows that strain is the main determinant of microbiota β-diversity in these plots.

Overall then, the β-diversity plots showed that mouse strain has major implications on the overall β-Diversity of the microbiota, and these differences were totally robust to the distance

metric used. Furthermore, all of the FDR corrected PERMANOVA P values were significant and R^2 values very high showing that the differences observed were both significant and large.

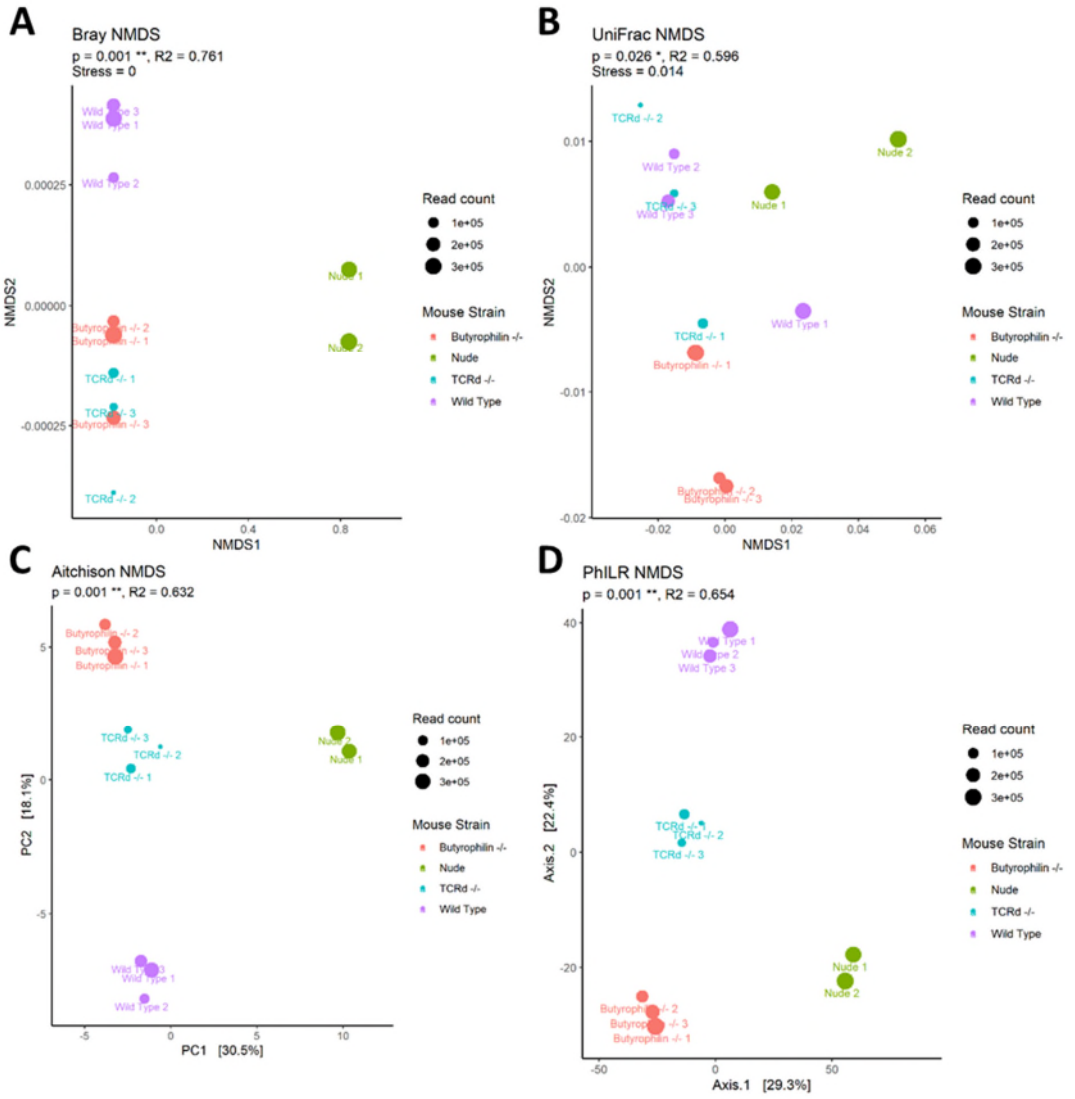


Figure 7.15 – 16S β-Diversity Plots by Strain and Tissue reveal clear Differences between the Caecal Floral of each of the Mouse Strains

β-diversity plots were generated using **A)** Bray-Curtis, **B)** Weighted UniFrac, **C)** Aitchison's and **D)** PhILR distances calculated from the 16S data. NMDS was performed using 20 random starts and the stress associated with each plot is indicated below the title. The percentage of variation explained by each of the RDA axes is displayed in brackets. Data for calculation of Aitchison's and PhILR distances was first center log ratio (clr) transformed. The phylogenetic trees used by UniFrac and PhILR were generated by first aligning with MAFFT, followed by computation of pairwise distances, production of an initial tree using Neighbour joining, optimisation of model parameters and production of the final tree using maximum likelihood. Different mouse strains are indicated by colour and point size represents the number of reads associated with that sample, giving a crude measure of reliability. The strain and mouse number are indicated below each point. PERMANOVA testing was performed using 999 permutations and the P values FDR corrected. Significance indicators were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05 and R2 values are displayed beneath the title. PERMANOVA did not have sufficient power to give significant results for individual paired comparisons and so these were omitted.

7.3.13 Microbiome Taxonomic Composition Plots Reveal Differences between the Mouse Strains at the Phylum and Order Level

In order to identify the source of these community differences, stacked barplots were generated at both the Phylum and Order levels (Fig 7.16). The taxonomic identity of a 16S sequence is determined by comparing it to an annotated, pre-existing 16S database (Ribosomal Database Project). In some cases, the sequence is absent from the database and therefore cannot be identified. Predictably, this becomes more likely the lower the taxonomic level, to the extent that species assignment is rare in 16S studies. Unfortunately, at both the Phyla and Order levels, an unpaired Wilcoxon signed rank test did not have the power to detect significant differences in these plots, despite the fact that some of the differences we observed were large.

Initial examination of the microbiota at the Phylum level reveals that all of the mice in this study contain the same phyla and the rough proportions match previous work[38–40] (Fig 7.16A). The majority of the microbiota in these mice was made up of *Bacteroidetes*, *Firmicutes* and smaller amounts of *Proteobacteria*. As well as these, several other less common phyla make up the remaining minority of the microbiota.

Comparing the mouse strains, though the ratios of *Firmicutes* and *Bacteroidetes* are broadly similar, some differences are present between them. Comparing first the Wild Type to the Butyrophilin -/- and TCRδ -/- strains, there is a slight increase in *Bacteroidetes* at the expense of *Firmicutes* in both, however this difference is particularly pronounced in the Butyrophilin -/- mice. Additionally, Butyrophilin -/- and TCRδ -/- have substantially less *Proteobacteria* than the Wild Type mice. The Butyrophilin -/- also has far greater levels of *Verrucomicrobia*, which is entirely absent from both Wild Type and TCRδ -/-. The Wild Type and Nude strains appear more similar in terms of *Firmicutes* and *Bacteroidetes*, but the Nude mice have far less *Proteobacteria* and slightly more *Tenericutes*.

Comparing the strains at the order level reveals more about these taxonomic differences (Fig 7.16B). Most strikingly, the decreases in *Proteobacteria* seen in all of the strains are clearly due to a marked decrease in the proportion of *Campylobacterales* when compared to Wild Type. This effect is further emphasised in the Nude mouse due to an additional decrease in *Desulfovibrionales* which is also a member of the *Proteobacteria*. The changes to the ratio of *Bacteroidetes* and *Firmicutes* in the Butyrophilin -/- and TCRδ -/- mice is clearly characterised by the Orders *Bacteroidales* and *Clostridiales* respectively, with the *Bacteroidales* being increased in the Butyrophilin -/- mice in particular.

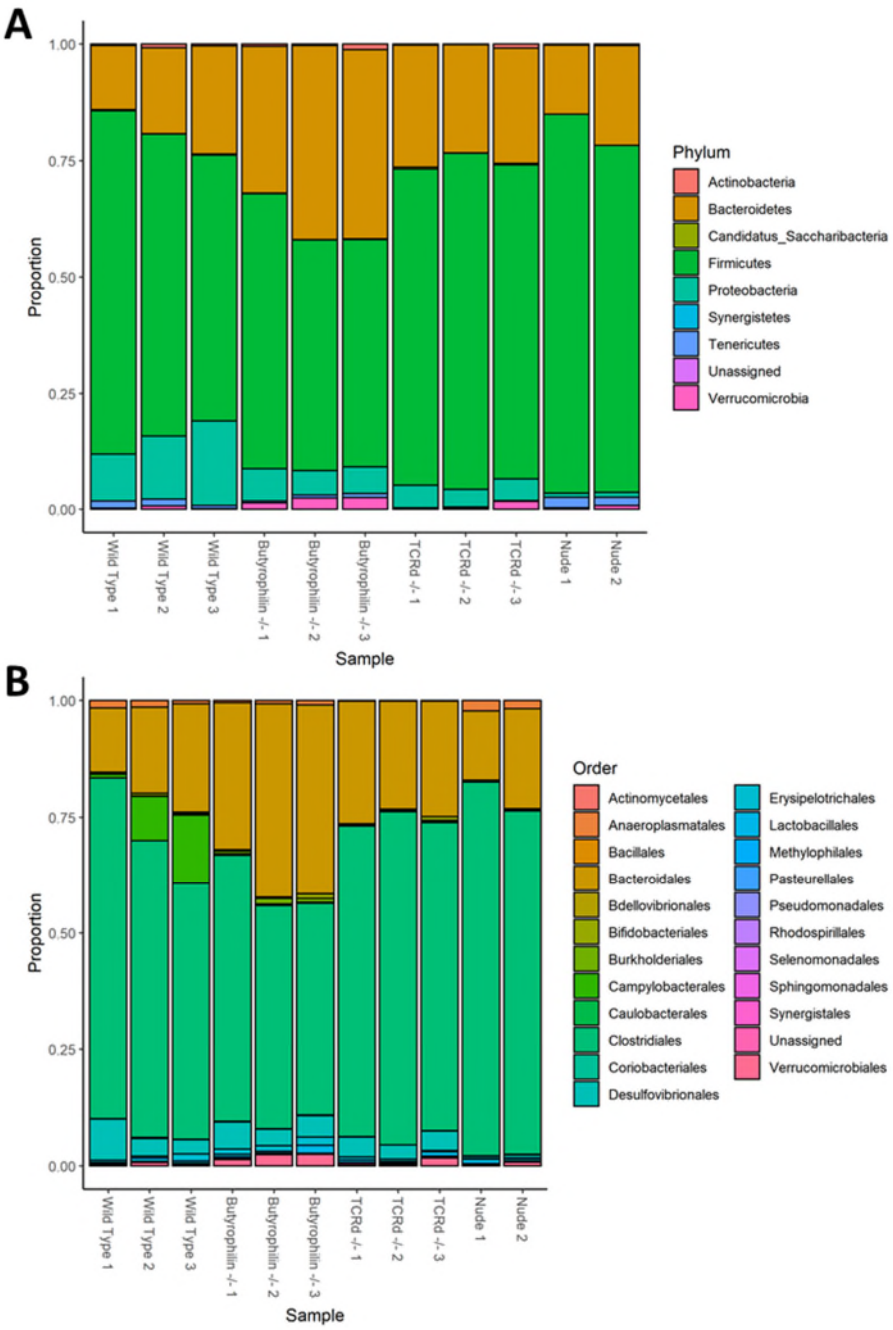


Figure 7.16 – 16S Microbiome Composition Plots reveal a large Decrease in *Campylobacterales* in all Mouse Strains except Wild Type

Stacked barplots show the relative proportion of the overall microbiota made up of each taxa plotted at both the **A)** Phylum and **B)** Order level. Phylogenetic information was assigned to each 16S sequence using the DADA2 assignTaxonomy Function. These taxa are then plotted in alphabetical order revealing the overall composition of the microbiota. Each colour is assigned to a different taxa as shown in the legend. At the phylum level, the visible taxa from top to bottom are; Red - Actinobacteria, Brown - Bacteroidetes, Green - Firmicutes, Turquoise - Proteobacteria, Blue - Tenericutes, Pink – Verrucomicrobia. At the Order level, the visible taxa from top to bottom are Orange - Anaeroplasmatales, Brown - Bacteroidales, Green - Campybacterales, Turquoise - Clostridales, Cyan - Desulfovibrionales, Blue - Lactobacillales, Pink – Verrucomicrobiales. Sequences for which taxonomy could not be assigned are grouped into the 'unassigned' category. Wilcoxon signed rank testing was performed but results are not shown as power was insufficient.

7.3.14 Individual Genus Plots Reveal Further Differences between the Mouse Strains, Particularly the Decrease in *Helicobacter* in all but the Wild Type mice

Genus level bar plots were produced in order to facilitate an examination of the previously observed taxonomic differences and gain a deeper insight into their potential implications. The literature regarding the effects of individual genera on the microbiome and host health is fairly well developed in mouse and this analysis allows us to draw conclusions based upon this literature.

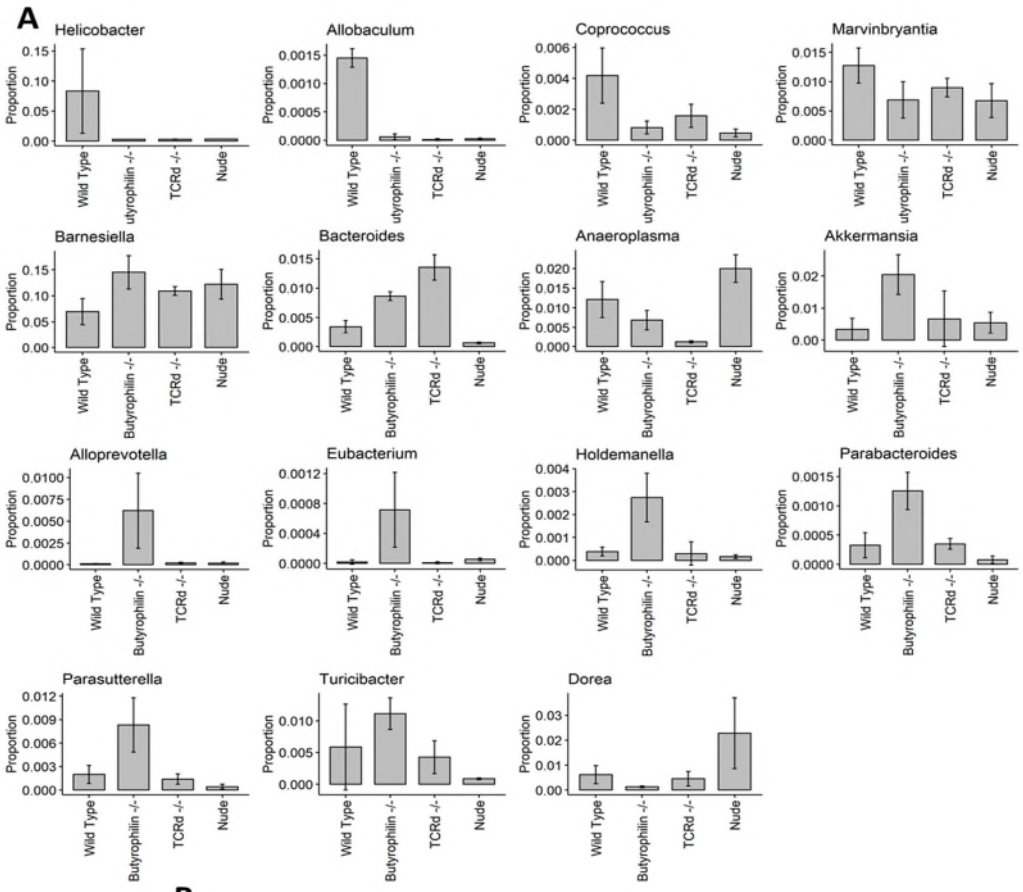
Comparing all of the strains to Wild Type, several differences in Genera are evident (Fig 7.17). The biggest single difference between the mouse strains is in *Helicobacter,* which is the sole member of the Order *Campylobacterales*, mentioned previously in the Order plots (Fig 7.16B). *Helicobacter* is present in the Wild Type at a proportion of ~9%, but it is barely detectable in the Butyrophilin -/-, TCRδ -/- and Nude mice. Similarly, *Allobaculum* is markedly decreased in the other 3 strains compared to Wild Type, though it does make up a far lower proportion of the Wild Type microbiota at only ~0.15%. *Coprococcus* again shows a similar pattern to *Helicobacter* and *Allobaculum* and is greatly decreased in the other 3 strains compared to the Wild Type where it is present at 0.4%. *Marvinbryantia* makes up ~1.2% of the microbiota in Wild Type but is also decreased in the Butyrophilin -/-, TCRδ -/- and Nude mice down to 0.7, 0.9 and 0.7% respectively. However, this decrease is far less striking than that observed in *Helicobacter, Coprococcus* or *Allobaculum.* Conversely, some genera are also reduced in the Wild Type mice compared to the mutant mouse strains. *Barnesiella* is substantially increased in Butyrophilin -/-, TCRδ -/- and Nude mice, where it is present at ~11% ~15% and ~12% respectively, compared to ~7% in Wild Type.

There are also genera differences which tend to be similar in the Butyrophilin -/- and TCRδ-/- mice. For instance, *Bacteroides* makes up ~8% of the microbiota in Butyrophilin -/- and ~13% in the TCRδ -/- but just 3% in the Wild Type and ~1% in the Nude. *Anaeroplasma* shows similar patterns but in the opposite direction, making up ~1.2% of the microbiota in the Wild Type and ~2% of the

microbiota in the Nude mice, but is reduced down to ~0.6 and 0.1% in the Butyrophilin -/- and TCRδ -/- mice respectively.

Interestingly, some Genera are increased in just the Butyrophilin -/- strain such as the *Akkermansia* which is present at ~2% of the microbiota compared to ~0.8% in the TCRδ -/-. Likewise, *Alloproteovella* is present in the Butyrophilin -/- at ~0.6% but drops down substantially to ~0.02% in TCRδ -/-. Additionally, a similar drop is also present in *Anaeroplasma,* which decreases from 0.6% in Butyrophilin -/- to 0.1% in TCRδ -/-. *Eubacterium, Holdemanella, Parabacteroides, Parasutterella, and Turicibacter* all show similar patterns and are markedly decreased in TCRδ -/- compared to Butyrophilin -/-. Additionally, *Dorea* is substantially reduced in all strains except the Nude, where it makes up ~2.3% of the microbiota.

These genera level differences suggest that the fine balance of microbes in the mouse gut is partially maintained by the γδ T cells and their perturbance can have large impacts on individual genera within the microbiota.



B

Phylum	Order	Genus
Proteobacteria	*Campylobacterales*	*Helicobacter*
Erysipelotrichia	*Erysipelotrichales*	*Allobaculum*
Firmicutes	*Clostridales*	*Coprococcus*
Firmicutes	*Clostridales*	*Marvinbryantia*
Bacteroidetes	*Bacteroidales*	*Barnesiella*
Bacteroidetes	*Bacteroidales*	*Bacteroides*
Tenericutes	*Anaeroplasmatales*	*Anaeroplasma*
Verrucomicrobia	*Verrucomicrobiales*	*Akkermansia*
Bacteroidetes	*Bacteroidales*	*Alloprevotella*
Firmicutes	*Clostridiales*	*Eubacterium*
Firmicutes	*Erysipelotrichales*	*Holdemanella*
Bacteroidetes	*Bacteroidales*	*Parabacteroides*
Proteobacteria	*Burkholderiales*	*Parasutterella*
Firmicutes	*Erysipelotrichales*	*Turicibacter*

Figure 7.17 – Individual Genera Plots reveal Fine Scale Microbiota Changes in each of the Mouse Strains
A) Plots show the mean proportions of each selected genera split by mouse strain. Genera of interest were selected by SIMPER analysis, which allows calculation of the contribution of each sequence to the overall Bray-Curtis distances. The genera to which these sequences belong were then examined and selected for display based on the differences in their proportions between the mouse strains. Proportions displayed are of the mean total microbiota for each strain and error bars are standard deviation. Wilcoxon signed rank testing was performed but results are not shown, as power was insufficient to obtain a significant P value.
B) Table shows the Phylum and Order that each plotted genus belongs to allow easy interpretation and reference back to Fig 7.16.

7.4 Discussion

This chapter aimed to first confirm and further investigate the previously reported phenotype that TRGV7+ γδ T cells are decreased in the gut in the absence of Btnl1 in comparison to Wild Type. This work was done alongside a nude mouse strain, allowing parallel investigation of the impacts of a lack of thymic development on the TCR gamma repertoire. Furthermore, to explore the effects of these γδ T cell perturbations on the microbiota, 16S sequencing was used to probe the caecal microbiota of these strains, alongside an additional group of TCRδ -/- mice lacking all functional γδ T cells.

As was observed previously in the literature, we confirmed TRGV7 to be the dominant productively rearranged V gene expressed in the mouse gut[7,41]. This TRGV7 dominance was more evident in the Duodenum and Jejunum and became less apparent as we descended down the gut, with TRGV2 becoming the dominant gene in the Ileum and Caecum. Though some previous literature has stated that TRGV2 is a rare rearrangement, not known to be associated with the gut[42], there is also evidence that it is actually relatively common[41]. The reason for this disparity is not known, since to our knowledge, this subset has not been examined in the gut in the detail the TRGV7+ γδ T cells have.

Furthermore, the levels of publicity observed in the Wild Type mouse TCR gamma repertoire were generally very high. TRGV5 and particularly 6 had the highest levels of publicity, which is unsurprising considering their extremely limited diversity, which was evident in our data and has previously been observed in the literature[43,44]. γδ T cells expressing these highly public TRGV gene segments originate from the first 2 waves of thymic export and are thought to be important in early immunity while TdT expression is low and therefore cannot enable the generation of large amounts of additional diversity through the addition/deletion of N nucleotides[44,45]. γδ T cells expressing TRGV5 are known to localise to the skin, where they are also known as Dendritic Epidermal T Cells

(DETCs), whereas TRGV6 is found mostly in the female reproductive tract[46]. The function of these subsets in the gut is unknown, but they may play a role in early immunity before they are diluted out by TRGV7 export from the thymus in later waves.

The repertoire data presented in this chapter, confirmed that the proportion of TRGV7 is significantly reduced in the absence of Btnl1, as has been shown previously by flow cytometry[7]. However, the proportional reduction was around 6 fold, whereas others have reported a ~90% reduction. This difference is likely attributable to the proportionality of the data, especially considering that the overall γδ T cell number will also be greatly decreased and so the reduction in proportion will underestimate the actual numerical decrease. As well as replicating this finding, this work also provides a strong validation for the TCR gamma repertoire sequencing method and confirms its ability to detect repertoire differences at a coarse scale, supporting work in previous chapters, as well as providing more detailed information. Furthermore, we also confirmed the finding that the rest of the repertoire is unaffected in the absence of Btnl1, and showed that this was true of the clonal structure as well as simply the number of TRGV7-ve γδ T cells, as was shown previously[7].

Focusing in on the TRGV7+ repertoire specifically, we showed that a lack of Btnl1 also led to a change in the population structure. This corresponded to slight differences in diversity and a specific reduction of the most prominent, germline TRGV7 expressing clone. These findings provide a deeper insight into the role of Btnl1, and suggest that as well as being a mediator of overall localisation of TRGV7 to the gut, it appears to act in a more specific way, and may have a stronger effect on certain TRGV7+ γδ T cell subsets. Though the mechanism of action of Btnls is yet to be fully elucidated, some interesting findings have recently been published in this area. Multiple authors have found that Btnls seem to exert their effects as heterodimers[47–49]. In the case of Btnl1 it forms a heterodimer with Btnl6. Curiously however, it seems that there is a regulatory hierarchy in these heterodimers. Btnl6 -/- mice display a similar but less pronounced phenotype compared with Btnl1 -

/- mice, with just a 50% reduction in gut TRGV7+ γδ T cells[50]. Additionally, Btnl6 seems to bind directly to a TRGV7 TCR in vitro, suggesting a mechanism that operates through direct binding[51,52]. Our findings suggest that this direct binding may favour some clones over others and selectively increase the rate of localisation or retention. The repertoire data suggests that the clones more likely to be favoured by this binding tend to be more germline, public clones, though more in depth in vitro binding assays would need to be done to confirm this. Although this same Btnl6 binding study suggests that interface occurs independent of the CDR3 region, studies on human Vγ9Vδ2+ T cell reactivity to phosphoantigens shows that this interaction requires BTN2A1, in addition to the BTN3A1/BTN3A2 heterodimer, in order to facilitate activation. While it has been shown that BTN2A1 binds to TRGV9 framework regions, another interaction site exists, involving one of these other 2 BTNs, and this second interacting region includes the TCR gamma chain CDR3 loop[47]. It has been argued that this may also be the case with the mouse Btnl1/6 heterodimer which may interface with the CDR3 region by other third party molecules, potentially also Btnls, and that these isolated interactions may be of insufficient affinity to be detected alone[53].

Increases in TRGV2 were also observed in the Butyrophiliin -/- strain. This is potentially an artefact of the proportional nature of the repertoire data and represents one downside to this form of repertoire analysis. Due to the fact that the sequence data obtained from the Illumina MiSeq is compositional, a decrease in one group of sequences must lead to an increase in another group of sequences. In this instance, as there is strong prior evidence that the reduction in TRGV7 is associated with a reduction in cells expressing TRGV7, it may therefore be the case that this corresponding observed increase in TRGV2 simply reflects a proportionate increase rather than an increase in actual cell numbers. However, the fact that increases in the other TRGV segments were absent, supports this increase being at least in part associated with a compensatory specific increase in TRGV2 expressing cells. Although repertoire studies provide more detailed information, they should be followed up or combined with other quantitative methods, such as histology, flow cytometry or qPCR.

The role of the thymus in development of gut γδ T cells remains unclear, with some investigators reporting reduced cell numbers in its absence[7,54] and others reporting that γδ T cell numbers are unchanged[55]. In terms of the TCR gamma repertoire itself, previous reports show that TRGV5 in the skin is absent from Nude mice, however, the proportion of gut TRGV7+ γδ T cells appears unaffected[7]. Our data supports the idea that TRGV7 rearrangement occurs independent of the thymus as its proportions remains unchanged in the Nude mice compared with the Wild Type. Interestingly, the remaining TRGV gene proportions were almost identical between the Wild Type and Nude mice, suggesting the remainder of the gut repertoire is also able to rearrange extrathymically, or at least independent of the nude mutation. Despite the similarities between the TRGV segment usage profiles, Nude mice did show a large reduction in every diversity metric in every tissue. One possible explanation for this is that the reduced gut γδ T cell numbers previously reported are responsible for this reduced diversity. Concordant with this, reductions in Nude UPA and TRGV7 band intensity on gels and subsequently reduced read numbers were observed, though this should be interpreted with caution. Additionally, Nude mice showed increased proportions of the canonical TRGV7 rearrangement consistent with the reduced diversity observed in tissues. Taken together, this data suggests that the classically functional thymus is not required for TRGV7 rearrangement, suggesting it may occur extrathymically as previously suggested. Though in the absence of the thymus, rearrangements appear to possess less junctional diversity and favour canonical CDR3s. How this may affect γδ T cell function in the gut is as yet unclear.

The 16S compositional and taxonomic plots revealed differences in the microbial communities of each mouse strain. A deficiency in γδ T cells or a change to the enteric TCR gamma repertoire therefore clearly influences the microbiota and may do so through various different mechanisms. γδ T cells are known to produce a range of cytokines, involved in induction and suppression of inflammation at the gut. They produce pro-inflammatory cytokines such as IFN-γ, IL17 and TNF-α[56], but also the anti-inflammatory cytokines TGF-β and IL-10[57,58].

Additionally, γδ T cells are reported to influence IgA production, though whether they result in upregulation or downregulation of IgA responses is currently disputed in the literature[59–61]. Interestingly, a recent finding from Melo-Gonzalez et al. 2019[62] observed that in the absence of ILC3s, a lineage of innate-like lymphoid cells related to γδ T cells[63], the profile of IgA bound bacteria was altered. The data presented here indicate a reduction in *Helicobacter* which is paralleled by data presented in the Melo-Gonzalez paper, which also observed decreased *Helicobacter,* as well as *Mucispirillum*, in the absence of ILC3s and went on to show that the amount of IgA binding to *Helicobacter* was increased. They propose a mechanism by which ILC3s impose a checkpoint in the gut IgA response, reducing its targeting of potentially beneficial microbes. Though antibody responses are most often regarded as a mechanism by which pathogens are eliminated, low affinity IgA in the gut has been shown to play a homeostatic role, and is involved in retaining beneficial microbes, wheras higher affinity, specific IgA is involved in punitive responses against harmful microbes[64–68]. This work and the interplay between γδ T cells and IgA, hints at potential mechanism by which γδ T cells may influence the gut flora. The recent development of IgA-seq studies provide a robust technique by which to explore this relationship.

Though historically classified as an obligate pathogen and the causative agent of gastritis, stomach ulcers and stomach cancer[69,70], more recent studies have challenged this paradigm, and suggested that under certain circumstances *Helicobacter* may have beneficial effects. Some studies have shown that certain non-pathogenic strains of *Helicobacter* have beneficial effects in the stomach, controlling stomach acid secretion[71,72] and regulating appetite[73]. Furthermore, in the majority of infected individuals, *Helicobacter* has no adverse effects and has even been shown to be protective against certain diseases including multiple sclerosis[74], celiac disease[75,76], gastroesophageal reflux disease[77] and oesophageal cancer[78]. This mix of both beneficial and detrimental effects suggests a more complex role for *Helicobacter* than was first appreciated. In the murine gut, equally mixed and complex relationships between *Helicobacter* and the host exist. Various studies suggest that while *Helicobacter* induces a pro-inflammatory colitogenic response through T effector subsets

in the context of pre-existing inflammation, in the context of gut homeostasis *Helicobacter* presence produces an anti-inflammatory response[79–84]. This *Helicobacter* mediated downregulation of inflammation during homeostasis has been shown to be facilitated through IL10 production by CD4+ve Tregs[79–81,83,84] and macrophages[82]. Therefore, while *Helicobacter* can exacerbate pre-existing inflammation, in the healthy gut it clearly plays a beneficial role in combatting chronic inflammation and thus our findings suggest that γδ T cells may play a role in facilitating colonisation and survival of microbial species which can confer benefits to the host during gut homeostasis. Furthermore, *Helicobacter* is a prime example of the fact that microbe host interactions are rarely simple and often dependant on context.

Interestingly, in the microbiota β-diversity plots, the Butyrophilin -/- mice appeared to be a more extreme version of the TCRδ -/- phenotype, clustering further away from Wild Type and showing a greater increase in *Bacteroidetes* at the expense of *Firmicutes* in the taxonomic Class stacked barplots. This seems to suggest that TRGV7+ γδ T cells are important for inducing downregulation of responses against certain species, consistent with their status as an IFNγ producing subset[85], whereas other γδ T cell subsets present in the gut may play more of a role in response and elimination to potentially harmful microbes.

As the cost of sequencing technologies has rapidly decreased, the number of microbiome studies has increased exponentially over the past 15 years. As microbiome studies have become more and more widespread a parallel literature warning of the pitfalls and advising the best practises has emerged. One of the major criticisms of studies looking to associate a gene with an effect on the microbiome is failure to account for the maternal effect in study design. The maternal effect was shown in a paper by Ubeda *et al*.[86] to be a very common flaw in much of the previous literature involving immune knockouts. This paper highlights the impact of the maternal effect by comparing the microbiotas of MyD88 and TLR knockouts with Wild Type mouse strains from two different breeding strategies. One set of strains were bred in isolation from one another for 5+ years

and the other set bred from F1 heterozygotes. Consistent with previous work, the authors did find differences in the strains bred in isolation but no differences in the F2 mice. Therefore, this paper and others emphasise the point that maternal effects should be considered in microbiota studies and accounted for in experimental design[86,87]. Additionally, animal housing can have a profound effect on the microbiota. Cage effects refer to the seemingly random simultaneous microbiota drift of mice living in the same cage[88]. Though its impact on the microbiota can be huge, the cause of this effect is as yet unclear and may be driven by simple stochastic events and coprophagy by co-housed mice. Current best practises recommend separating mice into as many cages as is feasible in order to dilute out potential cage effects. Additionally, and of particular relevance to the observed differences, *Helicobacter* has been shown to be a major driver of these cage effects[89,90].

This study was primarily designed to confirm previous findings that the proportion of TRGV7 rearrangements is substantially reduced in Butyrophilin -/- mice and assess the ability of repertoire sequencing to replicate findings from flow cytometry data. As a result, it does not account for the possibility that the observed differences arose from maternal or cage effects. Therefore, in order to show that these findings are robust to maternal and cage effects, the experiment should be replicated in a greater number of Wild Type and Butyrophilin/TCRδ knockout mice bred from heterozygous parents.

7.5 Conclusions

The work in this chapter aimed to replicate and further investigate the impacts of Btnl1 on mouse γδ T cells and also look at the effects of these perturbations on the microbiota. The findings presented here support the role of Btnls as mediators of the localisation of γδ subsets to epithelial sites and also suggests that they may exert a stronger effect on certain canonical subsets. Our data also supports the case for extrathymically derived gut γδ T cells, though we showed a massive reduction in diversity in athymic Nude mice. We also discovered large changes to the caecal mucosal microbiota using 16S sequencing, most notably the reduction of *Helicobacter* in the Butyrophilin -/-,

TCRδ -/- and Nude mouse strains. This indicates a potential role for γδ T cells as a regulator of responses against potentially beneficial microbes in the gut, though subsequent experiments using heterozygous mutants are needed to confirm this finding.

7.6 References

1. Holderness J, Hedges JF, Ramstead A, Jutila MA. Comparative Biology of γδ T Cell Function in Humans, Mice, and Domestic Animals. Annu Rev Anim Biosci. 2013;1(1):99–124.
2. Girardi M, Oppenheim DE, Steele CR, Lewis JM, Glusac E, Filler R, et al. Regulation of cutaneous malignancy by γδ T cells. J Immunol. 2018;200(9):3031–5.
3. Girardi M. Immunosurveillance and immunoregulation by γδ T cells. J Invest Dermatol. 2006;126(1):25–31.
4. Lefranc M. IMGT [Internet]. 2021.
5. Guy-Grand D, Vassalli P, Eberl G, Pereira P, Burlen-Defranoux O, Lemaitre F, et al. Origin, trafficking, and intraepithelial fate of gut-tropic T cells. J Exp Med. 2013;210(9):1839–54.
6. Hayday A, Vantourout P. A Long-Playing CD about the γδ TCR Repertoire. Immunity. 2013;39(6):994–6.
7. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.
8. Boyden LM, Lewis JM, Barbee SD, Bas A, Girardi M, Hayday AC, et al. Skint1 , the prototype of a newly identified immunoglobulin superfamily gene cluster , positively selects epidermal γδ T cells. 2008;40(5).
9. Afrache H, Gouret P, Ainouche S, Pontarotti P, Olive D. The butyrophilin (BTN) gene family : from milk fat to the regulation of the immune response. 2012;1.
10. Collins M, Ling V, Carreno BM. The B7 family of immune-regulatory ligands. Genome Biol. 2005;6(6):1–7.
11. Compte E, Pontarotti P, Collette Y, Lopez M, Olive D. Frontline : Characterization of BT3 molecules belonging to the B7 family expressed on immune cells. 2004;(34):2089–99.
12. Arnett HA, Escobar SS, Viney JL. Regulation of costimulation in the era of butyrophilins. Cytokine. 2009;46(3):370–5.
13. Chen Y, Chou K, Fuchs E, Havran WL, Boismenu R. Protection of the intestinal mucosa by intraepithelial γδ T cells. Proc Natl Acad Sci U S A. 2002;99(22):14338–43.
14. Komano H, Fujiura Y, Kawaguchi M, Matsumoto S, Hashimoto Y, Obana S, et al. Homeostatic regulation of intestinal epithelia by intraepithelial gamma delta T cells. Proc Natl Acad Sci U S A. 1995;92(13):6147–51.
15. Boismenu R, Havran WL. Modulation of epithelial cell growth by intraepithelial γδT cells. Science (80-). 1994;266(5188):1253–5.
16. Dalton JE, Cruickshank SM, Egan CE, Mears R, Newton DJ, Andrew EM, et al. Intraepithelial

γδ+ Lymphocytes Maintain the Integrity of Intestinal Epithelial Tight Junctions in Response to Infection. Gastroenterology. 2006;131(3):818–29.

17. Edelblum KL, Shen L, Weber CR, Marchiando AM, Clay BS, Wang Y, et al. Dynamic migration of γδ intraepithelial lymphocytes requires occludin. Proc Natl Acad Sci U S A. 2012;109(18):7097–102.

18. Ismail AS, Severson KM, Vaishnava S, Behrendt CL, Yu X, Benjamin JL, et al. Γδ Intraepithelial Lymphocytes Are Essential Mediators of Host-Microbial Homeostasis At the Intestinal Mucosal Surface. Proc Natl Acad Sci U S A. 2011;108(21):8743–8.

19. Fahrer AM, Konigshofer Y, Kerr EM, Ghandour G, Mack DH, Davis MM, et al. Attributes of γδ intraepithelial lymphocytes as suggested by their transcriptional profile. Proc Natl Acad Sci U S A. 2001;98(18):10261–6.

20. Shires J, Theodoridis E, Hayday AC. Biological insights into TCRγδ+ and TCRαβ+ intraepithelial lymphocytes provided by serial analysis of gene expression (SAGE). Immunity. 2001;15(3):419–34.

21. Hoytema van Konijnenburg DP, Reis BS, Pedicord VA, Farache J, Victora GD, Mucida D. Intestinal Epithelial and Intraepithelial T Cell Crosstalk Mediates a Dynamic Response to Infection. Cell. 2017;171(4):783-788.e13.

22. Fiala GJ, Gomes AQ, Silva-Santos B. From thymus to periphery: Molecular basis of effector γδ-T cell differentiation. Immunol Rev. 2020;298(1):47–60.

23. Kohyama M, Nanno M, Kawaguchi-Miyashita M, Shinichiro S, Watanabe M, Hibi T, et al. Cytolytic and IFN-γ -Producing Activities of γ δ T Cells in the Mouse Intestinal Epithelium Are T Cell receptor-β -chain Dependent. 1999;96(13):7451–5.

24. Silva-Santos B, Serre K, Norell H. γδT cells in cancer. Nat Rev Immunol. 2015;15(11):683–91.

25. Hiromatsu K, Yoshikai Y, Matsuzaki G, Ohga S, Muramori K, Matsumoto K, et al. A protective role of γ/δ T cells in primary inaction with listelqa monocytogenes in mice. J Exp Med. 1992;175(1):49–56.

26. Sheridan BS, Romagnoli PA, Pham QM, Fu HH, Alonzo F, Schubert WD, et al. γδ T Cells Exhibit Multifunctional and Protective Memory in Intestinal Tissues. Immunity. 2013;39(1):184–95.

27. Stark MA, Huo Y, Burcin TL, Morris MA, Olson TS, Ley K. Phagocytosis of apoptotic neutrophils regulates granulopoiesis via IL-23 and IL-17. Immunity. 2005;22(3):285–94.

28. Shibata K, Yamada H, Hara H, Kishihara K, Yoshikai Y. Resident Vδ1 + γδ T Cells Control Early Infiltration of Neutrophils after Escherichia coli Infection via IL-17 Production . J Immunol. 2007;178(7):4466–72.

29. McKenzie DR, Comerford I, Silva-Santos B, McColl SR. The emerging complexity of γδT17 Cells. Front Immunol. 2018;9(APR).

30. Papotto PH, Ribot JC, Silva-Santos B. IL-17 + γδ T cells as kick-starters of inflammation. Nat Immunol. 2017;18(6):604–11.

31. Ismail AS, Behrendt CL, Hooper L V. Reciprocal Interactions between Commensal Bacteria and γδ Intraepithelial Lymphocytes during Mucosal Injury. J Immunol. 2009;182(5):3047–54.

32. Khairallah C, Chu TH, Sheridan BS. Tissue Adaptations of Memory and Tissue-Resident Gamma Delta T Cells. Front Immunol. 2018;9:1–22.

33. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1).

34. Atsuta N, Nishimura H, Nakamura N, Emoto M, Iwatsuki T, Yoshikai Y. Diversity of Vγ gene segments rearranged to the Jγ4 gene in mice. J Immunol. 1995;154(2):676–84.

35. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol Evol. 2016;7(12):1451–6.

36. Lozupone C, Knight R. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. Appl Environ Microbiol. 2005;71(12):8228–35.

37. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8:1–6.

38. Riboulet-Bisson E, Sturme MHJ, Jeffery IB, O'Donnell MM, Neville BA, Forde BM, et al. Effect of lactobacillus salivarius bacteriocin ABP118 on the mouse and pig intestinal microbiota. PLoS One. 2012;7(2).

39. Sun BL, Li WW, Wang J, Xu YL, Sun HL, Tian DiY, et al. Gut Microbiota Alteration and Its Time Course in a Tauopathy Mouse Model. J Alzheimer's Dis. 2019;70(2):397–410.

40. Ward NL, Pieretti A, Dowd SE, Cox SB, Goldstein AM. Intestinal aganglionosis is associated with early and sustained disruption of the colonic microbiome. Neurogastroenterol Motil. 2012;24(9):874–85.

41. Takagaki Y, Decioux A, Bonneville M, Tonegawa S. Diversity of γδ T -cell receptors on murine intestinal intra- epithelial lymphocytes. Lett To Nat. 1989;339:712–4.

42. Lalor SJ, McLoughlin RM. Memory γδ T Cells–Newly Appreciated Protagonists in Infection and Immunity. Trends Immunol. 2016;37(10):690–702.

43. Asarnow DM, Goodman T, LeFrancois L, Allison JP. Distinct antigen receptor repertoires of two classes of murine epithelium-associated T cells. Nature. 1989;341(6237):60–2.

44. Carding SR, Egan PJ, Walter T, Hall E. γδ T Cells : Functional Plasticity and Heterogeneity. Immunology. 2002;2(May):336–45.

45. Aono A, Enomoto H, Yoshida N, Yoshizaki K, Kishimoto T, Komori T. Forced expression of terminal deoxynucleotidyl transferase in fetal thymus resulted in a decrease in γδ T cells and random dissemination of Vγ3Vδ1 T cells in skin of newborn but not adult mice. Vol. 99, Immunology. 2000. p. 489–97.

46. Vantourout P, Hayday A. Six-of-the-best: Unique contributions of γδ T cells to immunology. Nat Rev Immunol. 2013;13(2):88–100.

47. Rigau M, Ostrouska S, Fulford TS, Johnson DN, Woods K, Ruan Z, et al. Butyrophilin 2A1 is essential for phosphoantigen reactivity by γδ T cells. Science (80-). 2020;367(6478).

48. Vantourout P, Laing A, Woodward MJ, Zlatareva I, Apolonia L, Jones AW, et al. Heteromeric interactions regulate butyrophilin (BTN) and BTN-like molecules governing γδ T cell biology. Proc Natl Acad Sci. 2018;115(5):1039–44.

49. Melandri D, Zlatareva I, Chaleil RAG, Dart RJ, Nussbaumer O, Polyakova O, et al. The γδ T cell receptor combines innate with adaptive immunity by utilizing spatially distinct regions for agonist-selection and antigen responsiveness. Nat Immunol. 2019;19(12):1352–65.

50. Jandke A, Melandri D, Monin L, Ushakov DS, Laing AG, Vantourout P, et al. Butyrophilin-like

proteins display combinatorial diversity in selecting and maintaining signature intraepithelial γδ T cell compartments. Nat Commun. 2020;11(1).

51. Willcox CR, Vantourout P, Salim M, Zlatareva I, Melandri D, Zanardo L, et al. Butyrophilin-like 3 Directly Binds a Human Vγ4+ T Cell Receptor Using a Modality Distinct from Clonally-Restricted Antigen. Immunity. 2019;51(5):813-825.e4.

52. Melandri D, Zlatareva I, Chaleil RAG, Dart RJ, Chancellor A, Nussbaumer O, et al. The γδTCR combines innate immunity with adaptive immunity by utilizing spatially distinct regions for agonist selection and antigen responsiveness. Nat Immunol. 2018;19(12):1352–65.

53. Hayday AC, Vantourout P. The Innate Biologies of Adaptive Antigen Receptors. Annu Rev Immunol. 2020;38:487–510.

54. Lefrançois L, Puddington L. The role of the thymus in intestinal intraepithelial T-cell development. Ann N Y Acad Sci. 1996;778:36–46.

55. Bandeira A, Mota-Santos T, Itohara S, Degermann S, Hausser C, Tonegawa S, et al. Localization of γ/δ T cells to the intestinal epithelium is independent of normal microbial colonization. J Exp Med. 1990;172(1):239–44.

56. Simpson SJ, Holländer GA, Mizoguchi E, Allen D, Bhan AK, Wang B, et al. Expression of pro-inflammatory cytokines by TCRαβ+ and TCRγδ+ T cells in an experimental model of colitis. Eur J Immunol. 1997;27(1):17–25.

57. Kühl AA, Pawlowski NN, Grollich K, Blessenohl M, Westermann J, Zeitz M, et al. Human peripheral γ;δ; T cells possess regulatory potential. Immunology. 2009;128(4):580–8.

58. Inagaki-Ohara K, Chinen T, Matsuzaki G, Sasaki A, Sakamoto Y, Hiromatsu K, et al. Mucosal T Cells Bearing TCRγδ Play a Protective Role in Intestinal Inflammation. J Immunol. 2004;173(2):1390–8.

59. Fujihashi K, McGhee JR, Yamamoto M, Hiroi T, Kiyono H. Role of γδ T cells in the regulation of mucosal IgA response and oral tolerance. Ann N Y Acad Sci. 1996;778:55–63.

60. Han HJ, Jang YS, Seo GY, Park SG, Kang SG, Yoon S Il, et al. Murine γδ T cells render B cells refractory to commitment of IgA isotype switching. Immune Netw. 2018;18(4):1–14.

61. Huang Y, Heiser RA, Detanico TO, Getahun A, Kirchenbaum GA, Casper TL, et al. γδ T cells affect IL-4 production and B-cell tolerance. Proc Natl Acad Sci U S A. 2015;112(1):E39–48.

62. Melo-Gonzalez F, Kammoun H, Evren E, Dutton EE, Papadopoulou M, Bradford BM, et al. Antigen-presenting ILC3 regulate T cell-dependent IgA responses to colonic mucosal bacteria. J Exp Med. 2019;216(4):728–42.

63. Sutton CE, Mielke LA, Mills KHG. IL-17-producing γδ T cells and innate lymphoid cells. Eur J Immunol. 2012;42(9):2221–31.

64. Bunker JJ, Erickson SA, Flynn TM, Henry C, Koval JC, Meisel M, et al. Natural polyreactive IgA antibodies coat the intestinal microbiota. Science. 2017;358(6361).

65. Bunker JJ, Flynn TM, Koval JC, Shaw DG, Meisel M, McDonald BD, et al. Innate and Adaptive Humoral Responses Coat Distinct Commensal Bacteria with Immunoglobulin A. Immunity. 2015;43(3):541–53.

66. Peterson DA, McNulty NP, Guruge JL, Gordon JI. IgA Response to Symbiotic Bacteria as a Mediator of Gut Homeostasis. Cell Host Microbe. 2007;2(5):328–39.

67. Pabst O, Slack E. IgA and the intestinal microbiota: the importance of being specific. Mucosal Immunol. 2020;13(1):12–21.

68. Bunker JJ, Bendelac A. IgA Responses to Microbiota. Immunity. 2018;49(2):211–24.

69. Laird-Fick HS, Saini S, Hillard JR. Gastric adenocarcinoma: The role of helicobacter pylori in pathogenesis and prevention efforts. Postgrad Med J. 2016;92(1090):471–7.

70. Salama NR, Hartung ML, Müller A. Life in the human stomach : persistence strategies of the bacterial pathogen Helicobacter pylori. 2013;11:385–99.

71. Calam J. Helicobacter pylori modulation of gastric acid. Yale J Biol Med. 1999;72(2–3):195–202.

72. Waldum HL, Kleveland PM, Sørdal ØF. Helicobacter pylori and gastric acid: An intimate and reciprocal relationship. Therap Adv Gastroenterol. 2016;9(6):836–44.

73. Jeffery PL, McGuckin MA, Linden SK. Endocrine impact of Helicobacter pylori: Focus on ghrelin and ghrelin o -acyltransferase. World J Gastroenterol. 2011;17(10):1249–60.

74. Efthymiou G, Dardiotis E, Liaskos C, Marou E, Tsimourtou V, Scheper T, et al. Anti-hsp60 antibody responses based on Helicobacter pylori in patients with multiple sclerosis: (ir)Relevance to disease pathogenesis. J Neuroimmunol. 2016;298:19–23.

75. Lucero Y, Oyarzún A, O'Ryan M, Quera R, Espinosa N, Valenzuela R, et al. Helicobacter pylori cagA+ Is associated with milder duodenal histological changes in chilean celiac patients. Front Cell Infect Microbiol. 2017;7:1–10.

76. Narang M, Puri AS, Sachdeva S, Singh J, Kumar A, Saran RK. Celiac disease and Helicobacter pylori infection in children: Is there any Association? J Gastroenterol Hepatol. 2017;32(6):1178–82.

77. Kandulski A, Malfertheiner P. Helicobacter pylori and gastroesophageal reflux disease. Curr Opin Gastroenterol. 2014;30(4):402–7.

78. Islami F, Kamangar F. Helicobacter pylori and esophageal cancer risk: A meta-analysis. Cancer Prev Res. 2008;1(5):329–38.

79. Kullberg MC, Jankovic D, Gorelick PL, Caspar P, Letterio JJ, Cheever AW, et al. Bacteria-triggered CD4+ T regulatory cells suppress Helicobacter hepaticus-induced colitis. J Exp Med. 2002;196(4):505–15.

80. Kullberg MC, Jankovic D, Feng CG, Hue S, Gorelick PL, McKenzie BS, et al. IL-23 plays a key role in Helicobacter hepaticus-induced T cell-dependent colitis. J Exp Med. 2006;203(11):2485–94.

81. Chai JN, Peng Y, Rengarajan S, Solomon BD, Ai TL, Shen Z, et al. Helicobacter species are potent drivers of colonic T cell responses in homeostasis and inflammation. Sci Immunol. 2017;2(13).

82. Danne C, Ryzhakov G, Martínez-López M, Ilott NE, Franchini F, Cuskin F, et al. A Large Polysaccharide Produced by Helicobacter hepaticus Induces an Anti-inflammatory Gene Signature in Macrophages. Cell Host Microbe. 2017;22(6):733-745.e5.

83. Neumann C, Blume J, Roy U, Teh PP, Vasanthakumar A, Beller A, et al. c-Maf-dependent T reg cell control of intestinal T H 17 cells and IgA establishes host–microbiota homeostasis. Nat Immunol. 2019;20:471–81.

84. Campbell C, Dikiy S, Bhattarai SK, Chinen T, Matheis F, Calafiore M, et al. Extrathymically Generated Regulatory T Cells Establish a Niche for Intestinal Border-Dwelling Bacteria and Affect Physiologic Metabolite Balance. Immunity. 2018;48(6):1245-1257.e9.

85. Lee AJ, Ashkar AA. The dual nature of type I and type II interferons. Front Immunol. 2018;9:1–10.

86. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. J Exp Med. 2012;209(8):1445–56.

87. Robertson SJ, Lemire P, Maughan H, Goethel A, Turpin W, Bedrani L, et al. Comparison of Co-housing and Littermate Methods for Microbiota Standardization in Mouse Models. Cell Rep. 2019;27(6):1910-1919.e2.

88. Laukens D, Brinkman BM, Raes J, De Vos M, Vandenabeele P. Heterogeneity of the gut microbiome in mice: Guidelines for optimizing experimental design. FEMS Microbiol Rev. 2015;40(1):117–32.

89. Hildebrand F, Nguyen AT, Brinkman B, Yunta RG, Cauwe B, Vandenabeele P, et al. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. Genome Biol. 2013;R4.

90. Taylor NS, Xu S, Nambiar P, Dewhirst FE, Fox JG. Enterohepatic Helicobacter species are prevalent in mice from commercial and academic institutions in Asia, Europe, and North America. J Clin Microbiol. 2007;45(7):2166–72.

Department of Zoology, University of Oxford

Chapter 8

General Discussion and Future Work

8.1 General Discussion

This thesis has presented a study of the interactions between γδ T cells and the microbiota in the gut, which were dissected using T Cell Receptor (TCR) gamma repertoire and 16S sequencing respectively. As the two are in close prolonged contact and two-way interactions have been observed in other gut lymphocytes this presented an appealing area of study[1]. Overall, this work offers an insight into these interactions in chicken and mouse and presents interesting findings by taking a population level approach to examine these systems. Furthermore, this thesis provides a robust justification for the analysis methods used to generate data from repertoire and 16S reads, which have been carefully considered and discussed in detail.

The work in this thesis first presents a more detailed look at the chicken TCR gamma locus, which appears to have evolved by tandem duplication. Perplexingly however, despite the large number of T Receptor (TRG) V gene segments, the chicken TCR gamma locus has very low sequence diversity within TRGV families, as has been previously discussed[2]. In a marked contrast to what has been previously observed in mouse[3], a single TRGV gene segment was dominant in all of the tissues examined, which include gut, thymus , bursa and spleen. Furthermore, chicken γδ T cells appear to have some limited influence over the gut microbiota and may be involved in limiting responses to potentially beneficial genera, such as *Bifidobacterium,* in the caeca*.* However, the microbiota appears to have a limited effect on developing gut γδ T cells, which appear to develop normally in its absence. This insensitivity to the microbiota is perhaps due to the importance of γδ T cells in early immunity, as these lymphocytes begin to rearrange during foetal development, when the gut is still sterile. Colonisation with different co-infecting strains of *Campylobacter* appears to affect the microbiota differently depending on the strain composition of the infecting consortia. Certain strain mixtures also appeared to be able to stimulate an invariant γδ T cell response, though this finding will require more rigorous investigation. Finally, the influence of γδ T cells over the microbiota in mouse, a γδ low animal appeared to be far greater, and an almost total loss of *Helicobacter* was

observed in these animals when γδ T cells were perturbed or entirely absent. This suggests a more important role for gut γδ T cells in regulating the microbiota in mouse compared to chicken and provides an intriguing contrast.

8.1.1 Methods for the Processing and Analysis of 16S and Immune Repertoire Sequence Data

16S rRNA and immune repertoire NGS studies are both new and rapidly evolving techniques that enable investigation of the microbiota and immune system in detail unattainable prior to the sequencing era. These techniques produce huge amounts of data, on which an almost endless supply of analyses which can be performed. There is already an overwhelming number of different packages and methods for processing of this type of data with more being produced all the time. Though these are both exciting fields of study, as is often the case with more modern techniques, it is extremely important to carefully consider the advantages and disadvantages of a particular analysis method and ensure regular review of the literature to monitor new developments and changing conventions. Where conventions are questioned in the literature, it presents the issue of whether to follow the pre-established practises and risk an inappropriate analysis or to follow the newly presented methods and compromise one's ability to compare results with the previous literature. Until a new convention has been properly established as the consensus method in the literature, the safest approach is to do both as with the β-diversity analysis in this thesis, though this can present issues with displaying an excessive number of figures.

One example of how a lack of careful consideration of analysis methods can lead to issues is in calculation of diversity indices from data processed using the DADA2 pipeline, and many other 16S processing packages. Calculation of these indices can routinely be observed in the literature despite the fact that DADA2 does not call singletons due to the possibility of them being introduced due to sequencing artefacts, such as base miscalls, chimeras and contaminants and therefore not representing real variants. This particularly affects diversity indices which are highly influenced by

rare species, such as species richness and can lead to massively inflated diversity scores. Furthermore, the literature surrounding the development of these techniques and conventions often involves complex mathematics and can therefore feel impenetrable to the average user, making it difficult to know if the correct technique is being used and whether the data complies with its assumptions. Further confounding this issue is the potential for experimental set up to have extremely large effects on this type of sequencing data, often greater than that of the actual factors being studied[1].

In many ways, repertoire sequencing emphasises these issues because, in contrast with microbiome data, very few conventions are in place and researchers are therefore far more on their own in terms of deciding on an analysis pipeline[2,3]. Unfortunately, the establishment of these conventions and standards will take time and until then researchers are somewhat left to their own devices and must be self-reliant and ensure their own analyses are supported by logical and carefully thought out reasoning. As these issues make it extremely easy to make mistakes and analyse data incorrectly, the importance of careful consideration of the analysis methods to be used cannot be understated and any user should bear in mind, with great power, comes great responsibility.

8.1.2 TCR Gamma Repertoire of Complex vs Germ Free Chickens showed only Minor Changes, however, this does not Rule Out Microbiome Interactions

When comparing the TCR gamma repertoire of Complex flora and Germ Free (GF) birds, some differences between the two groups were observed. However, these were relatively small in magnitude in comparison to differences previously observed in other enteric immune factors of GF chickens. GF birds have been shown to display reduced cytokine expression[7], immature goblet cells and mucin profiles[8,9], reduced IgA, reduced B cells and a lack of caecal tonsil germinal centres[10,11]. The lack of major differences between Complex and GF birds does not necessarily mean that the microbiota does not interact with γδ T cells. Far greater differences in TCR gamma repertoire were observed when comparing *Campylobacter* infected birds with uninfected controls (See Chapter 6).

Campylobacter infection was associated with significant dysbiosis of the microbiota and changes to the proportion of a number of microbes in this and other work[12–14]. Though we showed that the precise nature of changes to the microbiota induced by *Campylobacter* colonisation is at least partially dependant on co-infecting strain composition, it is clear that its presence in the chicken gut is sufficient to cause microbiota changes. While it has been predicted that a γδ T cell response to *Campylobacter* may exist[15,16], it is yet to be directly demonstrated.

It may therefore be the case that the observed changes to the TCR gamma repertoire are instead a response to these *Campylobacter* induced effects on the microbiota and that γδ T cells are responding to markers of dysbiosis which are not present in GF mice. One possible method for testing this would be an experiment in which GF birds are gavaged with either *Campylobacter* or the microbes enriched by *Campylobacter* infection, such as *Romboutsia* or *Bacteroides*. A positive response in one of these groups and the lack of a response in the other would enable a positive identification of the reason for the TCR gamma repertoire changes we observed. One other possibility is that due to the chicken γδ T cells being seemingly more generalist in terms of cytokine expression, compared with those of mouse and human[17], a difference may instead be observed in cytokine expression profile in response to infection, instead of an expansion in a specific TRGV subset. This appears to be the case for mouse macrophages, which show similar populations, but differing activities in the absence of a microbiota[18] and may be another possible line of study.

8.1.3 Mouse vs Chicken TCR Gamma Repertoire

Mouse TRGV gene usage is well known to vary by tissue and different epithelial sites tend to be dominated by certain TRGV gene segments. During the mouse TCR gamma repertoire sequencing experiment (See Chapter 7), the finding that the gut is dominated by TRGV7 was also evident in our data and is well supported by the literature[19–21]. Our repertoire sequencing also revealed that when moving through the intestine from Duodenum to Caecum, the dominance of TRGV7 declined and TRGV2 appeared to take over as the dominant rearrangement. Additionally, adult mouse spleen and

thymus were both dominated by TRGV1 or TRGV4[22–25]. In contrast, during analysis of tissues from the chicken gut, as well as thymus, spleen and bursa, we showed that in adult birds the same TRGV gene segment, TRGV3.3, was dominant in terms of proportion in every tissue. Follow up studies involving further chicken tissues such as skin and vagina would form an interesting area for additional study, and further comparison with mouse, to confirm whether or not they too are dominated by TRGV3.3. Due to the results presented in Chapters 4, 5 and 6, we would predict that they would be.

Human γδ T cell subsets, though less studied than those of mouse, tend to be classified by Vδ usage, and it has been shown that Vδ1s are most commonly found in tissue, whereas Vδ2 is more prevalent in blood, particularly in the well-studied Vγ9Vδ2+ T cells, which express an invariant TCR and tend to make up a large proportion of circulating lymphocytes[26–29]. On the other hand, circulating γδ T cells of cattle and chickens have been classified by their expression of CD8 co-receptors[17,30,31]. CD8 appears to be a marker of activation for avian γδ T cells and has been used previously because chicken TRGV genes have been said to have a 'restricted IGV gene repertoire'[2,17]. However, while this is true of within family sequence diversity (IMGT subgroup definition), chicken TRGV genes have between family diversity comparable to that of mouse. While high sequence identity presents a problem for repertoire sequencing, it is likely a greater problem in terms of antibody specificity and therefore methods such as flow cytometry may have to analyse the repertoire on the level of TRGV family instead of individual gene.

As demonstrated then, the classification of γδ T cell subsets is currently non-standard between animals and it is therefore difficult to judge how widespread the phenomenon of localisation of TRGV gene segments to different tissues is, outside of the mouse. Certainly, in chicken this does not appear to be the case from our preliminary studies, however, additional analysis of further tissue types will be needed to confirm this. Moreover, this phenomena of a single TRGV gene segment dominating every tissue in chicken may be the standard state for gamma delta high

animals, though with the current paucity of evidence concerning TCR gamma repertoire studies, this remains speculative. Additionally, studies into the cytokine expression of chicken γδ T cells classified by CD8 showed that there is a large amount of overlap between these subsets and the lack of a characteristic profile[17]. This is in contrast to what has been reported in γδ T cells of mice and humans, which appear to display distinct cytokine expression profiles between subsets, at least when sorted based on location[32]. Overall then, this suggests a more generalist role for chicken γδ T cells compared to those of mammals, which express specific homogenous TCRs at different tissues. Chicken γδ T cells appear not to possess this tissue tropism, but may be more heterogeneous in terms of their cytokine expression potential and therefore effector function.

8.1.4 Mouse vs Chicken γδ T Cell Influence over the Microbiome

Chapters 5 and 7 showed that γδ T cells play a role in regulating the composition of the microbiota in the caeca of both mouse and chicken. Interestingly though, these effects were far greater in the mouse. Considering the difference in the numbers of circulating γδ T cells in both of these animals[33], this is the opposite of what we would expect. Considering the evidence presented in this thesis for interactions between γδ T cells and the microbiota, when removing these lymphocytes, one would naively expect to observe larger changes in animals which put more emphasis on the γδ T cell lineage, the γδ high chickens. Though the studies underpinning these differences in γδ T cell numbers have focussed only on the circulating population, it has been shown that contrary to the previous consensus, certain γδ subsets migrate between circulation and tissue in both γδ high and low animals[34,35]. The differences between the circulating population of γδ T cells in mouse and chicken suggests a fundamental biological difference between these lymphocytes, however it should be noted that this does not necessarily correspond with an increase in numbers of IELs[34]. One seemingly universal feature of γδ T cells is their localisation to epithelial tissues, including the gut, where they are present as IELs. Despite this similarity, the IEL γδ populations of mouse and chicken have additional differences, such as the presence of CD8aβ+ γδ T cells, which are present in chicken but not mouse[36].

In mouse, the 16S β diversity plots revealed highly significant separation between TCRδ -/- knockout and Wild Type mice. Remarkably, mouse strain was shown to explain the majority of the variance in these plots. We observed differences between the TCRδ -/- and Wild Type mice at every phylogenetic level, culminating in the discovery that the TCRδ -/-, Butyrophilin -/- and Nude mice, had greatly reduced *Helicobacter* as well as a swathe of other more minor taxonomic differences. Though a follow-up experiment to replicate this finding in mice originating from heterozygous parents is planned, the proportion of *Helicobacter* in the gut presented a stark contrast between Wild Type and TCRδ -/- mice. The differences we observed in the chicken caeca on the other hand were far more subtle. The TCRγ -/- chicken displayed weak separation on β-diversity plots which did not reach significance by any of the indices used. However, significant differences were observed at the order and genera level with a reduction in *Bifidobacterium* in the TCRγ -/- chicken strain.

One non-biological explanation for this disparity may be due to the difference in sample processing. 16S rRNA analysis in mouse was performed on DNA extracted from tissue samples, and therefore is likely to be more representative of the mucosal microbiota. The chicken 16S sequencing on the other hand, was performed on caecal contents and may therefore be more representative of the lumen. In chicken the caecal lumen and mucosal microbiotas have been shown to be different in multiple studies[37–39]. Despite the clear disparity in the magnitude of these differences between chicken and mouse, in both animals the lack of γδ T cells did seem to be accompanied by the loss of "beneficial" microbes. In the context of homeostasis, *Helicobacter* has been shown to induce dampening of damaging inflammatory responses through induction of IL-10 in mouse[40], while *Bfidobacterium* is a commonly administered probiotic associated with improved chicken gut health and reduced *Campylobacter*[41,42]. Despite the potential for a beneficial role in the gut for both of these microbes, any microbe-host interaction should be treated with caution as they are rarely simple and often context dependant. Therefore, the predicted effect of any one microbe should account for the wider microbial community and the current immune environment.

Low affinity IgA has been shown to associate with commensals and increase gut survival in mouse[43–47]. This may also be the case for chicken, though avian commensal IgA interactions are ill defined compared to those of mammals. Low affinity IgA binding to commensals is thought to be beneficial for both the epithelia, through limitation of microbial penetration of the mucosa[48], and the flora itself, by protection from the enteric immune system[49,50] and increased adhesion[51]. γδ T cells have been shown to impact IgA expression in mouse[52–54], and a similar mechanism may exist in chicken, though again due to a lack of study, to our knowledge there is no evidence for this.

In mouse it has been suggested that other gut lymphocytes, such as the ILC3s are able to reduce specific, high affinity, T dependant IgA binding to certain microbes, such as *Helicobacter,* thus inducing their survival. While low affinity IgA is beneficial, this high specificity T dependant IgA has been shown to be associated with responses to pathogens and elimination from the gut[55]. We would hypothesise that a similar mechanism may be at play in γδ T cells, and may be one mechanism by which they are able to influence the microbiota. Potentially consistent with a similar mechanism involving IgA in avians, *Bifidobacterium* and other beneficial microbes have been shown to increase levels of IgA in chicken[56–59]. Though this is beneficial for the host, some regulatory mechanism would need to be in place to protect the commensals from T-dependant specific IgA responses. IgA-Seq is a technique by which the IgA bound fraction of the microbiota is selected by FACs and sequenced[44,60] and would provide a route by which to test this hypothesis, though it would need to be first adapted for use in chickens. Preliminary work on this has already been conducted (Fig. A7).

Additionally, the microbiota differences we observed when the TRGV7+ γδ T cells were selectively reduced using the Btnl1 knockout strain, appeared to constitute a more severe version of what we observed when all γδ T cells were removed in the TCRδ -/- strain. In the chicken repertoire studies, we observed some small changes to the TCR gamma repertoire over time, which seemed to be characterised by a shift from more invariant to more diverse TRGV gene rearrangements. Though we were unable to identify a specific gut TRGV subset, this does indicate some degree of succession

suggesting TRGV gene segment profile changes occur over time. It may be the case, as it seems to be with mouse, that individual gut TRGV subsets have different effects on the microbiota, though further study would be needed to confirm this. One potentially useful tool in investigating this may be the chicken BG genes. These are a highly polymorphic group of genes in chicken which display homology to mouse Btnls[38,61]. It may be the case that this family play some role in facilitating this succession process by mediating the localisation of certain TRGV subsets, though this may occur in a less tissue specific manner compared to mouse.

One flaw of the mouse TCRδ -/- experiments is a failure to account for the maternal effect in experimental design[4] although they were all bred in the same facility for a long period. Chickens however, are not colonised during laying and remain sterile until hatch. Therefore, the differences we observed in the chicken experiment provide some evidence that γδ T cells are in fact able to influence the microbiota and increases the likelihood that the same, albeit exaggerated, result we observed in mouse constitutes a real phenotype.

8.2 Future Work

8.2.1 Common to all Chapters

8.2.1.1 Sequencing from Sorted γδ T Cells

In order to gain more information about how expression of CD surface molecules relates to TCR gamma repertoire, and thus link the recent findings about avian γδ T cell cytokine expression to our data, γδ T cells were sorted into CD8-, CD8αα+, CD8αβ+ populations using FACs[17]. RNA extraction was attempted numerous times alongside tissue lysates as positive controls, but the obtained concentrations were unusably low or in most cases too low to be detectable at all. The cells were sorted directly into a common volume of RLT lysis buffer (Qiagen), but the FACS buffer volume ejected from the nozzle varied according to sample cell count. Therefore, additional RLT buffer was added to each sample in order to mitigate the sharp decrease in extracted RNA concentration as the RLT lysis buffer dilution drops below x0.5[62]. However despite multiple attempts, troubleshooting did

not result in successful RNA extraction and the samples were abandoned due to time constraints. Completion of this experiment would provide a link with cytokine expression studies as well as a deeper insight into chicken γδ T cell biology and should therefore be a priority moving forwards.

8.2.1.2 TCR Delta Repertoire Sequencing

We focussed on the gamma repertoire due to the fact that it is the TCR chain present on γδ T cells that is known to undergo localisation of different V genes to different tissues. Furthermore, as the gamma chain rearranges after the delta chain, the likelihood of getting 2 different TCR deltas with the same TCR gamma is less than the reverse situation and so TCR gamma better represents the entire γδ TCR repertoire. Focussing on one particular chain of the γδ TCR enables us to undergo deeper sequencing and analysis. However, our discovery of the lack of TRGV gene localisation to different tissues within the chicken is an unexpected result. Though human γδ T cells have been less studied, there is known to be a degree of localisation of different TRGV delta gene segments, TRDVδ1s in tissue, TRDVδ2s in blood. Therefore, it is possible there may be some localisation to different tissues based instead on TRDV in chicken. Furthermore, analysis of the delta repertoire would provide a more complete picture of γδ T cells in general. Additionally, newer techniques such as single cell sequencing would enable a more in depth look at some of the highlighted gamma clones and rigorous characterisation of TRGV and TRDV chain pairing[63].

8.2.2 Chapter 5

8.2.2.1 Earlier TCR Gamma Repertoire Time Points in Germ Free Chickens

We observed some small changes to the chicken TCR gamma repertoire over time, characterised by the more public TRGVs becoming less public over time, and becoming supplanted by more diverse TRGV rearrangements (See Chapter 5). The youngest birds in this analysis were 14 days old and it may therefore be interesting to investigate this phenomena further in younger, possibly pre hatch birds. Furthermore, we observed an almost totally public, invariant TRGV, TRGV2.17/2.8 which is of particular interest and likely plays a role during early life, before TdT

becomes highly expressed and is able to generate junctional diversity. Later timepoints, beyond the 55 days considered in Chapter 5 could also be sampled, though we would predict that the proportions observed in these 55 day old birds represent the stable adult TCR gamma repertoire and that no further changes would be observed.

8.2.2.2 Development of IgA-Seq in Chicken

We hypothesise that one possible mechanism used by γδ T cells to influence the microbiota and bring about the microbiota changes we observed may be through IgA. One way to investigate this would be through IgA-seq. IgA-seq has become a relatively well established technique in mouse, but is yet to be adapted for use in chicken. This adaptation was planned to form part of this thesis but was abandoned due to time constraints. However, work on this so far has identified a potential fluorescent antibody which specifically binds to chicken IgA and shows an absence of non-specific binding (Fig. A7). However, this antibody is yet to be tested in an actual FACS experiment. Completing development of this technique would enable an experiment in which the IgA labelled profile of TCRγ -/- chicken faecal samples could be compared to that of Wild Type birds. Identification of an increase in specific IgA labelling of some of the genera we observed were reduced in the TCRγ -/- strain, such as *Bifidobacterium,* would provide confirmation of the ability of γδ T cells to act as a checkpoint for specific IgA responses aimed at eliminating potentially harmful microbes from the gut and preventing them from targeting beneficial commensals.

8.2.2.3 γδ T cells may Influence the Microbiota only in the Context of Dysbiosis

Though we showed that the microbiota is only slightly affected in the absence of γδ T cells in chicken, this does not necessarily rule out a role for γδ T cells in facilitating host control over the microbiota. As we only analysed healthy birds, it may be the case that γδ T cells only play a role in control over the microbiota in the context of dysbiosis. This seems to be the case in NOD1 and NOD2 receptor knockout mice, which are associated with colitis and dysbiosis[64,65] leading to inflammatory bowel disease (IBD). However, the absence of these receptors is not sufficient to induce microbiota

changes alone[66]. In order to investigate this, some mechanism for causing dysbiosis could be used to alter the microbiota in TCRγ -/- birds such as antibiotics. If γδ T cells only play a role in returning the microbiota to homeostasis after perturbation, this method would identify the retardation of this processes in the absence of these lymphocytes.

8.2.3 Chapter 6

8.2.3.1 Further Investigation of the γδ T Cell Response to *Campylobacter*

We observed differences in the TCR gamma repertoire potentially indicative of a γδ T cell response against *Campylobacter.* However, as it has been shown chicken, TRGV3.3 appeared to be dominant in every tissue examined and despite being significant in terms of β-diversity, the repertoires of the *Campylobacter* infected strains showed a TRGV gene expression profile which was slightly altered compared to that of the uninfected controls. There were also some significant differences in individual clones, but no overwhelming increases which would be diagnostic of an adaptive type response involving expansions of a clear subset of γδ T cells. Further study of these samples would be necessary to confirm the differences we observed as a response to *Campylobacter* infection. Cytokine expression studies on γδ T cells isolated from these infected birds may provide a potential method to provide this confirmation. Additionally, as we observed large microbiota differences in these birds it is also possible that the gamma repertoire changes we observed were a result of these perturbations and not a direct result of *Campylobacter* colonisation. Study of *Campylobacter* colonisation of GF birds may provide the simplest way to resolve this, especially considering we have shown that the gamma repertoire is virtually unaffected in these animals (See Chapter 5).

8.2.4 Chapter 7

8.2.4.1 Eliminating the Maternal Effect with Mutants from F2 Heterozygous Crosses

One issue with the experimental design of the microbiota work in Chapter 4 is that it does not account for either maternal or cage effects on the microbiome. These phenomena are well

documented in the literature and have been shown to have a greater impact than some enteric immune mutations[4,67,68]. Furthermore, despite the large effect sizes present in these analyses, the small numbers of individuals were not sufficient to obtain significant results. In order to address these issues, a follow up experiment involving far greater numbers of Btnl1 +/+, Btnl1, +/- and Btnl1 -/- mice generated from heterozygote crosses split across a greater number of cages has been planned and grant funding has been applied for.

8.2.4.2 IgA-Seq in Mice

Furthermore, a follow up experiment in order to test the reasons for the massively decreased proportions of *Helicobacter* in the Butyrophilin -/-, TCRδ -/- and Nude mice would be interesting and may uncover the potential mechanisms behind the ability of mouse γδ T cells to influence the microbiota. As our findings are similar to those seen previously in work involving ILC3s, which went on to show that the reduction in *Helicobacter* in an ILC3 knockout mouse strain was as a result of increased specific IgA binding to these microbes[69]. As IgA-seq is now a well-established technique[44,60], performing this analysis on faecal samples from the follow up heterozygote mouse experiment may lead to similar results. Particularly since IgA binding has already been shown to be altered in the absence of γδ T cells[52,70].

8.2.4.3 Investigation of the impact of GF mice on the gamma repertoire

In chickens, we investigated the γδ T cell microbiome interaction from both sides, looking at the impacts on one in the absence of the other. However, in mice we only examined this interaction from the perspective of γδ T cell control over the microbiota. We did perform gamma repertoire sequencing experiments in GF mice, however, this was performed by a student and when the triplicate PCRs were analysed they did not resemble one another and the TRGV expression profiles showed unexpected and seemingly random patterns between mice and tissues. Upon closer inspection of the clone profiles, it was shown that in many of these samples, the overwhelming majority of the TCR gamma repertoire was made up of a small number of clones, in some cases just

a single clone. This is strongly indicative of a lack of template in the cDNAs these PCRs were performed on and subsequent jackpotting resulting in massive overrepresentation of a few clones. This lack of cDNA template is often down to an initial lack of RNA. These RNA extractions were found to be extremely variable in terms of success and RNA itself is notoriously difficult to work with due to its fragility and susceptibility to destruction by RNases, which are ubiquitous in the environment[71]. This work should therefore be repeated from the initial tissue samples.

8.3 Closing Remarks

In conclusion, this thesis has for the first time, revealed that the TCR gamma repertoire of chickens seems to be dominated by a single TRGV segment, TRGV3.3. In contrast to our expectations, and previous findings in mouse, this gene segment was found to be dominant in every tissue examined, including small and large intestine, bursa, thymus and spleen and at every time point examined (14-55 days). Furthermore, this dominance was not altered in the absence of a microbiota or in the presence of the enteric pathogen *Campylobacter*. This work then went on to characterise γδ T cell interactions with the enteric flora. The presented data reveals that in the chicken gut, though some interplay exists between the two, with minor differences in the TCR gamma repertoire of GF and the microbiota of TCRγ -/- birds, γδ T cells and enteric microbes develop and operate mostly independent of one another, which is surprising given their close proximity. However, a TCR gamma repertoire shift was detected in the presence of co-infecting strains of *Campylobacter,* which were also shown to induce significant dysbiosis, though interestingly both of these effects were dependant on the composition and origin of the co-infecting strains. This may be indicative of a γδ T cell response to either dysbiosis or the *Campylobacter* strains themselves, though additional work will be required to fully dissect this interaction. Finally, this thesis presents data to suggest that the influence of γδ T cells over the microbiota is far greater in mouse, and various TCR gamma repertoire perturbations, including the selective depletion of dominant gut TRGV7+ γδ T cells, showed a total loss of *Helicobacter* compared to Wild Type. This suggests that γδ T cells of

mouse have a far greater influence over the microbiota that those of chicken, in contrast to what we expected given the different emphasis on γδ T cells in these animals. Finally, this thesis highlights the value of TCR repertoire and 16S rRNA NGS based studies to survey both T cells and the microbiota.

8.4 References

1. Smith K, McCoy KD, Macpherson AJ. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin Immunol. 2007;19(2):59–69.
2. Su C, Jakobsen I, Gu X, Nei M. Diversity and evolution of T-cell receptor variable region genes in mammals and birds. Immunogenetics. 1999;50(5–6):301–8.
3. Di Marco Barros R, Roberts NA, Dart RJ, Vantourout P, Jandke A, Nussbaumer O, et al. Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. Cell. 2016;167(1):203-218.e17.
4. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. J Exp Med. 2012;209(8):1445–56.
5. Heather JM, Ismail M, Oakes T, Chain B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform. 2017;19(4):554–65.
6. Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlsen TH, Franke A. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol. 2017;17(1):1–16.
7. Shanahan F. The host-microbe interface within the gut. Bailliere's Best Pract Res Clin Gastroenterol. 2002;16(6):915–31.
8. Cheled-Shoval SL, Gamage NSW, Amit-Romach E, Forder R, Marshal J, Van Kessel A, et al. Differences in intestinal mucin dynamics between germ-free and conventionally reared chickens after mannan-oligosaccharide supplementation. Poult Sci. 2014;93(3):636–44.
9. Forder REA, Howarth GS, Tivey DR, Hughes RJ. Bacterial modulation of small intestinal goblet cells and mucin composition during early posthatch development of poultry. Poult Sci. 2007;86(11):2396–403.
10. Kaspers B, Lettermann S, Roell S. DEVELOPMENT OF THE GUT ASSOCIATED IMMUNE SYSTEM. In: 20th European Symposium on Poultry Nutrition. 2015. p. 12–5.
11. Broom LJ, Kogut MH. The role of the gut microbiome in shaping the immune system of chickens. Vet Immunol Immunopathol. 2018;204:44–51.
12. Kaakoush NO, Sodhi N, Chenu JW, Cox JM, Riordan SM, Mitchell HM. The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. Gut Pathog. 2014;6(1):1–10.
13. Thibodeau A, Fravalo P, Yergeau E, Arsenault J, Lahaye L, Letellier A. Chicken Caecal Microbiome modifications induced by campylobacter Jejuni colonization and by a non-antibiotic feed additive. PLoS One. 2015;10(7):1–14.
14. McKenna A, Ijaz UZ, Kelly C, Linton M, Sloan WT, Green BD, et al. Impact of industrial production system parameters on chicken microbiomes: Mechanisms to improve

performance and reduce Campylobacter. Microbiome. 2020;8(1):1–13.

15. Reid WDK, Close AJ, Humphrey S, Chaloner G, Lacharme-Lora L, Rothwell L, et al. Cytokine responses in birds challenged with the human food-borne pathogen Campylobacter jejuni implies a Th17 response. R Soc Open Sci. 2016;3(3).

16. Shaughnessy RG, Meade KG, McGivney BA, Allan B, O'Farrelly C. Global gene expression analysis of chicken caecal response to Campylobacter jejuni. Vet Immunol Immunopathol. 2011;142(1–2):64–71.

17. Pieper J, Methner U, Berndt A. Heterogeneity of avian γδ T cells. Vet Immunol Immunopathol. 2008;124(3–4):241–52.

18. Yi P, Li LJ. The germfree murine animal: An important animal model for research on the relationship between gut microbiota and the host. Vet Microbiol. 2012;157(1–2):1–7.

19. Guy-Grand D, Vassalli P, Eberl G, Pereira P, Burlen-Defranoux O, Lemaitre F, et al. Origin, trafficking, and intraepithelial fate of gut-tropic T cells. J Exp Med. 2013;210(9):1839–54.

20. Asarnow DM, Goodman T, LeFrancois L, Allison JP. Distinct antigen receptor repertoires of two classes of murine epithelium-associated T cells. Nature. 1989;341(6237):60–2.

21. Ltohara S, Farr AG, Lafaille JJ, Bonneville M, Takagaki Y, Haas W, et al. Homing of a ,a thymocyte subset with homogeneous T -cell receptors to mucosal epithelia. Lett To Nat. 1990;343:754–7.

22. Balasubramaniam V, Appasamy PM. Development of murine pre-T cells into γδ T-cell receptor bearing cells. Dev Comp Immunol. 1994;18(3):179–91.

23. Haas W. Gamma / Delta Cells. Annu Rev Immunol. 1993;11:637–85.

24. Ikuta K, Kina T, MacNeil I, Uchida N, Peault B, Chien Y hsiu, et al. A developmental switch in thymic lymphocyte maturation potential occurs at the level of hematopoietic stem cells. Cell. 1990;62(5):863–74.

25. Havran WL, Allison JP. Origin of Thy-1 + dendritic epidermal cells of adult mice from fetal thymic precursors. Lett To Nat. 1990;344:68–70.

26. Deroost K, Langhorne J. Gamma/Delta T Cells and Their Role in Protection Against Malaria. Front Immunol. 2018;9:2973.

27. Lawand M, Déchanet-Merville J, Dieu-Nosjean MC. Key features of gamma-delta T-cell subsets in human diseases and their immunotherapeutic implications. Front Immunol. 2017;8:1–9.

28. Khairallah C, Chu TH, Sheridan BS. Tissue Adaptations of Memory and Tissue-Resident Gamma Delta T Cells. Front Immunol. 2018;9:1–22.

29. Fonseca S, Pereira V, Lau C, Teixeira M dos A, Bini-Antunes M, Lima M. Human Peripheral Blood Gamma Delta T Cells: Report on a Series of Healthy Caucasian Portuguese Adults and Comprehensive Review of the Literature. Cells. 2020;9(3):729.

30. Berndt A, Pieper J, Methner U. Circulating γδ T cells in response to Salmonella enterica serovar enteritidis exposure in chickens. Infect Immun. 2006;74(7):3967–78.

31. Wilson E, Aydintug KM, Jutila MA. A Circulating Bovine γδ T Cell Subset, Which Is Found in Large Numbers in the Spleen, Accumulates Inefficiently in an Artificial Site of Inflammation: Correlation with Lack of Expression of E-Selectin Ligands and L-Selectin. J Immunol.

1999;162:4914–9.

32. Kabelitz D, Wesch D. Features and Functions of γδ T Lymphocytes: Focus on Chemokines and Their Receptors. Crit Rev Immunol. 2003;23(5–6):339–70.

33. Holderness J, Hedges JF, Ramstead A, Jutila MA. Comparative Biology of γδ T Cell Function in Humans, Mice, and Domestic Animals. Annu Rev Anim Biosci. 2013;1(1):99–124.

34. Shekhar S, Milling S, Yang X. Migration of γδ T cells in steady-state conditions. Vet Immunol Immunopathol. 2012;147(1–2):1–5.

35. McKenzie DR, Kara EE, Bastow CR, Tyllis TS, Fenix KA, Gregor CE, et al. IL-17-producing γδ T cells switch migratory patterns between resting and activated states. Nat Commun. 2017;8.

36. Tregaskes CA, Kong FK, Paramithiotis E, Chen CL, Ratcliffe MJ, Davison TF, et al. Identification and analysis of the expression of CD8 alpha beta and CD8 alpha alpha isoforms in chickens reveals a major TCR-gamma delta CD8 alpha beta subset of intestinal intraepithelial lymphocytes. 1995;(154):4485–94.

37. Awad WA, Mann E, Dzieciol M, Hess C, Schmitz-Esser S, Wagner M, et al. Age-related differences in the luminal and mucosa-associated gut microbiome of broiler chickens and shifts associated with Campylobacter jejuni infection. Front Cell Infect Microbiol. 2016;6:1–17.

38. Salomonsen J, Chattaway JA, Chan ACY, Parker A, Huguet S, Marston DA, et al. Sequence of a Complete Chicken BG Haplotype Shows Dynamic Expansion and Contraction of Two Gene Lineages with Particular Expression Patterns. PLoS Genet. 2014;10(6).

39. Olsen KN, Henriksen M, Bisgaard M, Nielsen OL, Christensen H. Investigation of chicken intestinal bacterial communities by 16S rRNA targeted fluorescence in situ hybridization. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2008;94(3):423–37.

40. Bodger K, Bromelow K, Wyatt JI, Heatley R V. Interleukin 10 in Helicobacter pylori associated gastritis: Immunohistochemical localisation and in vitro effects on cytokine secretion. J Clin Pathol. 2001;54(4):285–92.

41. Estrada A, Wilkie DC, Drew M. Administration of Bifidobacterium bifidum to chicken broilers reduces the number of carcass condemnations for cellulitis at the abattoir. J Appl Poult Res. 2001;10(4):329–34.

42. Baffoni L, Gaggìa F, Di Gioia D, Santini C, Mogna L, Biavati B. A Bifidobacterium-based synbiotic product to reduce the transmission of C. jejuni along the poultry food chain. Int J Food Microbiol. 2012;157(2):156–61.

43. Bunker JJ, Erickson SA, Flynn TM, Henry C, Koval JC, Meisel M, et al. Natural polyreactive IgA antibodies coat the intestinal microbiota. Science. 2017;358(6361).

44. Bunker JJ, Flynn TM, Koval JC, Shaw DG, Meisel M, McDonald BD, et al. Innate and Adaptive Humoral Responses Coat Distinct Commensal Bacteria with Immunoglobulin A. Immunity. 2015;43(3):541–53.

45. Peterson DA, McNulty NP, Guruge JL, Gordon JI. IgA Response to Symbiotic Bacteria as a Mediator of Gut Homeostasis. Cell Host Microbe. 2007;2(5):328–39.

46. Pabst O, Slack E. IgA and the intestinal microbiota: the importance of being specific. Mucosal Immunol. 2020;13(1):12–21.

47. Bunker JJ, Bendelac A. IgA Responses to Microbiota. Immunity. 2018;49(2):211–24.

48. Macpherson AJ, Uhr T. Induction of Protective IgA by Intestinal Dendritic Cells Carrying Commensal Bacteria. Science (80-). 2004;303(5664):1662–5.

49. Jiang HQ, Thurnheer MC, Zuercher AW, Boiko N V., Bos NA, Cebra JJ. Interactions of commensal gut microbes with subsets of B- and T-cells in the murine host. Vaccine. 2004;22(7):805–11.

50. Kroese FGM, De Waard R, Bos NA. B-1 cells and their reactivity with the murine intestinal microflora. Semin Immunol. 1996;8(1):11–8.

51. McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. Host Selection of Microbiota via Differential Adhesion. Cell Host Microbe. 2016;19(4):550–9.

52. Fujihashi K, McGhee JR, Yamamoto M, Hiroi T, Kiyono H. Role of γδ T cells in the regulation of mucosal IgA response and oral tolerance. Ann N Y Acad Sci. 1996;778:55–63.

53. Han HJ, Jang YS, Seo GY, Park SG, Kang SG, Yoon S Il, et al. Murine γδ T cells render B cells refractory to commitment of IgA isotype switching. Immune Netw. 2018;18(4):1–14.

54. Huang Y, Heiser RA, Detanico TO, Getahun A, Kirchenbaum GA, Casper TL, et al. γδ T cells affect IL-4 production and B-cell tolerance. Proc Natl Acad Sci U S A. 2015;112(1):E39–48.

55. Fagarasan S, Kawamoto S, Kanagawa O, Suzuki K. Adaptive immune regulation in the gut: T cell-dependent and T cell-independent IgA synthesis. Annu Rev Immunol. 2010;28:243–73.

56. Yurong Y, Ruiping S, Shimin Z, Yibao J. Effect of probiotics on intestinal mucosal immunity and ultrastructure of cecal tonsils of chickens. Arch Anim Nutr. 2005;59(4):237–46.

57. Haghighi HR, Gong J, Gyles CL, Hayes MA, Sanei B, Parvizi P, et al. Modulation of antibody-mediated immune response by probiotics in chickens. Clin Diagn Lab Immunol. 2005;12(12):1387–92.

58. Haghighi HR, Gong J, Gyles CL, Hayes MA, Zhou H, Sanei B, et al. Probiotics stimulate production of natural antibodies in chickens. Clin Vaccine Immunol. 2006;13(9):975–80.

59. Bobíková K, Revajová V, Karaffová V, Levkutová M, Levkut M. IgA gene expression and quantification of cecal IgA+, IgM+, and CD4+ cells in chickens treated with EFAL41 and infected with Salmonella Enteritidis. Acta Histochem. 2015;117(7):629–34.

60. Palm NW, De Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014;158(5):1000–10.

61. Chen L, Fakiola M, Staines K, Butter C, Kaufman J. Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. Front Immunol. 2018;9:1–17.

62. Loontiens S, Depestel L, Vanhauwaert S, Dewyn G, Gistelinck C, Verboom K, et al. Purification of high-quality RNA from a small number of fluorescence activated cell sorted zebrafish cells for RNA sequencing purposes. BMC Genomics. 2019;20(1):1–16.

63. De Simone M, Rossetti G, Pagani M. Single cell T cell receptor sequencing: Techniques and future challenges. Front Immunol. 2018;9:1–7.

64. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshiftmutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001;411:603–6.

65. Natividad JMM, Petit V, Huang X, De Palma G, Jury J, Sanz Y, et al. Commensal and probiotic bacteria influence intestinal barrier function and susceptibility to colitis in Nod1-/-;Nod2-/-

Mice. Inflamm Bowel Dis. 2012;18(8):1434–46.

66. Robertson SJ, Yu Zhou J, Geddes K, Rubino SJ, Ho Cho J, Girardin SE, et al. Nod1 and Nod2 signaling does not alter the composition of intestinal bacterial communities at homeostasis. Gut Microbes. 2013;4(3).

67. Robertson SJ, Lemire P, Maughan H, Goethel A, Turpin W, Bedrani L, et al. Comparison of Co-housing and Littermate Methods for Microbiota Standardization in Mouse Models. Cell Rep. 2019;27(6):1910-1919.e2.

68. Laukens D, Brinkman BM, Raes J, De Vos M, Vandenabeele P. Heterogeneity of the gut microbiome in mice: Guidelines for optimizing experimental design. FEMS Microbiol Rev. 2015;40(1):117–32.

69. Melo-Gonzalez F, Kammoun H, Evren E, Dutton EE, Papadopoulou M, Bradford BM, et al. Antigen-presenting ILC3 regulate T cell-dependent IgA responses to colonic mucosal bacteria. J Exp Med. 2019;216(4):728–42.

70. Fujihashi K, McGhee JR, Kweon MN, Cooper MD, Tonegawa S, Takahashi I, et al. gamma/delta T cell-deficient mice have impaired mucosal immunoglobulin A responses. J Exp Med. 1996;183(4):1929–35.

71. Nielson H. RNA Methods and Protocols. Methods in molecular biology. 2011.

8.5 Appendices

Number	Sequence
1	TTCTCCA
2	AACACCT
3	CTCTCAC
4	TGCGCAA
5	GACACAG
6	ACCCCAT
7	TACACGA
8	ATCTCGT
9	CACACTC
10	GTCTCTG
11	AGCGCTT
12	TCCTACC
13	GTCGACA
14	AGCAACG
15	CACCACT
16	TTCGAAC
17	AACCAAG
18	CCCTAAT
19	TGCAAGC
20	GACCAGA
21	ACCTAGG
22	CTCGAGT
23	TACCATC
24	GCCTATA
25	ATCGATG
26	CGCAATT
27	ATCAGCA
28	TACTGCT
29	GTCAGAC
30	AGCCGAA
31	CACTGAG
32	TCCGGAT
33	AACTGGA

Number	Sequence
34	TTCAGGT
35	GACTGTC
36	ACCGGTA
37	CTCAGTG
38	TGCCGTT
39	ACCATCC
40	CTCCTCA
41	TGCTTCG
42	GACGTCT
43	ATCCTAC
44	CGCTTAA
45	TACGTAG
46	GCCATAT
47	AGCTTGC
48	CACGTGA
49	TCCATGG
50	GTCCTGT
51	AACGTTC
52	CCCATTA
53	TTCCTTG
54	GGATCCA
55	AAAGCCG
56	CCAACCT
57	TGATCAC
58	GAAGCAA
59	ACAACAG
60	CTACCAT
61	TAAGCGC
62	GCAACGA
63	ATACCGG
64	CGATCGT
65	TCAACTC
66	GTACCTA

Number	Sequence
67	AGATCTG
68	CAAGCTT
69	GTATACC
70	AGAGACA
71	TCACACT
72	GGAGAAC
73	CCACAAG
74	ACACAGA
75	CTATAGG
76	TGAGAGT
77	GCACATC
78	CGAGATG
79	ATAGGCC
80	CGAAGCA
81	TAACGCG
82	GCATGCT
83	AGAAGAC
84	CAACGAA
85	TCATGAG
86	GTAGGAT
87	AAACGGC
88	CCATGGA
89	GGAAGGT
90	ACATGTC
91	CTAGGTA
92	TGAAGTG
93	GAACGTT
94	CTAATCC
95	TGACTCA
96	GAATTCG
97	ACAGTCT
98	CGACTAC
99	GCAGTAG

Figure A1 – **Sequencing Primer Barcodes**
Primer barcodes are added onto the 5' end of the reverse primer sequence. These barcodes allow identification of each individual PCR from a pool of products. Barcode sequences are designed based on Hamming codes and thus allow error correction or a single base change which may occur due to PCR or sequencing error.

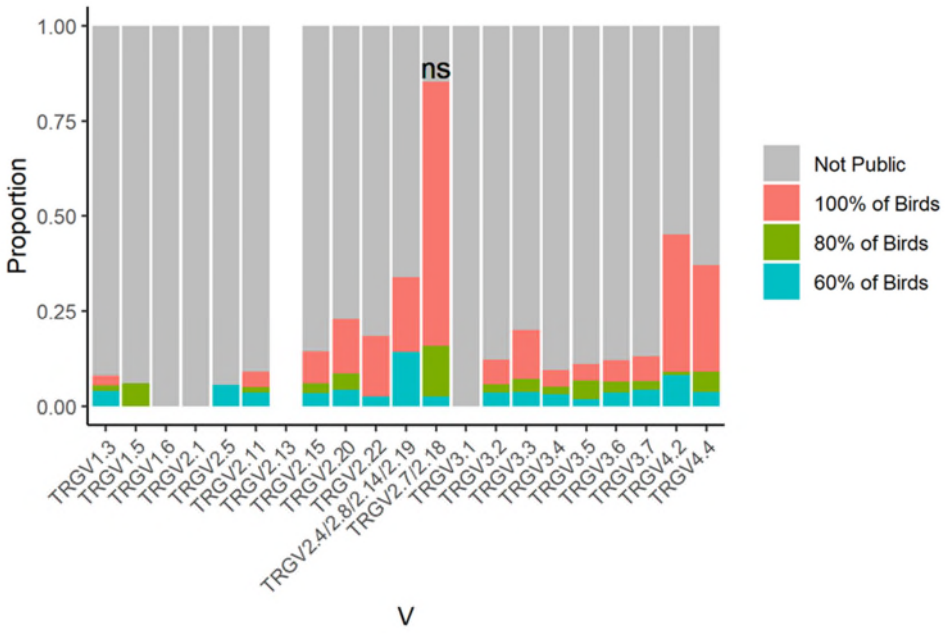


Figure A2 - **Levels of Publicity in each TRGV gene Segment in PA-12 White Leghorn Chickens**
Publicity is defined as a CDR3 nucleotide sequence found in 60%, 80% or 100% of birds (n=5) in any tissue at any frequency. The proportions of public clones in a given TRGV were calculated and displayed as a proportion of total sequences in the respective TRGV. Data is presented as stacked bars representing the proportion of the total repertoire for the particular TCRV. Due to a high level of identity, some TRGV segments have been grouped. P values are comparisons between the '100% of Birds' category of TRGV2.7/2.18 and the next most common TRGV gene segment, in this case TRGV4.2, and were calculated using an unpaired Wilcoxon signed rank test. *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

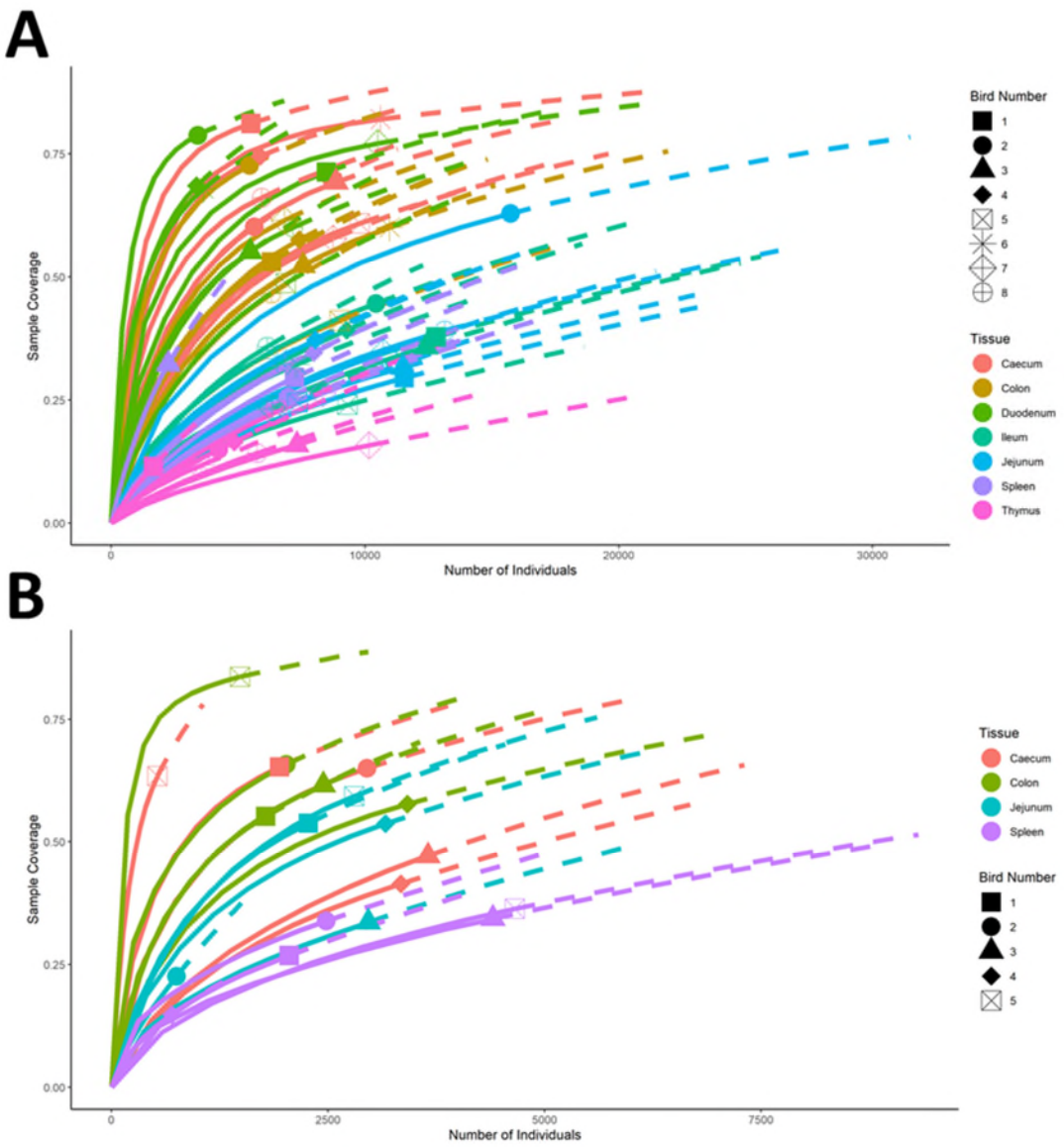


Figure A3 - Rarefaction Plots Generated using iNEXT

Plots show rarefaction curves of numbers of individuals, in this case represented by read number, against sample coverage for the **A)** ISA Brown and **B)** PA12 chicken lines. Plots were generated using iNEXT and as the number of individuals sampled increases, the coverage also increases but at a diminishing rate. Solid lines represent interpolated data, whereas dashed lines are extrapolations up to 2x the read number of the original sample. Coverage is calculated using Chao coverage. Line colour represents tissue, and the point shape, which is plotted at the intersection between the actual number of sequences and coverage of any given sample, represents bird number.

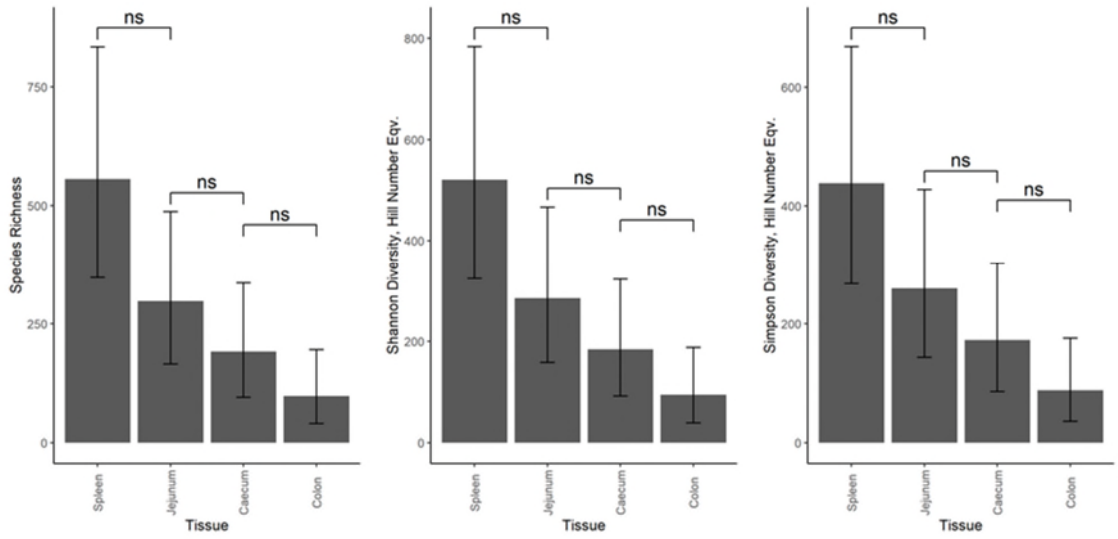


Figure A4 - Variation in the Patterns of TCR Gamma Diversity in different Tissues of PA12 White Leghorn Chickens.

Diversity indices were calculated from sample coverage rarefied abundance data for all clones in a tissue using the iNEXT platform. Diversity indices are displayed as Hill number variants and represent effective species counts - the number of evenly distributed species required to achieve the same diversity score. Indices are ordered by increasing Hill number and thus decreasing sensitivity to rare species. 95% confidence intervals were calculated using iNEXT. A linear model was constructed from the diversity index data using a cube root transform, which was confirmed to best approximate normality using Box Cox plots. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values. Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05. Thymus is most diverse, followed by Spleen. Unexpectedly, with the exception of Duodenum, the small intestine is most diverse.

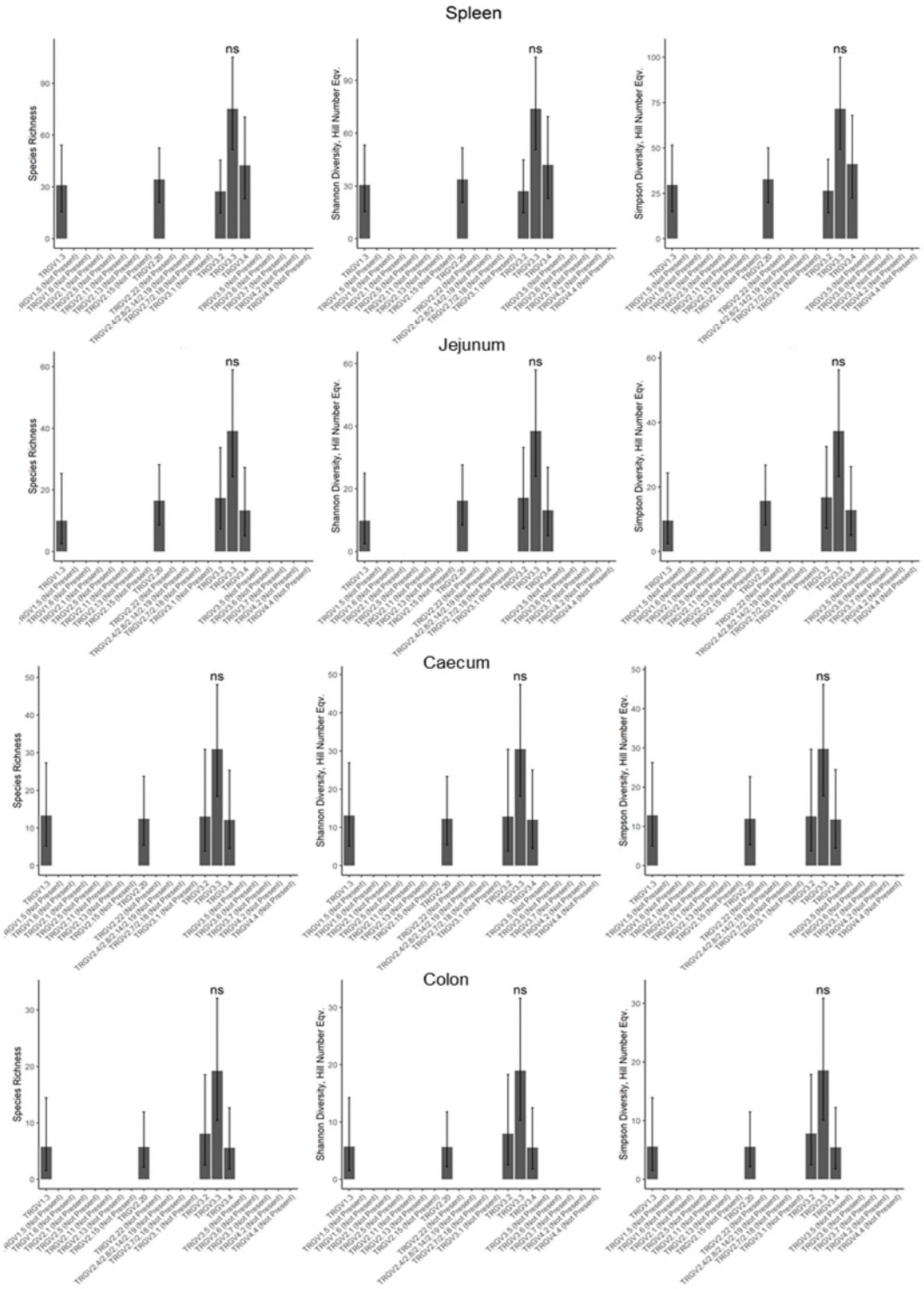


Figure A5 - Diversity of TRGV Gene Segment Rearrangements in different Tissues
Diversity indices were calculated as for the aggregate tissue diversity plots using all of the clones in a V segment. Due to their widely varying expression levels, gene segments below a read count of 100 were omitted. This overcomes the issue of rarefying down to extremely low sample coverage and therefore discarding an unacceptable amount of data. 95% confidence intervals were calculated using iNEXT. Indices were calculated with samples of each tissue from 5 birds. A linear model was constructed from the diversity index data using a cube root transform, which was confirmed to best approximate normality using Box Cox plots. Error bars are 95% confidence intervals and were extracted directly from the model along with the P values. P values are comparisons between TRGV3.3 and the next most common TRGV gene segment in that tissue Significance thresholds were as follows *** p value < 0.001, ** p value < 0.01, * p value < 0.05.

Lib 1 Plate 1 Rep 1			
Sample	Primer	Primer #	Plate Coord
B55 D55 conv Spl	ChCg_26.2	1	A1
B55 D55 conv Bur	ChCg_26.4	2	B1
B55 D55 conv Jej	ChCg_26.5	3	C1
B55 D55 conv Cae	ChCg_26.6	4	D1
B55 D55 conv Col	ChCg_26.7	5	E1
B56 D55 conv Spl	ChCg_26.8	6	F1
B56 D55 conv Bur	ChCg_26.10	7	G1
B56 D55 conv Jej	ChCg_26.12	8	H1
B56 D55 conv Cae	ChCg_26.13	9	A2
B56 D55 conv Col	ChCg_26.15	10	B2
B57 D55 conv Spl	ChCg_26.16	11	C2
B57 D55 conv Bur	ChCg_26.17	12	D2
B57 D55 conv Jej	ChCg_26.18	13	E2
B57 D55 conv Cae	ChCg_26.19	14	F2
B57 D55 conv Col	ChCg_26.20	15	G2
B58 D55 conv Spl	ChCg_26.21	16	H2
B58 D55 conv Bur	ChCg_26.23	17	A3
B58 D55 conv Jej	ChCg_26.24	18	B3
B58 D55 conv Cae	ChCg_26.25	19	C3
B58 D55 conv Col	ChCg_26.26	20	D3
B59 D55 conv Spl	ChCg_26.27	21	E3
B59 D55 conv Bur	ChCg_26.28	22	F3
B59 D55 conv Jej	ChCg_26.29	23	G3
B59 D55 conv Cae	ChCg_26.30	24	H3
B59 D55 conv Col	ChCg_26.31	25	A4
B67 D55 tetr Spl	ChCg_26.32	26	B4
B67 D55 tetr Bur	ChCg_26.34	27	C4
B67 D55 tetr Jej	ChCg_26.36	28	D4
B67 D55 tetr Cae	ChCg_26.37	29	E4
B67 D55 tetr Col	ChCg_26.38	30	F4
B68 D55 tetr Spl	ChCg_26.39	31	G4
B68 D55 tetr Bur	ChCg_26.40	32	H4
B68 D55 tetr Jej	ChCg_26.42	33	A5
B68 D55 tetr Cae	ChCg_26.44	34	B5
B68 D55 tetr Col	ChCg_26.45	35	C5
B69 D55 tetr Spl	ChCg_26.46	36	D5
B69 D55 tetr Bur	ChCg_26.47	37	E5
B69 D55 tetr Jej	ChCg_26.48	38	F5
B69 D55 tetr Cae	ChCg_26.49	39	G5
B69 D55 tetr Col	ChCg_26.50	40	H5
B1 D14 germ Spl	ChCg_26.51	41	A6
B1 D14 germ Bur	ChCg_26.52	42	B6
B1 D14 germ Jej	ChCg_26.53	43	C6
B1 D14 germ Cae	ChCg_26.54	44	D6
B1 D14 germ Col	ChCg_26.55	45	E6
B2 D14 germ Spl	ChCg_26.56	46	F6
B2 D14 germ Bur	ChCg_26.57	47	G6
B2 D14 germ Jej	ChCg_26.58	48	H6
B2 D14 germ Cae	ChCg_26.59	49	A7
B2 D14 germ Col	ChCg_26.60	50	B7
B3 D14 germ Spl	ChCg_26.61	51	C7
B3 D14 germ Bur	ChCg_26.62	52	D7
B3 D14 germ Jej	ChCg_26.63	53	E7
B3 D14 germ Cae	ChCg_26.66	54	F7
B3 D14 germ Col	ChCg_26.67	55	G7
B4 D14 germ Spl	ChCg_26.68	56	H7
B4 D14 germ Bur	ChCg_26.69	57	A8
B4 D14 germ Jej	ChCg_26.70	58	B8
B4 D14 germ Cae	ChCg_26.71	59	C8
B4 D14 germ Col	ChCg_26.72	60	D8
B5 D14 germ Spl	ChCg_26.73	61	E8
B5 D14 germ Bur	ChCg_26.74	62	F8
B5 D14 germ Jej	ChCg_26.75	63	G8
B5 D14 germ Cae	ChCg_26.76	64	H8
B5 D14 germ Col	ChCg_26.77	65	A9
B13 D14 Comp Spl	ChCg_26.78	66	B9
B13 D14 Comp Bur	ChCg_26.79	67	C9
B13 D14 Comp Jej	ChCg_26.80	68	D9
B13 D14 Comp Cae	ChCg_26.81	69	E9
B13 D14 Comp Col	ChCg_26.82	70	F9
B14 D14 Comp Spl	ChCg_26.84	71	G9
B14 D14 Comp Bur	ChCg_26.85	72	H9
B14 D14 Comp Jej	ChCg_26.87	73	A10
B14 D14 Comp Cae	ChCg_26.90	74	B10
B14 D14 Comp Col	ChCg_26.91	75	C10
B15 D14 Comp Spl	ChCg_26.92	76	D10
B15 D14 Comp Bur	ChCg_26.93	77	E10
B15 D14 Comp Jej	ChCg_26.95	78	F10
B15 D14 Comp Cae	ChCg_26.97	79	G10
B15 D14 Comp Col	ChCg_26.98	80	H10
B16 D14 Comp Spl	ChCg_26.99	81	A11
B16 D14 Comp Bur	ChCg_26.100	82	B11
B16 D14 Comp Jej	ChCg_26.101	83	C11
B16 D14 Comp Cae	ChCg_26.102	84	D11
B16 D14 Comp Col	ChCg_26.103	85	E11
B17 D14 Comp Spl	ChCg_26.104	86	F11
B17 D14 Comp Bur	ChCg_26.105	87	G11
B17 D14 Comp Jej	ChCg_26.106	88	H11
B17 D14 Comp Cae	ChCg_26.108	89	A12
B17 D14 Comp Col	ChCg_26.109	90	B12
B25 D28 germ Spl	ChCg_26.110	91	C12
B25 D28 germ Bur	ChCg_26.111	92	D12
B25 D28 germ Jej	ChCg_26.112	93	E12
B25 D28 germ Cae	ChCg_26.113	94	F12
B25 D28 germ Col	ChCg_26.114	95	G12
Water Ctrl 1	ChCg_26.115	96	H12

Lib 2 Plate 2 Rep 1			
Sample	Primer	Barcode	Plate Coord
B26 D28 germ Spl	ChCg_26.2	1	A1
B26 D28 germ Bur	ChCg_26.4	2	B1
B26 D28 germ Jej	ChCg_26.5	3	C1
B26 D28 germ Cae	ChCg_26.6	4	D1
B26 D28 germ Col	ChCg_26.7	5	E1
B27 D28 germ Spl	ChCg_26.8	6	F1
B27 D28 germ Bur	ChCg_26.10	7	G1
B27 D28 germ Jej	ChCg_26.12	8	H1
B27 D28 germ Cae	ChCg_26.13	9	A2
B27 D28 germ Col	ChCg_26.15	10	B2
B28 D28 germ Spl	ChCg_26.16	11	C2
B28 D28 germ Bur	ChCg_26.17	12	D2
B28 D28 germ Jej	ChCg_26.18	13	E2
B28 D28 germ Cae	ChCg_26.19	14	F2
B28 D28 germ Col	ChCg_26.20	15	G2
B29 D28 germ Spl	ChCg_26.21	16	H2
B29 D28 germ Bur	ChCg_26.23	17	A3
B29 D28 germ Jej	ChCg_26.24	18	B3
B29 D28 germ Cae	ChCg_26.25	19	C3
B29 D28 germ Col	ChCg_26.26	20	D3
B37 D28 Comp Spl	ChCg_26.27	21	E3
B37 D28 Comp Bur	ChCg_26.28	22	F3
B37 D28 Comp Jej	ChCg_26.29	23	G3
B37 D28 Comp Cae	ChCg_26.30	24	H3
B37 D28 Comp Col	ChCg_26.31	25	A4
B38 D28 Comp Spl	ChCg_26.32	26	B4
B38 D28 Comp Bur	ChCg_26.34	27	C4
B38 D28 Comp Jej	ChCg_26.36	28	D4
B38 D28 Comp Cae	ChCg_26.37	29	E4
B38 D28 Comp Col	ChCg_26.38	30	F4
B39 D28 Comp Spl	ChCg_26.39	31	G4
B39 D28 Comp Bur	ChCg_26.40	32	H4
B39 D28 Comp Jej	ChCg_26.42	33	A5
B39 D28 Comp Cae	ChCg_26.44	34	B5
B39 D28 Comp Col	ChCg_26.45	35	C5
B40 D28 Comp Spl	ChCg_26.46	36	D5
B40 D28 Comp Bur	ChCg_26.47	37	E5
B40 D28 Comp Jej	ChCg_26.48	38	F5
B40 D28 Comp Cae	ChCg_26.49	39	G5
B40 D28 Comp Col	ChCg_26.50	40	H5
B41 D28 Comp Spl	ChCg_26.51	41	A6
B41 D28 Comp Bur	ChCg_26.52	42	B6
B41 D28 Comp Jej	ChCg_26.53	43	C6
B41 D28 Comp Cae	ChCg_26.54	44	D6
B41 D28 Comp Col	ChCg_26.55	45	E6
B49 D55 germ Spl	ChCg_26.56	46	F6
B49 D55 germ Bur	ChCg_26.57	47	G6
B49 D55 germ Jej	ChCg_26.58	48	H6
B49 D55 germ Cae	ChCg_26.59	49	A7
B49 D55 germ Col	ChCg_26.60	50	B7
B50 D55 germ Spl	ChCg_26.61	51	C7
B50 D55 germ Bur	ChCg_26.62	52	D7
B50 D55 germ Jej	ChCg_26.63	53	E7
B50 D55 germ Cae	ChCg_26.66	54	F7
B50 D55 germ Col	ChCg_26.67	55	G7
B51 D55 germ Spl	ChCg_26.68	56	H7
B51 D55 germ Bur	ChCg_26.69	57	A8
B51 D55 germ Jej	ChCg_26.70	58	B8
B51 D55 germ Cae	ChCg_26.71	59	C8
B51 D55 germ Col	ChCg_26.72	60	D8
B52 D55 germ Spl	ChCg_26.73	61	E8
B52 D55 germ Bur	ChCg_26.74	62	F8
B52 D55 germ Jej	ChCg_26.75	63	G8
B52 D55 germ Cae	ChCg_26.76	64	H8
B52 D55 germ Col	ChCg_26.77	65	A9
B53 D55 germ Spl	ChCg_26.78	66	B9
B53 D55 germ Bur	ChCg_26.79	67	C9
B53 D55 germ Jej	ChCg_26.80	68	D9
B53 D55 germ Cae	ChCg_26.81	69	E9
B53 D55 germ Col	ChCg_26.82	70	F9
B61 D55 Comp Spl	ChCg_26.84	71	G9
B61 D55 Comp Bur	ChCg_26.85	72	H9
B61 D55 Comp Jej	ChCg_26.87	73	A10
B61 D55 Comp Cae	ChCg_26.90	74	B10
B61 D55 Comp Col	ChCg_26.91	75	C10
B62 D55 Comp Spl	ChCg_26.92	76	D10
B62 D55 Comp Bur	ChCg_26.93	77	E10
B62 D55 Comp Jej	ChCg_26.95	78	F10
B62 D55 Comp Cae	ChCg_26.97	79	G10
B62 D55 Comp Col	ChCg_26.98	80	H10
B63 D55 Comp Spl	ChCg_26.99	81	A11
B63 D55 Comp Bur	ChCg_26.100	82	B11
B63 D55 Comp Jej	ChCg_26.101	83	C11
B63 D55 Comp Cae	ChCg_26.102	84	D11
B63 D55 Comp Col	ChCg_26.103	85	E11
B64 D55 Comp Spl	ChCg_26.104	86	F11
B64 D55 Comp Bur	ChCg_26.105	87	G11
B64 D55 Comp Jej	ChCg_26.106	88	H11
B64 D55 Comp Cae	ChCg_26.108	89	A12
B64 D55 Comp Col	ChCg_26.109	90	B12
B65 D55 Comp Spl	ChCg_26.110	91	C12
B65 D55 Comp Bur	ChCg_26.111	92	D12
B65 D55 Comp Jej	ChCg_26.112	93	E12
B65 D55 Comp Cae	ChCg_26.113	94	F12
B65 D55 Comp Col	ChCg_26.114	95	G12
Water Ctrl 2	ChCg_26.115	96	H12

Lib 3 Plate 3 Rep 2			
Sample	Primer	Primer #	Plate Coord
B17 D14 Comp Cae	ChCg_26.2	1	A1
B17 D14 Comp Col	ChCg_26.4	2	B1
B25 D28 germ Spl	ChCg_26.5	3	C1
B25 D28 germ Bur	ChCg_26.6	4	D1
B25 D28 germ Jej	ChCg_26.7	5	E1
B25 D28 germ Cae	ChCg_26.8	6	F1
B25 D28 germ Col	ChCg_26.10	7	G1
Water Ctrl 1	ChCg_26.12	8	H1
B55 D55 conv Spl	ChCg_26.13	9	A2
B55 D55 conv Bur	ChCg_26.15	10	B2
B55 D55 conv Jej	ChCg_26.16	11	C2
B55 D55 conv Cae	ChCg_26.17	12	D2
B55 D55 conv Col	ChCg_26.18	13	E2
B56 D55 conv Spl	ChCg_26.19	14	F2
B56 D55 conv Bur	ChCg_26.20	15	G2
B56 D55 conv Jej	ChCg_26.21	16	H2
B56 D55 conv Cae	ChCg_26.23	17	A3
B56 D55 conv Col	ChCg_26.24	18	B3
B57 D55 conv Spl	ChCg_26.25	19	C3
B57 D55 conv Bur	ChCg_26.26	20	D3
B57 D55 conv Jej	ChCg_26.27	21	E3
B57 D55 conv Cae	ChCg_26.28	22	F3
B57 D55 conv Col	ChCg_26.29	23	G3
B58 D55 conv Spl	ChCg_26.30	24	H3
B58 D55 conv Bur	ChCg_26.31	25	A4
B58 D55 conv Jej	ChCg_26.32	26	B4
B58 D55 conv Cae	ChCg_26.34	27	C4
B58 D55 conv Col	ChCg_26.36	28	D4
B59 D55 conv Spl	ChCg_26.37	29	E4
B59 D55 conv Bur	ChCg_26.38	30	F4
B59 D55 conv Jej	ChCg_26.39	31	G4
B59 D55 conv Cae	ChCg_26.40	32	H4
B59 D55 conv Col	ChCg_26.42	33	A5
B67 D55 tetr Spl	ChCg_26.44	34	B5
B67 D55 tetr Bur	ChCg_26.45	35	C5
B67 D55 tetr Jej	ChCg_26.46	36	D5
B67 D55 tetr Cae	ChCg_26.47	37	E5
B67 D55 tetr Col	ChCg_26.48	38	F5
B68 D55 tetr Spl	ChCg_26.49	39	G5
B68 D55 tetr Bur	ChCg_26.50	40	H5
B68 D55 tetr Jej	ChCg_26.51	41	A6
B68 D55 tetr Cae	ChCg_26.52	42	B6
B68 D55 tetr Col	ChCg_26.53	43	C6
B69 D55 tetr Spl	ChCg_26.54	44	D6
B69 D55 tetr Bur	ChCg_26.55	45	E6
B69 D55 tetr Jej	ChCg_26.56	46	F6
B69 D55 tetr Cae	ChCg_26.57	47	G6
B69 D55 tetr Col	ChCg_26.58	48	H6
B1 D14 germ Spl	ChCg_26.59	49	A7
B1 D14 germ Bur	ChCg_26.60	50	B7
B1 D14 germ Jej	ChCg_26.61	51	C7
B1 D14 germ Cae	ChCg_26.62	52	D7
B1 D14 germ Col	ChCg_26.63	53	E7
B2 D14 germ Spl	ChCg_26.66	54	F7
B2 D14 germ Bur	ChCg_26.67	55	G7
B2 D14 germ Jej	ChCg_26.68	56	H7
B2 D14 germ Cae	ChCg_26.69	57	A8
B2 D14 germ Col	ChCg_26.70	58	B8
B3 D14 germ Spl	ChCg_26.71	59	C8
B3 D14 germ Bur	ChCg_26.72	60	D8
B3 D14 germ Jej	ChCg_26.73	61	E8
B3 D14 germ Cae	ChCg_26.74	62	F8
B3 D14 germ Col	ChCg_26.75	63	G8
B4 D14 germ Spl	ChCg_26.76	64	H8
B4 D14 germ Bur	ChCg_26.77	65	A9
B4 D14 germ Jej	ChCg_26.78	66	B9
B4 D14 germ Cae	ChCg_26.79	67	C9
B4 D14 germ Col	ChCg_26.80	68	D9
B5 D14 germ Spl	ChCg_26.81	69	E9
B5 D14 germ Bur	ChCg_26.82	70	F9
B5 D14 germ Jej	ChCg_26.84	71	G9
B5 D14 germ Cae	ChCg_26.85	72	H9
B5 D14 germ Col	ChCg_26.87	73	A10
B13 D14 Comp Spl	ChCg_26.90	74	B10
B13 D14 Comp Bur	ChCg_26.91	75	C10
B13 D14 Comp Jej	ChCg_26.92	76	D10
B13 D14 Comp Cae	ChCg_26.93	77	E10
B13 D14 Comp Col	ChCg_26.95	78	F10
B14 D14 Comp Spl	ChCg_26.97	79	G10
B14 D14 Comp Bur	ChCg_26.98	80	H10
B14 D14 Comp Jej	ChCg_26.99	81	A11
B14 D14 Comp Cae	ChCg_26.100	82	B11
B14 D14 Comp Col	ChCg_26.101	83	C11
B15 D14 Comp Spl	ChCg_26.102	84	D11
B15 D14 Comp Bur	ChCg_26.103	85	E11
B15 D14 Comp Jej	ChCg_26.104	86	F11
B15 D14 Comp Cae	ChCg_26.105	87	G11
B15 D14 Comp Col	ChCg_26.106	88	H11
B16 D14 Comp Spl	ChCg_26.108	89	A12
B16 D14 Comp Bur	ChCg_26.109	90	B12
B16 D14 Comp Jej	ChCg_26.110	91	C12
B16 D14 Comp Cae	ChCg_26.111	92	D12
B16 D14 Comp Col	ChCg_26.112	93	E12
B17 D14 Comp Spl	ChCg_26.113	94	F12
B17 D14 Comp Bur	ChCg_26.114	95	G12
B17 D14 Comp Jej	ChCg_26.115	96	H12

Lib 4 Plate 4 Rep 2			
Sample	Primer	Primer #	Plate Coord
B64 D55 Comp Cae	ChCg_26.2	1	A1
B64 D55 Comp Col	ChCg_26.4	2	B1
B65 D55 Comp Spl	ChCg_26.5	3	C1
B65 D55 Comp Bur	ChCg_26.6	4	D1
B65 D55 Comp Jej	ChCg_26.7	5	E1
B65 D55 Comp Cae	ChCg_26.8	6	F1
B65 D55 Comp Col	ChCg_26.10	7	G1
Water Ctrl 2	ChCg_26.12	8	H1
B26 D28 germ Spl	ChCg_26.13	9	A2
B26 D28 germ Bur	ChCg_26.15	10	B2
B26 D28 germ Jej	ChCg_26.16	11	C2
B26 D28 germ Cae	ChCg_26.17	12	D2
B26 D28 germ Col	ChCg_26.18	13	E2
B27 D28 germ Spl	ChCg_26.19	14	F2
B27 D28 germ Bur	ChCg_26.20	15	G2
B27 D28 germ Jej	ChCg_26.21	16	H2
B27 D28 germ Cae	ChCg_26.23	17	A3
B27 D28 germ Col	ChCg_26.24	18	B3
B28 D28 germ Spl	ChCg_26.25	19	C3
B28 D28 germ Bur	ChCg_26.26	20	D3
B28 D28 germ Jej	ChCg_26.27	21	E3
B28 D28 germ Cae	ChCg_26.28	22	F3
B28 D28 germ Col	ChCg_26.29	23	G3
B29 D28 germ Spl	ChCg_26.30	24	H3
B29 D28 germ Bur	ChCg_26.31	25	A4
B29 D28 germ Jej	ChCg_26.32	26	B4
B29 D28 germ Cae	ChCg_26.34	27	C4
B29 D28 germ Col	ChCg_26.36	28	D4
B37 D28 Comp Spl	ChCg_26.37	29	E4
B37 D28 Comp Bur	ChCg_26.38	30	F4
B37 D28 Comp Jej	ChCg_26.39	31	G4
B37 D28 Comp Cae	ChCg_26.40	32	H4
B37 D28 Comp Col	ChCg_26.42	33	A5
B38 D28 Comp Spl	ChCg_26.44	34	B5
B38 D28 Comp Bur	ChCg_26.45	35	C5
B38 D28 Comp Jej	ChCg_26.46	36	D5
B38 D28 Comp Cae	ChCg_26.47	37	E5
B38 D28 Comp Col	ChCg_26.48	38	F5
B39 D28 Comp Spl	ChCg_26.49	39	G5
B39 D28 Comp Bur	ChCg_26.50	40	H5
B39 D28 Comp Jej	ChCg_26.51	41	A6
B39 D28 Comp Cae	ChCg_26.52	42	B6
B39 D28 Comp Col	ChCg_26.53	43	C6
B40 D28 Comp Spl	ChCg_26.54	44	D6
B40 D28 Comp Bur	ChCg_26.55	45	E6
B40 D28 Comp Jej	ChCg_26.56	46	F6
B40 D28 Comp Cae	ChCg_26.57	47	G6
B40 D28 Comp Col	ChCg_26.58	48	H6
B41 D28 Comp Spl	ChCg_26.59	49	A7
B41 D28 Comp Bur	ChCg_26.60	50	B7
B41 D28 Comp Jej	ChCg_26.61	51	C7
B41 D28 Comp Cae	ChCg_26.62	52	D7
B41 D28 Comp Col	ChCg_26.63	53	E7
B49 D55 germ Spl	ChCg_26.66	54	F7
B49 D55 germ Bur	ChCg_26.67	55	G7
B49 D55 germ Jej	ChCg_26.68	56	H7
B49 D55 germ Cae	ChCg_26.69	57	A8
B49 D55 germ Col	ChCg_26.70	58	B8
B50 D55 germ Spl	ChCg_26.71	59	C8
B50 D55 germ Bur	ChCg_26.72	60	D8
B50 D55 germ Jej	ChCg_26.73	61	E8
B50 D55 germ Cae	ChCg_26.74	62	F8
B50 D55 germ Col	ChCg_26.75	63	G8
B51 D55 germ Spl	ChCg_26.76	64	H8
B51 D55 germ Bur	ChCg_26.77	65	A9
B51 D55 germ Jej	ChCg_26.78	66	B9
B51 D55 germ Cae	ChCg_26.79	67	C9
B51 D55 germ Col	ChCg_26.80	68	D9
B52 D55 germ Spl	ChCg_26.81	69	E9
B52 D55 germ Bur	ChCg_26.82	70	F9
B52 D55 germ Jej	ChCg_26.84	71	G9
B52 D55 germ Cae	ChCg_26.85	72	H9
B52 D55 germ Col	ChCg_26.87	73	A10
B53 D55 germ Spl	ChCg_26.90	74	B10
B53 D55 germ Bur	ChCg_26.91	75	C10
B53 D55 germ Jej	ChCg_26.92	76	D10
B53 D55 germ Cae	ChCg_26.93	77	E10
B53 D55 germ Col	ChCg_26.95	78	F10
B61 D55 Comp Spl	ChCg_26.97	79	G10
B61 D55 Comp Bur	ChCg_26.98	80	H10
B61 D55 Comp Jej	ChCg_26.99	81	A11
B61 D55 Comp Cae	ChCg_26.100	82	B11
B61 D55 Comp Col	ChCg_26.101	83	C11
B62 D55 Comp Spl	ChCg_26.102	84	D11
B62 D55 Comp Bur	ChCg_26.103	85	E11
B62 D55 Comp Jej	ChCg_26.104	86	F11
B62 D55 Comp Cae	ChCg_26.105	87	G11
B62 D55 Comp Col	ChCg_26.106	88	H11
B63 D55 Comp Spl	ChCg_26.108	89	A12
B63 D55 Comp Bur	ChCg_26.109	90	B12
B63 D55 Comp Jej	ChCg_26.110	91	C12
B63 D55 Comp Cae	ChCg_26.111	92	D12
B63 D55 Comp Col	ChCg_26.112	93	E12
B64 D55 Comp Spl	ChCg_26.113	94	F12
B64 D55 Comp Bur	ChCg_26.114	95	G12
B64 D55 Comp Jej	ChCg_26.115	96	H12

Lib 5 Plate 5 Rep 3			
Sample	Primer	Primer #	Plate Coord
B16 D14 Comp Spl	ChCg_26.2	1	A1
B16 D14 Comp Bur	ChCg_26.4	2	B1
B16 D14 Comp Jej	ChCg_26.5	3	C1
B16 D14 Comp Cae	ChCg_26.6	4	D1
B16 D14 Comp Col	ChCg_26.7	5	E1
B17 D14 Comp Spl	ChCg_26.8	6	F1
B17 D14 Comp Bur	ChCg_26.10	7	G1
B17 D14 Comp Jej	ChCg_26.12	8	H1
B17 D14 Comp Cae	ChCg_26.13	9	A2
B17 D14 Comp Col	ChCg_26.15	10	B2
B25 D28 germ Spl	ChCg_26.16	11	C2
B25 D28 germ Bur	ChCg_26.17	12	D2
B25 D28 germ Jej	ChCg_26.18	13	E2
B25 D28 germ Cae	ChCg_26.19	14	F2
B25 D28 germ Col	ChCg_26.20	15	G2
Water Ctrl 1	ChCg_26.21	16	H2
B55 D55 conv Spl	ChCg_26.23	17	A3
B55 D55 conv Bur	ChCg_26.24	18	B3
B55 D55 conv Jej	ChCg_26.25	19	C3
B55 D55 conv Cae	ChCg_26.26	20	D3
B55 D55 conv Col	ChCg_26.27	21	E3
B56 D55 conv Spl	ChCg_26.28	22	F3
B56 D55 conv Bur	ChCg_26.29	23	G3
B56 D55 conv Jej	ChCg_26.30	24	H3
B56 D55 conv Cae	ChCg_26.31	25	A4
B56 D55 conv Col	ChCg_26.32	26	B4
B57 D55 conv Spl	ChCg_26.34	27	C4
B57 D55 conv Bur	ChCg_26.36	28	D4
B57 D55 conv Jej	ChCg_26.37	29	E4
B57 D55 conv Cae	ChCg_26.38	30	F4
B57 D55 conv Col	ChCg_26.39	31	G4
B58 D55 conv Spl	ChCg_26.40	32	H4
B58 D55 conv Bur	ChCg_26.42	33	A5
B58 D55 conv Jej	ChCg_26.44	34	B5
B58 D55 conv Cae	ChCg_26.45	35	C5
B58 D55 conv Col	ChCg_26.46	36	D5
B59 D55 conv Spl	ChCg_26.47	37	E5
B59 D55 conv Bur	ChCg_26.48	38	F5
B59 D55 conv Jej	ChCg_26.49	39	G5
B59 D55 conv Cae	ChCg_26.50	40	H5
B59 D55 conv Col	ChCg_26.51	41	A6
B67 D55 tetr Spl	ChCg_26.52	42	B6
B67 D55 tetr Bur	ChCg_26.53	43	C6
B67 D55 tetr Jej	ChCg_26.54	44	D6
B67 D55 tetr Cae	ChCg_26.55	45	E6
B67 D55 tetr Col	ChCg_26.56	46	F6
B68 D55 tetr Spl	ChCg_26.57	47	G6
B68 D55 tetr Bur	ChCg_26.58	48	H6
B68 D55 tetr Jej	ChCg_26.59	49	A7
B68 D55 tetr Cae	ChCg_26.60	50	B7
B68 D55 tetr Col	ChCg_26.61	51	C7
B69 D55 tetr Spl	ChCg_26.62	52	D7
B69 D55 tetr Bur	ChCg_26.63	53	E7
B69 D55 tetr Jej	ChCg_26.66	54	F7
B69 D55 tetr Cae	ChCg_26.67	55	G7
B69 D55 tetr Col	ChCg_26.68	56	H7
B1 D14 germ Spl	ChCg_26.69	57	A8
B1 D14 germ Bur	ChCg_26.70	58	B8
B1 D14 germ Jej	ChCg_26.71	59	C8
B1 D14 germ Cae	ChCg_26.72	60	D8
B1 D14 germ Col	ChCg_26.73	61	E8
B2 D14 germ Spl	ChCg_26.74	62	F8
B2 D14 germ Bur	ChCg_26.75	63	G8
B2 D14 germ Jej	ChCg_26.76	64	H8
B2 D14 germ Cae	ChCg_26.77	65	A9
B2 D14 germ Col	ChCg_26.78	66	B9
B3 D14 germ Spl	ChCg_26.79	67	C9
B3 D14 germ Bur	ChCg_26.80	68	D9
B3 D14 germ Jej	ChCg_26.81	69	E9
B3 D14 germ Cae	ChCg_26.82	70	F9
B3 D14 germ Col	ChCg_26.84	71	G9
B4 D14 germ Spl	ChCg_26.85	72	H9
B4 D14 germ Bur	ChCg_26.87	73	A10
B4 D14 germ Jej	ChCg_26.90	74	B10
B4 D14 germ Cae	ChCg_26.91	75	C10
B4 D14 germ Col	ChCg_26.92	76	D10
B5 D14 germ Spl	ChCg_26.93	77	E10
B5 D14 germ Bur	ChCg_26.95	78	F10
B5 D14 germ Jej	ChCg_26.97	79	G10
B5 D14 germ Cae	ChCg_26.98	80	H10
B5 D14 germ Col	ChCg_26.99	81	A11
B13 D14 Comp Spl	ChCg_26.100	82	B11
B13 D14 Comp Bur	ChCg_26.101	83	C11
B13 D14 Comp Jej	ChCg_26.102	84	D11
B13 D14 Comp Cae	ChCg_26.103	85	E11
B13 D14 Comp Col	ChCg_26.104	86	F11
B14 D14 Comp Spl	ChCg_26.105	87	G11
B14 D14 Comp Bur	ChCg_26.106	88	H11
B14 D14 Comp Jej	ChCg_26.108	89	A12
B14 D14 Comp Cae	ChCg_26.109	90	B12
B14 D14 Comp Col	ChCg_26.110	91	C12
B15 D14 Comp Spl	ChCg_26.111	92	D12
B15 D14 Comp Bur	ChCg_26.112	93	E12
B15 D14 Comp Jej	ChCg_26.113	94	F12
B15 D14 Comp Cae	ChCg_26.114	95	G12
B15 D14 Comp Col	ChCg_26.115	96	H12

Lib 6 Plate 6 Rep 3			
Sample	Primer	Primer #	Plate Coord
B63 D55 Comp Spl	ChCg_26.2	1	A1
B63 D55 Comp Bur	ChCg_26.4	2	B1
B63 D55 Comp Jej	ChCg_26.5	3	C1
B63 D55 Comp Cae	ChCg_26.6	4	D1
B63 D55 Comp Col	ChCg_26.7	5	E1
B64 D55 Comp Spl	ChCg_26.8	6	F1
B64 D55 Comp Bur	ChCg_26.10	7	G1
B64 D55 Comp Jej	ChCg_26.12	8	H1
B64 D55 Comp Cae	ChCg_26.13	9	A2
B64 D55 Comp Col	ChCg_26.15	10	B2
B65 D55 Comp Spl	ChCg_26.16	11	C2
B65 D55 Comp Bur	ChCg_26.17	12	D2
B65 D55 Comp Jej	ChCg_26.18	13	E2
B65 D55 Comp Cae	ChCg_26.19	14	F2
B65 D55 Comp Col	ChCg_26.20	15	G2
Water Ctrl 2	ChCg_26.21	16	H2
B26 D28 germ Spl	ChCg_26.23	17	A3
B26 D28 germ Bur	ChCg_26.24	18	B3
B26 D28 germ Jej	ChCg_26.25	19	C3
B26 D28 germ Cae	ChCg_26.26	20	D3
B26 D28 germ Col	ChCg_26.27	21	E3
B27 D28 germ Spl	ChCg_26.28	22	F3
B27 D28 germ Bur	ChCg_26.29	23	G3
B27 D28 germ Jej	ChCg_26.30	24	H3
B27 D28 germ Cae	ChCg_26.31	25	A4
B27 D28 germ Col	ChCg_26.32	26	B4
B28 D28 germ Spl	ChCg_26.34	27	C4
B28 D28 germ Bur	ChCg_26.36	28	D4
B28 D28 germ Jej	ChCg_26.37	29	E4
B28 D28 germ Cae	ChCg_26.38	30	F4
B28 D28 germ Col	ChCg_26.39	31	G4
B29 D28 germ Spl	ChCg_26.40	32	H4
B29 D28 germ Bur	ChCg_26.42	33	A5
B29 D28 germ Jej	ChCg_26.44	34	B5
B29 D28 germ Cae	ChCg_26.45	35	C5
B29 D28 germ Col	ChCg_26.46	36	D5
B37 D28 Comp Spl	ChCg_26.47	37	E5
B37 D28 Comp Bur	ChCg_26.48	38	F5
B37 D28 Comp Jej	ChCg_26.49	39	G5
B37 D28 Comp Cae	ChCg_26.50	40	H5
B37 D28 Comp Col	ChCg_26.51	41	A6
B38 D28 Comp Spl	ChCg_26.52	42	B6
B38 D28 Comp Bur	ChCg_26.53	43	C6
B38 D28 Comp Jej	ChCg_26.54	44	D6
B38 D28 Comp Cae	ChCg_26.55	45	E6
B38 D28 Comp Col	ChCg_26.56	46	F6
B39 D28 Comp Spl	ChCg_26.57	47	G6
B39 D28 Comp Bur	ChCg_26.58	48	H6
B39 D28 Comp Jej	ChCg_26.59	49	A7
B39 D28 Comp Cae	ChCg_26.60	50	B7
B39 D28 Comp Col	ChCg_26.61	51	C7
B40 D28 Comp Spl	ChCg_26.62	52	D7
B40 D28 Comp Bur	ChCg_26.63	53	E7
B40 D28 Comp Jej	ChCg_26.66	54	F7
B40 D28 Comp Cae	ChCg_26.67	55	G7
B40 D28 Comp Col	ChCg_26.68	56	H7
B41 D28 Comp Spl	ChCg_26.69	57	A8
B41 D28 Comp Bur	ChCg_26.70	58	B8
B41 D28 Comp Jej	ChCg_26.71	59	C8
B41 D28 Comp Cae	ChCg_26.72	60	D8
B41 D28 Comp Col	ChCg_26.73	61	E8
B49 D55 germ Spl	ChCg_26.74	62	F8
B49 D55 germ Bur	ChCg_26.75	63	G8
B49 D55 germ Jej	ChCg_26.76	64	H8
B49 D55 germ Cae	ChCg_26.77	65	A9
B49 D55 germ Col	ChCg_26.78	66	B9
B50 D55 germ Spl	ChCg_26.79	67	C9
B50 D55 germ Bur	ChCg_26.80	68	D9
B50 D55 germ Jej	ChCg_26.81	69	E9
B50 D55 germ Cae	ChCg_26.82	70	F9
B50 D55 germ Col	ChCg_26.84	71	G9
B51 D55 germ Spl	ChCg_26.85	72	H9
B51 D55 germ Bur	ChCg_26.87	73	A10
B51 D55 germ Jej	ChCg_26.90	74	B10
B51 D55 germ Cae	ChCg_26.91	75	C10
B51 D55 germ Col	ChCg_26.92	76	D10
B52 D55 germ Spl	ChCg_26.93	77	E10
B52 D55 germ Bur	ChCg_26.95	78	F10
B52 D55 germ Jej	ChCg_26.97	79	G10
B52 D55 germ Cae	ChCg_26.98	80	H10
B52 D55 germ Col	ChCg_26.99	81	A11
B53 D55 germ Spl	ChCg_26.100	82	B11
B53 D55 germ Bur	ChCg_26.101	83	C11
B53 D55 germ Jej	ChCg_26.102	84	D11
B53 D55 germ Cae	ChCg_26.103	85	E11
B53 D55 germ Col	ChCg_26.104	86	F11
B61 D55 Comp Spl	ChCg_26.105	87	G11
B61 D55 Comp Bur	ChCg_26.106	88	H11
B61 D55 Comp Jej	ChCg_26.108	89	A12
B61 D55 Comp Cae	ChCg_26.109	90	B12
B61 D55 Comp Col	ChCg_26.110	91	C12
B62 D55 Comp Spl	ChCg_26.111	92	D12
B62 D55 Comp Bur	ChCg_26.112	93	E12
B62 D55 Comp Jej	ChCg_26.113	94	F12
B62 D55 Comp Cae	ChCg_26.114	95	G12
B62 D55 Comp Col	ChCg_26.115	96	H12

Figure A6 – Libraries used in GF TCR Gamma Repertoire Experiment (See Chapter 5)
Figure shows a breakdown of the samples in each sequencing library constructed during the GF TCR gamma repertoire experiment presented in Chapter 5. The primer and primer number used to amplify each sample is shown in the 2nd and 3rd column respectively.

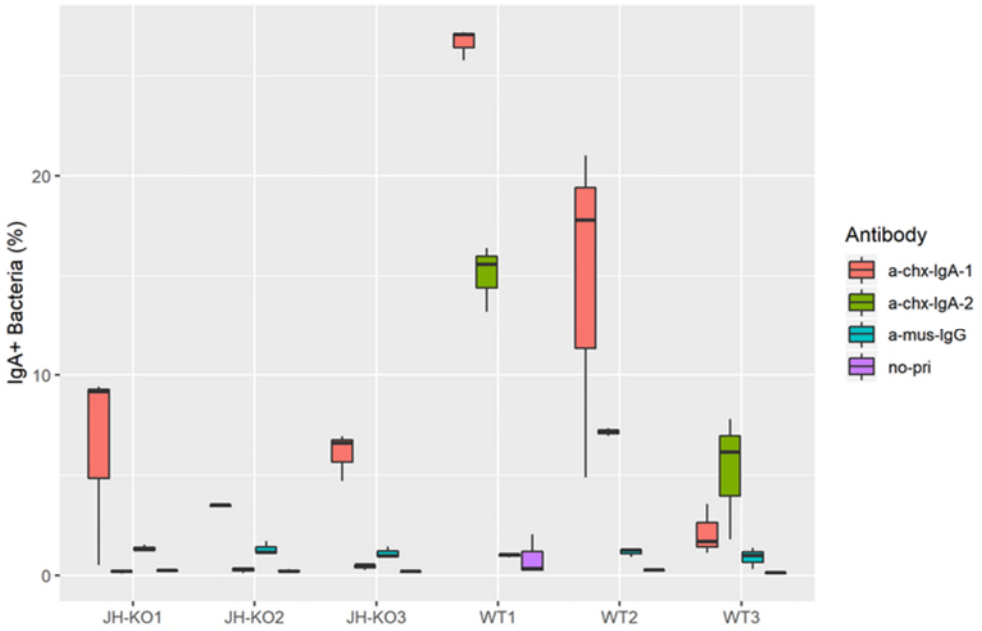


Figure A7 – Antibody Testing for Chicken IgA-Seq
Figure shows flow cytometric testing of two fluorescent anti-chicken IgA antibodies which could potentially be used in chicken IgA-seq. JH-KO1 is a chicken strain which is unable to produce IgA and was therefore used as a test of nonspecific binding to bacteria and WT is a wild type chicken strain. a-chx-IgA-1 appears to have issues with non-specific binding whereas a-chx-IgA-2 does not. A-mus-IgG is a control to highlight any bacterial binding to antibodies and no-pri is an unstained control. Samples were prepared from chicken faecal samples which were frozen at -80°C immediately after collection. Staining was performed and various washing stages before the samples were run an a BD Accuri™ C6 Plus Flow Cytometer. Y axis shows the percentage of events that were found to be fluorescent and therefore IgA labelled, each sample was assayed in triplicate. X axis is chicken strain and bird number within that strain.