The Spatial Dynamics of Microbiota-Immune Interactions



Kirstie McLoughlin

Brasenose College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

Acknowledgements

I would like to thank Adrian Smith and Kevin Foster for their supervision over the course of this DPhil. I am grateful to the BBSRC for their funding of my project, and to the Zoology Department for hosting me.

I would like to thank Mike Bonsall for his kind support and encouragement, despite not actually being my supervisor. Thank you also to Bernd Kaspers for hosting the animal experiments, and me, in Munich. My thanks in particular to Amy Boyd for your sage advice and dark humour, for reading my thesis, for pushing me to finish this project, and for all the burgers and milkshakes.

My thanks to the whole Smith and MERG groups for your continued support. I would like to especially thank Rob Paton, Chris Terry and John Fresen for their time and help with statistics, and Tom Brewer for reading and proofing sections. I also thank Elspeth Garman for her continued support and advice, and Emily Banks and Ellie Page for helping put me back together again.

I will forever be grateful to my family, for shaping and supporting me. I am extremely thankful for Robyn Clarke for your understanding, love, and wine. Thank you too to Margaret and Bob Norington for always pushing me forward. To Alice Chadwick, Eva Bryant and Isa Grübner thank you for holding my hands (both literally and metaphorically).

I must also thank Martin Peeks, who has supported me throughout my DPhil. The best part of my degree remains meeting you.

I could not have completed this thesis without any of you.

Table of Contents

Acknowledgements i

TABLE OF CONTENTS iii

ABSTRACT 1

GENERAL INTRODUCTION 3
[A] The microbiota and its host 3
[B] Interactions between the microbiota and immune system 4
[B.1] The innate immune system and the microbiota 4
[B.2] The adaptive immune system and the microbiota 5
[C] The avian model 7
[C.1] Avian gastrointestinal and lymphoid structures 8
[C.2] The avian microbiota 9
[C.3] Avian adaptive immune receptor formation 10
[D] Thesis structure 11

CHAPTER 1 13

Host selection of microbiota via differential adhesion 13
[A] Introduction 13
[A.1] Importance of the microbiota composition 13
[A.2] Evidence for host control of microbiota 14
[A.3] Chapter 1 aims 16
[B] Methods 16
[B.1] The model 16
[B.2] Implementation of adhesion as a host secreted compound 18
[B.3] Mucus flow induced displacement 19
[B.4] Feast-famine implementation 19
[B.5] Diversity calculations 20
[C] Results 20
[C.1] Adhesion as a mechanism of host selection 22
[C.2] Adhesion stabilises diversity across feast-famine cycles 24
[C.3] Host matrix (mucus) secretion and selection 26
[D] Discussion 27
[D.1] Candidate host-secreted adhesive molecules 30
[D.2] Theoretical predications 31
[E] Conclusions 33

CHAPTER 2 35

Fine-scale differences in the distribution of adaptive immune and microbiota components 35
[A] Introduction 35
[A.1] Spatial heterogeneity of the microbiota 35
[A.2] Factors that affect communities vary according to scale 36

[A.3] The chicken as a model species 37
[A.4] Chapter 2 aims 38
[B] Methods 39
[B.1] Sample collection 39
[B.2] Tissue extractions for primer design and sample processing 39
[B.3] cDNA synthesis 40
[B.4] DNA amplification 41
[B.5] Contaminant control 44
[B.6] Sequence creation and processing 45
[B.7] Quality checking sequences 46
[B.8] Statistics 46
[C] Method Development 48
[C.1] Accounting for sequencing error using microbiome analysis tools 49
[C.2] Comparing microbiome analysis pathways 50
[C.3] Identification of potential biases within the sampling process 52
[C.4] Identification of potential biases in sample amplification and barcoding 53
[C.5] Comparisons accounting for incomplete sampling 54
[D] Results 56
[D.1] Non-functional and contaminating sequences 56
[D.2] Population structures 57
[D.3] Population compositions 63
[D.4] Dissimilarity between segments 72
[D.5] Spatial structuring across populations 74
[E] Discussion 78
[E.1] Microbiota processing pipelines increase inter-study variation 79
[E.2] Large-scale sampling masks fine-scale heterogeneity 80
[E.3] Spatial structuring in caecal microbiota and immune populations 81
[E.4] Preliminary work and limitations 81
[F] Conclusions 82

CHAPTER 3 85

The role of rare species in microbiota community interactions 85
[A] Introduction 85
[A.1] Defining ecological stability 85
[A.2] Stability within complex communities 86
[A.3] Estimating species interactions 88
[A.4] Microbiota stability 89
[A.5] Chapter 3 aims 90
[B] Methods 90
[B.1] Microbial community sequence data 90
[B.2] Species interaction estimates by Lotka-Volterra derivation 91
[B.3] Species excluded from interaction estimates 93
[B.4] Table of terms 94
[B.5] Calculations and statistics 94
[C] Results 94
[C.1] Interactions within the microbiota 95
[C.2] Defining rarity 96
[C.3] Interspecific interactions 100
[C.4] Interactions in combined rarity groups 104
[C.5] Interspecific interactions between rarity groups 106

[D] Discussion 108
[D.1] Interactions within communities 110
[D.2] Spatial patterning with the microbiota 111
[D.3] Interaction environments vary between rarity groups 111
[D.4] Limitations of methodology 112
[D.5] A role for rarity in stabilising communities 113
[E] Conclusions 114

CHAPTER 4 115

Repeatability of fine-scale structure in the adaptive immune and microbiota components 115
[A] Introduction 115
[A.1] Microbiota variability 115
[A.2] Adaptive immune variability 115
[A.3] Temporal variation 116
[A.3] Chapter 4 aims 116
[B] Methods 117
[B.1] Sample collection 117
[B.2] Tissue RNA/DNA extraction 117
[B.3] cDNA synthesis 118
[B.4] DNA amplification 118
[B.5] Sequence creation and processing 118
[B.6] 16S contamination control 119
[C] Results 120
[C.1] Non-functional sequences 120
[C.2] Large-scale heterogeneity in wild-type population structures 121
[C.3] Fine-scale variation in population structure 123
[C.4] Large-scale population compositions 125
[C.5] Fine-scale variation in population composition 132
[C.6] Dissimilarity between segments 134
[D] Discussion 136
[D.1] Inter-individual variation 137
[D.2] Community stability 138
[E] Conclusions 138

CHAPTER 5 139

B-cells mediate immune and microbiota community dynamics 139
[A] Introduction 139
[A.1] Interactions between the adaptive immune system and microbiota 139
[A.2] Immunoglobulin A 140
[A.3] B-cell and T-cell interactions 141
[A.4] Evidence of microbiota-immune interactions in chicken 142
[A.5] Avian B-cell receptor knock-out 142
[A.6] Chapter 5 aims 143
[B] Methods 143
[B.1] Sample collection 143
[B.2] Tissue RNA/DNA extraction, cDNA synthesis and DNA amplification 144
[B.3] Sequence creation and processing 145
[B.4] 16S contamination control 147

[C] Results 147
[C.1] Functional sequences 148
[C.2] B-cell sufficiency and large-scale population structure 148
[C.3] Large-scale community composition of B-cell sufficient and knock-out birds 151
[C.4] Fine-scale population structure in B-cell competent birds 158
[C.5] Fine-scale population composition in B-cell sufficient and knock-out birds 160
[C.6] Spatial structure in B-cell sufficient and knock-out birds 165
[D] Discussion 167
[D.1] Mild B-cell knock-out phenotype 167
[D.2] Spatial variation exists in all treatment types 168
[D.3] Phenotypic variation in B-cell knock-outs at a fine-scale 168
[D.4] Selection within the microbiota 169
[D.5] Limitations 169
[E] Conclusions 170

GENERAL DISCUSSION 173
[A] Summary of findings 173
[B] Future work 175
[B.1] Replicability of fine-scale analysis 175
[B.2] The role of rarity in community stabilisation 175

SUPPLEMENTARY INFORMATION 177

APPENDICES 231
Appendix 1 – Sequence Processing 231
Appendix 2 - Sequences retained to abundance tables 231
Appendix 3 - Repeatability across technical replicates 231
Appendix 4 – Sample degradation during sampling 232
Appendix 5 – Comparing species interaction estimation methods 233
[A] Methods 233
[B] Consistency of species interaction estimates 235
[C] Interspecific interactions 235
[D] Interspecific interactions between rarity groups 237

BIBLIOGRAPHY 241

Abstract

In recent years, the modulating role of the microbiota on host health has become increasingly apparent. The composition of this diverse community is intrinsic to the host-associated benefits of a microbiota and is regulated by the host's immune system. Understanding the dynamics of microbiota-immune interactions is therefore necessary before modulating microbiota community composition for therapeutic purposes is viable. Ecological theory holds that the scale at which a community is considered changes the forces by which it is shaped. However, to date, microbiota-immune interactions have been assessed predominantly at large scales.

Within this thesis the selective pressures which control microbiota community composition are explored. An individual-based model of biofilm growth is initially used to interrogate the role of adhesion in maintaining microbiota diversity. A fine-scale method is then developed to investigate the model's predictions of spatial heterogeneity within the microbiota, allowing consideration of whether the scale at which the microbiota and adaptive immune populations are sampled changes the apparent dynamics of these communities.

At small scales species interactions are important in community dynamics. To investigate species interactions within the microbiota, a method of estimating species interactions across space is developed. By grouping species according to their relative abundance, the role of rare individuals in community stability is investigated.

Finally, an avian B-cell knockout is used to assess the role of immunoglobulin A (IgA) in modulating microbiota composition. By comparing the effect of IgA absence upon the microbiota at large and fine scales, the importance of sampling scale in elucidating immune and microbiota interactions is highlighted.

General Introduction

'Different patterns emerge at different scales of investigation of virtually any aspect of any ecological system.' J. A. Wiens[1].

When one considers an ecological system the first thing brought to mind is an island or a rainforest; a complex community of individuals interacting across species types. In recent years another ecological community has emerged into the spotlight as a result of its effects upon our health: the microbiota. Attempts to apply ecological theory to the microbiota community are beginning to emerge[2–4]. However one of the most important variables to affect community dynamics has yet to be considered: scale.

The scale at which a community is considered not only changes its apparent membership[1,5,6], but also its stability[1,7] and dynamics[8–10]. Manipulation of the microbiota as a mechanism of modulating health is an exciting new frontier, which requires an understanding of both the microbes involved in host health benefits and how these microbes interact with other components of the microbiota and the host's immune system. However our current understanding of these dynamics is restricted by the relatively large scale at which these interactions are considered. The aim of this thesis is to explore the dynamics of the microbiota and the adaptive immune system communities across spatial scales.

Each of the chapters within this thesis is independently structured. However there are some concepts which underlie every chapter: the influence of the microbiota upon its host; the relationship between the microbiota and the immune system; the structure of the avian gut; and the formation of adaptive immune receptors (specifically within chickens), each of which are introduced below.

[A] The microbiota and its host

The microbiota affects its host's evolutionary fitness in a variety of ways: providing nutrients (e.g. through the fermentation of non-digestible substances)[11,12], influencing

metabolic pathways (e.g. intestinal glucogenesis, cholesterol and fat metabolism, and insulin production)[12,13], stimulating normal tissue development (e.g. gastrointestinal villi and epithelial barrier formation), modulating the development of immune tissues (e.g. lymph node formation)[14,15] and the immune profile (e.g. inducing immune tolerance)[16,17], and even influencing the development of the nervous system (e.g. central glial development)[18] and brain (e.g. serotonin production)[18,19]. The impact of the microbiota across all these systems generates long-term effects upon a host as well, including pathogen protection (e.g. via colonisation resistance)[20,21] and chronic disease development (e.g. cardiovascular disease[22], cancer[23] and diabetes[24]).

Understanding the dynamics of the microbiota community, and how to modulate these systems, is therefore intrinsic to understanding the mechanisms of both human and livestock health.

[B] Interactions between the microbiota and immune system

The central mechanism by which a host has been suggested to modulate its microbiota is the immune system. The interactions between the immune system, both innate and adaptive, and the microbiota have been reviewed extensively elsewhere[25–28] so this relationship is only briefly summarised here.

[B.1] The innate immune system and the microbiota

Innate immune cell modulation of the microbiota occurs via barrier functions, the secretion of antimicrobial peptides, and the activation of the adaptive immune system (Figure 0.1)[26]. Both innate immune cells and intestinal epithelial cells sense microbial presence via Toll-like receptors (TLRs) and NOD-like receptors (NLRs). Microbiota presence induces tight-junction formation, in a TLR2-dependent manner[29], within the intestinal epithelium which is important for maintaining the spatial segregation of the microbiota from the host[30]. Spatial separation is also maintained through the secretion of mucus[31], antimicrobial peptides (e.g. REgIII-γ[32,33]) and cytokines (e.g. IL-10[34] and TGFβ[35]).

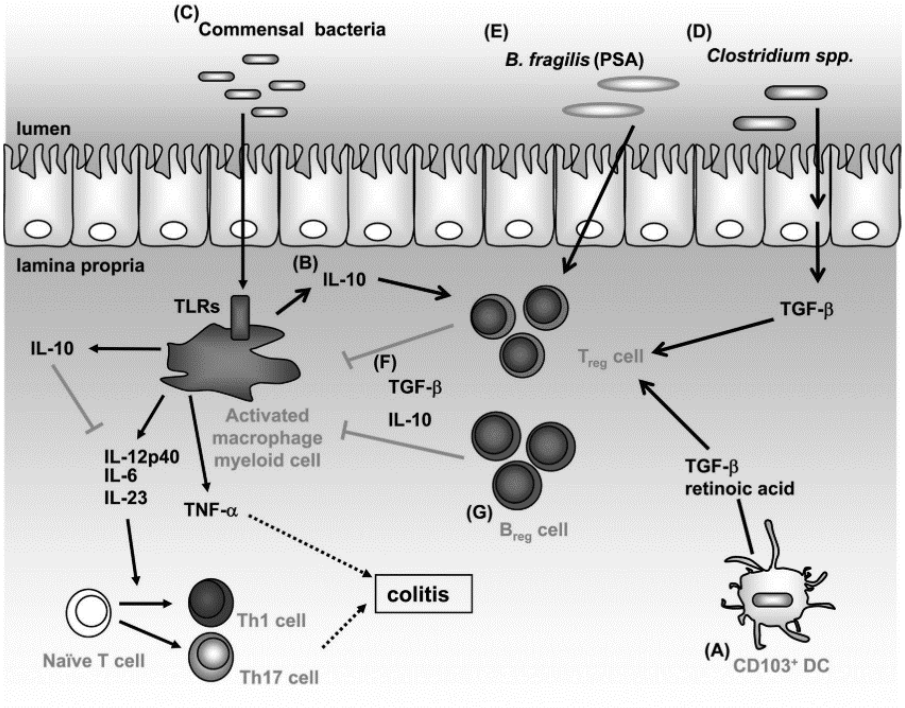


Figure 0.1 – Innate and adaptive immune mechanisms regulating the gastrointestinal microbiota. A) Dendritic cells (DCs) induce T-regulatory cells (Tregs) which induce an anti-inflammatory immune profile. B) Macrophages maintain Treg expression in an IL-10-dependent manner. C-E) Microbiota derived compounds activate innate immune receptors and induce different adaptive immune profiles. F) Tregs inhibit cytokine production in intestinal macrophages and myeloid cells which modulates the T_H1/T_H17 profile which in turn controls the intestinal inflammatory state. G) B-regulatory cells (Bregs) suppress intestinal inflammation by down-regulating the activation of intestinal myeloid cells. Figure adapted from Kayama et al. (2012)[26].

Dendritic cells modulate the adaptive immune response to the microbiota, either activating or suppressing T-cell responses to microbial colonisation[27]. For example, TLR5-activated DCs can polarise the T_H1/T_H17 profile to push the mucosal immune response towards or away from inflammation[36].

[B.2] The adaptive immune system and the microbiota

In jawed vertebrates, the immune system is comprised of the generalised innate immune system and the highly specific adaptive immune system. There are two arms of the adaptive immune system: B-cells and T-cells. Both parts of the adaptive immune system regulate, and are in turn modulated by, the gastrointestinal microbiota[37,38]. The interaction between the adaptive immune system and the microbiota is dependent upon both the simple

presence or absence of microbial stimulation[14,15,39–41], its diversity[42–45], and the specific composition of the microbial or immune components[17,21,46–50].

The normal development of immune tissues (e.g. Peyer's patches and mesenteric lymph nodes) is dependent upon the presence of a microbiota community[51,52], as is the recruitment of adaptive immune components to the lamina propria[53]. Germ-free mice show a reduction in T-follicular helper cells within the Peyer's patches[39] and an accumulation of natural killer T-cells[14] leading towards a pro-inflammatory state within the gastrointestinal tract. Exposure to microbes early in life changes the Th1/Th2 balance in T-cells, inducing an anti-inflammatory state[54], and is required for T-cell education[55,56]. The microbiota thus induces immune tolerance, reducing the probability of inflammatory disease (e.g. colitis).

Similarly, microbiota-derived antigens promote B-cell plasma differentiation[57] and class-switching, generating a diverse resting BCR repertoire. Low levels of microbial stimulation increases immunoglobulin E class-switching[58], increasing the probability of immunogenic diseases (e.g. asthama, allergies and diabetes). Germ-free animals show reduced numbers of immunoglobulin A (IgA)-producing plasma cells[40,41], which may reduce colonisation resistance within the gut.

The diversity of the microbiota has also been linked to specific immune phenotypes, for example IgA class-switching in B-cells depends on a diversified gut microbiota[43,44], and a reduction in microbiota diversity is correlated with a skew towards a pro-inflammatory T-cell response[42]. Shifts in microbiota composition are also associated with changes in the adaptive immune phenotype within the gastrointestinal tract[49,59–63]. For example, murine weaning stimulates microbiota compositional changes that increase the presence of peripherally stimulated inducible T regulatory cells[59] (Tregs).

The ratio of pro-inflammatory to anti-inflammatory signals within the mucosal immune system is dependent upon specific microbes (e.g. Lachnospiraceae members induce TGFβ production inhibiting Th2 subset proliferation[60] and Clostridia induce the expansion of Tregs[61,62]) pushing the mucosal immune system towards an anti-inflammatory response.

Conversely, segmented filamentous bacteria induce Th17 proliferation, pushing the gastrointestinal tract towards a pro-inflammatory state[49,63–65].

Microbe specific immune stimulation can also change the cytokine profile of T-cell subsets. For example, homeostatic Th17 cells express significantly fewer pro-inflammatory transcripts (e.g. TNF and IL23a), and have increased anti-inflammatory transcription (e.g. ICOS and IL22)[66] which is predominantly stimulated by segmented filamentous bacteria[49,63,67] or other epithelial cell adhesive bacteria[65].

[B.2.1] Microbiota diversity and the adaptive immune system

Microbe-host relationships can be seen across the tree of life; from Rhizobium-legume associations facilitating nutrient uptake in plants[68] to light production by *Vibrio fischeri* providing camouflage for the bobtailed squid[69]. Gut-associated microbial communities are also phylogenetically diverse, with gastrointestinal microbiota identified within hydra[70], insects[71], fish[72,73], reptiles[74], birds[75,76] and mammals[77,78]. However the extensive community diversity characteristic of the mammalian microbiota is restricted to jawed-vertebrates and has been suggested to be a product of the adaptive immune system[79].

Given the association between the beneficial properties of a microbiota and its diversity, understanding the dynamics of the interaction between the adaptive immune system and the microbiota is therefore extremely important.

[C] The avian model

Though evidence of microbiota-adaptive immune interactions is extremely abundant in mouse and human models, relatively few studies have attempted to unravel the interactions between the microbiota and mucosal immune system in birds. However, the global dependence upon the chicken as a food source, and the burgeoning restrictions upon antibiotic-usage in farm animals mean that extensive work (> 890 papers since 2002) looking at the effects of probiotics and prebiotics has been performed in chickens.

Much of the literature examined within this thesis arises from mammalian models, however where information exists suggesting distinct behaviours within avian microbiota dynamics (or where mammalian dynamics have been corroborated) this information is also provided.

[C.1] Avian gastrointestinal and lymphoid structures

The avian gastrointestinal tract is distinct from the mammalian gut in several features (both physical and immunological). The foregut of the chicken has additional compartments compared with the murine gut, specifically the presence of a crop, proventriculus and gizzard. The crop acts as an initial storage compartment, the proventriculus for initial digestion and the gizzard for mastication.

In the hindgut, the chicken is initially relatively similar to the murine intestine. The duodenum extends off the gizzard, transitioning to the jejunum and ileum. However, unlike mice, chicken have two caeca which extend from the central digestive tract between the ileum and the colon (Figure 0.2). The point from which the caeca emerge is called the caecal tonsil and is a lymphoid organ.

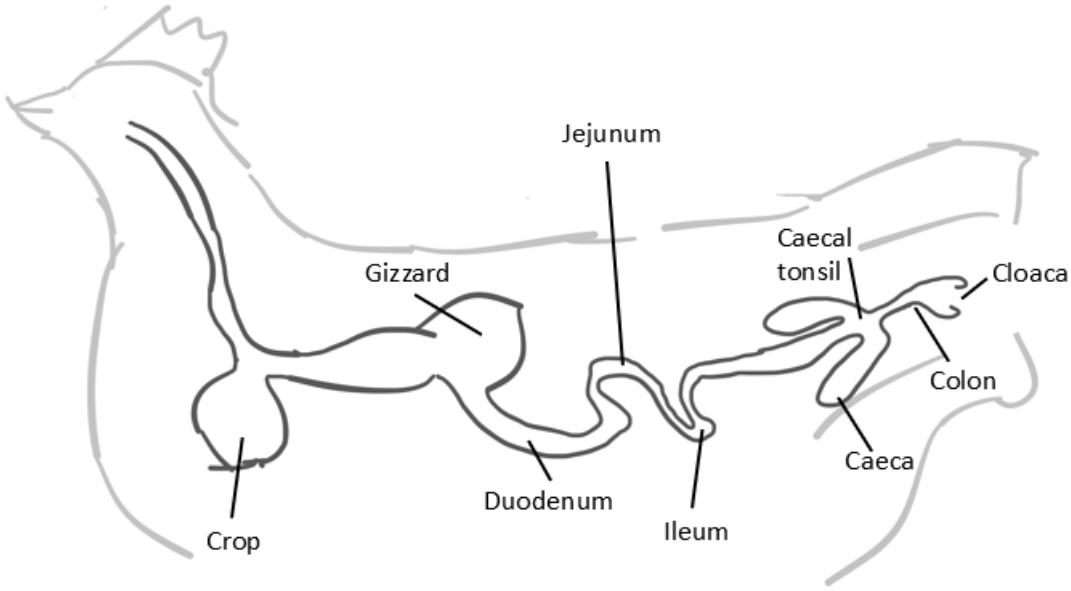


Figure 0.2 – Schematic of the avian gastrointestinal tract.

[C.1.1] Gut-associated lymphoid tissue

A major difference between mammals and birds is the absence of encapsulated lymph nodes in birds[80]. Although chicken lack lymph nodes, there is evidence of lymphoid structures throughout the gastrointestinal tract, including Peyer's patches, isolated lymphoid follicles and the caecal tonsil[81].

The chicken foregut has relatively few lymphoid structures compared with the hindgut, with the greatest concentration of isolated lymphoid follicles in the apical region of the caecum[81]. However, like mammals, the maturation of lymphoid structures within birds is dependent upon antigen stimulation (i.e. microbiota presence)[82].

[C.2] The avian microbiota

At a phylum level, avian microbiota composition remains relatively constant across the chicken gastrointestinal tract. The Firmicutes phylum dominates the avian gut microbiota, although the extent of its dominance appears to vary between birds and studies[83–85]. The family-level composition of the microbiota in mice and birds, however, varies across the length of the gastrointestinal tract[85–87]. In chickens this translates to a predominance of *Lactobacillales* in the crop, gizzard and small intestinal compartments of the gut, whilst *Clostridiales* dominate the caecum and colon. At a genus-level these differences are even more pronounced, with facultative anaerobes dominating the foregut and strict anaerobes restricted to the caecum and colon.

As in mammals, the avian microbiota develops temporally, becoming more diverse as the host ages[83,88]. Specifically, the chicken gut is initially dominated by Proteobacteria (one-week-old) and transitions through a predominance of Firmicutes (six-weeks-old) to a roughly equal ratio of Firmicutes and Bacteriodetes (> six-weeks-old)[88]. In birds this progression is more rapid than seen in humans, with the chicken microbiota stabilising between six and eight weeks of age[83,88–90].

[C.3] Avian adaptive immune receptor formation

Chickens, like all jawed vertebrates, have an adaptive immune system comprising two main cell subsets: B-cells and T-cells. The receptors on these cell types that are responsible for recognition of antigen and are formed from the somatic rearrangement of genomically encoded segments (constant (C), diversity (D), junction (J) and variable (V) genes). The process of T-cell receptor (TCR) formation in chickens is consistent with that of mammals[91,92] (though the number of genes contributing to each of segments varies[93,94]), however B-cell receptor (BCR) formation in birds has several distinct features.

[C.3.1] B-cell receptor formation

B-cell development in birds occurs within a gut-associated organ called the Bursa of Fabricius[95]. As in mammals, functional immunoglobulin (Ig) receptors are formed by the successful rearrangement of Ig gene segments in both the heavy (IgH) and light (IgL) chains. In chickens, the IgH locus consists of a single functional V_H, approximately 15 D_H, a single J_H and three C_H[96], whilst the IgL locus is made up of a single V_L, a single J_L and one C_L.

As is apparent from the restricted diversity of the BCR locus, diversity in the BCR population does not emerge as a result of the rearrangement of a large number of genomic segments, nor does random nucleotide addition to these rearranged segments expand the BCR repertoire (unlike in mammals). Instead, avian BCR diversity is generated through the insertion of sections of pseudo-V genes into the V or VD region of the rearranged BCR (a process known as gene conversion). Gene conversion from the 58 pseudo V_H or *c.* 25 pseudo V_L can occur multiple times upon a single chain of the BCR, however only one of the two BCR chains undergoes gene conversion to form each unique receptor.

The expression of different BCR C segments occurs in the peripheral germinal centres[97,98], in response to B-cell activation (through a process known as class switching). Class-switched BCR have undergone irreversible, non-homologous recombination between one of the C-segments and the switch region (the constant switch region keeps the BCR sequence in-

frame). In chickens there are only three antibodies produced: immunoglobulin M, immunoglobulin Y, and immunoglobulin A.

After activation, BCR diversity can also be generated via selection of receptors that have undergone rapid mutation induced within the BCR variable region (affinity maturation). However the presence of affinity maturation within chickens has proven challenging to detect given the random nature of gene conversion.

[C.3.2] T-cell receptor formation

TCR formation in birds is more consistent with that seen in mammals: T-cells develop from precursors within the thymus and TCR repertoire diversity is generated through somatic gene rearrangement with random nucleotide addition at the gene segment junctions[91,92]. Functional TCR are then selected within the thymus and remain fixed after antigen recognition. As in mammals, there are four TCR chains which combine to form two TCR types (αβ and γδ). The larger TCR chains (β and δ) are made up from V, D and J segments whilst the smaller (α and γ) are formed only of V and J segments.

[D] Thesis structure

This thesis comprises five research chapters; two theoretical and three empirical. Chapter 1 is an extended version of McLoughlin *et al.* (2016)[99], describing an individual-based model of biofilm dynamics used to predict the effect of host-provided selective adhesion upon the microbiota. The model presented in Chapter 1 is used to generate hypotheses which Chapters 2, 4 and 5 assess empirically.

Chapter 2 is a method-development chapter focused upon a single bird. The microbiota, T-cell receptor (TCR), and B-cell receptor (BCR) populations are assessed for differences between the large and fine scales. The core prediction of Chapter 1, that spatial structuring exists within the microbiota at a fine spatial scale, is considered. The sampling structure developed in Chapter 2 is then used to predict the types of interactions present within the microbiota, and the stratification of these interactions across different species based upon

their relative abundance within the sampled microbial population (Chapter 3). The work within Chapter 3 led to the development of a hypothesis that rare species are dynamically distinct from abundant species and may maintain stability within the microbiota.

Until Chapter 4 the work presented is either theoretical or based upon a single sampled individual. Chapter 4 therefore considers the repeatability of the dynamics observed in Chapter 2 across other wild-type birds. The genetic backgrounds and housing conditions of the birds compared are different from that of the preliminarily interrogated bird in Chapter 2, which provides some challenges when considering Chapter 2's repeatability. These birds do, however, form the basis of an experiment considering the effects of B-cells upon the avian microbiota and TCR populations (Chapter 5). The effects of B-cell presence are assessed at both the large-scale (comparative to previous literature) and the fine-scale.

Chapter 1

Host selection of microbiota via differential adhesion

This chapter is adapted from McLoughlin *et al.* (2016)[99], with additional discussion of the data in the context of this thesis. Supplementary information can be found within the manuscript supplementary (MSI). KM provided the initial hypothesis. KM, JS and KRF designed analyses. JS and KM performed preliminary analyses, JS ran those presented in the paper. All authors contributed to interpretation and writing.

[A] Introduction

Symbioses with microorganisms are central to the biology of plants and animals. For example, in plants, symbioses underlie the ability to photosynthesise[100] and take up nutrients[101]. In animals these associations can fulfil similar functions of nutrient acquisition[102,103], prevention of disease[104] and camouflage[105]. Ultimately symbioses are thought to have facilitated even the evolution of multi-cellularity[100,106].

The microbes within the gastrointestinal tracts of many organisms provide similar symbiotic functions. In humans, a healthy microbiota is composed of a complex community of some 100 trillion microbes, which predominantly colonise the lower gastrointestinal tract[16]. The species in these communities are important for normal tissue and immune development[39,51,62,63,107,108], provide metabolic functions[11,15,109–111] and help prevent pathogen colonisation[20,48,112–116].

[A.1] Importance of the microbiota composition

The beneficial properties of the microbiota are highly dependent upon its composition[17,21,46–49,117–120]. Specifically beneficial effects like colonisation resistance are often dependent upon both the complexity of a microbiota and its specific community members[45,48,121]. The exact composition of a microbiota can change its effect on immune

tolerance and development; for example a host's Th17:Treg profile is associated with the abundance of Bacteroidetes found within the small intestine[49,62,122].

However evolutionary and ecological dynamics continually threaten to disrupt a given community whenever non-beneficial species can establish themselves[17,123–126]. The maintenance of complexity in large communities[127,128] and of specific beneficial species in the presence of more rapidly growing contemporaries is a problem. This suggests that there is strong selection on hosts to control, and manage, the composition of their microbiota[129].

[A.2] Evidence for host control of microbiota

There is extensive evidence that hosts exert some control over their microbiota in humans and other systems[70,72,137,104,130–136]. For example, although the zebrafish and mouse gastrointestinal tracts contain similar bacterial taxa, reciprocal transference of the complete microbiota ultimately results in communities which resemble the normal gut microbiota of the host organism[72]. In vertebrates, the typical model of host control is a punitive one, whereby a host suppresses harmful species using the immune system.

[A.2.1] Punitive control of microbiota

The role of the immune system, both adaptive and innate, is well defined with respect to its control of pathogenic infections. The mechanisms behind immune control of pathogenic infection extend to non-pathogenic microbes – for example those of the gut microbiota. In the mammalian gut, intestinal epithelial cells act as both a physical barrier between microbes and the host's body, and a mediator of mucosal immune responses through the direct sensing of the microbiota[25,53,138–140]. The microbes of the gut community are kept from overwhelming the intestinal epithelial cell pattern recognition receptors by separating these cells from the majority of the community using a dense mucus layer[141].

Hosts commonly secrete a matrix that surrounds their symbionts at epithelial surfaces. Plants secrete "mucilage" polysaccharides from their roots into the rhizosphere, while animals make mucins, which are heavily glycosylated proteins. In the gut, these mucins form a coat over the epithelium that is broadly divided into two layers[141–143]. The inner, epithelium associated,

layer is formed of dense, interlocking mucins, which progressively unravel further away from the epithelium to form an upper layer that is much more loosely packed[142,143]. The inner layer is largely free from microbes while the outer layer can contain large numbers of microbes that appear to be both protected from sloughing, and fed, by the mucins around them[144–147].

The sterility of the dense mucus layer is maintained through the production of anti-microbial compounds[148] (including RegIIIγ[32,33] and defensins[149]) and the constant production of mucus[31]. Adaptive immune compounds, including immunoglobulin A (IgA)[40,150,151], have also been demonstrated to modulate the composition of the gut microbiota.

[A.2.2] Positive control of microbiota

While punitive host mechanisms have the potential to influence the microbiota, an alternative way for a host to influence their microbiota is via positive control. In positive control, a host acts in a way that promotes beneficial microbes rather than inhibits harmful ones. In fact, theoretical work suggests that positive control can be more effective than negative control in modulating the gut microbiota, because the former encourages growth of beneficial species near the epithelium[129,152]. This growth at the biofilm base thereby pushes harmful species away from the epithelium and into the lumen, where the microbes can be washed out of the gut[129]. A key candidate mechanism for positive control is the feeding of host-preferred microbial species via epithelial derived nutrients including fucose[153–155]. Consistent with this, a growing body of empirical work suggests that host-secreted nutrients can influence the species composition at the gut epithelium[129,151,155–157].

There is a second potentially general mechanism of positive control: adhesion. Adhesion to host epithelial cells and mucus has long been considered a key property underlying colonisation by both pathogenic and beneficial bacteria[158–160]. Recent work also suggests that microbes in surface associated communities can outcompete other genotypes simply by being more adhesive[161]. Adhesion allows cells to keep their position better and push other cells up towards the biofilm top and therefore out of the community. Moreover, many host-secreted factors have the potential to affect adhesion of microbial cells, both to each other and to host

factors like mucus. In mammals, these factors include the glycan residues attached to the mucin backbone of mucus and IgA proteins. Both glycans and IgA molecules come in a vast diversity of forms, which confer specificity with particular forms binding certain microbiota species more strongly than others[162–168].

[A.3] Chapter 1 aims

With a vast and rapidly growing body of data on the microbiome, there is a need for complementary theory that both identifies general principles and make testable predictions to evaluate these principles. With this in mind, the aim of this chapter is to use an individual-based model of host-microbe interactions across the epithelial surface to investigate whether the host might engage in positive selection of the microbiota by affecting microbial adhesion.

Specifically the chapter will address the role of adhesion in maintaining microbiota diversity and how coupling two potential immune factors (adhesion and mucus flow) could generate differential selection within the gut microbiota. The context of this work with respect to published empirical data will be discussed, and a number of testable predictions will be outlined.

[B] Methods

This work extends an extensively tested individual-based simulation framework that has been successfully used to investigate the behaviours of microbial communities growing on inert surfaces[169–171] and the interplay between host and symbionts in gut-associated microbial communities[129,161]. The assumptions of the model have been empirically validated and reviewed elsewhere[171–174].

[B.1] The model

The model is a hybrid between an individual-based simulation of a multi-strain bacterial community and a continuum model of solutes. To simulate microbiota communities, the simulation space is sectioned into host epithelium (bottom boundary), a biofilm domain where microbial cells live, a diffusion layer in which the concentration of solutes is exclusively governed

by diffusion (above the biofilm domain), and a bulk phase in the gut lumen where other transport processes dominate diffusion and concentrations of solutes are set to a constant value[129].

[B.1.1] Cellular growth

Cells are modelled as stiff spheres that can grow and divide depending on local nutrient concentrations (see B.1.2). Cells grow according to this consumption (Equation 1.1) and upon growth or division, neighbouring cells are pushed aside (see B.1.3) and the overall biofilm domain expands. Beyond a distance of 40 µm, cells are sloughed off and removed from the simulation in accordance with a previous model of gut epithelium attached microbial communities[129,161].

$$\mu = \frac{N\mu_{max}}{N + K_s}$$

Equation 1.1 – Monod equation describing cellular growth within the model; N is the local steady-state nutrient conditions, μ_{max} the maximum growth rate of the cell and K_s the solute Monod constant.

[B.1.2] Nutrient consumption

Consumption of nutrients informs the continuum model of solutes and is used to update the nutrient concentration fields. The model implements a multigrid solver for the reaction diffusion partial differential equations. Here, consumption of nutrients functions as local sinks for the respective solute. It is assumed that diffusion takes place on faster time-scales than cell-based events such as growth and division. Therefore, for each time-step, steady-state concentration gradients are calculated. See Schluter *et al*. (2012) for relevant equations[129].

Nutrients are provided from both the top and bottom of the biofilm to simulate luminal and host secretions. It is assumed that nutrient availability is high, with predominantly luminal nutrient provision, to mimic conditions in the gastrointestinal tract[129].

[B.1.3] Adhesion

The work of Schluter et al. (2015), which implemented differential adhesion between cells during the growth and pushing phase of the algorithm, was extended here to include vertical and horizontal gradients of adhesion (B.2). Schluter et al. (2015)'s model accurately

predicted behaviour of two differentially adhesive strains in biofilm experiments grown in flow chambers.

During each iteration, as a cell increases in diameter or divides, overlap is generated between neighbouring cells. Each overlapping cell is randomly selected and the overlap resolved, using a movement vector, until no overlap remains within the model. When overlap occurs between cells of different adhesiveness, adhesion is included within the model through scaling the movement vector by the relative adhesiveness of the focal cell (Equation 1.2). Therefore a more adhesive cell moves less far upon overlap with a less adhesive neighbour.

$$dV = 2\frac{\sigma_n}{\sigma_f + \sigma_n}$$

$$v_{new} = vdV$$

Equation 1.2 – Scaling movement vectors by relative adhesiveness of the cells; where dV is the scaling vector, v is the original movement vector, σ_f is the adhesiveness of the focal cell and σ_n the adhesiveness of the neighbouring cell.

[B.2] Implementation of adhesion as a host secreted compound

By extending the work of Schluter et al. (2015) to include horizontal and vertical gradients, we simulate different host secretions from the gut epithelium that convey the differential ability of cells to resist displacement. Adhesion strength is simulated as a relative ability to resist displacement, similar to spheres moving through a viscous liquid that experience viscous drag. Initially it is assumed that the secreted adhesions have the same effect everywhere within the model ('no gradient' simulations), simulating the situation that adhesiveness is a microbial property.

[B.2.1] Vertical gradients

Gradients of adhesiveness are then explicitly modelled, simulating the provision of adhesion as an epithelium-secreted compound. It is assumed that such secretions from the epithelium are at equilibrium between the source (epithelial cells) and a sink in a bulk phase deeper in the lumen ('vertical gradient' simulations, see MSI Table 1 for parameter values). The

local concentration then is a linearly decreasing function with its maximum at the source (bottom boundary) and zero in the lumen (bulk phase).

Vertical gradient simulations then simulate a scenario where the relative ability of cells to resist displacement gradually decreases further away from the source. The inclusion of a vertical gradient only mildly exacerbates the effects of generalised adhesiveness (MSI Figure 1), so subsequent simulations are run using the simpler model that lacks a vertical gradient. The impact of the vertical solutions is included only to demonstrate that this particular aspect of host-secreted adhesiveness does not change the effects of adhesiveness within a biofilm.

[B.2.2] Horizontal gradients

A horizontal gradient is similarly included in the model, to simulate the condition that a certain region of the epithelium secretes a specific adhesive compound (e.g. a singular epithelial cell translocating secretory immunoglobulin A molecules specific to a single surface epitope and therefore a singular microbial genotype). Left and right of this source, concentrations of this secreted product decrease (parameter values in MSI Table 1).

Cells that are not targeted by adhesive secretions are unaffected by the adhesion gradients.

[B.3] Mucus flow induced displacement

To model the effect of cells translocating with mucus secreted by the epithelium, a displacement function that has an effect at each iteration is implemented (parameter values in MSI Table 1). This discretises the translocation effect which occurs continuously in the real gut environment. However, as the models rely on discrete time steps, the parameter values must be viewed as approximations, which carry meaning in relative rather than absolute terms.

[B.4] Feast-famine implementation

The model was further extended to consider the effect of temporary availabilities of nutrients specific for some microbial species. During periods where the nutrients are available (feast) the system is flushed with nutrients and concentrations are saturating throughout the

simulation space. These periods are of various lengths, as indicated, within a 48h period (i.e. 4h means a four hour long period of feast begins every 48h). Outside these periods (famine) the concentrations of these nutrients are set to zero throughout the simulation space.

[B.5] Diversity calculations

Diversity within the model is calculated using the Shannon-Wiener index (Equation 1.3), which measures the uncertainty associated with predicting the identity of a species when an individual is drawn randomly from the population[9,175]. Therefore, the diversity index is maximized when all genotype frequencies are equal. When one or two species dominate the population Shannon-Wiener's index will decrease.

$$H' = -\sum_{i} p_i \log p_i$$

Equation 1.3 – Shannon-Wiener index (H'), used to calculate the diversity of individual simulations; p_i is the proportion of species i in the whole population.

[C] Results

The model was intended to capture known conditions within the gut: host-ingested compounds will typically provide the majority of the microbiota nutrients, but various compounds that can feed microbes such as mucins and attached sugars such as fucose are being secreted by the epithelium[145–147,153]. Although this study considered interactions across the gastrointestinal tract, the model should capture comparable processes whenever microbes grow upon a host epithelial surface, including plant roots[101,133], corals[176] and other symbioses[70,71,104,113].

The microbiome is an ecological and evolutionary system where bacterial populations undergo a continual turnover. The key implication of this is that strains that are faster dividing, or better at surviving, will gradually replace strains that divide more slowly or persist less well (Figure 1.1 B). As a result, whenever the strains and species of microbes that provide the greatest benefit to the host are not the fastest growers, the host will tend to lose its most beneficial strains and lose the advantages of a healthy microbiota.

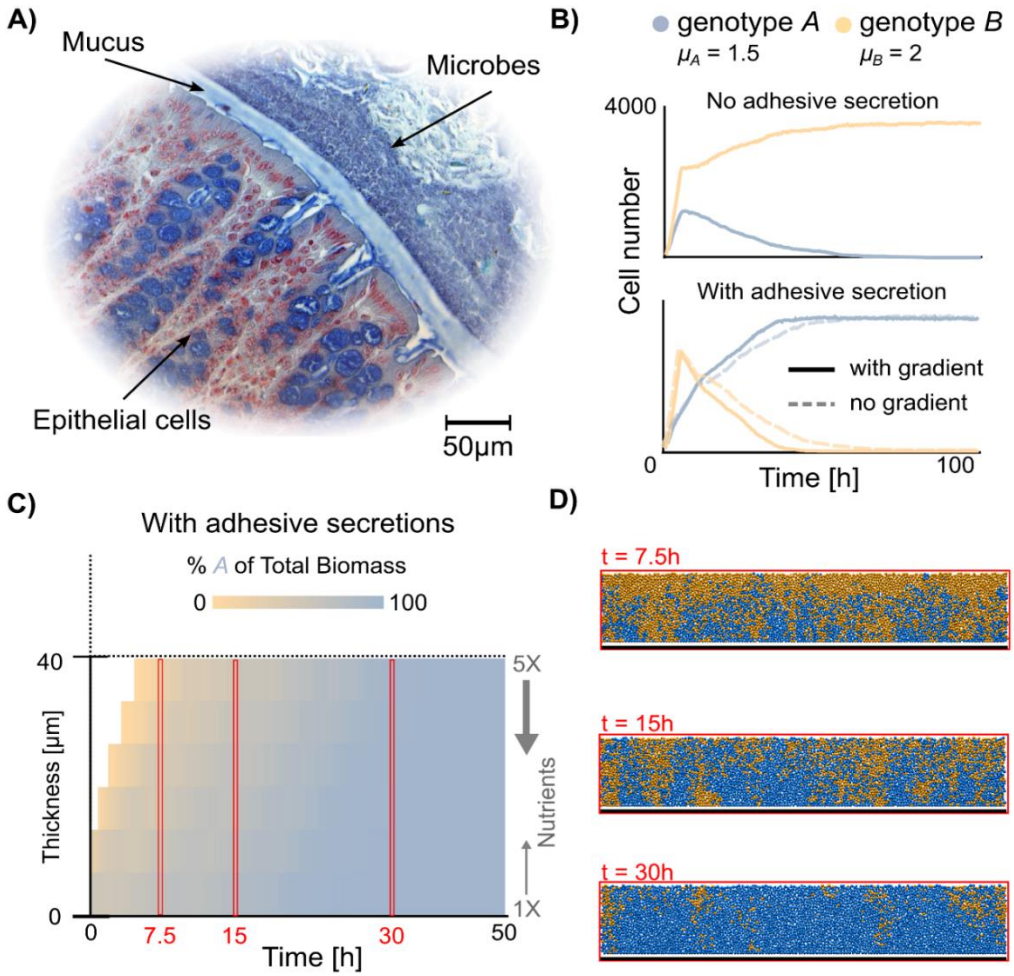


Figure 1.1 – Host-secreted adhesion can select for specific microbial species (genotypes). The competition between a slow-growing genotype A (blue) and a fast-growing genotype B (orange) on the gut epithelium is simulated. A) The biological context of the model: micrograph of a health mouse large intestine (C57Bl/6 mouse, transverse colon stained with Alcian Blue), photo credit: Lev Lichtenstein and Eugene Chang. B) Top: the faster growing orange species (maximum growth rate $\mu_B = 2$) outcompetes the slower growing species B ($\mu_A = 1.5$). Bottom: when a host-secreted factor increases the relative ability of species A to resist displacement, species A is able to outcompete faster growing species B. The impact of adhesion on species A occurs regardless of whether adhesion is a host-provided (gradient present) or microbial-intrinsic (no gradient) factor. C) Average biomass distribution across the entire simulation with shows that host provision of a species A specific adhesion factor causes the two strains to separate vertically from each other (species A localises below species B). D) Snapshots from the simulation at time points corresponding to the red sections in (C). In these simulations, it is assumed that the concentration of luminal nutrients is five times the epithelial nutrient concentrations. However, the conclusions are robust to other ratios of nutrient supply (see MSI Figure 1).

[C.1] Adhesion as a mechanism of host selection

Schluter *et al.* (2015) demonstrated that microbes could use adhesion as a strategy to outcompete other strains (recapitulated in Figure 1.1 B). The process of cell division means that cells will push into each other as numbers increase. In our model, we capture these collisions and more adhesive cells are better able to resist displacement relative to non-adhesive cells, according to a physical model of viscous drag experienced by cells within a mucus layer. As adhesion allows cells to resist displacement by other cells, an adhesive strain can better colonise the base of an expanding biofilm community as it is less likely to be pushed away from the epithelial surface by another strain. So long as there is growth at the base of the biofilm, an adhesive strain will be in a dominant position to divide and push all other strains up and out of the system (Figure 1.1 C and D). Therefore, adhesiveness may be a bacterial strategy for survival within a nutrient-saturated biofilm that enables commensals to thrive within the gastrointestinal tract[177–179].

Given the advantages that adhesion can provide to a microbe, differential adhesion might also be employed as a host strategy to positively select for particular strains or species. Specifically, rather than adhesion being a property of the microbes themselves (Figure 1.1 B, no gradient condition), what will happen if instead the host secretes factors into the mucus layer that promote adhesion (Figure 1.1 B, gradient condition)? To capture such host-secreted factors, the diffusion of a molecule that is continuously secreted from the epithelial surface and preferentially sticks to specific strains in the microbial community, limiting their movement, was simulated. In this model, the host-derived adhesiveness decays with greater distance from the source of secretion (Figure 1.1 B, see MSI Figure 1).

Does an adhesion gradient function as an effective host selection mechanism? A key distinction for this hypothesis is whether adhesion is a property of the microbes themselves or whether the host provides a factor that influences the adhesion of particular strains or species of microbe. It is possible to compare the effect of a host-secreted molecule that affects microbial adhesion to the case where differential adhesion is a property of the microbes themselves

(Figure 1.1 B). Specifically, a host derived molecule is expected to have the most effect close to the epithelial surface, whereas a microbe-based factor will be a property of the microbial cell that, all else being equal, will have the same effect at any position. In order to compare these two scenarios, the maximum adhesive effect at the epithelial surface in the host-derived adhesion model was set to be the same as the effect that occurs throughout the microbial community in the microbial-adhesion model.

Despite the fact that the total adhesive effect was much weaker in the host-derived adhesion model due to the gradual decrease in strength away from the epithelium, a comparable effect on microbial competition was observed in both cases. In fact, the host secreted case can even perform slightly better than an intrinsic microbial adhesion (Figure 1.1 B gradient vs no gradient conditions respectively). The reason for the enhanced effect of host-secreted adhesion is that an adhesion gradient maximises the effect of adhesion at the point where it is most important: at the epithelial surface itself. Thus, cells closest to the surface are the least likely to be displaced by any other cells in the system, ensuring that the favoured cells rapidly conquer the surface and push all others out of the system. The effectiveness of host adhesion for a range of lumen nutrient concentrations and growth rates of the host-favoured strain was explored in MSI Figure 1. These simulations showed the intuitive result that the slower a strain grows, the greater the host-supplied adhesion that is required to maintain it. In addition, they showed that adhesion-based selection was effective across a wide range of lumen nutrient concentrations.

An important potential feature of microbial selection via adhesive molecules is that it can be both specific and variable. Any one adhesive molecule can have a specific target and if the host can generate a wide variety of molecules, they can target a diverse set of strains and species. As discussed above, two clear candidates for such specificity in mammalian systems are the glycans of mucus molecules and IgA. Both come in a vast diversity of forms and, importantly, there is evidence that both preferentially associate with certain microbiome species[162–168].

Both IgA and mucosal glycans can act as candidate molecules that could support slow growing species, preventing competitive exclusion of these species and thereby maintain a diverse set of strains at the epithelial surface. Without the provision of selective adhesive secretions, maintaining diversity is difficult: starting from a mixed set of species on the epithelial surface typically leads to the loss of diversity during the simulation. This occurred even when all strains had the same growth rate through stochastic processes (Figure 1.2 A) but it was particularly problematic when different strains had different growth rates (Figure 1.2 B). However, in the case where a host non-uniformly secretes a range of adhesive molecules across the epithelial surface, each with their own unique focus of secretion, these secretions had a powerful stabilising effect on diversity (Figure 1.2 A and B). Positive selection by the host is effectively creating a set of niches along the epithelial surface that guarantees that each strain has at least one position where it can establish and thrive.

[C.2] Adhesion stabilises diversity across feast-famine cycles

So far, the model has simplistically assumed that microbial species have a constant growth rate over time. In practice, the growth rate of a particular strain or species will change according to the amount and types of nutrients that are available[145,180–184]. Species-specific nutrient availability can be modelled by introducing two food types a host may consume. The first food type was only available periodically and species A microbes relied exclusively on this food type. The second food type was always available and it contained nutrients for a generalist species B that can survive on both food types. An example of species B is *Bacteroides thetaiotaomicron*, which can consume wide range of complex carbohydrates, some of which are food-derived whereas others may come from host secretions[145].

Modelling the interaction between microbes consuming two different food sources highlights the fact that variability in diet can be difficult for a host: sudden shifts in host diet bring the risk that certain microbial species will be lost if their preferred nutrients are in short supply. These dietary shifts might result in a loss of metabolic potential, or other benefits, to the host. Specifically, the model predicted that species A will often be lost during the fasting periods

where the first food type was not available (Figure 1.3 A). These extinction events, however, can be prevented through the use of selective host adhesion (Figure 1.3 B). Adhesion can create a region of the gut where species A always persists by providing a large enough competitive benefit despite the lack of growth during a fasting period.

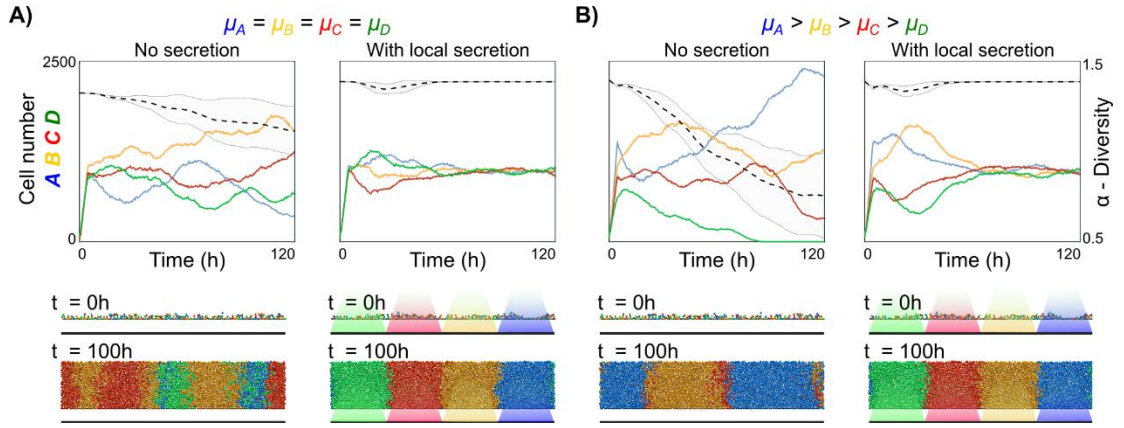


Figure 1.2 – Horizontal gradients of host-secreted adhesion maintains diversity. Here, four genotypes compete and the host may secrete four different factors in four different horizontal regions. Time-series data show cell number (left y-axis) of four genotypes over time for one representative simulation and are accompanied by alpha-diversity data across the whole simulation space from 50 independent simulations (right y-axis, black dotted line, the grey lines above and below show the standard error of the mean). Snapshots show the simulation at t=0h and t=100h, where adhesion-promoting host secretions form horizontal gradients and are indicated by shaded areas with the maximum effectiveness of the secretion in the centre of each region. Colours of host secretions correspond to colours of genotypes for which they are specific. The *x* and *y* axes in the snapshots are distance along and away from the epithelial surface respectively. A) When the four genotypes grow at identical rates, stochastic events can lead to uneven genotype abundances over time. Accordingly, diversity tends to decrease slowly over time. When the host secretes factors that promote the ability of a species to resist displacement in a confined horizontal region, this provides a refugium for this genotype and diversity remains high. B) When species differ in their maximum growth rates, diversity loss occurs rapidly and dramatically. Diversity loss can be prevented by host secretions that create ecological refugia for the slow growing species.

In ecological terms, selective adhesion creates an ecological refugium for species A that prevents its extinction. This refugium may facilitate specific resource usage, for example through digestion of the specific adhesive compound, which would favour the persistence of the targeted species. The creation of refugia by a host may be a way to ensure that a diverse and desirable

set of species can be maintained at epithelial surfaces at all times in the face of environmental fluctuations caused by diet and other factors.

[C.3] Host matrix (mucus) secretion and selection

The existence of a mucin network is consistent with a model of differential adhesion as different genotypes have differing abilities to aggregate within mucin[185,186] and some microbial species even attach directly to mucins[186,187]. However, mucins are constantly produced by the host and this creates a continual movement of mucus away from the epithelial surface which this model does not yet capture. Moreover, the rate of mucin production and consequent mucus flow has the potential to vary both within and between individuals. During inflammation and infection, for example, mucin production rates can increase substantially[31,188–191].

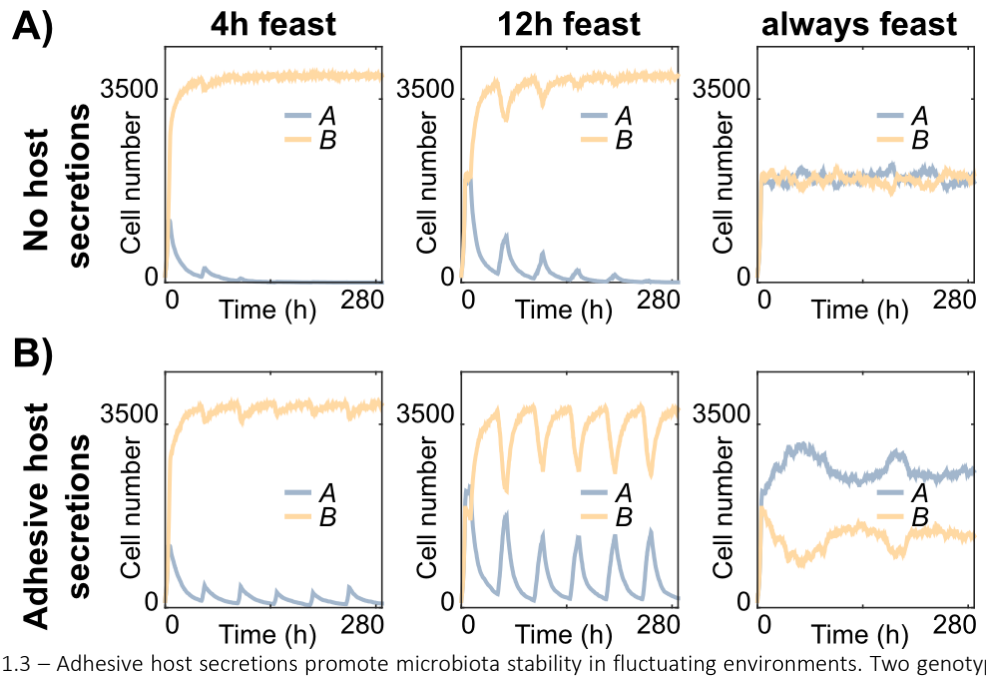


Figure 1.3 – Adhesive host secretions promote microbiota stability in fluctuating environments. Two genotypes are simulated: a generalist genotype B that can consume lumen nutrients which are available at all times and a specialist genotype A which may only consume a second nutrient that is only available periodically. The intervals of feast periods may vary: at 4h and 12h feast durations per 48h period; "always feast" is a control where the second nutrient is also available at all times. A) Without adhesive host secretions, genotype A is lost from the community if "feast" periods are rare and short. B) When the host secretes an adhesion-promoting factor that creates a horizontal region in which genotype A resists displacement better than genotype B, both genotypes can be maintained even when the environment fluctuates. The variability in the "always feast" condition is due to stochastic fluctuations in population size. These results are robust to changes in duration and periodicity of feasts (see MSI Figure 2).

Given that the many of the effects of differential adhesion within this model are due to differential movement of cells away from the epithelium, it is important to understand what happens when cells are translocated along with the secreted mucus. Mucus-induced translocation was incorporated into the model by assuming that cells are contained within a mucus gel that moves upwards a fixed distance each time step, where the magnitude of this movement vector increased for an increased rate of mucus flow. Countering this effect, cell division near the epithelial surface will generate new biomass that repopulates the space created by mucus flow. Thus, it is possible to ask whether mucus flow rate influences the benefits of host-provided adhesion.

For low rates of mucus flow, the benefits of adhesion were unaffected and adhesiveness provided a competitive advantage to microbes (Figure 1.4). However, increasing the rate of mucus flow changed the prediction. As flow rates increase, cells were carried more rapidly away from the epithelial surface and it became more and more challenging to repopulate the space created at the epithelial surface. Repopulation was more challenging for adhesive cells that tend to move less within the mucus gel than non-adhesive cells (Figure 1.4 C). The result was that, on average, the adhesive cells were carried more rapidly up and out of the system than non-adhesive cells and, importantly, this allowed non-adhesive cells to dominate. However, this effect only occurred for a relatively narrow parameter window because, with too much mucus flow, no cell type can repopulate the epithelial surface rapidly enough and all microbes are swept away and into the lumen (Figure 1.4, B.3). In sum, increasing the rate of mucus flow can make adhesion shift from a mode of positive selection – keeping strains close to the epithelium – to negative selection where adhesive strains are flushed out of the system. Too much flow, however, and all strains will be flushed out.

[D] Discussion

The composition of the microbiota associated with a host is central to a host's health and, ultimately, its evolutionary fitness. A host can benefit, therefore, from strategies that allow it to influence which microbes thrive at its epithelial surfaces[79,192]. In this chapter, we have shown

how host-provided adhesiveness can act as a mechanism to affect the composition of a host's microbiota (Figure 1.1 and Figure 1.2). In particular, these models predict that, under certain conditions, a host can use adhesion to favour the colonisation and competitiveness of beneficial microbial genotypes. While the focus of these models has been on increasing adhesion, a corollary of these predictions is that host might also secrete factors that reduce adhesion in order limit a strain's colonisation.

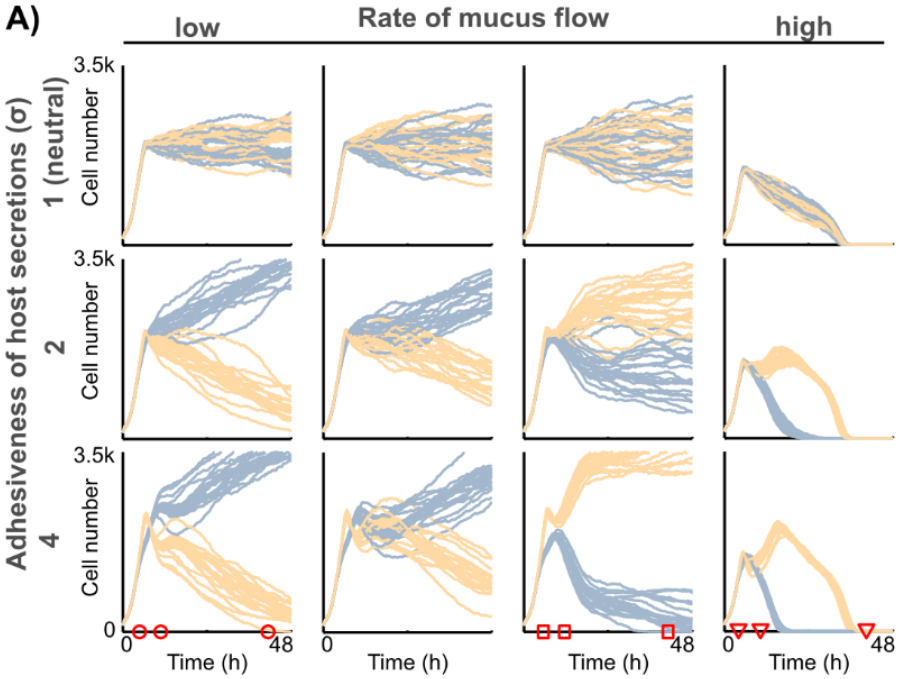


Figure 1.4 A – Hosts can combine adhesion and mucus flow to select for, or against, a particular microbial species. Each plot shows cell numbers of two genotypes, A (blue) and B (orange) over a 48h simulation period for 50 independent simulations. Without adhesion-promoting secretions from the host, both types are equal in all parameters and grow at the same rate (top row). At each timestep, cells are moved upwards and away from the epithelium simulating a mucus matrix in which cells are embedded and which pulls them along (B.3). If this flow rate is faster than the recolonization of now freed space below, all cells get washed out (top right). When the host secretes a factor that increases the ability to resist displacement of genotype A, low flow-rates recapitulate the findings in from Figure 1.1 that show how being more adhesive can convey a competitive advantage (two far left columns). However, at higher flow rates that still allow persistence of cells in the mucus layer (3rd column), the fate of the more adhesive genotype A is reversed and it is preferentially removed from the system relative to the non-adhesive genotype B.

The composition of the microbiota associated with a host can change widely in response to changes in the environment[193–196], chiefly host diet in the case of gut communities[180–183]. Host-provided factors that positively select for particular genotypes have the potential to limit the

scale of these fluctuations. In particular, a host can provide ecological refugia for symbionts that might otherwise be lost during shifts in composition[116,154,156,184,197,198] (Figure 1.3). Along with feeding, adhesion seems a promising mechanism for refugia generation due to the potential for specific binding to target symbionts.

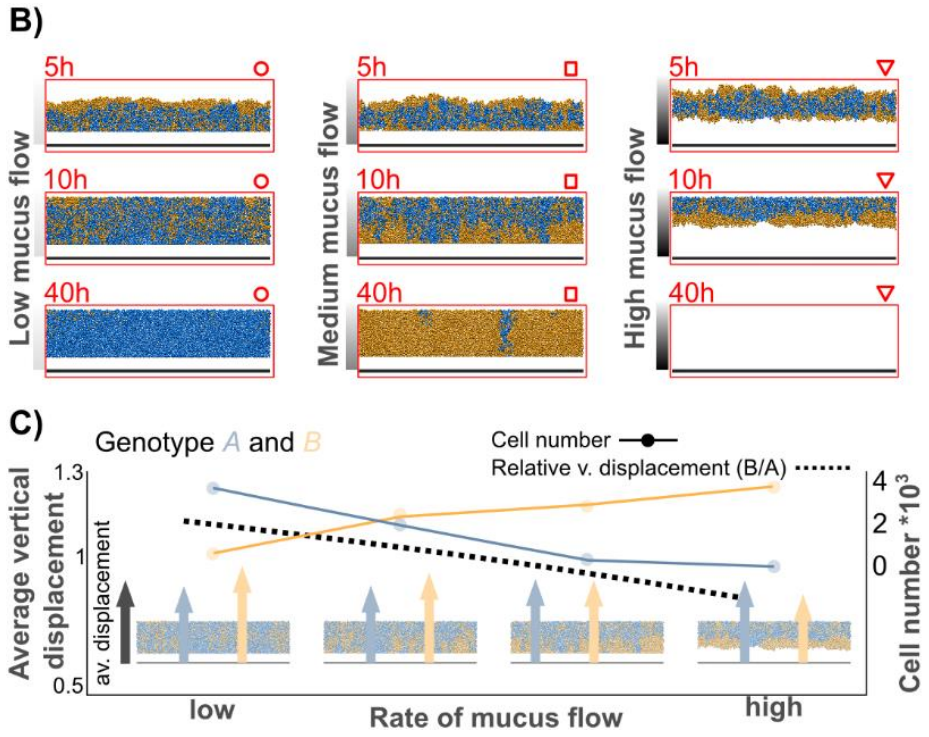


Figure 1.4 B and C – B) Snapshots of representative simulations show that less adhesive cells of genotype B are pushed out at low mucus flow rates (left) but tend to persist better at higher mucus flow rates (middle), while both are flushed out with higher flow rates (right). C) High fitness is associated with the ability to resist vertical displacement; quantification of relative vertical displacement (dotted black line) between the genotypes and the resulting effects on fitness (solid lines). The arrows indicate average vertical movement rates of the two genotypes in µm/h. At low mucus flow rates, cells of genotype A experience less vertical movement relative to cells of genotype B. This is reversed at higher flow rates as the less adhesive B cells can more easily repopulate the void space created through translocation of the whole community with mucus flow. This leads to higher fitness (measured as cell number at t=22h) of the adhesive genotype A at low mucus flow, and lower fitness at high flow.

Symbiotic microbes often sit encased in a matrix – mucus in animals and mucilage in plants – that flows outwards under host control. The rate of flow of mucus in animals is known to be extremely variable. In particular, during infection or dysbiosis the inflammatory response is associated with mucus hyper-production by goblet cells within the epithelial layer[188–190]. This model predicts that the flow rate away from the epithelial surface is critical to whether increased

adhesion acts to benefit or inhibit a particular genotype (Figure 1.4). Under low flow rates, adhesion helps cells to displace less adhesive genotypes and stay close to the epithelium. However, adhesive populations are less able to counter the mucus flow by expanding back towards the epithelium. The result is that, at high flow rates, adhesive genotypes are more readily carried away from the epithelium than less adhesive genotypes.

[D.1] Candidate host-secreted adhesive molecules

Many host-secreted factors have the potential to affect adhesion of microbial cells, including the glycan residues attached to the mucin backbone of mucus and IgA. Both mucin-associated glycans and IgA molecules are highly diverse, which could confer specificity to individual microbiota species.

There are multiple host compounds that might enable hosts to affect microbial adhesion. Mucins are heavily glycosylated and many species of microbes have the ability to attach to these glycans, both as a way to digest them[145–147] but also seemingly as way to anchor themselves[146,163,186,187,199,200]. These diverse moieties then can serve as a nutrient source that selects for particular symbionts[129,156], particularly species that have the necessary enzymes to remove glycans from mucins and other macromolecules[145–147]. However, the extreme structural diversity of these glycans also raises the possibility that they act as specific attachment targets for certain symbionts[163,168]. Glycans may serve then to select for particular strains both by nutrient provision but also by affecting adhesion; our model suggests that the two can work well together (see MSI Figure 3). Consistent with this prediction, there is growing evidence that host-secreted glycans are important for which species occur where in the gut[157,201] as well as in providing resistance to infection[116].

Another candidate for the manipulation of microbial adhesion is IgA, with its well-documented ability to bind to bacterial epitopes and other molecules[16,202–206]. The model presented here raises the possibility for a positive effect of IgA binding on targeted microbes. Under normal conditions, IgA is produced in large amounts at the gut epithelial surface in response to the presence of symbionts[17,150,207,208], and it appears to coat the majority of bacteria

in the gut[165,209–212]. These immunoglobulins stay close to the epithelial surface by binding mucins within the mucus layer[213–215].

Broadly consistent with being a mechanism of positive selection, IgA has been shown to promote bacterial adhesion and biofilm formation in vitro[216,217]. IgA adhesion is possible via both by a hyper-variable region that enables different IgA forms to target specific microbial epitopes, and a non-specific binding region that adheres both to microbes[179,218] and to the host-produced mucins that form the mucus matrix at the epithelial surface[199,213–215]. As a result IgA co-localises with epithelial microbes[219] both via non-specific binding[206,220] and by the classical hyper-variable, complementary determining region-3 (CDR3) facilitated binding[164]. It is this latter specific binding that raises the possibility that hosts can enrich for a particular microbial strain by secreting a particular IgA form into the epithelial mucus. However, under conditions of high mucus flow, particularly during infection, IgA may act in an opposite manner and help to pull strains away from the epithelial surface leading to their control and clearance[221–224].

[D.2] Theoretical predications

These models make a number of testable predictions that can be used to reject the hypothesis that host-derived adhesive molecules are important in modulating the composition of the microbiota.

The first prediction of these models is that the composition of the microbiota, particularly at the epithelial surface, will shift when the abundance of host-derived adhesive molecules is altered. This prediction is supported by studies showing compositional shifts in the microbiota when fucosylation patterns are altered[116,155,157,167], IgA production is inhibited[131,205,225,226] or its specificity changed[41,206]. A related prediction is that a reduction in the abundance, or diversity, of adhesive molecules should be associated with a loss of diversity in the microbiota. This prediction can be extended to the expectation that over small scales within the microbiota the production of an adhesive compound, and its target species, should be varied to facilitate the formation of refugia.

The previous predictions, however, do not consider whether adhesion is functioning as a mode of positive or negative selection. If a particular class of adhesive molecule is functioning in positive selection, then a more specific prediction is that the molecule should be preferentially expressed when the target focal strain is disadvantaged, as might occur during a famine period or diet switch.

Consistent with this, there is evidence of up-regulation in IgA production and secretion in malnourished individuals[227,228], as well as up-regulation of glycan production during famine periods[153,154]. The available data, therefore, does not reject the hypothesis that host-derived adhesive molecules are important for microbiota composition. However, while the data do not reject this model, the results could also be explained by other effects of the glycans and IgA that are not associated with adhesion, such as their nutritional value to the microbiota.

A more specific test of this model would first need to identify specific species that are preferentially bound by a particular adhesive molecule. With this information, the model predicts that removing (or adding) the focal molecule at the epithelial surface will preferentially affect the frequency of the targeted species over other species. In particular, if adhesion is functioning in positive selection, when the adhesive molecule is removed the targeted symbiont should suffer a loss of abundance.

There is currently little data relevant to this prediction. However, it has been observed that anti-*Helicobacter pylori* IgA is associated with increased *H. pylori* colonisation of the stomach epithelium in a manner consistent with the effects of increased adhesion[229]. This data is consistent with this adhesion model but not with alternative models such as IgA functioning as a way to target and destroy members of the microbiota via the immune system.

Finally, this work suggests that mucus flow rate has the potential to shift the sign of selective effects. For a given adhesive molecule that benefits a symbiont under low mucus flow, that same benefit can turn into a cost to the symbiont under conditions of sufficiently high mucus flow. There is no current data available to corroborate, or contest, this prediction.

[E] Conclusions

This work suggests that adhesion, along with the secretion of a matrix like mucus, can allow a host to exert control over its microbiota. The effect of adhesive molecules is flexible, with the potential to act in both a positive and negative manner on their target species. A strength of this model is the potential for specificity in a host's secretion of adhesive compounds, particularly relative to selection by growth inhibitors, like defensins, or indeed host-epithelial feeding. However, the possibility that adhesion acts as a mechanism of positive selection is perhaps the most surprising prediction of this work. Further, it appears that positive selection can be powerful as an anchored strain can then divide and push other strains away from the surface. This study emphasises how hosts can benefit just as much from helping beneficial strains as harming pathogens.

Chapter 2

Fine-scale differences in the distribution of adaptive immune and microbiota components

[A] Introduction

The beneficial effects of a microbiota upon its host are highly dependent upon both the composition[17,21,48,49] and diversity[42–45] of the microbial community. The drive of both ecological and evolutionary dynamics to disrupt communities is therefore problematic for a hosts' maintenance of microbiota community composition[17,123–126]. Moreover the specific factors that modulate community composition can be expected to vary depending upon the scale at which the community is examined[6,8].

Chapter 1 outlines how host provision of selectively adhesive molecules can facilitate the maintenance of community diversity by spatially structuring the microbiota. Spatial patterning of communities preserves diversity by isolating small groups of species. Where this spatial segregation is driven by environmental selection, a globally disadvantaged species can be preserved within a community through the creation of refugia in which the species is selected for by the environment[230,231]. In the absence of environmental selection, spatial structuring still improves the maintenance of diversity by limiting the number of intraspecific species interactions (see Chapter 3 A.2).

[A.1] Spatial heterogeneity of the microbiota

Microbiota composition varies between individuals[232,233], geographic regions[234,235] and species[72,73,76,77]. This compositional variation has been associated with host genetics[236,237], housing[238,239], sex[240,241] and dietary intake[183,242]. The exact composition of the microbiota is not simply origin dependent but is maintained in response to perturbation, such as the introduction

of an alternative species' microbiota[72]. The mechanism behind this maintenance has been suggested to be the host's immune system[243–246].

The immune system, particularly the adaptive immune system, is similarly variable between individuals and species. Different species have distinct gene fragments that rearrange to form T-cell and B-cell receptors[93,94], and even the process by which these receptors are formed can vary between species[95] (e.g. the avian system process of gene rearrangement is distinct from the mammalian; see General Introduction C.3). The process of gene rearrangement and cell proliferation over an individual's life leads to a huge variation in the resting potential of the adaptive immune system between individuals.

Even within individuals, the composition of the microbiota and adaptive immune systems is heterogeneous. The skin microbiota composition is distinct from that of both the gastrointestinal and vaginal tracts[233,247]; cells within the systemic immune system are functionally distinct from those of the mucosal immune system[248]. Even within tissues microbiota and immune composition vary; mucus secretion rates[143] and types of Toll-like receptors (TLRs) vary between different sections of the gut[140], the concentrations of immune cells vary depending upon the number of Peyer's patches and lymphoid follicles within the gut epithelium[81]; the composition and density of the microbiota similarly changes between the small and large intestine with the total number and proportion of anaerobes increasing distally[85–87,201].

The spatial scale at which both the microbiota and immune system is considered might therefore change both community abundance and composition. In fact, ecological theory has shown that communities, and the dynamics that shape them, are dependent upon the scale at which the population is examined[1,5,6,8].

[A.2] Factors that affect communities vary according to scale

Community composition is dependent upon three main factors: the physical environment in which it is found; the species interactions it comprises; and the dispersal

limitations of its individuals[6,8]. The relative importance of these factors is dependent upon the scale at which the community is considered[1,5,6].

At large scales environmental conditions can be expected to dictate species survival[1,6]. For example, in traditional ecological studies, the climate in which a community exists varies relatively little at small scales whereas climate effects on community composition are pronounced at larger scales. Within the context of the microbiota the effects of climate can be compared with the physical processes that vary along the length of the gastrointestinal tract (e.g. oxygen-availability, pH and immune compounds). Conversely at fine scales community composition is more dependent upon species interactions (e.g. competition and mutualism) and factors limiting dispersion (e.g. motility, adhesion)[8–10].

The scale at which a system is studied can also affect its apparent stability[1,7]. Whilst small-scale populations may fluctuate due to local extinctions, the community composition at a larger scale remains constant with locally extinct species re-populated from the global population. Thus, spatial structuring at fine scales (decreased local diversity), mediated by species interactions or environmental constraints, can maintain community composition at larger scales (increased global diversity)[249–251].

Large-scale interrogations of the immune-microbiota relationship may characterise the immune system's role as an environmental modulator, whilst missing a more specific interactions between microbes and immune molecules. Understanding the mechanisms of factors involved in modulating microbiota composition, and its stability, therefore requires consideration across scale.

[A.3] The chicken as a model species

The chicken has been an invaluable model since the advent of the modern field of immunology, especially in the identification of the adaptive immune cells[252]. For example, the role of B-cells as antibody-producing lymphocytes and of T-cells in cell-mediated functions was initially identified within Bursa of Fabricius deficient chickens that failed to produce antibody in response to *Salmonella* challenge[252].

Though the avian and mammalian immune systems share many immunological features (e.g. the presence of an adaptive immune system) the mechanisms through which the immune cells are generated varies. In particular chicken B-cells are derived from the Bursa of Fabricius rather than the bone marrow and utilise only a single V-segment from which B-cell receptors (BCRs) are formed[252]. The hyper-variability associated with the adaptive immune system is still maintained within chickens, but it is generated through the insertion of V segment pseudogenes into the pre-formed BCR[253,254].

Chickens are a major global food source and the majority of chickens consumed are sacrificed before the effective establishment of the adaptive immune system (by *c.* six weeks of age)[255,256]. Protecting flocks from infectious disease is therefore an important challenge for food security. Traditionally antibiotics have been used to prophylactically protect flocks against infectious disease, though antibiotic resistance has led to regulations restricting the use of antibiotics in treating livestock[257]. As the microbiota has been demonstrated to have a protective role against infection, the use and investigation of probiotics and prebiotics in infection prevention has become an integral field of research[258,259].

[A.4] Chapter 2 aims

One of the predictions of the individual-based model presented in Chapter 1 is that host secretion of a specific adhesive molecule can lead to the formation of refugia for targeted species (Chapter 1 C.1 and D.2). This prediction, in association with theoretical work demonstrating the necessity of spatial patterning in facilitating stability within the microbiota[230,231,249,260], suggests the existence of fine scale variation in community composition within the microbiota.

This chapter therefore aims to develop a method for fine scale analysis of a chicken caeca, and to investigate the adaptive immune repertoire and the microbiota for evidence of spatial variation. First the methodology behind sampling a fine scale grid is assessed for the introduction of biases from sampling, amplification, sequencing or processing. Secondly, whether there is evidence of a difference between large-scale and small-scale samples in terms

of population structure, composition and diversity is considered. Finally the fine-scale grid is used to assess whether there is evidence of spatially structured micro-niches as predicted in Chapter 1.

[B] Methods

[B.1] Sample collection

A male inbred Line 6 (L6), 6-week-old chicken was sacrificed and the intestinal tract removed, and stored in cold phosphate buffered saline (PBS) on ice for transfer to the lab for processing. Once in the lab, a 2 x 2 mm^2 (4 mm^2) grid (Figure 2.1) was cut from the right caecum, starting with columns (A to E) that were separated into 4 rows (1 to 4) using the same sterile scalpel. The full grid encompassed the entire circumference of the caecal section. The grid sections were placed individually in 200 µl of RNAlater, and stored at 4 °C overnight before being transferred to a -80 °C freezer for long term storage.

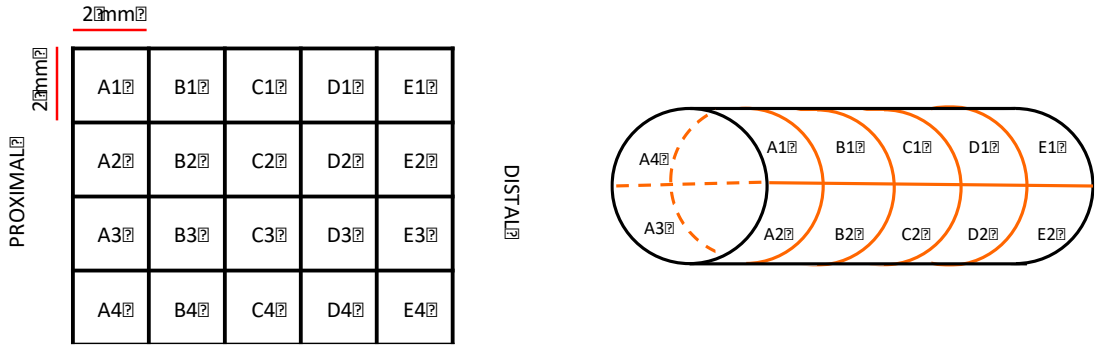


Figure 2.1 – Schematic of a L6 6-week old chicken caecal grid sampling structure. The sampling structure as a 2D schematic (left) where from left to right the columns indicate movements laterally down through caeca towards the caecal sack base (i.e. column A is closest to the caecal tonsil). The rows cover the entire circumference of the right caeca to form a cylinder of the caeca (right). Segments are named relative to their location in terms of column and row (e.g. column A row 1 is segment A1).

[B.2] Tissue extractions for primer design and sample processing

16S primer amplification and sensitivity was tested using *Escherichia coli*, *Salmonella enterica* and *Enterococcus faecalis* genomic DNA (gDNA). Immune cell (immunoglobulin A (IgA)-producing B-cells and T-cells) primer amplification and sensitivity was established using

complementary DNA (cDNA), produced as in B.3, from splenic RNA. To amplify both microbial 16S and immune fragments from the same sample required dual DNA/RNA extractions from sample tissue.

[B.2.1] Genomic DNA extraction

Bacterial gDNA was extracted using Promega Wizard Genomic 4 DNA Purification Kit. Single clones, selected by plating, were cultured overnight in LB broth at 37 °C with 225 rpm shaking. gDNA was extracted from 1 mL of starting culture.

[B.2.2] RNA extraction

Splenic RNA was extracted using QIAGEN RNeasy Mini extraction columns, according to the manufacturer's instructions. The tissue (< 1 mg) was disrupted and homogenised using a bead beater with 0.1 mm sterile glass beads in 600 µl buffer RLT for 2 minutes. The sample was eluted with 30 µl nuclease-free water.

[B.2.3] RNA/DNA dual extraction

RNA and DNA were extracted from samples using QIAGEN AllPrep RNA/DNA extraction columns, according to the manufacturer's instructions. The tissue was disrupted and homogenised using a bead beater with 0.1 mm sterile glass beads in 350 µl Buffer RLT for 2 minutes. The weight of the tissue was checked as below the capacity of the columns for each sample. The processed samples were eluted with 14 µl RNase-free water and 50 µl (pre-heated to 70 °C) Buffer EB for the RNA and DNA respectively.

Sample RNA and DNA concentrations were measured by Qubit (RNA BR and dsDNA BR respectively), according to the manufacturer's instructions.

[B.3] cDNA synthesis

5' RACE cDNA was made from RNA, according to the SMARTer reverse transcriptase protocol. 2.75 µl RNA was combined with 1 µl 5' CDS Primer A (12 µM) and incubated at 72 °C (3 min), then 42 °C (2 min). 2 µl 5X first strand buffer, 1 µl DTT (20mM), 1 µl dNTPs (10 mM), 0.25 µl RNase Inhibitor (40 U/µl), 1 µl SMARTer 3R IIA oligo (12 µM) and 1 µl SMART reverse

transcriptase were combined and added to each reaction before incubation at 42 °C (90 min), then 70 °C (10 min). Product was diluted in 20 µl tricine EDTA buffer.

[B.4] DNA amplification

[B.4.1] Primer design

As immune cell receptors are generated through gene rearrangement with a variety of variable region starting points, 5' RACE was used to attach a universal DNA sequence to all RNA fragments during the cDNA synthesis. Primers were then designed to complement the universal sequence (UPA long) and the UPA long primer to which barcodes were attached (UPA short). A distinct reverse primer, designed within the first 100 base pairs (bp) of the C-region, conferred specificity to each TCR and BCR chain. Primer sequences are shown in Table 2.1.

Table 2.1 – Information on all primers used, including sequence, target regions and the size of the amplified product. The primer sequences are shown using standard base and redundancy codes. 'NA' indicates the amplicon is of variable length.

Primer name	Primer Sequence (5'->3')	Target region	Amplicon size
E338	ACTCCTACGGGAGGCAGCAGT	16S gene	462 bp
UPA long	TAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	5' RACE target	*NA*
UPA short	CTAATACGACTCACTATAGGGC	5' RACE target	*NA*
β-actin F1	TGCTACGTCGCACTGGATTT	β-actin	170 bp
16SCR45F	GCGGTGAAATGCGTAGAGAT	16S gene	108 bp
R806	TAATCTWTGGGVHCATCAGG	16S gene	462 bp
Cβ7	GAACCAGATGTGGTCATCTTTTC	TCRβ C-region	*NA*
Cγ25	AGGAACATGAAAAACAAGTGGTGTATGTTTG	TCRγ C-region	*NA*
IgA H1	GGAGGAGCTGGTAGAGGGTC	BCRα C-region	*NA*
β-actin R1	CTTGGGTATGGAGTCCTGTGG	β-actin	170 bp
16SCR45R	TCTAATCCTGTTTGCTCCCCAC	16S gene	108 bp

16S primers, used for their ability to reliably amplify product and distinguish taxa across the prokaryotic kingdom[261], were chosen which fell within the conserved regions of the 16S gene (Figure 2.2) and which would amplify variable regions 3 and 4[262,263].

The 16S amplicon selected covers variable region 4; as this region showed the lowest error rates for taxonomic classification[264]. As over 400 bp fragments provide highly accurate microbial identities, and variable region 3 has been shown to be most efficient at identifying microbial genus[265], the 16S amplicons were extended to also cover variable region 3.

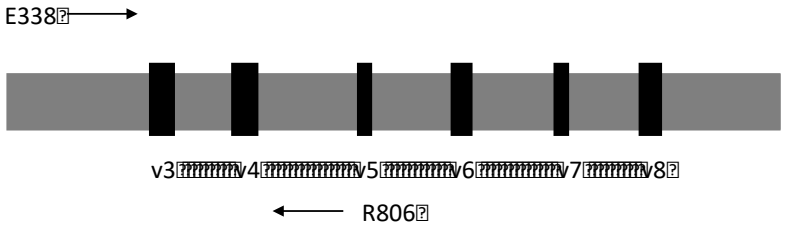


Figure 2.2 – Schematic of the 16S gene from variable region 3 to 8. The 16S gene is made up of conserved (grey) and highly variable (black) regions. The variable regions differ in size and are represented proportionately. The locations of the PCR primers used in this analysis are indicated.

[B.4.2] Barcode addition

In-house designed barcodes, based on the Hamming code principles[266], were used to identify specific samples in order to increase the number of samples run per Illumina MiSeq run. The barcodes are made up of 7 bp; a 4 bp unique barcode and 3 bp quality control to avoid misidentification in the case of mutation introduction. Of all possible barcode sequences only those that were not redundant and had a low GC content were used, leaving 209 potential barcodes.

The barcodes were added to the forward primer of each reaction set (UPA short or E338) and validated by PCR using chicken splenic cDNA or *E.coli* genomic DNA respectively. The amplification ability of the barcoded primers (sensitivity) was tested using titrations of the relevant DNA (SI Table 2.1). There was some variation in sensitivity between the barcoded primers, which did not appear to result from any particular commonality between the barcoded primers.

16S primer sensitivities were unable to be established, as at this time it was not possible to distinguish between low concentration test PCR and water controls due to contamination (see B.5 and C.4).

[B.4.3] PCR conditions

PCR was carried out to amplify specific DNA fragments (16S gDNA and TCRβ, TCRγ and BCRα chain cDNA, primers and fragment sizes in Table 2.1). Immune reactions required a touchdown method to ensure UPA short primer binding to the UPA long complementary

section of the 5' RACE cDNA synthesis, whilst the microbial reactions used a basic PCR format (Table 2.2).

Table 2.2 – PCR conditions for PCRs. Conditions for microbial (16S) and immune (T-cell receptor β, T-cell receptor γ and immunoglobulin A producing B-cell receptors) PCRs are indicated separated as the latter required a touchdown step to facilitate universal primer A (UPA) long binding. 'x' indicates the number of cycles for which the step was run. ∞ represents the store step after PCR completion until the electrophoresis gel was run.

Microbial PCR			Immune PCR		
			98°C	(1 min)	x1
98°C	(1 min)	x1	98°C 72°C	(15 sec) (30 sec)	x5
98°C 60°C 72°C	(15 sec) (15 sec) (30 sec)	x35	98°C 70°C 72°C	(15 sec) (15 sec) (30 sec)	x5
72°C 4°C	(2 min) (∞)	x1	98°C 60°C 72°C	(15 sec) (15 sec) (30 sec)	x35
			72°C 4°C	(2 min) (∞)	x1

Microbial PCR were set up using a master mix containing 5 µl 5X Phusion Buffer, 0.25 µl Phusion polymerase, 0.5 µl dNTPs (10 mM), 0.5 µl E338 forward primer (10 µM), 0.5 µl R806 reverse primer (10 µM), 1 µl gDNA and 17.25 µl PCR-grade water for each 25 µl reaction. All immune PCRs were set up with a master mix containing 5 µl 5X Phusion Buffer, 0.25 µl Phusion polymerase, 0.5 µl dNTPs (10 mM), 0.5 µl UPA short forward primer (10 µM), 0.1 µl UPA long forward primer (10 µM), 0.5 µl R806 reverse primer (10 µM), 1 µl 5' RACE cDNA and 17.25 µl nuclease-free water for each reaction.

The presence of cDNA in each sample was validated by β-actin PCR, using the same ratio of reagents and PCR conditions as the microbial PCRs, prior to specific primer amplification in order to determine which barcoded UPA primers to associate with the samples. The lowest volume cDNA samples were matched with higher sensitivity barcoded UPA primers, otherwise the primers were allocated randomly.

[B.4.4] 16S qRTPCR

qRTPCR was performed on 16S gDNA using the 16SCR45F and 16SCR45R primers in Table 2.1. The standards were 10-fold serially diluted linearized EcoRI-digested (Promega) plasmid standard, constructed by cloning the 108bp V4-5 PCR amplification product of Bifidobacterium Lactis (ATCC Number: SD5219) into the pMiniT 2.0 Vector (NEB) using NEB 5-alpha Competent E. coli.

Reactions were made up as a master mix containing; 9 µl Fast SYBR Green Master Mix (Applied Biosystems), 0.05 µl of each primer (100 µM), 1 µL template DNA and 9.9 µl PCR-grade water in each 20 µl reaction. The qPCR was run in 96-well format on a ViiA7 and analysed using the ViiA7 software. PCR conditions are shown in Table 2.3.

Table 2.3 – PCR conditions for 16S qRTPCR. 'x' indicates the number of cycles for which the step was run.

	95°C (10 min)
	95°C (15 sec) x40 60°C (60 sec)
Melting curve:	95°C (15 sec) 60°C (1 min)

[B.5] Contaminant control

Contaminants were apparent in the water controls of some 16S PCR. The presence of the contamination was inconsistent, although other researchers found the same contaminants in multiple laboratories, which made assigning the source of the contamination difficult. The incidence of contamination was, however, reduced through the use of a sole-purposed still air hood and the use of PCR-grade water.

[B.5.1] Contamination control during sample preparation

16S PCRs were set up in a still air hood that was UV treated before use, along with all non-biological reagents, for 30 min. The hood was then cleaned using DNAZap according to the manufacturer's instructions. All reagents and samples were only opened within the still air

hood. Water controls were included with every sample 16S PCR and sequenced to identify any contaminants.

[B.5.2] Identification of contaminating sequences

Contaminating sequences were designated as those found solely in the water controls, in every water control or in > 50% of the water controls. Sequences found non-trivially (> 10 sequences in any single water control) were also removed as contaminating sequences.

[B.6] Sequence creation and processing

Each sample within the grid was subject to three independent PCRs, each amplified with different barcodes (see B.4). Each triplicated set was processed on a single MiSeq run, with each library made up of a complete grid with unique barcode and reverse primer combinations.

[B.6.1] Sequencing

The Illumina MiSeq platform, with 300 bp paired-end reads, was used for sequencing, using facilities at the Zoology Department, University of Oxford. MiSeq libraries were prepared using NEBnext Ultra DNA Library Prep Kit for Illumina, according to the manufacturer's instructions. Size selection was performed for only a minimum cut off using 0.6 volumes of Agencourt AMPure XP beads.

[B.6.2] Processing raw sequences

All sequence data was initially processed by sorting sequences by barcode, and forward and reverse primer identity. Immune sequences were dereplicated, quality filtered (Appendix 1) and the clones identified, using an in-house sequence processing pipeline. Identity of V, D, J and C regions was determined using BLAST against dictionaries of the respective gene segments for the TCR datasets. Those sequences not identified by BLAST were discarded. TCR sequences had the CDR3 extracted and unique clones identified to form abundance tables. BCR sequences were trimmed to a consistent quality and the sequence length standardised to facilitate clustering. BCR sequences were then aligned by reverse read location and clonal clusters were

determined using single-linkage agglomerative clustering (Appendix 1). These clonal clusters were then used to form BCR abundance tables.

Microbial sequences were processed using the Quantitative Insights Into Microbial Ecology (QIIME)[267] or Divisive Amplicon Denoising Algorithm 2 (DADA2)[268] pipelines, starting with the introduction of demultiplexed files. Operational taxonomic units (OTUs) were formed through QIIME, using an open reference database and the standard 97% similarity threshold. OTUs and DADA2-produced independent sequence variants (ISVs) were taxonomically identified by comparison with the Greengenes 16S rRNA database (version 13_8)[269,270] using the RDP classifer[264].

[B.6.3] Phylogenies

OTU phylogenies were built as part of the OTU picking stage of the pipeline within QIIME. The phylogenies were constructed from the aligned most common sequence of each OTU using FastTree, which approximates a maximum-likelihood tree. Phylogeny creation for the DADA2 dataset, visualisation and image modification was achieved within R (version 3.31)[271], using the R packages ape[272], phytools[273] and ggtree[274].

[B.7] Quality checking sequences

The quality of base-calls along each read of an Illumina MiSeq run can be assessed using DADA2's *plotQualityProfile.r* function. To merge sequence reads the sequences must be long enough to overlap with reliable base-calls; using these primer sets there is potential for a 116 bp overlap. 16S sequences were trimmed using DADA2's *fastqPairedFilter.r* function, to 230 bp and 210 bp on read 1 and read 2, respectively, to compensate for decreased quality of the sequence reads (SI Figure 2.1).

[B.8] Statistics

Sequence data was compared from the different samples using multiple statistical analyses. All statistical analyses were performed using R software (version 3.3.1 and 3.5.0)[271]. Some visualisations involved modification of functions within a putative RepertoiR package

(currently under construction), although predominately figures were generated with ggplot2[275] and cowplot[276] packages.

Packages used to implement statistical analyses include; lme4[277], iNEXT[278], vegan[279] and HH[280]. Simpson's diversity index was calculated according to Equation 2.1, using the iNEXT package. Shannon's diversity index (also known as the Shannon-Wiener index and Shannon's information metric, Equation 1.3) was calculated using ggfortify[281] or iNEXT depending of the requirements of the calculation. The product of the iNEXT call is information on an observed value with its associated upper and lower 95% confidence interval bands. To convert this to a standard error estimate Equation 2.2 was used.

$$\lambda = -\sum_{i} {p_i}^2$$

Equation 2.1 – Simpson's index (λ); p_i is the proportion of species i in the whole population.

$$SD = \sqrt{N} \times (qD.UCL - qD.LCL)/3.92$$

Equation 2.2 – Conversion from 95% upper and lower confidence intervals, as given from iNEXT rarefaction data, into standard deviation estimates needed for calculation of one-way ANOVA. N is the number of individuals in the sample at the end time point, qD.UCL is the upper 95% confidence interval, qD.LCL the lower confidence interval. The devisor (3.92) is the value used for the 95% confidence interval.

Euclidean distance was calculated according to Pythagoras' theorem using Equation 2.3.

$$Dist = \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2}$$

Equation 2.3 – Euclidean distance equation; $\mathbf{y}$ refers to the $\mathbf{y}$-axis co-ordinate of the location, $\mathbf{x}$ to the $\mathbf{x}$-axis co-ordinate. $\mathbf{x}$ and $\mathbf{y}$ are indexed by their relevant locations.

Dissimilarity metrics used within this study were Hellinger's dissimilarity, Chi-squared distance and Kulczynski's dissimilarity (Equations 2.4-2.6 respectively). Kulczynski's and Hellinger's similarities were calculated using vegan[279], and the Chi-squared distance between samples was calculated using analogue[282].

$$K = 1 - \frac{1}{2}\left(\frac{\sum \min(x_{ij}, x_{ik})}{x_{j+}} + \frac{\sum \min(x_{ij}, x_{ik})}{x_{k+}}\right)$$

Equation 2.4 – Kulczynski's dissimilarity metric; x_{ij} refers to the count of species i in site j, x_{j+} to the sum of all species counts in site j.

$$H = \sqrt{\sum_{i=1}^{p}\left(\sqrt{\frac{x_{ij}}{x_{j+}}} - \sqrt{\frac{x_{ik}}{x_{k+}}}\right)^2}$$

Equation 2.5 – Hellinger's dissimilarity metric; x_{ij} refers to the count of species i in site j, x_{j+} to the sum of all species counts in site j. p is the largest count of any species within the dataset. This metric does not make comparisons where there is no incidence of a species in both comparison sites.

$$X = \sqrt{\sum \frac{1}{x_{i+}}\left(\frac{x_{ij}}{x_{j+}} - \frac{x_{ik}}{x_{k+}}\right)^2}$$

Equation 2.6 – Chi-squared distance metric; x_{ij} refers to the count of species i in site j, x_{j+} to the sum of all species counts in site j and x_{i+} to the sum of all counts of species i across all sites.

The similarity between a given distribution and the normal distribution was assessed using least squares analysis. A normal distribution of the same mode was fitted to the data and the distributions compared between the minimum and maximum observed values. The standard deviation chosen to define the normal distribution was that which minimised the sum of least squares deviation. Within this chapter the combined TCRβ normal distribution mode was 42 with a standard deviation of 37, the TCRγ mode was 54 with a standard deviation of 36.

[C] Method Development

There are multiple pipelines available for the analysis of 16S sequence data, including QIIME[267], mothur[283], UPARSE[284] and DADA2[268], that are designed to minimise the probability of sequencing errors affecting results. This section initially compares two of the major, but theoretically distinct, 16S analysis pipelines to consider how 16S sequence data should be processed with respect to fine-scale samples. Both immune and microbial sequence data were then interrogated to determine whether there were any inherent biases in the sampling

process used to generate the fine scale grid samples. Finally contaminating sequences were identified and removed from the microbial dataset.

[C.1] Accounting for sequencing error using microbiome analysis tools

Sequence data suffers from the inherent problem of sequencing errors[285–288]. These errors can occur during the preparation of samples for sequencing, specifically during the PCR amplification steps, or during the sequencing itself. The presence of sequencing error is particularly problematic when considering ecological community information – especially diversity estimates – as the inclusion of sequence or experimentally produced errors artificially increases diversity estimates[287].

The effect of PCR-introduced errors, particularly those that are differentially introduced by varied primers, can be reduced using technical replicates which are amplified using different barcoded primers.

[C.1.1] Quality filtering using microbiome pipelines

Errors introduced during the sequencing stage can be accounted for, to an extent, through quality filtering using Phred quality information produced by Illumina for each nucleotide call in every sequence[289]. Read length cut offs, average Phred score across the whole sequence and sequential low-quality scores can all be used to inform quality filtering. Bokulich et al. (2013) have optimised these parameters for microbial data processed through the QIIME pathway to reduce the effect of sequencing errors on microbial community diversity estimates[289]. These optimised parameters were used within this study.

[C.1.2] Operational taxonomic units

A secondary method to reduce the potential for increased diversity estimates is the incorporation of raw unique sequence counts into groups based on a sequence similarity threshold – in microbial analyses referred to as the creation of operational taxonomic units (OTUs). These OTUs can be created using any given threshold, although a 'species-level' threshold of 97% identity across the full sequence length is often used. An abundance cut-off

threshold can then be applied to the resulting OTUs to reduce the probability of diversity estimates based on sequencing error.

There are a number of pipelines which have been developed to analyse sequence data and attempt to reduce sequencing error effects using quality filtering and OTU formation, including; QIIME[267], mothur[283] and UPARSE[284]. At the time of analysis QIIME was the most commonly used method so this was selected as a method of data processing within this project.

[C.1.3] Probability models

More recently a different method has been developed to attempt to control for Illumina generated sequencing errors based upon amplicon-specific error-correction models[268,290]. A full description of how the error-model is implemented to generate sequence-error corrected independent sequence variants (ISVs), using DADA2, can be found in Callahan's publication (2016) of the updated Divisive Amplicon Denoising Algorithm (DADA)[268].

Simply, an error model built on the assumption of errors occurring independently, both within and between reads, quantifies the rate at which a given focal sequence can be generated from a comparison sequence. A Poisson model of the abundance of the focal sequence, parameterised by the rate at which the comparison sequence might generate the focal sequence, is used to generate a p-value to divide focal sequences into those that could be derived from the comparison sequence, which are corrected to the comparison sequence, or those which could not be derived. This process is repeated with each sequence being assigned the role of focal sequence until there are no more sequences that could be derived from any other sequence in the sample. The remaining sequences can be considered to not contain any sequence error and are designated ISVs.

[C.2] Comparing microbiome analysis pathways

Two microbiome analysis pipelines (QIIME and DADA2) were selected for method comparison, representing the different methods of accounting for sequencing error. The default settings with respect to quality filtering were used for each pipeline[268,289]. However even within each dataset there are different pathways through which sequence data can be

processed to produce either OTUs or ISVs. The analysis might use only one of the reads or the reads might be merged together; chimeras may be included or removed, and taxonomy might be assigned using different thresholds, methods or reference databases.

Both QIIME and DADA2 receive sequence data after demultiplexing, barcode identification and primer removal. This read-in step in QIIME can use forward, reverse or merged sequences, whilst DADA2 reads in either (or both) forward or reverse reads and as part of the pipeline will merge the sequences. This step is important as the DADA2 pipeline builds an independent sequence model for *both* the forward and reverse reads, and then corrects the sequences before merging. Once sample sequences have been identified into sequencing error removed reads, either OTUs or ISVs, these sequences can be searched for the formation of chimera sequences which may then be removed from the dataset. Finally taxonomy can be assigned to the remaining sequences using UCLUST or a RDP classifier, QIIME and DADA2 defaults respectively, with reference to one of a variety of databases.

Comparisons between the two pipelines based on sequence retention (Appendix 2), repeatability (Appendix 3) and the ability to assign taxonomy led to the selection of DADA2 defaults and QIIME forward reads as the methods used for analysis. Greengenes[269,270] was selected as the reference database as it assigned the greatest proportion of sequences to taxonomy. Chimeras were removed from both pipelines to minimise inter-replicate variation.

[C.2.1] Utility of assigned taxonomic units

Assigned taxonomic units (ATUs) are a mechanism of clustering sequences into groups of similarity – much like the formation of OTUs. In the case of OTUs similarity is determined based on sequence identity. ATUs should work in a similar way: sequences that are closely related are assigned to the same ATU, with shared taxonomy becoming increasingly reduced as the sequences diverge. ATUs therefore provide a coarser level of grouping based on a supposed hereditary relationship that can then be assigned functional properties.

'Unassigned' ATUs might be allocated based on sequence novelty (and thus absence from the reference database) or alternatively as a result of intermediate status of a sequence

between two ATUs. The utility of ATU groupings can therefore be considered by overlaying ATU identity with sequence phylogenies (i.e. how well ATUs approximate sequence relatedness).

Maximum likelihood phylogenies were built for both the QIIME and DADA2 datasets. The spread of 'unassigned' sequences throughout the DADA2 phylogeny suggests these sequences are intermediates between two or more reference sequences. The distribution of ATUs within both the DADA2 and QIIME phylogenies is best visualised in sections, so three random subtrees were selected from each phylogeny (subtree locations shown in SI Figure 2.2).

Although there are subsections in the phylogenies of both the DADA2 and QIIME datasets (Figure 2.3 and SI Figure 2.3) where a particular ATU dominates, there are essentially no clades within the phylogeny made up of a single ATU or a single ATU in combination with unassigned sequences. As phylogenetic rankings increase (e.g. at order level) the relatedness of sequences overlays better with the assigned taxonomies. However, even when ATU overlays of phylogeny are considered at the coarsest phylogenetic ranking (i.e. phylum) there are clear interruptions of the association of relatedness and ATU allocation (SI Figure 2.4). Thus, grouping sequences into ATUs does not reflect sequence relatedness.

Thus, although it may be useful to consider ATUs as a mechanism of discerning functional properties of interesting sequences, grouping these sequences for comparison may artificially modify the differences between samples.

[C.3] Identification of potential biases within the sampling process

The development of a method for fine-scale analysis of the gut tissue requires that the sampling process does not introduce artificial bias. These biases might emerge within a sample's diversity estimates or the rate at which RNA or DNA within a sample becomes degraded. Sectioning of the grid takes time, and samples are necessarily added to RNAlater at different time points from complete tissue extraction. Degradation of sample RNA or DNA could result in reduced diversity in the latterly stored samples. Moreover, diversity estimates could be biased if cells are transferred between segments. In this case we would expect to see an

increase in diversity over the course of sample collection, which would be reflected in all of the studied populations.

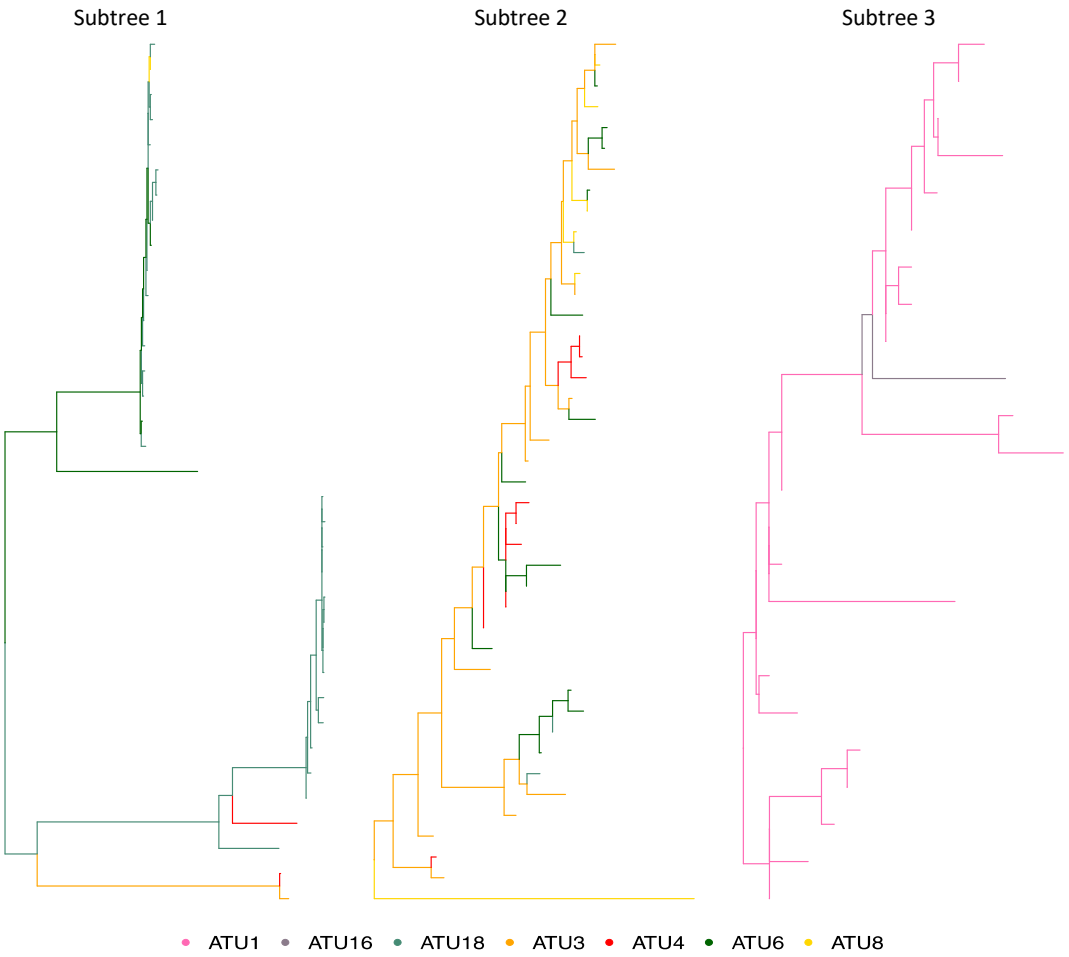


Figure 2.3 – Subtrees extracted from the approximate maximum likelihood tree for the DADA2 dataset. The ISV tree is produced from the merged, chimera removed, DADA2 dataset. The branches are coloured by the assigned taxonomic unit (ATU) to which each ISV was assigned using the Greengenes database. ATU1 (Clostridia; Clostridiales; Lachnospiraceae; Ruminococcus), ATU16 (Clostridia; Clostridiales; Lachnospiraceae), ATU18 (Alphaproteobacteria; Rickettsiales; mitochondria), and ATU8 (Alphaproteobacteria). ATU3 was unable to be assigned past the kingdom of Bacteria, ATU6 was unable to be assigned past phylum level as Proteobacteria and ATU4 was completely unassigned.

Fine-scale segments were taken from the caecum as in B.1. There was no evidence of sample degradation resulting from any function of sampling order (Appendix 4), nor was there any evidence of cell transfer between segments as a result of the sampling method.

[C.4] Identification of potential biases in sample amplification and barcoding

PCR was used to amplify sample DNA for sequencing (see B.5). Barcodes were added to these amplicons to facilitate sample identification (B.5). As the barcodes are intrinsically non-

identical, it is plausible that different barcodes might be better at amplifying the target DNA than others or that different barcodes are better able to amplify some specific sequences than others.

Different barcodes were better able to amplify low starting concentrations of cDNA. Barcode sensitivity (amplification ability) classifications were developed to reflect the lowest sample concentration at which a barcode was able to produce a PCR product (SI Table 2.1). A barcode's sensitivity was dependent upon the reverse primer with which it was paired. Despite some variable barcode sensitivity, there was no association between primer barcode and sequencing depth in any studied population.

There was no significant effect of barcode on any diversity estimate in any population studied. There was, however, a significant correlation between the sensitivity of the barcode and the diversity of its samples (SI Table 2.2), though this effect appeared to be an artefact of the deliberate association of low starting concentration samples with high sensitivity primer barcodes. Subsequent samples (i.e. Chapter 4 and Chapter 5 samples) were randomly assigned barcodes between amplification groups 3 and 4. Any intrinsic effect of barcode was further minimised by using multiple barcodes per sample in the triplicate study design.

[C.5] Comparisons accounting for incomplete sampling

Comparing samples of different sampling depths is challenging for every discipline, particularly with respect to diversity comparisons. It is possible to account for relative differences between differentially sampled samples where comparable sample coverage has been achieved[291]. Sample coverage in the microbial datasets was not significantly different between any of the segments (Figure 2.4), however variation in sample coverage in the immune datasets must be taken into consideration when comparing sample diversity estimates. The variation in sample coverage in the immune datasets may simply reflect that these datasets have lower sample coverage, or alternatively that these populations are more skewed to rarity than the microbial data.

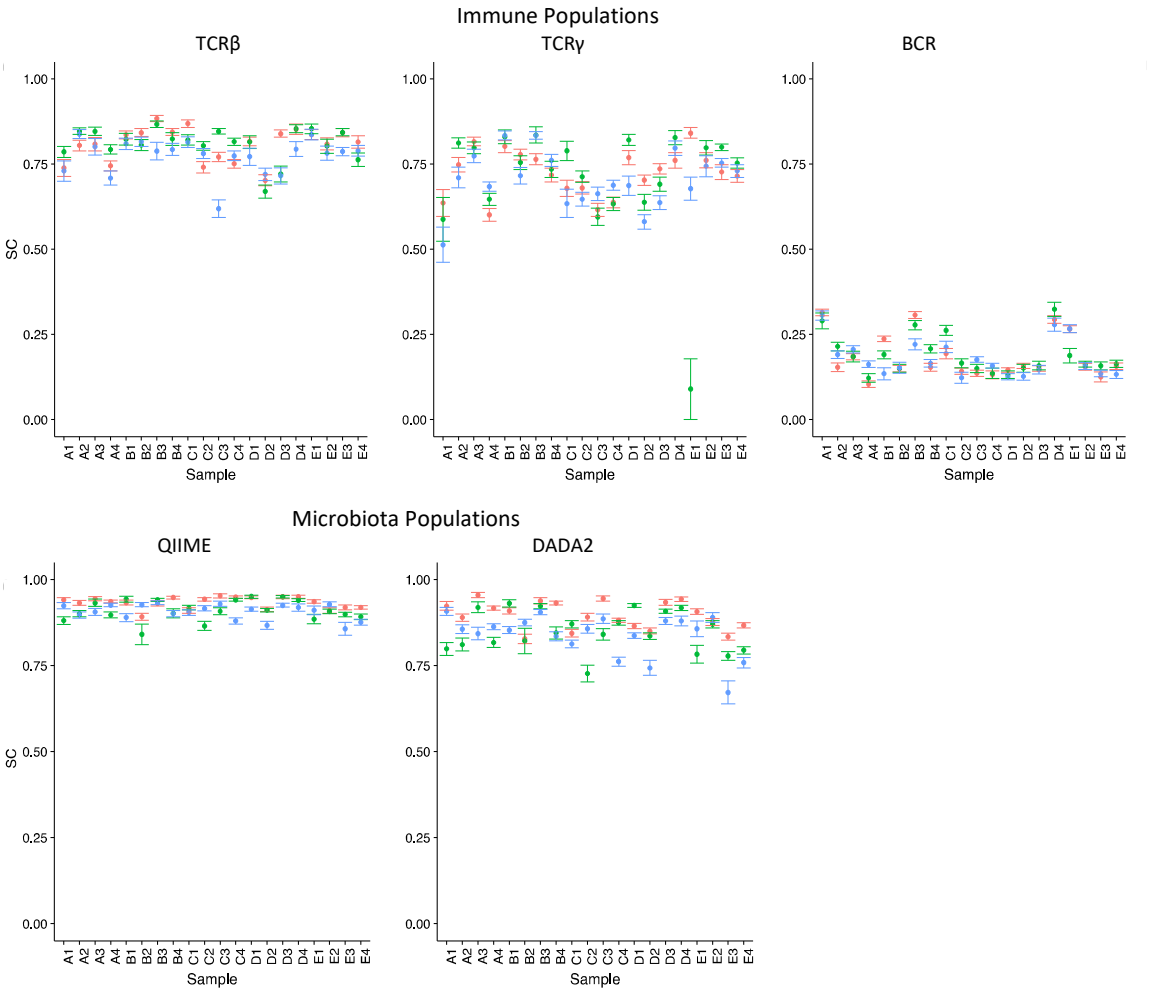


Figure 2.4 – Sample coverage estimates (SC), with associated standard error, for each population and replicate. There was significant variation between the coverage estimates in each of the immune populations (p-value: 0.002, 0.04 and $8.0e^{-12}$, R^2 adjusted: 0.381, 0.236 and 0.800 and for TCRβ, TCRγ and BCR populations respectively). There was no variation in the coverage estimates between segments in the microbial dataset, processed by either QIIME or DADA2. Samples named by grid location (see Figure 2.1), replicates 1-3 shown in red, blue and green respectively.

Sample coverage was very low in the B-cell dataset, despite more extensive sampling than any other population. Low sample coverage can mislead sample comparisons; decrease reproducibility across technical replicates; and increase the noise in the dataset. However the structure of the B-cell population, in conjunction with the clustering necessary to associate naïve B-cell descendants, could invalidate the assumption of an underlying log-normal distribution.

This study encountered an additional challenge: high sample coverage of low quantity samples. Unlike ecological samples, where species are removed from an environment, counted

and returned, the replicates within this dataset have been created through what is in essence sampling without replacement. Sample coverage estimates rely upon an assumption of sampling with replacement[291]. If the population is large enough, the results of sampling without replacement approach those of sampling with replacement. However, given the small size of the segments, the number of cells within each is unlikely to approach this threshold. In this case the sample coverage should be treated as an estimate of the lower bound[292], and thus comparisons between samples of equivalent sample coverage are still possible.

Comparisons of diversity between samples were further facilitated through the triplicated design of this study; differences between samples that are greater than differences between the technical replicates can surmount the inherent problems with incomplete sampling.

[D] Results

The results presented here are from a single individual, and provide a preliminary dataset to investigate whether there is a difference between large-scale (formed by pooling all the small-scale segments) and small-scale segments when interrogating the microbiota and immune repertoire in L6 chickens. This preliminary dataset is further used to assess whether there is evidence of spatial variation on a fine-scale as suggested by the modelling in Chapter 1. The results and analyses developed here can be further used to determine whether the same trends are present in other wild-type birds (Chapter 4).

[D.1] Non-functional and contaminating sequences

During T and B-cell receptor formation gene segments are somatically rearranged to create novel combinations (see General Introduction C.3). In the process of this recombination nucleotide insertions/deletions can result in the formation of stop codons, thus rendering the receptor non-functional. These non-functional receptors should be removed within the thymus[293] or Bursa of Fabricius[294,295]. However, non-functional receptor sequences (i.e. non-protein producing), amplified from cDNA, are found in these gastrointestinal samples.

13-15% of the TCRβ and TCRγ datasets are non-functional sequences, 12-14% of the total clones in these populations. Similarly, 24-34% of BCR sequences (30-38% clones) are non-functional. In all the immune populations, non-functional clones are predominantly singletons, though the TCR datasets do contain a few abundant non-functional receptors (*c.* 600 counts in the TCRβ and *c.* 100 counts in the TCRγ populations).

Despite both 16S analysis pathways using the same initial sequence data, there was a significant difference in the proportion of contaminant sequences and species produced by each pipeline (paired t-test; p-value $2.89e^{-19}$). Specifically the QIIME-processed data suffers from higher levels of contamination than the DADA2 population. This enhanced contamination may be a result of more sequences passing its quality filters, or of unique sequences being clustered together and thus culminating in the appearance of non-trivial (see B.5) contamination in the water controls.

All of the results presented here are robust to the inclusion of non-functional or contaminating sequences, though only the data excluding non-functional and contaminating sequences is included.

[D.2] Population structures

The structure of a population can provide information about the dynamics of its components. For example, in a naïve adaptive immune repertoire the necessity of a diverse repertoire to facilitate pathogen recognition can be seen in an abundance of rare individuals. Conversely a responding repertoire would be expected to have a higher contribution from a smaller number of proliferating (abundant) clones.

[D.2.1] Large-scale variation in population structures

As shown previously, the gastrointestinal microbiota is dominated by relatively few, highly abundant taxa[83,84,296–298], which was recapitulated here using both processing pipelines (Figure 2.5).

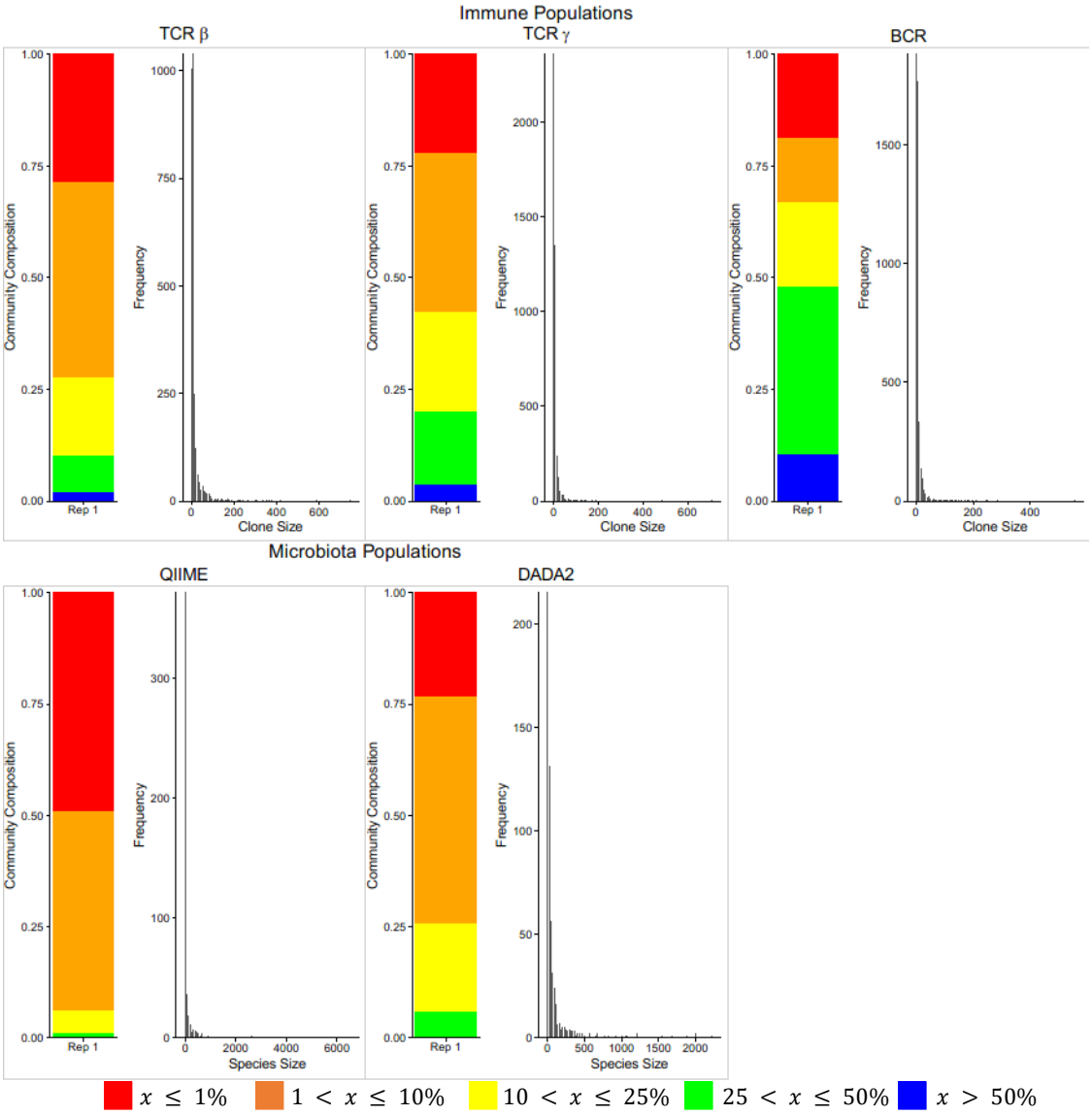


Figure 2.5 – The population structures of the immune and microbiota populations within a L6 6-week old chicken right caeca. Clonal homeostasis plots show the proportion of the population made up by species ranked by abundance (e.g. red group is the proportion of the population comprised of the top 1% most abundant species), whilst the histograms show the distribution of species sizes (excluding singletons). The microbiota sequence data was processed using either QIIME or DADA2. Contaminant and non-functional sequences have been removed prior to analysis.

There were significant differences between the population structures of the QIIME and DADA2 processed data (Figure 2.5, SI Table 2.3), particularly with respect to the proportional contribution of abundant, moderate and moderately rare clones (Figure 2.5, red, yellow and green respectively). The QIIME generated data showed an exaggerated presence of abundant individuals with *c.* 50% of the population composed of the top 1% of OTUs, whilst the top 1% of ISVs comprise only *c.* 25% of the DADA2 population. The difference between the pipeline

structures results from the existence of several extremely large OTUs in the QIIME dataset (4000-6000 sequence counts), which were absent from the DADA2 dataset. These highly abundant OTUs are a result of OTU clustering; grouping of similar sequences (by QIIME) that DADA2 has identified as distinct.

The differences between the QIIME and DADA2 produced population structures are robust to the inclusion of the contaminant sequences and whether percentage or numerical (top 10) ranking of clonal contribution of the clones was considered.

All the immune populations were skewed toward low abundance individuals, as would be expected of non-responding immune repertoires (Figure 2.5). The BCR repertoire was particularly dominated by rare clones, skewed largely because of the high proportion of singletons (70.6±1.0%). The increased proportional contribution of rare clones to the BCR population structure may be linked to the low coverage estimates of this dataset (Figure 2.4), though the repeatability of this pattern suggests this is not the case (Chapter 4 C.2).

The BCR population was much more clonally diverse than either of the two TCR populations (BCR: 63233±18113 unique clones, TCRβ: 6715±917 and TCRγ: 9702±70), which inherently reduces the proportional contribution of the abundant clones to the BCR population. However, the contribution of large clones to the BCR population was less than that of either TCR population even after accounting for this variation in clonal diversity. Conversely the TCRβ population comprised a much higher proportion of expanded clones than either of the other immune populations (Figure 2.5).

[D.2.2] Large-scale TCR CDR3 length distributions

Abundant individuals within the immune populations are formed by the proliferation of clones responding to antigen. Traditionally, the presence of these clonal expansions was identified using spectratypes, with non-normal distributions of sequence length indicating the existence of an expanding clone. It is possible to produce spectratypes of the TCR populations by simply grouping clones by sequence length. The clustering of the BCR population precludes this analysis for this cell type.

Both large-scale TCR spectratypes were effectively normally distributed (summed least squares TCRβ 0.09 and TCRγ 0.03, Figure 2.6), suggesting an absence of any large clonal expansions within the T-cell receptor repertoires.

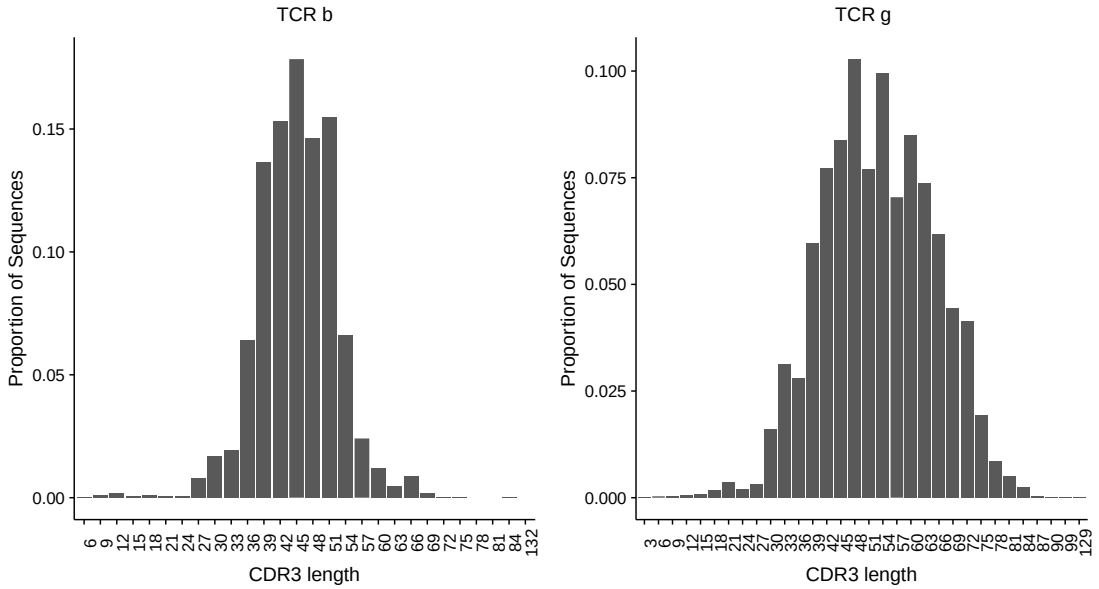


Figure 2.6 – T-cell receptor spectratypes (CDR3 length distributions) of a L6 6-week old chicken right caeca. CDR3 length refers to the length of the hypervariable region of the TCR sequence.

[D.2.3] Small-scale TCR spectratypes

The clearest way to assess whether the small-scale segment TCR populations are any different from their large-scale counter parts is to compare their spectratypes. Although the large-scale populations approximate a normal distribution (Figure 2.6), the small-scale segments are more variable in their distributions (e.g. the TCRβ distributions Figure 2.7) and fit less well to a normal distribution in both TCR datasets.

As is apparent in Figure 2.7, some sites deviated far more from a normal distribution than others (Figure 2.8). For example, segment A1 was extremely deviated from its comparative normal distribution for both the TCRβ and TCRγ populations, whilst segment C3 is more normally distributed. The TCRβ population spectratypes were always less normally distributed than those of the TCRγ population.

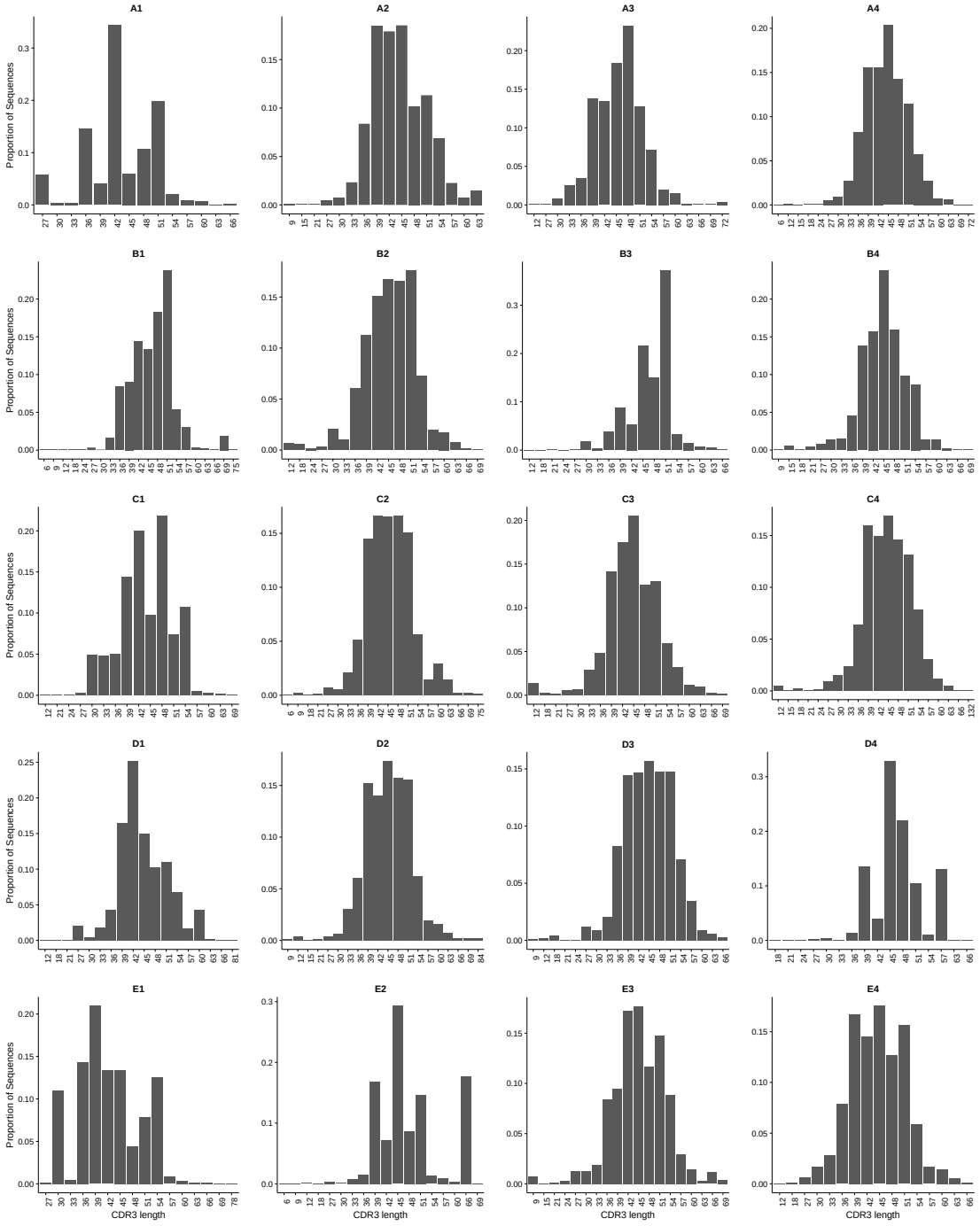


Figure 2.7 – Spectratypes depicting the CDR3 length distribution of the TCRβ repertoire across 2 mm^2 segments in the right caeca of a L6 6-week old chicken. CDR3 length refers to the length of the hypervariable region of the TCR sequence.

Using sums of least squares it was possible to compare the distributions of the spectratypes across the grid as a function of their distance (Figure 2.8). There was an essentially neutral correlation between difference in spectratype shape and their Euclidean distance in both TCR populations, which is robust to the inclusion of non-functional sequences.

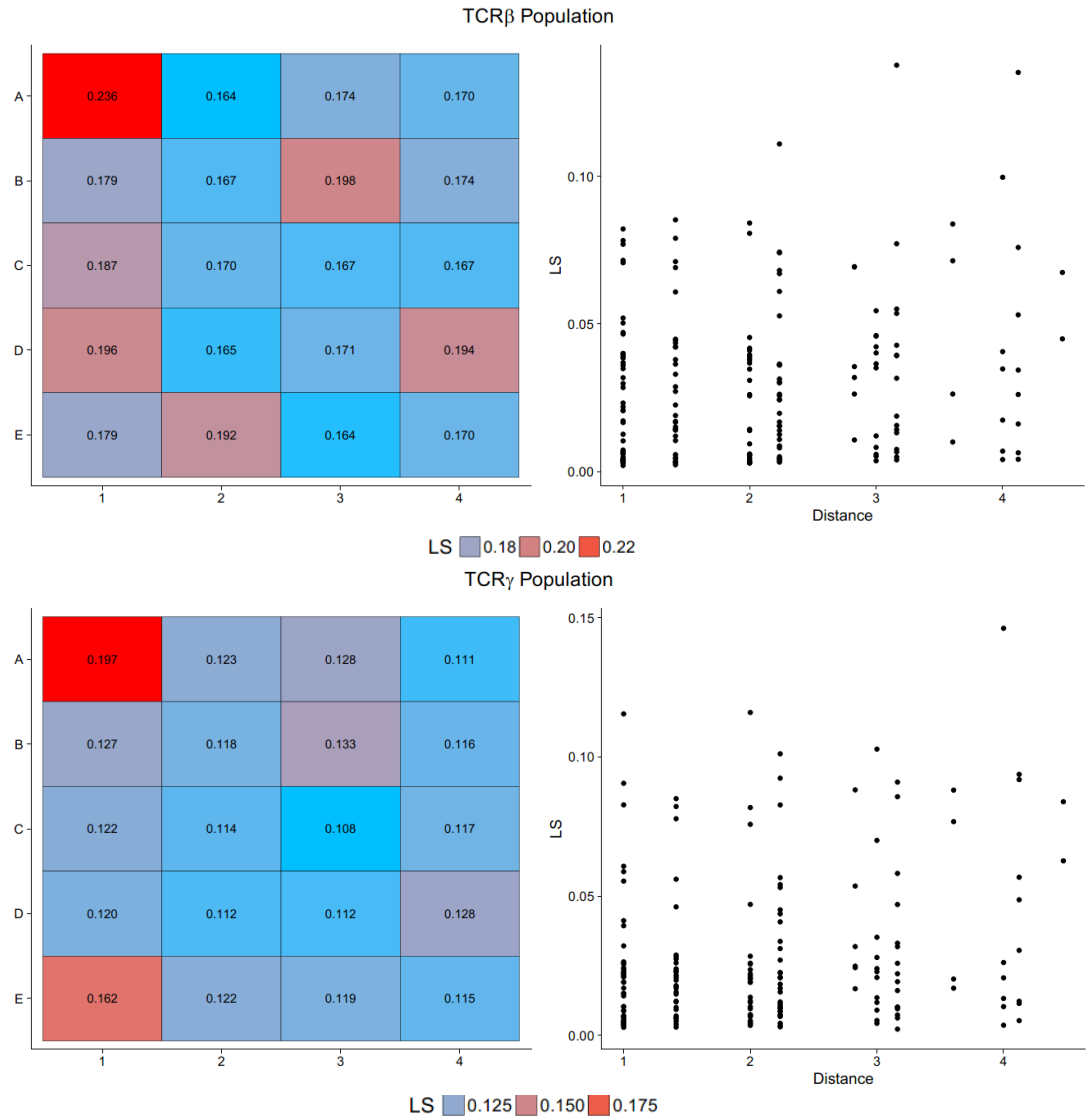


Figure 2.8 – Similarity of T-cell receptor (TCR) spectraypes to both their respective normal distributions and the other spectratypes within the grid as a function of distance, for both TCR populations. (Left) Heatmaps of the sum of least squares difference between each segment's TCRβ (upper) and TCRγ (lower) spectratypes and the normal distribution. Each normal distribution was fragmented to include only functional comparisons (lengths of a factor of three), and constrained by the mode, upper, and lower values of the observed spectratype distribution (see B.8). (Right) The difference between two spectratype distributions (LS) as a function of the Euclidean distance between the two sites (Distance). In both TCR populations there was a significant correlation between two segments spectratype shapes and the distance between the segments (TCRβ: p-value $2.28e^{-02}$, R^2 adjusted 0.0221; TCRγ: p-value $2.83e^{-03}$, R^2 adjusted 0.0414) though the slope of this association was essentially neutral (TCRβ 0.00455, TCRγ 0.00620).

[D.2.4] Small-scale variation in population structures

The population structures of all immune datasets varied between the segments (SI Table 2.4). Some small-scale segments were dominated by rare clones (e.g. in the BCR

population segments B1 and D1), whilst others were more evenly distributed (e.g. BCR segments D4 and E1). The variation between segments was robust to whether the top percentage (e.g. 1%) or count (e.g. 10) of abundant clones were considered. The size of the largest clones considered was variable between sites (e.g. *c.*10 – 70 sequence counts in the BCR dataset), which may account for some of the change in population structure. The variation between sites in population structure may, however, result from variable starting numbers of immune cells in each segment or sample coverage (SI Table 2.5).

The origin of the inter-segment variation in population structures appears to vary based on the dataset considered, although there are some consistently different segments (SI Figure 2.4). In the BCR dataset segments A1, B3, and D4 are significantly different, whilst in the TCRβ dataset differences stem from segments A4, B3, C1, D2, and D4. TCRγ dataset differences are based around segments B1, B3, and D4. There was no correlation between a segment's CDR3 length distribution deviation from normality and the proportional contribution of abundant clones within either TCR dataset.

There were significant differences in the proportional contribution of the moderately abundant and moderately rare groups within the DADA2-processed microbiota datasets, and in all but the rarest group in the QIIME-processed microbiota data, across the grid (Figure 2.9, SI Table 2.6 and SI Figure 2.5). The absence of segmental variation when contaminants were included in the dataset may reflect the increased noise resulting from contaminant inclusion.

The presence of segmental variation in population structure across segments in the immune and microbiota datasets suggest small-scale heterogeneity in the gastrointestinal populations.

[D.3] Population compositions

The composition of a population can be measured in many ways. Entropy estimates, such as species richness, Shannon, and Simpson's indices, can provide information about the number of community members or the evenness of their distributions. In microbiome analyses, the functional capacity of a community can be considered by assigning taxonomy to the

individual members where possible. In adaptive immune analyses, gene segment usage can be assessed for biases.

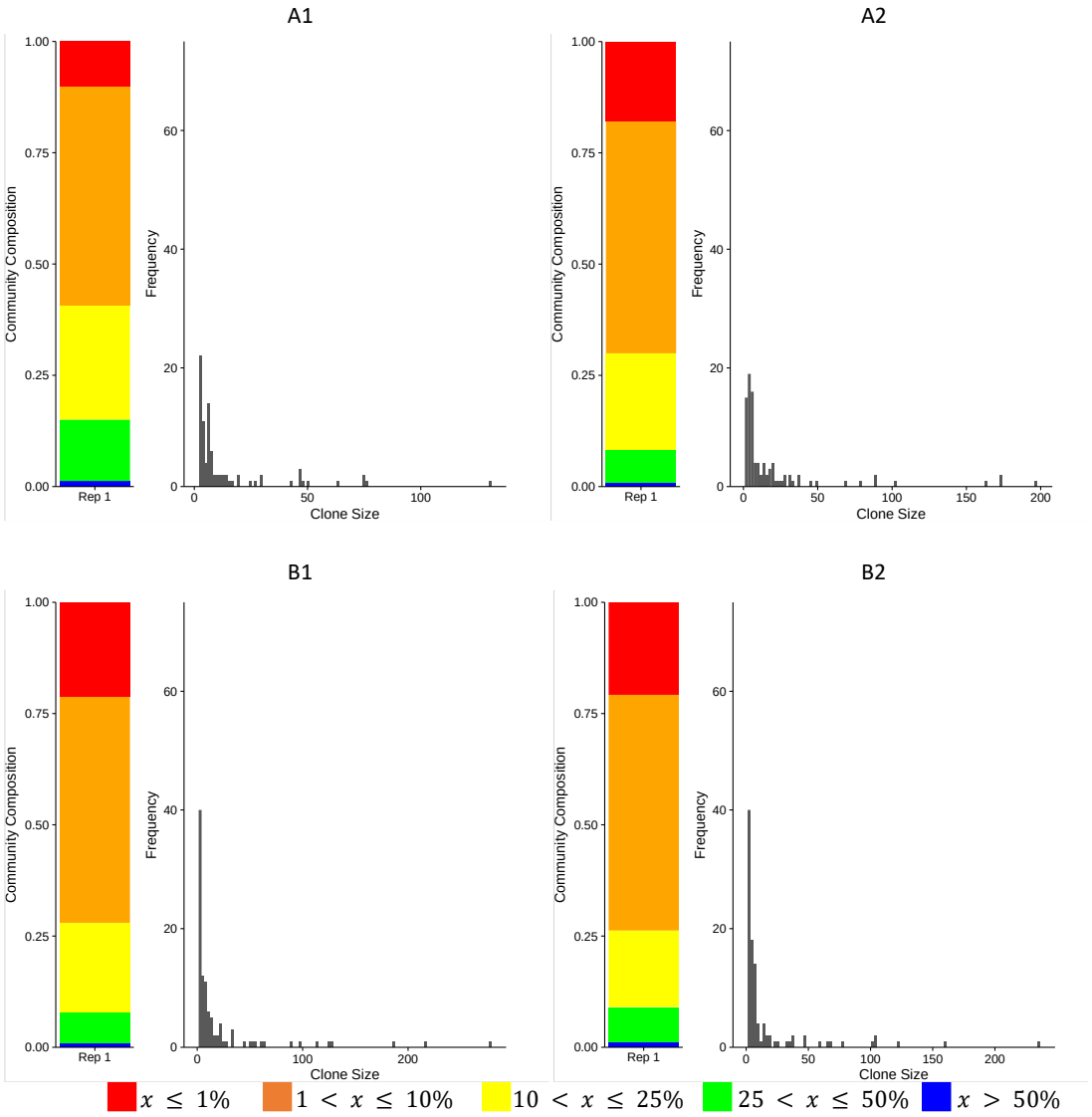


Figure 2.9 – Example segments showing the microbiota population structure of a L6 6-week old chicken right caeca, separated into 4 mm^2 segments. The sequence data has been processed using the QIIME pipeline, and contaminant sequences removed prior to analysis. Clonal homeostasis plots show the proportion of the population made up by species ranked by abundance, whilst the histograms show the distribution of species sizes (excluding singletons). Full figure can be seen in SI Figure 2.5.

In this section the composition of the large-scale tissue fragment will be considered in terms of its diversity estimates, assigned taxonomy and gene segment usage. The profiles produced will then be compared to those of the individual, small-scale segments and the segments themselves will be compared to examine evidence of spatial variation.

[D.3.1] Diversity between large-scale populations

Comparing community composition is a classic ecological problem[299] and many metrics exist which attempt such comparisons. Here three metrics are considered: species richness, Shannon's entropy, and Simpson's diversity index.

The large-scale microbiota populations, regardless of processing method, were significantly less diverse than any of the immune populations (Figure 2.10). Although the QIIME-processed microbiota population showed increased species richness relative to the DADA2-processed dataset, both β-diversity estimates (Shannon's and Simpson's indices) were reduced in the OTU population (Figure 2.10). The increased species richness in the QIIME processed dataset could be caused by the increased number of sequences retrieved for analysis using this pipeline. This is consistent with the decreased diversity using β-diversity estimates which account for sampling effort and previously reported incidences of OTU clustering artificially increasing species richness estimates[300].

All three immune populations had significantly different large-scale diversities. As suggested by the population structures, the IgA-producing B-cell population had a much greater diversity than either of the T-cell populations, regardless of the metric used (Figure 2.10). The TCRβ population showed the lowest diversity of any adaptive immune population.

[D.3.2] Taxonomic composition changes between processing pipelines

As is becoming unsurprising, the taxonomic profile of the microbiota varied depending on processing pipeline (Figure 2.11). The variation between the two pipelines was apparent even at the coarsest level of assignment (Figure 2.11), with QIIME able to assign a higher number of species (Appendix 2).

The differential taxonomic assignment between pipelines was not restricted to rare species –none of the top ten most abundant OTUs or ISVs shared the same assigned taxonomy. However the sequences comprising the top OTUs or ISVs are not the same: using a sequence-by-sequence comparison the assigned taxonomy was the same.

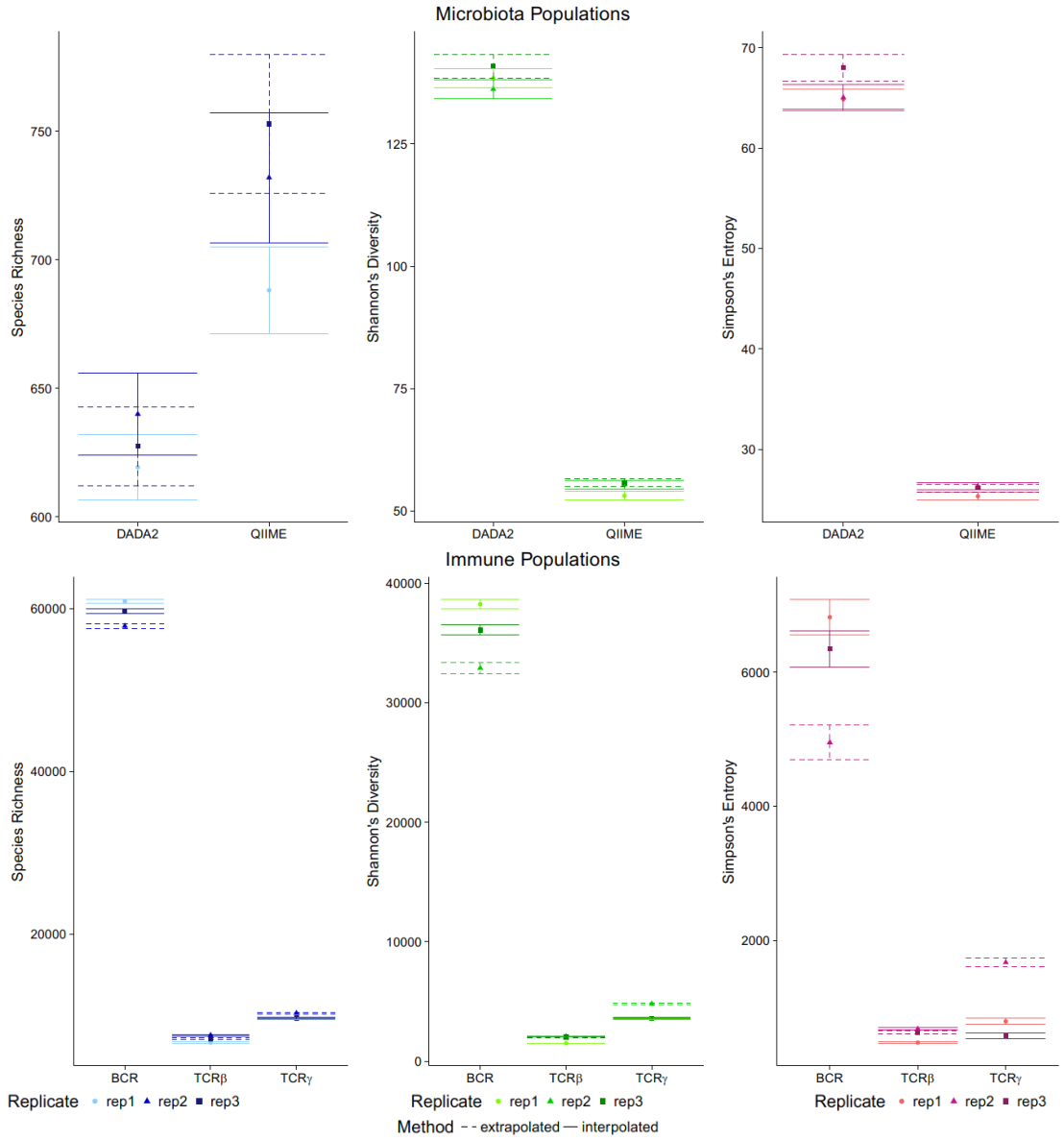


Figure 2.10 – Diversity estimates the functional B-cell receptor (BCR) and T-cell receptor (TCRβ and TCRγ), and contaminant-free microbiota, populations. Microbiota populations are processed using either DADA2 or QIIME. All estimates are taken at the same (median of the population) sampling depth, with the 95% confidence intervals of the estimate indicated as error bars. Dashed error bars indicate the sample required extrapolation to reach the median sampling depth, solid lines indicate the median sampling depth was either observed or interpolated.

Both microbial profiles were dominated by the Firmicutes phylum, though the predominance of Firmicutes was almost balanced by an abundance of Proteobacteria in the DADA2 dataset (Figure 2.11). Both pipelines produced microbiota compositions which were consistent with other processed chicken caecal samples of a similar age[83,84,296–298], though the proportion of Firmicutes in the QIIME dataset was higher than previously reported.

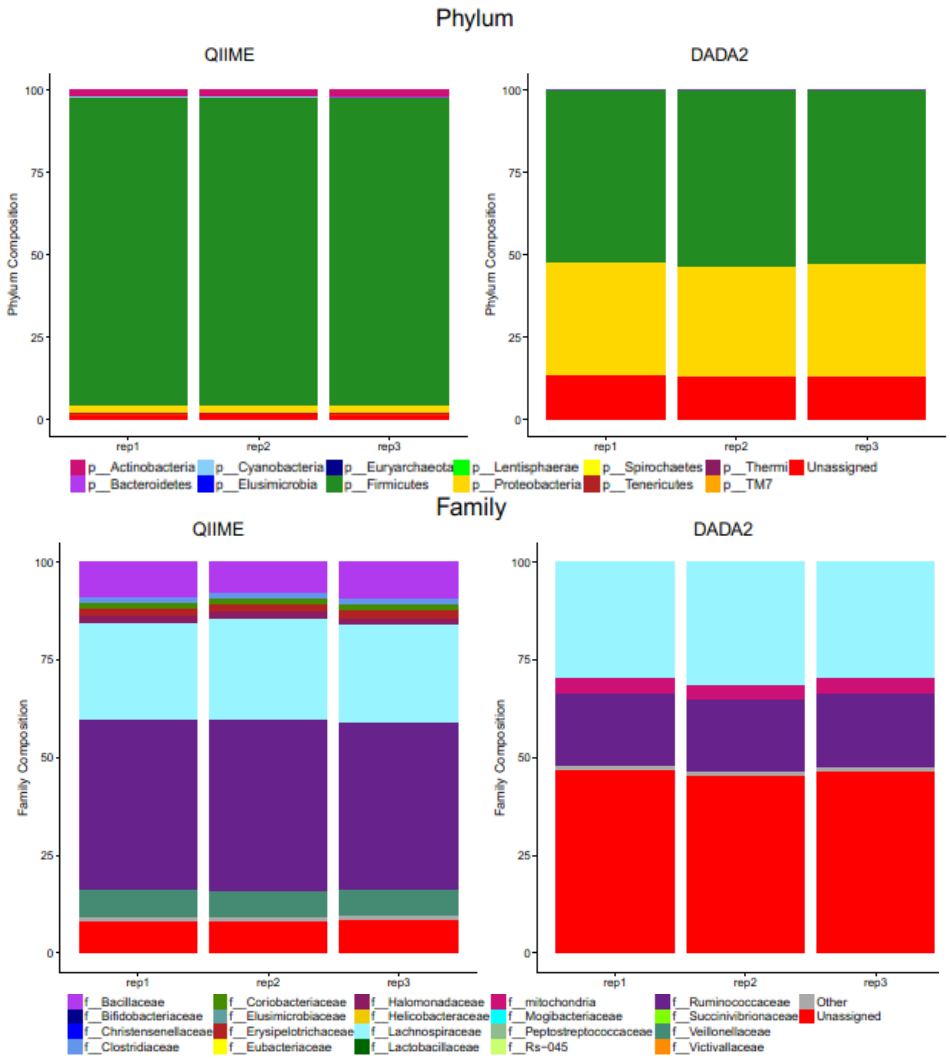


Figure 2.11 – Taxonomic assignment of large-scale microbiota samples, processed according to either QIIME or DADA2. Contaminant sequences were removed before analysis. Sequences were grouped according to the assigned taxonomy of their phylum or family. Family assignments that contributed less than 1% of the sequences were compressed into a single 'other' category (SI Table 2.7).

[D.3.3] Immune cell profiles

Previous work has demonstrated a diverse TCR repertoire in both TCRβ V-gene family subsets[301]. Here, comparison of Vβ-gene usage across the full TCRβ repertoire showed a bias towards Vβ1 usage in the large-scale caecal population (Figure 2.12) which was consistent with the reported scarcity of Vβ2 T-cells in avian mucosal tissues[302,303]. Within the Vβ1-associated Jβs there was a significant difference in Jβ gene usage with Jβ2 segment usage reduced relative to Jβ3 and Jβ4, whilst in the Vβ2-associated Jβs, Jβ3 is preferentially used (Table 2.4).

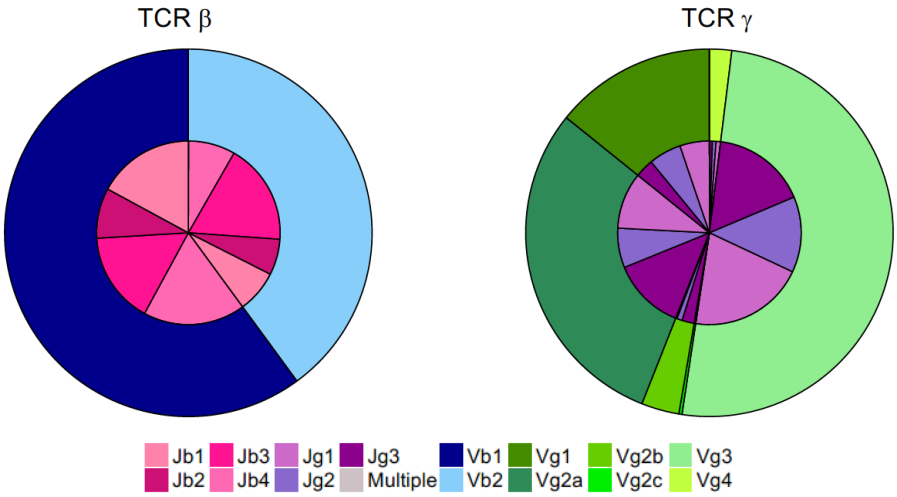


Figure 2.12 – T-cell receptor V-gene and J-gene usage of large-scale samples for both TCR populations, as indicated. V-gene usage in the TCRβ and TCRγ populations is displayed on the outer ring of the pie charts; J-usage associated with each V-gene is indicated within the V-gene ring. Gene usage defined as 'multiple' refers to sequences that were assigned, by BLAST, to more than one V-gene or J-gene. Gene segment assignment was achieved through BLAST comparison to an internal TCRβ or TCRγ gene dictionary.

To date there is no reported TCRγ repertoire in chickens. The TCRγ repertoire presented here was dominated by Vγ3 expression, with moderate Vγ1 and Vγ2a expression and restricted usage of the other Vγ's. Within the Vγ2a and Vγ3 associated Jγs there was a significantly reduced usage of Vγ2 segments, whilst Vγ2b segments were preferentially associated with Jγ3 segments (Table 2.4).

Table 2.4 – The variation in V-associated J-segment usage within the T-cell receptor β and γ populations. Differential J usage was assessed within each V family using a linear model and the J-segments whose proportional usage was different identified using Tukey's HSD test. All values are given to three significant figures.

Population	V	P-value	R^2 Adjusted	Different J
TCRβ	Vβ1	2.01e-06	0.962	Jβ2-Jβ3/Jβ4
	Vβ2	4.03e-10	0.987	Jβ1-Jβ3/Jβ4 Jβ2-Jβ3/Jβ4 Jβ3-Jβ4
TCRγ	Vγ1	4.63e-04	0.837	*NA*
	Vγ2a	8.17e-07	0.967	Jγ2-Jγ1/Jγ3
	Vγ2b	2.33e-05	0.962	Jγ3-Jγ1/Jγ2
	Vγ2c	3.60e-02	0.560	*NA*
	Vγ3	1.99e-07	0.977	Jγ2-Jγ1/Jγ3
	Vγ4	0.297	0.110	*NA*

As the BCR repertoire in chickens is formed through gene conversion[304,305], it was not possible to assess gene segment usage to form a BCR profile, although amino-acid usage could

be compared. Looking at the whole population of BCR sequences, it was immediately apparent that amino acid usage was extremely conserved from amino acid position 75 (SI Figure 2.6). The start of the conserved region in the IgA heavy chain can be seen from position 89, with the sequence 'ASASR'[306]. The further 14 consistent amino acids before the conserved region form part of the reported J_H segment[306]. There was remarkably little variability within the J-segment both in this study and within previously reported studies[306].

As expected there was a large diversity in amino acid usage across the BCR CDR3 region (SI Figure 2.6). Amino acid usage did appear more biased at particular amino acids however, for example towards leucine at amino acid positions 44-47, and towards alanine and glycine at positions 72 and 73, respectively.

[D.3.4] Estimating diversity in small-scale segments

Variation in the community composition of large tissue segments across the gastrointestinal tract is apparent from repertoire comparisons involving TCRβ populations[301,307]. Much smaller-scale tissue comparisons are possible through the grid structure used here.

There was a significant difference in community diversity between segments in all the populations in every diversity metric (SI Table 2.8). The segments driving the variation in diversity estimate varied between populations, though there were some consistently distinct segments (e.g. segment C4 Figure 2.13), with more distinct segments within the immune populations than in either microbiota population.

As expected, given the association between cell count and sample coverage, there was an association between the species richness estimates and starting cell count within the QIIME datasets (Figure 2.14 D and E panel 0). There was no association, however, between the starting cell count and either the Shannon's or Simpson's diversity estimates (Figure 2.14 D and E panels 1 and 2). The difference between sites seen by comparing the estimates is not therefore simply an artefact of variation in the abundance of the populations found in each of the segments.

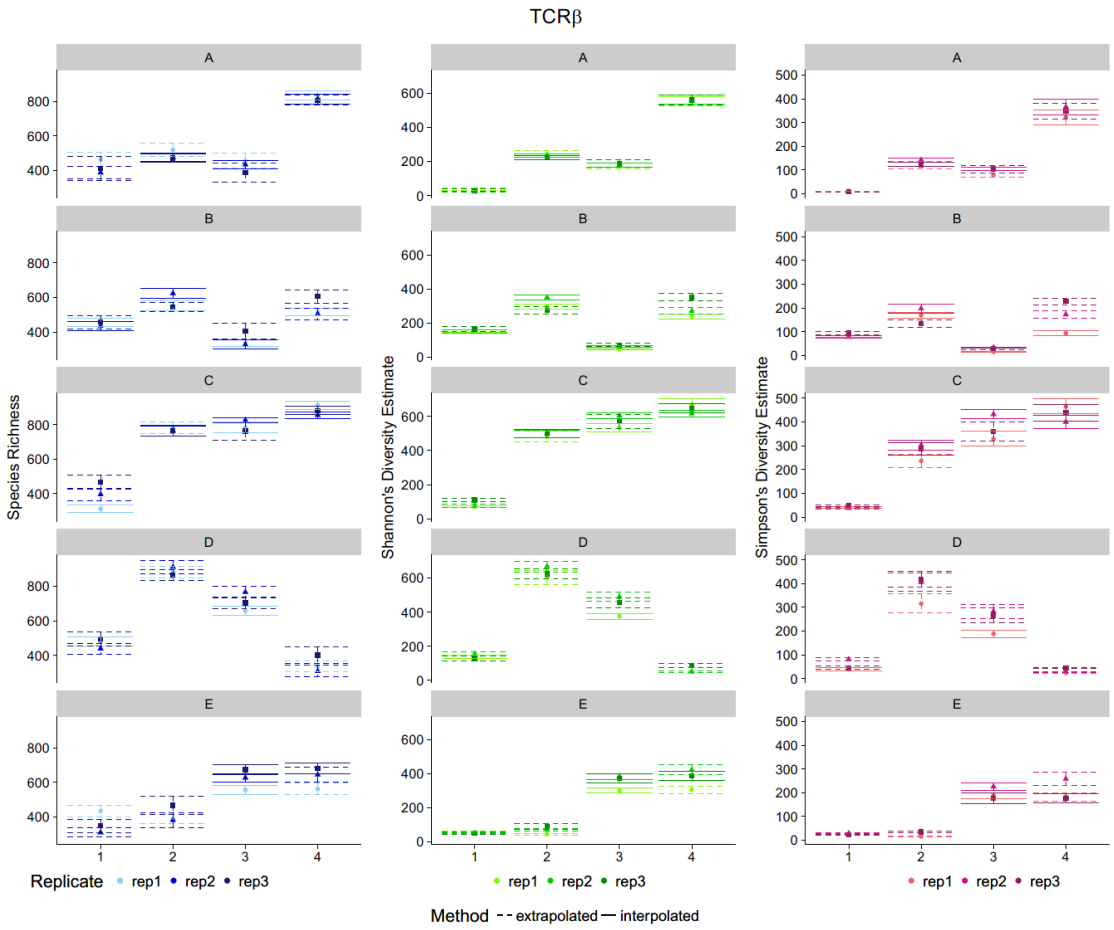


Figure 2.13 – Diversity estimates for each segment within the T-cell receptor β population; species richness estimates (left), Shannon's diversity (middle) and Simpson's entropy (right). All estimates are taken at the median sampling depth, with the 95% confidence intervals of the estimate indicated as error bars. Dashed error bars indicate the sample required extrapolation to reach the median sampling depth, solid lines indicate the median sampling depth was interpolated. The results for each segment are indexed by their column (letter facets) and row (x-axis).

[D.3.5] ATU profile and TCR gene usage in small-scale segments

Despite ATU groupings being a poor substitute for sequence similarity, disparity between segments in terms of taxonomy is a proxy for of community variation at the 4 mm^2 scale. There was significant variation in the proportional contribution of the Proteobacteria phylum between the grid segments regardless of which data processing method was used (Table 2.5). There was additional variation in the Actinobacteria, Cyanobacteria, Firmicutes and Tenericutes phyla between segments in the QIIME processed microbiota population.

Finer taxonomic scale interrogation showed an increased number of groups varied between segments (SI Table 2.9), though more families varied within the QIIME-processed

dataset than within the DADA2. The variation between segments was driven predominantly by a small number of segments within the grid (e.g. Table 2.5 segments A3 and E1) that showed a reduced proportional contribution of Ruminococcaceae and an expansion of the Veillonellaceae family. However the patterns of variation were not always consistent between the pipelines.

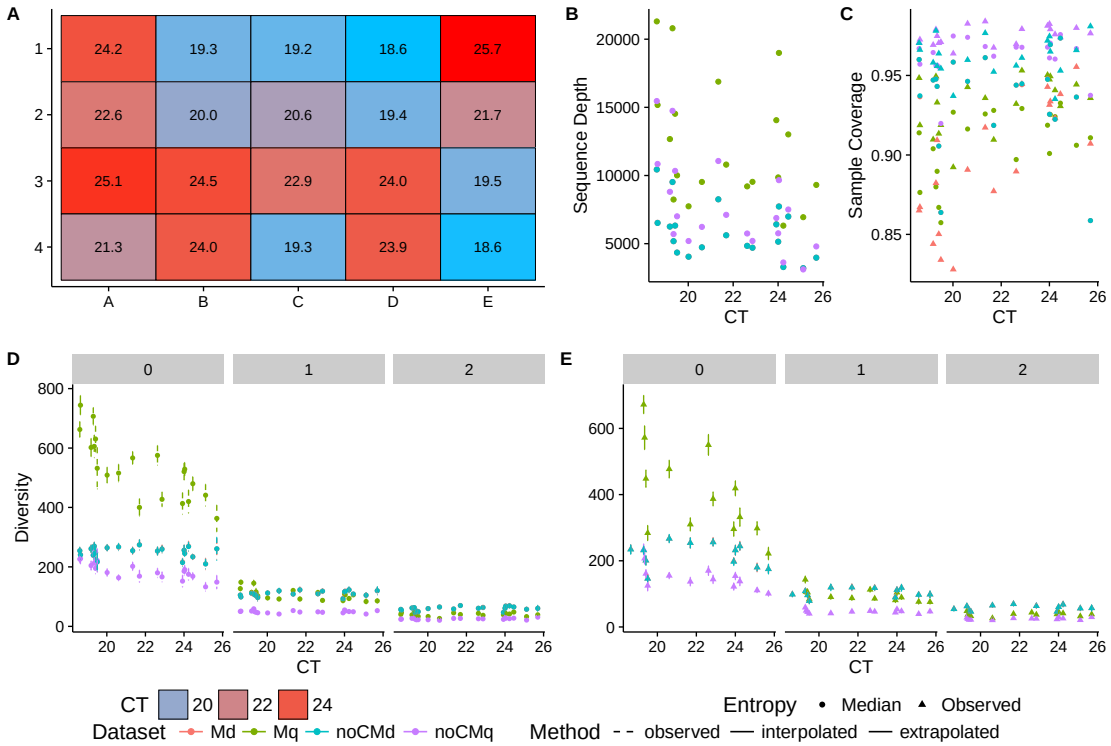


Figure 2.14 – Starting cell count variation within the right caeca microbiota of a L6 6-week old chicken, as assessed by 16S qPCR. Microbial data including entropy and sequence depth estimates are shown for both processing methods (QIIME and DADA2) and both with and without contaminating sequences. (A) Heatmap of the threshold cycle (CT) at which each sample could be identified by qPCR. Heatmap layout is the same as the grid shown in Figure 2.1, with the x-axis indicating a segment's column within the grid and the y-axis its row. (B) Association between CT value and sequence depth of a sample: there is no effect of CT on sequence depth. (C) Plot showing the association between CT and sample coverage. This correlation is only significant in the QIIME processed datasets. (D-E) Correlation between entropy estimates and CT values for the median sequence depth of all segments (D) and for the observed sequence depth of each segment (E). Entropy estimates are species richness (0), Shannon's diversity index (1) and Simpson's diversity index (2). The standard errors of the estimates are shown, the lines indicating whether the estimate has been observed, interpolated or extrapolated.

Comparing the TCR repertoire within each segment showed no change in V-gene usage between segments within the TCRβ population, though the proportion of Vγ3 gene usage

varied within the TCRγ population over space (p-value $4.19e^{-02}$, R^2 adjusted 0.0743). Within the TCRβ population, however, the usage of Vβ1 associated Jβ3 varied between segments (p-value $1.5e^{-03}$, R^2 adjusted 0.397), whilst there was no difference in any of the Vγ associated Jγs between segments.

Table 2.5 – The variation in the proportional contribution of different phyla between segments within the microbiota population, processed using either DADA2 or QIIME. The variation between segments was assessed using a linear model of the form lm(Prop~Location), and the segments which drove the difference identified using Tukey's HSD test. All values are given to three significant figures.

Processing Method	Phylum	P-value	R^2 Adjusted	Different Segments
DADA2	Proteobacteria	1.05e-06	0.740	A3-A4/B2/C1/C4/D1/D2/D3/E2/E3/E4 B2-C3 B3-B2/D2/E3/E4 D2-B4/C3/E1 E1-B2/E3
QIIME	Actinobacteria	2.84e-07	0.758	A3-A4/D1/D4/E2/E3/E4 B2-A4/C1/C4/D1/D4/E2/E3/E4 C2-D4/E2/E3 E1-A1/A2/A3/B1/B2/B3/B4/C2/C3/D2/D3
	Cyanobacteria	8.01e-03	0.540	D1-A1/A2/A3/B1/B2/B3/B4/C2/C3/D4/E1/E3
	Firmicutes	3.98e-08	0.784	B1-A3/B3/B4/D3 B4-C1/C4 D1-A1/A3/B3/B4/C3/D3 D2-A3/B3/B4/C3/D3 E1-B1/C1/C4/D1/D2/E3/E4 E3-A1/A2/B2/B3/C2/C3/D3 E4-A1/A3/B3/B4/D3
	Proteobacteria	1.14e-03	0.599	B3-D2/E3 B4-D2/E3/E4
	Tenericutes	1.27e-06	0.736	C3-B1/B2/C1/C4//D1/D2/E3 D3-B1/B2/C1/C2/C4/D1/D2/E1/E2/E3/E4

[D.4] Dissimilarity between segments

Diversity estimates, although informative about the evenness and complexity of a community, are not able to indicate whether species are shared between samples. Estimates of community similarity fall under the jurisdiction of another ecological metric: dissimilarity (or similarity).

Dissimilarity indices can be calculated in a variety of ways that account for compositional variation in different ways[308,309]. Dissimilarity indices can ignore, or account for, variation in species abundances. For example, Kulczynski's dissimilarity metric does not weight species by their overall abundance. Therefore estimates of Kulczynski's dissimilarity between

two sites will give more weight to rare species in determining the extent of site differences (as proportional changes in rare species will be inherently larger).

Three dissimilarity metrics were used to assess the average difference between segments in each population. These metrics were used to interrogate different aspects of the data, specifically the importance of rare, abundant and absent species (Kulczynski, Chi and Hellinger's dissimilarities respectively). The equations for calculating dissimilarity, and the packages used for the calculations, are presented in full in B.8.

As expected, given the diverse nature of a naïve adaptive immune repertoire and reported dominance of restricted species in the microbiota, segments within the immune populations were very dissimilar whilst those of the microbiota were relatively similar (Figure 2.15). When rare species are given more weight, however, the differences between sites in the microbiota were more apparent (i.e. Kulczynski's metric).

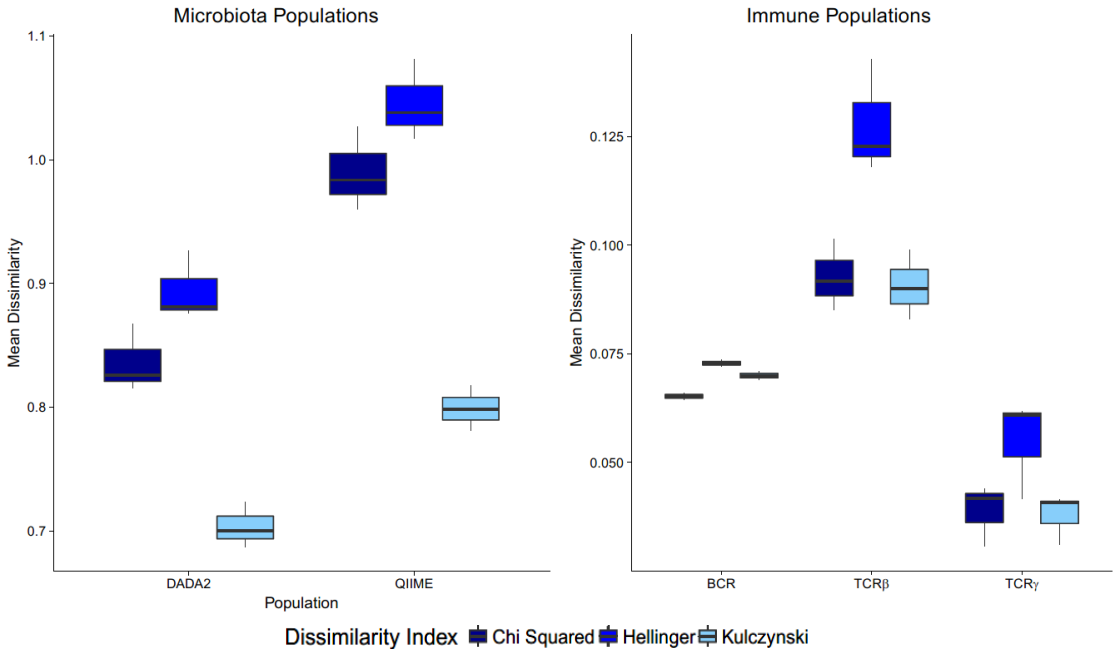


Figure 2.15 – Mean similarity between segments within the microbiota and immune populations, according to three different dissimilarity indices. A lower value indicates a lesser degree of similarity between segments. Boxplots indicate the mean value (inner line), upper and lower quartiles (outsides of box) and range (vertical lines) of the similarity estimates. The microbiota populations were processed using either DADA2 or QIIME as indicated.

There was a significant difference in the mean dissimilarity of the microbiota communities depending upon whether the data was processed by QIIME or DADA2 regardless

of dissimilarity metric (Tukey's HSD test; Chi Squared: p-value $2.15e^{-05}$, Hellinger: p-value $3.32e^{-05}$, Kulczynski: p-value $2.15e^{-05}$). The increased similarity between segments in the QIIME dataset was, however, likely an artefact of the clustering of individuals.

[D.5] Spatial structuring across populations

One of the predictions of the individual-based model in Chapter 1 is that diversity within the gut microbiota can be maintained through spatial structuring of the community (Chapter 1 D.2). Heterogeneity of individual clones (or species) across the grid can be seen in their distributions, the unevenness of which is more apparent in some clones than others (e.g. Figure 2.16).

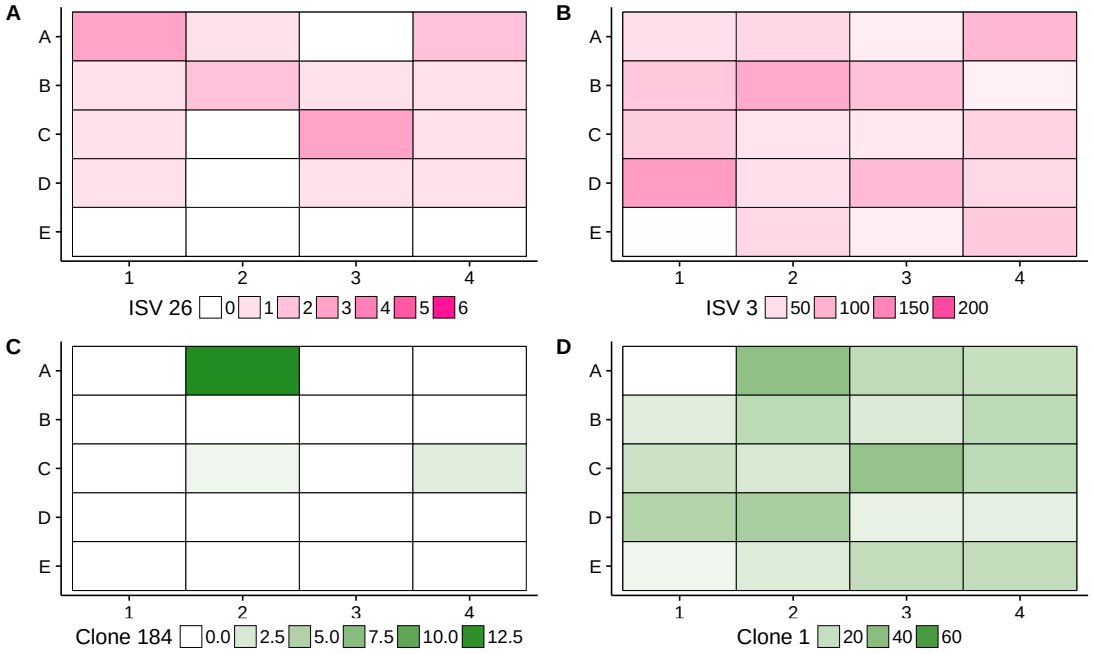


Figure 2.16 – Heat-maps showing the distribution of an example subset of microbe species (A-B) and B-cell receptor (BCR) clones (C-D). The identity of the specific species/clone used and its abundance across the grid are given beneath each heat-map.

Given the number of species (and clones) within all the datasets comparing the distributions of species in this way was not feasible. Instead it was possible to incorporate the information about the extent to which two segments differ as a function of their Euclidean distance (see B.8). As there will always be more possible dissimilarities between segments close to each other, than between those that are further away, the correlation between the

dissimilarity of segments and their respective distance is weighted by the incidence of any particular distance.

There was a weak negative correlation in the dissimilarity between two segments and their distance within both microbiota populations, regardless of the dissimilarity metric used (Figure 2.17, Table.26). The explanatory power of distance on community dissimilarity was, however, quite weak (Table 2.6). The dissimilarity between segments was not significant as a function of either row or column within the grid.

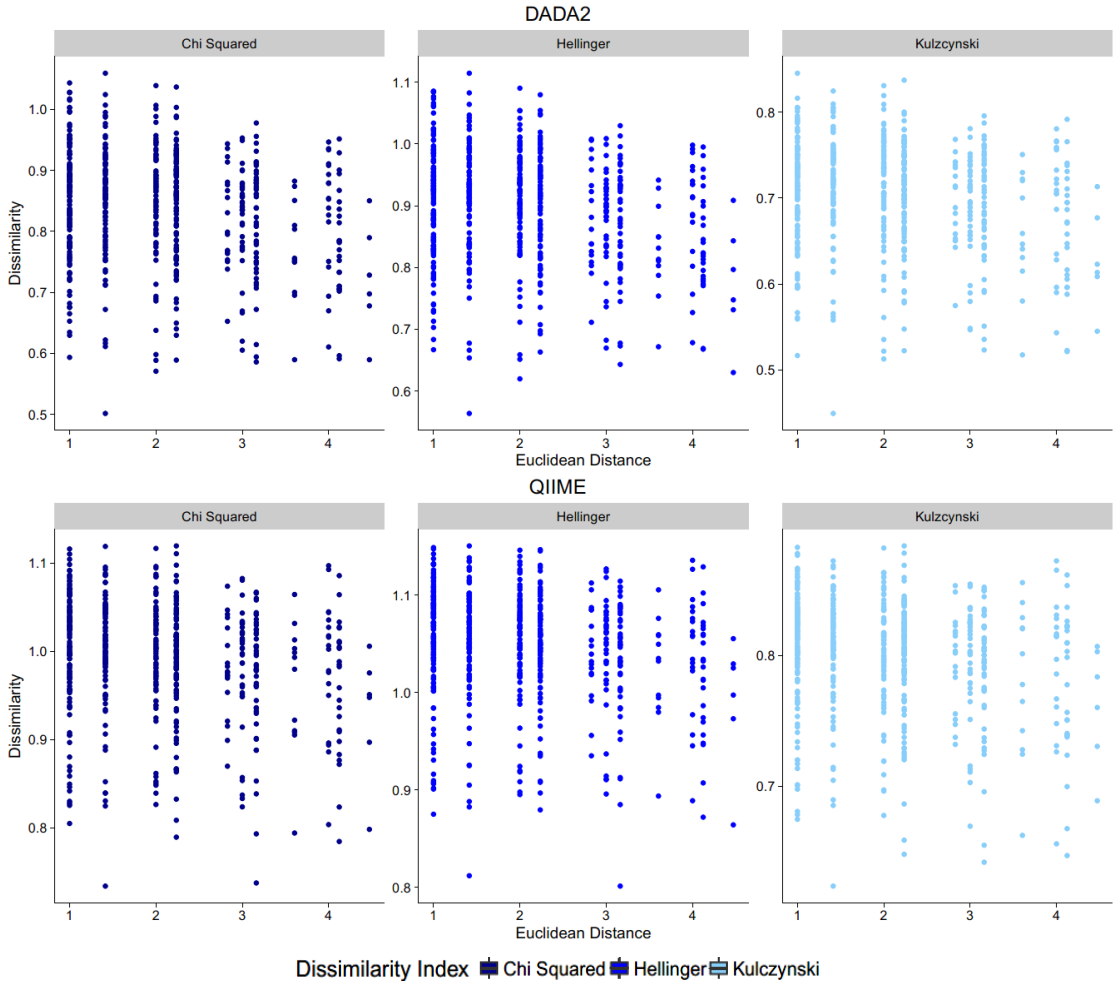


Figure 2.17 – The association between the dissimilarity of segments in terms of community composition and their distance in the microbiota population. The dissimilarity between segments was assessed using three different dissimilarity indices and the microbiota data processed according to either DADA2 or QIIME as indicated.

Only the IgA-producing BCR population showed a significant correlation between segment dissimilarity and distance within the immune datasets (Figure 2.18, Table 2.6). The association between dissimilarity and distance within the BCR population was negatively

correlated and explained a much higher proportion of variation within the BCR dataset than within either of the microbiota populations. These correlations suggest the presence of spatial structuring within the microbiota and BCR populations.

Table 2.6– The significance of the association between segment dissimilarity and distance within the microbiota and immune populations. The microbiota data was processed using either DAD2 or QIIME as indicated. The correlation was assessed using a linear model of the form glm(Dissim~Dist, family=Gaussian, weights) and values are given to three significant figures.

Population	Chi Squared		Hellinger		Kulczynski	
	P-value	R^2 Adjusted	P-value	R^2 Adjusted	P-value	R^2 Adjusted
BCR	1.77e-27	0.188	6.37e-28	0.191	6.80e-27	1.84
TCRβ	0.0778	0.00372	0.0961	0.00312	0.0853	0.00345
TCRγ	0.386	0.000434	0.279	0.000311	0.275	0.000342
DADA2	1.69e-05	0.0304	3.56e-05	0.0280	1.42e-05	0.0310
QIIME	1.18e-04	0.0241	2.02e-04	0.0223	1.97e-04	0.0224

[D.5.1] Specific species drive spatial structuring

The distribution of species within the caecal grid varied (Figure 2.16, Chapter 3 B.3), with some species found in every location whilst others were only present within a single segment. The spread of these species could be driven by environmental heterogeneity on a small scale within the gastrointestinal tract or by microbe dispersal rates increasing the proximal similarity of segments. Species that increase or decrease the correlation between segment dissimilarity and distance can be identified by examining this association in the absence of a given focal species.

There were microbial species in both the QIIME and DADA2 processed datasets that varied the dissimilarity-distance correlation coefficients upon removal (Figure 2.19). 23 species within the DADA2-processed microbial dataset exhibited correlation coefficients which varied significantly from the mean coefficient (more than three standard deviations from the mean; SI Table 2.10). Of these, 14 increased the effect of distance upon dissimilarity by their presence, and nine decreased the association. 19 OTUs within the QIIME-processed dataset significantly varied the correlation coefficient, with eight OTUs increasing, and 11 OTUs decreasing the correlation coefficient with their presence.

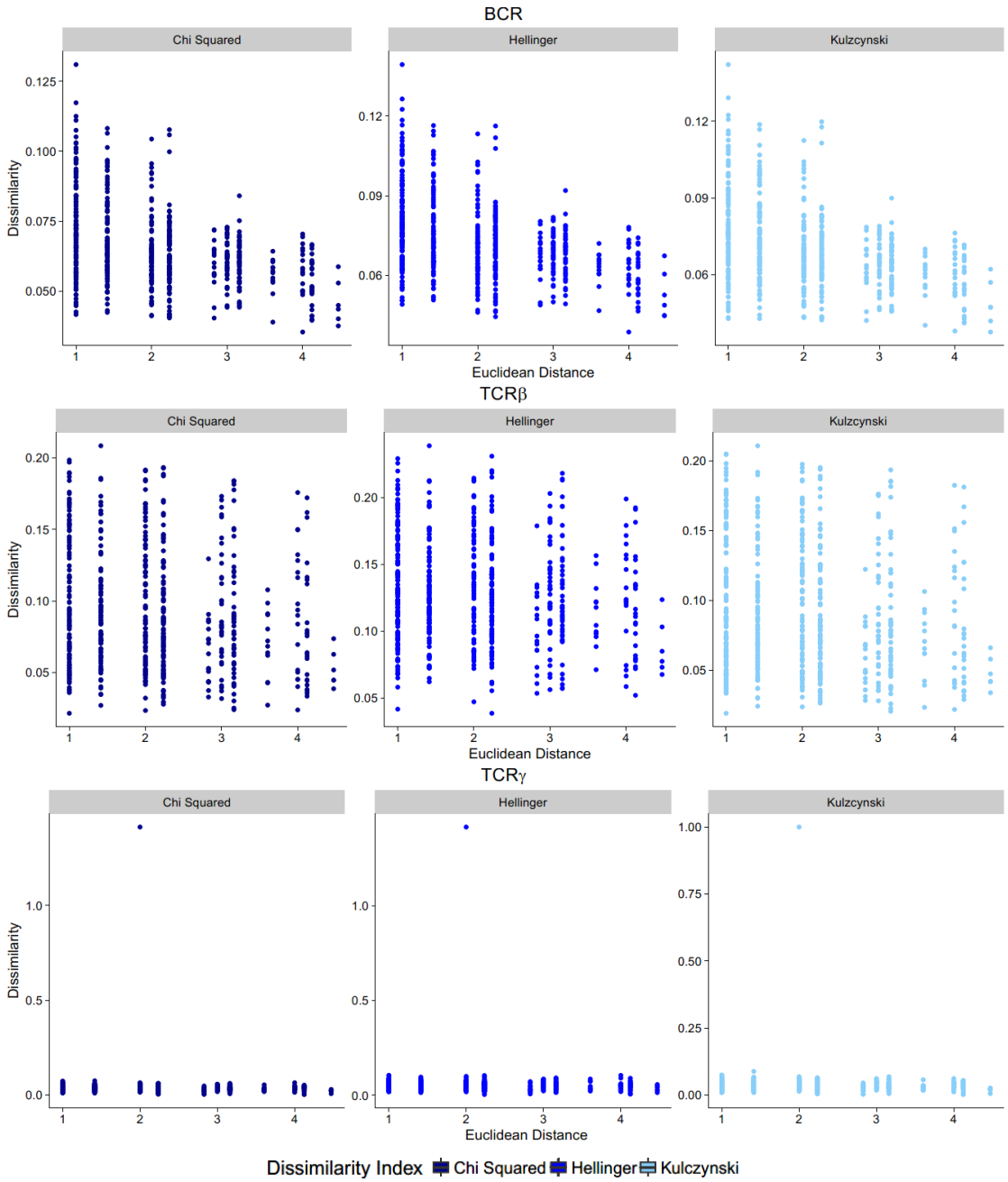


Figure 2.18 – The association between the dissimilarity of segments in terms of community composition and their distance in the immune populations. The dissimilarity between segments was assessed using three different dissimilarity indices.

There was no consistent pattern in species that either increased or decreased the correlation coefficient by their presence, though all species that significantly affect the association between dissimilarity and distance showed large changes between segments. Though there was no consistent pattern, species that improve the spatial structuring within the microbiota tended to have more variable abundance between segments. Within the species

that increased the effect of distance upon dissimilarity, two species showed the same distribution patterns (SI Figure 2.7 ISV5 and ISV8), which could reflect a co-dependence of these species.

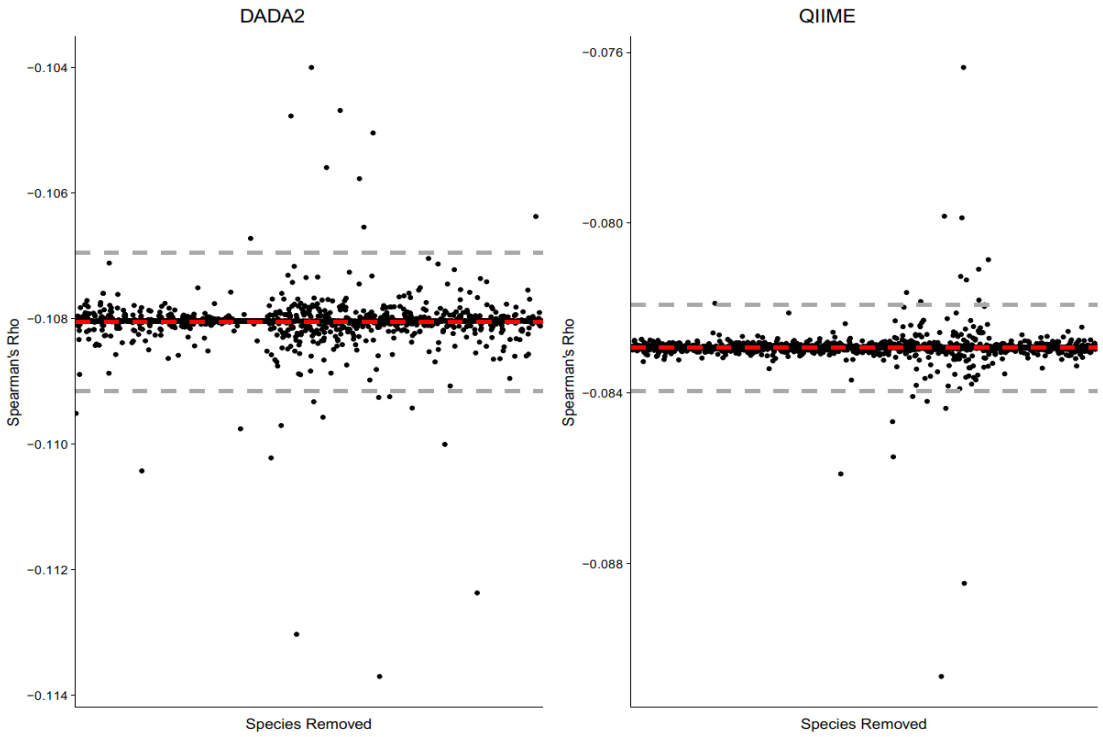


Figure 2.19 – The effect of individual species upon the correlation between segment's dissimilarity and distance within the microbiota, the dataset processed using either DADA2 or QIIME. The correlation between dissimilarity and distance was calculated with a Spearman's rank test and the correlation coefficient (Rho) plotted. All species, except the focal species removed, were retained within the community for the dissimilarity calculations.

[E] Discussion

The composition, diversity, and stability of the gut microbiota is central to its associated benefits[17,21,311,312,42–46,48,49,310]. However, ecological theory suggests a necessary trade-off between diversity and stability[127,311,313]. The absence of the diversity-stability trade-off has been suggested, theoretically, to be facilitated by the spatial structuring of communities[1,249–251]. To consider whether spatial structuring exists within the microbiota requires small-scale consideration of this community. In this Chapter, a method has been developed to assess the impact of scale on the composition of both the microbiota and the adaptive immune cells. The method has been interrogated for biases and used to show there is preliminary evidence that

spatial variation exists within both the microbiota and B-cell populations at the 4 mm^2 grid scale (Figure 2.16 – Figure 2.18).

The importance of factors influencing ecological populations vary depending upon the scale at which a community is considered[1,5,6]. Whilst local populations may fluctuate and suffer from extinctions of community members, these populations can be recolonized on a larger scale[1,7]. Thus, large-scale populations often lose evidence of local variation, as can be seen comparing fine-scale clonal expansions (Figure 2.7) within the TCR repertoires with the overall lack of response in the large-scale populations (Figure 2.6). Similarly the average abundance of microbes across a large-scale gut segment does not reflect the 120-fold difference in 16S sequence counts present between segments at a small-scale (Figure 2.14).

[E.1] Microbiota processing pipelines increase inter-study variation

Microbiota studies are plagued by the problem of inter-study variation[83,84,296–298,314]. This variation can result from different family histories of the participants (both genetically[236,237] and environmentally[234,235]), different housing[238,239] or nutritional conditions[183,242] and the time of sampling[83,88]. Another factor can be added to this variability profile – the processing pipeline that is used for analysis.

Despite using the same starting sequence data, the sequence populations produced by DADA2 and QIIME were fundamentally different; from population structure (Figure 2.5), to estimates of population diversity (Figure 2.10), and even to the taxonomy assigned to community members (Figure 2.11, SI Table 14). The discrepancy between the two pipelines could be driven by the intrinsic differences in how the methods account for the introduction of sequencing errors[267,268,289]; namely the difference in creating OTUs and ISVs. The microbial population as processed by QIIME was dominated by more abundance sequence groups, as a result of clustering sequences that DADA2 identifies as distinct. The DADA2 pipeline is therefore more able to identify rare, and marginally different, sequence types[268]. When asking questions of microbial data it is therefore important to assess which pipeline is more appropriate – including whether the method facilitates comparison with previous literature.

[E.2] Large-scale sampling masks fine-scale heterogeneity

As seen in Figures 2.6 - 2.7; variation within ecological communities depends upon the scale at which a population is considered[1,7]. However, choosing an appropriate scale at which to study a given population is difficult[1]. High variability in sampling attempts at fine-scales is a well-established phenomenon[315], and can be used as an indication that an appropriate sampling scale is being used. The high inter-replicate variability in each population within this study makes assessing differences between segment communities more difficult, however despite this effect there was significant variation between all segments in all the studied communities in terms of gene-usage (section D.3), assigned taxonomic profile (Table 2.5, SI Table 14), and individual sequence presence (Figure 2.16).

Heterogeneity between segment communities can arise as a result of environmental variation (e.g. fucosylation patterns[116,129,154] or mucus glycosylations[147,157,199]), segment species composition (e.g. presence of competitive or beneficial contemporary species) or even by the densities of adjoining communities[1,315]. At fine scales the interactions between species and the local abundance of populations are expected to be more influential in determining community composition[8–10]. Local abundance of the microbiota varies substantially across segments (Figure 2.14). Although assessing microbiota abundance using 16S qRTPCR carries the caveat that different microbe species contain different numbers of 16S copies[316], there is no reason to suspect that microbes will differentially aggregate into communities based on 16S copy number.

Similar assessments of local immune population abundance were not possible within this study due to the inherent challenge of designing qRTPCR primers that allow for the hyper-variability and genetic recombination of adaptive immune compounds whilst simultaneously assessing productive use of these compounds rather than simple genomic presence. It may be possible in further work to investigate immune population abundance across fine scales through coupling bisulfite sequencing with qRTPCR.

[E.3] Spatial structuring in caecal microbiota and immune populations

Both B-cells and T-cells are found within the gastrointestinal tract of chickens. Though TCRαβ, TCRγδ and BCR populations have been shown to be spatially structured transversely across the gut[317], there has been relatively little evidence of any lateral variation in immune populations across the gut[301,317] and no evidence of fine-scale differences. Similarly, although there is evidence of large-scale variation in the composition of the associated microbiota between[247] and within tissues[75,83,89], there has been no fine-scale interrogation of these differences.

Spatial heterogeneity maintains community diversity by facilitating the coexistence of species through the formation of refugia[99,249,315]. The communities within the gastrointestinal tract, both adaptive immune and microbial, vary across space (see D.5). However only the microbiota and BCR populations appear spatially structured (Figure 2.17 – Figure 2.18). Within both populations this spatial restriction may be a result of reduced migration between local communities or the aggregation of replicating individuals. The restricted movement may be exacerbated by interactions between the adaptive immune and microbial populations – specifically between the IgA-producing B-cells and the microbiota[99].

[E.4] Preliminary work and limitations

The aim of this chapter has been to develop a fine-scale method which can be used to interrogate the predictions produced theoretically in Chapter 1 (see Chapter 1 D.2). Specifically, this fine-scale sampling method can be used to assess whether there is evidence of fine-scale spatial heterogeneity within the microbiota which would be expected if host-secreted adhesion is facilitating the maintenance of microbiota diversity. There are, however, some limitations to how this methodology can address this hypothesis.

Firstly, all the work within this chapter should be caveated as an interrogation of a single L6 6-week-old chicken. Although it is possible to derive preliminary conclusions from this work further experiments are necessary to assess whether the trends seen here are replicable across multiple individuals, or even across multiple host strains or species.

Secondly, this methodology uses the IgA-producing B-cell receptor population as a proxy for IgA secretion across the epithelium. The individual-based model in Chapter 1 predicts that host-provided adhesion can spatially structure the microbiota and this thesis has adopted the assumption that this structuring is achieved by a host through the production of IgA. However IgA-producing B-cell presence does not necessarily equate to IgA production and secretion across the intestinal epithelium. Instead fluorescence-activated cell sorting (FACS) could be used to separate out IgA-bound and non-IgA bound bacteria from each segment location before 16S analysis is completed, although this could be challenging given the small size of the samples. The IgA amino acid sequences could be identified using protein sequencing and then compared with the B-cell receptor distribution to determine how well B-cell receptor presence represents IgA presence across locations within the grid.

Although the association between segment location and dissimilarity could be seen using linear regression, in the case of the microbiota and B-cell populations, it would be interesting to use techniques from mathematical ecology to assess whether this association is simply a function of distance. Ecological analyses often adjust for spatial autocorrelation (the idea that near communities are more likely to be similar than distant communities) to avoid problems of pseudoreplication[318,319]. In the association between dissimilarity and distance presented here this was accounted for using weights within the linear models built. However, an alternative would be to use the Moran's I statistic to infer whether segments are more distinct or similar across space. Future work might therefore investigate whether the spatial associations seen within this thesis persist using ecological assessments of spatial autocorrelation.

[F] Conclusions

The primary aim of the work presented here was to develop a method of investigating fine-scale dynamics within the microbiota and adaptive immune communities in order to assess whether there is any evidence of spatial structuring within the gastrointestinal microbiota. The small-scale grid method was developed and no bias shown through sampling, amplification,

sequencing or processing of the data demonstrated. The data produced was then used to show that, as in ecological communities, the scale at which a population is assessed influences the dynamics of that population and that by considering both microbiota and adaptive immune populations at small-scales it is possible to see the spatial heterogeneity in these communities. This spatial variability has been predicted to facilitate the diversity and stability of the gut microbiota, both important factors in maintaining the microbiota's beneficial function for its host.

Chapter 3

The role of rare species in microbiota community interactions

[A] Introduction

The microbiota is essential for the normal development and survival of the host. It affects host nutrient acquisition[11,15,109], pathogen protection[20,48,112], and the priming of the immune system[39,51,107]. However, the benefits of the gut microbiota depend upon its specific composition[17,21,48,49,310], complexity[42–45], and stability[46,311,312]. Thus factors that affect the composition and stability of the microbiota affect the evolutionary fitness of a host.

The exact composition of a beneficial microbiota is variable between individuals[232,233] and species[72,76,310]. However, the necessity of microbiota diversity and stability in retaining the benefits of the microbiota for the host remains constant[42–46,311,312]. Until the 1970s the widespread existence of diverse, stable communities in nature, such as the microbiota, was unremarkable, with increases in diversity generally believed to improve stability[313,320]. The rise of mathematical models in ecology, in particular Robert May's work using linear stability analysis (1973)[127], showed that increases in diversity do not intrinsically increase a community's stability. In fact, in sufficiently large ecological systems, the probability of a community persisting in the absence of additional stabilising factors is close to zero[127,321,322]. Explaining the relationship between complexity and stability has since become a central question in modern ecology.

[A.1] Defining ecological stability

Stability can be, and often is, defined in many different ways[322,323]. The multiple definitions of stability arise out of an attempt to capture the multifaceted nature of a community's dynamics and response to disturbance. Stability can be defined in terms of a community's asymptotic stability (whether it returns to equilibrium), variability (how much it

varies over time or space), persistence (length of time it retains the same state), resistance (whether it retains its initial state after disturbance), and resilience (the rate at which a system returns to equilibrium after disturbance).

Both theoreticians and empiricists are interested in ecological community stability[323]. However there has been a serious disconnect in the types of stability that the two camps study, with theoretical studies predominantly focused on asymptotic stability, while experimental studies use variability as a measure of stability. Within the human gut, the stability of the microbiota in terms of variability[324,325], resistance[72] and resilience[326] has been demonstrated. Less information is available within the avian microbiota, though its age-related compositional stability[88–90] and functional response to antibiotic-generated perturbation suggest similar dynamics to mammals[327].

[A.2] Stability within complex communities

Robert May's work in 1973 showed that community complexity cannot lead to increased stability in isolation and thus other factors must facilitate complex community stability in nature[127]. A large literature interrogating the factors that facilitate community stability has since arisen, with explanations including the strength[127,328–330], type[249,322] and structure[331–333] of species interactions and spatial heterogeneity[313,334]. Within the context of this thesis, the effect of within community interactions and the existence of spatial structuring are particularly interesting.

[A.2.1] Species interactions

Species interactions include both within species (intraspecific) and between species (interspecific) interactions. May's initial work predicted that a community would be stable only if the complexity of a community is less than the strength of the intraspecific interactions within it (Equation 3.1)[127]. The strength of intraspecific interactions within a community and its complexity are therefore hypothesised to be the driving factors in whether the community is stable.

$$\sigma\sqrt{SC} < \bar{d}$$

Equation 3.1 – Robert May's requirements for community stability. Where σ is the standard deviation of interaction strength, S is the species richness of the community, C its connectance (the probability of two species interacting) and $\overline{d}$ is the magnitude of intraspecific competition.

The effects of interspecific interaction upon community stability are similarly influenced by the strength of the interaction between species. However, unlike intraspecific interactions, which must be stronger than the community complexity for a stable community[127], interspecific interactions stabilise communities when there is a predominance of weak interactions[328–330]. Community perturbations destabilise communities when positive feedback loops are formed which drive the community away from its equilibrium, weak interactions dampen these positive feedback loops and so stabilise the system[328]. In real communities, interspecific interaction strengths are non-randomly structured, with long chains of species interactions dominated by weak interactions[332,335,336]. The structuring of interaction strengths enhances stability by reducing the amount of intraspecific competition needed to stabilise the community.

The types of interspecific interactions within a community also influences its stability. A competitor species may influence a focal species positively, negatively or neutrally. The pairwise interaction between a focal and competitor species can thus be cooperative (+/+), competitive (-/-), commensal (+/0) or amensal (-/0). As cooperative interactions form positive feedback loops, they can be treated as generally destabilising, whereas competitive interactions disrupt feed-forward loops and are thus stabilising[249,322]. However the effect of interaction types upon community stability is contextual, with the structure of a community influencing the role of interspecific interactions upon its stability[331–333]. In real communities the interactions within small groups of species (two to five species) can drive the entire interaction environment, and thus stability, of a community[337].

[A.2.2] Spatial heterogeneity

Environmental variation enhances community diversity by isolating small groups of species, facilitating the creation of refugia where less advantageous species can survive in the absence of more competitive contemporaries[230,231,260]. By changing a species' environment, either in terms of community interactions or selection pressures, community stability can be enhanced by reducing the strength of competition between individuals[313,338]. Even where spatial patches are identical, the isolation of groups of species reduces the potential for interspecific interactions which can stabilise communities by decoupling feedback loops[249].

[A.2.3] Stability is more than species interactions

While May predicted, using randomly generated species networks, that increasing complexity tended to destabilise systems, real communities do not exist in isolation, nor do they appear randomly structured[329,339]. The stability of a complex system may in fact depend upon both the scale at which the community is studied or upon background population variability[313]. Stability can be maintained through the differential responses of species within a community to different selective forces, suggesting species richness increases stability as there is greater potential for differential responses to perturbations. In this way community-level stability can decrease with diversity (in accordance with May's predictions) whilst population-level stability remains relatively unaffected.

The factors involved in driving stability also depend, to some extent, upon the definition of stability itself[340,341]. For example, diversity within a plant ecosystem is positively correlated with community persistence but has little effect upon population variability. Further, the association between diversity and stability may be clouded by a link to functional rather than species diversity. Diversity within a community can allow it to persist by providing either functional or evolutionary redundancy.

[A.3] Estimating species interactions

Species interactions can be estimated from experimental datasets (both field and laboratory) and observed datasets. Species interactions can be measured by introducing

species to, or removing them from, a community and detecting changes in community dynamics, or by looking at the growth of single species in isolation and then in competition with other species. There are advantages and disadvantages to these methods (reviewed in Wootton & Emmerson 2005[342]), though they share an assumption of measuring the changes in communities over time.

The effects of a species upon a community can be considered spatially as well as temporally, using the grid structure of Chapter 2. Using spatially, rather than temporally, distinct communities reduces the effects of experimental duration upon interaction strength estimates based upon species removal. However, the interaction estimates are still limited by the density of the focal species and indirect interactions with additional species may skew species interactions.

[A.4] Microbiota stability

The methods used within this chapter to estimate species interactions rely upon the assumption that the microbiota community is stable. Experimental work has shown that the beneficial functions of the microbiota depend upon both its complexity[42–45] and its stability[46,311,312]. The importance of microbiota diversity for host health has been shown through the use of single species or defined communities[343]. The stability of the microbiota, and its ability to confer a beneficial properties to a host, is harder to determine.

The resistance and resilience of the microbiota is contextual, depending upon the individual host[311,324], age[327,344], type of disturbance[183,345–347], and the number of perturbations[348,349]. For example, though the adult microbiota is resilient to a single application of antibiotics[349–351], multiple rapid perturbations of the microbiota during infancy are unrecoverable and lead to long-term shifts in microbiota composition[54].

Even in the absence of direct perturbation, microbiota composition varies over a host's life[311]. However this variability may result from temporary fluctuations in species abundance, coupled with sampling limitations with respect to rare species, simulating the apparent loss of

rare individuals despite the community composition remaining constant. This is consistent with the observation that many variable taxa appear to be reoccurring within the community[352].

The stability of the microbiota, defined by its asymptotic stability, variability, persistence, resistance or resilience, is thus somewhat contextual.

[A.5] Chapter 3 aims

The aim of this chapter is to use the microbiota data produced in Chapter 2 to estimate species interactions within the gastrointestinal microbiota of a six-week-old L6 chicken. The sampling grid provides a natural perturbation experiment, allowing the comparison of focal species dynamics in the presence and absence of competitor species. The intra and interspecific interactions of focal species will then be examined as a function of the species local population abundance and geographic spread to assess whether rare and abundant species experience the same interaction environments.

Ultimately these interactions could be used to parameterise a model to assess the role of rare species in community stability, though this is outside the scope of the work presented here.

[B] Methods

This study uses the grid presented in Chapter 2 as a naturally occurring perturbation experiment. Species interactions are defined as the combined effect of a competitor species population upon a focal species over the entire grid. The ability to estimate species interactions within this framework relies upon the assumption that the 6-week-old chicken microbiota is stable, and therefore represents a community at equilibrium.

[B.1] Microbial community sequence data

The grid format and sequences were produced as in Chapter 2 sections B.1-B.7. Only data from the DADA2 pipeline with chimeras and contaminants removed was used within this analysis. DADA2 was used in order to utilise its increased sensitivity to, and robust assignment

of, low abundance sequences. 'Species' within this chapter refers to the independent sequence variants retrieved from DAD2, not taxonomic assignment.

[B.2] Species interaction estimates by Lotka-Volterra derivation

The dynamics of a focal species' growth can be defined in terms of its intrinsic growth rate and the strength of the intraspecific (α) and interspecific (β) interactions it experiences (Equation 3.2).

$$\frac{dx}{dt} = rx(1 - \alpha x - \beta y)$$

Equation 3.2 – Equation describing a focal species (x) growth. α is the strength of intraspecific competition of the focal specific, β is the strength of the interspecific interactions between the focal species (x) and the competitor species (y), r is the intrinsic growth rate of the focal species, x is the abundance of the focal species and y is the abundance of the competitor species.

The species interactions estimated here are always determined within the context of a community. The interaction coefficients estimated here are not produced through the traditional pairwise comparison experiments. Intraspecific interactions are usually experimentally determined by growing a focal species in isolation (with the growth dynamics defined in Equation 3.3), however there is no case within the grid structure in which a single species occurs in isolation. The intraspecific interactions presented within this study are therefore calculated with respect to a specific competitor species.

$$\frac{dx}{dt} = rx(1 - \alpha x)$$

Equation 3.3 – Lotka-Volterra equation describing a focal species (x) growth. α is the strength of intraspecific competition of the focal specific, r is the intrinsic growth rate of the focal species and x is the abundance of the focal species.

Since the microbiota is assumed to be at an equilibrium, $\frac{dx}{dt} = 0$. Therefore $0 = r\tilde{x}(1 - \alpha\tilde{x})$. Equation 3.4 can thus be rearranged and used to estimate a focal species' intraspecific interactions.

$$\alpha = \frac{1}{\tilde{x}}$$

Equation 3.4 – The localised intraspecific interaction (α) calculated for a focal species (x), where $\tilde{x}$ is the abundance of the focal species in a given location at equilibrium.

In the case of each focal-competitor species pair, all instances in which the focal species was found in the absence of the competitor species were used to calculate its intraspecific interactions according to Equation 3.4. Where a focal species occurred more than once in the absence of its competitor species its intraspecific interactions were calculated for each instance of isolation and the mean of these estimates taken as the focal species' intraspecific interaction (Equation 3.5).

$$\bar{\alpha} = \frac{\sum_{i=1}^{n} \frac{1}{\tilde{x}_i}}{n}$$

Equation 3.5 – The intraspecific interaction ($\bar{\alpha}$) for a specific focal species at equilibrium ($\tilde{x}$). α can be calculated for each occurrence (i) where a species is not found in the presence of a specified competitor species (see Equation 3.4). The intraspecific interaction of a species is taken to be the mean of α. So $\tilde{x}_i$ is the abundance of a focal species in a given location and n is the number of locations in which the focal species is found without its competitor.

The interspecific interactions between a focal-competitor species pair were estimated by rearranging Equation 3.2 to give Equation 3.6. As in the case of the intraspecific interactions, where a focal-competitor species pair co-located more than once, the impact of the competitor species upon the focal species is taken as the mean of all the individually calculated interspecific interactions.

$$\beta = \frac{1 - \bar{\alpha}\tilde{x}_j}{\tilde{y}_j}$$

Equation 3.6 – The interspecific interaction (β) calculated between a focal species ($\tilde{x}$) and a competitor species ($\tilde{y}$) at equilibrium. Where $\tilde{x}$ is the abundance of the focal species in a given location (j) and $\tilde{y}$ is the abundance of the competitor species in the same location. $\bar{\alpha}$ used here is the mean of all the possible intraspecific interactions of a focal species in the absence of the competitor species.

Within this study the estimated interspecific interactions are unidirectional and therefore described as positive (+), negative (-) or neutral (0) rather than mutualistic (+,+), competitive (-,-), commensal (+,0) or amensal (-,0). The interspecific interactions described

here only refer to the effect of the competitor species upon the focal species, not vice-versa. However, interspecific interactions are calculated for any focal-competitor species pair possible. If species A is the focal species and species B the competitor species in one calculated interspecific interaction, species B may also be the focal species and species A the competitor species in another calculation.

[B.3] Species excluded from interaction estimates

There are limitations to using the grid structure to estimate species interactions, especially with respect to extremely abundant and rare individuals. The restrictions of this calculation method means that a large number of focal-competitor species interactions could not be assessed (Table 3.1). The estimates of interaction between a competitor species and a focal species requires that the focal species be found in the absence of the competitor species (to calculate α), and in the presence of the competitor species (to calculate β).

Table 3.1 – The percentage of each technical replicate for which interactions were not able to be calculated by problem group. The percentages of each technical replicate that are a problem are shown as both the percentage of the total species ('Species') and of the total abundance within the community ('Abundance') they represent. All values are given to two decimal places.

Problem Group	Replicate 1		Replicate 2		Replicate 3	
	Species	Abundance	Species	Abundance	Species	Abundance
Ubiquitous	2.16	63.80	1.07	51.99	1.50	55.61
Single Instance	47.78	4.61	47.48	4.33	49.00	6.04

Species that are found ubiquitously throughout the grid cannot be competitor species, as there is no instance in which a focal species can be found in isolation. Therefore it was not possible to calculate the interactions (either intraspecific or interspecific) between ubiquitous species. Conversely species that are only found once within the grid cannot be focal species, as they cannot be found in both the presence and absence of a competitor species.

As is apparent from Table 3.1 there was inconsistency between the replicates in which species were ubiquitous, single instance or non-co-locating. It was possible to calculate interactions for some species pairs in one technical replicate but not in others. The analyses

presented here were not restricted to focal-competitor species pairs that were observed across technical replicates.

[B.4] Table of terms

Several terms are used repeatedly within this chapter, particularly with respect to rarity categories and the types of interactions experienced. All terms are defined upon first use and are summarised in Table 3.2.

Table 3.2 – Explanation of abbreviations and terms used throughout this chapter. Log(A) means log(total abundance) of the species within the whole grid. Rarity group abbreviations are 'AC' abundant by count, 'MC' moderate by count, 'RC' rare by count, 'SC' single incidence count, 'UP' ubiquitously present, 'AP' abundantly present, 'RP' rare presence and 'SP' single presence. Where '*NA*' is indicated there was no abbreviation used.

Abbreviation	Term	Definition
AC	Abundant	Species where log(A) ≥ 5
MC	Moderately abundant	$2 <$ Species where log(A) and < 5
RC	Rare	Species where log(A) ≤ 2
UP	Global	Species found in every segment
AP	Widespread	Species found in $7 <$ segments ≤ 20
RP	Localised	Species found in ≤ 7 segments
α	Intraspecific interaction	Interactions between individuals of the same species
β	Interspecific interaction	Interactions between individuals of different species
-	Negative interaction	$\beta \geq 1$
n	Weak negative interaction	$0.01 < \beta < 1$
0	Neutral interaction	$-0.01 \leq \beta \leq 0.01$
p	Weak positive interaction	$-1 < \beta < -0.01$
+	Positive interaction	$\beta \leq -1$
NA	Total community size	The total number of individuals within a community, as assessed by qPCR

[B.5] Calculations and statistics

All calculations were performed within R (versions 3.3.1 and 3.5.0)[271] using the packages dplyr and reshape2[353] for the calculations in addition to ggplot2[275], cowplot[276] and lme4[277] for data display and statistical analysis.

[C] Results

The ecological community considered here is that of the gastrointestinal microbiota of a single six-week-old L6 male chicken (see Chapter 2 Methods). The interaction environment of

different focal species is considered in terms of its interspecific interactions, as a function of the focal species' rarity.

[C.1] Interactions within the microbiota

As seen in other ecological communities[329,330,336], weak interactions dominated within the chicken gastrointestinal microbiota (Figure 3.1) both in terms of intra and interspecific interactions. Weak interactions have been associated with the stabilisation of communities by reducing the number of feed-forward loops[328] and facilitating spatial structuring[313,338]. However, the skew towards weak interactions within the chicken microbiota was reduced relative to other ecological communities[336]. Moreover, unlike what is expected of stable communities, the interspecific interactions estimated here are largely positive (Figure 3.1 C), though the strength of the positive interactions varied between technical replicates.

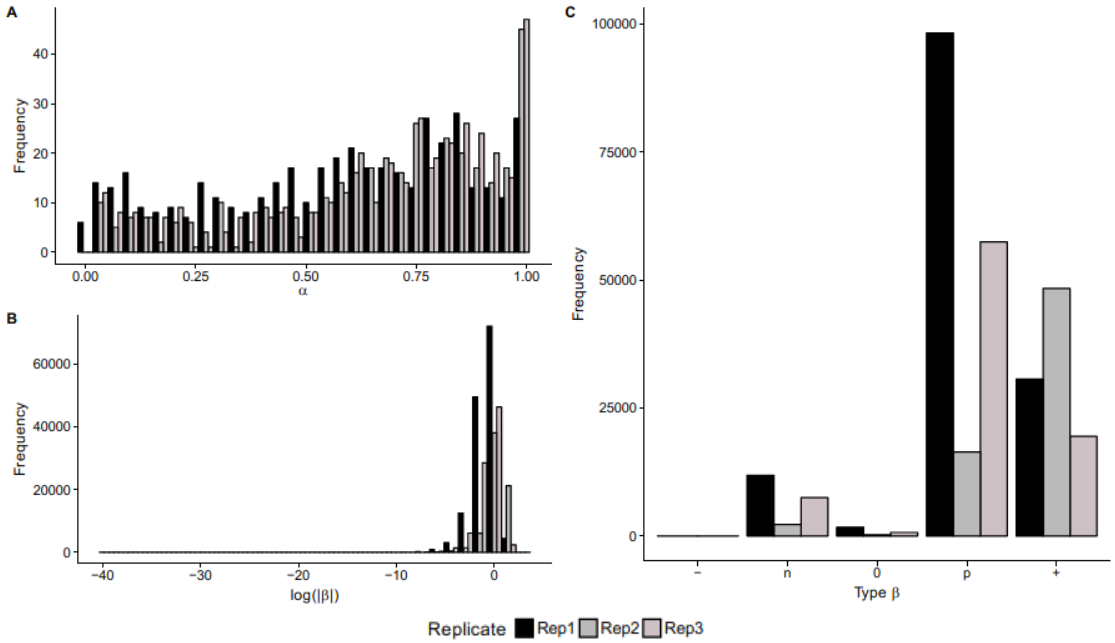


Figure 3.1– Interaction strength and type within the chicken gastrointestinal microbiota. A) The distribution of intraspecific interaction strengths across the whole chicken gut microbiota community, taken as the mean of a focal species' intraspecific interactions in each grid segment. B) Distribution of interspecific interaction strengths across the microbiota community, taken as the mean of each species pair interactions in each grid segment. C) Frequency of interaction types experienced across the whole microbiota, for all three technical replicates. Interaction types are; Strong negative ('-'), weak negative ('n'), neutral ('0'), weak positive ('p') and strong positive ('+').

Considering the distribution of interaction strengths and types within the whole microbiota community does not, however, give any indication of the role of specific species within the community. Whether there is a link between the abundance of a focal species and its interaction environment is considered through the rest of this chapter.

[C.2] Defining rarity

In order to consider the role of rarity in species interactions it is first important to determine what it means to be rare. Intuitively, a rare species is one that is found in a low abundance. However a species might also be rare in terms of location; a species that is found abundantly on a single island might still be considered rare globally. Rabinowitz (1981) characterised rarity along three axes: local population size, geographic isolation and habitat specificity[354]. A species might be rare on one, or several, of these axes.

[C.2.1] Rarity within this study

Within this experiment, rarity can be considered in terms of geographic isolation and local population size (Figure 3.2). Specifically, a particular species is rare if it appears in few segments within the grid (geographic isolation) or is found at low frequency (local population size). For consistency between grid segments a species' frequency is defined as its summed abundance across all segments. Thus, a focal species may be 'rare' in total abundance within the grid, whilst simultaneously abundant within a single segment.

In this study 'rare' species are those found in fewer than eight segments (of a possible 20) or with a log(Abundance) of less than two (summarised in Table 3.2). It is important to note that, using these definitions, while a species that is geographically rare can be both locally rare and locally abundant, locally rare species are intrinsically geographically restricted (Figure 3.2). For clarity, rarity groups based on local abundance are termed 'rare' (RC), 'moderately abundant' (MC) or 'abundant' (AC), whilst those based on geographic restriction are 'localised' (RP), 'widespread' (AP) or 'global' (UP).

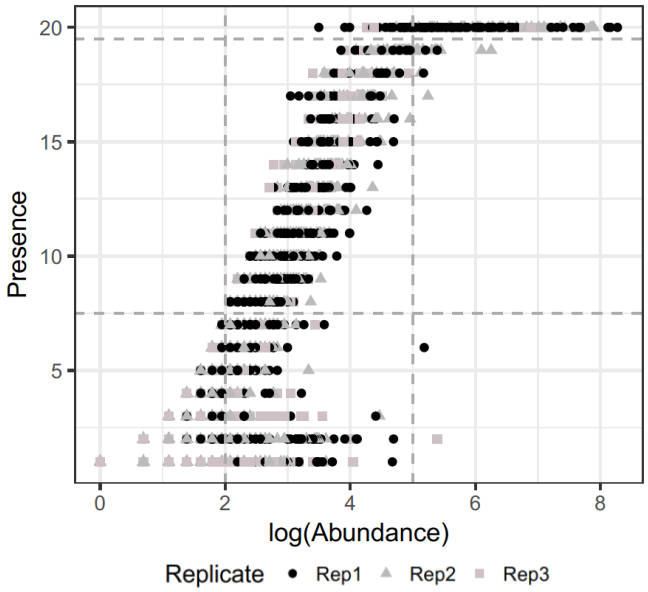


Figure 3.2 – Determining species rarity across two of Rabinowitz's rarity axes. Microbial population data distributed across axis of rarity; two axes are presented corresponding to Rabinowitz's geographic isolation (presence) and local abundance (log(abundance)). Presence data is comprised of the incidence of each species within the grid i.e. the number of segments within the grid within which the species is found in any abundance. Abundance data is the count of a given species summed across the full grid. Points are plotted for all replicates (by colour) and the breaks from each group are indicated by dotted lines.

[C.2.2] Characterising rarity groups

The number of species, and the proportion of the community, that made up each rarity group varied (Table 3.3). The moderately abundant (MC) and widespread (AP) groups contained the greatest proportion of species. As would be expected, the abundant (AC) and global (UP) groups contributed most to each community in terms of abundance. Together, the widespread and global groups contained almost all of the species found within the grid. The abundant and global groups contained most of the taxonomically assigned species, whilst rare species were almost entirely unable to be taxonomically assigned, regardless of how rarity is defined (Table 3.3). The inability to assign rare species is likely a restriction of the database used to assign taxonomy.

As in other ecological communities, although abundant species dominate the chicken microbiota numerically (Table 3.3) more species are rare than abundant. The small numerical contribution of rare species to the microbiota community is consistent at the fine-scale with

rare species contributing relatively few members to a given community regardless of how rarity is defined (e.g. Figure 3.3). Within the abundance rarity groups only the proportional contribution of rare (p-value $2.39e^{-06}$, R^2 adjusted 0.597) and single count (p-value $2.66e^{-03}$, R^2 adjusted 0.373) species varied between segments within the grid. Although the localised proportion of the microbiota also varied between locations (p-value $4.43e^{-03}$, R^2 adjusted 0.350), there was a significant variation in the proportion of each segment's community that was widespread (p-value $7.25e^{-03}$, R^2 adjusted 0.326).

Table 3.3 – The characteristics of each rarity group, defined by either local population abundance or geographic spread. 'Unassigned' refers to the percentage of species within each group that were unable to be taxonomically assigned, 'Species Richness' to the number of species within each rarity group as a percentage of the total number of species in the grid and 'Proportion of Community' to the sequence count within each rarity group as a percentage of the sequence depth. All values are the mean of the three technical replicates, with the standard deviation of the mean displayed, given to two decimal places. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'SC' single count, 'UP' global, 'AP' widespread, 'RP' localised and 'SP' single instance. For full explanation of abbreviations see Table 3.2.

Rarity Basis	Group	Unassigned (%)	Species Richness (%)	Proportion of Community (%)
Local Abundance	AC	9.89±0.31	8.16±2.82	60.37±3.34
	MC	59.58±0.97	53.91±7.63	29.72±1.00
	RC	96.49±0.65	31.92±6.67	8.64±2.77
	SC	99.41±0.22	6.01±3.79	1.26±0.92
Geographic Spread	UP	11.02±0.15	5.39±1.86	52.24±2.21
	AP	54.55±1.47	84.48±4.84	44.83±0.27
	RP	94.67±3.80	2.64±1.73	0.59±0.42
	SP	97.97±1.30	7.49±5.02	2.34±1.85

The variation in the proportional community contribution of the rarity groups was not associated with either the sequence depth of the segments or their sample coverage (SI Table 3.1), with the exception of the widespread group (AP) where the proportion of widespread species increased with sequencing depth.

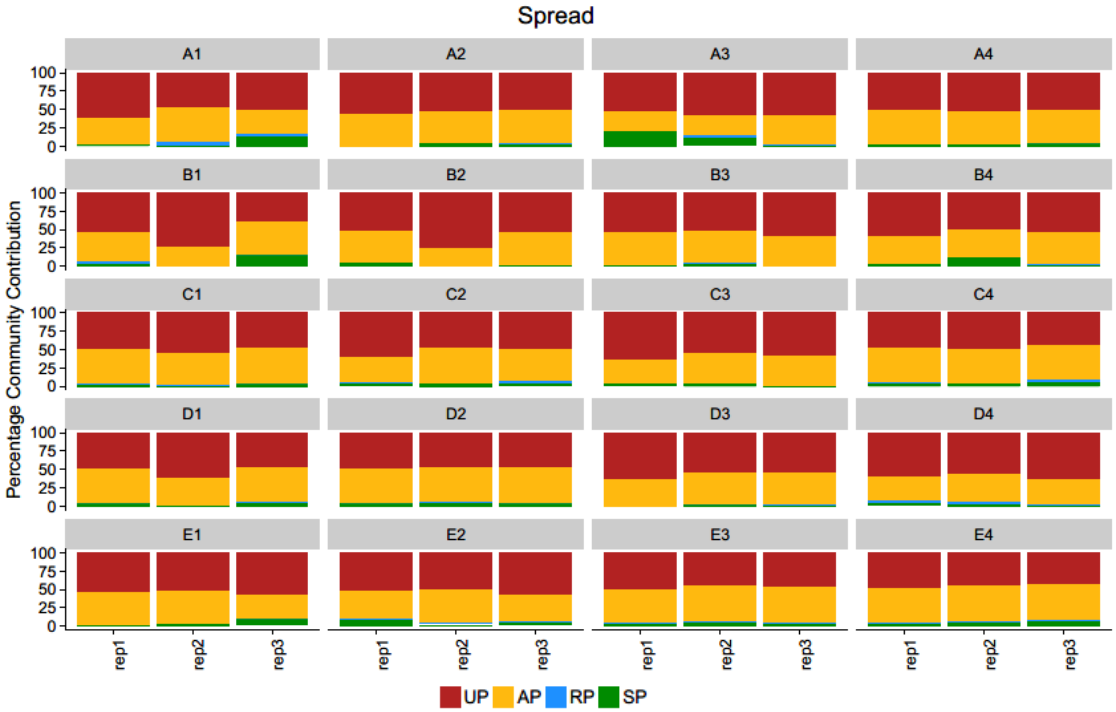


Figure 3.3 – Contribution of rarity groups, defined by geographic spread, to community composition across space. Segments are named after their location within the caeca, with segments A apical and segments E distal, the cylinder of the gut means segments 1 are next to segments 4 (see Figure 2.1). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'SC' single count, 'UP' global, 'AP' widespread, 'RP' localised and 'SP' single instance. For full explanation of abbreviations see Table 3.2.

[C.2.3] Community diversity and rarity

The species richness, Shannon's diversity and Simpson's entropy of segments varied across the grid (SI Figure 3.1). There was a significant association between the contributions of some rarity groups to a segment and its diversity estimates (Table 3.4), although the effect varied between the indices used. The strongest associations were between the moderately abundant and widespread groups, with the proportional contribution of these groups positively correlated with species richness and Shannon's diversity.

The size of a segment's community, assessed using qPCR, affected the proportional contribution of the abundant, rare, global and widespread groups (SI Table 3.2). As community size increased the proportional contribution of the abundant group increased whilst the rare group became less abundant. There was no association between a segment's community size and either Shannon's or Simpson's diversity of the segment, though species richness decreased

as community size increased (see Chapter 2 D.3). Segments with a larger microbiota population were not more populous as a result of an increased number of species, but rather due to an expansion of abundant species.

Table 3.4 – The significance of the association between the proportional contribution of each rarity group to a given segment and the segment's diversity indices. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'SC' single count, 'UP' global, 'AP' widespread, 'RP' localised and 'SP' single instance. For full explanation of abbreviations see Table 3.2.

Diversity Estimate	Rarity Basis	Group	P-value	R^2 Adjusted
Species Richness	Local Abundance	AC	3.38e-02	0.0594
		MC	2.56e-03	0.132
		RC	0.480	-0.00845
		SC	0.692	-0.0147
	Geographic Spread	UP	1.06e-02	0.0921
		AP	1.02e-03	0.157
		RP	0.693	-0.0147
		SP	0.168	0.0162
Shannon's Diversity	Local Abundance	AC	0.269	0.00416
		MC	2.28e-03	0.135
		RC	0.697	-0.0146
		SC	2.39e-02	0.0704
	Geographic Spread	UP	5.29e-02	0.0469
		AP	3.54e-02	0.0581
		RP	0.883	-0.0172
		SP	0.398	-0.00474
Simpson's Entropy	Local Abundance	AC	0.723	-0.0150
		MC	0.0947	0.0310
		RC	0.0883	0.0329
		SC	1.30e-02	0.0877
	Geographic Spread	UP	0.846	-0.0166
		AP	0.220	0.00899
		RP	0.943	-0.0175
		SP	4.89e-02	0.0500

[C.3] Interspecific interactions

Interspecific interactions are one of the main forces influencing community stability, especially in fine-scale communities[8–10]. Communities with dominating weak or competitive interactions are more resilient than those with higher levels of cooperative interactions[249,330,339,355]. However the six-week-old chicken gut microbiota described here had a high proportion of positive (*c.* 48%) and slightly positive (*c.* 35%) interactions (Figure 1 B and C), which is unusual in a stable community.

There was a significant difference in the strength of the interspecific interactions experienced in the different rarity groups, with the abundant (AC) and global (UP) groups containing much stronger interactions than the moderate or rare groups (Figure 3.4).

There was also a significant difference in the proportional contribution of different types of interspecific interactions to each of the different rarity groups (Figure 3.4, SI Table 3.5). Despite the exact contribution of each interspecific interaction type being different between the rarity groups, the abundant, moderate (MC), global and widespread (AP) groups all showed a predominance of positive interactions (Table 3.5). The localised (RP) and rare (RC) groups, conversely, showed a high proportion of competitive interactions with only a few, small positive interactions (Figure 3.4). The predominance of weak and competitive interactions within the rare groups suggests that these species may play a role in stabilising the microbiota community[249,322,336].

Table 3.5 – Percentage of interspecific interaction types found in the each of the rarity groups, as assessed by local population abundance or geographic spread, within the chicken microbiota. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and defined by the rarity of the focal species. For full explanation of abbreviations see Table 3.2. All values presented are the mean of the three technical replicates, with the associated standard deviations, and given to two decimal places.

Interaction Type	Local Abundance			Geographic Spread		
	AC	MC	RC	UP	AP	RP
+	41.97±24.47	62.23±34.59	0.03±0.02	34.03±19.98	61.31±34.47	0.11±0.05
p	42.40±16.35	22.33±22.23	2.82±0.63	49.45±13.37	24.69±22.63	4.52±0.86
0	1.20±0.84	1.68±0.97	13.49±0.51	1.32±0.82	1.23±0.93	14.01±0.46
n	14.29±7.36	12.99±11.82	72.04±4.17	15.20±5.92	12.21±11.19	70.37±3.71
-	0.15±0.17	0.78±0.61	11.63±4.08	0.00±0.00	0.55±0.54	11.00±3.95

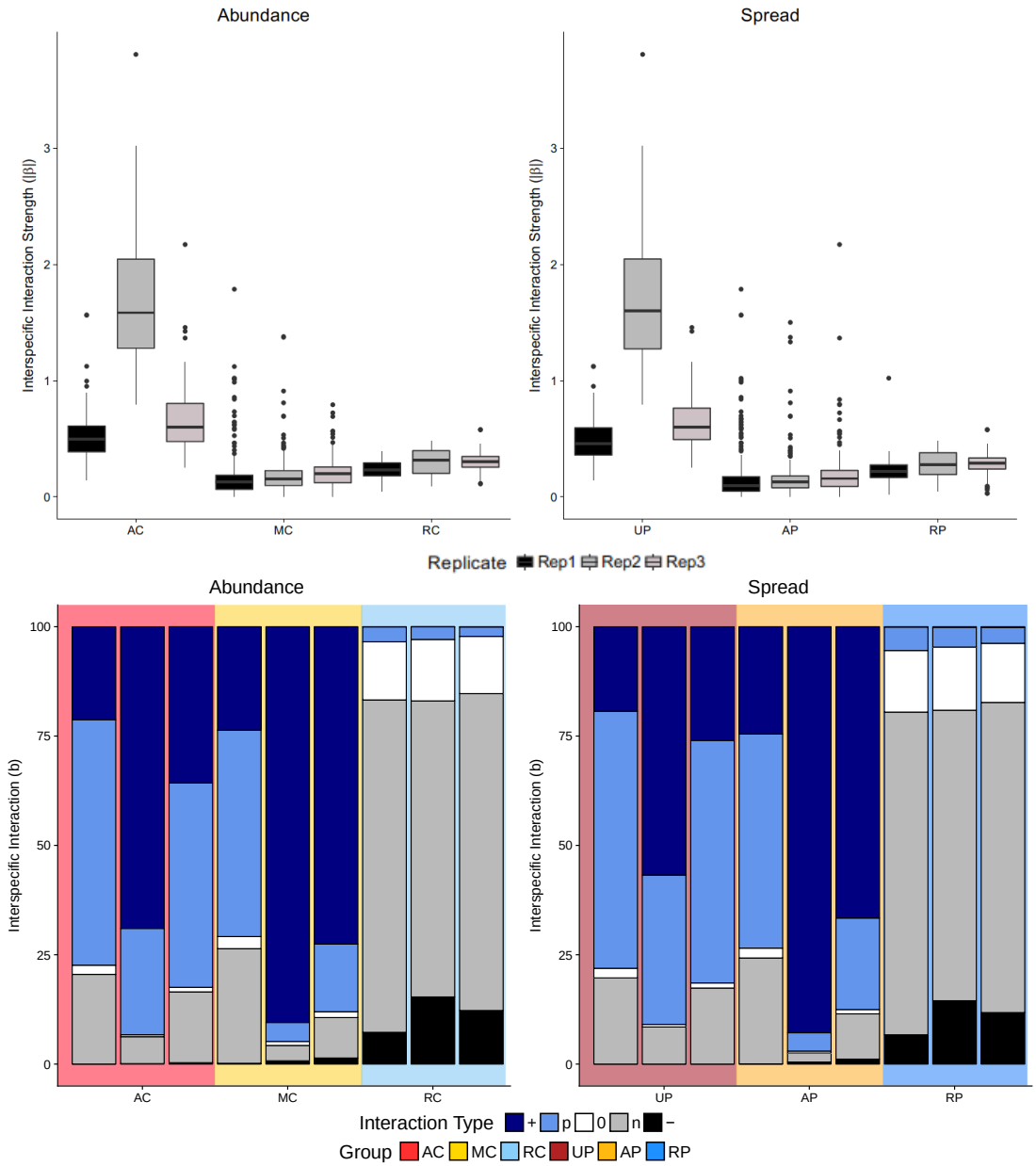


Figure 3.4 – Strengths (upper) and types (lower) of interspecific interactions experienced within different rarity groups, defined by local population abundance or geographic spread. There was a significant difference between the interaction strengths of each rarity group (abundance: p-value < 2.2e-16, R^2 adjusted 0.411; spread: p-value < 2.2e-16, R^2 adjusted 0.365); the boxes of the plot indicate the first and third quartiles of each rarity group interspecific interaction strengths, the whiskers extend to the highest or lowest value within 1.5 times the interquartile range, outliers are indicated as dots. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and are defined by the rarity of the focal species. For full explanation of abbreviations see Table 3,2.

[C.3.1] Interspecific interactions across space

The strength of interspecific interactions within each rarity group varied between locations (SI Figure 3.4, Table 3.6), as did the proportional contribution of the different types of interactions which each rarity group experienced (SI Figure 3.5, Table 3.6).

Table 3.6 – The comparison of linear models of the strength and type of interspecific interactions within each rarity groups in isolation (LM1) or in addition with segment information (LM2). The linear models were of the form LM!; lm(β~Group) and LM2; lm(β~Group+Location). The degrees freedom (DF) and model comparison (AIC) are indicated. The type of interactions are strong positive ('+'), weak positive ('p'), neutral ('0'), weak negative ('n') and strong negative ('-'). All values are rounded to complete integers.

Rarity Basis		Strength			AIC				
	Model	DF	AIC	DF	+	p	0	n	-
Local Abundance	LM1	4	51915	3	791295	385137	26864	515874	41173
	LM2	23	48262	22	660495	342693	24390	356573	26132
Geographic spread	LM1	4	52772	3	874729	426306	28198	531131	38163
	LM2	23	49119	22	739743	382620	25902	372539	23611

The variation of the interspecific interaction strengths between locations resulted from a few distinct segments within each of the rarity groups (SI Table 3.6). There was no correlation between the interaction strengths of any of the rarity groups and the proportional contribution that each group made to the community. There were, however, correlations between the proportion of different interspecific interaction types and the proportional contribution of the different rarity groups to the microbiota. Specifically the rare and localised species showed a significant positive correlation between the proportion of slightly positive or neutral interactions in the group and the proportional contribution of that group to a segment (Table 3.7).

The variation between segments in the proportional contribution of each type of interspecific interaction was correlated with multiple sampling factors (SI Table 3.7), in association with rarity group type. The best explanatory factors varied between the types of interspecific interactions examined. For example, the proportional contribution of strong positive interactions was best explained by rarity group in combination with the interaction

between segment and total community size. In contrast, the proportional contribution of weak positive, or strong negative, interactions was best explained by the rarity group in combination with a site's species richness.

Table 3.7 – The significance of the correlation between the types of interactions found within a rarity group, as defined by local population abundance or geographic spread, and the contribution of that rarity group to its segment community. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and defined by the rarity of the focal species. For full explanation of abbreviations see Table 3.2. Significant correlations are highlighted in grey. All values are given to three significant figures.

Interaction Type	Coefficients	Local Abundance			Geographic Spread		
		AC	MC	RC	UP	AP	RP
+	P-value	0.183	0.761	0.919	0.260	1.36e-02	0.423
	R^2 Adjusted	0.0137	-0.0156	-0.0171	0.00501	0.0850	-0.00597
p	P-value	0.528	0.257	6.92e-05	0.218	0.993	1.12e-02
	R^2 Adjusted	-0.0102	0.00527	0.228	0.00929	-0.0172	0.0903
0	P-value	0.168	0.390	5.94e-06	0.191	5.56e-03	4.18e-03
	R^2 Adjusted	0.0158	-0.00423	0.288	0.0126	0.110	0.118
n	P-value	0.120	0.435	1.49e-02	0.0540	2.84e-03	0.169
	R^2 Adjusted	0.0246	-0.00653	0.0825	0.0464	0.129	0.0156
-	P-value	4.05e-03	2.53e-04	0.461	*NA*	1.41e-04	0.376
	R^2 Adjusted	0.119	0.194	-0.00767	*NA*	0.209	-0.00346

[C.4] Interactions in combined rarity groups

As highlighted in section C.2, the two axes of rarity investigated within this chapter, though dependent, were not restrictive. Although species which were rare along one axis were likely to be rare along the second, this was not necessarily the case. An abundant group species might be globally spread or widespread; global group species might be abundant or moderately abundant (Table 3.8). However, rare species were always localised, and localised species were always rare (Table 3.8, Figure 3.2). The rarity groups that could be broken down into further definitions of rarity might therefore be expected to vary in terms of their combined rarity group in their interspecific interactions.

The interspecific interaction strengths were similarly variable between combined rarity groups (p-value < $2.2e^{-16}$, R^2 adjusted 0.439), with only the combined global groups (AC:UP and MC:UP) not being different from each other (Figure 3.5). The contribution of different

interspecific interaction types to each combined group were significantly different for the strong positive, weak positive and neutral interactions (SI Table 3.8). The contribution of negative interactions was less variable, with the combined global groups showing no difference in proportion of strong negative interactions and the combined widespread groups showing no difference in the proportional contribution of weak negative interactions (SI Table 3.8).

Table 3.8 – Number of unique species within each focal species rarity group pair. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised.

	AC	MC	RC	Total
UP	34	1	0	35
AP	18	400	0	418
RP	0	239	965	1204
Total	52	640	965	1657

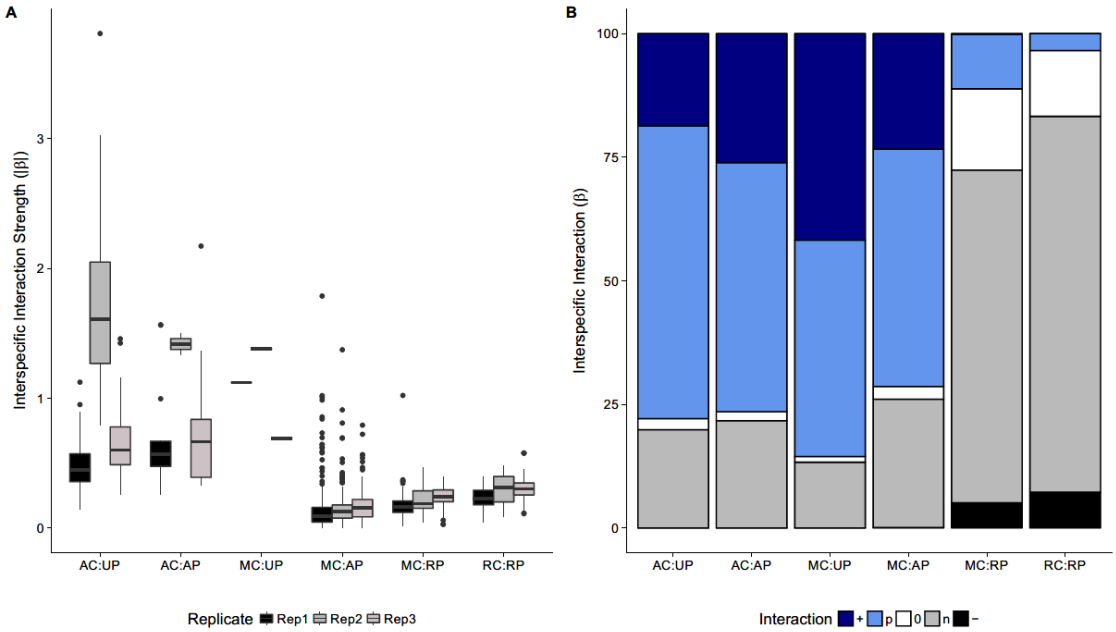


Figure 3.5 – The interspecific interaction strengths (A) of each combined rarity group, along with the proportional contribution of different interspecific interaction types within each combined rarity group (B). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and defined by the rarity of the focal species. For full explanation of abbreviations see Table 3.2. The boxes of the boxplots indicate the first and third quartiles of each rarity group interspecific interaction strengths, the whiskers extend to the highest or lowest value within 1.5 times the interquartile range, outliers are indicated as dots.

As in the case of the individual rarity groups, including location improved the fit of the models for interspecific interaction strengths and interspecific interaction types (SI Table 3.9).

[C.5] Interspecific interactions between rarity groups

As highlighted in section B.3, not all interactions could be estimated between all pairs of species within the grid. For example, global species could not act as competitor species using this method (Figure 3.6). Given that interactions occur between only two species, not necessarily of the same rarity, it is worth considering whether there is a difference in the interaction environment of a focal species with a competitor species of a given rarity group.

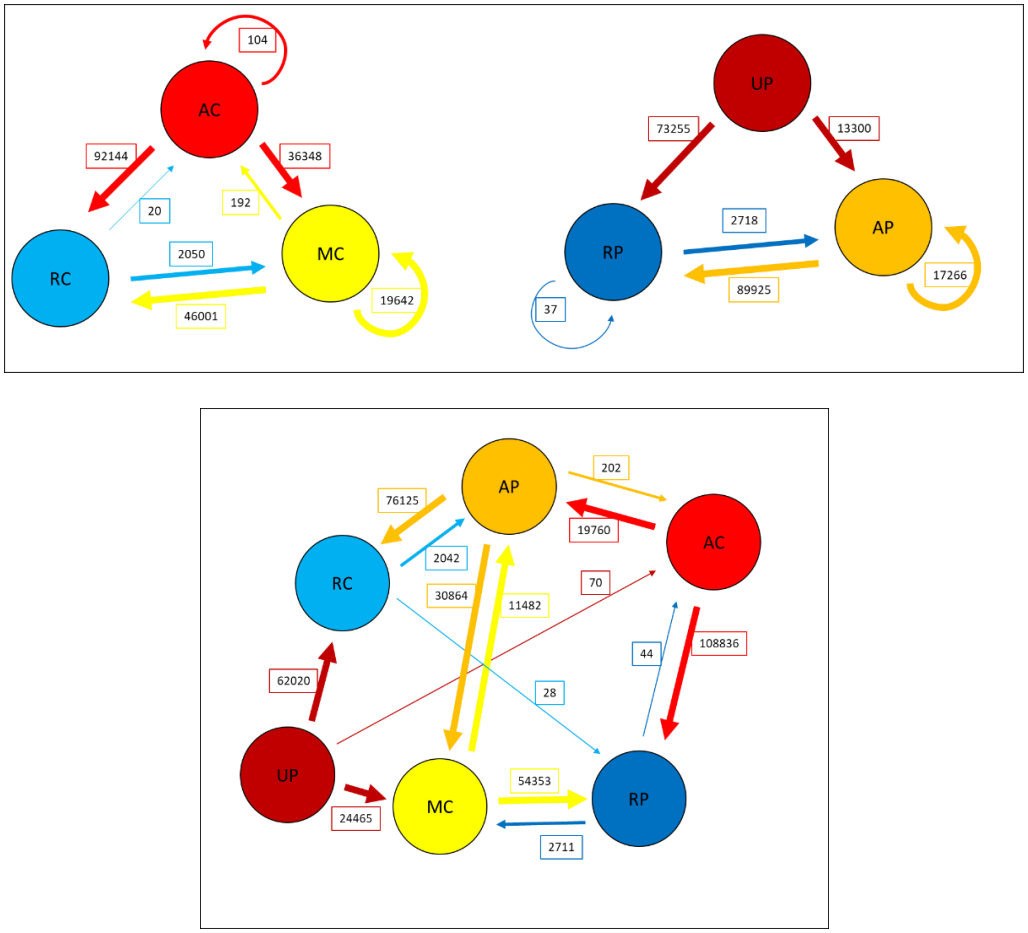


Figure 3.6 – Schematic of the number of species from each rarity group interacting with other rarity groups, either within rarity definition (upper) or between rarity definitions (lower). Rarity groups were defined by either local population abundance or geographic spread. Arrows go from the focal species rarity group to the competitor species rarity group, and are weighted in width by the number of species involved in that interaction with the exact number indicated. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised.

The majority of competitor species were moderately abundant, widespread, rare or localised (SI Table 3.10). Qualitatively, rare species did not compete with other rare species and there were relatively few interactions between the rare and localised or between localised species (Figure 3.6).

The strength of interspecific interactions varied between focal species rarity groups (Figure 3.5, SI Table 5) and within each focal species rarity group by competitor species rarity (Figure 3.7, SI Table 3.11). Interactions within the abundant and global focal species groups were strongest, although interactions within and between these groups were relatively weak. Within each focal species, rarity group interactions with rare and localised competitor species were different from most other rarity groups (SI Table 3.11) with the rare and localised competitors associated with stronger interspecific interactions.

The types of interspecific interactions experienced within each focal-competitor species rarity group pair also varied in the majority of cases (Figure 3.8, SI Table 3.12) though not within interaction types that contributed relatively little to any given focal-species competitor pairs (SI Table 3.13). Specifically, there was no influence of competitor species upon strong or weak negative interactions (- or n, respectively) in abundant or global focal species or within strong positive interactions (+) of rare focal species.

Within each focal species group there was a significant effect of competitor species rarity on interspecific interaction strength and type (SI Table 3.11, SI Table 3.12). Including location as an explanatory factor further improves the model fit, with respect to interspecific interaction strength, though the model is only marginally better than the fit with location and without competitor species rarity (SI Table 3.13).

The combination of competitor species rarity with location best fit the model of interspecific interaction type with the exception of strong negative interactions (-) within global focal species (UP) and strong positive interactions (+) in the rare focal species (RC) group (SI Table 3.14). Within the global focal species the proportion of strong negative interspecific interactions was most dependent upon the competitor species rarity, whilst the rare focal

species proportional strong positive interaction contribution was best explained by location in isolation.

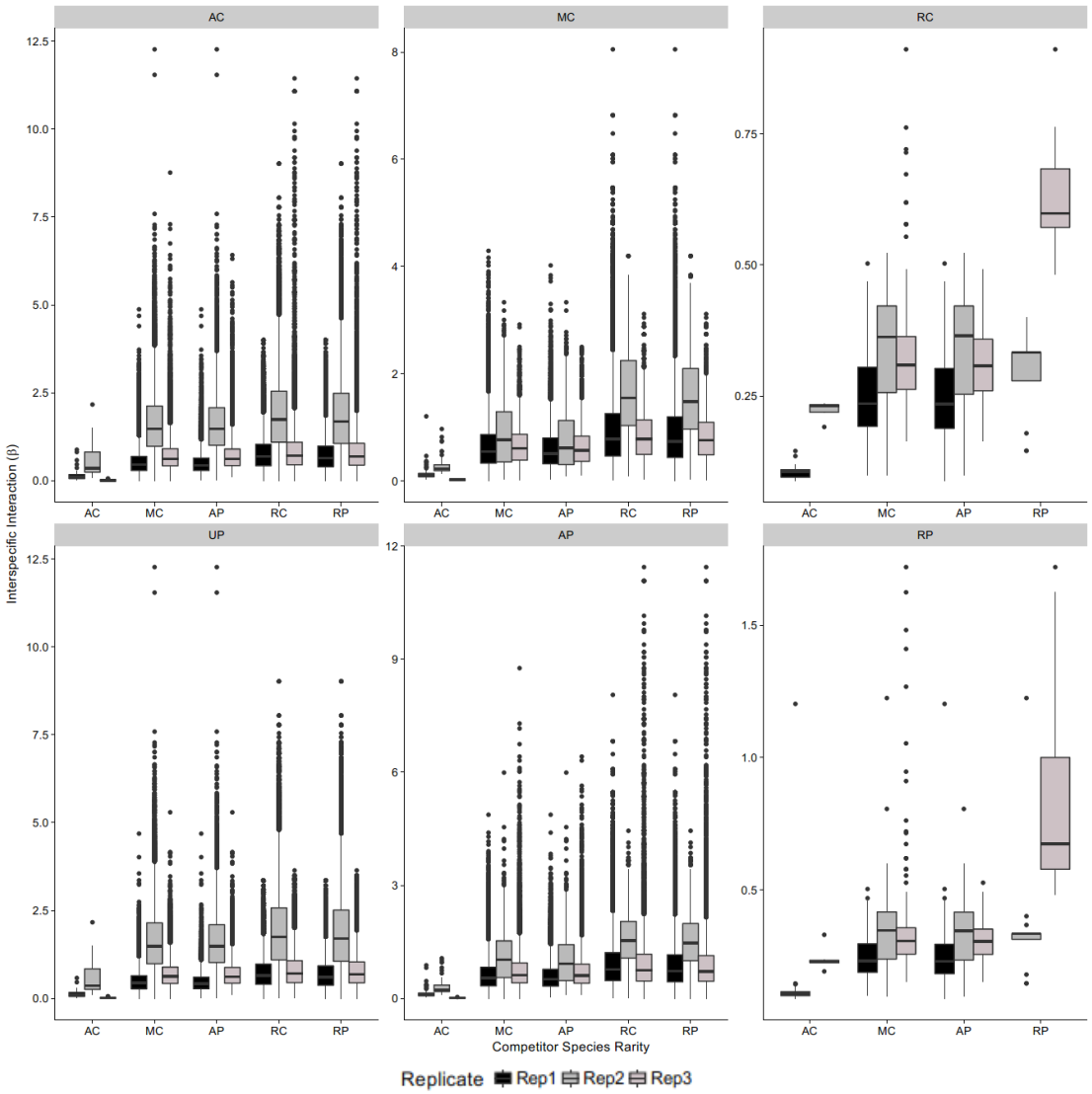


Figure 3.7 – The strength of interspecific interactions between focal-competitor species rarity groups. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. The boxes of the plots indicate the first and third quartiles of each rarity group interspecific interaction strengths, the whiskers extend to the highest or lowest value within 1.5 times the interquartile range, outliers are indicated as dots. All three technical replicates are indicated.

[D] Discussion

The gastrointestinal microbiota is a complex community intrinsic to the evolutionary fitness of its host organism. The effect of the microbiota on host health is dependent upon its complexity[42–45] and stability[46,311,312], however complex communities are not intrinsically stable[127]. Aligning the observed associations between complexity and stability within natural

communities with theoretical predictions is a central question in ecology with explanations including species interactions[313,323].

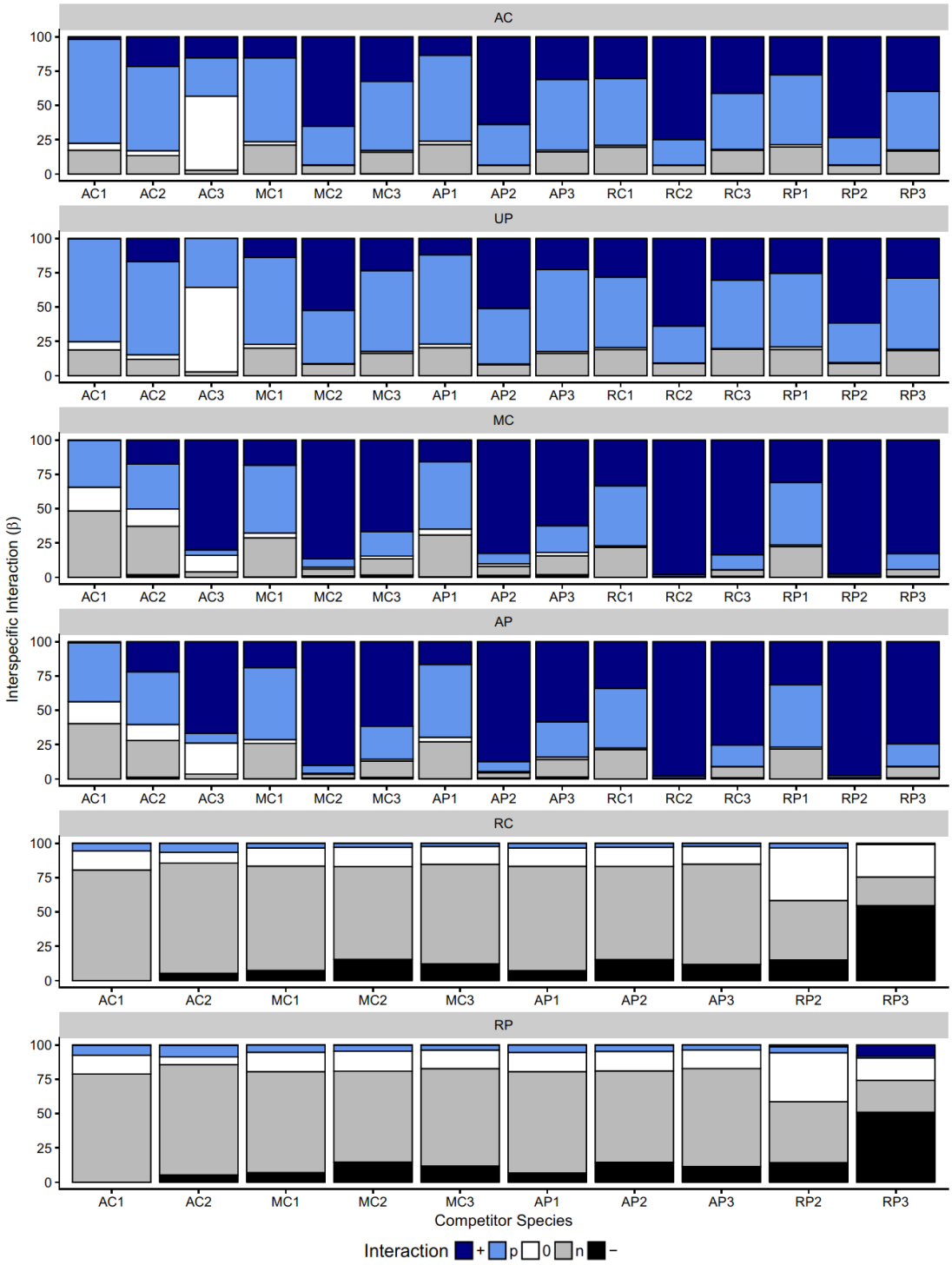


Figure 3.8 – The proportional contribution of different interspecific interaction types as a function of the focal and competitor species rarity groups. Each plot represents a different focal species rarity group. Within each plot the x-axis is a combination of the competitor species plot and the technical replicate number (e.g. competitor group AC within technical replicate 1 is labelled AC1). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific

interactions, ‘p’ to weak positive, ‘0’ to neutral, ‘n’ to weak negative and ‘-‘ to strong negative. For full explanation of abbreviations see Table 3.2.

In this chapter the natural perturbation experiment of separating the caecum of a L6 six-week-old male chicken into segments was used to facilitate estimates of interspecific interactions within the host organism’s microbiota. By separating the species found within the microbiota into rarity groups based on either total population size or geographic restriction, it was possible to show that rare species experience different interaction environments to abundant species. It was also possible to characterise the types of interactions found within the chicken gastrointestinal microbiota.

[D.1] Interactions within communities

The strength[127,328–330], type[249,322] and structure[331–333] of species interactions play a role in facilitating the stability of complex ecological communities. Stable communities within nature are dominated by weak interspecific interactions[328,330] (particularly within long interaction chains, and negative feedback loops[329,332]) whereas intraspecific interactions would be expected to be strong in more complex communities[127]. The distribution of interaction strengths within the chicken gastrointestinal microbiota presented here was consistent with previous literature[329,330,336], showing a skew towards weak interspecific interactions and a mild skew towards strong intraspecific interactions (Figure 3.1), although the range of the interspecific interaction data is more restricted than in previously reported communities[337].

The discrepancy between the interactions presented here and those measured in traditional ecological communities could arise as a result of how the interactions are measured. Interaction measurements in natural ecological communities are largely constrained to measurements of biomass[342]. Conceptually, biomass measurements represent the βy parameters of Equation 3.2 rather than the sole β interaction as estimated in this work. Direct comparisons to the ecological literature, therefore, would be better facilitated using βy values as a proxy for interaction strength and could be integrated into future work.

Both theoretical and observed communities have demonstrated that an increased proportion of negative interactions in an ecosystem increases the system's stability[249,322,336]. However the microbial community investigated here appeared to be dominated by positive interactions (Figure 3.1). The predominance of cooperative interactions within the microbiota suggests that the community should be unstable in the absence of additional stabilising factors, such as host-mediated spatial structuring.

[D.2] Spatial patterning with the microbiota

Chapter 1 predicts that community diversity can be maintained by host-mediated spatial structuring in the gastrointestinal microbiota[99]. Spatial variability can facilitate community stability through the generation of local communities with different selection pressures[338], isolating small groups of species[249,313], and creating refugia where less advantageous species can survive, in the absence of more competitive contemporaries[230,231,260]. These communities can then act as source populations for other communities where specific species may be disadvantaged.

Differential selection pressure upon local populations may arise due to resource heterogeneity (e.g. the carbohydrate landscape of the gut). This can have direct effects (e.g. which species can access resources) and indirect effects (e.g. through modifying the interaction environment) that influence community composition. Chapter 2 showed preliminary evidence for the existence of spatial heterogeneity within both the microbiota and adaptive immune populations, and thus local communities of microbes and differential selection from 'predators'. In this chapter, the strength and types of interspecific interactions were found to vary between gastrointestinal segments (SI Figure 3.4 and SI Figure 3.5), which may generate, or result from, differences in community composition and selection between sites.

[D.3] Interaction environments vary between rarity groups

Every observed community has species that are rare and which are common. Within most communities, as in the microbiota, relatively few species dominate the population numerically and a large number of species are present at lower frequency. This consistent

structure raises the question; why are species rare? Do rare species remain rare as they struggle to exist just above extinction or do rare species persist by limiting the expansion of their more abundant contemporaries?

The abundant and global species experienced much stronger interspecific interactions than any other rarity group, though as competitor species they provided the weakest interactions regardless of focal species rarity (Figure 3.8). The strength of interspecific interactions appears to be driven predominantly by rarity groups defined through local population abundance, though the types of interspecific interactions were mainly influenced by a species' geographic spread (Figure 3.5).

Surprisingly, both the abundant and moderately abundant groups (regardless of rarity definition) experienced predominantly positive interactions (Figure 3.4). Within the moderately abundant species almost all the competitive interactions were experienced in relation to abundant competitor species, with rare species contributing almost exclusively positive interactions. Almost all negative interactions within the gastrointestinal microbiota were restricted to the rare focal species groups (Figure 3.4) with rare-rare interactions, though unusual, most strongly negative (Figure 3.8). The increased proportional contribution of rare species in segments with relatively more positive and neutral interactions suggests that the predominance of competitive interactions within the rare groups could limit their local abundance.

[D.4] Limitations of methodology

The limitations of this mechanism of estimating species interactions include the assumption of equilibrium, and the number of species for which interaction estimates could not be calculated (see B.3). In order to use the spatial structure of the caecal grid as a natural perturbation experiment it was necessary to assume the microbiota community was at equilibrium which may not be true[88–90] (see Chapter 4 D.2). The assumption that intraspecific and interspecific interaction strength is a constant value may be similarly questionable[355], given

that interaction strength may be influenced by the density of the interacting species and the surrounding community composition.

As the method of estimating interactions between species required that a focal species exists in both the presence and absence of its competitor species, interactions with ubiquitous species as competitor species were not possible. The work presented within this chapter could be extended to include these interactions by using linear convergence to estimate the intraspecific and interspecific interactions between consistently co-locating species (see Appendix 5). Estimates for species that never co-locate will still not be possible. Given that species that never co-exist are likely to be highly competitive (or requiring completely distinct environmental conditions) the lack of these species interaction estimates may explain the skew towards positive interactions within the community presented here.

The interactions within this chapter were assessed in the absence of any functional information about the species involved or predator-prey dynamics. The structure of interactions within communities (e.g. between trophic levels) are often important in community stability[331–333]. Specifically, predator species being rare relative to their prey reduces the weight of feedback loops within communities, facilitating their stability[332]. The absence of any information with respect to higher order interactions within this framework is therefore another limitation of the chapter.

[D.5] A role for rarity in stabilising communities

Despite the limitations of this methodology, there is a clear distinction in the strength and types of interactions experienced by species in different rarity groups. That theoretical models have linked a predominance of weak and competitive interactions with increased community stability[249,322,336] suggests that the relatively increased number of weak and competitive interactions within the rare species may act as a stabilising force within the gastrointestinal tract.

[E] Conclusions

The beneficial effects of the gastrointestinal microbiota are depend on both its diversity and stability, a trade-off that has continued to receive attention within ecology. Conversely the importance of rare species in community dynamics is relatively unstudied, though much literature has been devoted to attempts at conserving these species. Using a grid structure sampling method within the gastrointestinal microbiota as a natural perturbation experiment, it was possible to estimate species interactions and consider the role of rare species within the microbiota. A species' interaction environment is dependent upon both its relative rarity and the local community in which it was found. Specifically, rare species were much more likely to experience weak and negative interspecific interactions, suggesting a stabilising role for these species within the microbiota community.

Chapter 4

Repeatability of fine-scale structure in the adaptive immune and microbiota components

[A] Introduction

Both the adaptive immune system and the microbiota are unique to their individual hosts. The composition of a host's microbiota depends upon microbe colonisation[42,54,356] and is modified during host development[88–90,351]. The adaptive immune system is similarly unique as it is formed over the course of an individual's life in response to pathogen stimulation. The inter-individual variation within both these communities' compositions can arise from a variety of factors and thus mean community dynamics may similarly change between individuals.

[A.1] Microbiota variability

Microbiota composition is influenced from birth, depending upon birth method and feeding[54]. The infant microbiota is even more variable between individuals than that of adults, and is highly vulnerable to disruption (e.g. by antibiotics)[54,88,327]. During maturity the microbiota can be modified due to perturbation (e.g. antibiotics[349–351,357] or infection[195]) and dietary shifts[198,242,358], but remains largely stable[324,325].

The time at which the microbiota is sampled[83,88], the host's genetic background[236,237], and rearing conditions[238,239], are all likely to affect microbiota composition.

[A.2] Adaptive immune variability

T-cell receptor (TCR) formation in birds is the same as in mammals; gene segments from the constant (C), diversity (D), junction (J) and variable (V) regions are rearranged to form a single TCR. The random rearrangement of these genes, in combination with the stochastic

addition and deletion of nucleotides around the joining sites, leads to a hyper-diverse TCR repertoire which even in its naïve state varies between individuals.

Unlike in mammals, and the avian TCR, the chicken B-cell receptor locus is remarkably sparse in terms of V, D and J segments for rearrangement (see General Introduction C.3). However the BCR repertoire of birds is not restricted relative to the mammalian repertoire; instead the hyper-diversity is generated through the random insertion of fragments of V-region pseudogenes into the rearranged BCR.

The adaptive immune repertoires of birds are initially inactive after hatch, maturing at roughly five-to-eight weeks of age[81,252]. The process of adaptive immune receptor formation (see General Introduction C.3), however, means that the immune repertoire has the potential to vary substantially between individuals.

[A.3] Temporal variation

Although the gut microbiota is relatively stable within individuals throughout their lives (see Chapter 3 A.4)[88,324,325], hosts are not born with an intact adult microbiota. Instead the microbiota develops with its host, becoming more diverse and stabilising as the organism ages[88–90,351]. The age at which the microbiota stabilises varies between species, contemporaneously with the maturation of the adaptive immune system. In birds the microbiota stabilises between six and eight weeks of age[88–90].

Post-hatch, the size and complexity of the intraepithelial lymphocytes and lamina propria increases[81], until peak lymphocyte numbers are reached at *c.* eight-weeks of age[359]. During this development period the adaptive immune system undergoes substantial changes, including the colonisation of the periphery with naïve T-cells and B-cells and their subsequent maturation.

[A.3] Chapter 4 aims

This chapter contains a comparison between the preliminary L6 bird and four JH wild-type birds. The JH wild-type birds were from a B-cell knock-out experiment in 6 week-old chickens, examining immune and microbiota variation in wild-type, heterozygote and knock-

out genotypes (see Chapter 5). The grids taken within this experiment were all 1 mm^2 segments, and varied in total segment number, so true comparison to the preliminary bird was restricted by both genetic background and sampling strategy.

[B] Methods

[B.1] Sample collection

Four wild-types on the same mixed genetic background (Commerical Brown and White Leghorns chickens and Minnesota Marker Line chickens)[360], were sacrificed as part of an experiment looking at the effect of B-cell absence upon the microbiota (Chapter 5). The animals were provided in collaboration with Professor Bernd Kaspers' laboratory (Ludwig-Maximilians University, Munich) and Dr Benjamin Schusser.

[B.1.1] Grid creation

After sacrifice, the intestinal tract was removed from each bird and grids taken *in situ*. A 1 mm^2, 4 column, grid was taken from the right caecum of each bird. The columns of the grid were taken moving from the anterior to posterior of the caecum (A-D respectively), then the rows of the grid excised at 1 mm intervals until the whole circumference of the caecum had been sampled (numerical notation of rows, alphabetic notation for columns). A maximum grid size of 4 x 12 was retrieved.

The segments were stored in 100 µl of RNAlater at -20 ^{o}C overnight before shipping to the UK and stored at -80 ^{o}C until processing.

[B.2] Tissue RNA/DNA extraction

RNA and DNA were extracted from the samples using QIAGEN 96-well AllPrep DNA/RNA extraction kit, according to the manufacturer's instructions. Plate extractions were designed to contain an individual from each treatment (see Chapter 5) whilst minimising the number of individual grids split over a two 96-well plates.

The tissue was disrupted and homogenised using a bead beater with 0.1 mm sterile glass beads in 350 µl Buffer RLT for 2 minutes, in sets of 8 samples. The weight of the tissue

was validated as below the capacity of the columns for each sample. The processed samples were eluted in 45 µl RNase-free water and 60 µl (pre-heated to 70 °C) Buffer EB, in two elution steps, for the RNA and DNA respectively.

[B.3] cDNA synthesis

5' RACE cDNA was made from the AllPrep extraction RNA, according to the SMARTer reverse transcriptase protocol (as in Chapter 2 B.3). Master mixes were set up for each individual bird grid from every 96-well extraction. Sample extraction and synthesis were validated by β-actin PCR (Chapter 2 B.4).

[B.4] DNA amplification

Samples within each 96-well plate were amplified, in triplicate, using randomly assigned non-repeating barcodes for each population type (microbiota [16S], IgA-producing B-cells [BCR] and T-cells [TCRβ]). Samples that were not amplified with their initially assigned barcode were assigned a subsequent (non-repeated) barcode. Each sample's technical replicate was amplified with a different barcode. Barcodes and primers were designed, and PCR conditions used, as in Chapter 2 B.4.

[B.5] Sequence creation and processing

Samples from the wild-type birds were run upon 4 MiSeq runs and extraction plates. All samples from a single 96-well extraction plate were pooled into libraries by technical replicate (see Chapter 5 B.3 for full details).

[B.5.1] Processing raw sequences

Initial sequence processing was performed as in Chapter 2 B.6 and sequence quality assessed as in Chapter 2 B.7. 16S sequences amplified within MiSeq22 library 3 (wild-type bird WT5) were significantly degraded (SI Figure 4.1) and thus excluded from further analyses. This appears to be a problem with the PCRs rather than the MiSeq library itself as there was no evidence of excessive degradation in the immune populations.

As a comparison to Chapter 5, only the raw data from the BCR populations is presented here (Chapter 5 B.3). There were differences between the raw and trimmed BCR datasets, where these changed the results of the analysis the deviation is indicated.

[B.5.2] 16S sequence identification

Microbial sequences were processed using the Divisive Amplicon Denoising Algorithm 2 (DADA2)[268] pipeline, using the DADA2 defaults, starting with the introduction of demultiplexed files. Chimera sequences were removed to reduce inter-replicate variation. DADA2-produced independent sequence variants (ISVs) were taxonomically identified by comparison with the Greengenes 16S rRNA database (version 13_8)[269,270] using the RDP classifer[264].

[B.6] 16S contamination control

118 controls were run over the 4 MiSeq runs, comprising 18 extraction blanks and 100 water controls. Extraction blanks were controls processed through from the RNA/DNA extraction stage as samples but without sample addition. The blanks were amplified by PCR and sequenced in triplicate for each population. All extraction blanks contained at least 1 independent sequence variant (ISV). 2 water controls were run per PCR reaction. Environmental contamination during PCR amplification was controlled for as in Chapter 2 B.5. Only 17 water controls sequenced contained contaminating ISVs, within these water controls none of the ISVs were present in more than 50% of the water controls.

[B.6.1] Contaminant identification thresholds

663 ISVs were identified within the blanks (of all types), 9 of which were not present in any of the samples and were therefore immediately removed from all the MiSeq runs. ISVs that were present in both the samples and extraction blanks were removed if they were found in more than 50% of the blanks. ISVs that were found more than incidentally (>10 sequences) in any one extraction blank or water control were designated ‘contaminants’ and removed. Within this study, 13 ISVs (of a total 1609 ISVs) were designated contaminating sequences and removed from further analysis.

354 samples contained at least one contaminating sequence. In a single sample, the highest count of a single contaminating sequence was 2216 reads and, though only four of the contaminating sequences were found at more than 10 counts in any given sample, the majority of samples (344/354) in which contaminating sequences were found contained one of these abundant contaminating sequences.

[C] Results

The preliminary experimental work presented in Chapter 2 was highly repeatable across technical replicates but only performed within a single bird. The consistency of the result – that fine-scale analysis can be important for distinguishing community dynamics and the existence of spatial structuring within the immune and microbiota populations – is therefore investigated here.

The JH wild-type birds used here as a comparison to the preliminary L6 bird were not sacrificed specifically for this purpose so there are flaws in the design of the comparison. Specifically there is only one preliminary bird that, although sacrificed at the same age as the JH wild-type birds, was of a distinct genetic background. The grid segments taken within the JH wild-type birds were also taken at 1 mm^2 intervals rather than the 4 mm^2 segments collected in Chapter 2. With these caveats, the same analyses performed upon the preliminary bird were replicated within the JH wild-type to explore the consistency of the initial results.

[C.1] Non-functional sequences

The adaptive immune populations in all birds contained both functional and non-functional sequences (SI Table 4.1).

Within the JH wild-type TCRβ population, 11-19% of the unique sequences, and 12-19% of the clones, were non-functional. There was a significant effect of bird identity upon the percentage of both TCRβ clones (p-value $5.77e^{-14}$, R^2 adjusted 0.998) and TCRβ sequences (p-value $8.15e^{-05}$, R^2 adjusted 0.848) that were non-functional, which was largely due to JH wild-type bird WT1. Although the preliminary L6 bird had fewer non-functional sequences than

some of the JH wild-type birds, WT4 and WT5, there was no difference between the preliminary L6 bird and the JH wild-type birds as a group.

Within the JH wild-type immunoglobulin A (IgA)-producing BCR population 20-38% of the unique sequences, and 25-39% of the clones, were non-functional. However, unlike the TCRβ population, there was no difference in the proportion of non-functional BCR sequences or BCR clones between birds. Regardless of which adaptive immune population was considered, there was no difference between the JH wild-type birds collectively and the preliminary L6 bird in the proportion of non-functional sequences or clones retrieved.

[C.2] Large-scale heterogeneity in wild-type population structures

The microbiota populations in all of the JH wild-type birds were skewed toward rarity. There was no difference between the preliminary L6 bird and the wild-type JH birds in terms of the proportion highly abundant microbes (Figure 4.1, SI Table 4.2). However there was variation between the birds in terms of all other abundance groups, with the preliminary L6 bird having more moderately abundant and fewer moderate species than the JH wild-type birds (Figure 4.1 orange and yellow respectively).

As in the preliminary L6 bird, the immune populations were skewed further toward rarity than the microbiota populations in all JH wild-type birds (Figure 4.1), with the IgA-producing B-cell receptor population containing the highest proportion of rare sequences. The TCRβ population showed variation between birds in the proportional contribution of all the abundance groups (SI Table 4.2). JH wild-type bird WT1 appeared to be driving the difference in TCRβ population structures, with a significant difference to all other birds in every abundance group except group 4 (Figure 4.1 yellow).

There was no evidence of bias within the TCRβ CDR3 length distributions in either the preliminary L6 bird or the JH wild-type birds at a large-scale. All large-scale spectratypes, regardless of bird, approximated normality, though the mode of the distribution varied (least squares values SI Table 4.3, SI Figure 4.2), possibly reflecting a difference in TCRβ repertoires between the wild-type birds.

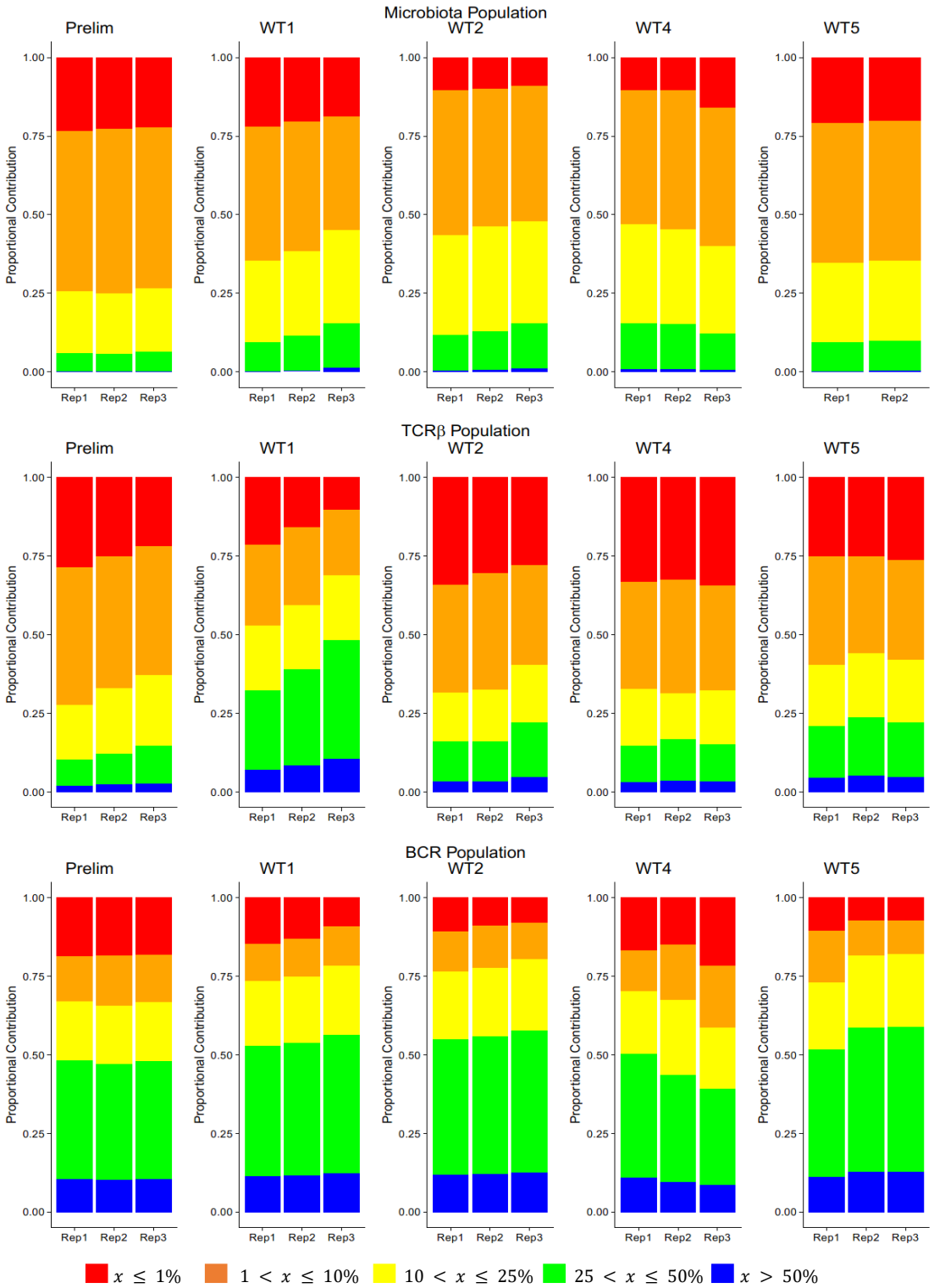


Figure 4.1 – Population structures of the microbiota, T-cell receptor β and IgA-producing B-cell receptor immune repertoire for each wild-type (WT) JH bird compared with the preliminary wild-type L6 bird. All three technical replicates are indicated. Within the 16S microbiota population replicate 3 of WT5 sequence quality was poor (SI Figure 12) so this was excluded from the analysis. Abundant species/clones are in red, moderately abundant species/clones in orange, moderate species/clones in yellow, moderately rare in green, and rare in blue.

There was a significant difference between the birds in the extent to which the spectratypes approximated the normal distribution (p-value 9.24e-06, R^2 adjusted 0.930). The preliminary L6 bird and JH wild-type bird WT1 appear to drive the difference in the spectratypes as they deviate significantly less from normality than the other JH wild-type birds.

The population structure of the BCR dataset showed a similar pattern to that of the TCRβ population; there was a significant effect of bird upon proportional contribution of all abundance groups (except group 3) that was driven largely by JH WT4 and to a lesser extent the preliminary L6 bird (SI Table 4.2).

[C.3] Fine-scale variation in population structure

The L6 preliminary bird in Chapter 2 showed no evidence of non-normal CDR3 distributions at a large-scale (Chapter 2 D.2), which was replicated in the JH wild-type birds. Within the preliminary bird these dynamics were distinct, in terms of spectratype shape and the proportional contribution of differentially abundant immune clones, at fine scales.

[C.3.1] Microbiota abundance across space

The preliminary bird showed no variation in any of the microbiota population structure abundance groups and this was similarly true in the JH wild-type birds WT1 and WT5. JH wild-type WT2 showed segmental variation in the proportional contribution of moderately rare individuals (group 5; p-value $2.50e^{-07}$, R^2 adjusted 0.672), whilst JH wild-type 4 shows significant variation in almost all abundance groups (group 3; p-value $4.27e^{-02}$, R^2 adjusted 0.548, group 4; p-value $2.08e^{-02}$, R^2 adjusted 0.573, group 5; p-value $2.74e^{-03}$, R^2 adjusted 0.629 and group 5; p-value $1.17e^{-02}$, R^2 adjusted 0.590).

[C.3.2] Immune population structures between segments

There was significant variation in the population structures between segments for the TCRβ population of each bird except WT4 (SI Table 4.4). For the preliminary bird, the proportional contribution of each abundance group varied between segments, while the variation between locations in the JH wild-type birds (especially birds WT1 and WT2) was restricted to specific abundance groups (SI Table 4.4). For example in WT1 the variation

between segments was predominantly in the most abundant and rare groups and largely driven by segment D2 varying from segments B2 and A4 (Figure 4.2).

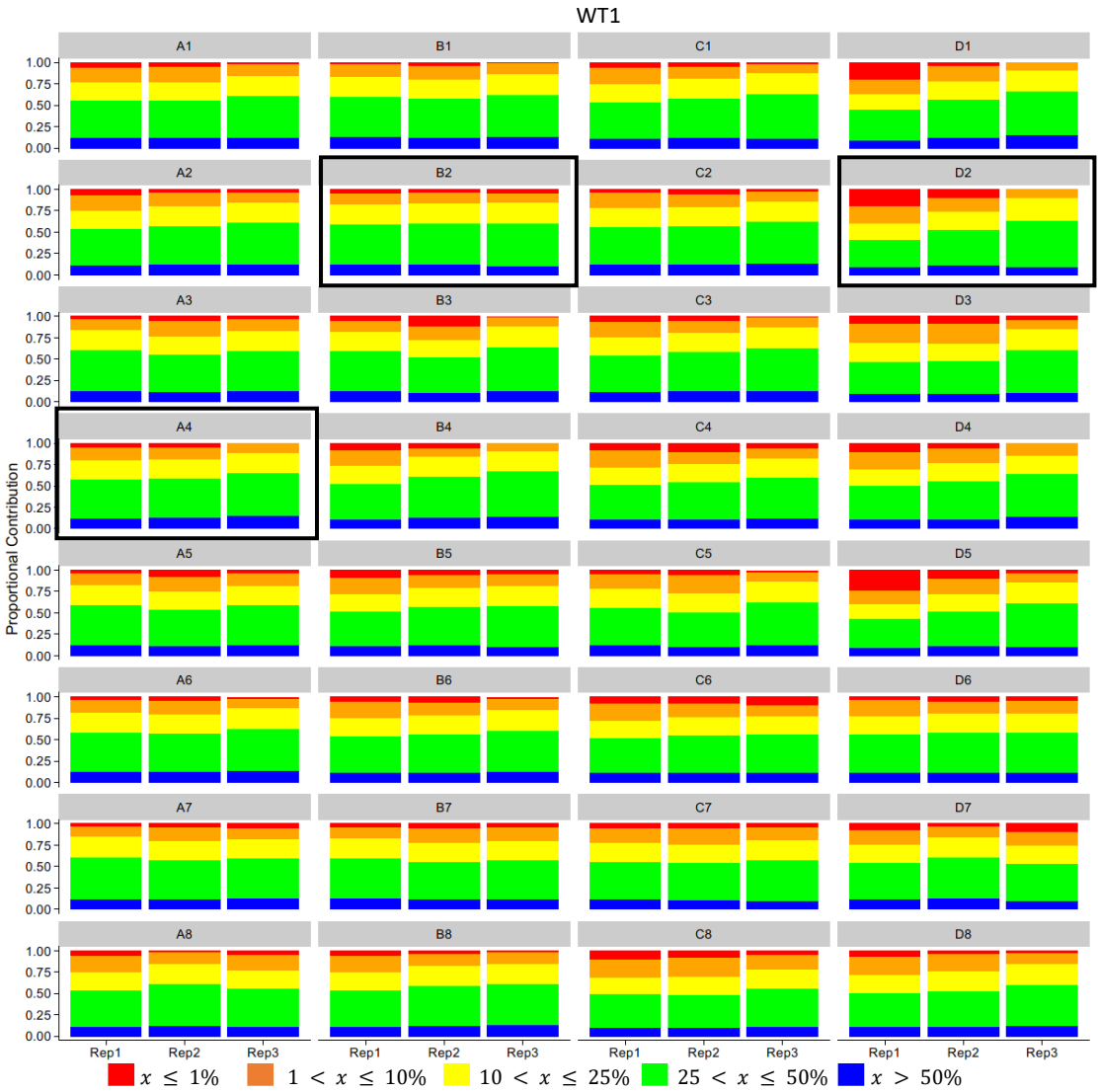


Figure 4.2 – The proportional contribution of different population abundance groups of the TCRβ population between segments within the wild-type bird WT1. Abundant species/clones are in red, moderately abundant species/clones in orange, moderate species/clones in yellow, moderately rare in green, and rare in blue. Segments referred to in the text are highlighted in black. All technical replicates are indicated.

Unlike in the preliminary L6 bird (see Chapter 2 D.2), there was no evidence of segmental variation within any of the JH wild-type bird BCR population abundance groups, except in WT5's moderate clones (group 4: p-value $3.45e^{-02}$, R^2 adjusted 0.512). The variation within WT5 was driven almost exclusively by segment A2 showing an expansion of moderate clones relative to most other segments.

[C.3.3] TCRβ CDR3 spectratypes

In the preliminary bird there was a significant effect of location upon the length distributions of the TCRβ population. This effect was only replicated in the JH wild-type bird WT4, and the association between spectratype shape and segment distance was not repeated in any of the wild-type birds (SI Table 4.5).

[C.4] Large-scale population compositions

As would be expected the immune populations were more diverse than the microbiota communities in all the wild-type birds. In each bird the BCR population is the most clonally diverse, as reflected by the increased proportional contribution of rare species within the B-cell repertoire.

The microbiota population showed a significant effect of bird identity upon all three diversity estimates interrogated (Figure 4.3, Table 4.1) with a significant difference between the preliminary bird and the JH wild-types. The TCRβ population was also significantly different in all three diversity estimates (Figure 4.3, Table 4.1) predominantly due to the difference between the preliminary and JH wild-type bird WT1 and all other JH wild-type birds. Despite WT1 being similar to the preliminary bird, there was a significant difference between the preliminary and JH wild-type birds, collectively, in both species richness and Shannon's diversity (Table 4.1), though not in the Simpson's diversity estimates.

The raw BCR dataset was similarly variable between birds in each of the diversity estimates (Figure 4.3, Table 4.1). Unlike the TCR population, the differences in the BCR population appeared to emerge from JH wild-type birds WT2 and WT5 differing from the other two JH wild-type birds and the preliminary L6 bird. As would therefore be expected, there was no collectively difference between the preliminary L6 bird's BCR diversity and the diversity of the JH wild-type birds.

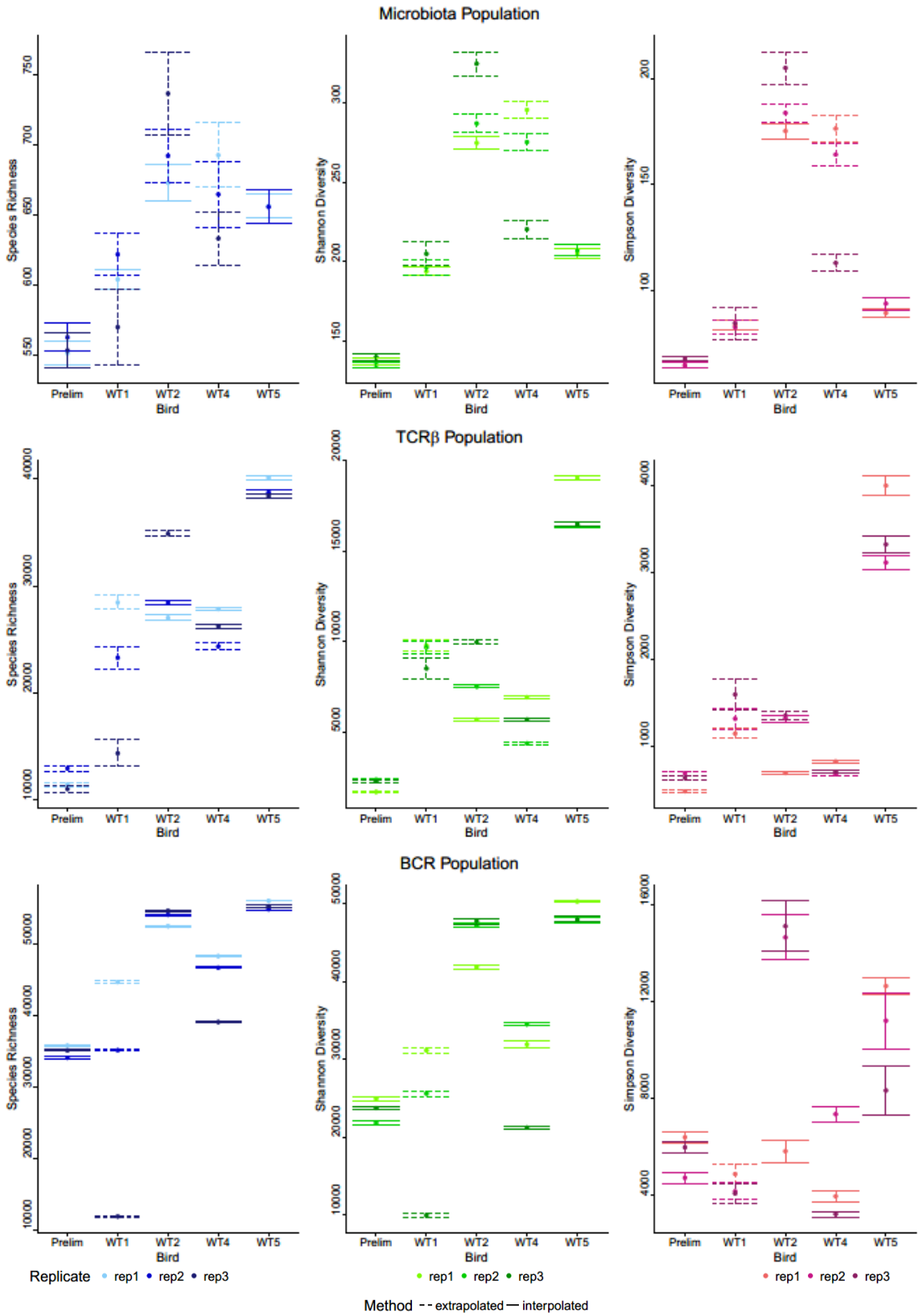


Figure 4.3 – Diversity estimates for the preliminary (L6) and JH wild-type (WT) birds. Species richness (blue), Shannon's diversity index (green) and Simpson's diversity index (red) of each technical replicate are displayed for each bird. Estimates were calculated to the same end point either through interpolation or extrapolation as indicated. The standard error of the estimates (95% confidence interval) is shown.

Table 4.1 – The significance of the effect of bird upon diversity estimates and difference between the preliminary L6 bird and the JH wild-type (WT) birds as a collective. Estimates were calculated to the same end point either through interpolation or extrapolation as indicated in Figure 4.3 using the R package iNEXT. The linear models used to generate the p-value and R^2 adjusted were of the form lm(Estimate~Bird). Values are given to three significant figures. BCR data presented is the raw dataset.

		Effect of Bird			Prelim *vs* WT	
Population	Estimate	P-value	R^2 Adjusted	Different Birds	P-value	R^2 Adjusted
Microbiota	Species Richness	3.69e-04	0.879	Prelim-WT2/WT4/WT5 WT1-WT2	3.68e-03	0.518
	Shannon's Diversity	1.00e-04	0.910	Prelim-WT4/WT5 WT1-WT4 WT2-Prelim/WT1/WT5	2.66e-03	0.542
	Simpson's Entropy	6.31e-05	0.919	WT2/WT4-Prelim/WT1/WT5	4.05e-02	0.305
TCRβ	Species Richness	4.07e-03	0.794	Prelim-WT2/WT4	1.44e-03	0.654
	Shannon's Diversity	7.95e-04	0.864	Prelim-WT1/WT2/WT4 WT1-WT4	1.25e-03	0.664
	Simpson's Entropy	1.33e-02	0.720	WT1-Prelim/WT4	0.0507	0.330
BCR	Species Richness	1.07e-02	0.701	WT1-WT2/WT5	0.166	0.142
	Shannon's Diversity	6.38e-04	0.835	WT2/WT5-WT1/WT4/Prelim	0.116	0.180
	Simpson's Entropy	1.92e-02	0.661	WT1-WT2	0.404	0.0541

[C.4.1] Taxonomic variation

The taxonomic composition of the microbiota was consistent between the L6 and JH wild-type birds, although the exact contribution of each taxa to the microbiota community varied between birds. The proportional contribution of the Actinobacteria, Bacteriodetes, Cyanobacteria, Elusimicrobia, Firmicutes and Proteobacteria phyla varied between all the wild-type birds (Figure 4.4, SI Table 4.6). However, the only collective difference between the JH wild-type birds and the preliminary bird at the phylum level was in the increased contribution of the Cyanobacteria phylum in the L6 bird (p-value $3.39e^{-02}$, R^2 adjusted 0.323).

At the family-level, seven families varied in their proportional contribution between birds; the Bifidobacteriaceae, Christensenellaceae, Clostridiaceae, Elusimicrobiaceae, Lachnospiraceae, Ruminococcaceae, and Veillonellaceae (full data in SI Table 4.7). These differences often arose from the preliminary L6 bird, for example in the case of the Clostridiaceae, Lachnospiraceae, and Ruminococcaceae. However, only the Lachnospiraceae

family varied significantly between the preliminary L6 bird and the JH wild-type birds (p-value $2.71e^{-08}$, R^2 adjusted 0.9930), with its contribution significantly increased in the preliminary bird's microbiota community relative to the JH wild-type birds (Figure 4.4).

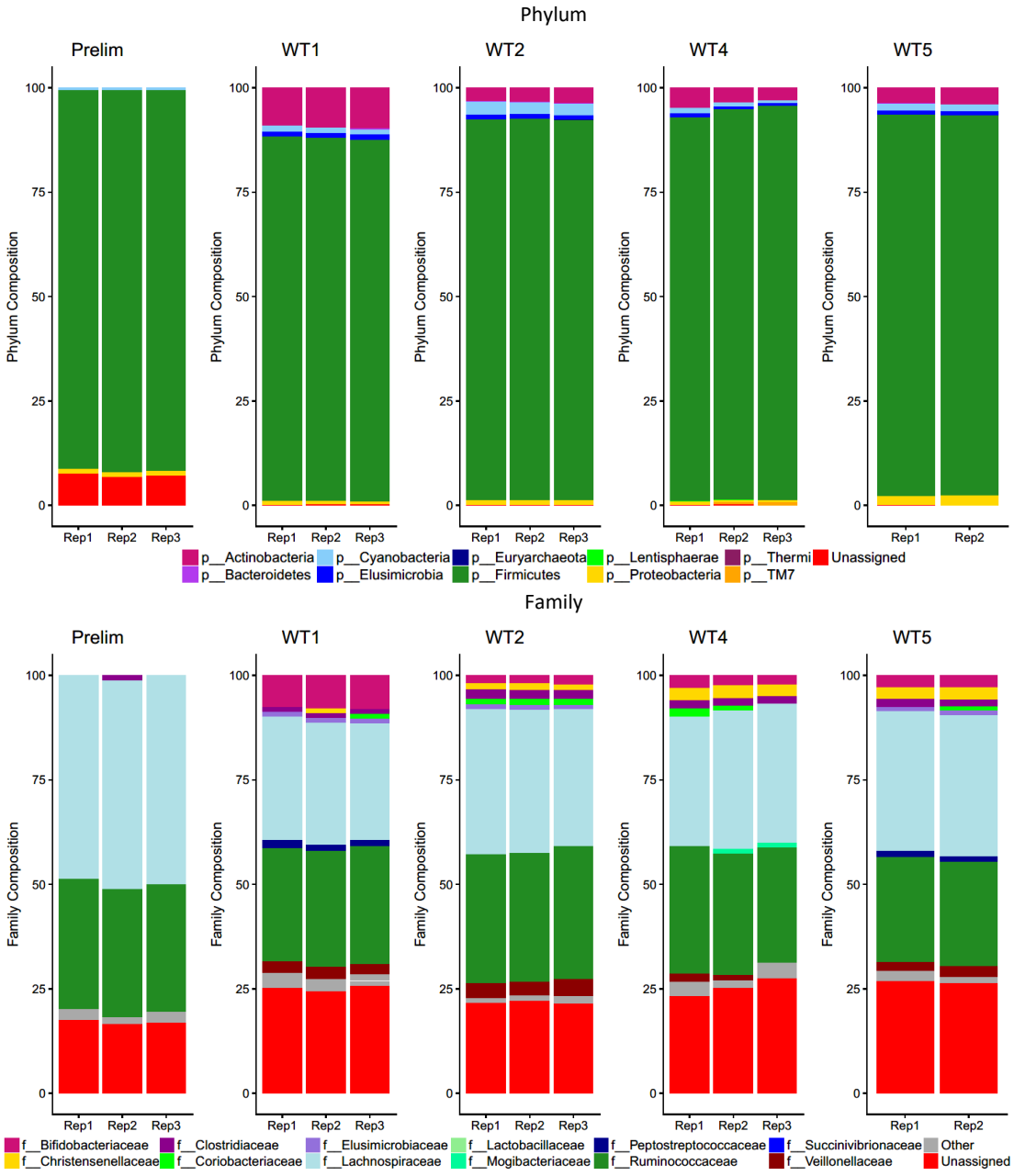


Figure 4.4 – Phylum (upper) and family (lower) taxonomic assignments for the microbiota of the preliminary (L6) and JH wild-type (WT) birds. Family assignments that contributed less than 1% of the total sequences were compressed into a single 'other' category (SI Table 4.8 for full family list). All three technical replicates are indicated. Replicate 3 of WT5 sequence quality was poor (SI Figure 4.1) so this was excluded from the analysis.

[C.4.2] TCRβ gene usage

Within the TCRβ population there was a significant difference in the utilisation of Vβ2in the preliminary L6 bird with the JH wild-type birds (Figure 4.5). Although all the wild-type birds showed the expected skew towards Vβ2 usage, it was significantly higher within the preliminary bird.

Vβ1-associated Jβ segment usage varied between birds (Figure 4.5). The difference in Jβ segment usage was largely driven by the preliminary bird showing an increased utility of Jβ4 compared with the JH wild-type birds (Table 4.2). Though there was some difference in Vβ2-associated Jβ usage between birds, there was substantially less variation in Vβ2-Jβ utilisation between birds compared with Vβ1-associated Jβ segment usage (Table 4.2).

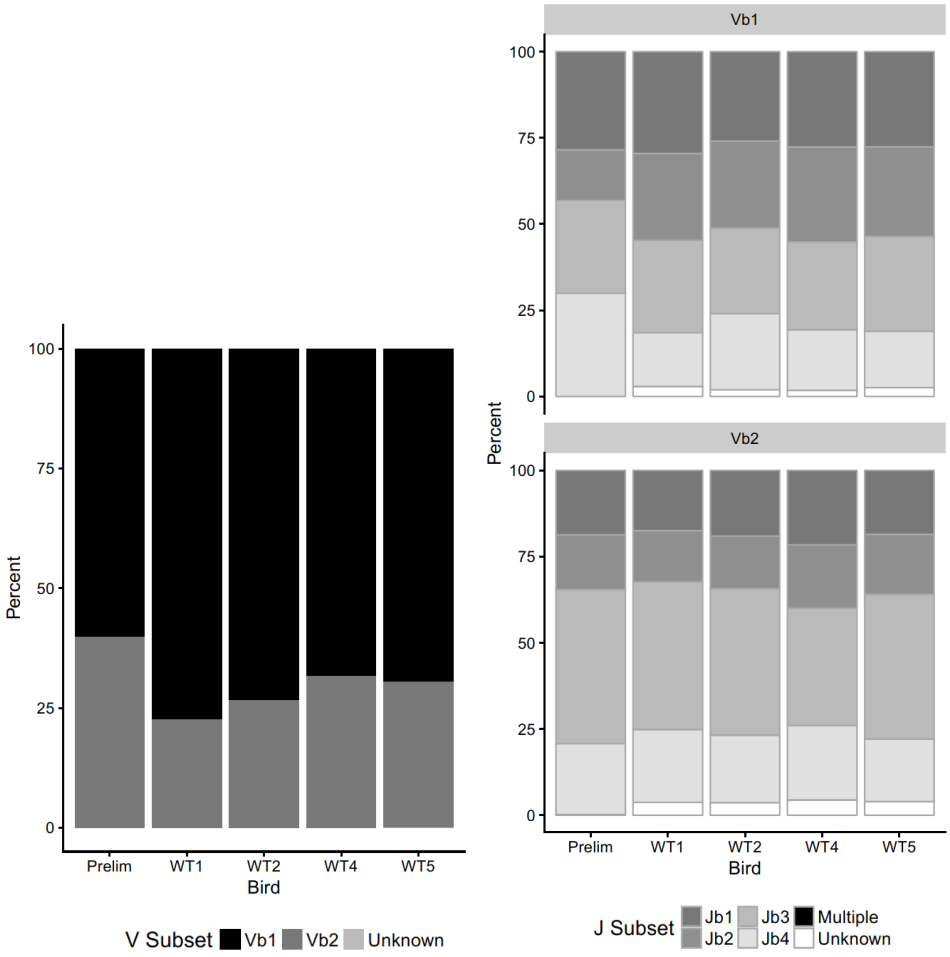


Figure 4.5 – TCRβ gene segment usage in the first technical replicate of the preliminary L6 and JH wild-type (WT) birds. Vβ family segment usage (left) varied between the preliminary L6 and JH wild-type birds. Jβ segment usage, by associated Vβ gene segments (right), varied in some cases and not in others (see Table 4.3).

Table 4.2 – The significance of differential gene usage between birds, in terms of both Vβ gene segments and their associated Jβ segments. Where 'NA' is indicated, the effect of bird upon segment usage was not significant. Values are given to three significant figures. The same significance levels are seen when the JH wild-type birds are collectively compared to the preliminary L6 bird. Birds which are significantly different from each other were assessed using Tukey's HSD test and are indicated as 'Different bird – birds distinct from'. For example in the Vβ1 group the preliminary L6 bird ('Prelim') is significantly different from all the other JH wild-type birds.

V Subset	J Subset	P-value	R^2 Adjusted	Different Birds
Vβ1	*All*	1.35e-07	0.970	Prelim-WT1/WT2/WT4/WT5 WT1-WT2/WT4/WT5
	Jβ1	5.69e-03	0.739	Prelim-WT4/WT2 WT2-WT1
	Jβ2	2.53e-07	0.966	Prelim-WT1/WT2/WT4/WT5 WT1-WT4
	Jβ3	*NA*	*NA*	*NA*
	Jβ4	3.18e-05	0.911	Prelim-WT1/WT2/WT4/WT5
Vβ2	*All*	1.29e-07	0.971	Prelim-WT1/WT2/WT4/WT5 WT1-WT2/WT4/ST5
	Jβ1	*NA*	*NA*	*NA*
	Jβ2	*NA*	*NA*	*NA*
	Jβ3	2.53e-02	0.640	Prelim-WT4
	Jβ4	1.21e-02	0.693	Prelim-WT5

[C.4.3] Public immune repertoire

19 TCR nucleotide sequences were found in all the JH wild-type birds, of which six were also found within the preliminary L6 bird (Table 4.3). All of the shared nucleotide sequences were found infrequently within each bird (Figure 4.6). As none of the wild-type birds were sequenced within the same MiSeq run these are sequences do not result from sequencing cross-over and could represent a shared naïve TCR repertoire.

At the amino acid level 22 TCRβ sequences were shared between the preliminary L6 bird and the JH wild-type birds and 119 across all JH wild-type birds (SI Figure 4.3). No BCR amino acid sequences were shared between the preliminary L6 bird and the JH wild-type birds, nor were any amino acid sequences shared between all the JH wild-type birds. A single BCR amino acid sequence was found in three of the four JH wild-type birds (PGKGLEFVAAINRFGNSTGH-GAAVKGRVTISRDDGQSTVRLQLSNLRAEDTATYYCAKGAYGYCGSGSWCAAGNIDAWGHGTEVIVSSA-SASR), though this was absent from JH wild-type bird WT4.

Table 4.3 – Nucleotide, and associated amino acid, sequences that were present in all the JH wild-type birds. Sequences also found within the preliminary L6 bird are highlighted in grey.

Nucleotide Sequence	Amino Acid Sequence
TGC	C
TGTGCCAGC	CAS
TGCGCTAAA	CAK
ACACCACTGAACTTT	TPLNF
TGTGCCAGCAGTTTG	CASSL
TGCGCTAAGCAAGACGAAAGACTGATCTTT	CAKQDERLIF
TGCGCTAAGCAAGGGGACACACCACTGAACTTT	CAKQGDTPLNF
TGTGCCAGCAGACAGTACAGTAACATGATTTTC	CASRQYSNMIF
TGCGCTAAGCAAGACAGGGACGTAAGACTGATCTTT	CAKQDRDVRLIF
TGTGCCAGCAGGGGGATCGTTAATATCCAGTATTTT	CASRGIVNIQYF
TGCGCTAAGCAAGATCAGGGGATCGAAAGACTGATCTTT	CAKQDQGIERLIF
TGCGCTAAGCAAGATCGAATCGGGACAGGGGTAATTTTC	CAKQDRIGTGVIF
TGCGCTAAGCAAGATACAGGGGGATACAGTAACATGATTTTC	CAKQDTGGYSNMIF
TGCGCTAAGCAAGGGACAGGGGAGTACAGTAACATGATTTTC	CAKQGTGEYSNMIF
TGCGCTAAGCAACGGGACAGGGGGTACAGTAACATGATTTTC	CAKQRDRGYSNMIF
TGCGCTAAGCAAGGGACAGGGGGATACAGTAACATGATTTTC	CAKQGTGGYSNMIF
TGCGCTAAGCAAGACAGGGGGATCGACACACCACTGAACTTT	CAKQDRGIDTPLNF
TGCGCTAAGCAAGATGGGACAGGGGGATCGTACAGTAACATGATTTTC	CAKQDGTGGSYSNMIF
ATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGGGACTGAACTTT	IGQAVVSTQSTWGLNF

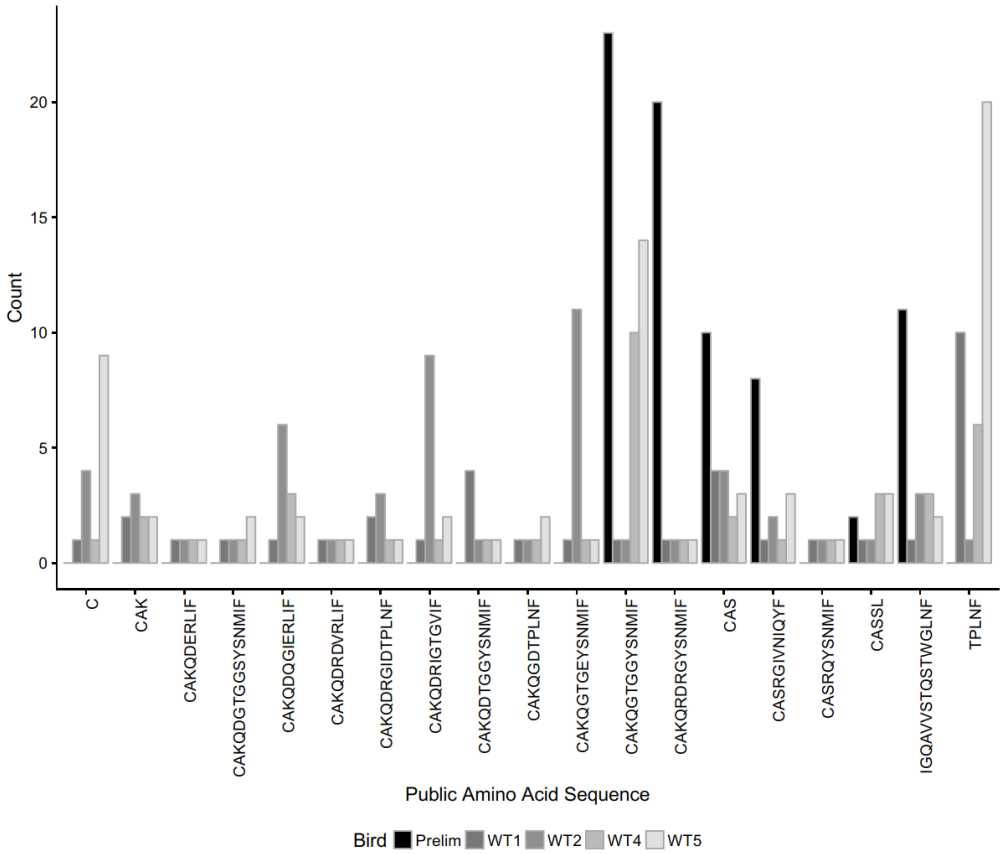


Figure 4.6 – The abundance of shared T-cell receptor β nucleotide sequences, presented using their associated amino acid sequences. The nucleotide sequences presented are those found in all the JH wild-type birds. Counts given are those found across the entire gut grid of each wild-type bird.

[C.5] Fine-scale variation in population composition

The variation in population diversity estimates between birds at a large scale was recapitulated in the mean fine scale diversity estimates (Figure 4.7). Collectively, the preliminary L6 bird was significantly different from the JH wild-type birds in all populations (Table 4.4); specifically the preliminary bird was less diverse in its microbiota and TCRβ populations and more diverse in its IgA-producing B-cell population (Figure 4.7).

Table 4.4 – The summary statistics for the association between bird and the mean diversity indices for each population. 'Effect of Bird' refers to the significance of the association between a given diversity estimate and the comparison bird identity (both preliminary and JH wild-type Birds which are significantly different from each other were assessed using Tukey's HSD test and are indicated as 'Different bird – birds distinct from'. For example in the Microbiota species richness group the preliminary L6 bird ('Prelim') is significantly different from all the other JH wild-type birds. 'WT-Prelim Difference' refers to the significance of the difference between the preliminary bird and all of the JH wild-type birds for a given diversity estimate. The complete BCR data is in SI Table 4.9. All values are given to three significant figures.

		Effect of Bird			WT-Prelim Difference	
Population	Estimate	P-value	R^2 Adjusted	Different Birds	P-value	R^2 Adjusted
Microbiota	Species Richness	7.28e-04	0.859	Prelim-WT1/WT2/WT4/WT5	1.99e-04	0.698
	Shannon's Diversity	2.04e-04	0.894	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT4/WT5	2.14e-03	0.558
	Simpson's Entropy	3.85e-05	0.928	Prelim-WT4 WT2-Prelim/WT1/WT4/WT5	2.69e-02	0.346
TCRβ	Species Richness	3.15e-03	0.802	Prelim-WT1/WT2/WT5	9.18e-04	0.614
	Shannon's Diversity	1.08e-04	0.908	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT5 WT1-WT4	4.03e-03	0.512
	Simpson's Entropy	3.24 e-03	0.801	Prelim-WT1/WT5 WT1-WT2	3.34e-02	0.325
BCR	Species Richness	5.81e-06	0.937	Prelim-WT1/WT2/WT4/WT5 WT4-WT2/WT5	2.73e-06	0.826
	Shannon's Diversity	6.93e-04	0.832	WT4-Prelim/WT2/WT5 Prelim-WT1	3.03e-03	0.504
	Simpson's Entropy	6.29e-03	0.733	WT4-Prelim/WT2/WT5	0.167	0.141

The preliminary L6 bird exhibited variability in each segment's community diversity (Chapter 2 D.3), regardless of population type. The effect of location on community diversity was inconsistently replicated within the JH wild-type birds. In the microbiota, only JH wild-type birds WT1 and WT2 showed any evidence of segmental variation in diversity (SI Table 4.10)

whilst WT2 and WT5's TCRβ segment communities and WT4's BCR communities varied between locations (SI Table 4.10).

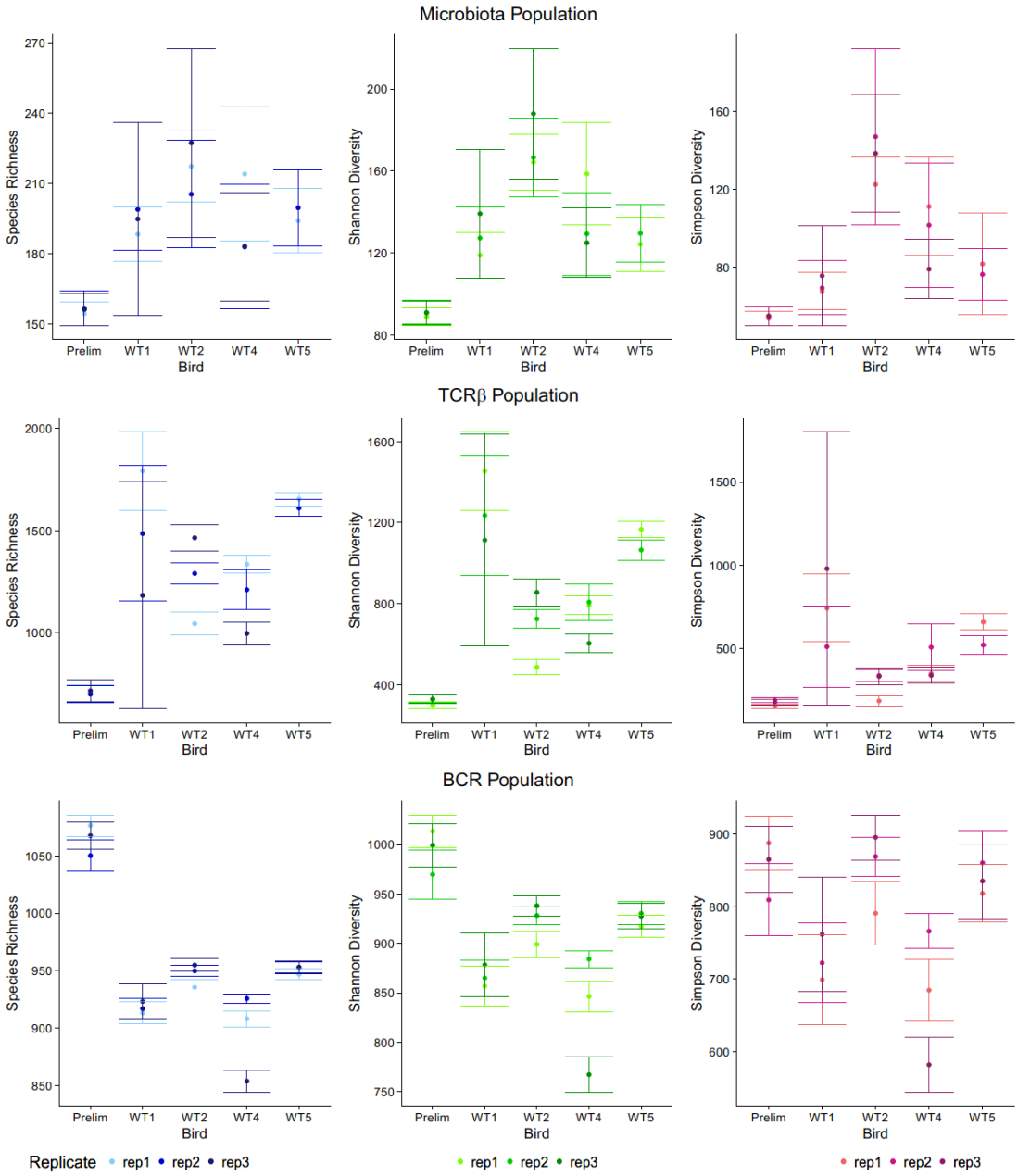


Figure 4.7 – Mean diversity estimates of individual segments for the preliminary (L6) and JH wild-type (WT) birds for every interrogated population. Species richness (blue), Shannon's diversity index (green) and Simpson's diversity index (red) of each technical replicate are displayed for each bird. Estimates were calculated to the same end point either through interpolation or extrapolation. The standard error of the estimates (95% confidence interval) is shown.

In each case, the variation in diversity between segments was driven by a single, or few, segments being significantly different from the other segments, regardless of population type.

[C.5.1] Taxonomic variation

Each of the wild-type birds showed evidence of microbiota community variation between segments, though the level at which the segmental variation could be seen varied between birds (SI Tables 4.11 and 4.12).

Firmicutes was the most dominant phylum in any of the wild-type birds. In each bird, except JH wild-type bird WT4, the proportional contribution of Firmicutes to the microbiota varied between segments (SI Table 4.11). Similarly the Proteobacterium phylum varied in its community contribution between segments in most of the JH wild-type birds but not in the preliminary bird.

Although JH wild-type bird WT4 showed no variation in the proportional contribution of different phyla to the microbiota, there was segmental variation in the contribution of the Clostridiaceae and Veillonellaceae families (SI Table 4.12).

[C.5.2] TCRβ gene usage

Unlike the consistency of microbiota variation between segments, the variability of TCRβ gene utility between locations was only seen within the preliminary L6 bird and the JH wild-type bird WT1 (SI Table 4.13). Moreover, the genes that changed in usage between segments were not repeated between the preliminary L6 bird and the JH wild-type bird WT1 (Vβ1-Jβ3 and Vβ1-Jβ1 respectively).

[C.6] Dissimilarity between segments

Segment dissimilarity was reduced in the microbiota populations of both the preliminary L6 bird and the JH wild-type birds. Similarly in all the wild-type birds the BCR populations varied most between segments (Figure 4.8).

Although the trends between the populations considered were consistent between the preliminary L6 and the JH wild-type birds, there was a significant difference between the mean microbiota dissimilarity between segments of the L6 and JH wild-type birds according to any of the dissimilarity indices used (Kulczynski; p-value $6.8e^{-03}$, R^2 adjusted 0.443, Hellinger; p-value $3.5e^{-03}$, R^2 adjusted 0.494, Chi Squared; p-value $2.5e^{-03}$, R^2 adjusted 0.517). Specifically the

preliminary bird's microbiota varied less between segments than any of the JH wild-type birds, which could be due to the difference in sampling scale between these birds.

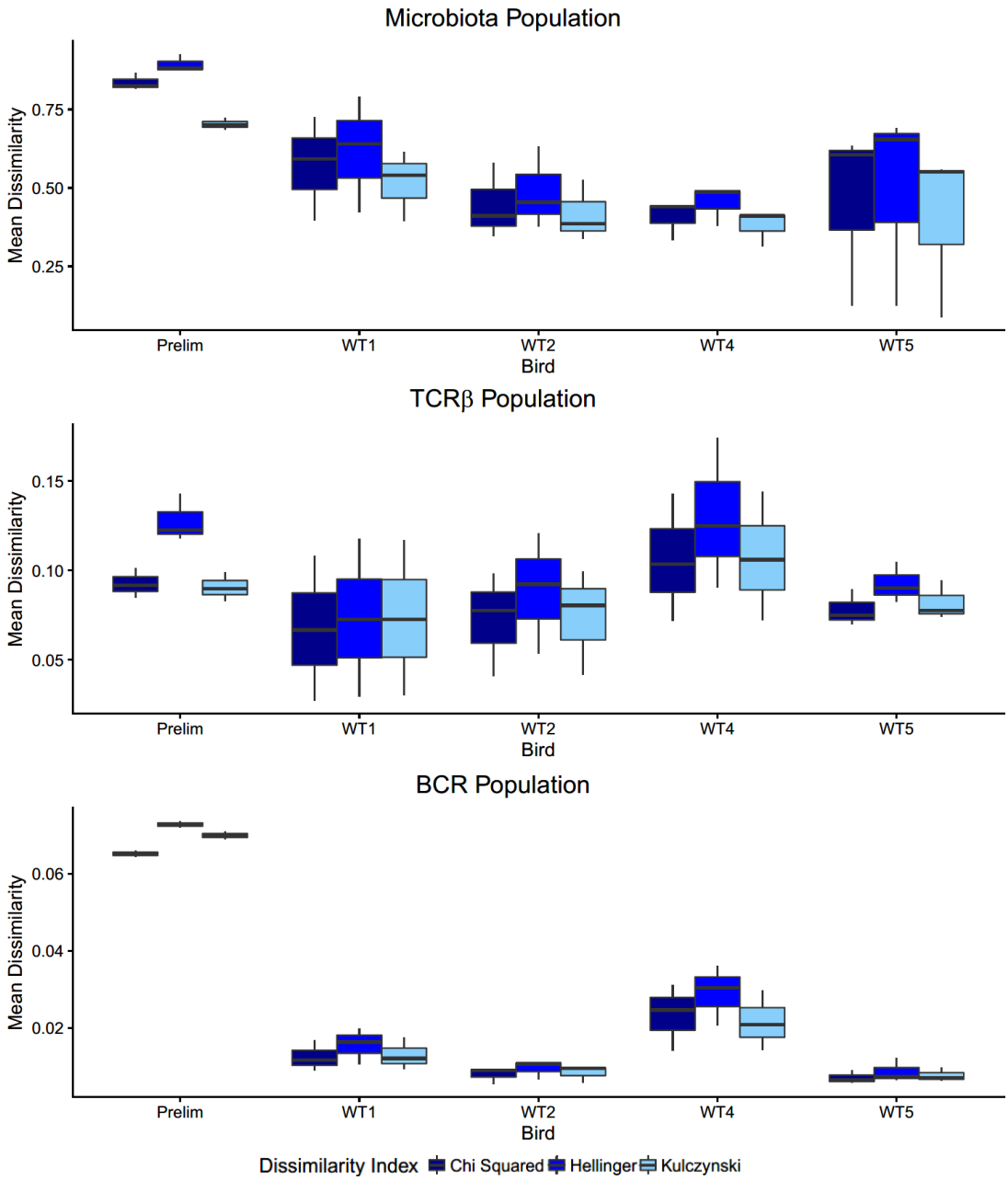


Figure 4.8 – Mean dissimilarity estimates of individual segments for the preliminary (L6) and JH wild-type (WT) birds. Dissimilarity was calculated using the Chi Squared, Hellinger and Kulczynski indices as indicated.

Both the BCR mean diversity (Figure 4.7) and dissimilarity of the preliminary bird was reduced relative to the JH wild-type birds (Figure 4.8), though there was no significant difference between the preliminary L6 and JH wild-type birds in the TCRβ population (Figure 4.8).

[C.6.1] Spatial structure within the microbiota and immune populations

Within the microbiota population, the preliminary L6 bird showed a positive correlation between the dissimilarity of paired segments and their Euclidean distance (Chapter 2 D.5). However, this correlation was not replicated within any of the JH wild-type birds. Nor was there an association between segment's dissimilarity and distance within any of the JH wild-type BCR populations.

In the L6 preliminary bird TCRβ population there was no correlation between dissimilarity and distance. JH wild-type bird WT2, however, showed some evidence of a positive association, though the effect size was small and only significant in two of the dissimilarity indices (Kulczynski: p-value $1.90e^{-02}$, R^2 adjusted $1.97e^{-03}$; Chi Squared: p-value $3.70e^{-02}$, R^2 adjusted $1.49e^{-03}$).

[D] Discussion

The aim of this chapter was to consider the reproducibility of the initial results from the preliminary L6 bird, in particular the discrepancy in community appearances at large and fine scales and the existence of spatial structuring within the immune and microbiota populations. However the comparison performed here is caveated by the different experimental designs of the preliminary L6 bird and the JH wild-type birds: specifically their distinct genetic backgrounds and segment sizes.

The overall trends seen within the preliminary L6 bird could be largely replicated within the JH wild-type birds: the microbiota had relatively fewer rare species than the immune populations; there was variation in the contribution of different species to distinct segment communities; and the immune populations were more dissimilar between segments than the microbiota populations. However, the differences between the large-scale and fine-scale population dynamics were inconsistent – with the patterns of the preliminary L6 bird repeated more in some JH wild-type birds (e.g. WT4 and WT1) than others – and evidence of spatial structuring was absent.

[D.1] Inter-individual variation

Both the microbiota and adaptive immune system are highly variable between individuals[232,233], particularly before their maturation[88–90,351]. As stated previously, the wild-type birds compared here are of two distinct genetic backgrounds: the L6 background being more inbred than the modified Leghorn-Minnesota Marker background[360]. The birds investigated were also housed in two distinct locations (Oxford and Munich) and thus raised in fundamentally distinct conditions. These factors have been shown previously to effect the outcome of microbiota studies[234–238].

As Chapter 2 highlighted, there is evidence that the scale at which ecological communities are considered can affect the dynamics which shape communities[1,5–7]. The difference in scale between the 4 mm^2 preliminary grid and subsequent 1 mm^2 grids may therefore affect the differences seen between these two populations; especially the presence of spatial structuring. The grouping at the larger scale (4 mm^2) may exaggerate a continuous gradient of changing community structure, facilitating the emergence of spatial structuring that becomes less apparent over the finer-scale, particularly given the inconsistency of the distance unit.

To compensate for the discrepancy in scale between the two sampling attempts it could be possible to repeat these comparisons by grouping the finer scale segments into comparative fragments (sets of adjacent 4 segments), although selecting which combinations of segments to combine and accounting for sequencing variability may prove challenging.

The reduction in scale between the two sampling structures may also affect the extent to which each segment's community has been effectively sampled and therefore the dissimilarity between segment communities. The finer scale segments will have been sampled to a potentially greater depth than for the 4 mm^2 segments, meaning the sampling attempt is more likely to retrieve numerically rare individuals. As the PCR process is a non-replacing sampling structure, truly rare species may be absent from subsequent sampling attempts and therefore create highly variable technical replicates. The high inter-replicate variability of the 1

mm^2 segments relative to the preliminary 4 mm^2 segments suggests that this could be occurring, and thus obscuring true differences between the segment populations.

[D.2] Community stability

In chickens, the microbiota stabilises at *c.* six to eight weeks of age[88–90], concordant with the maturation of the adaptive immune system[81,252,359]. The variability in the trends between the six-week-old wild-type birds could, therefore, reflect the stabilisation of the microbiota in some birds and not others. Some of the JH wild-type birds look more like the preliminary bird than the other JH wild-type birds (i.e. WT4 in many cases), which may indicate that these birds are at a similar point of maturation distinct to the other interrogated birds. Repeating these analyses with older birds may therefore produce more repeatable results.

Given the instability of the immature microbiota, and the importance of microbiota stability in its beneficial properties, a corollary of these results is that manipulating the microbiota of chickens with probiotics or prebiotics in an effort to reduce the burden of infectious disease upon flocks is likely to be challenging.

[E] Conclusions

Microbiota studies are notoriously variable, due to the effects of strain and rearing conditions upon community composition. The consistency of some of the trends presented in Chapter 2 between individuals of different strains and locations within this chapter is thus notable. There are, however, some results from Chapter 2 which were not repeatable in the JH wild-type birds interrogated here. It is possible that the difference between the JH wild-type birds and the preliminary L6 bird emerge as a result of strain differences, sampling differences or inherent variability of both the microbiota and adaptive immune system at the age at which the birds were sacrificed. Nonetheless this is a valuable and broadly applicable methodology for interrogating host-microbiota interactions.

Chapter 5

B-cells mediate immune and microbiota community dynamics

[A] Introduction

As has been mentioned previously, the beneficial properties of the gastrointestinal microbiota to its host are dependent upon its composition[17,21,48,49,310] and stability[46,311,312]. The factors that affect community composition and stability are varied[6,8] and depend upon the scale at which a community is studied[1,5,6]. In the case of the microbiota one of the strongest selection pressures upon its community comes from the host itself in the form of the host's immune system.

Both the innate and the adaptive immune systems interact with the gut microbiota (see General Introduction B). However the adaptive immune system appears to be key in maintaining the microbiota diversity necessary for its host-beneficial properties.

[A.1] Interactions between the adaptive immune system and microbiota

The interaction between the adaptive immune system and the microbiota is dependent upon both the presence of microbial stimulation[14,39–41], microbiota diversity[42–45] and the specific composition of the interacting microbial and immune components[17,21,46–50] . Changes in the microbiota have been associated with biases in T-cell usage[14,42,54–56], T-cell cytokine profiles[49,63,65–67], B-cell differentiation[54] and B-cell class-switching[40,41,43,44], all of which influence the inflammatory state of the mucosal immune system.

The inflammatory state of the gut influences the ability of microbes to survive within the microbiota, for example *Klebsiella* species drive the microbiota to a colitic profile[361] and Salmonella has been shown to induce inflammation within the gut to reduce colonisation resistance[345,362]. Conversely, some microbes specifically promote an anti-inflammatory state within the gastrointestinal tract (e.g. Lachnospiraceae[60] and Clostridia[61,62] species). Regardless

of the direction of the relationship, a less diverse microbiota has been correlated with an increased pro-inflammatory profile within the gut immune system[42,363,364].

Whilst microbiota composition and diversity influences both arms of the adaptive immune system, the component of the adaptive immune system that is most involved in the modulation of the microbiota is immunoglobulin A (IgA).

[A.2] Immunoglobulin A

IgA is the most highly produced immunoglobulin in both mammals and birds[365,366]. IgA is predominantly produced within the gastrointestinal tract[203] in response to microbial colonisation[40,41,207]. The abundance of IgA-producing B-cells and IgA production itself have both been linked with the presence of a microbiota[40,41], suggesting that the high levels of IgA secretion within the gut are generated in response to the microbiota. Moreover specific bacteria, and bacterial products, vary in their abilities to stimulate IgA production and its specificity[367].

IgA has been shown to modulate both the microbiota and the local and systemic immune responses to the microbiota. Secretory IgA coats a remarkable amount of the microbes within the gastrointestinal tract and its secretion effects both microbiota composition[368,369] and diversity[219], though the mechanisms of IgA's modulation of the microbiota remain unclear.

IgA control of the microbiota has been suggested to occur in many ways, including through barrier functions[370,371], neutralisation of bacterial toxins[372,373] and enhanced clearance of commensals both from the basolateral surface of the epithelium back to the lumen[374,375] and from the lumen itself[371]. Although IgA is produced specifically in response to microbiota components[376], the barrier functions of IgA are not dependent upon its specificity[377].

As suggested in Chapter 1, the agglutination of microbes by IgA may also facilitate the survival of non-advantageous components of the microbiota and improve biofilm formation[217]. IgA KO mice show compositional changes within the microbiota[369,378], with restoration of affinity matured specific IgA restoring the initial composition[369,379]. IgA secretion can also effect surface epitope expression in microbes, instigating an anti-inflammatory phenotype in

Bacteroides thetaiotamicron[151]. Thus IgA may act in a two-fold manner, both preventing bacterial access to the host and ensuring a beneficial composition within the microbiota itself.

The effect of IgA upon the local and systemic immune systems is no less important than its effect upon the microbiota, though the knock-out phenotype is relatively mild[380]. At a local level, IgA influences the uptake of specific microbial components into the mesenteric lymph nodes[381] and thus dendritic cell exposure to the microbiota. The specific antigen exposure within the mesenteric lymph nodes influences the activation of both B-cell class-switch recombination to IgA[382] and the cytokine profiles of local T-cells[383]. It has also been suggested that IgA may drive affinity maturation within the gastrointestinal germinal centres by restricting the access of low-affinity immunoglobulin to antigen, thereby accelerating B-cell affinity maturation[384]. IgA can also instigate an anti-inflammatory mucosal immune profile through impairing the activation of the innate immune system.[151]

Systemically IgA down-regulates T-cell and B-cell responses to commensal microbes; for example in the absence of IgA transcytosis microbiota-specific serum immunoglobulin G[385] and CD4+ T-cell production is increased[370].

[A.3] B-cell and T-cell interactions

Although IgA is an important component in microbiota regulation it does not act in isolation. The composition of the microbiota, and its surrounding immune environment, is also dependent upon T-cell presence and subset, though this dependence may be a result of T-cell activation of IgA-producing B-cells.

T-regulatory cells (Treg)[386], T follicular helper cells (Tfh)[387] and CD4+ T-cells[388] have all been demonstrated to be involved in maintaining gastrointestinal homeostasis, both by modulating microbiota composition and by inducing an anti-inflammatory immune environment. Treg control of the microbiota is both TGFβ and IL-10 dependent[389], though other sources of IL-10 and TGFβ within the gut are not sufficient to restore the phenotype, which given the import of both TGFβ[390] and IL-10 in the activation of IgA-producing B-cells[391,392] supports a co-ordinating role for these subsets. Tregs also prevent effector T-cell induction[393]

and induce T follicular regulator cells[37,391] which maintain an anti-inflammatory immune environment within the gut.

Most gastrointestinal IgA is produced via T-cell dependent processes[244], with the absence of T-cells reducing the volume of IgA production without changing its affinity[370,381]. T-cell subsets are involved in all the stages of B-cell activation through class-switching to IgA production and plasma cell longevity. T-cells can influence both the class-switching of B-cell receptors to IgA (Th17 and Tfh via IL21)[394], its affinity maturation (Tfh via IL-21)[205], the production of IgA (Th17 via IL17, Tregs)[370] and the longevity of the plasma cell itself (Tregs)[370].

[A.4] Evidence of microbiota-immune interactions in chicken

Though evidence of microbiota-adaptive immune interactions is extremely abundant with respect to mouse and human models, relatively few studies have attempted to unravel the interactions between the microbiota and mucosal immune system in birds[256,301]. Work within avian models has been largely restricted to the interrogation of the effects of specific probiotics and prebiotics in the modulation of immune responses and microbiota[259].

As in mammals, microbial colonisation of the avian intestinal tract increases the integrity of the epithelial cell walls[395] and changes mucus production[396]. The size and complexity of the intestinal epithelial lymphocytes and lamina propria also increases post hatch and stabilises at roughly the same time as the intestinal microbiota[256,359]. In terms of the adaptive immune system germ-free birds show a reduction in the abundance of both B-cells and T-cells, mucosally and systemically, with no evidence of IgA or IgY production within the gastrointestinal tract. Despite a reduction in the overall T-cell abundance, germ-free birds show an increased in the diversity of their T-cell populations[301], possibly as a result of retaining their naïve polyclonal state.

[A.5] Avian B-cell receptor knock-out

In mice two B-cell knock-out lines exist; μMT[397] and CD79b-mutant strains[398]. However, although both these lines have been shown to knock-out mature B-cells and antibody

production[397,398], there is also evidence of immunoglobulin production within the μMT mouse[399,400] and in the human counter-part of the CD79b-mutant strain[401].

The distinct BCR development process in chickens (see General Introduction C.3), centred around the restricted genetic diversity of the BCR loci (in particular the existence of a single J_H gene), has facilitated the creation of an avian BCR homozygote knock-out line where the J_H segment has been completely removed[360]. The removal of the J_H segment means that BCR IgH formation is blocked regardless of V_H or C_H usage. B-cell progenitors still colonise and proliferate within the Bursa of Fabricius in these JH knock-outs and IgL chain formation remains normal. However post-hatch, IgH expression becomes necessary for B-cell survival and by four weeks of age the knock-out is completely depleted of B-cells.

[A.6] Chapter 5 aims

This chapter comprises a B-cell knock-out experiment within six-week-old chickens, with wild-type, heterozygote and homozygous knock-out genotypes. The aim of this chapter was to assess whether there is a difference in the T-cell receptor repertoire and microbiota population between B-cell sufficient and B-cell knock-out birds. The difference between B-cell sufficient (either wild-type or heterozygous) and knock-out birds was assessed at both the large-scale (aggregated 1cm grid) and fine-scale (segmented grid).

[B] Methods

Sample collection, amplification, sequencing and initial processing were as in Chapter 4 section B. The extensions and key features of the methods are summarised here.

[B.1] Sample collection

11 homozygote B-cell receptor knockout (BCR KO) and heterozygote chickens of the JH-KO line, and wild-types (JH-intact) on the same mixed genetic background (see Chapter 4 B.1)[360], were sacrificed for this experiment. The animals were provided in collaboration with Professor Bernd Kaspers laboratory (Ludwig-Maximilians University, Munich) and Benjamin

Schusser. The BCR KO was formed using homologous gene-targeting in primordial germ cells to delete the J segment of the BCR heavy-chain[360].

[B.1.1] Grid creation

After sacrifice, the intestinal tract was removed from each bird and grids taken in situ. A 1 mm^2, 4 column, grid was taken from the right caecum of each bird. The columns of the grid were taken moving from the anterior to posterior of the caecum (A-D respectively), then the rows of the grid excised at 1 mm intervals until the whole circumference of the caecum had been sampled (numerical notation of rows, alphanumeric notation for columns). A maximum grid size of 4 x 12 was retrieved.

The segments were stored in 100 µl of RNAlater at -20 ^oC overnight before shipping to the UK and stored at -80 ^oC until processing.

[B.2] Tissue RNA/DNA extraction, cDNA synthesis and DNA amplification

RNA and DNA were extracted from the samples using QIAGEN 96-well AllPrep DNA/RNA extraction kit, according to the manufacturer's instructions. Plate extractions were designed to contain an individual from each treatment whilst minimising the number of individual grids split over a two 96-well plates. Tissue disruption, homogenisation and RNA/DNA extraction was performed as in Chapter 4 B.2.

5' RACE cDNA was made from the AllPrep extraction RNA, according to the SMARTer reverse transcriptase protocol (as in Chapter 2 B.3, and extended in Chapter 4 B.3). Master mixes were set up for each individual bird grid from every 96-well extraction.

Sample extraction and synthesis were validated by β-actin PCR (Chapter 2 B.4, and extended in Chapter 4 B.4). During extraction and synthesis validation it was noted that the genomic DNA extraction from a single JH knock-out bird (KO5) had failed. The microbiota from KO5 could not therefore be assessed. There was no problem with any of the other birds on the extraction plate or with any of the RNA extractions.

Samples within each 96-well plate were amplified as in Chapter 2 and Chapter 4. Barcodes and primers were designed and PCR conditions as in Chapter 2 B.4.

[B.3] Sequence creation and processing

The complete BCR knock-out experiment was run across 4 MiSeq runs, with each MiSeq run containing all the technical replicates of a single 96-well extraction plate.

[B.3.1] Library formation and sequencing

All samples from a single 96-well extraction plate were pooled into libraries by technical replicate. Each library consisted of all population PCRs of a single technical replicate (16S, BCR and TCRβ). The volume of PCR product pooled was adjusted to reflect the relative strength of the sample bands from the electrophoresis gels. Samples were initially pooled by population type and the DNA concentration of the pool quantified by dsDNA BR Qubit (according to the manufacturer's instructions). The pooled populations were then combined in a 1:5:5 ratio (16S, BCR and TCRβ, respectively).

The libraries were prepared and sequenced as in Chapter 2 B.6, quantified by qPCR and pooled in equal ratios to produce 20 µl of a 4 nM mixture for Illumina sequencing.

[B.3.2] Processing raw sequences

Initial sequence processing was performed as in Chapter 2 B.6 and sequence quality assessed as in Chapter 2 B.7. 16S sequences amplified within MiSeq22 library 3 were significantly degraded (SI Figure 5.1) and thus excluded from further analyses (see Chapter 4 B).

[B.3.3] Data trimming

During data exploration it was discovered that the JH knock-out birds were retrieving B-cell receptor sequences. As the knock-out birds themselves have been validated as not producing antibody these sequences cannot be IgA-producing B-cells. Amino acid fragments from the Jβ segments were found in all the birds, though the corresponding nucleotide sequences were far less abundant (Table 5.1). As would be expected through gene rearrangement there were very few complete Jβ segments within the sequences retrieved.

Table 5.1 – The number of unique sequences containing Jβ segment fragments in either the nucleotide or its corresponding amino acid sequence. Fragment lengths identified are indicated. The MiSeq run on which the birds data was run is highlighted in shades of grey (white on MiSeq19, light grey on MiSeq20, dark grey on Miseq21 and black on MiSeq22), bird HT5 was split across MiSeq runs 21 and 22. Fragments identified were Jβ1 (4 amino acids: FGDG, 6 amino acids: FGDGTK, 12 nucleotides: TTCGGTGATGGC), Jβ2 (4 amino acids: FGTG, 6 amino acids: IFGTGT, 12 nucleotides: TTTGGCACTGGG), Jβ3 (4 amino acids: FGQG, 6 amino acids: FGQGTR, 12 nucleotides: TTTGGACAGGGC), Jβ4 (4 amino acids: FGEG, 6 amino acids: YFGEGT, 12 nucleotides: TTTGGAGAAGGA), and centred on the conserved F amino acid. No Jβ segments of 18 nucleotides were identified in any of the birds in full.

	Central 4 Amino Acids				Central 6 Amino Acids				Central 12 Nucleotides			
Bird	Jβ1	Jβ2	Jβ3	Jβ4	Jβ1	Jβ2	Jβ3	Jβ4	Jβ1	Jβ2	Jβ3	Jβ4
WT1	14	5	5	13	0	0	2	0	0	0	0	0
WT2	28	211	6	11	0	1	3	0	0	0	0	0
WT4	58	117	5	17	0	0	2	0	0	0	0	1
WT5	178	1134	20	96	0	0	12	0	2	0	0	0
HT1	44	92	6	16	0	0	6	0	0	0	0	0
HT3	28	6	5	19	0	0	1	0	0	0	0	0
HT5	18	204	3	10	0	0	1	0	0	0	0	0
KO1	78	698	14	41	1	2	10	0	0	0	0	1
KO2	31	192	3	20	0	0	2	0	2	0	0	0
KO3	27	61	1	9	0	0	0	0	0	0	0	1
KO5	14	3	4	12	0	0	0	0	0	0	0	0

Jβ segment fragments were found in both the B-cell sufficient and B-cell knock-out birds, although the knock-out birds did not contain any sequences that were not also found in B-cell sufficient birds on their shared MiSeq run. It is likely that cross-contamination between samples has therefore occurred, either as part of the PCR process or as a result of cross-over during the Illumina MiSeq sequencing.

To attempt to compensate for this experimental error two methods were used in data analysis. Firstly the data produced in the BCR knockout PCRs was ignored and all other datasets were left intact. Secondly sequences were removed systematically from the samples in all population PCRs according to Figure 5.1 and SI Figure 5.1. The second method left 12642 unique nucleotide sequences (12589 amino acid sequences) across all the KO birds, though only 208 were found ≥6 times in any sample. There was no significant difference between the trimmed and raw TCRβ dataset so only the trimmed dataset was used for further analysis.

The data presented is predominantly the raw BCR data, though where the trimmed dataset conclusions varied is highlighted.

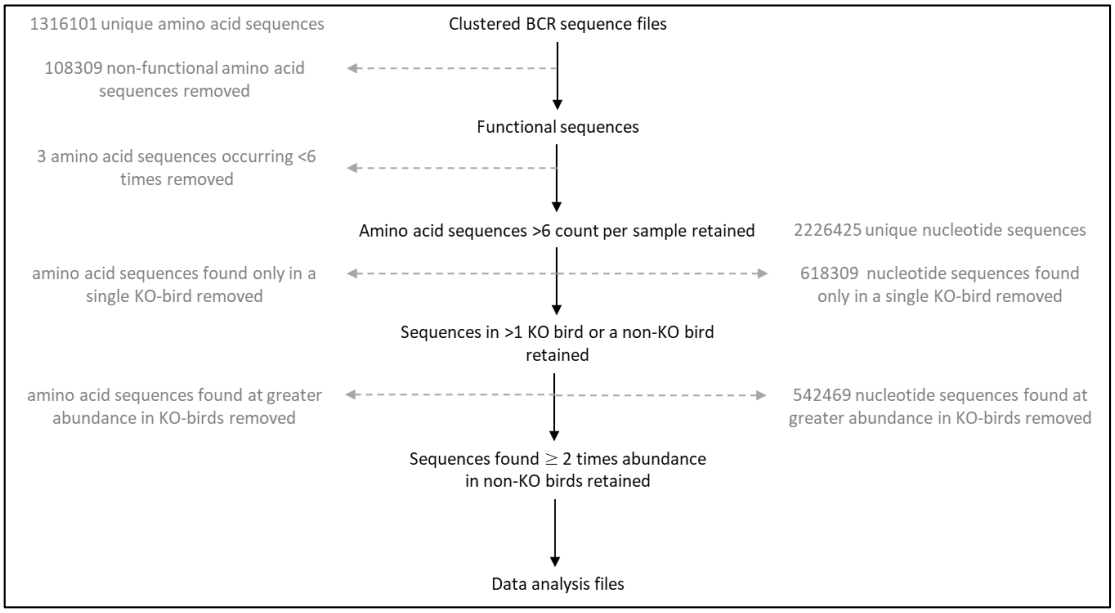


Figure 5.1 – Schematic of trimming decisions for the IgA-producing B-cell receptor (BCR) dataset. 'KO' refers to JH knock-out. Retaining amino acid sequences of >6 count was based upon error rates reported by Schirmer et al. (2015,2012)[285,288].

[B.4] 16S contamination control

118 controls were run over the 4 MiSeq runs, comprising 18 extraction blanks and 100 water controls. Contamination sequences were designated as in Chapter 4 based on extraction blanks, water controls and identification thresholds. Within this study, 121 ISVs were designated contaminating sequences and removed from further analysis.

[C] Results

The microbiota and adaptive immune system are interdependent[37,38] and their interaction mediated by IgA[40,41,376]. IgA production is largely T-cell dependent[244,402] which links the B-cell, T-cell and microbiota populations. Within this chapter, the effect of B-cell presence and absence upon the structure and composition of the microbiota and T-cell receptor β populations was considered at both a large (comparative to previous murine studies) and fine-scale.

Surprisingly, at a large-scale there was no difference in the diversity of the microbiota or T-cell receptor β population between B-cell sufficient and B-cell knock-out birds. However at a fine-scale there were significant differences in the population structures of both TCRβ clones and the microbiota. Specifically, rare microbial species contributed significantly more to microbiota of B-cell sufficient birds. The variation in the contribution of rare microbiota species to B-cell sufficient and B-cell knock-out birds could be seen at a fine-scale in both the phylum and family level assigned taxonomic units.

As mentioned in B.3.3 the presence of BCR sequences within the knock-out birds led to trimming of the datasets. In the case where there was a significant difference between the raw and trimmed populations any comparison, statistical or otherwise, was performed for both the trimmed and raw datasets.

[C.1] Functional sequences

Both the immune populations interrogated contained sequences that could not produce transcripts (either containing a stop codon or translating out-of-frame). In all birds there were more non-functional B-cell receptor (BCR) than T-cell receptor β (TCRβ) sequences and clones, though the exact proportion of non-functional sequences/clones in each population varied between birds (SI Table 5.1).

Within the TCRβ population neither genotype nor B-cell status effected the proportion of non-functional sequences/clones in each bird (SI Table 5.1). However both genotype and B-cell status significantly affected the presence of functional clones within the BCR population (Figure 5.2, SI Table 5.1). Specifically the knock-out birds had significant fewer productive BCR sequences than either the heterozygote or wild-type birds.

[C.2] B-cell sufficiency and large-scale population structure

Each of the grids produced within this experimental framework completely encircled the caeca. When the segments of the grid are pooled as a ‘large-scale’ comparison between the genotypes they therefore represent a complete section of the caeca.

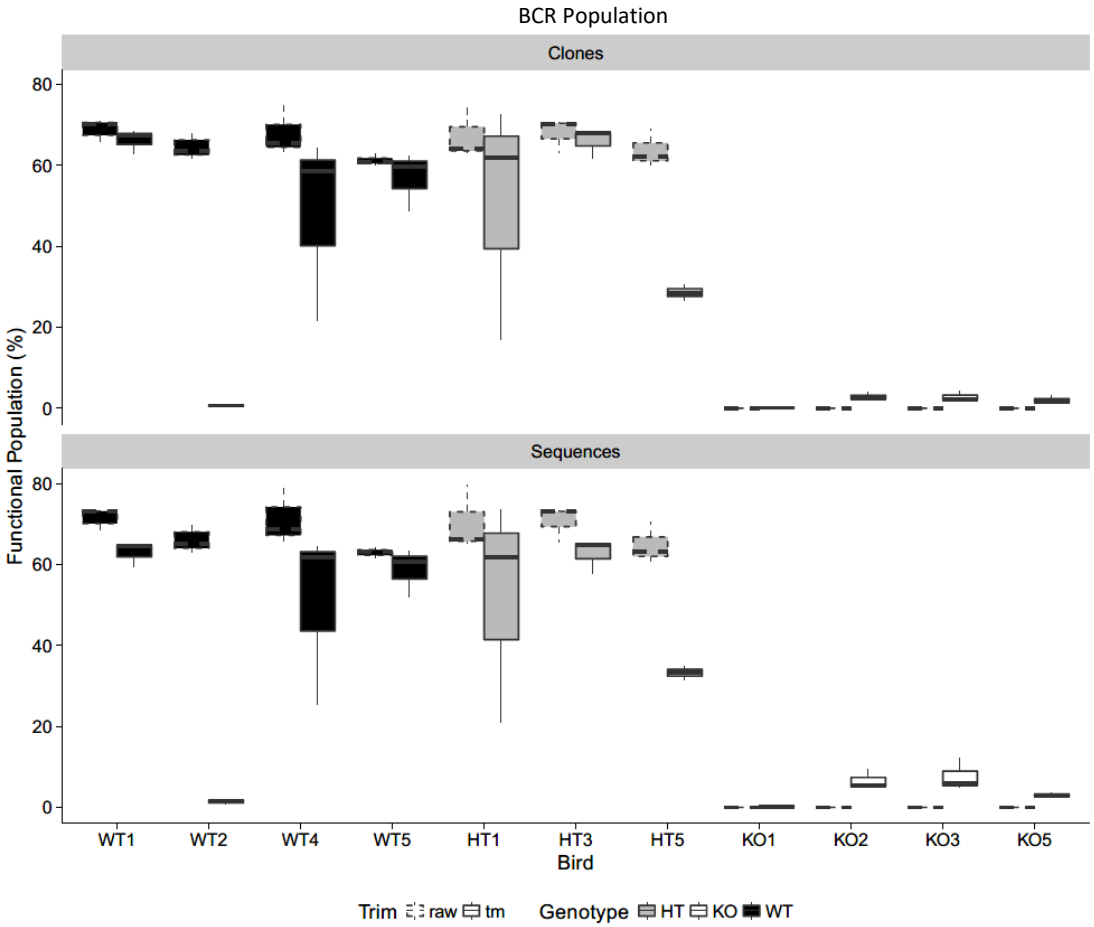


Figure 5.2 – The proportional contribution of functional sequences and clones within each bird for the IgA-producing B-cell receptor (BCR) populations. Boxplots are coloured by the genotype of the bird; wild-type (WT) in black, heterozygote (HT) in grey and knock-out (KO) in white, and outlined according to the dataset treatment; the raw data dotted and the trimmed data as solid lines. The trimming conditions implemented can be seeing in B.3 Figure 5.1, SI Figure 5.1.

Although there are different numbers of segments between the different birds (indicating the diameter of the caeca varies) there was no effect of genotype upon caecal size. B-cell status did, however, effect caecal diameter with the knock-out birds' caeca significantly smaller than those of the B-cell sufficient birds (Welch two sample t-test: p-value 0.046, population means 31.0 and 37.1 respectively).

Though the proportional contribution of differentially abundant microbe species to the microbiota varied between birds (SI Table 5.2), there was no significant effect of either genotype or B-cell status upon the population structure of the microbiota (Figure 5.3). The TCRβ population similarly varied between birds in the abundance of different size clones (SI

Table 5.2) and the deviation of CDR3 length distributions from normality (p-value $1.03e^{-04}$, R^2 adjusted 0.641), though this variation was not linked to either bird B-cell status or genotype (Figure 5.3 and SI Figure 5.2).

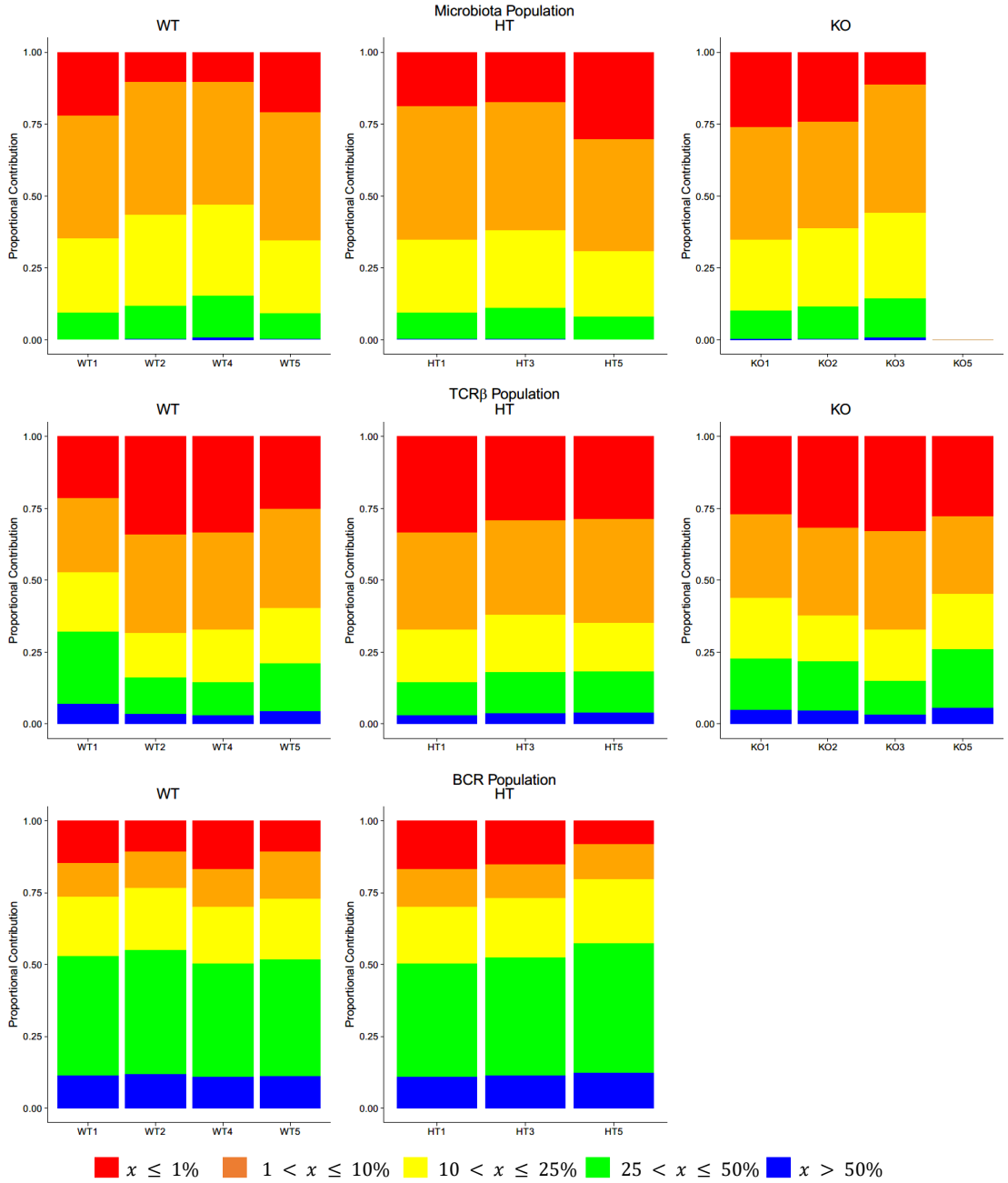


Figure 5.3 – Population structures of the microbiota and immune populations for each bird by genotype. 'HT' refers to the heterozygote birds, 'WT' to the wild-type birds and 'KO to the knock-out birds. The BCR data presented is the raw dataset. The allocation of specific clones to each abundance group is performed within each bird. Only the first technical replicate is presented.

The birds that were significantly different from each other in terms of population structure varied between the immune populations and the microbiota (Figure 5.3). In the

microbiota the heterozygote bird HT5 grouped with the knock-out birds KO1 and KO2 in a relative expansion of abundant species, however in the TCRβ population variation between the birds was driven largely by the knock-out bird KO5 and wild-type bird WT1 showing a reduction in the contribution of large clones and an increased importance of rare clones to the TCRβ community.

As would be expected there was an effect of bird, genotype and B-cell status upon the relative importance of differentially abundant BCR clones to the B-cell population. However the effect of genotype was driven entirely by the knock-out birds differing from the B-cell sufficient birds (SI Table 5.3,Figure 5.3).

[C.3] Large-scale community composition of B-cell sufficient and knock-out birds

The species richness of the microbiota was consistent between all 11 birds interrogated, although both the microbiota's Shannon's diversity (p-value $7.79e^{-03}$, R^2 adjusted 0.637) and Simpson's entropy (p-value $3.49e^{-07}$, R^2 adjusted 0.886) indices varied between birds (Figure 5.4). Similarly, the diversity of the TCRβ and BCR populations varied between birds, in all three diversity estimates (SI Table 5.4).

The birds driving the difference in microbiota diversity varied depending upon the diversity estimate used (SI Table 5.4), though the microbiota of wild-type bird WT2 was consistently more diverse than most other birds (Figure 5.4). There was more consistency in which birds drove the differences in diversity of the immune populations (SI Table 5.4), though there was no consistent pattern between the immune populations or between the microbiota and either the TCRβ or BCR populations.

In the case of the BCR population the variation in species richness between birds was driven entirely by the lack of BCR in the knock-out birds. Different B-cell sufficient birds showed variation in their BCR diversity in terms of Shannon's diversity (p-value $1.37e^{-02}$, R^2 adjusted 0.638) and Simpson's entropy (p-value $4.06e^{-02}$, R^2 adjusted 0.566). Although there was a significant difference in the BCR diversity of B-cell sufficient and knock-out birds, as would be

expected, there was no difference between the BCR diversity of the wild-type and heterozygote birds in any of the diversity indices.

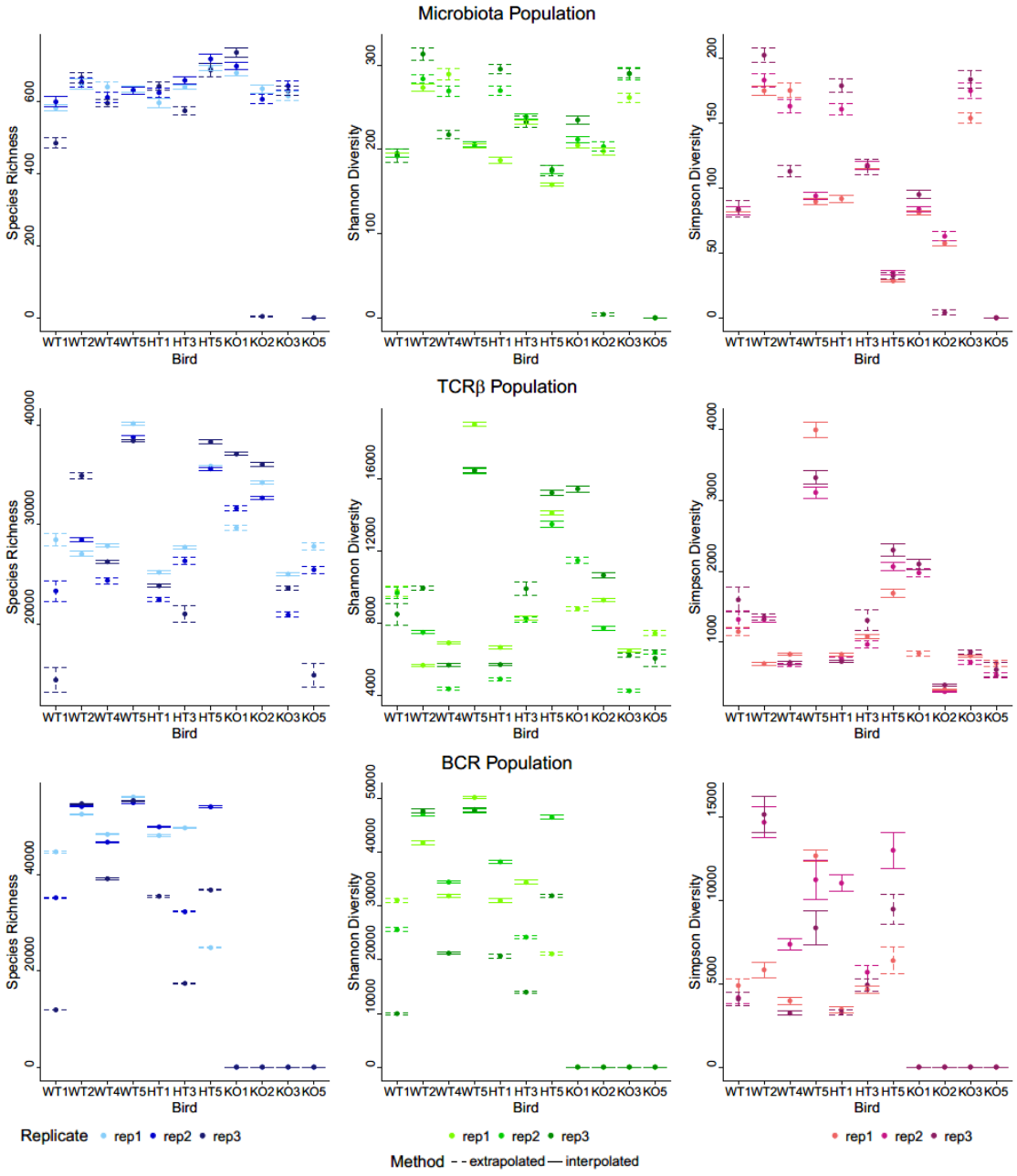


Figure 5.4 – The diversity estimates of the microbiota population of each of the birds at a large-scale. Species richness (blue), Shannon's diversity index (green) and Simpson's diversity index (red) of each technical replicate are displayed for each bird. 'WT' refers to the wild-type birds, 'HT' to the heterozygotes and 'KO' to the homozygote knock-out birds. Estimates were calculated to the same end point either through interpolation or extrapolation, as indicated. The standard error of the estimates (95% confidence interval) is shown.

Similarly, though there was a significant effect of bird upon the diversity of the TCRβ population, there was no effect of genotype upon a bird's TCRβ diversity. The B-cell status of a

bird did significantly affect the Simpson's entropy of its TCRβ repertoire (Welch's t-test p-value 1.82e-02), but not its species richness or Shannon's diversity.

[C.3.1] Microbiota taxonomic composition

The Firmicutes phylum consistently dominated the microbiota of every bird, contributing *c.* 90% of the microbes to the community. The proportional contribution of the rarer phyla within the microbiota (including the Actinobacteria, Bacteroidetes, Cyanobacteria, Euryarchaeota and Proteobacteria phyla; SI Table 5.5), however, varied between birds. The variation in microbiota composition between birds was largely independent of genotype or B-cell status; for example the Cyanobacteria phylum was significantly more abundant in wild-type bird WT2 than in any of the other wild-type birds but didn't vary in abundance between any of the heterozygote or knock-out birds (Figure 5.5).

Only a single phylum (Euryarchaeota) varied between birds of different genotype (p-value $1.26e^{-03}$, R^2 adjusted 0.931) and B-cell status (p-value $1.10e^{-04}$, R^2 adjusted 0.930). The Euryarchaeota phylum was significantly expanded within the knock-out birds compared with the B-cell sufficient birds (Figure 5.5). The difference in Euryarchaeota between genotypes, however, resulted from variation between single birds of each genotype; knock-out bird KO1 showing an increased presence of Euryarchaeota relative to the heterozygote bird HT5 and wild-type bird WT2.

Seven families varied significantly between birds, though only three showed any variation between the genotypes. The Elusimicrobiaceae (p-value $1.39e^{-22}$, R^2 adjusted 1), Mogibacteriaceae (p-value $2.94e^{-02}$, R^2 adjusted 0.999) and Ruminococcacea (p-value $3.99e^{-02}$, R^2 adjustsed 0.227) families contributed differentially to the microbiota between the genotypes. The Ruminococcaceae family genotypic variation was between the heterozygote and wild-type birds (Figure 5.5) with an increased presence within the heterozygote population, whilst the Mogibacteriaceae family was only consistently present within the wild-type bird WT4 (Figure 5.5). The Elusimicrobiaceae presence, however, varied between the knock-out birds and the B-cell sufficient birds (p-value $5.06e^{-25}$, R^2 adjusted 1). Elusimicrobiaceae was

significantly more abundant within the microbiota community of the knock-out birds than in either of the B-cell sufficient bird genotypes (Figure 5.5).

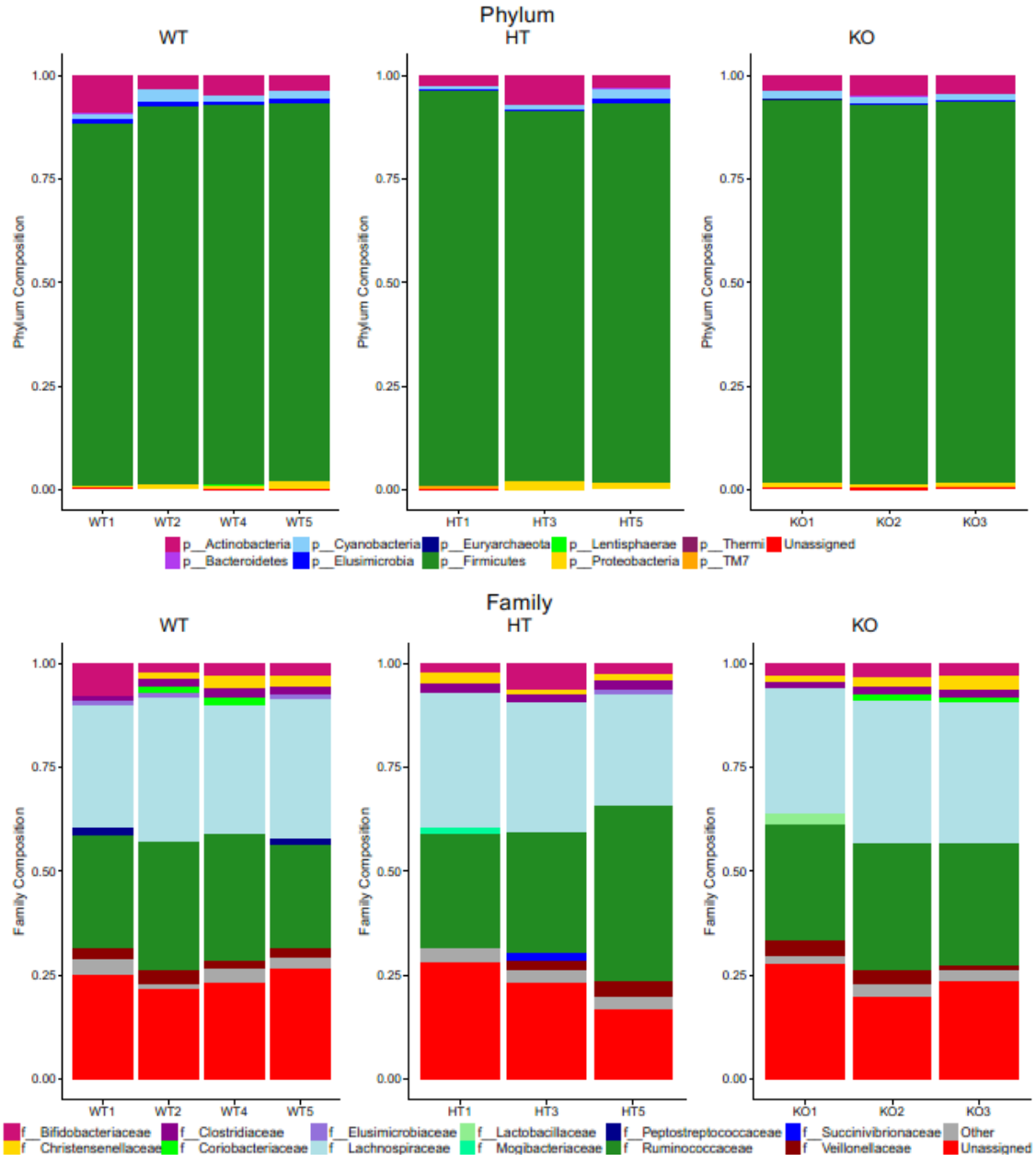


Figure 5.5 – The proportional contribution of different taxa to the microbiota of the wild-type ('WT'), heterozygote ('HT') and homozygote knock-out ('KO') birds. The taxonomic assignment was at the Phylum (upper) or Family (lower) level. Family assignments that contributed less than 1% of the total sequences, within a bird, were compressed into a single 'other' category (SI Table 4.8 for full family list). Only the first technical replicate of each bird is indicated.

[C.3.2] Vβ and Jβ usage in the TCRβ populations

In every bird the TCRβ repertoire showed the previously reported bias towards Vβ2 family usage[301]. There was significant variation between birds, however, in the extent of the skew towards Vβ2 usage. The variation in Vβ2 utilisation between birds was not associated with genotype or B-cell status but driven by a single bird of each genotype. Specifically wild-type bird WT1, heterozygote HT3 and knock-out bird KO5 all showed reduced utilisation of the Vβ1 gene relative to the other birds (SI Figure 5.3).

Within each Vβ family there was a skew in the Vβ-associated Jβ usage (e.g. Jβ3 was preferential utilised within the Vβ2 family). Within Vβ Jβ usage also varied between birds in the Vβ1-Jβ2, Vβ1-Jβ4 and Vβ2-Jβ3 families (Table 5.2). In each case the between bird variation was due to a single distinct bird (e.g. knock-out bird KO2 had significantly higher usage of Vβ2-Jβ3 relative to wild-type bird WT4, heterozygote bird HT1 and knock-out bird KO3). There was no effect of genotype or B-cell status upon Vβ or Jβ gene usage within the TCRβ repertoire.

[C.3.3] Public adaptive immune receptors

At the nucleotide level of the TCRβ population, 19 sequences were shared in all four wild-type birds, 181 sequences were found in all three heterozygote birds (roughly consistent with the wild-type and knock-out birds; 187 and 181 sequences shared respectively) and 26 nucleotide sequences were shared between all the knock-out birds. Only four of the shared heterozygote sequences, and two of the shared knock-out sequences, were moderately abundant (> 50 counts) and these were only moderately abundant within a single individual. None of the shared wild-type sequences comprised more than 20 counts in any bird.

Table 5.2 – The significance of differential gene usage between birds, in terms of both V gene segments and their associated J segments. Values are given to three significant figures. Birds that are significantly different from each other are indicated, and those that are different from more than four other birds are highlighted in grey. Values are given to three significant figures.

V Subset	J Subset	P-value	R^2 Adjusted	Different Birds
Vβ1	*NA*	4.44e-08	0.884	WT1-WT2/WT4/WT5/HT1/HT5/KO1/KO2/KO3 HT3/KO5-WT2/WT4/WT5/HT1/HT5/KO1/KO2 /KO3
	Jβ2	4.41e-03	0.633	WT4-HT3/KO5
	Jβ4	9.70e-03	0.598	KO1-WT5
Vβ2	*NA*	4.95e-08	0.883	WT1-WT2/WT4/WT5/HT1/HT5/KO1/KO2/KO3 HT3-WT2/WT4/WT5/HT1 HT3-HT5/KO1/KO2/KO3 KO5-WT2/WT4/WT5/HT1/HT5/KO1/KO2/KO3
	Jβ3	3.00e-03	0.648	KO2-WT4/HT1/KO3

11 nucleotide sequences were found in all the B-cell sufficient birds and three sequences (Figure 5.6 sequences 11, 17 and 19) were found in all the birds (SI Table 5.7). Though numerically rare, these three sequences may represent a public TCRβ repertoire.

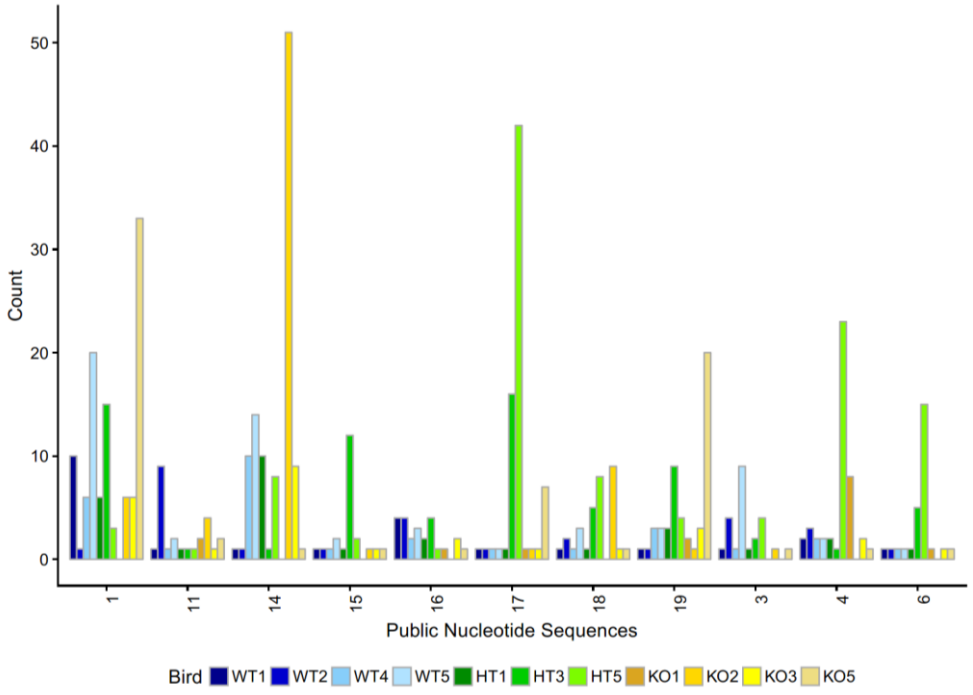


Figure 5.6 – TCRβ nucleotide sequences counts for sequences present in all B-cell sufficient birds. Corresponding sequence IDs are in SI Table 5.7.

At the amino acid level 119 TCRβ sequences were shared between the wild-type birds, 571 within the heterozygote birds and 161 between the knockout birds. Of the shared amino acid sequences relatively few were found moderately abundantly (> 50 counts) and in each case usually only within a single bird. 12 shared amino acid sequences were moderately

abundant within the wild-type birds, though only one was found at moderate levels within more than one bird (CAKQGGSNTPLNF in birds WT4 and WT5). Interestingly the same amino acid sequence was moderately abundant within three of the four knock-out birds (excluding KO5) and a similar derivation moderately abundant within most of the heterozygote birds (CAKQGGSSTPLNF in HT3 and HT5).

The B-cell sufficient birds shared 95 public TCRβ amino acid sequences, of which 53 were also found within all the knockout birds. 23 of these public amino acids were moderately abundant (> 50 counts), and five of these sequences were moderately abundant in more than one bird (Table 5.3). Three amino acid sequences were moderately abundant within at least one of the B-cell sufficient birds but none of the knock-out birds.

Table 5.3 – Moderately abundant TCRβ amino acid sequences found within the B-cell sufficient or all the birds. Sequences that were moderately abundant within one of the B-cell sufficient birds, but none of the knock-out birds, are highlighted in grey.

Amino Acid Sequence	HT1	HT3	HT5	WT1	WT2	WT4	WT5	KO1	KO2	KO3	KO5
CAKQDERLIF	1	9	10	1	1	1	102	2	18	1	1
CAKQDRGIDTPLNF	3	8	10	3	7	3	111	20	33	2	5
CAKQDRGINNERLIF	65	10	22	4	7	58	26	48	17	53	21
CAKQDRGITPLNF	2	143	24	29	123	2	7	19	11	1	18
CAKQGDNERLIF	3	5	50	1	3	3	205	3	47	3	2
CAKQGGSNTPLNF	73	5	34	3	28	71	95	51	62	52	8
CAKQGGSSTPLNF	7	71	51	10	12	7	10	11	7	6	12
CAKQGGSYSNMIF	4	25	19	18	61	4	28	24	5	1	11
CAKQGTGGNTPLNF	17	120	9	24	3	17	6	9	4	13	18
CAKQGTGGSNERLIF	1	5	16	2	134	1	9	50	3	1	2
CAKQGTGGYSNMIF	20	35	38	7	12	19	93	9	90	18	11
CAKQGTGGYTPLNF	13	77	22	5	16	12	22	44	9	12	9
CASSLGTGGYSNMIF	9	4	6	1	59	9	12	25	8	8	5
CAKDRGIDIQYF	66	3	21	1	5	66	5	18	6	6	0
CAKQDRGIYSNMIF	3	58	6	13	2	2	4	4	2	0	7
CAKQGTGGNMIF	1	3	51	1	23	1	3	24	2	0	17

As shown in Chapter 4, using the raw BCR data, no BCR amino acids were shared between all four wild-type birds. 2106 BCR amino acids were found in more than one wild-type bird, though only one amino acid sequence was found in three of the four wild-type birds (see Chapter 4 C.4). The shared wild-type amino acids sequences varied in abundance between the wild-type birds (SI Figure 5.4), and were usually more abundant within a single individual. 85

amino acid sequences were moderately abundant (> 50 counts) in the wild-type birds, though only 11 of these sequences were found at moderate abundance in more than one bird.

Within the heterozygote birds there were 88 BCR amino acid sequences found in multiple birds, though none were found in all three heterozygote birds. The shared heterozygote amino acid sequences were generally less abundant than those found shared between the wild-type birds (SI Figure 5.4); only four heterozygote amino acid sequences were moderately abundant, and none of these were found in more than one bird.

57 amino acid sequences were found in multiple wild-type and heterozygote birds; almost all of which were found in wild-type bird WT5. However only a single amino acid sequence was found in the majority of the B-cell sufficient birds (SI Figure 5.4 ID 1108; see Chapter 4 C.4 for sequence).

The shared BCR sequences were generally more abundant than the shared TCRβ amino acid sequences, though less consistently present between B-cell sufficient birds. As the knock-out birds should not be producing BCR sequences, those sequences retrieved were not investigated.

[C.4] Fine-scale population structure in B-cell competent birds

At a large-scale there were no significant differences in the population structures of the microbiota and TCRβ populations between B-cell sufficient and knock-out birds. However, as indicated in Chapter 2 D.2 and Chapter 4 C.3, the scale of the samples considered is important when understanding community dynamics and variation between communities can arise at a small scale that are not apparent with global sampling.

The variation in population structure between segments was most consistent within the TCRβ population of the wild-type birds (see Chapter 4 C.3), excluding only wild-type bird WT4. Segmental variation was similarly present in the TCRβ populations of the heterozygote and knock-out populations, in each case only absent in a single bird (heterozygote HT3 and knock-out bird KO5).

The effect of segment upon microbiota population structure was less apparent. Only knock-out bids KO1 and KO2 and wild-type bird WT4 showed any evidence of location changing community structure (SI Table 5.8), and this effect was restricted to the contribution of the moderately abundant species (abundance group 4). None of the B-cell sufficient birds showed evidence of segmental variation in the contribution of differentially abundant BCR clones.

Genotype did not significantly affect the mean proportional contribution of either TCRβ clones or microbial species (SI Table 5.9). There was, however, a significant difference in the mean contribution of rare species and abundant TCRβ clones to their respective communities between the B-cell sufficient and knock-out birds (Table 5.4 and Figure 5.7). Specifically the moderately abundant TCRβ clones (abundance group 3) contribute significantly more to the TCRβ repertoire in the B-cell sufficient birds and rare microbial species significantly less to the microbiota of the knock-out birds.

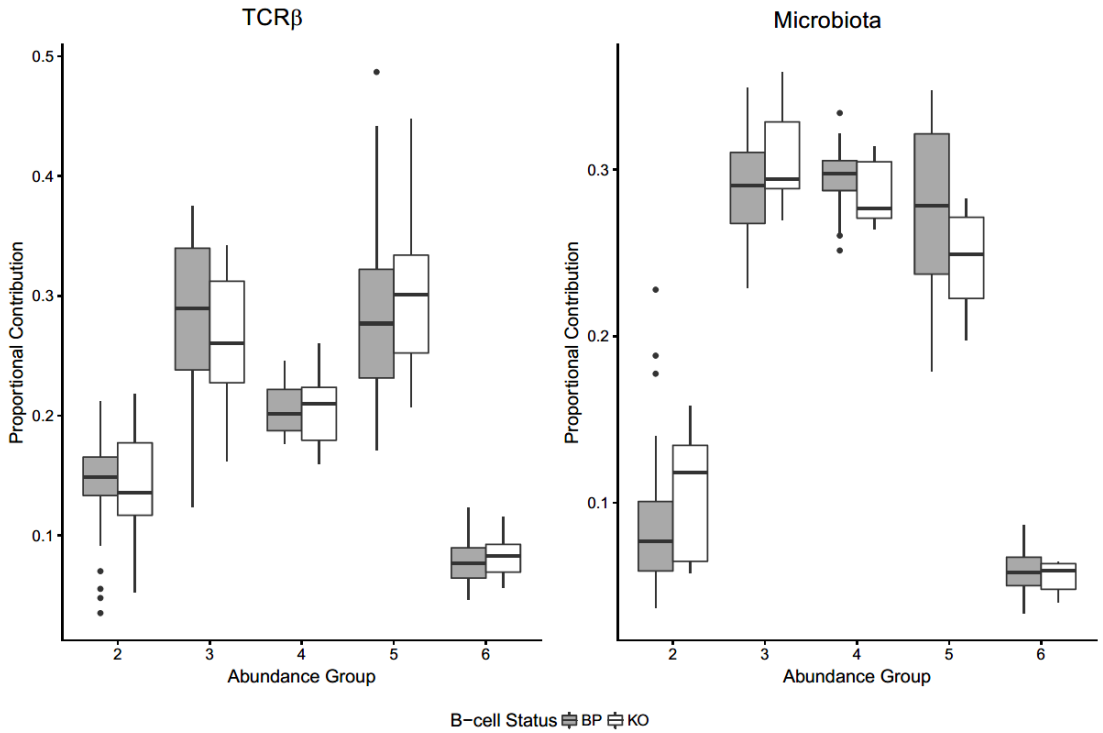


Figure 5.7 – The difference between the mean proportional contribution of different population abundance groups within the microbiota and immune populations between the knock-out birds (KO; white) and the B-cell sufficient (BP; grey) birds. Abundance groups refer to the relative abundance of the clones/species (groups defined from the most abundant species/clones in group 2 [$x \leq 1\%$], through group 3 [$1 < x \leq 10\%$], group 4 [$10 < x \leq 25\%$] and group 5 [$25 < x \leq 50\%$] to the least abundant species/clones in group 6 [$x > 50\%$]). Summary statistics in Table 5.4.

[C.5] Fine-scale population composition in B-cell sufficient and knock-out birds

The mean segment species richness of both immune populations varied between birds (Table 5.5), although this difference in the BCR population was driven almost exclusively by the BCR repertoire being significantly reduced in the knock-out birds (Figure 5.8). Only wild-type bird WT4 and heterozygote HT1 of the B-cell sufficient birds showed any distinction in BCR species richness (Figure 5.8) and this may result from the greater inter-replicate variation within the BCR repertoire of these birds. There was no significant variation between birds in the species richness of the microbiota (Table 5.5).

Table 5.4 – The difference between the mean proportional contribution of different population abundance groups within the microbiota and immune populations between the knock-out birds (KO) and the B-cell sufficient (BP) birds. Group refers to the relative abundance of the clones/species (groups defined from the most abundant species/clones in group 2 [$x \leq 1\%$], through group 3 [$1 < x \leq 10\%$], group 4 [$10 < x \leq 25\%$] and group 5 [$25 < x \leq 50\%$] to the least abundant species/clones in group 6 [$x > 50\%$]). The difference between the B-cell status groups was assessed using Welch's two sample t-test for each abundance group. BCR population data presented is the raw dataset, the trimmed data can be found in SI Table 5.10. Values are given to three significant figures.

	Microbiota			TCRβ			BCR		
Group	P-value	KO	BP	P-value	KO	BP	P-value	KO	BP
2	6.90e-05	0.110	0.0952	0.182	0.147	0.140	< 2.2e-16	0.000	0.0443
3	1.17e-06	0.309	0.292	2.25e-02	0.267	0.279	< 2.2e-16	0.000	0.114
4	2.55e-08	0.284	0.295	0.615	0.202	0.203	< 2.2e-16	0.000	0.239
5	9.08e-03	0.252	0.272	0.452	0.301	0.297	< 2.2e-16	0.000	0.538
6	1.98e-03	0.0551	0.0591	0.397	0.0827	0.0814	< 2.2e-16	0.000	0.128

Table 5.5 – The significance of the effect of bird upon the diversity estimates within both the microbiota and immune populations. The effect of bird was assessed using a linear model of the form lm(Estimate~Bird), where 'Estimate' is the specific diversity index used. The birds that are significantly different from each other, as assessed using Tukey's HSD test, are indicated. Birds that are different from more than four other birds are highlighted in grey. The BCR population presented here is the raw dataset, a comparison with the trimmed BCR dataset can be seen in SI Table 5.11. All values are given to three significant figures.

Population	Estimate	P-value	R^2 Adjusted	Different Birds
Microbiota	Species Richness	0.232	0.211	*NA*
	Shannon's Diversity	4.66e-06	0.567	WT2-WT1/WT5/HT1/HT5/KO1/KO2 HT3-HT5/KO2/KO3 KO2-KO3
	Simpson's Entropy	6.18e-14	0.820	WT1-HT1/KO3 WT2-WT1/WT4/WT5/HT1/HT3/HT5/KO1/KO2/KO3 HT5-WT1/WT4/WT5/HT1/HT3/KO1/KO3 KO1-HT5/HT1/KO3 KO2-WT4/HT3/HT1/KO3
TCRβ	Species Richness	3.91e-02	0.526	*NA*
	Shannon's Diversity	6.05e-05	0.766	WT1-WT2/WT4/HT1/KO2/KO3 HT1-WT5/HT3/KO1/KO5 WT5-WT2
	Simpson's Entropy	1.99e-04	0.736	WT1-WT2/WT4/HT1/KO2/KO3/KO5 KO2-WT5/KO1
BCR	Species Richness	9.86e-33	0.999	WT1/WT2/WT4/WT5/HT1/HT3/HT5-KO1/KO2/KO3/KO5 WT4-WT2/WT5/HT5 HT1-WT5
	Shannon's Diversity	7.87e-27	0.997	WT1/WT2/WT4/WT5/HT1/HT3/HT5-KO1/KO2/KO3/KO5 WT4-WT2/WT5
	Simpson's Entropy	6.26e-19	0.986	WT1/WT2/WT4/WT5/HT1/HT3/HT5-KO1/KO2/KO3/KO5 WT2-HT3 WT4-WT2/WT5

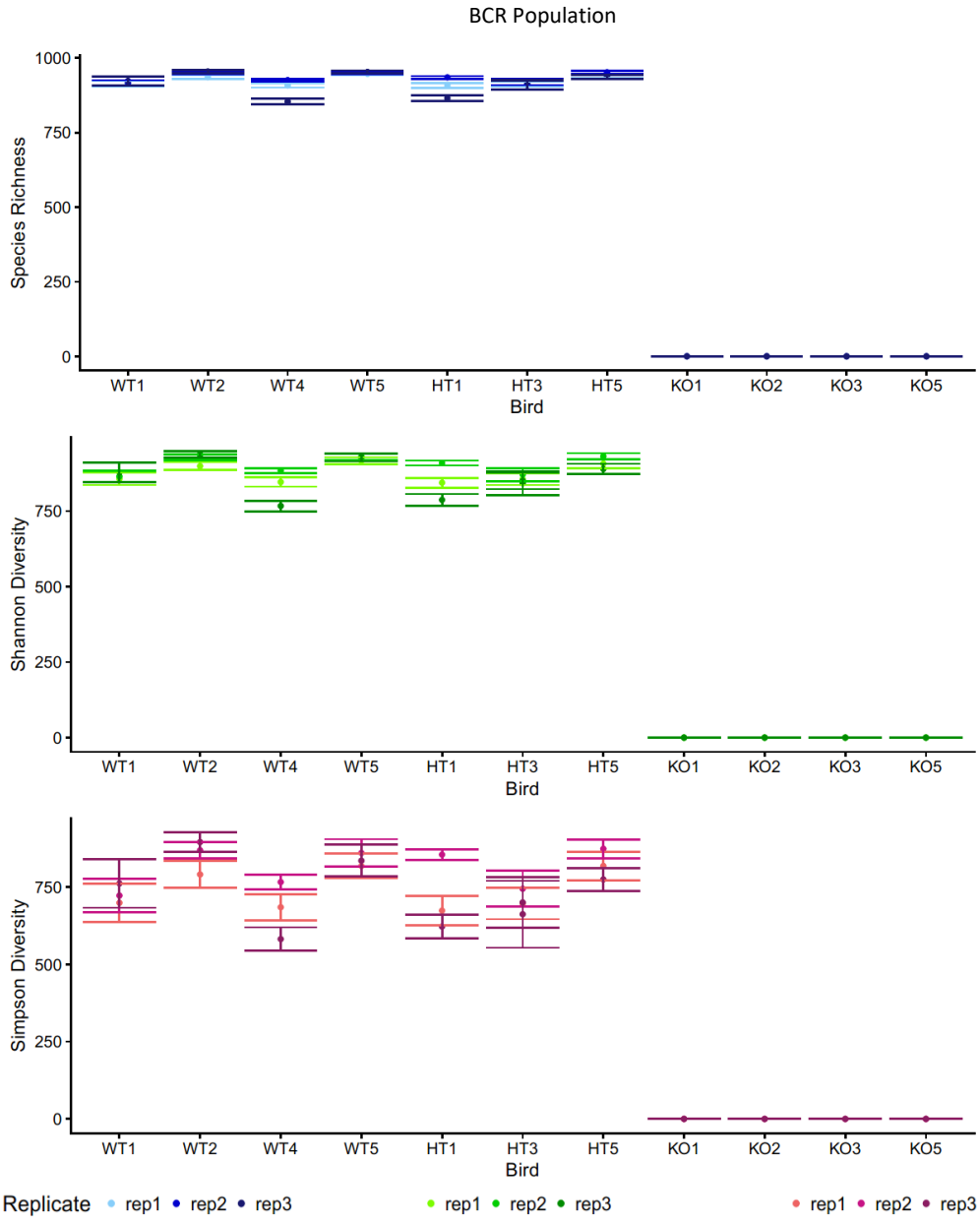


Figure 5.8 – The mean diversity estimates of individual segments within the raw BCR population of each bird. Species richness (blue), Shannon's diversity index (green) and Simpson's diversity index (red) of each technical replicate are displayed for each bird. 'WT' refers to the wild-type birds, 'HT' to the heterozygotes and 'KO' to the knock-out birds. Estimates were calculated to the same end point either through interpolation or extrapolation. The standard error of the estimates (95% confidence interval) is shown. The trimmed dataset can be found in SI Figure 5.5.

Although both Shannon's diversity and Simpson's entropy varied between birds in all the populations considered (microbiota, TCRβ and BCR) there was no association between any of the diversity indices and genotype in the TCRβ and microbiota populations. The BCR repertoire was intrinsically significantly reduced (i.e. absent) within the knock-out birds, though

there was no significant difference in diversity of the BCR populations of the B-cell sufficient genotypes.

[C.5.1] Fine-scale microbiota taxonomy

Although there was no difference in microbiota diversity between the B-cell sufficient and knock-out birds, the mean contribution of six phyla was significantly different depending on B-cell sufficiency (Figure 5.9). As suggested by the variation in the fine-scale microbiota population structure, the contribution of the Actinobacteria (Welch's two sample t-test: p-value $6.19e^{-09}$), Cyanobacteria (p-value $1.35e^{-05}$), Proteobacteria (p-value $1.26e^{-12}$), Lentisphaerae (p-value $4.50e^{-04}$) and TM7 (p-value $3.02e^{-02}$) phyla was significantly higher in the B-cell sufficient birds whilst the Firmicutes phylum was expanded in the knock-out birds (p-value $1.71e^{-11}$).

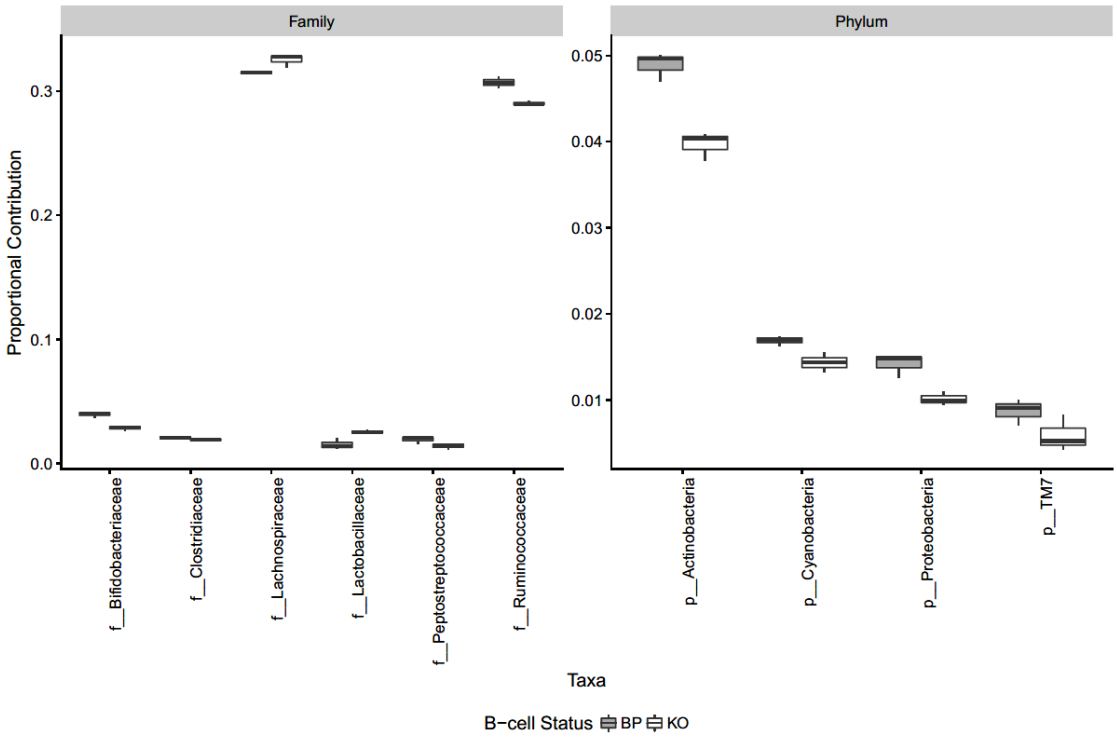


Figure 5.9 – The difference between the mean proportional contribution of different microbiota taxa (phylum and family level assignments) between the knock-out birds (KO; white right of pair) and the B-cell sufficient (BP; grey left of pair) birds. The taxa shown are those that had a significant difference between the B-cell sufficient and knock-out birds.

The contribution of different phyla to the microbiota community varied between segments in all birds except knock-out bird KO3 (SI Table 5.12). The contribution of

Proteobacteria varied between segments in almost all birds, and the Firmicutes phylum was variable across locations within the wild-type birds. The effect of location upon the contribution of different phyla to the microbiota was driven by a single, or small number, of segments that were significantly different to most other segments (e.g. within the Proteobacteria phylum of bird WT1 segment A8 is significantly different from all other segments except C8).

All birds, including KO3, showed an effect of location upon the proportional contribution of different families to the microbiota community (SI Table 5.13). Again segmental variation resulted from a small number of distinct sites in most birds. The variation of Elusimicrobiaceae between segments within knock-out bird KO3, however, came from multiple sites including A2, A5, A6 and B4, B5 and B7. In birds where more than one family showed evidence of spatial variation, the segments driving the effect of location depend upon the specific family interrogated (SI Table 5.13).

The families varying between the B-cell sufficient and knock-out birds reflected the phylum splits as would be expected with the Lachnospiraceae (Welch's two sample t-test: p-value $8.11e^{-03}$) and Lactobacillaceae (p-value $1.43e^{-21}$) families of the Firmicutes phylum expanded within the knock-out birds. The Biffidobacteriaceae (Welch's two sample t-test: p-value $9.68e^{-15}$), Clostridiaceae (p-value $1.21e^{-02}$), Ruminococcaceae (p-value $1.14e^{-06}$) and Peptostreptococcaceae (p-value $2.63e^{-02}$) families contribution to the microbiota were all reduced within the knock-out birds compared to the B-cell sufficient birds (Figure 5.9).

[C.5.2] TCRβ gene family variation between locations

As seen at a large-scale, there was no significant effect of genotype of B-cell status upon the Vβ or Vβ-associated Jβ usage of the TCRβ repertoire. There was evidence of differential usage between segments in three of the B-cell sufficient birds; wild-type bird WT1 and heterozygotes HT1 and HT5. Heterozygote HT5 showed the most variation in TCRβ gene usage between segments, with the proportion of Vβ1, Vβ1-associated Jβ4 and Vβ2 utilised differently in D7 than in almost all the other segments. Wild-type bird WT1 varied in terms of Vβ1-associated Jβ4 usage between segment D5 and a few other segments, and heterozygote

segment A1 difference in its usage of Vβ1-associated Jβ4 and Vβ2-associated Jβ3. In all cases where TCRβ gene usage varied between segments the difference was driven by a single segment varying (Table 5.6).

Table 5.6 – V gene segments and their associated J segments in birds where there was significant variation in gene-usage between segments. The specific segments which differ from each other are indicated, and highlighted in grey where more than four segments differ from the focal segment. Values are given to three significant figures. 'All' J-subset refers to the combined V-family usage regardless of the associated J segments.

V Subset	J Subset	Bird	P-value	R^2 Adjusted	Different Segments
Vb1	*All*	HT5	1.97e-04	0.356	D7-A1/A2/A3/A4/A5/A6/A7/A9/B1/B2/B3/B4/B5/B7 /B8/B9/C1/C2/C3/C4/C5/C6/C7/C8/C9/D1/D2/D3 /D4/D5/D6/D8/D9
	Jb1	WT1	4.30e-02	0.179	D5-A6/A7/C4/C5/C6/D3
	Jb4	HT1	2.64e-02	0.210	A1-A2/A4/A6/A7/A8/B1/B3/B5/B6/B7/B8/C1/C5/C6 /C7/C8/D1/D2/D3/D5
		HT5	3.59e-03	0.269	*NA*
Vb2	*All*	HT5	2.39e-04	0.351	D7-A1/A2/A3/A4/A5/A6/A7/A9/B1/B2/B3/B4/B5/B7 /B8/B9/C1/C2/C3/C4/C5/C6/C7/C8/C9/D1/D2/D3 /D4/D5/D6/D7/D9
	Jb3	HT1	1.76e-02	0.229	A1-A4/C5/C6/D2

[C.6] Spatial structure in B-cell sufficient and knock-out birds

With the evidence of taxonomic variation within the microbiota apparent, the existence of spatially structured differences was interrogated. As in Chapter 2 and Chapter 4, the association between two segments similarity and their distance was used as evidence of spatial structuring across the grid. Three dissimilarity indices, which differentially account for species presence and abundance, were used to assess the similarity between segments (see Chapter 2 D.5).

The mean dissimilarity of the microbiota and TCRβ populations did not vary between birds, genotypes or B-cell sufficiencies (Figure 5.10). As would be expected, given the lack of BCR data in the knock-out birds, the BCR population was more dissimilar between segments of the B-cell sufficient birds. There was no difference in the mean dissimilarity of the BCR population within the B-cell sufficient genotypes, though wild-type bird WT4 and heterozygote bird HT1's BCR population were significantly more dissimilar between segments than wild-type birds WT2 and WT5.

None of the wild-type or knock-out birds showed any evidence of spatial structure in the microbiota. However heterozygote birds HT1 and HT5 showed the same positive correlation between dissimilarity and distance seen in the preliminary L6 bird (SI Table 5.14). Similarly, unlike in the preliminary bird, none of the JH birds showed any evidence of a spatially structured BCR population.

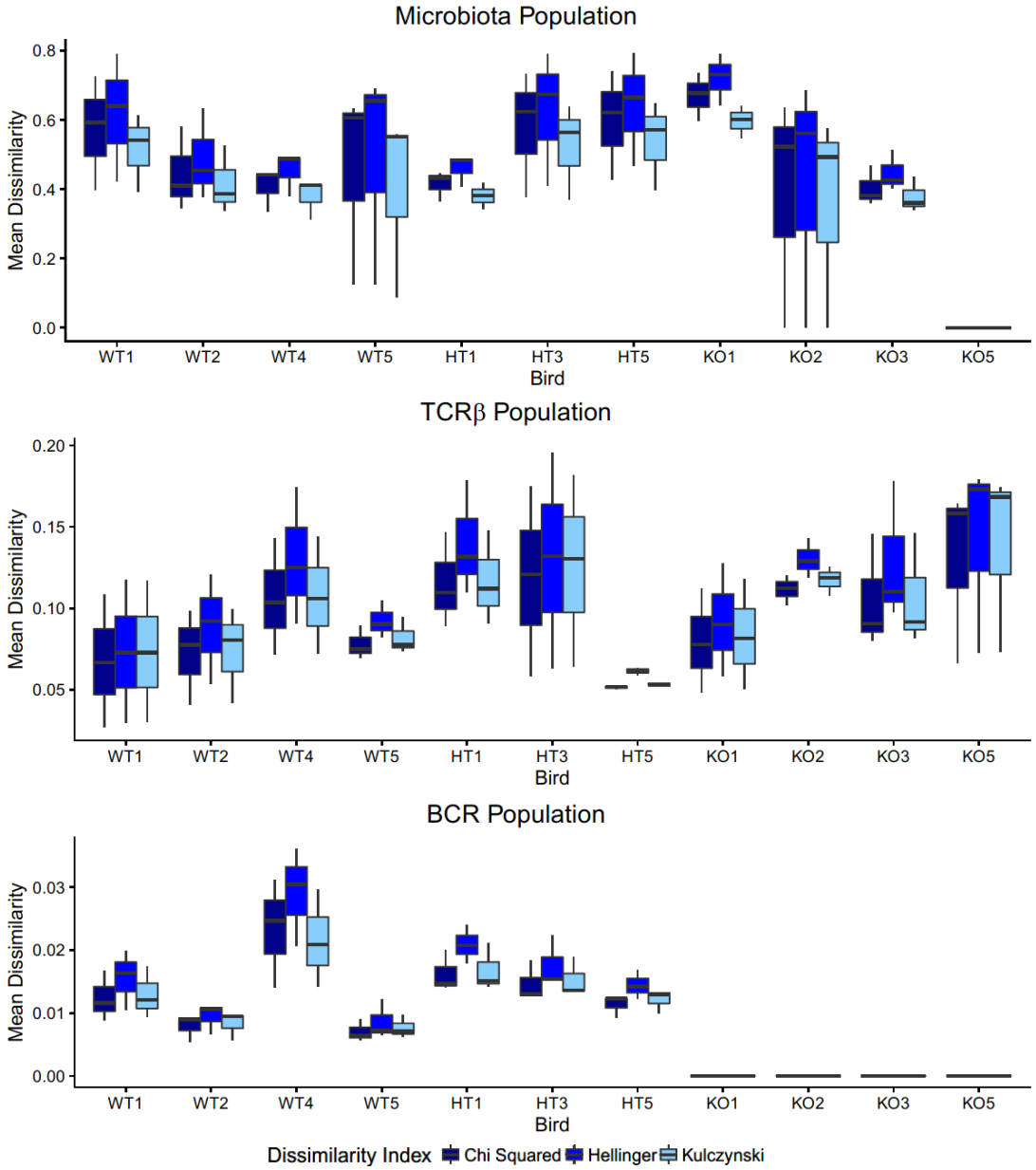


Figure 5.10 – Mean dissimilarity estimates of individual segments, in both the microbiota and immune populations, for all the birds. Dissimilarity was calculated using the Chi Squared, Hellinger and Kulczynski indices as indicated. 'WT' refers to wild-type birds, 'HT' to heterozygote birds and 'KO' to homozygote knock-out birds. Bird KO5 had an incomplete DNA extraction (see B.4) and as such did not produce any data for the microbiota population. The BCR population displayed is the raw dataset.

As shown in Chapter 4, there was an association between dissimilarity and distance in the TCRβ population of wild-type bird WT2. Furthermore two of the heterozygote birds, HT1 and HT5, showed evidence of spatial structuring using all three of the dissimilarity indices (SI Table 5.14). However only a single homozygote knock-out bird, KO2, showed any evidence of a correlation between segment TCRβ dissimilarity and distance (SI Table 5.14).

There was no apparent effect of B-cell status or genotype on the potential for spatial structuring within any of the populations considered (SI Table 5.15).

[D] Discussion

The aim of this chapter was to investigate whether there was any difference within the immune repertoires and microbiota of birds that are B-cell sufficient or lack B-cells. Immunoglobulin A (IgA) has been demonstrated to effect the composition of the microbiota within mice[219,368,369], though the phenotype presented is mild[403,404]. The avian model presented here is of a complete B-cell knock-out[360], so the effects upon the microbiota and the other adaptive immune components could be expected to be more substantial. However at large-scale (1 cm^2) there appears to be only mild differences between the B-cell knock-out and B-cell sufficient birds; including a reduced caecal diameter and differential contribution of the Euryachaeota phylum (Figure 5.5). Instead the knock-out shows variation within the microbiota and TCRβ populations that can only be seen when small-scale structures (1 mm^2) between the treatment types are interrogated (C.4 and C.5).

[D.1] Mild B-cell knock-out phenotype

As would be expected the proportion of functional B-cell receptors was reduced within the knock-out birds relative to the B-cell sufficient birds (Figure 5.2), and there were consistent differences between the birds BCR populations based upon B-cell sufficiency in terms of population structure and diversity at both scales (Figure 5.3 and Figure 5.8). The microbiota of knock-out birds showed an increased presence of Euryachaeota at a large scale (Figure 5.5),

but no other variation within the microbiota or TCRβ community's population structures or diversity.

The B-cell knock-out phenotype in mice is mild, both in terms of its impact upon the immune system and microbiota[403,404]. However the absence of a strong IgA knock-out phenotype in mice has been suggested to arise as a result of the murine B-cell knock-out (μMT[397]) retaining some B-cell functionality or transporting immunoglobulin M across the luminal epithelial instead[399,400]. The B-cell knock-out interrogated here is a complete B-cell knock-out, with no evidence of antibody production or B-cell receptor formation[360]. The consistency of the mild phenotype displayed by the B-cell knock-out birds with that reported in mice suggests, therefore, that the phenotype displayed is not an experimental flaw but may be a function of the scale at which the effects are considered.

[D.2] Spatial variation exists in all treatment types

Each of the birds showed evidence of spatial variation at a small-scale within the microbiota, TCRβ and BCR population structures and/or compositions. The variation between grid locations was driven almost entirely by a single or few distinct segments. During the sampling, efforts were made to avoid sections of the gut which had clear Peyer's patches, however it is possible that these unusual segments contained isolated lymphoid follicules that were unapparent. This is supported by the immune populations within these segments being more diverse than in other locations within the grid.

[D.3] Phenotypic variation in B-cell knock-outs at a fine-scale

There was no difference in the knock-out birds and the B-cell sufficient birds in terms of diversity. In particular, within the microbiota species richness did not even vary between birds let alone genotype. The mean dissimilarity between segments was similarly unaffected by genotype and B-cell status within the microbiota and TCRβ populations.

However, despite almost no evidence of B-cell effects within the large-scale microbiota and TCRβ populations, there was a significant difference in the population structures of both these populations at a small-scale. Specifically the contribution of rare species within the

microbiota of knock-out birds was significantly reduced whilst the TCRβ population shows an increase in the contribution of moderately abundant clones within B-cell sufficient birds.

The decrease presence of rare species within the knock-out bird's microbiota was similarly present taxonomically at a fine-scale; specifically the absence of B-cells lead to an expansion in the contribution of the Firmicutes phylum to microbiota. The reduced presence of rare microbial species in B-cell knock-out birds supports the hypothesis that the absence of B-cells leads to a reduced survival of rare phenotypes within the microbiota community. The mechanism by which rare species are lost is not, however, immediately clear. The reduced contribution of rare species to the microbiota could arise as a result of the expansion of dominant microbes normally kept in check by IgA or from a lack of positive selection for less evolutionarily viable microbes. The expansion of the moderately abundant TCRβ clones in B-cell intact birds could indicate evidence of T-cell mediated activation of B-cells present in these birds.

[D.4] Selection within the microbiota

Chapter 3 predicted that the relative rare species may be important in facilitating stability within the microbiota, whilst Chapter 1 suggested that a host may maintain rare species through positive selection. The relative expansion of rare species within the microbiota of B-cell sufficient birds compared to knock-outs corroborates these predictions.

However there is an alternative interpretation of these results: negative-frequency dependence. Negative-frequency dependence might facilitate rare species persistence through selecting against abundant species rather than selecting for rare species. Specifically by selecting against abundant species, rarity becomes an advantageous trait. This process has been shown to drive strain diversification by the immune system in some pathogenic species[405] and maintain polymorphisms within ecological communities[406].

[D.5] Limitations

As highlighted in Chapter 4, the microbiota and immune populations of birds mature at *c.* six to eight weeks of age. The birds sampled within this studied were all sacrificed at six-

weeks, consistent with market-age. However the sampling of birds at six-weeks may be premature when considering the dynamics of the relationship between the microbiota and the immune system.

The method of investigating the presence of spatial structuring may also be flawed in its reliance on a continuous variation in populations along the caeca. If the patterning in the microbiota and immune communities is not continuous but rather repeating across space, the association between dissimilarity and distance would not capture this spatial structuring. As mentioned in Chapter 2, it could be better to use ecological methods of spatial autocorrelation.

The difference in the mean proportional contribution of the microbiota rarity groups between B-cell sufficient and knock-out birds was small (always less than 3%) but significant. This difference in effect size is small which could indicate that there are other factors more significant factors involved in this difference or alternatively that the number of samples use to compute this comparison is large (which could call into question the reliability of the results). Given the larger number of samples present at a fine-scale, due consideration should be taken of the small difference between the B-cell sufficient and knock-out birds.

[E] Conclusions

The aim of this chapter was to consider whether there is a difference in the microbiota and adaptive immune communities of B-cell sufficient and B-cell knock-out birds. Both large-scale and small-scale interrogation of all these communities was used to assess the differences in population structure and composition between the treatment types. At a large-scale the effect of B-cell sufficiency upon the microbiota and TCRβ populations, in terms of either structure or diversity, was mild. However at small-scales there were differences in the population structures and compositions of both the microbiota and TCRβ communities. Specifically, B-cell sufficient birds showed an increased presence of rare species within the microbiota at both a phylum and family level.

The existence of population variation resulting from B-cell sufficiency at small-scales, whilst absence at large-scales, suggests that the community effects of immune status upon

other immune components and microbiota interactions are scale dependent and may not be immediately apparent at the scale usually studied. As apparent in ecology, the scale at which a system is studied is important when considering the factors that influence community dynamics. This seems particularly important when considering the long-term effects that the microbiota composition and stability have upon the host. In unravelling the effects of the immune system upon the microbiota and the microbiota upon the immune system, it is therefore integral to consider changes that significantly affect the fitness of a host may not be apparent at the scales normally studied.

General Discussion

Understanding the dynamics of the microbiota community is incredibly important in any attempt to modulate host health through microbiota modification. Given the extensive effects of the microbiota upon every aspect of host health[11,12,21,13–20], the potential for prophylactic treatments via the microbiota is huge. Even focusing simply upon developing treatments for individuals with microbiota dysbiosis, an understanding of the dynamics of a healthy and unhealthy microbiota is necessary. By only considering the effects of treatments and immune interventions upon large-scale microbiota composition, community interaction information is lost. This may be especially important when considering the importance of specific immune molecules upon microbiota composition.

The microbiota, and the adaptive immune system, are simply ecological communities. Ecological theory has shown that the scale at which a community is considered changes the importance of selection pressures upon community composition[1,5,6]. The work presented within this thesis has shown that the scale at which the microbiota and adaptive immune system are considered can reveal distinct information about microbiota interactions and the effects of the immune system upon microbiota composition.

[A] Summary of findings

Chapter 1 of this thesis considered the effect of host-provided adhesion (e.g. immunoglobulin A (IgA)) upon microbiota dynamics within an individual-based model of biofilm growth. The provision of specific adhesion within the microbiota facilitated spatial structuring and the maintenance of disadvantaged species within the biofilm in a mucus-flow dependent manner. One of the predictions of this model is that fine-scale spatial structuring exists within the microbiota.

Given the hypothesis of spatial structuring within the microbial community, a method was developed to investigate whether there is evidence of fine scale heterogeneity in the

microbiota and adaptive immune system (Chapter 2). The methodology developed within Chapter 2 was interrogated for sampling and processing biases and used to provide preliminary evidence of spatial heterogeneity within the microbial and IgA-producing B-cell receptor populations of a L6 chicken caecum. Importantly, in line with ecological theory, both the immune and microbiota communities varied between large and fine scales.

At large scales, ecological communities are shaped by global effects, such as environmental variation[1,6]. The use of a fine-scale grid structure within this thesis allowed the development of a method estimating community interactions at a scale at which species interactions would be expected to shape community dynamics. By grouping species in terms of their global rarity (defined by either population abundance or geographic spread), it was possible to see that rare species experience distinct interaction environments from their more abundant contemporaries. Specifically, rare species experience more weak and negative interactions than abundant species, which suggests these species could be stabilising to the microbiota community[249,322,336].

The analysis within Chapter 3 relies upon the assumption of community stability within the six-week-old chicken microbiota. However, the inconsistency of the patterns within the preliminary L6 bird and the JH wild-type birds (Chapter 4), and previous literature, suggest that the dynamics within the avian microbiota and immune system may not have stabilised at six-weeks[88–90]. With this caveat, a comparison between the large and fine-scale microbiota and TCRβ communities in B-cell sufficient and knock-out birds was conducted (Chapter 5).

At a large-scale, the effect of B-cell sufficiency upon the microbiota was mild, with only a single phylum varying between the B-cell competent and knock-out birds. However, at smaller scales, there were significant differences in the population structures and compositions of both the microbiota and TCRβ communities. In conjunction with the predictions of Chapter 3, that rare species are experiencing fundamentally different interaction environments in a way that may maintain community stability, and Chapter 1, that the secretion of IgA may facilitate the maintenance of disadvantaged species within the microbiota, the finding of Chapter 5 that, at

small-scales, B-cell sufficient birds showed an increased contribution of rare species within the microbiota suggests the mildness of the B-cell knock-out phenotype belies a fundamental difference in the microbiota community dynamics within these birds.

[B] Future work

There are a set of apparent extensions to this work that could be explored to refine the conclusions presented.

[B.1] Replicability of fine-scale analysis

The trends presented in Chapter 2 varied between the L6 preliminary bird and the JH wild-type birds (Chapter 4). The inconsistency of these trends could result from the differences in the birds' genetic background, the scale at which the samples were taken, or an inherent variability in the timing of the stabilisation of the chicken gastrointestinal microbiota. A set of experiments interrogating the replicability of spatial structuring in birds sacrificed at six-weeks-old and then at two-week intervals until 20 weeks of age[88], in a single genetic background, would allow the effect of microbiota stabilisation upon the presence of spatial structuring to be assessed. Samples from birds with a stabilised microbiota could then be used to assess the effect of scale upon the presence of spatial heterogeneity within the caecal microbiota, focusing on comparing the 4 mm^2 and 1 mm^2 scales. Depending upon whether age and scale accounts for the variability shown here (Chapter 4), further experiments comparing birds of distinct genetic backgrounds (in particular inbred and outbred lines) could be interrogated.

Having determined a useful scale and age at which to assess the spatial patterning of the microbiota, the comparison between B-cell sufficient and B-cell knock-out birds could be repeated to determine whether the B-cell knock-out phenotype remains mild once the microbiota has stabilised.

[B.2] The role of rarity in community stabilisation

Rare species, within the preliminary L6 bird, appear dynamically distinct from abundant species. However the estimation of interactions has only been performed within a single

individual. Therefore, exploring the reproducibility of these results in multiple individuals would be important. Understanding how these interactions change across spatial scales may also be informative given the current debate on the utility of this method of species interaction estimation at large ecological scales. The predictive ability of this method with respect to microbiota community dynamics could be assessed through removal experiments within a controlled microbiota environment – for example the Schaedler flora[86].

Given that rare species appear to experience distinct (perhaps more stabilising) interaction environments from their more abundant contemporaries, the effect of rare species upon community stability could then be assessed. Specifically, using a model in which the proportional contribution of rare species to the community can be modified, with the interactions within and between the rarity groups parameterised to reflect the findings of Chapter 3, it is possible to perform linear stability analyses to compare the stability of the microbiota community in the presence and absence of rare species. This analysis could then be extended to compare the effects of B-cell presence and absence upon the interaction environment of the microbiota. Specifically, to consider whether the effect of IgA is to stabilise the microbiota community through the maintenance of rare species.

Taking this work a step further, it could be possible to use the same fine-scale sampling method to assess the interactions between different gastrointestinal communities. For example, the grid structure could be used to predict which microbes could be positively or negatively interacting with specific BCR. This information could be used to synthesise molecules whose association with either positive or negatively influenced microbes (assessed via Western blot) can then shed light on the role of IgA within the microbiota. Understanding the association between specific IgA molecules and microbes could then allow for the development of IgAs that select for specific beneficial microbes within the microbiota and act as a therapeutic treatment for individuals with microbiota dysbiosis.

Supplementary Information

SI Table 2.1 – The PCR barcode sequences, with their associated names, and DNA amplification ability. The barcode sequences are all followed by the UPA short sequence (CTAATACGACTCACTATAGGGC). The sensitivity of the UPA-associated barcode was assessed in combination with different reverse primers. The initial barcode groupings (grey) used a DNA dilution series where with a starting concentration of 1µl of 1302 ng/mL. As such, a sensitivity of 1 indicates amplification of 1µl of 1302 ng/mL, 2: 130.2 ng/mL, 3: 13.02 ng/mL and 4: 1.302 ng/mL. Subsequent barcode groups, used a starting DNA concentration of 1 µl of 8.0 ng/mL, and compared with barcodes from the initial groupings (Cγ25 barcodes: 6, 27, 175, Cβ7 barcodes: 47, 80, 30) to standardise across the pool 3 and pool 4 boundary. Pool 3 and 4 barcodes were used for further analysis.

Barcode	Barcode Sequence	Sensitivity		
		Cβ7	Cγ25	IgA
2	TTCTCCA	3	4	3
4	AACACCT	3	3	4
5	CTCTCAC	3	3	3
6	TGCGCAA	3	2	4
7	GACACAG	3	4	3
13	CACACTC	3	3	3
15	GTCTCTG	3	4	3
16	AGCGCTT	3	3	3
24	CCCTAAT	2	2	3
27	ACCTAGG	3	3	3
30	GCCTATA	4	3	4
31	ATCGATG	3	2	2
34	ATCAGCA	3	4	2
37	GTCAGAC	4	3	2
44	TTCAGGT	3	3	3
45	GACTGTC	4	3	3
47	CTCAGTG	2	3	3
52	GACGTCT	3	4	4
56	GCCATAT	3	3	3
66	GGATCCA	3	4	3
67	AAAGCCG	3	3	3
68	CCAACCT	3	2	3
69	TGATCAC	3	3	3
70	GAAGCAA	3	4	2
71	ACAACAG	2	3	1
75	ATACCGG	2	3	4
80	CAAGCTT	3	3	3
85	GGAGAAC	2	1	3
87	CCACAAG	2	3	3
91	CTATAGG	2	3	3
93	GCACATC	2	4	3
97	ATAGGCC	3	4	3
100	GCATGCT	3	3	3
104	GTAGGAT	2	3	4
106	CCATGGA	3	3	3
113	CTAATCC	2	3	2
130	AAGTCCA	3	3	3
133	GAGTCAC	2	3	3
138	ATGACGA	3	3	3
142	AGGCCTA	2	3	3
143	CAGTCTG	3	3	3
175	GAGAGTG	4	4	3
197	ACTTCAC	3	3	3
198	CTTGCAA	3	3	3
202	CGTACGA	3	2	2
8	ACCCCAT	3	2	NA
10	TACACGA	4	2	NA
12	ATCTCGT	3	2	NA

Barcode	Barcode Sequence	Sensitivity	
		Cβ7	Cγ25
17	TCCTACC	3	2
18	GTCGACA	3	2
19	AGCAACG	3	2
20	CACCACT	4	2
21	TTCGAAC	3	2
23	AACCAAG	3	3
25	TGCAAGC	3	3
26	GACCAGA	4	2
28	CTCGAGT	4	2
29	TACCATC	2	3
32	CGCAATT	2	2
36	TACTGCT	4	4
38	AGCCGAA	2	3
39	CACTGAG	3	3
40	TCCGGAT	2	4
42	AACTGGA	2	4
46	ACCGGTA	2	4
48	TGCCGTT	3	4
49	ACCATCC	2	3
50	CTCCTCA	2	4
51	TGCTTCG	2	4
53	ATCCTAC	2	4
54	CGCTTAA	3	4
55	TACGTAG	3	4
57	AGCTTGC	3	4
58	CACGTGA	3	4
59	TCCATGG	2	3
60	GTCCTGT	4	3
61	AACGTTC	4	2
62	CCCATTA	4	4
63	TTCCTTG	4	4
72	CTACCAT	3	2
73	TAAGCGC	3	4
74	GCAACGA	4	3
76	CGATCGT	4	4
77	TCAACTC	4	4
78	GTACCTA	2	2
79	AGATCTG	2	4
81	GTATACC	2	4
82	AGAGACA	4	4
84	TCACACT	2	2
90	ACACAGA	4	3
92	TGAGAGT	4	4
95	CGAGATG	3	4
98	CGAAGCA	3	4
99	TAACGCG	2	4
101	AGAAGAC	4	4
102	CAACGAA	2	2

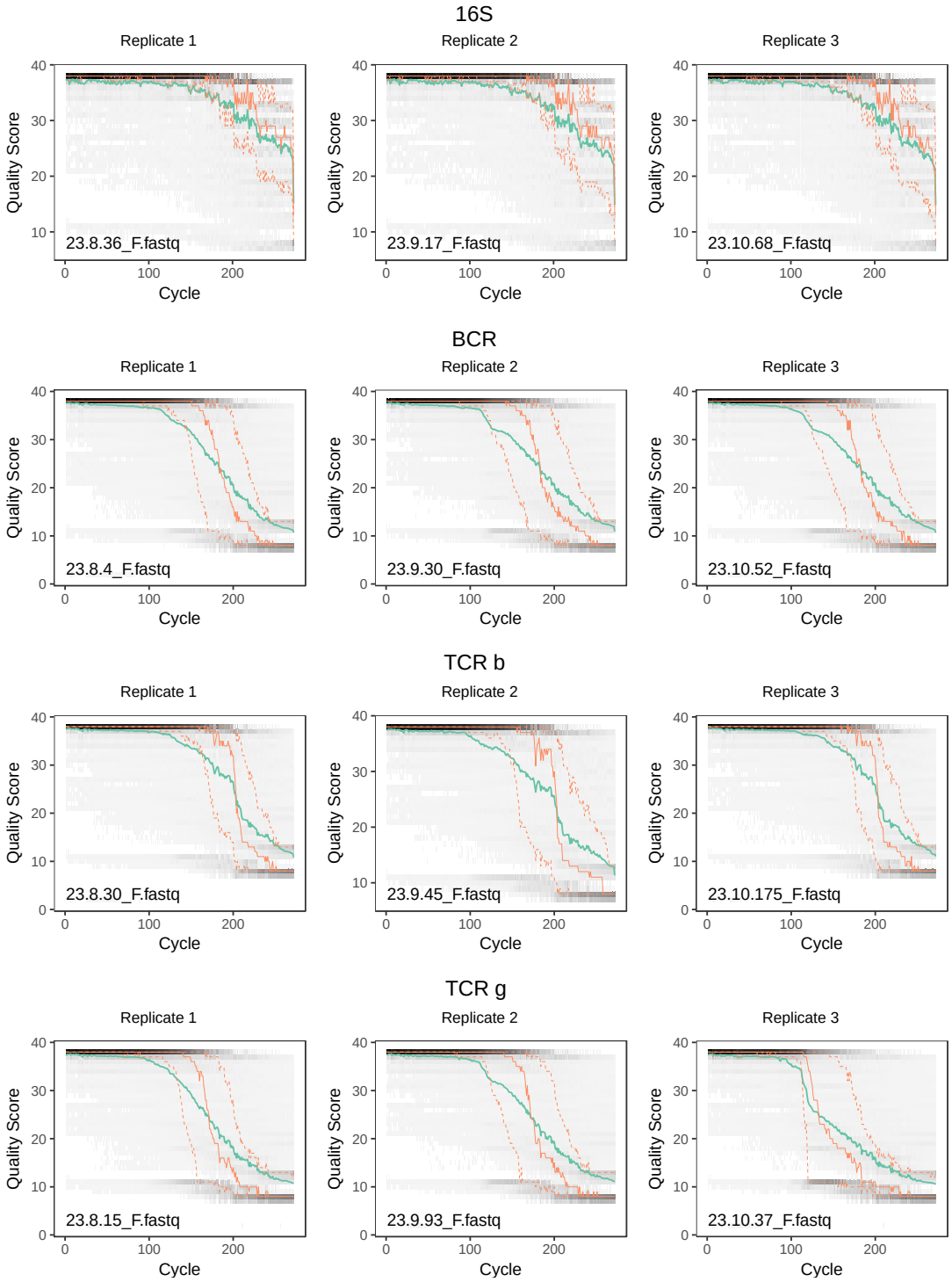
Barcode	Barcode Sequence	Sensitivity	
		Cβ7	Cγ25
103	TCATGAG	3	2
105	AAACGGC	2	2
108	GGAAGGT	4	3
109	ACATGTC	4	4
110	CTAGGTA	2	2
111	TGAAGTG	4	2
112	GAACGTT	4	4
114	TGACTCA	2	2
115	GAATTCG	4	2
116	ACAGTCT	2	2
117	CGACTAC	4	2
119	GCAGTAG	4	4
121	CAATTGC	4	4
122	TCAGTGA	4	4
123	GTAATGG	4	2
124	AGACTGT	4	3
125	CCAGTTC	4	4
127	GGACTTG	2	4
132	TTGACCT	2	4
134	ACGGCAA	3	3
135	CTGACAG	3	4
136	TGGCCAT	4	2
140	TAGTCGT	4	4
141	GTGACTC	2	2
144	TCGGCTT	2	4
145	AGGTACC	2	2
146	CAGGACA	2	2
147	TCGAACG	2	4
148	GTGCACT	4	2
149	AAGGAAC	3	3
151	TTGCAAG	4	3
152	GGGTAAT	4	2
153	ACGAAGC	4	4
154	CTGCAGA	3	2
155	TGGTAGG	2	4
156	GAGGAGT	4	2
157	ATGCATC	4	4
158	CGGTATA	3	4
159	TAGGATG	4	2
160	GCGAATT	4	4
162	TAGAGCA	4	4
164	ATGTGCT	4	4
165	CAGAGAC	4	4
166	TCGCGAA	4	4
167	GTGTGAG	4	4
170	TTGTGGA	2	2
172	AAGAGGT	2	2
173	CTGTGTC	2	2

SI Table 2.1 continued.

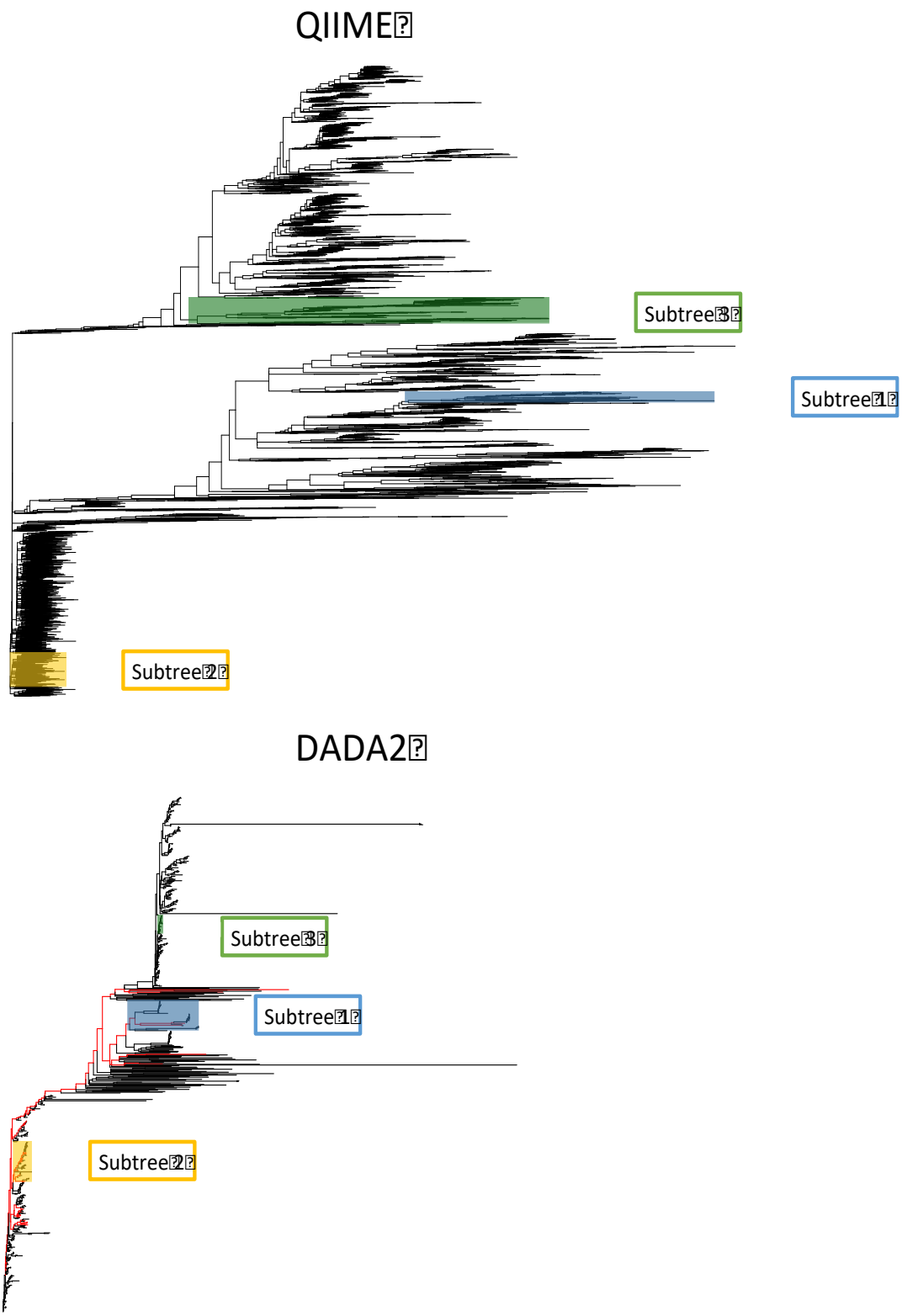
Barcode	Barcode Sequence	Sensitivity	
		Cβ7	Cγ25
176	ACGCGTT	4	2
177	TGGATCC	2	3
178	GAGCTCA	2	2
179	ACGTTCG	2	3
180	CTGGTCT	2	3
181	TAGCTAC	2	3
182	GCGTTAA	2	2
183	ATGGTAG	2	4
184	CGGATAT	2	4
185	TCGTTGC	2	3
186	GTGGTGA	2	4
187	AGGATGG	2	4
188	CAGCTGT	2	3
189	TTGGTTC	3	3
190	GGGATTA	3	2
191	AAGCTTG	3	4

Barcode	Barcode Sequence	Sensitivity	
		Cβ7	Cγ25
194	CCTTCCA	2	2
195	TTTGCCG	3	4
196	GGTACCT	3	4
199	TGTACAG	3	4
200	GATCCAT	2	4
201	ATTGCGC	3	4
203	TATCCGG	2	4
204	GCTTCGT	2	4
205	AGTACTC	2	2
207	TCTTCTG	3	3
208	GTTGCTT	3	4
210	TCTGACA	2	4
212	AGTCACT	4	4
215	GGTCAAG	4	4
218	TGTCAGA	4	2
221	CGTCATC	3	4

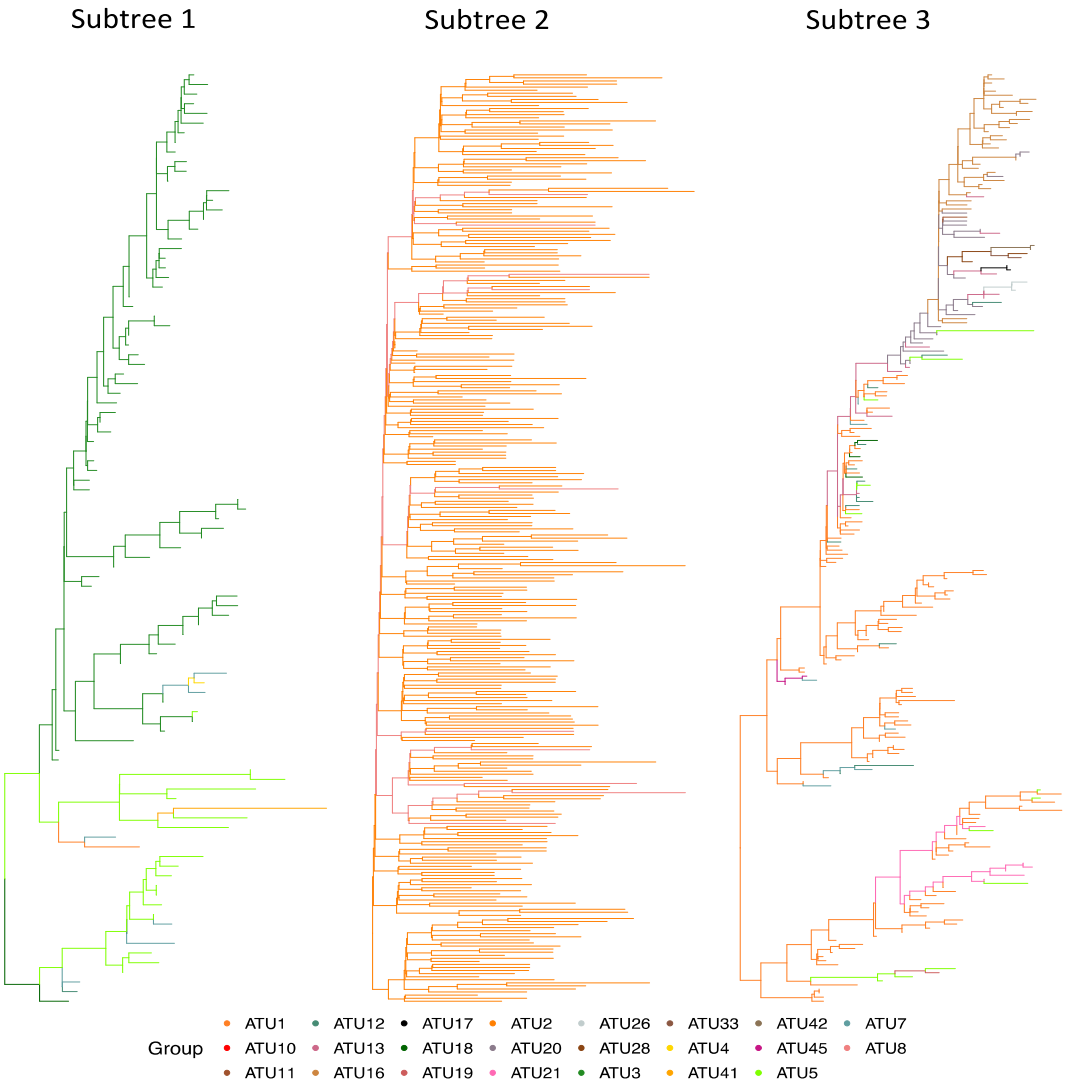
Barcode	Barcode Sequence	Sensitivity	
		Cβ7	Cγ25
223	GCTGATG	4	4
226	GCTAGCA	4	4
228	CGTTGCT	4	2
229	TCTAGAC	2	3
230	GTTCGAA	2	4
231	AGTTGAG	3	2
233	TTTCGGC	3	4
234	GGTTGGA	2	3
237	TGTTGTC	2	3
240	CTTCGTT	4	4
241	GATATCC	4	4
242	ACTCTCA	2	4
244	TGTGTCT	4	3
245	GCTCTAC	2	4



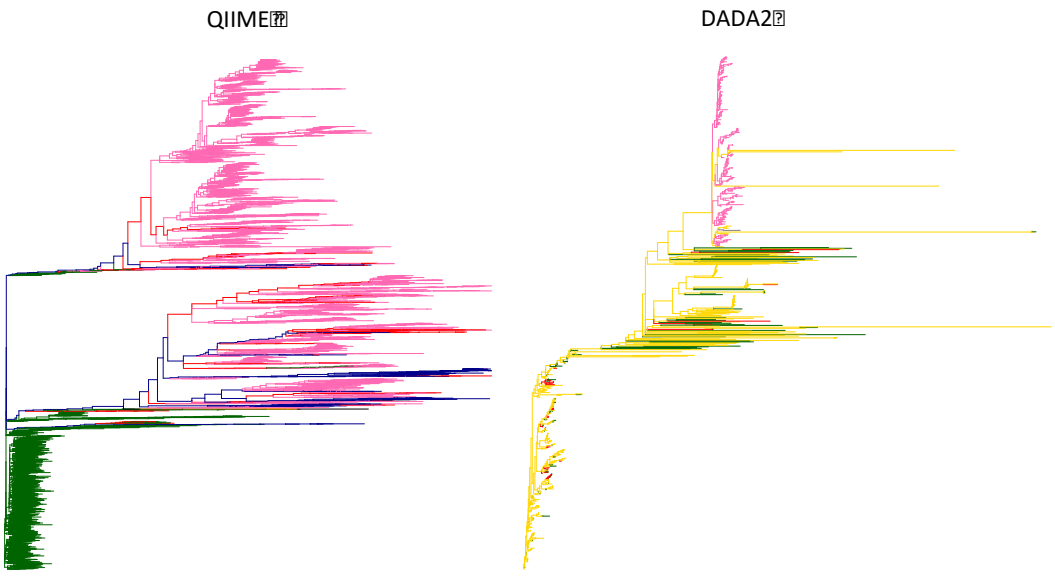
SI Figure 2.1 – Plots of quality scores for segment A1 forward reads, from every population type and replicate. Sequence type and replicate is indicated above the plots. Plots are in the standard DADA2 format. The grey-scale heatmap indicates the distribution of quality scores at each position along the sequence, with darker colours indicating a higher frequency. The green line indicates the mean of the quality scores are each position, orange the median and the dashed orange lines are the upper and lower quartiles (25th and 75th respectively). All sequences were demultiplexed to allow quality comparison between samples. 16S sequences were not dereplicated before quality was assessed, as there are different qualities associated with the different repeated reads of a sequence. The quality comparison for the immune datasets was performed on both demultiplexed and dereplicated, as necessitated by the in-house processing pipeline.



SI Figure 2.2 – Full approximate maximum likelihood tree for the OTUs (upper) and ISV (lower) datasets. The OTU tree is generated using the forward only, chimera removed QIIME dataset, whilst the ISV tree is produced from the merged, chimera removed DADA2 dataset. The OTU tree was formed using the most common sequence in each OTU. Comparison subtrees are highlighted in blue, (Subtree 1), yellow (Subtree 2) or green (Subtree 3). Branches that end in 'unassigned' sequences are red.



SI Figure 2.3 – Comparison subtrees extracted from the approximate maximum likelihood tree QIIME generated OTUs. The OTU tree is generated using the forward only, chimera removed QIIME dataset and was formed using the most common sequence in each OTU. The branches are coloured by the assigned taxonomic unit (ATU) to which each OTU was assigned using the Greengenes database. ATU1 (Clostridia; Clostridiales), ATU2/ATU3 (Clostridia; Clostridiales; Lachnospiraceae), ATU4 (Clostridia; Clostridiales; Ruminococcaceae; Ruminococcus), ATU5 (Clostridia; Clostridiales; Ruminococcaceae), ATU7 ('Unassigned'), ATU8 (Gammaproteobacteria), ATU10 (Clostridia; Clostridiales; Lachnospiraceae; Blautia), ATU11 (Flavobacteriia; Flavobacteriales; Flavobacteriaceae; Flavobacterium), ATU12 (Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus), ATU13 (Clostridia; Clostridiales; Ruminococcaceae; Anaerotruncus), ATU16 (Bacilli; Lactobacillales; Lactobacillaceae; Lactobacillus; reuteri), ATU17 (Clostridia; Clostridiales; Lachnospiraceae; Ruminococcus), ATU18 (Gammaproteobacteria; Alteromonadales; Alteromonadaceae), ATU19 (Gammaproteobacteria; Alteromonadales; Shewanellaceae; Shewanella), ATU20 (Mollicutes; RF39), ATU21 (Bacilli; Bacillales; Bacillaceae), ATU26 (Clostridia; Clostridiales; Peptostreptococcaceae), ATU28 (Clostridia; Clostridiales; Lachnospiraceae; Coprococcus), ATU33 (Erysipelotrichi; Erysipelotrichales; Erysipelotrichaceae), ATU41 (Clostridia; Clostridiales; Lachnospiraceae; Roseburia), ATU42 (Gammaproteobacteria; Oceanospirillales; Oceanospirillaceae), and ATU45 (Clostridia; Clostridiales; Clostridiaceae; Clostridium).



SI Figure 2.4 – Full approximate maximum likelihood trees using the QIIME and DADA2 processed sequences. The QIME OTU tree is generated using the forward only sequences, whilst the DADA2 ISV tree is produced from the merged sequences. Both sequence datasets have had the chimera sequences removed. The QIIME tree was formed using the most common sequence in each OTU. The branches are coloured by the phylum to which each terminal OTU/ISV was assigned using the Greengenes database. Phyla key: 'Unassigned' in red, Actinobacteria in orange, Bacteroidetes in yellow, Cyanobacteria in green, Elusimicrobia in blue, Euryarcgaeita in purple, Firmicutes in pink, Lentisphaerae in gold, Proteobacteria in dark pink, Tenericutes in dark blue and TM7 in dark purple.

SI Table 2.2 – Wilcoxon's test p-values for the association between barcode sensitivity and diversity estimates for each of the immune populations. All values are given to three significant figures.

Population	Species Richness	Shannon's Diversity	Simpson's Entropy
TCRβ	1.85e-03	9.52e-06	1.35e-05
TCRγ	8.14e-02	3.27e-02	0.0602
BCR	2.91e-02	1.16e-02	5.74e-03

SI Table 2.3 – Large-scale differences in the population structure of the microbiota and immune populations. The significance of the difference in proportional contribution of each of the abundance groups between populations according to a linear model of the form lm(Prop~Population) for each abundance group. Group 2 contains the most abundant clones (top 1%) and group 6 the smallest 50%. Populations compared are the IgA-producing B-cell receptors (including [B] and excluding [proB] non-functional nucleotide sequences), T-cell receptor β (Tb) and γ (Tg) and the microbiota datasets. The microbiota populations were included both containing and excluding contaminating sequences (upper case and lower case respectively) and processed with either QIIME (MQ) or DADA2 (MD). All values are given to three significant figures.

CH Group		P-value	R^2 Adjusted	Different Populations	Different Trim
2	$x \leq 1\%$	3.75e-15	0.972	B-MD/MQ/Md/Mq/proTb/Tb MD-MQ/Mq MQ-Md/Mq/proB/proTb/proTg/Tb/Tg Mq-Md/proB/proTb/proTg/Tb/Tg	MQ-Mq
3	$1 < x \leq 10\%$	3.24e-18	0.986	B-MD/MQ/Md/Mq/proTb/proTg/Tb/Tg MD-MQ/proB/proTb/proTg/Tb/Tg MQ-Md/Mq/proB/proTb/proTg/Tb/Tg Md-proB/proTb/proTg/Tb/Tg Mq-proB/proTb/proTg/Tb/Tg proB-proTb/proTg/Tb/Tg proTb-proTg/Tg Tb-proTg/Tg	MQ-Mq

SI Table 2.3 continued.

CH Group		P-value	R^2 Adjusted	Different Populations	Different Trim
4	$10 < x \leq 25\%$	1.90e-12	0.948	B-MQ/Mq MD-MQ/Mq/Tg MQ-Md/proB/proTb/proTg/Tb/Tg Md-Mq/proB/proTb/proTg/Tb/Tg proB-proTg/Tg	
5	$25 < x \leq 50\%$	2.53e-21	0.993	B-MD/MQ/Md/Mq/ proTb/proTg/Tb/Tg MD-Mq/proB/proTb/ProTg/Tb/Tg MQ-proB/proTb/proTg/Tb/Tg Md-Mq/proB/proTb/proTg/Tb/Tg Mq-proB/proTb/proTg/Tb/Tg proB-proTb/proTb/Tb/Tg proTb-proTg/Tg/Tb Tg-Tb	proTb-Tb
6	$x > 50\%$	9.69e-25	0.997	B-MD/MQ/Md/Mq/proB/proTb/proTg/Tb/Tg MD-proB/proTb/proTg/Tb/Tg MQ-proB/proTb/proTg/Tb/Tg Md-proB/proTb/proTg/Tb/Tg Mq-proB/proTb/proTg/Tb/Tg ProB-proTb/proTg/Tb/Tg proTb-proTg/Tg/Tb Tg-Tb	proB-B proTb-Tb

SI Table 2.4 – The variation in the proportional contribution of different population structure groups between segments within the immune populations. Variation was assessed using a linear model of the form lm(Prop~Segment) for each abundance group. see SI Figure 6 for full group identities. The segments driving the fine-scale differences in population structure within the immune and microbiota populations and indicated for each abundance group. All values are given to three significant figures.

Population	Group	P-value	R^2 Adjusted	Different Segments
BCR	2	0.243	0.0856	*NA*
	3	5.87e-12	0.803	A1-A2/A3/A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 B3-A2/A3/A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 C1-A4/C2/D1/D2/E3 D4-A2/A3/A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 E1-A4/B2/C2/C3/C4/D1/D2/D3/E2/E3/E4
	4	4.25e-06	0.583	A1-A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 B3-A4/C2/C4/D1/E2/E3 D4-A4/B2/C2/C4/D1/D2/D3/E2/E3/E4
	5	3.25e-06	0.589	A1-A4/B1/B2/B4/C2/C3C4/D1/D2/D3/E2/E3/E4 B3-A4/C2/D1/E3 D4-A4/B2/C2/C4/D1/D2/D3/E2/E3/E4
	6	6.09e-07	0.628	A1-A2/A3/A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 B3-A4/B2/C2/C4/D1/D3/E2/E3/E4 D4-A4/B2/C2/C3/C4/D1/D2/D3/E2/E3/E4
TCRβ	2	3.82e-07	0.637	A1-A2/A3/A4/B1/B2/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B3-A4/C3/C4/D2/E4
	3	3.35e-10	0.755	B1/B3-A1/A4/C2/C3/C4/D2/D3/E4 C1-A1/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 D2-C3/D1/D4/E1/E2 D4-A1/A2/A3/A4/B2/B4/C2/C4/D3/E3/E4 E1-A1/A4/B2/B4/C2/C4/D3/C3/E3/E4 E2-A1/A4/C2/C4/D3/C3/E4
	4	3.08e-20	0.927	A1-A2/A3/A4/B1/B2/B4/C2/C3/C4/D1/D2/D3/E3/E4 A2/A3/A4/B2/B4-B1/B3/C1/D1/D4/E1/E2 B1-B2/B3/B4/C1/C2/C3/C4/D2/D3/D4/E1/E3/E4 B3/C1-B4/C2/C3/C4/D1/D2/D3/E3/E4 C2/C3/C4/D2/D3-D4/E1/E2 D1-C2/C3/C4/D2/D3/D4/E1/E3/E4 D3/E1/E2-E3/E4

SI Table 2.4 continued.

Population	Group	P-value	R^2 Adjusted	Different Segments
TCRβ	5	4.61e-09	0.717	A4-A1/A2/B1/B3/C1/D1/D4/E1/E2 B1-C2/C3/C4 C3/C4-A1/A2/D1/D4/E1/E2 D2-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3 B3/C1/E1/D4-C2/C3/C4/D3/E4
	6	5.018e-08	0.677	A4-A1/B1/B3/C1/D4/E1/E2 C2/C4/D3-B3/C1/D4/E1 C3-A1/B1/C1/B3/D4/E1/E2 D2-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3 E4-D4
TCRγ	2	0.291	0.0659	*NA*
	3	1.522e-06	0.607	B1-A4/C3/C4/D2 B3/D4-A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 C1-A4/B4/C2/C3/C4/D2/D3/E3
	4	4.46e-08	0.679	A1-A2/A3/A4/B2/B4/C2/C3/C4/D1/D2/D3/E3/E4 A2-B1/B3/C1/E1 B3-A2/A4/B2/B4/C2/C3/C4/D1/D2/D3/E3/E4 C1/E1-A2/A4/B2/C2/C3/C4/D1/D2/D3/E3/E4
	5	2.62e-03	0.373	*NA*
	6	7.10e-06	0.570	A4-A3/B1/B3/D4 B1-C2/C3/C4/D2/D3 B3-C2/C3/C4/D2/D3/D4/E2 C3-A3 D4-C4/D2

SI Table 2.5 – The association between sample coverage and proportional contribution of difference population abundance groups in the microbiota and immune populations. The correlation coefficients were taken from a linear model of the form lm(Prop~SC) for each abundance group. All values are given to three significant figures.

Group	BCR		TCRβ		TCRγ		QIIME		DADA2	
	P-value	R^2 Adj	P-value	R^2 Adj	P-value	R^2 Adj	P-value	R^2 Adj	P-value	R^2 Adj
2	2.43e-01	0.0856	3.82e-07	0.637	2.91e-01	0.0659	2.38e-05	0.537	5.84e-02	0.205
3	5.87e-12	0.803	3.35e-10	0.755	1.52e-06	0.607	2.65e-02	0.256	5.41e-04	0.437
4	4.25e-06	0.583	3.08e-20	0.927	4.46e-08	0.679	4.83e-07	0.633	9.24e-02	0.171
5	3.25e-06	0.589	4.61e-09	0.717	2.62e-03	0.373	3.73e-03	0.358	1.19e-02	0.301
6	6.09e-07	0.628	5.02e-08	0.677	7.10e-06	0.570	4.01e-01	0.0262	2.55e-01	0.0803

SI Table 2.6 – The variation in the proportional contribution of different population structure groups between segments within the microbiota populatinos, processed using either QIIME or DADA2. Variation was assessed using a linear model of the form lm(Prop~Segment) for each abundance group. see SI Figure 6 for full group identities. The segments driving the fine-scale differences in population structure within the immune and microbiota populations and indicated for each abundance group. All values are given to three significant figures.

Population	Group	P-value	R^2 Adjusted	Different Segments
DADA2	2	0.0584	0.205	*NA*
	3	5.41e-04	0.437	A3-A4/B1/B3/C1/C4/D1/E4 E1-C4/D1/E1
	4	0.0945	0.171	*NA*
	5	1.19e-02	0.301	D1-A3
	6	0.255	0.0803	*NA*
QIIME	2	2.38e-05	0.537	A1-C2/C4/D1/E4 D1-A3/C3/E1 D2-A1/A3/B3/B4/C3/D4/E1 E1-E4
	3	2.65e-02	0.256	C1-A3/B2
	4	4.83e-07	0.633	A1/A3/B4/C3-C4/D1/D2/E4 B3-C1/C4/D1/D2/E4 D1-B2/D3/D4 D4-D2 E1-C1/C4/D1/E4
	5	3.73e-03	0.358	D1-A1/A3
	6	0.401	0.0262	*NA*

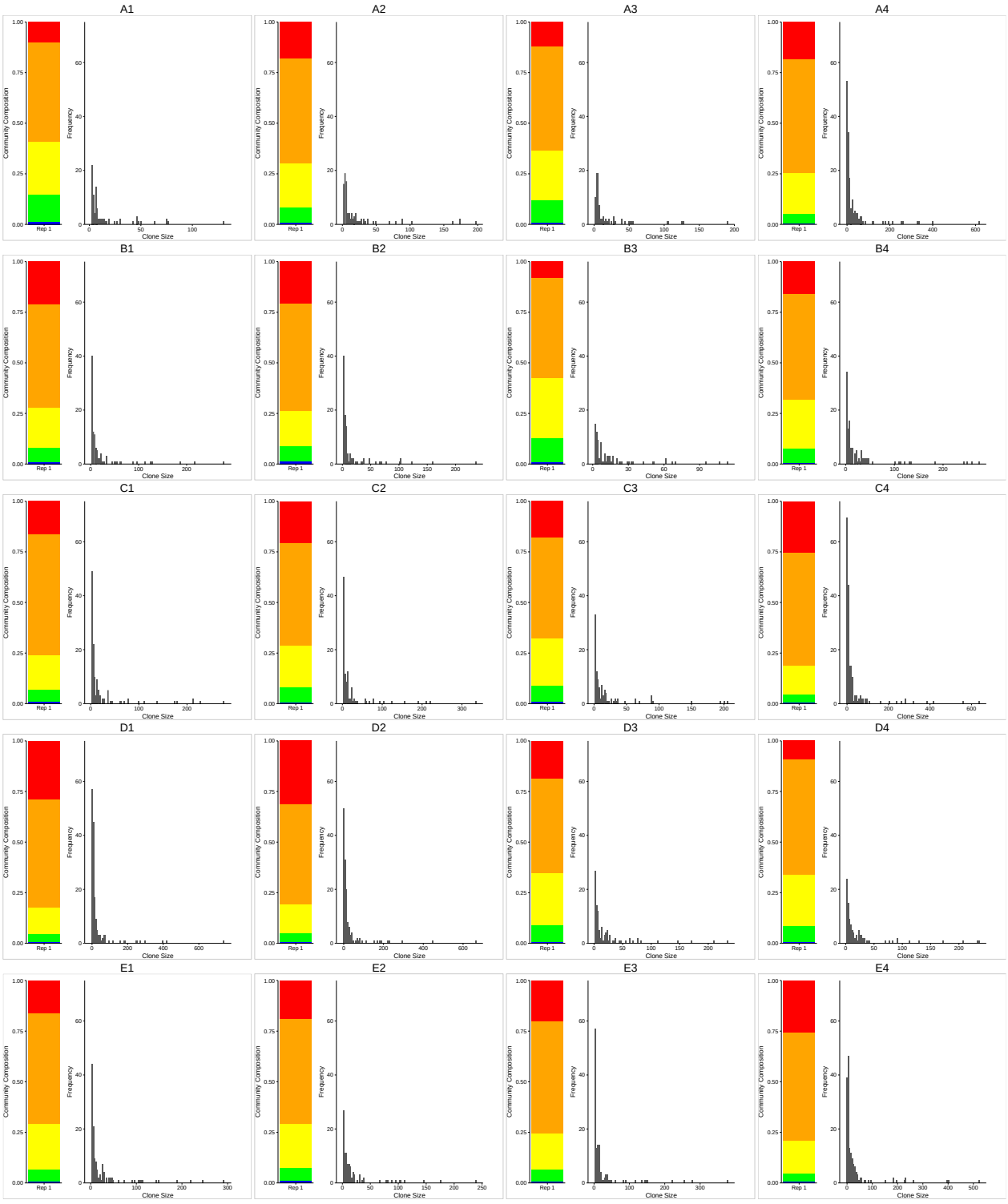
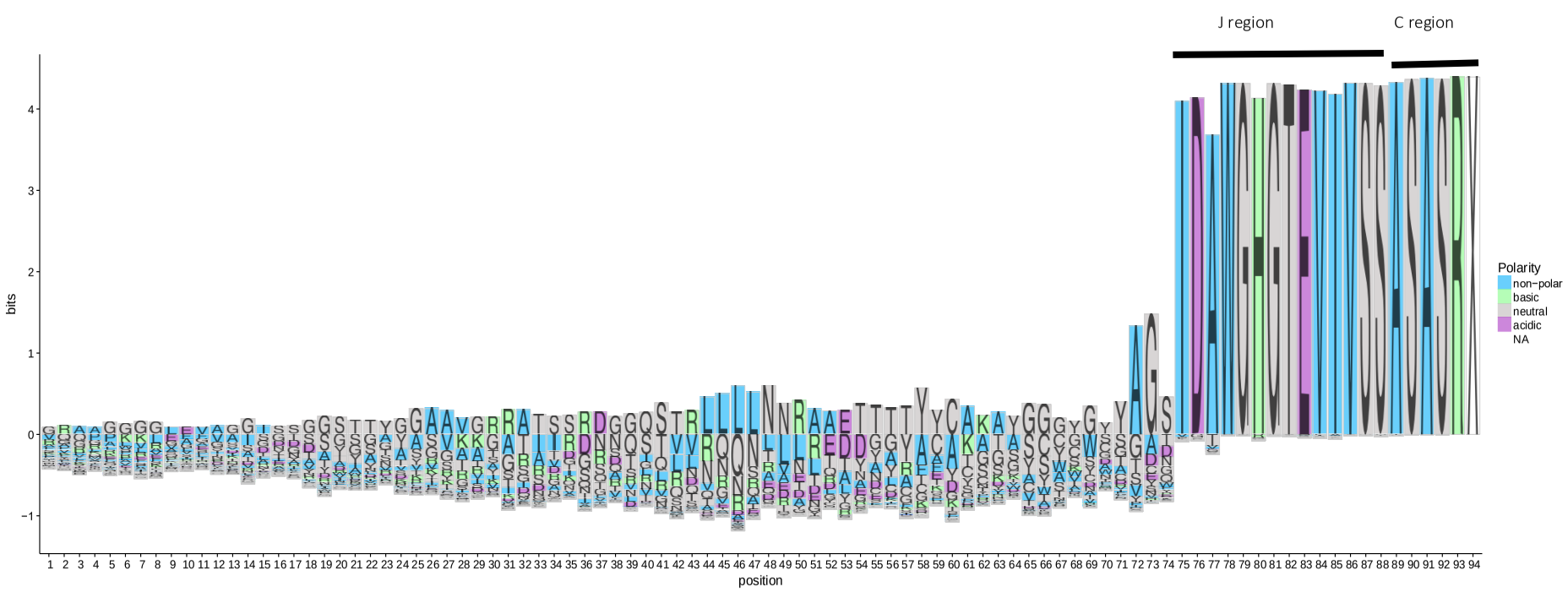


Figure 2.5 – Microbiota population structure of a L6 6-week old chicken right caeca, separated into 4 mm^2 segments. The sequence data has been processed using the QIIME pipeline, and contaminant sequences removed prior to analysis. Clonal homeostasis plots show the proportion of the population made up by species ranked by abundance, whilst the histograms show the distribution of species sizes (excluding singletons).

SI Table 2.7 – Assigned families contributing to less than 1% of the total microbiota composition, processed using either QIIME or DADA2, represented in Figure 21 as the 'other' category.

Processing Method	Families contributing < 1% of the sequences
QIIME	f__Bacteroidaceae, f__[Barnesiellaceae], f__Bifidobacteriaceae, f__Cardiobacteriaceae, f__Christensenellaceae, f__Dehalobacteriaceae, f__Desulfovibrionaceae, f__Enterobacteriaceae, f__EtOH8, f__Helicobacteraceae, f__Lactobacillaceae, f__Micrococcaceae, f__[Mogibacteriaceae], f__Nitrosomonadaceae, f__Peptococcaceae, f__Peptostreptococcaceae, f__Porphyromonadaceae, f__Rikenellaceae, f__Shewanellaceae, f__Streptococcaceae, f__Succinivibrionaceae, f__Turicibacteraceae
DADA2	f__Acetobacteraceae, f__Bacillaceae, f__Christensenellaceae, f__Clostridiaceae, f__Erysipelotrichaceae, f__Halomonadaceae, f__Lactobacillaceae, f__[Mogibacteriaceae], f__Peptostreptococcaceae, f__Shewanellaceae, f__Sphingomonadaceae, f__Streptococcaceae



SI Figure 2.6 – Amino acid (AA) usage across the large-scale immunoglobulin A producing B-cell population. Sequences displayed are from the reverse reads, clustered into clones and the consensus sequences extracted. AA' s are displayed using the standard single letter AA code, and coloured by the AA' s polarity. The y-axis gives the amount of the mixture that each AA contributes, as measured by Shannon' s Information metric (see Chapter 2 B.8) and given in ' bits' . The reverse C-primer is at the far right of the plot and would start at position 95.

SI Table 2.8 – The significance of variation in diversity indices (species richness, Shannon's diversity and Simpson's entropy) between segments within the immune populations and the microbiota populations, processed using either DADA2 or QIIME. Variation was assessed using a linear model of the form lm(Diversity~Segment) for each index. The segments driving the fine-scale differences in population composition within the immune and microbiota populations and indicated for each diversity metric. All values are given to three significant figures.

Estimate	Population	P-value	R^2 Adjusted	Different Segments
Species Richness	BCR	2.30e-14	0.853	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1 /D2/D3/E1/E2/E3/E4 B3-A4/B2/C2/C3/C4/D1/D2/D3/E2/E3/E4 C1-A4/D1/D2/D3/D4/E2/E3/E4 D4-A2/A3/A4/B1/B2/B3/B4/C2/C3/C4/D1/D2/D3/E2/E3/E4 E1-A4/B2/C2/C3/C4/D1/D2/D3/E2/E3/E4
	TCRβ	2.34e-22	0.943	A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 B2-A1/A3/B3/D4/E1/E2 B4-B3/D4/E1 C1-A4/B2/B4/C4/E3/E4 C2/C3/C4/D2-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 D3-A1/A2/A3/B1/B3/B4/C1/C4/D1/D2/D4/E1/E2/E3/E4 E3/E4-A1/A2/A3/B3/B1/D1/E1/E2
	TCRγ	6.47e-17	0.892	A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2 B2-B1/B3/D4 B3-A1 B4-B1/B3/C3/D2/D4/E2 C2/C3-A1/A2/A3/B1/B2/B3/C1/D1/D4/E1/E2 C4-A1/A2/A3/B1/B2/B3/B4/D1/D4/E1/E2/E4 D2/D3-A1/A2/A3/B1/B2/B3/C1/D1/D4/E1/E2 E3/E4-A2/A3/B1/B3/C1/D1/D4/E1/E2
	DADA2	1.029e-01	0.164	A3-A4/C1/C4/D1/D2/E3/E4 E4-A1/B3/D4 C4-B3/D4
	QIIME	9.450e-06	0.562	A1-A3
Shannon's Diversity	BCR	5.68e-14	0.846	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2 /D3/E1/E2/E3/E4 A4-C1 B3-A4/B2/C3/C4/D1/D2/D3/E2/E3/E4 D4-A2/A3/A4/B1/B2/B4/C1/C2/C3/C4/D1/D2/D3/E1/E2/E3/E4 E1-C4/D1/D2/D3/E1/E3/E4 E2-A2/A3/A4/C1
	TCRβ	1.49e-30	0.978	A2-A1/B3/E1/E2 A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 B2/B4-A1/B3/C1/D1/D4/E1/E2 C2-A1/A2/A3/B1/B2/B3/B4/C1/C4/D1/D2/D4/E1/E2 C3/C4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 D2-A1/A2/A3/B1/B2/B4/C1/D1/D3/D4/E1/E2/E3/E4 D3-A1/A2/A3/B1/B3/C1/C3/C4/D1/D4/E1/E2/E3/E4 E3/E4-A1/A3/B1/B3/C1/D1/E1/E2
	TCRγ	1.35e-13	0.839	A4-A1/A2/A3/B1/B2/B3/C1/D1/D4/E1/E2 B4-A1/D4 C1-B1/B3/B4 C2-A1/A2/A3/B1/B3/C1/D1/D4/E2 C3-A1/A2/A3/B1/B3/C1/D1/D4/E1/E2 C4-A1/A2/A3/B1/B3/B2/B4/D1/D4/E1/E2 D2/D3-A1/A4/B1/B3/C1/D1/D4/E1/E2 E3-A1/A2/A3/B1/B3/C1/D4/E2 E4-A1/B1/B2/B3/C1/D4/E2
	DADA2	2.59e-04	0.463	A1-B1/D1/D2/D4/E3/E4
	QIIME	4.88e-06	0.579	A3-A1/A4/B3/B4/C4/D3/E1/E4 B4-C2 C4/E1-B1/B2/C2

SI Table 2.8 continued.

Estimate	Population	P-value	R^2 Adjusted	Different Segments
Simpson's Entropy	BCR	9.82e-10	0.740	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2 /D3/E1/E2/E3/E4 A2-C4 D4-A4/B1/B2/B4/C3/C4/D1/D2/D3/E3/E4 E2-A2/A3/B2/B3/B4/C1/C2/D4/E1
	TCRβ	1.24e-22	0.945	A2-A1/B3/E1/E2 A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 B2-A1/B3/C1/D1/D4/E1/E2 B3-B4/D3/E3/E4 B4-A1/C1/D1/D4/E1/E2 C2-A1/A2/A3/B1/B2/B3/B4/C1/C4/D1/D2/D4/E1/E2 C3/C4/D2- A1/A2/A3/B1/B2/B3/B4/C1/D1/D3/D4/E1/E2/E3/E4 D3-A1/A2/A3/B1/B3/C1/D1/D4/E1/E2 A1/A3/B1/C1/D1/D4/E1/E2-E3/E4
	TCRγ	1.25e-07	0.660	A1-A4/C2/C3/D2/D3/E3 A4-A2/A3/B1/B3/C1/D1/D4/E2 C3-B1/B2/C1/D4/E2 C4-A1/A2/A3/B1/B2/C1/D4/E2 D2-A3/B1/B2/C1/D4/E2 D3-B1/B2/C1 D4-E3
	DADA2	4.85e-07	0.632	A1-A3/A4/B1/B2/D1/D2/D4/E3/E4 B3-A3/B1/D1/D2/D4/E3 D3/E2-A3/B1/D2/D4/E3 E3-A2/C3
	QIIME	1.09e-12	0.820	A1-B1/D1 A3-A1/A2/A4/B3/B4/C1/C3/C4/D3/E2 B2-A1/A4/C1/C3/D3/E2 C2-A1/A2/A4/C1/C3/D3/E2 B3-A2/A4/B1/B2/C1/C2/C3/D1/D2/D4/E3/E4 B4/C4-B1/B2/C2/D1/D2/E3 D2-A1/A2/A4/C1/C3/D3/E2 D3-B1/D1 E1-A2/A3/A4/B1/B2/C1/C2/C3/D1/D2/D4/E2/E3/E4 E2-E1 E3-A1/B3/B4/C4/E1

SI Table 2.9 – The variation in the proportional contribution of different families between segments within the microbiota population, processed using either DADA2 or QIIME. The variation between segments was assessed using a linear model of the form lm(Prop~Location), and the segments which drove the difference identified using Tukey's HSD test. All values are given to three significant figures.

Processing Method	Phylum	Family	P-value	R^2 Adjusted	Different Segments
DADA2	Firmicutes	Lachnospiraceae	5.85e-05	0.669	B1-B4/C1/C4/D1/E1/E3/E4 B3-C1/C4/E1/E4
		Ruminococcaceae	4.85e-04	0.621	A3-A4/B1/2B/C1/D2/E3 B2-B3/B4/E1
	Proteobacteria	Mitochondria	4.64e-02	0.472	*NA*
QIIME	Actinobacteria	Coriobacteriaceae	3.34e-05	0.696	A3-A4/C1/D1/D4/E2/E3/E4 B2-A4/C1/C4/D1/D4/E2/E3/E4 C2-E3
	Firmicutes	Bacillaceae	5.26e-06	0.714	B3-A3/B2/C2/C3/D3 B2-C4 C2-A1/A2/A4/B1/B3/B4/C1/C4/D1/D4/E1/E2/E3/E4
		Clostridiaceae	6.51e-06	0.726	A2-A1/A3/C2/C4/D1/D2/E3/E4 C3-C4/D2/E3/E4 E1-C4/D2/E3/E4 E2-D2/E4
		Erysipelotrichaceae	1.45e-08	0.847	A4-A3/B1/B2/C1/C2/C4/D1/D2/E3/E4 B3-B1/B2/C1/C4/D1/D2/E4 B4-A2/A3/B1/B2/C1/C2/C4/D1/D2/E3/E4 D1-D4 D3-A2/A3/B1/B2/C1/C2/D1/D2/E3/E4 E1-A2/A3/B1/B2/C1/C2/D1D2/E2/E3/E4 E4-C3/D4
		Lachnospiraceae	2.35e-03	0.578	C1/C4-C2/C3/D3
		Ruminococcaceae	9.61e-10	0.824	A3-A2/A4/B1/B2/B3/C1/C4/D1/D2/D4/E2/E3/E4 B1-A1/B4/C2/C3/D3/E1 D1-A1/A2/B3/B4/C2/C3/D3/E1 D2-A1/B4/C2/C3/D3/E1 E1-B2/D2/E4 E3-B4/C2/C3/E1 E4-B4/C2
		Veillonellaceae	4.38e-12	0.868	A3-A1/A2/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2 /D3/D4/E1/E2/E3/E4 B4-A4/B1/B2/D1/D2/D3/E2 D1-A2/C1/C2/C3/D4/E1 C2-B2/D2/E2 D4-A4/B1/D2/D3/E2
	Proteobacteria	Halomonadaceae	1.43e-02	0.527	*NA*

SI Table 2.10 – The species which significantly change the correlation coefficient of the association between dissimilarity and distance, within the microbiota population process using either DADA2 or QIIME.

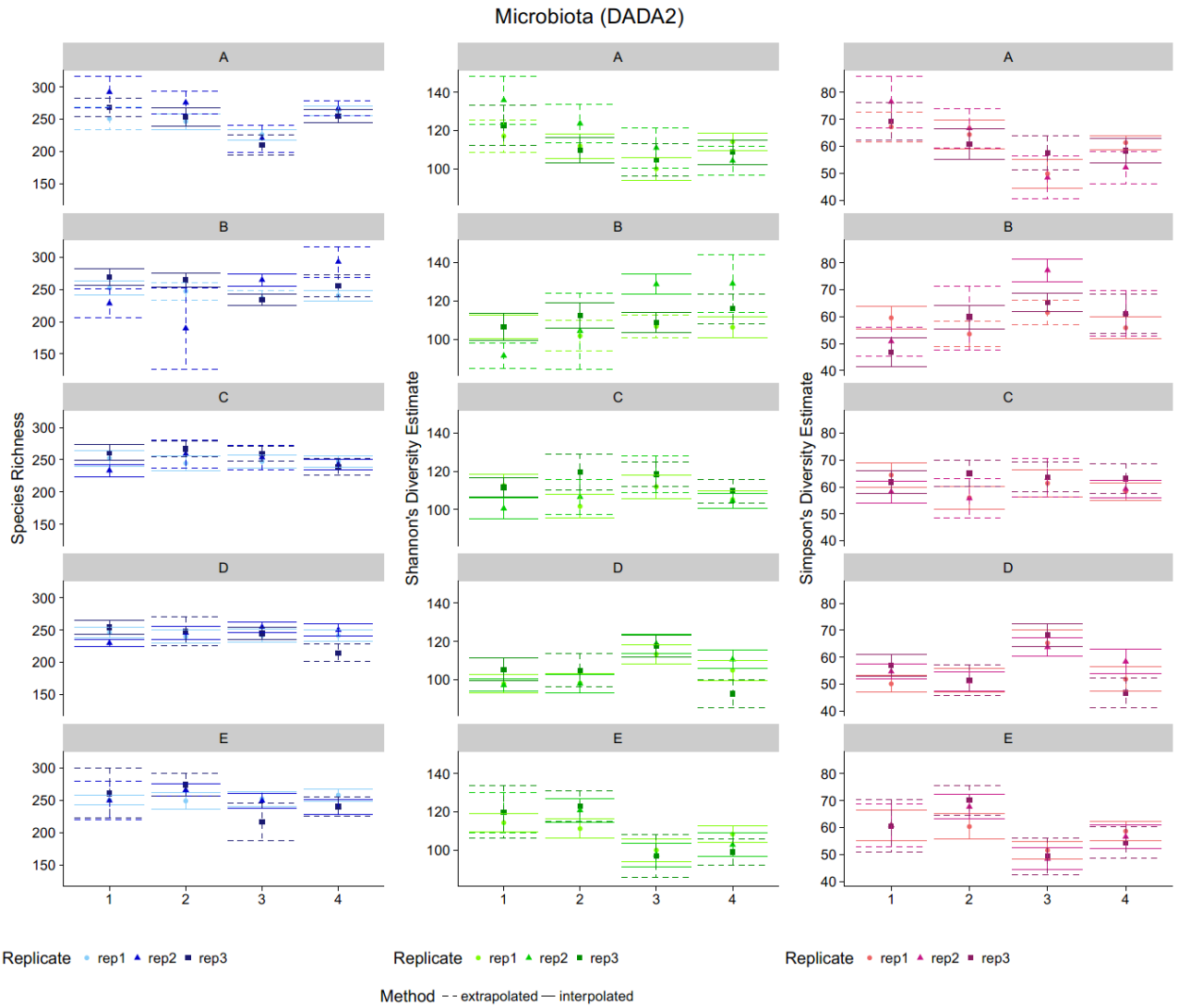
DADA2		QIIME	
Correlation increased by absence	Correlation decreased by absence	Correlation increased by absence	Correlation decreased by absence
ISV23	ISV3 ISV20	OTU264 OTU583	OTU18
ISV38	ISV5 ISV53	OTU319 OTU1319	OTU21
ISV48	ISV6 ISV247	OTU409	OTU212
ISV98	ISV7 ISV327	OTU472	OTU288
ISV153	ISV8 ISV497	OTU477	OTU343
ISV292	ISV10	OTU483	OTU398
ISV338	ISV12	OTU494	OTU414
ISV438	ISV15	OTU542	OTU486
ISV451	ISV17	OTU543	



SI Figure 2.7 – The abundance distributions of selected microbial species that significantly affect the correlation between segment dissimilarity and distance. Data presented was processed using the DADA2 pipeline.

SI Table 3.1 – The significance of sampling factors and diversity indices upon the proportional contribution of different rarity groups, as defined by either local population abundance or geographic spread, to each microbiota community. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'SC' single count, 'UP' global, 'AP' widespread, 'RP' localised and 'SP' single instance. For full explanation of abbreviations see Table 3.2. All values are given to three significant figures. Significant associations are highlighted in grey.

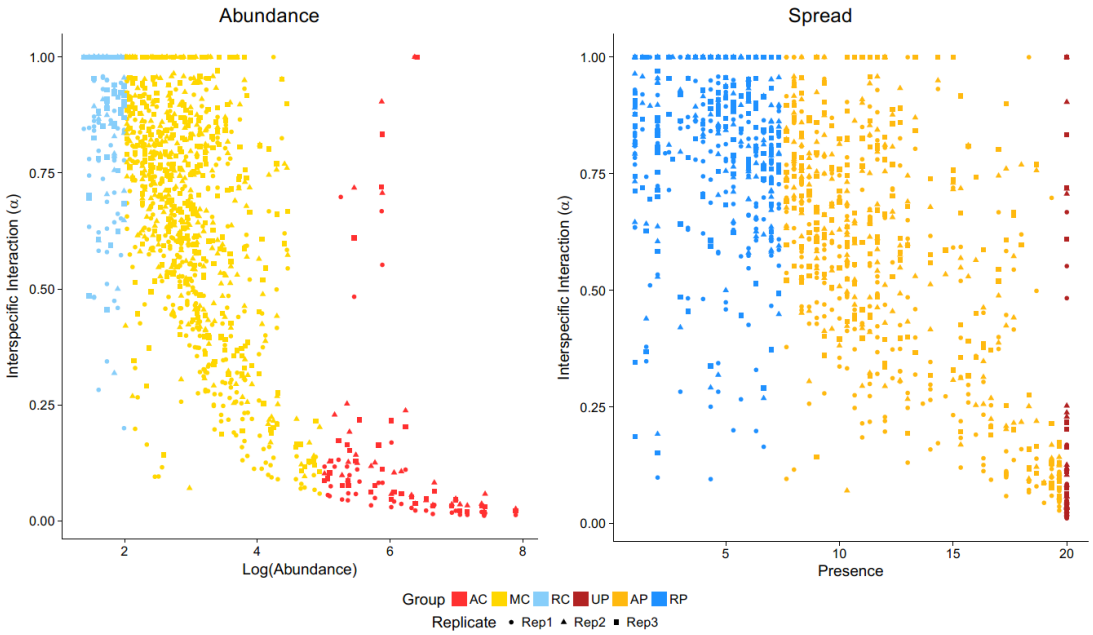
Explanatory Variable	Coefficient	Local Abundance				Geographic Spread			
		AC	MC	RC	SC	UP	AP	RP	SP
Sequence Depth	P-value	0.427	0.138	0.816	0.817	0.324	2.90e-02	0.503	0.0508
	R^2 Adjusted	-0.00613	0.0209	-0.0163	-0.0166	-0.000162	0.0637	-0.00950	0.0489
Sample Coverage	P-value	0.329	0.0562	0.791	0.684	0.138	3.25e-02	0.458	0.0525
	R^2 Adjusted	-0.000516	0.0452	-0.0160	-0.0146	0.0209	0.0605	-0.00767	0.0480
Species Richness	P-value	3.38e-02	2.56e-03	0.480	0.692	1.06e-02	1.02e-03	0.693	0.168
	R^2 Adjusted	0.0594	0.132	-0.00845	-0.0147	0.0921	0.157	-0.0147	0.0162
Shannon's Diversity	P-value	0.269	2.28e-03	0.697	2.39e-02	0.0529	3.54e-02	0.883	0.398
	R^2 Adjusted	0.00416	0.135	-0.0146	0.0704	0.0581	0.0469	-0.0172	-0.00474
Simpson's Entropy	P-value	0.723	0.0947	0.0883	1.30e-02	0.220	0.846	0.943	4.89e-02
	R^2 Adjusted	-0.0150	0.0310	0.0329	0.0877	0.00899	-0.0166	-0.0175	0.0500



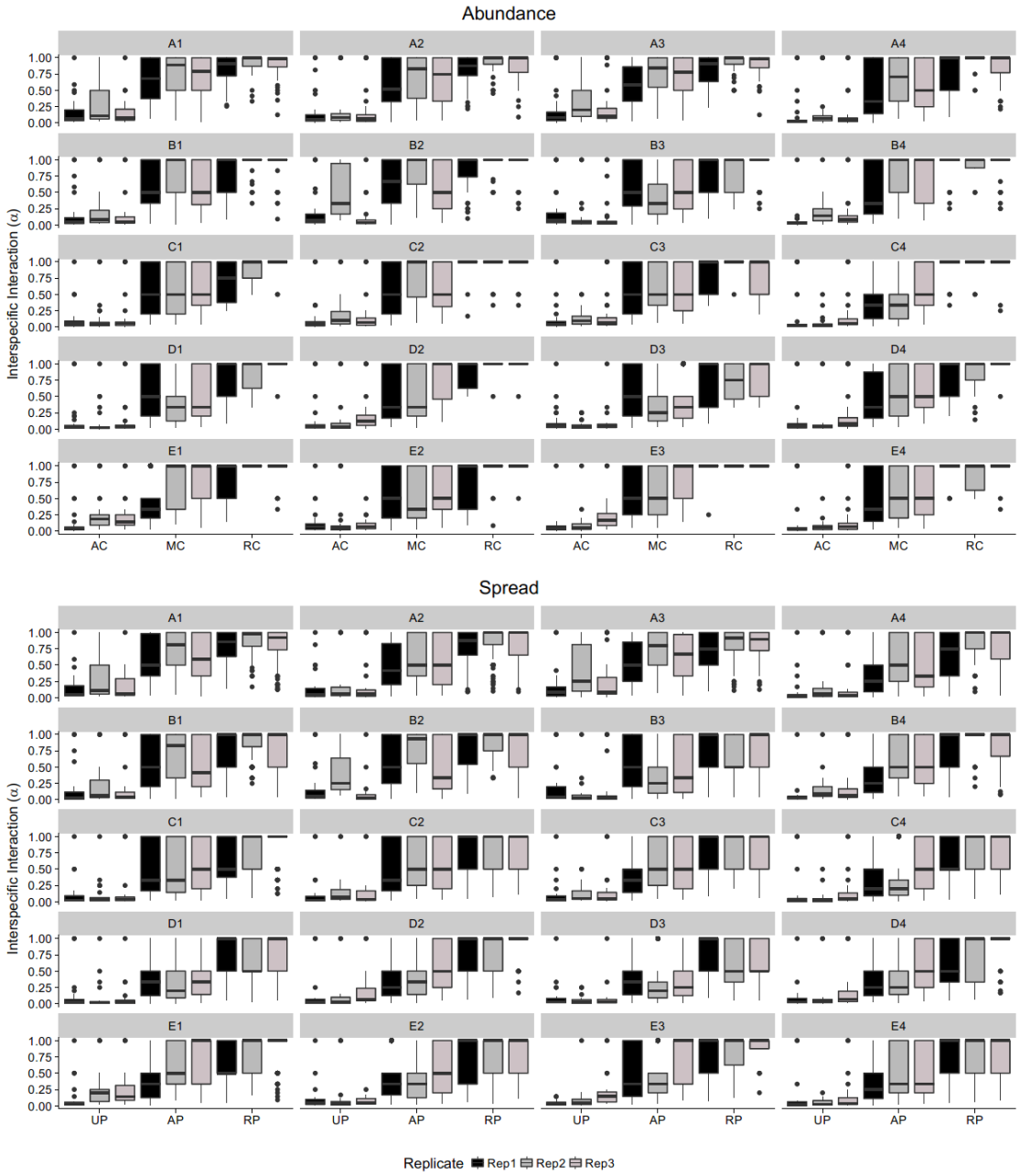
SI Figure 3.1 – Diversity estimates for each segment within the DADA2 processed microbial population. Species richness estimates are given in blue (left), the Shannon's diversity in green (middle) and Simpson's entropy in red (right). All estimates are taken at the same (median of the population) sampling depth, with the 95% confidence intervals of the estimate indicated as error bars. Dashed error bars indicate the sample required extrapolation to reach the median sampling depth, solid lines indicate the median sampling depth was either observed or interpolated. The results for each segment are indexed by their column (letter facets) and row (x-axis).

SI Table 3.2 – The significance of the association between the total community size, as assessed by qPCR, and the proportional contribution of each rarity group, defined by either local population abundance or geographic spread, to the microbiota community. All values are given to three significant figures. Significant associations are highlighted in grey.

Rarity Basis	Group	P-value	R^2 Adjusted
Local Abundance	AC	5.82e-04	0.172
	MC	0.279	0.00325
	RC	1.93e-04	0.201
	SC	2.92e-08	0.404
Geographic Spread	UP	6.08e-03	0.108
	AP	3.73e-04	0.184
	RP	0.867	-0.0168
	SP	0.399	-0.00472



SI Figure 3.2 – The association between the total abundance, and spread, of a species and its intraspecific interaction strength. The rarity groups, as defined by local population abundance or geographic spread, are indicated. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 2. All three technical replicates are displayed.



SI Figure 3.3 – Intraspecific interactions grouped by rarity, as defined by local population abundance or geographic spread, across space. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 2. Different technical replicates are indicated. Values outliers are indicated as dots.

SI Table 3.3 – The significance of the variation between intraspecific interaction strength within rarity groups, as defined by local population abundance or geographic spread, between different segments. Linear models of the form lm(α~Location) for each rarity group independently. Segments that were significantly different from each other were identified using Tukey's HSD test. Segments different then more than four other segments are highlighted in grey. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 3.2.

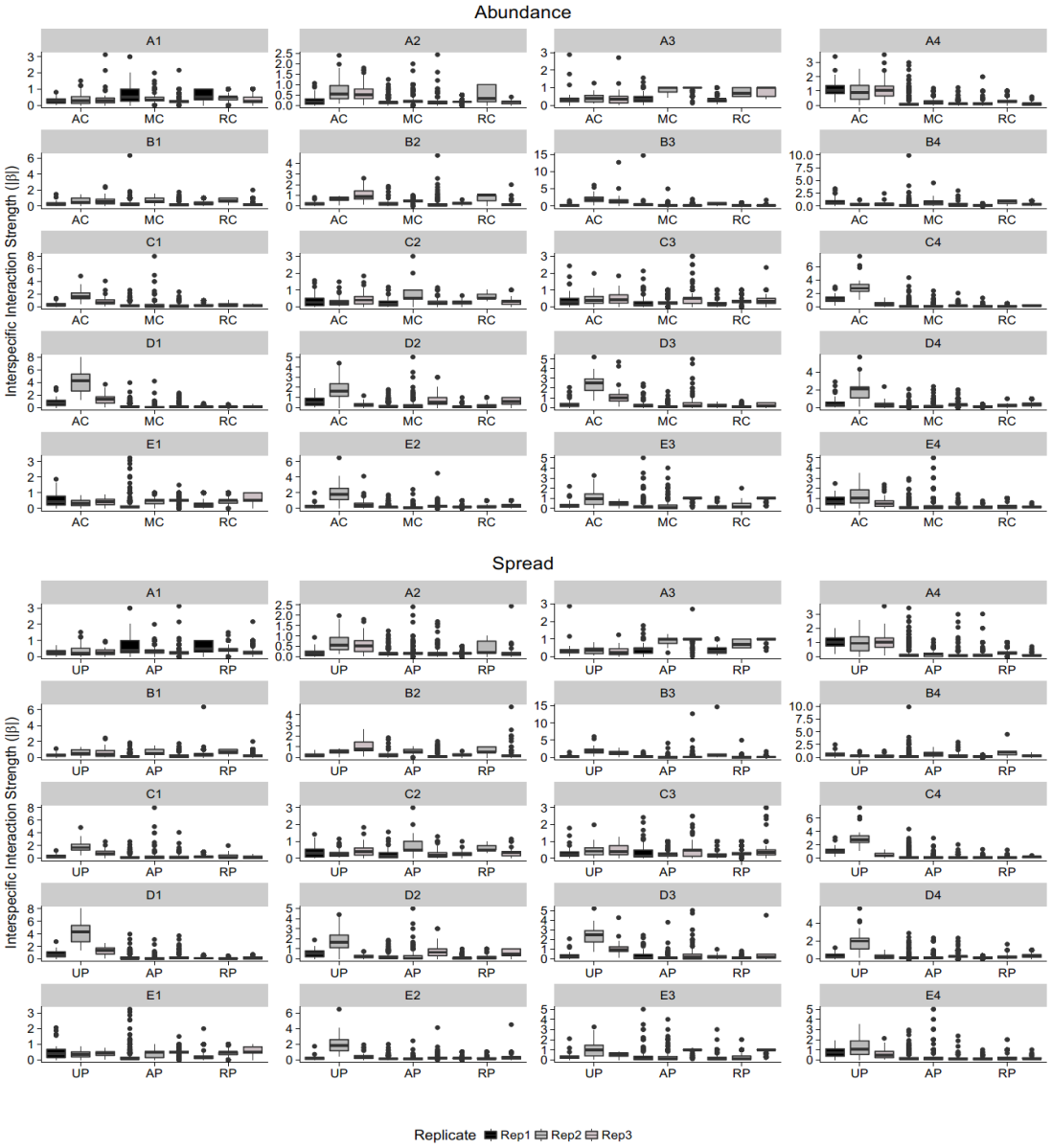
Rarity Basis	Group	P-value	R^2 Adjusted	Different Segments
Local Abundance	AC	1.32e-10	0.0283	A3-A4/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E2/E3/E4 B2-C1/C4/D1/D3/E2/E4 D3-A1/E1 E1-D1
	MC	1.00e-129	0.0431	A1-A4/B3/B4/C1/C2/C3/C4/D1/D2/D4/E1/E2/E3/E4 A2-A4/B3/C1/C3/D2/D4/E2/E4 A3-A4/B3/B4/C1/C2/C3/C4/D1/D2/D4/E2/E3/E4 B1-A4/B3/C1/C3/D2/D4/E2/E3/E4 B2-B3/C1/C2/C3/D2/D4/E2/E3/E4 B4-A4/D4/E2/E4 C1-B3 C2-B3/D4/E4 C4-A2/A4/B1/B2/B4/C1/C2/C3/D2/E1/E2/E3 D1-A2/A4/B1/B2/B4/C1/C2/C3/D2/E1/E3 D3-A1/A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/D2/D4/E1/E2/E3 E1-A4/B3/C1/D2/D4/E2/E4 E4-C2/E1/E3
	RC	8.33e-04	0.0149	D3-A1/B4/C4/E3
Geographic Spread	UP	7.93e-05	0.0210	A3-B3/C1/C3/C4/D1/D3/D4/E2/E3/E4
	AP	2.35e-11	0.0129	A1/A3-A2/A4/B3/B4/C1/C2/C3/C4/D1/D2/D4/E2/E3/E4 A2-A4/B3/C1/D2/D4/E2/E4 B1-A4/B3/C1/C3/D2/D4/E2/E4 B2-A4/B3/C1/C3/D2/D4/E2/E3/E4 B4/C2-A4/B3/D4/E2/E4 C4/D1-A2/A4/B1/B2/B4/C1/C2/C3/D2/E1/E2/E3 D3-A1/A2/A3/A4/B1/B2/B4/C1/C2/C3/D2/E1/E2/E3/E4 E1-D2/D4/E2/E4 E4-D3
	RP	1.74e-117	0.0492	B3-A1/A2/B1/B2 B1-D4 D3-A1/A2/A3/A4/B1/B2/B4/C1/D2/E3

SI Table 3.4 – The model fit of the strength of intraspecific interactions as explained by the rarity groups, as assessed by either local population abundance or geographic spread, and additional factors. The additional factors included segment location, the interaction between location and sequencing depth (Location*N), sequencing depth (N), sample coverage (SC), a segment's species richness (SR), Shannon's diversity (SD), Simpson's entropy (SE) and total community size as assessed by qPCR (qPCR). All values are given to two decimal places and the degrees of freedom (DF) of each model indicated.

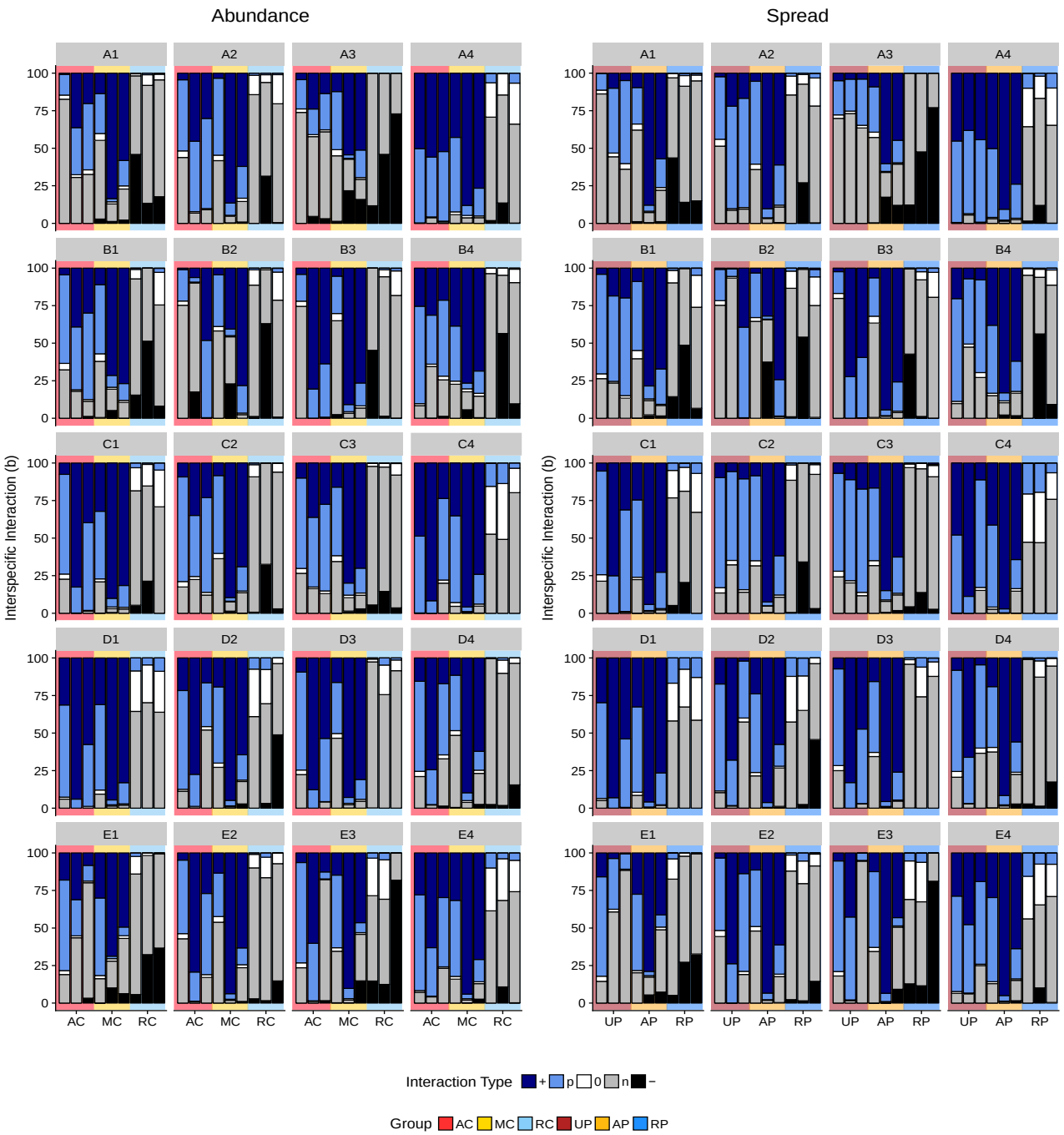
Model	Group+X	DF	AIC	
			Abundance	Spread
LM1	*NA*	4	12060.63	12038.33
LM2	Location	23	11363.54	11408.66
LM3	Location*N	43	10681.00	10732.91
LM4	N	5	10972.57	10991.68
LM5	Location*SC	43	10681.00	10732.91
LM6	SC	5	10975.76	10991.68
LM7	Location*SR	43	10716.74	10769.26
LM8	SR	5	10975.76	10989.70
LM9	Location*SD	43	10786.69	10832.12
LM10	SD	5	11625.18	11613.00
LM11	Location*SE	43	10880.21	10935.33
LM12	SE	5	11983.42	11964.51
LM13	Location*qPCR	43	11034.34	11095.32
LM14	qPCR	5	12001.54	11991.99

SI Table 3.5 – The significance (p-value) of the different proportional contribution of interspecific interaction types to each rarity group, as assed by local population abundance or geographic spread. The variation in interaction types in each rarity group was assessed using a generalised linear model, weighted by interaction number and a possion distribution. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

Different Groups	+	p	0	n	-
AC-MC	0	0	8.63e-198	2.84e-39	0
AC-RC	2.53e-125	0	0	0	0
AP-UP	0	0	1.33e-05	0	0.920
AP-RP	0	0	0	0	0



SI Figure 3.4 – Interspecific interaction strengths within the chicken gut microbiota rarity groups, as defined by local population abundance or geographic spread, as a function of their location within the grid. Rarity groups are defined by the relative rarity of the focal species. There was a significant difference between the interspecific interaction strengths of each rarity group (abundance: p-value < 2.2e-16, R^2 adjusted 0.411; spread: p-value < 2.2e-16, R^2 adjusted 0.365). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 3.2. Different technical replicates are indicated. Values outliers are indicated as dots.



SI Figure 3.5 – Types of interspecific interactions experienced within different rarity groups, as defined by local population abundance or geographic spread, as a proportion of all β interactions across space. The rarity groups are defined by the rarity of the focal species. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2. All three technical replicates are displayed, replicate 1 to 3 from left to right respectively.

SI Table 3.6 – The significance of the variation between interspecific interaction strength within rarity groups, as defined by local population abundance or geographic spread, between different segments. Linear models of the form lm(β~Location) for each rarity group independently. Segments that were significantly different from each other were identified using Tukey's HSD test. Segments different then more than four other segments are highlighted in grey. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 3.2.

Rarity Basis	Group	P-value	R^2 Adjusted	Different Segments
Local Abundance	AC	2.49e-95	0.169	A4-A1/A2/A3/B1/B2/B4/C2/C3/E3 B3-A1/A2/A3/B1/B2/B4/C2/C3/D2/D4/E1/E2/E3/E4 C1-A1/A2/A3/B1/B4/C2/C3/C4 C4-A1/A2/A3/B1/B2/B4/C1/C2/C3/D2/D4/E1/E2/E3/E4 D1-A1/A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D2/D3 /D4/E1/E2/E3/E4 D2/E2/E4-A1/A3/C2 D3-A1/A2/A3/B1/B2/B4/C2/C3/E1/E3 D4-A1/A3/C2/C3 E1-A4/B3/C1/D2/D4/E2
	MC	0.00	0.0815	A1-A2/A3/A4/B1/B2/B3/C1/C2/C3/C4/D1/D2/D3/D4/E1 /E2/E4 A3-A2/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1 /E2/E3/E4 A4-B1/B2/B3/C1/D2/D3/D4/E1/E3 B1-A2/B2/B3/C1/C4/D1/D3/D4/E2/E3/E4 B2-A2 B3-A2/D1/E2/E4 B4-A2/A4/B1/B2/B3/C1/C3/C4/D1/D2/D3/D4/E1/E2/E4 C2-A2/A4/B2/B4/C1/C4/D1/D3/E1/E2/E3/E4 C3-A2/A4/B2/B3/C1/C4/D1/D3/D4/E2/E3/E4 C4-B2/B3/C1/D2/D3/D4/E1/E3/E4 D2-A2/B2/B3/C1/D1/D3/D4/E2/E3/E4 D4-A2/D1/E1/E2/E3/E4 E1-A2/B2/B3/C1/D1/D3/E2/E3/E4 E3-A2/B2/B3/C1/D1/D3/E2/E4
	RC	0.00	0.173	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4 /E1/E2/E3/E4 A3-A2/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1 /E2/E3/E4 B1-A2/B2/B4/C1/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B2-B3/B4/C2/C3/C4/D1/D2/D3/D4/E1/E3/E4 B3-A2/B2/B4/C1/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B4-A2/B2/C1/C2/C3/C4/D1/D2/D3/D4/E2/E4 C2-A2/B2/C1/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 C3-A2/B2/C1/C4/D1/D2/D3/D4/E1/E2/E3/E4 C4/D1-A2/B2/C1/D2/D3/D4/E1/E2/E3/E4 D3-A2/C1/D2/D4/E1/E2/E3 E1/E3-A2/B2/C1/D2/D4/E2/E4 E2-B2 E4-A2/C1/D2/D4/E2

SI Table 3.6 continued.

Rarity Basis	Group	P-value	R^2 Adjusted	Different Segments
Geographic Spread	UP	3.61e-97	0.212	A1/A3-A4/B3/C1/D1/D2/D3/D4/E2/E4 A2-A4/B3/C1/D1/D3/E2 A4-B1/B2/C2/C3/D1/E1 B1-C1 B3-B1/B2/B4/C2/C3/D1/E1/E3 B4-A4/C1/D2/E2 C2-C1/D2/D4/E2/E4 C3-C1/D2/E2 C4-A1/A2/A3/A4/B1/B2/B4/C1/C2/C3/D1/D2/D4/E1/E2/E3/E4 D1-B1/B2/B4/C1/C2/C3/D2/D3/D4/E1/E2/E3/E4 D3-B1/B2/B4/C2/C3/D4/E1/E3/E4 E1-C1/D2/E2/E4
	AP	6.75e-205	0.0531	A1-A2/A3/A4/B1/B2/B3/C1/C2/C3/C4/D1/D3/D4/E2/E4 A2-A3/B1/B4/C2/C3/D2/D4/E1/E3 A3-A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 A4-B1/B4/C2/C3/D2/D4/E1/E3 B1/C2/D4-C4/E2 B4-B1/B2/B3/C1/C2/C3/C4/D1/D3/D4/E1/E2/E4 C3-B2/B3/C1/C4/D1/D3/E2/E3/E4 D2-B1/B2/B3/C1/C2/C4/D1/D3/D4/E2/E4 E1-B2/B3/C1/C4/D1/D3/D4/E2/E3/E4 E3-B1/B2/B3/C1/C2/C4/D1/D3/D4/E2/E4
	RP	0.00	0.152	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 A3-A2/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 A4-A2/B2/B3/B4/C1/C2/C3/C4/D1/D2/D4/E1/E2/E3 B1-A2/A4/B2/B4/C1/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B3-A2/B2/B4/C1/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B4/C2-A2/B2/C1/C2/C3/D1/D2/D3/D4/E1/E2/E3/E4 C3-A2/B2/C1/C4/D1/D2/D3/D4/E1/E2/E3/E4 C4/D1-A2/B2/C1/D2/D3/D4/E1/E2/E3/E4 D2-A2/B2/C1 D3-A2/B2/C1/D2/D4/E1/E2/E3 D4-A2 E1-A2/B2/C1/D2/D4/E2/E4 E3-A2/B2/C1/D2/D4/E2 E4-A2/B2/C1/D2/D4/E2/E3

SI Table 3.7 – The effect of location, sampling depth ('n'), sample coverage ('SC') and diversity estimates (species richness 'SR', Shannon's diversity 'SD' or Simpson's entropy 'SE') upon the different proportional contribution of interspecific interaction types to each rarity group, as assed by local population abundance or geographic spread. The models were of the form lm(prop~Group+X), where 'Group' is the rarity group and X is the variable specified. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

Rarity Basis	Model	Group+X	DF	AIC				
				+	p	0	n	-
Local Abundance	LM1	*NA*	4	1643.120	1494.985	1168.919	1592.505	1418.863
	LM2	Location	23	1662.362	1516.992	1172.440	1597.636	1421.850
	LM3	Location*n	43	1651.856	1492.513	1190.090	1606.242	1392.136
	LM4	n	5	1638.339	1480.923	1145.686	1581.166	1376.548
	LM5	Location*SC	43	1656.999	1499.233	1190.190	1607.123	1386.581
	LM6	SC	5	1643.548	1472.235	1155.880	1591.285	1358.155
	LM7	Location*SR	43	1655.103	1492.572	1190.042	1610.306	1389.277
	LM8	SR	5	1638.204	1470.068	1143.515	1580.867	1349.370
	LM9	Location*SD	43	1653.793	1498.694	1190.933	1607.882	1389.867
	LM10	SD	5	1639.767	1489.066	1168.555	1592.422	1378.336
	LM11	Location*SE	43	1656.948	1498.423	1196.156	1613.913	1423.509
	LM12	SE	5	1641.955	1496.764	1170.795	1594.476	1408.721
	LM13	Location*qPCR	43	1616.761	1496.203	1201.325	1598.154	1430.009
	LM14	qPCR	5	1641.709	1492.690	1150.416	1577.795	1417.950
Geographic spread	LM1	*NA*	4	1648.641	1522.183	1158.510	1620.278	1410.811
	LM2	Location	23	1663.866	1537.572	1163.850	1619.747	1418.453
	LM3	Location*n	43	1659.769	1517.954	1181.386	1625.782	1402.089
	LM4	n	5	1636.774	1511.777	1135.896	1599.037	1375.350
	LM5	Location*SC	43	1663.812	1523.299	1181.492	1627.354	1398.106
	LM6	SC	5	1645.971	1501.974	1144.662	1611.992	1361.883
	LM7	Location*SR	43	1659.327	1517.439	1181.212	1630.007	1399.307
	LM8	SR	5	1638.176	1498.184	1132.611	1596.229	1353.005
	LM9	Location*SD	43	1657.575	1520.932	1182.081	1626.972	1400.306
	LM10	SD	5	1643.556	1514.905	1157.519	1616.346	1376.524
	LM11	Location*SE	43	1663.657	1523.980	1187.839	1631.468	1426.520
	LM12	SE	5	1647.117	1523.776	1160.483	1621.837	1402.156
	LM13	Location*qPCR	43	1632.694	1526.080	1192.813	1621.251	1432.237
	LM14	qPCR	5	1647.021	1518.670	1140.374	1604.089	1410.349

SI Table 3.8 – The significance of the difference between the proportional contribution of interspecific interaction types to each combined rarity group. The variation in interaction types in each rarity group was assessed using a generalised linear model, weighted by interaction number and a possion distribution. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

Combined Group	+	p	0	n	-
AC:AP	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16
AC:UP	< 2e-16	< 2e-16	< 2e-16	< 2e-16	0.877
MC:AP	< 2e-16	< 2e-16	< 2e-16	0.292	< 2e-16
MC:RP	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16
MC:UP	< 2e-16	< 2e-16	2.17e-02	< 2e-16	0.900
RC:RP	< 2e-16	< 2e-16	< 2e-16	< 2e-16	< 2e-16

SI Table 3.9 – The model fit of combined group (LM1) and combined group by location (LM2) for the intraspecific and interspecific interaction strengths. The models are of the form LM1; lm(α~CombinedGroup) and LM2; lm(α~CombinedGroup+Location) and compared using AIC with the degrees of freedom (DF) indicated. All values are given to two decimal places. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

Model	DF	Interaction Strength		DF	Interaction Type				
		α	β		+	p	0	n	-
LM1	7	10553.27	51723.06	6	915426.80	460467.90	31584.55	563823.30	42202.36
LM2	26	9906.13	48062.68	25	780450.90	416775.40	29275.49	405202.80	27608.01

SI Table 3.10 – Number of unique species within each rarity group pair. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. For full explanation of abbreviations see Table 3.2.

CS	AC	MC	RC	Total
UP	0	0	0	0
AP	1	379	0	380
RP	1	320	1772	2093
Total	2	699	1772	2473

SI Table 3.11 – The variation in interspecific interaction strength within each focal species rarity group as explained by the competitor species rarity group. Summary statistics are provided from a linear model of the form lm(|β|~CompetitorSpecies), where |β| is the strength of the interspecific interaction between a focal-competitor species pair, for each focal species rarity group. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. Complete abbreviations in Table 2.

Focal Species Group	P-value	R^2 Adjusted	Different Competitor Species Groups
AC	0.00	0.0114	AC-MC/AP/RC/RP MC-RC/RP AP-RC/RP RP-RC
UP	0.00	0.0101	AC-MC/AP/RC/RP MC-RC/RP AP-RC/RP RP-RC
MC	0.00	0.0451	AC-MC/AP/RC/RP MC-AP/RC/RP AP-RC/RP RP-RC
AP	0.00	0.0279	AC-MC/AP/RC/RP MC-AP/RC/RP AP-RC/RP RP-RC
RC	8.90e-39	0.0403	AC-MC/AP/RP RP-AC/MC/AP
RP	8.02e-109	0.0827	AC-MC/AP/RP RP-MC/AP

SI Table 3.12 - The variation in the proportional contribution of interspecific interaction type within each focal species rarity group as explained by the competitor species rarity group. The variation in interaction types in each rarity group was assessed using a generalised linear model, weighted by interaction number and a poisson distribution. Only non-significant p-values are provided in full. Focal-competitor rarity pairs coloured black do not occur. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 2. All values are given to three significant figures.

		Focal Species					
Competitor Species	+	AC	UP	MC	AP	RC	RP
	AC					0.997	
	UP						
	MC					0.998	
	AP					0.998	
	RC						
	RP					0.998	

		Focal Species					
	p	AC	UP	MC	AP	RC	RP
	AC						
	UP						
	MC						
	AP						
	RC						
	RP						

Competitor Species	-	AC	UP	MC	AP	RC	RP
	AC	0.909	0.999				
	UP						
	MC	0.941	1.00		0.131		
	AP	0.940	1.00				
	RC	0.941	1.00		0.185		
	RP	0.942	1.00		0.148		

	n	AC	UP	MC	AP	RC	RP
	AC						
	UP						
	MC	0.801	0.692				
	AP	0.939	0.707				
	RC	0.612	0.356				
	RP	0.558	0.431				

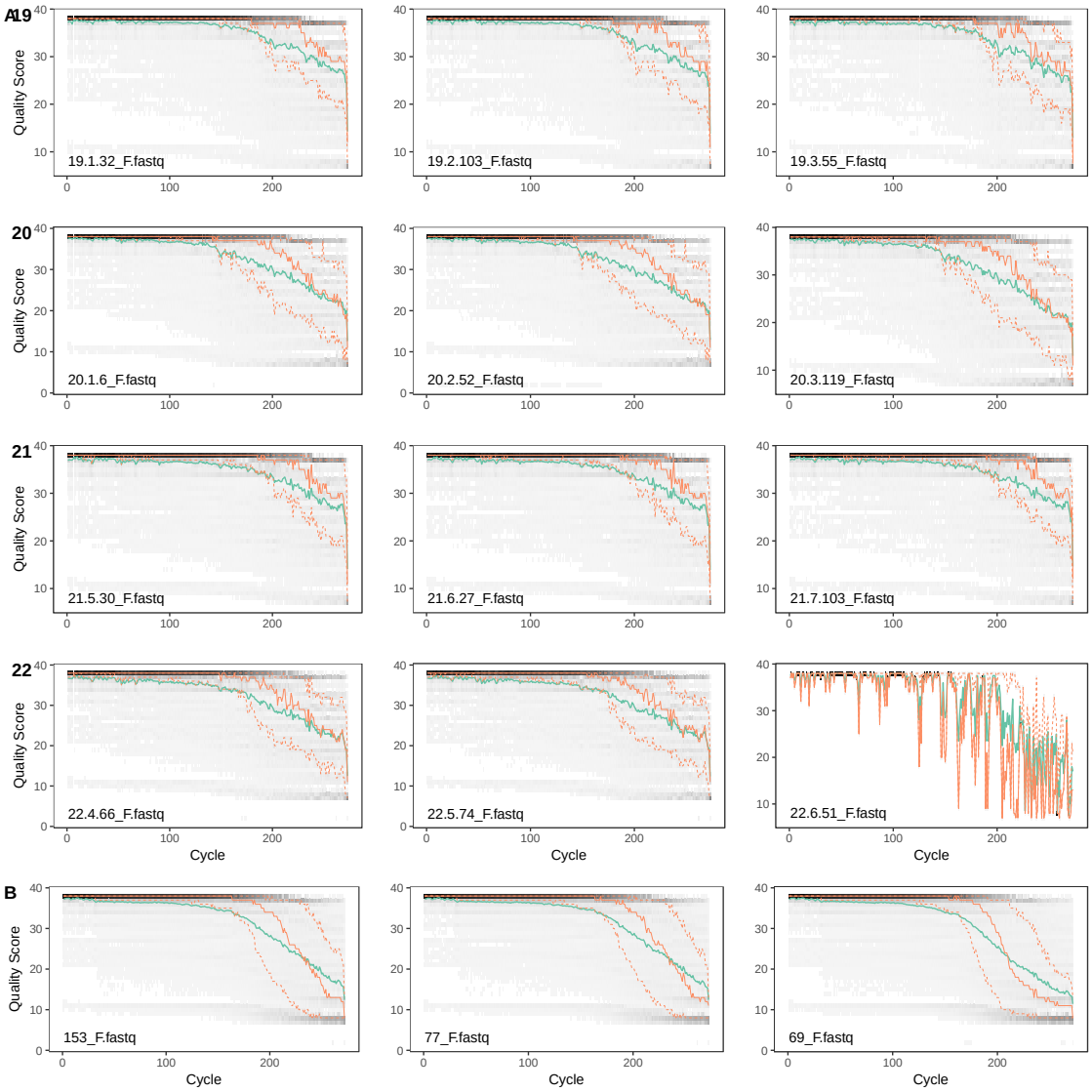
Competitor Species	0	AC	UP	MC	AP	RC	RP
	AC						
	UP						
	MC					0.680	
	AP					0.714	0.0534
	RC						
	RP						

SI Table 3.13 – The explanatory power of competitor species group (CS Group), location (Location) and a combination of competitor species group and location (CS Group + Location) on interspecific interaction strength within each focal species rarity group. The abundant, global, moderate and widespread degrees of freedom were 6, 25 and 21 for each interaction type model respectively, whereas the rare and localised degrees of freedom were 5, 24 and 21 respectively. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. The best model for each interaction type and focal species group was highlighted in grey. All values are given to three decimal places.

Model	Focal Species Group					
	AC	UP	MC	AP	RC	RP
CS Group	8214295	6519980	2062580	3689121	27686.54	48073.72
CS Group + Location	7918778	6240220	2026660	3620211	15222.13	33724.70
Location	7929185	6246564	2043068	3637671	15411.25	34218.87

SI Table 3.14 – The explanatory power of competitor species group (CS Group), location (Location) and a combination of competitor species group and location (CS Group + Location) on interaction type within each focal species rarity group. The abundant, global, moderate and widespread degrees of freedom were 5, 24 and 20 for each interaction type model respectively, whereas the rare and localised degrees of freedom were 4, 23 and 20 respectively. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread and 'RP' localised. The type of interactions are strong positive ('+'), weak positive ('p'), neutral ('0'), weak negative ('n') and strong negative ('-'). The best model for each interaction type and focal species group was highlighted in grey. All values are given to three decimal places.

Focal Species Group	Interaction Type	CS Group	CS Group + Location	Location
AC	+	972867.70	736373.50	762894.80
	p	481982.50	397432.20	429681.80
	0	38847.76	29176.43	30378.14
	n	786077.80	505543.90	507366.00
	-	23918.22	11251.81	11255.21
UP	+	728648.50	471582.30	488164.80
	p	352091.20	272010.00	294723.40
	0	29829.75	21815.32	22419.28
	n	619000.40	390875.10	394174.50
	-	10.00	48.00	40.00
MC	+	651473.90	612952.90	650238.10
	p	337869.60	309207.30	317843.30
	0	12017.86	11612.02	23876.10
	n	255295.50	189160.70	200121.40
	-	40979.10	24096.50	26283.99
AP	+	1065618.10	994618.30	1035216.20
	p	547070.40	511059.20	531237.00
	0	21068.96	19036.49	28297.49
	n	450322.80	332110.20	336935.60
	-	53299.78	29495.09	30202.44
RC	+	118.52	115.017	114.14
	p	3386.11	1316.14	1320.55
	0	8361.79	3468.38	3534.78
	n	10476.86	6316.68	6506.73
	-	18552.89	11884.73	12018.60
RP	+	374.90	344.10	425.23
	p	6251.11	2123.79	2134.99
	0	10188.86	4271.82	4313.67
	n	13593.37	7745.22	8002.07
	-	24371.55	15510.40	15735.90



SI Figure 4.1 – Quality plots for the A1 segment of all wild-type birds across all MiSeq runs with all technical replicates. A) 16S sequence quality scores. B) TCRβ sequence quality scores for MiSeq22. Technical replicates run left to right (1st to 3rd respectively), MiSeq identification to the right of the 1st replicate.

SI Table 4.1 – Percentage of immune sequences and clones lost during data trimming. 'Raw' datasets refer to MiSeq sequences that have had their primers removed, been demultiplexed, deprelicated and in the case of the IgA-producing B-cell receptor data clustered. 'Trimmed' datasets are raw datasets that have been further modified. Non-functional sequences refer to amino acid sequences including a frame-shift or stop codon. Values are given to two decimal places.

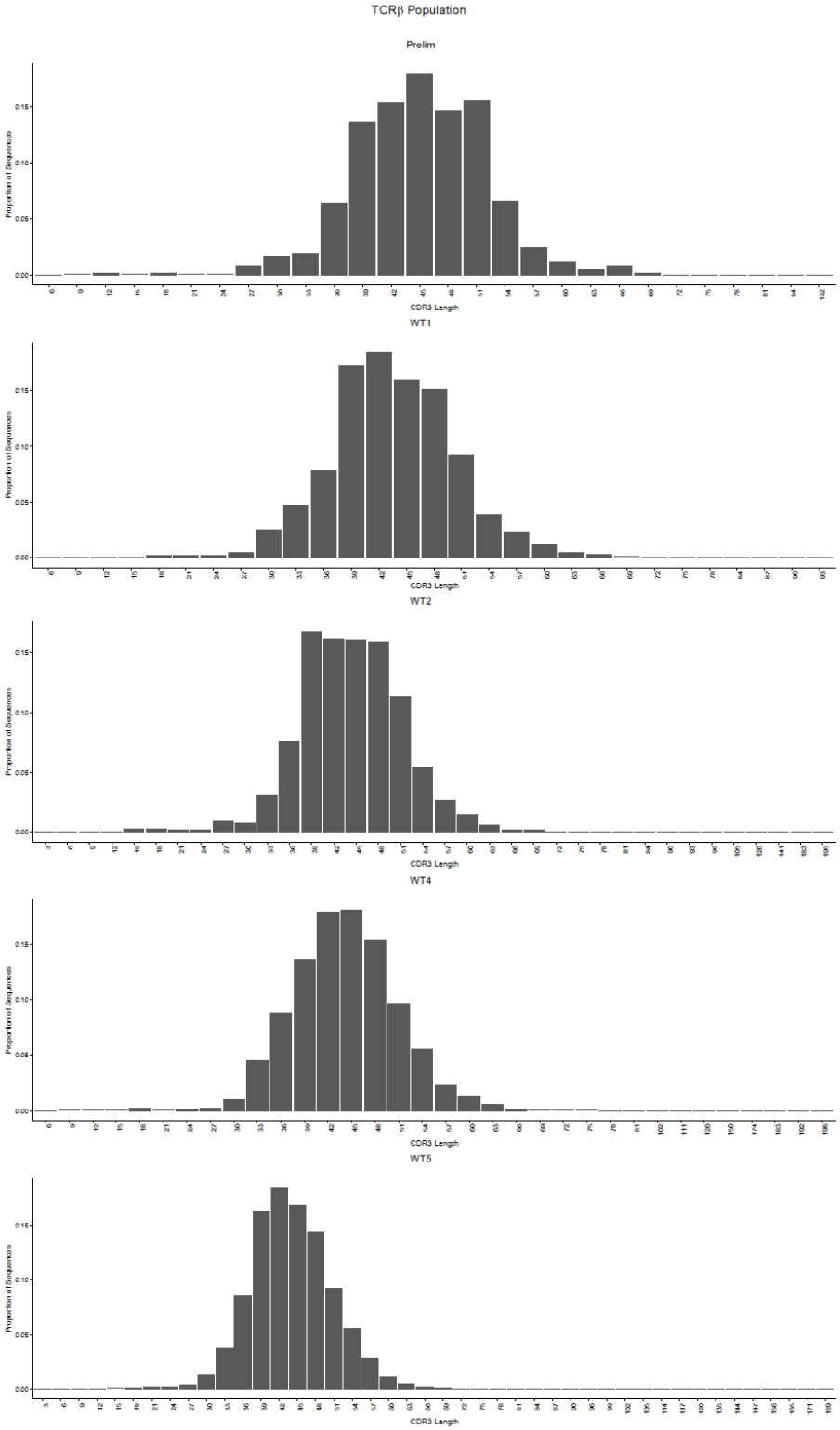
Bird	Rep	Sequences lost in trim (%)		Clones lost in trim (%)		Raw datasets				Trimmed datasets			
						Non-functional Sequences (%)		Non-functional Clones (%)		Non-functional Sequences (%)		Non-functional Clones (%)	
		TCRβ	BCR	TCRβ	BCR	TCRβ	BCR	TCRβ	BCR	TCRβ	BCR	TCRβ	BCR
WT1	1	0	11.72	0	2.62	11.06	32.84	12.15	35.40	11.06	40.71	12.15	37.09
	2	0	11.65	0	2.80	12.97	26.74	12.38	29.75	12.97	35.28	12.38	31.71
	3	0	11.79	0	4.87	12.40	26.91	12.22	29.15	12.40	35.53	12.22	32.60
WT2	1	0	97.97	0	99.07	15.46	29.06	16.49	31.01	15.46	98.56	16.49	99.36
	2	0	97.02	0	98.85	18.02	34.92	16.88	36.49	18.02	98.06	16.88	99.27
	3	0	98.82	0	99.35	17.00	37.07	16.74	38.36	17.00	99.26	16.74	99.60
WT4	1	0	61.35	0	65.67	18.78	34.22	18.89	36.54	18.78	74.58	18.89	78.21
	2	0.01	6.02	0.04	1.96	18.57	31.33	18.90	34.56	18.57	35.47	18.90	35.85
	3	0	22.49	0	21.78	17.08	20.34	18.98	25.25	17.08	38.25	18.98	41.53
WT5	1	0	17.44	0	19.68	17.56	36.99	17.86	39.27	17.56	47.98	17.86	51.22
	2	0	1.56	0	0.74	17.91	35.78	18.02	37.18	17.91	36.78	18.02	37.64
	3	0	1.23	0	0.68	17.56	38.47	18.01	39.85	17.56	39.22	18.01	40.26

SI Table 4.2 – Linear model results, and Tukey HSD summaries, for each rarity group's proportional contribution to each population (Chapter 4 Figure 4). The data presented for the BCR population is provided for both the raw and trimmed datasets. The trimmed data presented is modified according to Chapter 4 B.5.3. The linear models were of the form lm(Prop~Bird), where 'prop' is the proportional contribution of the population structure abundance groups calculated independently. Group 2 refers to the most abundant microbes and group 6 the rarest. Values are given to three significant figures.

Population		Group	P-value	R^2 Adjusted	Birds Different
Microbiota		2	2.63e-05	0.933	Prelim/WT1/WT5-WT2/WT4
		3	5.37e-04	0.868	Prelim-WT1/WT2/WT4/WT5
		4	1.17e-05	0.945	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT5 WT5-WT4
		5	1.59e-03	0.831	Prelim-WT1/WT2/WT4
		6	2.05e-01	0.450	*NA*
TCRβ		2	5.98e-04	0.837	WT1-Prelim/WT2/WT4/WT5
		3	1.29e-05	0.926	Prelim-WT1/WT2/WT4/WT5 WT1-WT2/WT4/WT5
		4	2.11e-02	0.654	*NA*
		5	9.03e-05	0.889	WT1-Prelim/WT2/WT4/WT5
		6	5.88e-05	0.899	WT1-Prelim/WT2/WT4/WT5 Prelim-WT5
BCR	Raw	2	6.00e-04	0.838	Prelim-WT1/WT2/WT5 WT4-WT2/WT5
		3	0.100	0.5101	*NA*
		4	3.78e-02	0.6072	Prelim-WT5
		5	3.20e-03	0.7685	WT4-WT1/WT2/WT5
		6	3.20e-03	0.7687	WT4-WT1/WT2/WT5
	Trimmed	2	2.66e-03	0.778	Prelim-WT1/WT2/WT5 WT4-WT1/WT2/WT5
		3	7.90e-03	0.720	WT1-WT2/WT4 WT5-WT2
		4	1.76e-02	0.668	Prelim-WT2
		5	6.90e-03	0.728	Prelim-WT1 WT1/WT5-WT4
		6	2.27e-05	0.917	Prelim-WT1/WT2 WT1/WT5-WT2/WT4

SI Table 4.3 – The sum of least squares values for the deviation from normality of the large-scale TCRβ population. Values are given to three decimal places. 'Prelim' refers to the preliminary L6 bird, 'WT' refers to the JH wild-type birds. Values are given to three significant figures.

Bird	Rep	Least Squares Sum
Prelim	1	0.089
	2	0.087
	3	0.082
WT1	1	0.075
	2	0.067
	3	0.061
WT2	1	0.098
	2	0.105
	3	0.106
WT4	1	0.105
	2	0.101
	3	0.097
WT5	1	0.107
	2	0.106
	3	0.100



SI Figure 4.2 – TCRβ spectratypes for the preliminary bird and the JH wild-type (WT). Displayed is the first technical replicate for each bird. Each spectratype approximates a normal distribution bird (sums of least squares regression data in SI Table 37).

SI Table 4.4 – The significance of variation in population structure abundance groups in the TCRβ population, with the driving segments highlighted, within the preliminary and JH wild-type (WT) birds. Only significant associations are shown. Group 2 refers to the most abundant microbes and group 6 the rarest. P-values and R^2 adjusted values are derived from a linear model of the form lm(Prop~Segment) where Prop is the proportional contribution of a specified abundance group. Within each of the abundance groups differences between segments, as assessed using Tukey's HD test, are indicated with the driving segments highlighted (grey). Values are given to three significant figures.

Bird	Group	P-value	R^2 Adjusted	Different Segments
WT1	2	2.68e-02	0.485	D2-B1
	6	1.44e-02	0.481	D2-A4
WT2	4	5.50e-03	0.497	*NA*
WT5	2	3.12e-10	0.6991	D10-A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/A11/B1/B3/B4/B5/B6/B7/B8/B9/B10/B11/C1/C2/C4/C5/C6/C7/C9/C10/C11/D1/D2/D3/D4/D5/D6/D7/D8/D9; B2-A2/A4/A5/B3/C7/D5/D8 D11-A2/A3/A4/A5/B3/B5/C7/C8/D5/D8 C3-A5/D8/D10; B11-A5
	3	5.08e-03	0.484	*NA*
	4	2.63e-12	0.7362	B2-A2/A3/A4/A5/A7/A8/A9/A11/B3/B5/B6/B9/B10/C1/C2/C4/C6/C7/C11/D1/D5/D6/D8/D9; D10-A1/A2/A3/A4/A5/A7/A8/A9/A10/A11/B1/B3/B5/B6/B8/B9/B10/C1/C2/C4/C6/C7/C9/C11/D1/D5/D6/D8/D9; B11-A1/A3/A4/A5/A7/A9/B5/B9/B10/C1/C7/C11/D1/D5/D6/D8; D11-A1/A2/A3/A4/A5/A7/A8/A9/A10/A11B1/B3/B5/B6/B8/B9/B10/C1/C2/C4/C6/C7/C9/C11/D1/D5/D6/D8/D9; C3-A3/A4/A5/A9/B5/B9/B10/C1/C7/D1/D5/D6/D8; D4-C7
	5	1.02e-05	0.589	*NA*
	6	9.42e-06	0.590	*NA*
Prelim	2	3.82e-07	0.754	A1-A2/A3/A4/B1/B2/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B3-A4/C3/C4/D2
	3	3.35e-10	0.834	B1-A1/A4/C2/C3/C4/D2/D3/E4 B3-A1/A4/C2/C3/C4/D2/D3/E4 C1-A1/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 D2-D1 D4-A1/A2/A3/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 E1-A1/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 E2-A1/A4/B2/C2/C3/C4/D2/D3
	4	3.075109e-20	0.950	A1-A2/A3/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 B1-A1/A2/A3/A4/B2/B4/C1/C2/C3/C4/D2/D3/D4/E1/E2/E3/E4 B3-A2/A3/A4/B1/B2/B4/C2/C3/C4/D2/D3/E2/E3/E4 C1-A2/A3/A4/B1/B2/B4/C2/C3/C4/D2/D3/E3/E4 D1-A1/A2/A3/A4/B2/B3/B4/C1/C2/C3/C4/D2/D3/E1/E3/E4 D4-A3/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 E1-A2/A3/A4/B2/B4/C2/C3/C4/D2/D3/E3/E4 E2-A2/A3/A4/B2/B4/C2/C3/C4/D2/D3/E1/E3/E4
	5	4.605284e-09	0.808	A4-A1/A2/B1/B3/C1/D1/D4/E1/E2 C2-B3/C1/D4/E1 C3-A1/A2/B1/B3/C1/C2/D1/D4/E1/E2 C4-A1/A2/B1/B3/C1/C2/D1/D4/E1/E2; D2-A1/A2/A3/B1/B2/B3/B4/D1/D4/E1/E2/E3 D3-B3/C1/D4/E1 E4-B3/C1/D4/E1
	6	5.017481e-08	0.781	A4-A1/B1/B3/C1/E1/E2 B3-A4/C2/C3/C4/D2/D3 C2-D4/E1 C3-B1/C1/D4/E1 C4-A4/C1/D4/E1 D2-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3 D3-C1/D4/E1

SI Table 4.5 – The significance of the effect of location upon the sums of least squares deviation from normality of the TCRβ spectratypes ('Segmental Vartiation'), and the correlation of CDR3 length distribution shape with the distance between two segments ('Dissimilarity by Distance'). The sums least squares calculations for the association with distance were calculated from the difference between the shapes of the distributions of the comparison segments spectratypes. Values are given to three significant figures.

	Segmental Variation		Dissimilarity by Distance	
Bird	P-value	R^2 Adjusted	P-value	R^2 Adjusted
Prelim	1.70e-03	0.587	2.10e-02	9.29e-03
WT1	0.869	0.251	0.213	1.04e-03
WT2	0.803	0.280	0.715	5.91e-05
WT4	1.64e-02	0.563	0.322	7.70e-04
WT5	0.634	0.307	0.882	7.78e-06

SI Table 4.6 – Phyla that show a significant effect of bird upon the phylum's proportional contribution to the community with birds driving the difference shown. 'Prelim' refers to the preliminary L6 bird analysed in Chapter 2, 'WT' refers to the wild-type JH birds investigated. Corresponding plots are shown in Figure 4.4. P-values and R^2 adjusted values are derived from a linear model of the form lm(Prop~Bird) where Prop is the proportional contribution of a specified phyla. The significantly different birds were identified using Tukey's HD test.

Phylum	P-value	R^2 Adjusted	Birds Different
Actinobacteria	8.47e-06	0.972	WT1-WT2/WT4/WT5
Bacteroidetes	1.62e-02	0.750	WT4-WT1/WT2
Cyanobacteria	1.70e-06	0.964	Prelim-WT1/WT2/WT5 WT2-WT1/WT4/WT5
Elusimicrobia	4.14e-03	0.833	WT4-WT1/WT2
Firmicutes	1.66e-05	0.940	Prelim-WT1/WT4 WT1-WT2/WT4/WT5 WT4-WT2
Proteobacteria	5.62e-07	0.972	Prelim-WT1/WT4/WT5 WT1-WT2/WT5 WT2-WT4/WT5 WT4-WT5
Unassigned	4.87e-11	0.997	Prelim-WT1/WT2/WT4/WT5

SI Table 4.7 – Families that show a significant effect of bird upon the family's proportional contribution to the community with birds driving the difference shown. 'Prelim' refers to the preliminary L6 bird analysed in Chapter 2, 'WT' refers to the wild-type JH birds investigated. Corresponding plots are shown in Figure 4.4. P-values and R^2 adjusted values are derived from a linear model of the form lm(Prop~Bird) where Prop is the proportional contribution of the specified family. The significantly different birds were identified using Tukey's HD test and are indicated as 'Different bird – birds distinct from'. For example in the Bifidobacteriaceae family the JH wild-type bird WT1 is significantly different from all the other JH wild-type birds. Values are given to three significant figures.

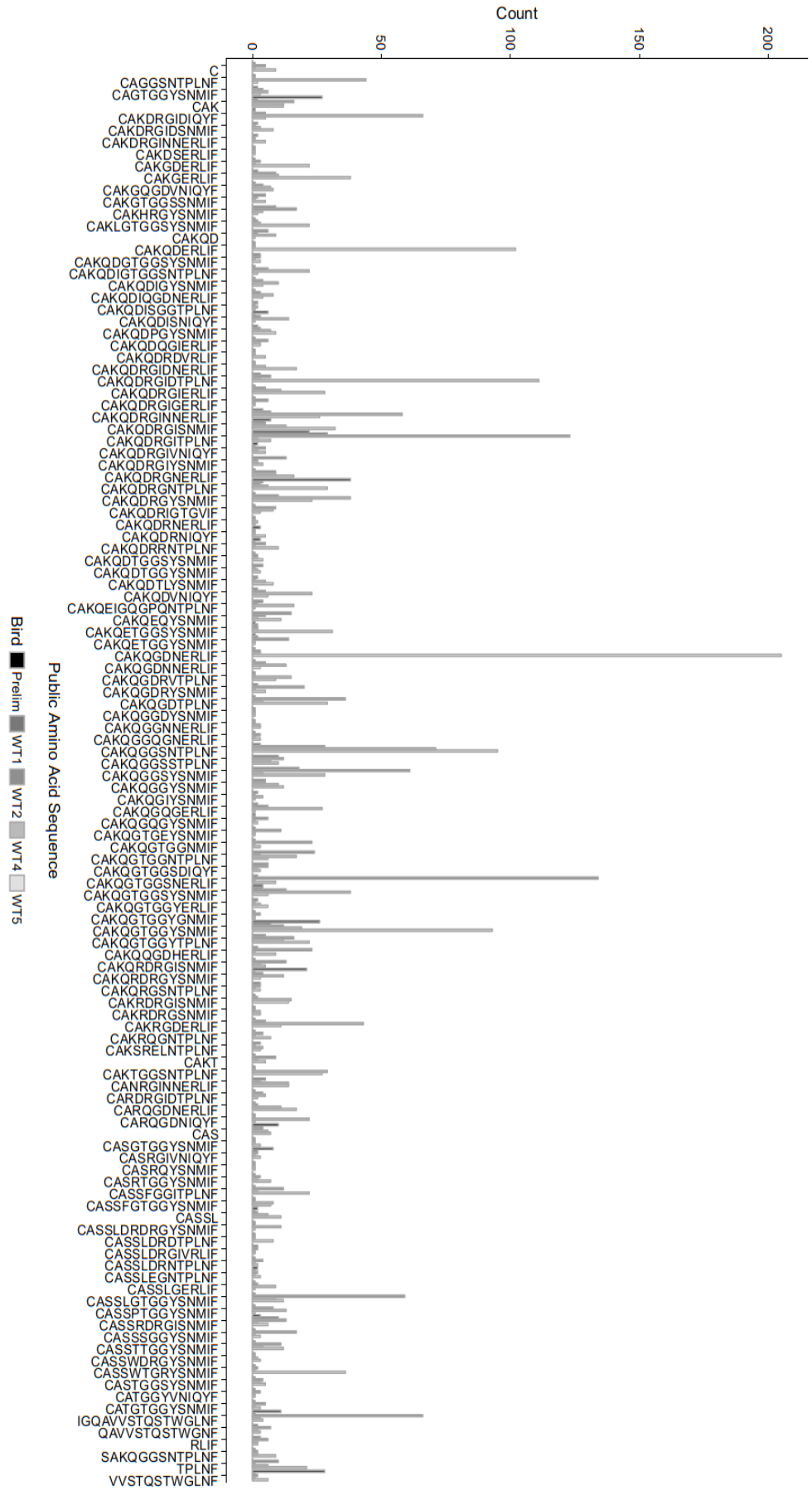
Family	P-value	R^2 Adjusted	Different Birds
Bifidobacteriaceae	4.83e-08	0.994	WT1-WT2/WT4/WT5 WT5-WT2
Christensenellaceae	2.88e-05	0.989	WT4/WT5-WT1/WT2
Clostridiaceae	3.17e-05	0.966	Prelim/WT1-WT2/WT4/WT5 WT1- WT2/WT4/WT5
Elusimicrobiaceae	1.97e-02	0.792	WT1-WT5
Lachnospiraceae	4.76e-09	0.990	Prelim-WT1/WT2/WT4/WT5 WT1-WT2/WT4/WT5
Ruminococcaceae	9.55e-05	0.911	Prelim-WT1/WT5 WT2-WT1/WT4/WT5 WT5-WT1/WT4
Veillonellaceae	5.49e-03	0.862	WT2-WT4/WT5
Unassigned	1.68e-05	0.940	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT4/WT5

SI Table 4.8 – The full list of family-level taxonomic assignments contributing to the 'other' category plotted for the JH wild-type and preliminary bird in Figure 4.4. The phylum-level identification is indicated, as is the naming of the group in which the proportional contribution of the family is displayed.

Phylum	**Family**
Actinobacteria	Coriobacteriaceae
	Corynebacteriaceae
	Micrococcaceae
Bacteroidetes	Barnesiellaceae
Deinococcus-Thermus	Deinococcaceae
Euryarchaeota	Methanomassiliicoccaceae
Firmicutes	Bacillaceae
	Erysipelotrichaceae
	EtOH8
	Eubacteriaceae
	Lactobacillaceae
	Peptococcaceae
	Peptostreptococcaceae
	Planococcaceae
	Streptococcaceae
Lentisphaerae	Victivallaceae
Proteobacteria	Alcaligenaceae
	Enterobacteriaceae
	Halomonadaceae
	Helicobacteraceae
	Mitochondria
	Shewanellaceae
	Succinivibrionaceae

SI Table 4.9 – The summary statistics of the association between bird and the diversity indices for the BCR population. 'Effect of Bird' refers to the significance of the association between a given diversity estimate and the comparison birds (both L6 preliminary and JH wild-type). The birds which show a significant difference are indicated. 'WT-Prelim Difference' refers to the significance of the difference between the preliminary bird and the JH wild-type birds as a collective for a given diversity estimate. All values are given to three significant figures.

Dataset	Estimate	Effect of Bird			WT-Prelim Difference	
		P-value	R^2 Adj	Different Birds	P-value	R^2 Adj
Raw	Species Richness	5.81e-06	0.937	Prelim-WT1/WT2/WT4/WT5 WT4-WT2/WT5	2.73e-06	0.826
	Shannon's Diversity	6.93e-04	0.832	WT4-Prelim/WT2/WT5 Prelim-WT1	3.03e-03	0.504
	Simpson's Entropy	6.29e-03	0.733	WT4-Prelim/WT2/WT5	1.67e-01	0.141
Trimmed	Species Richness	3.99e-09	0.995	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT4/WT5	3.03e-07	0.916
	Shannon's Diversity	1.51e-07	0.987	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT4/WT5	3.41e-06	0.870
	Simpson's Entropy	8.44e-06	0.964	Prelim-WT1/WT2/WT4/WT5 WT2-WT1/WT4/WT5	8.09e-05	0.770



SI Figure 4.3 - The shared T-cell receptor β amino acid counts in the JH wild-type birds and the preliminary bird.

SI Table 4.10 – The significance of variation between segments in the diversity indices of the microbiota and immune populations, within each JH wild-type bird and the preliminary bird. Segments that are significantly different, as assessed by Tukey's HSD test, are indicated and highlighted in grey where more than four other segments are different.

Estimate		Bird	P-value	R^2 Adj	Different Locations
Microbiota	SR	WT1	7.44e-04	0.589	A8-A1/A2/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6/B7/B8/C1/C2/C3/C4/C5/C6/C7/D1/D2/D3/D4/D5/D6/D7/D8 C8-A1/A2/A3/A5/A6/A7/B1/B2/B3/B4/B5/B6/B7/B8/C2/C3/C4/C7/D1/D2/D3/D5/D8
		WT2	8.88e-05	0.590	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/B1/B2/B3/B4/B5/B7/B9/B10/C1/C2/C3/C5/C7/C8/C9/C10/D1/D2/D3/D4/D5/D6/D7/D8/D9/D10
		Prelim	2.17e-02	0.503	*NA*
	SD	WT1	7.31e-03	0.534	A8-A1/A2/A3/A5/A6/B1/B2/B4/B5/C2/C4/C5/C7/D1/D2/D3/D5 C8-A2/A3/A6/C7/D3
		Prelim	2.79e-05	0.683	A1-B1/D1/D2/D4/E3/E4 E3-B3/B4/C3/D3/E1/E2
	SE	WT1	8.46e-04	0.587	C8-A2/A3/A4/A6/B5/C2/C5/C6/D1/D2/D3
		WT2	3.64e-06	0.638	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/B1/B2/B3/B4/B5/B7/B9/B10/C1/C2/C3/C5/C7/C8/C9/C10/D1/D2/D3/D4/D5/D6/D7/D8/D9/D10 B10-A3/A4/B7/D3/D4/D5/D7/D9/D10
		Prelim	3.75e-07	0.754	A1-A3/A4/B1/B2/D1/D2/D4/E3/E4 B3-A3/B1/D1/D2/D4/E3 D2-A2 D3-A2/A3/B1/D2/D4/E3 E2-A3/B1/D2/D4/E3 E3-A2/C3
TCRβ	SR	Prelim	5.44e-19	0.943	A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4/E1/E2/E3/E4 B2-D4/E1 B3-B2/B4/D4 C2/C3-A1/A2/A3/B1/B2/B3/C1/D1/D4/E1/E2 C4-B1/B2/B3/C1/D1/D3/D4/E1/E2/E3/E4 D2-A1/A2/A3/B1/B2/B3/C1/D1/D3/D4/E1/E2/E3/E4 D3-A1/A2/A3/B1/C1/D1/D4/E1/E2 E3/E4-A1/A3/C1/D4/E1/E2
		WT5	2.56e-13	0.752	A2-A6/A10/B1/B4/B5/B6/B7/B8/B11/C2/C3/C4/C5/C6/C8/C10/D2/D3/D4/D7/D9/D10/D11 A3-B4/B11/D4/D10/D11 A4-A6/A10/B1/B2/B4/B5/B6/B7/B8/B11/C2/C3/C4/C5/C6/C8/C10/D2/D3/D4/D10/D11 A5-A6/A10/B1/B2/B4/B5/B6/B7/B8/B11/C2/C3/C4/C6/C8/C10/D2/D3/D4/D10/D11 B2-A1/A2/A3/A7/A8/A9/A11/B10/C1/C7/C11/D1/D5/D6 C7-B4/C3/D4/D10/D11/B11 D8-A6/B2/B4/B5/B7/B8/C3/C8/D4/C10/D10/B11/D11
	SD	Prelim	4.90e-30	0.984	A1-B1/D1 A2-A1/B3/C1/E1/E2 A3-A1/B3/D4/E1/E2 A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D3/D4/E1/E2 B2/B4-A1/A3/B1/B3/C1/D1/D2/D3/D4/E1/E2 C2-A1/A2/A3/B1/B2/B3/B4/C1/C4/D1/D2/D4/E1/E2/E3/E4 C3/C4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D3/D4/E1/E2/E3/E4 D2-A1/A2/A3/B1/B3/C1/D1/D3/D4/E1/E2 D3-A1/A2/A3/B1/B3/C1/D1/D4/E1/E2 E3/E4-A1/A2/A3/A4/B1/B2/C1/D1/D2/D4/E1/E2
		WT1	1.97e-03	0.533	D1-A6/C3 D2-C3
		WT2	1.55e-02	0.471	*NA*

SI Table 4.10 continued.

Estimate		Bird	P-value	R^2 Adj	Different Locations
TCRβ	SD	WT5	5.62e-19	0.823	A2-A1/A6/A7/A8/A9/A10/A11/B1/B2/B3/B4/B5/B6/B7/B8/B9/B10/B11/C2/C3/C4/C5 /C6/C8/C9/C10/C11/D2/D3/D4/D7/D9 A3-A6/B2/B4/B11/C3/D5 A4/A5-A6/A7/A8/A10/A11/B1/B2/B3/B4/B5/B6/B7/B8/B9/B10/B11/C2/C3/C4/C5/C6 /C8/C9/C10/D2/D3/D4/D7/D9 B2-A1/A7/A9/A11/B3/B6/B9/B10/C1/C2/C4/C5/C9/C11/D2/D3/D5/D6/D7/D8/D9 C7-A6/B2/B4/B7/B11/C3/C8/D4 D8-A6/A7/A10/B1/B4/B5/B6/B7/B8/B11/C2/C3/C4/C5/C6/C8/C10/D2/D3/D4/D7 D1-B2/B4/D4 D5/D6-B4/B11/D4/D10/D11 D10-A1/A2/A3/A4/A5/A7/A9/A11/B3/B9/B10/C1/C2/C7/C9/C11/D1/D8/D9 D11-A1/A2/A3/A4/A5/A9/B10/C1/C7/C11/D1/D8
TCRβ	SE	Prelim	2.32e-22	0.961	A1-A2 A3-E4 A4-A1/A2/A3/B1/B2/B3/B4/C1/D1/D4 B2/B4-A1/B3/C1/D1/D4/E1/E2 C2-A1/A2/A3/B1/B2/B3/B4/C1/C3/C4/D1/D2/D4/E1/E2 C3/C4/D2-A1/A2/A3/B1/B2/B3/B4/C1/D1/D3/D4/E1/E2/E3/E4 D3-A1/A2/A3/B1/B3/C1/D1/D4/E1/E2 E3/E4-A1/B1/B3/C1/D1/D4/E1/E2
		WT2	2.89e-04	0.556	B1-A1/B5/B6/B7/B8/B10/C3/C4/C5/C6/C7/C8/D2/D4/D5/D6/D7/D10
		WT5	2.64e-15	0.780	A2-A1/A6/A7/A8/A9/A10/A11/B1/B3/B4/B5/B6/B7/B8/B9/B10/B11/C1/C2/C3/C4 /C5/C6/C8/C9/C10/C11/D2/D3/D4/D7/D9 A4-A6/A7/A8/A9/A10/A11/B1/B3/B4/B5/B6/B7/B8/B9/B10/B11/C2/C3/C4/C5/C6 /C8/C9/C10/D2/D3/D4/D7/D9 A5-A1/A6/A7/A8/A9/A10/A11/B1/B3/B4/B5/B6/B7/B9/B10/B11/C1/C2/C3/C4/C5 /C6/C8/C9/C10/C11/D2/D3/D4/D7/D9 B2-A2/A3/A4/C7/D8/D1/D5/D6 C7-A6/B2/B4/B7/B11/D4 D8-A7/A8/A10/A11/B1/B3/B4/B5/B6/B7/B8/B11/C2/C3/C4/C5/C6/C8/C10/D2 /D3/D4/D7/D9 D10-A2/A3/A4/A5/C7/D8/D5/D6 D11-A2/A4/A5/C7/D8/D6
BCR	SR	Prelim	9.05e-15	0.910	A1/D4-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/E1/E2/E3/E4 B3-A4/B2/C3/C4/D1/D2/D3/E2/E3/E4 E2-A2/A3/C1/E1
		tmWT4	1.04e-02	0.903	*NA*
	SD	Prelim	3.74e-14	0.903	A1-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4/E1/E2/E3/E4 B3-A4/C4/D2/D3/D4/E2/E3/E4 D4-A2/A3/A4/B1/B2/B4/C1/C2/C3/C4/D1/D2/D3/E1/E2/E3/E4 E2-A2/A3/C1
		tmWT4	2.41e-02	0.883	*NA*
	SE	Prelim	1.81e-11	0.864	A1/D4-A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/E1/E2/E3/E4 E2-A2/A3/B3/C1/C2

SI Table 4.11 – Birds where there is a significant effect of location on the proportional contribution of specified phyla. The effect of location was assessed using a linear model of the form lm(Prop~Location) within each bird, where 'Prop' refers to the proportional contribution of a specified phylum.

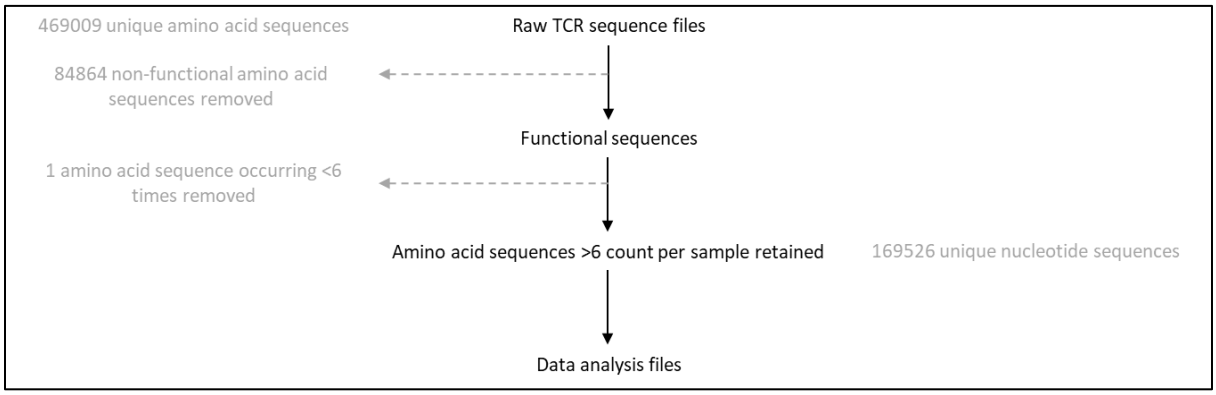
Phylum	Bird	P-value	R^2 Adj
Actinobacteria	HT5	2.45e-03	0.586
	WT1	6.91e-04	0.591
	WT2	7.47e-32	0.947
Cyanobacteria	HT5	1.38e-04	0.650
	WT1	2.06e-05	0.714
Firmicutes	KO1	1.27e-03	0.542
	WT1	2.28e-13	0.839
	WT2	2.17e-18	0.857
	WT5	9.20e-03	0.669
	Prelim	3.74e-02	0.481
Proteobacteria	HT3	2.00e-04	0.584
	HT5	4.03e-05	0.665
	KO1	2.74e-03	0.525
	KO2	1.80e-04	0.825
	WT1	1.01e-24	0.968
	WT2	7.25e-03	0.525
	WT5	2.24e-07	0.841
Unassigned	HT1	9.11e-03	0.852
	WT1	9.79e-13	0.947
	WT4	5.01e-05	0.981
	WT5	3.41e-04	0.972
	Prelim	1.31e-02	0.523

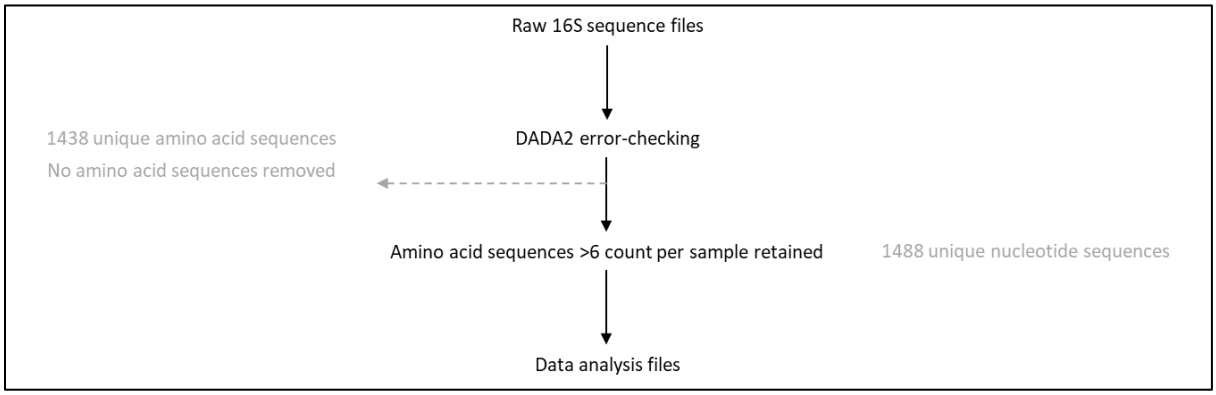
SI Table 4.12 – Birds where there is a significant effect of location on the proportional contribution of specified families. The effect of location was assessed using a linear model of the form lm(Prop~Location) within each bird, where 'Prop' refers to the proportional contribution of a specified family. All values are given to three significant figures.

Phylum	Family	Bird	P-value	R^2 Adj
Actinobacteria	Bifidobacteriaceae	HT5	2.88e-03	0.604
		WT1	6.44e-04	0.592
		WT2	3.81e-37	0.984
	Coriobacteriaceae	HT3	4.80e-02	0.965
		KO2	1.39e-02	0.802
		KO3	7.22e-04	0.913
	Micrococcaceae	WT1	4.18e-02	0.959
Elusimicrobia	Elusimicrobiaceae	HT3	7.47e-05	0.895
		KO2	8.34e-13	1.00
		KO3	1.84e-03	0.998
Firmicutes	Clostridiaceae	WT1	2.74e-03	0.823
		WT4	2.93e-02	0.621
		WT5	4.66e-02	0.660
		Prelim	1.61e-02	0.825
	Lachnospiraceae	KO1	5.14e-04	0.562
		WT1	1.06e-03	0.582
		WT2	5.36e-03	0.508
		WT5	1.87e-02	0.649
	Peptostreptococcaceae	WT1	6.49e-03	0.662
	Ruminococcaceae	KO1	2.44e-05	0.617
		WT1	4.90e-02	0.474
	Veillonellaceae	KO3	4.51e-02	0.450
		WT4	2.93e-03	0.850
		WT5	3.64e-02	0.655
Proteobacteria	Succinivibrionaceae	HT3	2.32e-05	0.685
		HT5	1.90e-03	0.903
Unassigned	Unassigned	HT1	3.65e-02	0.490
		KO1	1.58e-02	0.478
		KO2	7.48e-03	0.716
		WT1	2.76e-02	0.494
Other	Other	KO1	3.50e-02	0.073
		KO3	4.07e-03	0.138

SI Table 4.13 – The model fit of V gene segments and their associated J segments for models of the effect of bird (Model 1) and the effect of bird and location (Model 2). Values are given to three significant figures. 'NA' J-subset refers to the combined V-family usage regardless of the associated J segments.

V Subset	J Subset	Model	DF	AIC
Vb1	*NA*	Model 1	6	-1875.16
		Model 2	173	-1705.62
	Jb1	Model 1	6	-2025.75
		Model 2	173	-1871.56
	Jb2	Model 1	6	-2060.74
		Model 2	173	-1919.70
	Jb3	Model 1	6	-1873.00
		Model 2	173	-1853.72
	Jb4	Model 1	6	-1910.22
		Model 2	173	-1762.66
Vb2	*NA*	Model 1	6	-1874.98
		Model 2	173	-1705.29
	Jb1	Model 1	6	-2373.87
		Model 2	173	-2167.98
	Jb2	Model 1	6	-2585.13
		Model 2	173	-2470.19
	Jb3	Model 1	6	-2272.39
		Model 2	173	-2123.19
	Jb4	Model 1	6	-2445.99
		Model 2	173	-2281.18





SI Figure 5.1 – Schematic of trimming decisions for the microbiota dataset. Retaining amino acid sequences of >6 count was based upon error rates reported by Schirmer *et al*. (2015,2012)[285,288].

SI Table 5.1 – The significance of the association between the proportion of functional sequences or clones and bird, genotype or B-cell status in the IgA-producing B-cell receptor and T-cell receptor populations. The dataset used was either raw or trimmed (see B.5.3). The effect of treatment and bird was evaluated using a linear model of the form lm(Prop~Variable) where 'Prop' was the proportion of functional clones retained and Variable is the explanatory variable indicated. The effect of B-cell status was assessed using Welch's two-sample t-test. All values are given to three significant figures.

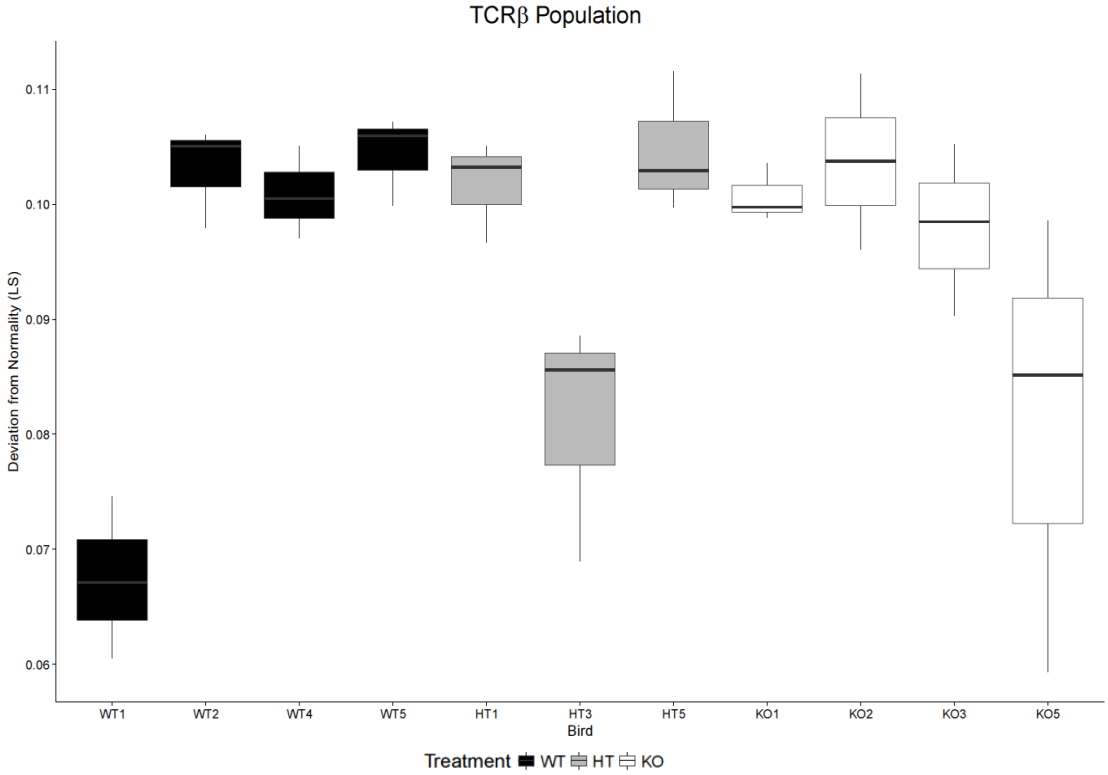
Functional Set	Population	Variable	P-value	R^2 Adjusted
Sequences	Raw BCR	Bird	< 2.20e-16	0.984
		Genotype	< 2.20e-16	0.982
		B-cell status	< 2.20e-16	*NA*
	Trimmed BCR	Bird	1.90e-08	0.844
		Genotype	8.49e-06	0.510
		B-cell status	9.67e-08	*NA*
	TCRβ	Bird	8.80e-10	0.884
		Genotype	0.937	-0.0621
		B-cell status	0.787	*NA*
Clones	Raw BCR	Bird	< 2.20e-16	0. 987
		Genotype	< 2.20e-16	0.986
		B-cell status	< 2.20e-16	*NA*
	Trimmed BCR	Bird	2.32e-08	0.841
		Genotype	1.16e-05	0.500
		B-cell status	1.61e-07	*NA*
	TCRβ	Bird	< 2.20e-16	0.995
		Genotype	0.897	-0.0590
		B-cell status	0.752	*NA*

SI Table 5.2 – The significance of the effect of bird upon the proportional contribution of different population structure abundance groups for both the microbiota and immune populations, with birds showing a significant differences from other birds indicated. The difference between birds was assessed using linear models of the form lm(Prop~Bird), where 'prop' is the proportional contribution of each abundance group. The exact birds which were different were assessed with Tukey HSD summary. Group 2 refers to the most abundant microbes and group 6 the rarest. BCR data presented is the raw dataset, as all the KO birds had no sequences these were significantly different from all other birds. All values are given to three significant figures.

Population	Group	P-value	R^2 Adjusted	Different Birds
Microbiota	2	2.10e-08	0.926	KO1-KO3/HT1/HT3/WT2/WT4 KO2-KO3/HT1/WT2/WT4 KO3-HT3/WT1/WT5 HT5-HT1/HT3/KO3/WT1/WT2/WT4/WT5 WT1-HT1/WT2/WT4 WT2-WT5/HT3 WT4-WT5
	3	9.72e-05	0.802	HT5-HT1/HT3/KO3/WT2/WT4/WT5 WT2-KO1 KO2-HT1/HT3/KO3/WT2/WT4/WT5
	4	2.62e-04	0.760	HT5-HT1/KO2/KO3/WT2/WT4 KO1-KO3/WT2 WT5-WT2
TCRβ	2	3.97e-04	0.716	WT1-HT1/KO2/KO3/WT2/WT4
	3	4.12e-05	0.775	KO5-HT1/HT5/KO3/WT2/WT4 WT1-HT1/HT5/KO3/WT2/WT4/WT5

SI Table 5.2 continued.

Population	Group	P-value	R^2 Adjusted	Different Birds
TCRβ	4	1.41e-04	0.745	HT3-HT1/KO2/WT2/WT4 KO1-HT1/KO2/WT2/WT4 WT1-KO2/WT4
	5	2.00e-04	0.736	WT1-HT1/HT5/KO1/KO2/KO3/WT2/WT4/WT5 KO5-HT1/WT4
	6	2.09e-04	0.734	WT1-HT1/HT5/KO1/KO2/KO3/WT2/WT4/WT5 KO5-HT1/WT4
BCR	2	1.37e-11	0.946	HT1-HT5/WT2/WT5 WT4-HT5/WT1/WT2/WT5
	3	1.56e-13	0.964	WT4-HT3/HT5/WT1
	4	1.91e-22	0.995	HT1-HT5/WT5
	5	3.24e-22	0.994	WT4-HT3/HT5/WT1/WT2/WT5 HT5-HT1
	6	3.18e-22	0.994	WT4-HT3/HT5/WT1/WT2/WT5 HT5-HT1



SI Figure 5.2 – The deviation from normality, as assessed by using the sums of least squares regression, of the TCRβ spectratype distributions. Points are coloured by the treatment type of the bird.

SI Table 5.3 – The significance of the variation in proportional contribution of different population structure abundance groups within the raw and trimmed B-cell receptor population based on either genotype of B-cell status of the birds. The difference between treatments was assessed using linear models of the form lm(Prop~Genotype) or lm(Prop~B-status), where 'prop' is the proportional contribution of each abundance group. Group 2 refers to the most abundant microbes and group 6 the rarest. All values are given to three significant figures.

Explanatory Factor	Group	Raw BCR		Trimmed BCR	
		P-value	R^2 Adjusted	P-value	R^2 Adjusted
Genotype	2	6.40e-10	0.756	9.69e-06	0.549
	3	1.29e-17	0.925	1.67e-13	0.859
	4	2.99e-30	0.989	3.48e-23	0.968
	5	3.31e-25	0.977	1.99e-21	0.958
	6	3.28e-25	0.977	6.76e-18	0.928
B-cell Status	2	5.23e-11	0.755	1.54e-06	0.542
	3	6.48e-19	0.924	9.91e-15	0.859
	4	1.03e-31	0.989	5.55e-23	0.959
	5	8.20e-27	0.977	5.90e-23	0.958
	6	8.11e-27	0.977	5.82e-19	0.925

SI Table 5.4 – The significance of the variation in diversity indices between birds in each microbiota and immune population. The difference between treatments was assessed using linear models of the form lm(Estimate~Bird) where 'Estimate is the diversity estimate in question. All values given to three significant figures.

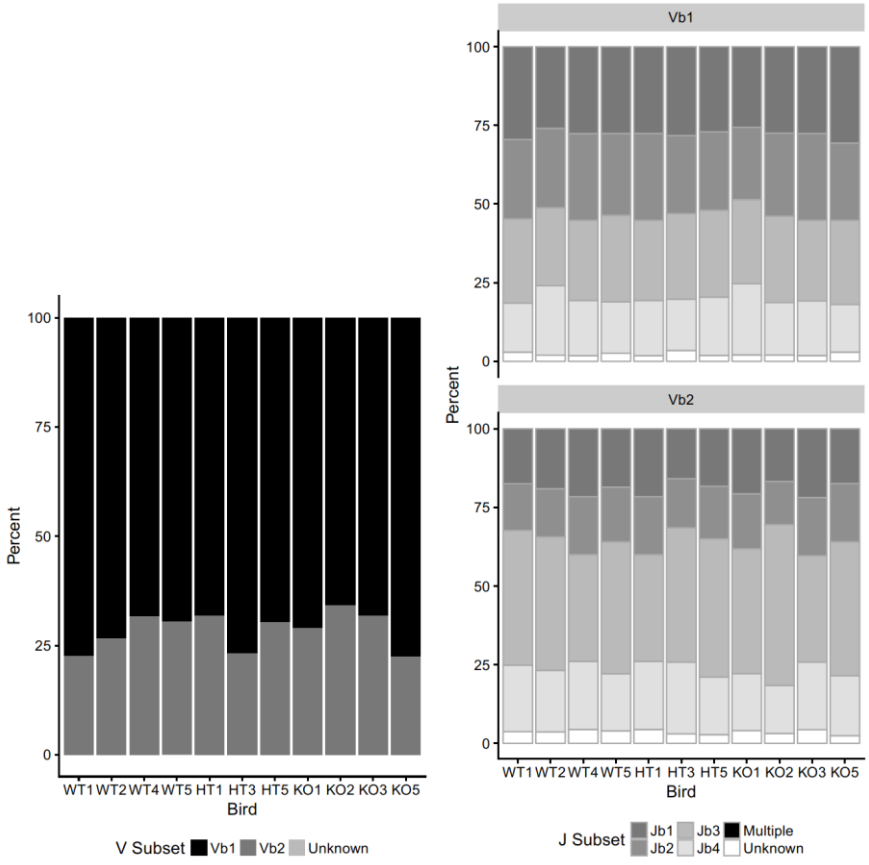
Diversity Estimate	P-value	R^2 Adjusted	Birds Different
Species richness	3.12e-05	0.780	WT1-WT5/HT5/KO2 WT5-WT4/HT1/HT3/KO3/KO5 HT5-HT1/HT3/KO3.KO5 KO2-KO3/KO5
Shannon's Diversity	2.39e-08	0.891	WT5-WT1/WT2/WT4/HT1/HT3/KO1/KO2/KO3/KO5 HT5-WT1/WT2/WT4/HT1/HT3/KO2/KO3/KO5 KO1-WT4/HT1/KO3/KO5
Simpson's Entropy	1.21e-09	0.918	WT5-WT1/WT2/WT4/HT1/HT3/HT5/KO1/KO2/KO3/KO5 HT5-WT4/HT1/HT3/KO2/KO3/KO5 KO1-WT4/KO2/KO5 KO2-WT1

SI Table 5.5 – The significance of microbe phyla in which bird effects the phylum's proportional contribution to the community. Birds that are significantly different to other birds are shown ('Different Birds') with the focal bird separated from birds it is distinct from (e.g. KO1 is significantly different to HT5 and WT2 in terms of the proportion of Euryachaeota found in its microbitoa). Birds different to more than four other birds are highlighted in grey. 'WT' refers to the wild-type birds investigated, 'HT' to the heterozygotes and 'KO' to the knock-out birds. P-values and R^2 adjusted values are derived from a linear model of the form lm(Prop~Bird) where Prop is the proportional contribution of a specified phylum. The significantly different birds were identified using Tukey's HSD test. Values are given to three significant figures.

Phylum	P-value	R^2 Adjusted	Different Birds
Actinobacteria	3.22e-10	0.954	*NA*
Bacteroidetes	2.25e-07	0.914	*NA*
Cyanobacteria	1.73e-09	0.944	WT2-WT1/WT4/WT5
Euryarchaeota	1.26e-03	0.931	KO1-HT5/WT2
Proteobacteria	2.36e-09	0.942	HT3-HT1/KO1/KO2/KO3/WT1/WT2/WT4 HT5-HT1/KO2/KO3/WT1/WT4/WT5 KO1-HT1/WT4 WT5-KO1/KO2/KO3/WT1/WT2/WT4 WT2-HT1/WT4
Unassigned	3.80e-04	0.766	KO1-HT1/HT3/WT4/WT5 KO2-HT1/HT3/WT2/WT4/WT5

SI Table 5.6 – Families that show a significant effect of bird upon the family's proportional contribution to the community with birds driving the difference shown. 'WT' refers to the wild-type birds investigated, 'HT' to the heterozygotes and 'KO' to the knock-out birds. Birds that are significantly different to more than four other birds are highlighted in grey. Corresponding plots are shown in Figure 7. P-values and R^2 adjusted values are derived from a linear model of the form lm(Prop~Bird) where Prop is the proportional contribution of a specified family. The significantly different birds were identified using Tukey's HSD test. Values are given to three significant figures.

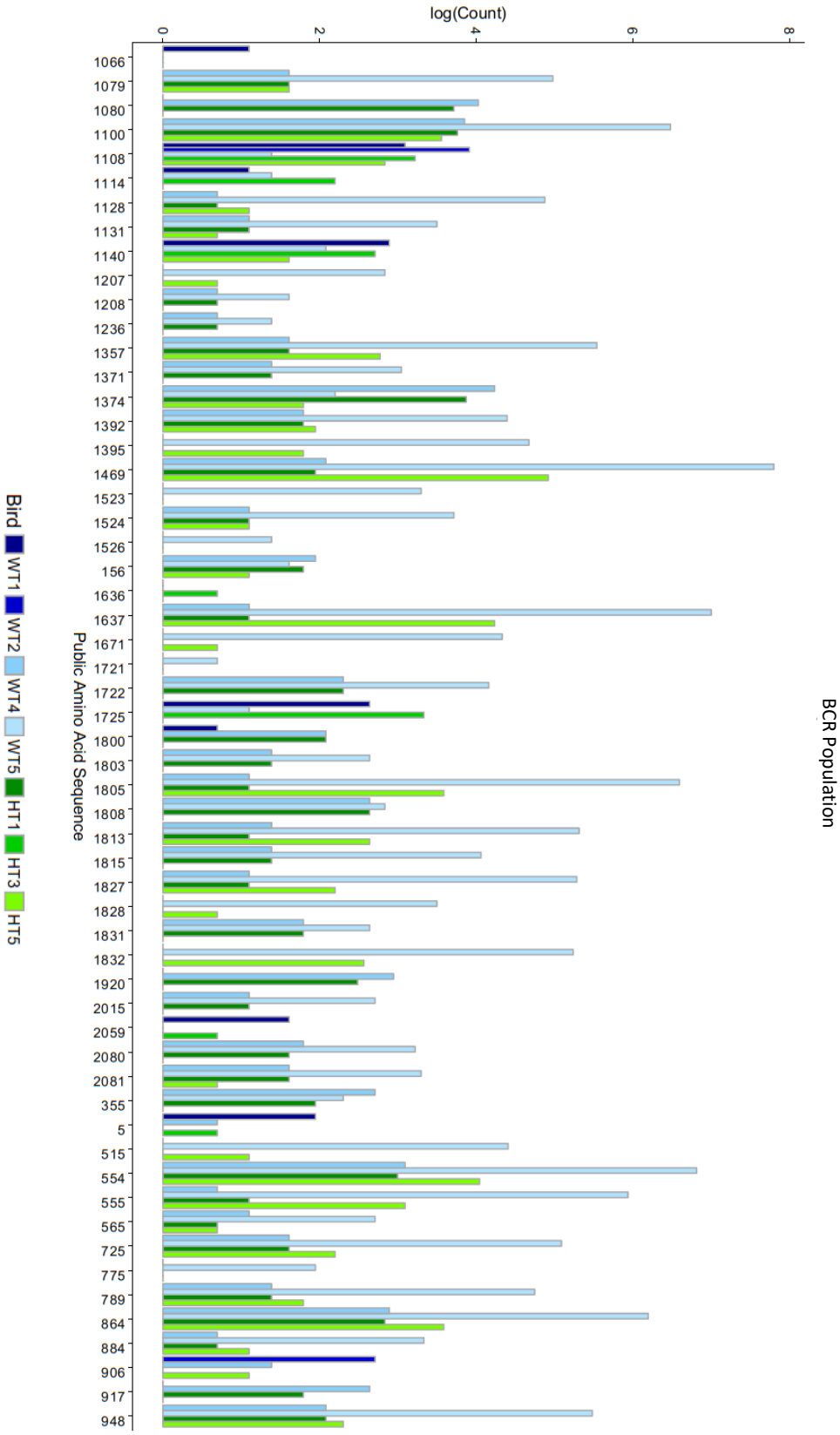
Phylum	Family	P-value	R^2 Adjusted	Different Birds
Actinobacteria	Bifidobacteriaceae	5.54e-16	0.990	*NA*
Elusimicrobia	Elusimicrobiaceae	1.60e-16	1.00	KO2-HT1/HT5/WT1/WT2/WT5
Firmicutes	Christensenellaceae	7.42e-11	0.973	*NA*
	Clostridiaceae	2.68e-07	0.901	HT1-WT5 HT3-HT5/WT2 KO1-HT1/HT5/KO3/WT2 WT1-HT1/HT3/HT5/KO2/KO3/WT2/WT4/WT5
	Lachnospiraceae	2.44e-08	0.915	HT5-HT1/HT3/KO1/KO2/KO3/WT2/WT4/WT5 HT3-KO2/WT2 KO1-KO2/KO3/WT2/WT5 WT1-HT1/KO2/KO3/WT2/WT4/WT5
	Mogibacteriaceae	2.94e-02	0.999	*NA*
	Ruminococcaceae	3.19e-14	0.984	HT5-HT1/HT3/KO1/KO2/KO3/WT1/WT2/WT4/WT5 KO2-KO1/WT1 WT5-HT1/HT3/KO1/KO2/KO3/WT1/WT2/WT4 WT2-HT1/KO2/WT1/WT4/WT5
Unassigned	Other	8.56e-03	0.650	WT2-HT1/HT3/HT5/KO2/WT1/WT4
	Unassigned	9.31e-08	0.912	HT3-KO1 HT5-HT1/HT3/KO1/KO3/WT1/WT2/WT4/WT5 WT2-HT1/KO1/WT4/WT5 KO2-HT1/HT3/KO1/KO3/WT1/WT4/WT5



SI Figure 5.3 – TCRβ gene segment usage in the first technical replicate of each of the JH birds. V family segment usage (left) varies between the JH birds. J segment usage, by associated V gene segments, varies in some cases and not others (right).

SI Table 5.7 – TCR nucleotide sequences, and their counts in each bird, that are public to the B-cell sufficient or all birds. Nucleotide sequences present in all birds are highlighted in grey.

Nucleotide Sequence	HT1	HT3	HT5	WT1	WT2	WT4	WT5	KO1	KO2	KO3	KO5
acaccactgaacttt	6	15	3	10	1	6	20	0	6	6	33
tgc	1	2	4	1	4	1	9	0	1	0	1
tgcgctaaa	2	1	23	2	3	2	2	8	0	2	1
tgcgctaagcaagacagggacgtaagactgatcttt	1	5	15	1	1	1	1	1	0	1	1
tgcgctaagcaagatcgaatcgggacaggggtaatttc	1	1	1	1	9	1	2	2	4	1	2
tgcgctaagcaagggacaggggatacagtaacatgattttc	10	1	8	1	1	10	14	0	51	9	1
tgcgctaagcaaggggacacaccactgaacttt	1	12	2	1	1	1	2	0	1	1	1
tgtgccagc	2	4	1	4	4	2	3	1	0	2	1
tgtgccagcagacagtacagtaacatgattttc	1	16	42	1	1	1	1	1	1	1	7
tgtgccagcaggggatcgttaatatccagtatttt	1	5	8	1	2	1	3	0	9	1	1
tgtgccagcagtttg	3	9	4	1	1	3	3	2	1	3	20



SI Figure 5.4 – BCR amino acid sequences counts for sequences present in at least two birds of each B-cell sufficient genotype.

SI Table 5.8 –The significance of the effect of location upon the proportional contribution of different population abundance groups to the microbiota community. Displayed are only those instances in which there was a significant effect of location. The segments involved in driving the effect of location upon proportional contribution are indicated, with segments that are different than more than four other segments highlighted in grey. The effect of location was assessed using a linear model of the form lm(Prop~Location) within each bird, where 'Prop' is the proportional contribution of a specified abundance group. Segments that were significantly different from others were identified using Tukey's HSD test. Values are given to three significant figures.

Bird	Group	P-value	R^2 Adjusted	Different Birds
HT5	2	2.10e-02	0.529	*NA*
KO1	4	7.00e-03	0.501	A4-A5/C8; B6-A5/C8
KO2	4	2.39e-02	0.680	A8-D1
WT2	5	2.50e-07	0.672	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/B1/B2/B3/B4/B5/B7/B9/B10 /C1/C2/C3/C5/C7/C8/C9/C10/D1/D2/D3/D4/D5/D6/D7/D8/D9/D10
WT4	6	1.17e-02	0.590	D5-A9/B2/B7/C1/D1/D9
	5	2.74e-03	0.629	D5-A1/A2/A3/A5/A6/A8/A9/B2/B3/B5/B6/B7/B8/C1/C2/C3/C4/C5 /C6/C8/D1/D3/D4/D7/D9
	4	2.08e-02	0.573	D5-A1/A3/B2/B3/B4/B5/B6/C1/C3/C4/C8/D1/D2/D3/D4
	3	4.27e-02	0.548	D5-A9/B7/B8/C1/D9

SI Table 5.9 – Model fit summaries for each population structure abundance group. Each model was of the form lm(Prop~X), where 'Prop' is the proportional contribution of a specified population abundance group. X varied between models, as indicated. Model 1 [lm(Prop~Bird)] results were always equivalent to Model 4 [lm(Prop~Bird+Treatment)], and Model 2 [lm(Prop~Bird+FullLocation)] equivalent to Models 3 [lm(Prop~Bird+FullLocation+Treatment)] and 5 [lm(Prop~FullLocation)], where multiple models produced the same answers only the simplest model is displayed.

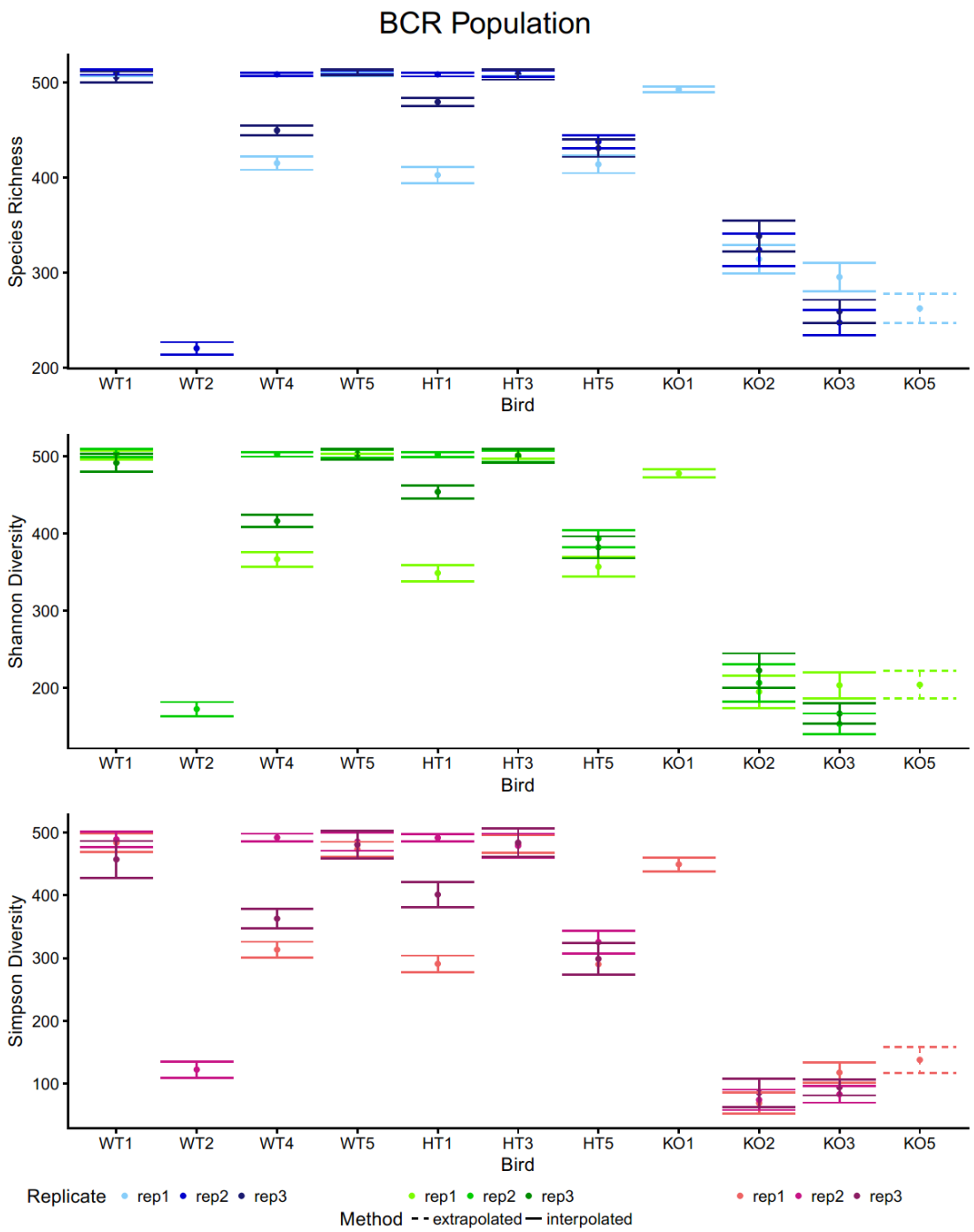
			Microbiota		**TCRβ**		**Raw BCR**		**Trimmed BCR**	
Group	**Model**	**X**	**DF**	**AIC**	**DF**	**AIC**	**DF**	**AIC**	**DF**	**AIC**
2	Model 1	Bird	11	-3611.054	12	-2601.401	12	-3422.0284	12	-2055.5141
	Model 5	Full Location	343	-3385.270	381	-2570.963	367	-3149.2151	275	-1945.4362
	Model 6	Treatment	4	-2693.300	4	-2320.447	4	-3336.1842	4	-1941.8529
	Model 7	B-cell Status	3	-2609.239	3	-2321.299	3	-3338.0353	3	-1942.7835
3	Model 1	Bird	11	-3029.187	12	-2794.783	12	-4306.6419	12	-2038.1059
	Model 5	Full Location	345	-2726.057	381	-2768.157	367	-3914.5471	333	-1932.1529
	Model 6	Treatment	4	-2840.204	4	-2204.857	4	-4161.5079	4	-1899.0193
	Model 7	B-cell Status	3	-2835.510	3	-2199.613	3	-4163.3770	3	-1901.0153
4	Model 1	Bird	11	-4269.826	12	-3598.784	12	-4738.4638	12	-3246.3712
	Model 5	Full Location	345	-4092.041	381	-3594.742	368	-4487.3534	337	-3206.1613
	Model 6	Treatment	4	-3858.855	4	-3476.364	4	-4649.4145	4	-3132.1027
	Model 7	B-cell Status	3	-3808.935	3	-3477.616	3	-4651.4071	3	-3130.3647
5	Model 1	Bird	11	-1639.428	12	-2856.991	12	-1053.6444	12	-953.6310
	Model 5	Full Location	346	-1355.798	381	-2697.524	377	-831.7091	357	-868.7106
	Model 6	Treatment	4	-1576.969	4	-2037.298	4	-1059.4575	4	-697.8120
	Model 7	B-cell Status	3	-1575.995	3	-2028.465	3	-1060.6152	3	-696.2015
6	Model 1	Bird	11	-4651.695	12	-5740.828	12	-5772.1043	12	-3382.8622
	Model 5	Full Location	345	-4281.551	381	-5626.638	367	-5410.7151	320	-3253.2516
	Model 6	Treatment	4	-4515.228	4	-4975.300	4	-5670.4349	4	-3118.7534
	Model 7	B-cell Status	3	-4495.369	3	-4967.867	3	-5671.4320	3	-3120.3453

SI Table 5.10 – The difference between the mean proportional contribution of different population abundance groups within the BCR dataset between the knock-out birds (KO) and the B-cell sufficient (BP) birds. The difference between the B-cell status groups was assessed using Welch's two sample t-test for each abundance group. Values are given to three significant figures.

	Raw BCR Population			Trimmed BCR Population		
Group	P-value	KO	BP	P-value	KO	BP
2	< 2.2e-16	0.000	0.0443	< 2.2e-16	0.0944	0.119
3	< 2.2e-16	0.000	0.114	2.18e-02	0.532	0.492
4	< 2.2e-16	0.000	0.239	5.34e-02	0.240	0.245
5	< 2.2e-16	0.000	0.538	< 2.2e-16	0.202	0.132
6	< 2.2e-16	0.000	0.128	< 2.2e-16	0.112	0.0427

SI Table 5.11 – The significance of the effect of bird upon the diversity estimates within the IgA-producing B-cell receptor (BCR) populations, both trimmed and raw. The effect of bird was assessed using a linear model of the form lm(Estimate~Bird), where 'Estimate' is the specific diversity index used. The birds that are significantly different from each other, as assessed using Tukey's HSD test, are indicated. Birds that are different from more than four other birds are highlighted in grey. All values are given to three significant figures.

Population	Estimate	P-value	R^2 Adjusted	Different Birds
Trimmed BCR	Species Richness	1.03e-08	0.953	WT2-WT1/WT4/WT5/HT1/HT3/HT5/KO1 WT1/WT4/WT5/HT1/HT3/HT5/KO1-KO2/KO3/KO5
	Shannon's Diversity	2.07e-08	0.948	WT1-WT2/HT5 WT2-WT1/WT4/WT5/HT1/HT3/HT5/KO1 WT1/WT4/WT5/HT1/HT3/HT5/KO1-KO2/KO3/KO5 HT5-WT5/HT3
	Simpson's Entropy	3.47e-08	0.945	WT1-WT2/HT5 WT2-WT1/WT4/WT5/HT1/HT3/KO1 WT1/WT4/WT5/HT1/HT3/KO1-KO2/KO3/KO5 HT5-WT5/HT3/KO2/KO3



SI Figure 5.5 – The mean diversity estimates of individual segments within the trimmed BCR population of each bird. Species richness (blue), Shannon's diversity index (green) and Simpson's diversity index (red) of each technical replicate are displayed for each bird. 'WT' refers to the wild-type birds, 'HT' to the heterozygotes and 'KO' to the knock-out birds. Estimates were calculated to the same end point either through interpolation or extrapolation. The standard error of the estimates (95% confidence interval) is shown.

SI Table 5.12 – Birds where there is a significant effect of location on the proportional contribution of specified phyla, with segments that are significantly different from each other indicated. The effect of location was assessed using a linear model of the form lm(Prop~Location) within each bird, where 'Prop' refers to the proportional contribution of a specified phylum. Segments that are different from more than four other segments, as assessed using Tukey's HSD test, are highlighted in grey. All values are given to three significant figures.

Phylum	Bird	P-value	R^2 Adj	Different Segments
Actinobacteria	HT5	2.45e-03	0.586	C7-A5/A6/B3/C4/D3 D6-D3
	WT1	6.91e-04	0.591	C7-A2/A3/A4/C5/C6/D2/D3 C6-C1/D6
	WT2	7.47e-32	0.947	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/B1/B2/B3/B4/B5/B7 /B9/C1/C2/C3/C5/C7/C8/C9/D1/D2/D3/D4/D5/D6 /D7/D8/D9
Cyanobacteria	HT5	1.38e-04	0.650	A2-B4/B5/C5/C6/C8/C9/D1/D4 D4-A3/A4/D5 A4-C9
	KO2	1.39e-02	0.802	*NA*
	WT1	2.06e-05	0.714	A8-A1/A2/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6/B7 /B8/C1/C2/C3/C4/C5/C6/C7/D1/D2/D3/D4/D5/D6 /D7/D8
Firmicutes	KO1	1.27e-03	0.542	B3-A4/B5/C1/D3 C3-A4/C1/D3
	WT1	2.28e-13	0.839	A8-A1/A2/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6/B7 /B8/C1/C2/C3/C4/C5/C6/C7/C8/D1/D2/D3/D4 /D5/D6/D7/D8 A1-C6/D3 C7-A4/B2/B8/C2/C3/C5/C6/D2/D3 C6-A5/B5/B6/B7/C1/D5/D6
	WT2	2.17e-18	0.857	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/B1/B2/B3/B4/B5 /B7/B9/C1/C2/C3/C5/C7/C8/C9/D1/D2/D3/D4 /D5/D6/D7/D8/D9
	WT5	9.20e-03	0.669	A2-B9/B10/D10 D10-B8
Proteobacteria	HT3	2.00e-04	0.584	A1/B5-B8/D7/B9 B8-A2/A3/A4/A5/A7/A9/B1/B2/B3/B4/B5/B6/B7/C1 /C2/C3/C4/C6/D1/D2/D3/D4/D5/D6/D8/D9 D7/B9-C2/D4
	HT5	4.03e-05	0.665	D3-A1/B1/C1/C2/C3/C4/C5/C9/D1/D4/D5 D6-A1/A2/A3/A4/A5/A7/A8/B1/B2/B3/B4/B5/B6/B7 /B9/C1/C2/C3/C4/C5/C6/C7/C8/C9/D1/D2/D4 /D5/D7/D8/D9
	KO1	2.74e-03	0.525	B3-A4/B1/B6/C1/D3
	KO2	1.80e-04	0.825	B9-A2/A3/A4/A5/A6/A7/A8/A9/B1/B2/B5/B6/B7/B8 /C1/C2/C3/C4/C5/C6/C7/C9/D1/D2/D3/D4 /D5/D6/D7/D8/D9
	WT1	1.01e-24	0.968	A8-A1/A2/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6 /B7/B8/C1/C2/C3/C4/C5/C6/C7/D1/D2/D3 /D4/D5/D6/D7/D8 D5-C5/D3
	WT2	7.25e-03	0.525	C7-B5
	WT5	2.24e-07	0.841	A2-A1/A3/A4/A5/A6/A7/A8/A9/A10/A11/B1/B2/B3 /B4/B5/B6/B7/B8/B9/B10/C1/C2/C3/C4/C5/C6 /C7/C8/C9/C10/C11/D1/D2/D3/D4/D5/D6/D7/D8 /D9/D10/D11 D8-A4/B3/B8 B2-A4/B8

SI Table 5.12 continued.

Unassigned	HT1	9.11e-03	0.852	A4-A1/A5/A6/C2/C3/C6/C8/D3/D4/D6/D7/D8
	WT1	9.79e-13	0.947	A8-A1/A2/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6 /B7/B8/C1/C2/C3/C4/C5/C6/C7/D1/D2/D3/D4 /D5/D6/D7/D8
	WT4	5.01e-05	0.982	A7-A1/A3/A6/A9/B1/B2/B5/B6/B9/C1/C3/C4/C6 /D1/D3/D4/D6/D9 B2-D1/C3
	WT5	3.41e-04	0.972	B2-A1/A2/A4/A5/A6/A7/A9/A10/A11/B1/B5/B6/B7 /B8/B11/C2/C7/C8/C9/C10/D2/D4/D5/D6/D7 B3-A2/A4/A5/A6/A9/A10/B5/B6/B11/C2/C7/C10 /D2/D6/D7 C7-A8 D9-A2/A4/A5/A6/A10/B5/B6/B11/C2/C7/C10/D2 /D6/D7

SI Table 5.13 – Birds where there is a significant effect of location on the proportional contribution of specified families, with segments that are significantly different from each other indicated. The effect of location was assessed using a linear model of the form lm(Prop~Location) within each bird, where 'Prop' refers to the proportional contribution of a specified family. Segments that are different from more than four other segments, as assessed using Tukey's HSD test, are highlighted in grey. All values are given to three significant figures.

Family	Bird	P-value	R^2 Adjusted	Different Segments
Bifidobacteriaceae	HT5	2.88e-03	0.604	*NA*
	WT1	6.44e-04	0.592	C7-A2/C5/C6/C8/D1/D2/D3 A8-C6/C8/D3 C6-A5
	WT2	3.81e-37	0.984	B8-A1/A2/A3/A4/A5/A6/A7/A8/A9/A10/B1/B2/B3/B4 /B5/B7/B9/B10/C1/C2/C3/C5/C7/C8/C9/C10/D1/D2 /D3/D4/D5/D6/D7/D8/D9/D10
Clostridiaceae	WT1	2.74e-03	0.823	A8-A1/A3/A4/A5/A6/A7/B1/B2/B3/B4/B5/B6/B7/B8 /C1/C2/C3/C4/C5/C6/D1/D2/D3/D4/D5/D6/D7/D8
	WT4	2.93e-02	0.621	D8-C3
	WT5	4.66e-02	0.660	D10-A3/A6/A9/B3/B7/B9/C3/C6/C7/D1/D2/D4/D8/D11
Coriobacteriaceae	HT3	4.80e-02	0.965	*NA*
	WT1	7.22e-04	0.913	C8-A1/A2/A5/A7/B1/B2/B3/B4/B5/B6/B7/B8/C1/C4/C5 /C6/C7/D1/D3/D5/D6/D7
Elusimicrobiaceae	HT5	7.47e-05	0.895	B6-A1/A2/A3/A4/A5/A7/A8/A9/B1/B2/B3/B5/B8/C2/C3 /C4/C5/C7/C8/C9/D1/D2/D4/D6/D7/D8/D9 D3-A1/A2/A3/A4/A5/A6/A7/A8/A9/B1/B2/B3/B4/B5 /B7/B8/B9/C2/C3/C4/C5/C7/C8/C9/D1/D2/D4/D5 /D6/D7/D8/D9
	KO2	8.34e-13	1.000	D1-A3/A4/A8/A9/B2/B5/B6/B9/C2/C4/C5/C7/C9/D2 /D3/D6/D7/D9
	KO3	1.84e-03	0.998	A5-A2/A3/A6/B1/B3/B4/B7/C1/C4/C5/C6/D2/D4/D5 B5-A2/A3/A6/B1/B3/B4/C1/C4/C5/C6/D2/D4/D5 B7-A2/A6/B4/C4/C5/C7/D3/D5 C3-A2/A6/B3/B4/C4/C5/C6/D5 C7/D3-A2/A3/A6/B1/B3/B4/ C1/C4/C5/C6/D2/D4/D5
Lachnospiraceae	KO1	5.14e-04	0.562	A4-A2/A7/B8/C8/D1/D2/D8 D3-A7/D1/D2/D8 D2-C1
	WT1	1.06e-03	0.582	C8-A1/A2/A3/A4/A5/A6/A7/A8/B1/B2/B3/B4/B6/B7 /B8/C1/C2/C3/C4/C6/D1/D2/D3/D4/D5/D6/D7/D8
	WT2	5.36e-03	0.508	B8-A1/A3/A7/A9/B1/B2/B3/B5/B10/C1/C2/C3/C8 /D5/D9/D10 B10-D7
	WT5	1.87e-02	0.649	*NA*

SI Table 5.13 continued.

Micrococcaceae	WT1	4.18e-02	0.959	C5-A1/B5/C1/D6
Peptostreptococcaceae	WT1	6.49e-03	0.662	C6-B3/B6/D3/D7
Ruminococcaceae	KO1	2.44e-05	0.617	A4/C1-A7/B4/B7/D2/D4 B6-A5/A7/B3/B4/B5/B7/C6/C8/D2/D4/D5
	WT1	4.90e-02	0.474	*NA*
Succinivibrionaceae	HT3	2.32e-05	0.685	B8-A1/A2/A3/A4/A9/B3/B4/B5/B7/C1/C2/C3/C4/C6 /D2/D3/D4/D5/D6/D9 D7-A1/A2/A3/A4/A5/A6/A9/B2/B3/B4/B5/B6/C1/C2 /C3/C4/C6/C9/D2/D3/D4/D5/D6/D9 B9-A1/A2/A4/A9/B3/B4/B5/B7/B9/C2/C3/C4/D2/D3 /D4/D9
	HT5	1.90e-03	0.903	D6-A2/A3/A4/A6/A8/A9/B2/B3/B5/B8/B9/C1/C4/C6 /C7/D1/D2/D3/D7
Veillonellaceae	WT2	4.51e-02	0.450	*NA*
	WT4	2.93e-03	0.850	D5-A1/A2/A3/A4/A5/A6/A7/A8/B1/B2/B3/B4/B5/B6 /B9/C1/C2/C3/C4/C5/C6/C8/C9/D1/D3/D6/D7/D9
	WT5	3.64e-02	0.655	B2-A1/A7/B3/B4/B5/C7/D3/D9/C11
Other	KO1	3.50e-02	0.0733	C1-D7
	KO3	4.07e-03	0.138	A4-A2/C3/D4/D7
Unassigned	HT1	3.65e-02	0.490	D1-*All*
	KO1	1.58e-02	0.478	A2-D3/D5
	KO2	7.48e-03	0.716	B9-A1/A2/A3/A4/A6/A7/A9/B1/B5/B6/B8/C1/C2/C3 /C4/C5/C6/D1/D2/D4/D5/D6/D7/D8/D9
	WT1	2.76e-02	0.494	C1-A5/C7/C8

SI Table 5.14 – Significant associations between dissimilarity and distance within the microbiota and immune populations. The model used was of the form lm(Dissimilarity~Distance), within each bird. All values are given to three significant figures.

Population	Bird	Dissimilarity Index	P-value	R^2 Adjusted
Microbiota	HT1	Chi Squared	1.66e-02	0.00362
		Hellinger	3.23e-02	0.00274
		Kulczynski	9.23e-03	0.00442
	HT5	Chi Squared	3.73e-10	0.0248
		Hellinger	3.86e-10	0.0248
		Kulczynski	1.32e-08	0.0203
	KO3	Kulczynski	4.78e-02	0.00299
TCRβ	HT1	Chi Squared	7.07e-08	0.0199
		Hellinger	8.29e-07	0.0166
		Kulczynski	5.22e-08	0.0203
	HT5	Chi Squared	1.20e-77	0.168
		Hellinger	1.23e-76	0.166
		Kulczynski	9.54e-78	0.168
	KO1	Chi Squared	3.46e-02	0.00234
		Hellinger	4.76e-02	0.00201
		Kulczynski	2.77e-02	0.00264
	KO2	Chi Squared	1.51e-04	0.00706
		Hellinger	1.90e-05	0.00912
		Kulczynski	4.92e-04	0.00589
	KO5	Hellinger	4.95e-02	0.00253
	WT2	Chi Squared	3.67e-02	0.00149
		Kulczynski	1.94e-02	0.00197
	WT5	Chi Squared	3.96e-02	0.00171
		Hellinger	2.67e-02	0.00207

SI Table 5.14 continued.

BCR	HT1	Chi Squared	1.66e-02	0.00362
		Hellinger	3.23e-02	0.00274
		Kulczynski	9.23e-03	0.00442
	HT5	Chi Squared	3.73e-10	0.0248
		Hellinger	3.86e-10	0.0248
		Kulczynski	1.33e-08	0.0203
	KO3	Kulczynski	4.78e-02	0.00299

SI Table 5.15 – The effect of bird, genotype and B-cell status upon the association between segment dissimilarity and distance. All values are given to two decimal places.

			Microbiota		TCRβ		BCR			
							Raw		Trimmed	
Dissimilarity	Model	X	DF	AIC	DF	AIC	DF	AIC	DF	AIC
Chi Squared	Model 1	*NA*	3	-6634.55	3	-56901.07	3	-61235.81	3	-27096.69
	Model 2	Bird	12	-10681.02	13	-62710.90	13	-61865.78	13	-30107.51
	Model 6	Genotype	5	-6781.09	5	-57899.00	5	-61600.20	5	-28558.54
	Model 7	B-cell Status	4	-6716.73	4	-57632.26	4	-61589.32	4	-28552.69
Hellinger	Model 1	*NA*	3	-5766.23	3	-52573.03	3	-59981.01	3	-25271.52
	Model 2	Bird	12	-9739.06	13	-58507.15	13	-60921.60	13	-28909.91
	Model 6	Genotype	5	-5902.04	5	-53423.61	5	-60509.65	5	-26986.83
	Model 7	B-cell Status	4	-5847.26	4	-53226.63	4	-60489.38	4	-26975.67
Kulczynski	Model 1	*NA*	3	-12006.39	3	-55282.56	3	-68782.82	3	-28349.33
	Model 2	Bird	12	-16040.62	13	-61034.09	13	-69770.85	13	-32016.51
	Model 6	Genotype	5	-12167.39	5	-56343.12	5	-69428.28	5	-30128.09
	Model 7	B-cell Status	4	-12107.25	4	-56065.42	4	-69397.37	4	-30118.75

Appendices

Appendix 1 – Sequence Processing

Quality filtering was based on barcode identity. All barcodes used have a hamming distance of 3 and the quality of the barcode was the cumulative quality score of each of the 7 base pairs (bp) within the barcode. If the barcode was perfect (all bases identified together as a true barcode) and the cumulative quality score of the barcode was above 95% then the sequence associated with the barcode was kept. If the cumulative quality score was 99% or higher, and the barcode had a hamming distance of 1 from a true barcode then the barcode was corrected and the sequence kept. All other sequences were discarded.

All libraries within each MiSeq run were pooled and all potential edges between every sequence calculated. Sequences which were different by a single bp were clustered into chains of 1 bp mutations. When no other sequences could be added to this group a cluster was formed. The clustering was performed by Kenneth Hoen.

Appendix 2 - Sequences retained to abundance tables

QIIME retained more sequences to the end abundance tables, using singular reads, than DADA2 (Appendix Table 2.1). Conversely DADA2 was able to retain more merged sequences than QIIME, as the sequences were trimmed and corrected to ISVs before attempting the merge. Chimera removal, as expected, universally reduced the number of sequences retained though a higher proportion of those retained could be assigned taxonomy.

More sequences could be taxonomically assigned using QIIME, possibly as the fragment length compared to the reference database was longer or a result of using the most abundant sequence within each OTU as a reference sequence.

Appendix Table 2.1 – Comparison of the number of sequences processed to an abundance table using either QIIME or DADA2. Read refers to the way in which sequences were read into the pipelines. Assigned taxonomic units (ATU) were designated with reference to the Greengenes database, using either UCLUST or the RDP classifier in QIIME and DADA2 respectively. QIIME sequences were grouped into OTUs, and DADA2 sequences into ISVs.

Pipeline	Read	Chimera	Total Sequences Produced	
			OTU/ISV	Assigned ATU
QIIME	Forward Only	Present	261119	254881
		Removed	244878	241537
	Reverse Only	Present	260222	259496
		Removed	262199	261443
	Merged	Present	5887	5819
		Removed	5825	5768
DADA2	Forward Only	Present	259921	29309
		Removed	207815	26045
	Reverse Only	Present	259921	29127
		Removed	210737	25917
	Merged	Present	168177	160108
		Removed	118250	112079

Appendix 3 - Repeatability across technical replicates

Consistency across technical replicates refers to both identification of a specific sequence within a sample and its relative abundance. All QIIME pathways perform have a reduced mean variance compared with the DADA2 single read pathways in terms of whether a sequence is present in a given replicate (Appendix Table 2.2). This increased consistency in QIIME is expected as an intrinsic property of the clustering effects of OTU formation, as is the

decreased consistency with respect to counts. Pipeline processing using single reads or merged reads makes no difference in the repeatability of a sequences' presence across technical replicates, though the default setting of DADA2 merging the reads substantially increases the consistency of independent sequence counts across the technical replicates. The merged sequences analysed via QIIME are similarly more consistent with respect to count, though that is likely to result from smaller starting sequence numbers.

Appendix Table 2.2 – Repeatability of the QIIME and DADA2 assignments to operational taxonomic units (OTUs) or independent sequence variants (ISVs) respectively across the three technical replicates, both with (left) and without singletons (right). 'Mean' refers to the mean of the variance across the technical replicates; 'standard error 'is the error of this mean variance. 'Count' is the calculations of variance across technical replicates with respect to the number of sequences assigned to a specific type of group (either OTU or ISV in QIIME and DADA2 respectively), 'presence' is these same variance calculations for only presence/absence binary data.

Pipeline	Read	Chimera Removal	Count		Presence		Percentage Singletons	Count		Presence	
			Mean	Standard Error	Mean	Standard Error		Mean	Standard Error	Mean	Standard Error
QIIME	Forward Only	Present	26.923	4.021	0.065	0.000399	8.96602698	32.823910	4.9320655	0.010	0.000192
		Removed	19.026	2.120	0.061	0.000369	7.61154534	22.133717	2.4842823	0.011	0.000189
	Reverse Only	Present	82.831	15.459	0.054	0.000497	5.18019230	100.010430	18.7018495	0.009	0.000233
		Removed	57.488	11.151	0.061	0.000419	6.57058189	68.538443	13.3221270	0.011	0.000212
	Merged	Present	11.691	4.408	0.069	0.003490	5.70749108	14.334433	5.4535705	0.012	0.001772
		Removed	12.900	4.833	0.068	0.003721	5.20171674	15.922510	6.0324850	0.011	0.001863
DADA2	Forward Only	Present	57.278	7.774	0.138	0.000995	4.89033206	89.555907	12.3784321	0.069	0.001035
		Removed	113.209	16.090	0.108	0.001361	2.25777735	156.443424	22.5274252	0.058	0.001307
	Reverse Only	Present	68.601	11.479	0.138	0.001024	4.62640572	107.679234	18.2708333	0.070	0.001071
		Removed	138.552	24.420	0.106	0.001415	1.99300550	191.306041	33.9853561	0.056	0.001340
	Merged	Present	3.611	0.248	0.065	0.000398	13.63682311	4.138997	0.2994668	0.017	0.000242
		Removed	23.660	1.867	0.104	0.001290	4.25792812	31.692818	2.5844024	0.051	0.001184

Removing chimeras from the dataset increases variation in count between technical replicates in all pathways of both pipelines. Most chimera sequences are present at low abundance; variance in count between technical replicates is therefore intrinsically low, pulling down the mean variance for each pathway. As expected removing singletons reduced the variation in presence in both pipelines and all pathways, whilst increasing variation in counts across technical replicates. of chimera sequences thus increased variance across technical replicates when assigned taxonomy was considered (Appendix 2 - Sequences retained to abundance tables).

Appendix 4 – Sample degradation during sampling

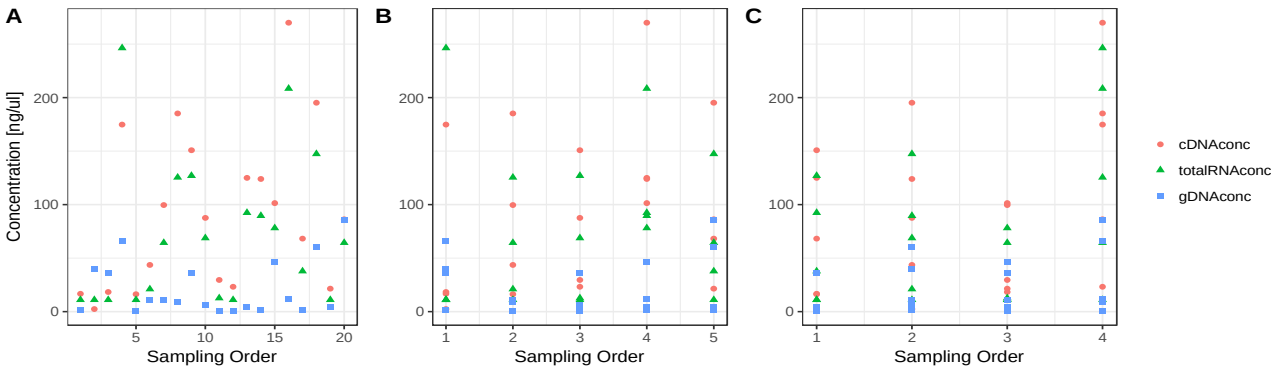
The grid itself is formed by taking columns then rows within the columns, so error might be introduced as a product of the sample creation order (columns), the order in which the samples were added to RNA later (sample storage order; rows) or a linear order of complete sample creation (sample collection order).

There was no correlation between RNA, DNA or cDNA concentration with any of the sampling process conditions (Appendix Figure 2.1) suggesting that samples are not degraded over the course of the sampling process. Any association of sequencing depth with the sampling order is therefore unlikely to result from sample degradation.

Sequence depth was not correlated with any of the sampling processes in the TCRβ, BCR or DADA2 microbial populations. There was a positive association between the QIIME microbial and TCRγ datasets and sample storage order (p-value: $3.37e^{-02}$, $2.14e^{-03}$, R^2 adjusted: 0.0600, 0.136 respectively). This is unlikely to be driven by sample degradation as the latterly stored samples retrieve a better sequencing depth.

Diversity estimates, especially species richness, can be influenced by both the differences in sequencing depth and potentially by the physical transfer of cells between samples during the sampling process. The association between sampling effort (i.e. sequencing

depth) and diversity estimates is well known and can be reduced by standardisation and compensation.



Appendix Figure 2.1 – The effect of sampling order upon sample degradation with respect to RNA, cDNA and DNA concentration. A) Sampling order is defined as the order in which samples were added to RNAlater (a function of both the row and column in which the sample was found). B) Sampling order is defined as the column in which the sample was found. C) Sampling order as defined by the row in which the sample was found. Points are coloured and shaped by the source of the concentration. There is no significant correlation between sampling order, by any definition, and the concentration of DNA, RNA or cDNA.

Both TCR datasets show a positive correlation between Shannon and Simpson's diversity estimates and sample storage order, driven by the decreased diversity in the first row. This association was not seen in the BCR or microbial datasets, which would be expected from a systemic sampling bias.

Appendix 5 – Comparing species interaction estimation methods

Chapter 3 investigated the types of interactions within the chicken microbiota estimated using a derivation of the Lodka-Volterra model. To consider the robustness of the conclusions within this chapter relating to the distinct nature of rare species interactions two further methods were compared.

[A] Methods

Three methods for estimating species interactions were used. The Method 1 and Method 2 are derived from the Lodka-volterra model, Method 3 was estimated using local species density. Methods 1 and 2 assume the community is at equilibrium.

Method 1 is the method using within Chapter 3.

Method 1 – Substitution

Estimates were calculated by substituting the abundance counts of a focal and competitor species across the grid, according to Appendix Equations 3.1 and 3.2.

$$\beta = \frac{1 - \bar{\alpha}\tilde{x}_j}{\tilde{y}_j}$$

Appendix Equation 3.1 – The interspecific interaction (β) calculated between a focal species ($\tilde{x}$) and a competitor species ($\tilde{y}$) at equilibrium. Where $\tilde{x}$ is the abundance of the focal species in a given location (j) and $\tilde{y}$ is the abundance of the competitor species in the same location. $\bar{\alpha}$ used here is the mean of all the possible intraspecific interactions of a focal species in the absence of the competitor species.

$$\bar{\alpha} = \frac{\sum_{i=1}^{n} \frac{1}{\tilde{x}_i}}{n}$$

Appendix Equation 3.2 – The intraspecific interaction ($\bar{\alpha}$) for a specific focal species at equilibrium ($\tilde{x}$). α can be calculated for each occurrence (i) where a species is not found in the presence of a specified competitor species (see Equation 3.4). The intraspecific interaction of a species is taken to be the mean of α. So $\tilde{x}_i$ is the abundance of a focal species in a given location and n is the number of locations in which the focal species is found without its competitor.

Method 2 – Non-linear convergence

The nls function was used to estimate α and β according to Appendix Equation 3.3. Data was only produced for species pairs where convergence was possible. Starting points of $\alpha = 0.1$ and $\beta = -0.5$ were used.

$$x = \frac{1 - \beta y}{\alpha}$$

Appendix Equation 3.3 – The abundance of focal species x explained by its intraspecific interactions (α) and interspecific interactions (β) with competitor species y.

Method 3 – Local density

The direction of the interaction between two species can also be estimated by the association between the local densities of each species. Linear models can be used to estimate the gradient and intercept of the association between two species. Using only those models which had a significant ($p < 0.05$) association, the interspecific interaction (m) between a focal and competitor species can be estimated (see Appendix Equations 3.4 and 3.5).

$$\text{(A)} \qquad x - \bar{x} = my + c$$

Appendix Equation 3.4 – The dependence of focal species x abundance upon the abundance of competitor species y, adjusted for the general abundance of focal species x across all locations within the grid. $\bar{x}$ is the mean abundance of x from every location across the grid. m is therefore reflects the interspecific interaction between x and y, and c the effect of all other species (including those of the focal species itself) upon the focal species' abundance.

$$\text{(B)} \qquad x_j - \bar{x}_i = my_j + c$$

Appendix Equation 3.5 – The dependence of focal species x abundance upon the abundance of competitor species y, adjusted for the general abundance of focal species x across all locations within the grid. $\bar{x}$ is the mean abundance of x from every location where y is not present. m is therefore reflects the interspecific interaction between x and y, and c the effect of all other species (including those of the focal species itself) upon the focal species' abundance. Location i is any segment in which species y is not found, location j any location where species y is present.

In Method 3 the m estimate is therefore comparable to the β estimates from methods 1 and 2. The interspecific estimate (gradient m) in Method 3 A is constant regardless of accounting for focal species' x averaged abundance.

Interaction Types

It should be noted that the sign of the interspecific interaction estimates produced by each of the methods does not correspond to the same type of interspecific interaction. A negative β in either Method 1 or Method 2 corresponds to a positive interaction between species, whilst a positive m indicates a positive interaction in Method 3.

[B] Consistency of species interaction estimates

As would be expected, given the different requirements of each method, there were different numbers of interactions able to be predicted between species dependent upon method. Method 2 was able to predict the greatest number of interactions and Method 3 the fewest (Appendix Table 3.1). Similarly the proportional contribution of the rarity groups varied between the methods, although Methods 1 and 2 showed the same pattern when the abundance groups (AC, MC, RC) were considered and Methods 1 and 3 the same using the presence groups (UP,AP,RP).

Appendix Table 3.1 – The number of unique species pairs and focal species within each rarity group for each estimation method. The proportional contribution of each rarity group relative to each rarity definition (either abundance or presence) is given in brackets to two decimal places.

	Method 1	Method 2	Method 3A	Method 3B
Species Pairs	196501	26028936	122124	
Abundant Species (AC)	52 (3.14%)	54 (1.78%)	43 (5.22%)	
Moderately abundant species (MC)	640 (38.62%)	833 (27.39%)	464 (56.38%)	
Rare species (RC)	965 (58.24%)	2154 (70.83%)	316 (38.40%)	
Global species (UP)	35 (2.11%)	35 (1.37%)	28 (3.63%)	
Widespread species (AP)	1589 (95.90%)	2029 (79.54%)	722 (93.65%)	
Localised species (RP)	33 (1.99%)	487 (19.09%)	21 (2.72%)	

There were only 12296 species pairs shared between all three methods and the direction of the interspecific interaction estimated was not always consistent. Only 3.57% (439 species pairs) of the consistent species pairs showed the same directionality of interspecific interaction between all three methods, though 17.38% (2137 species pairs) were consistent across Methods 1, 2 and at least one of the Method 3 sub-methods (Appendix Table 3.2).

Appendix Table 3.2 – The number of species pairs shared between each method with the same directionality to their interspecific interactions (including the sub-methods in method 3).

	Method 2	Method 3A	Method 3B
Method 1	10979	874	2137
Method 2			
Method 3A			12296

The majority of consistently estimated species pairs were either moderately abundant (70.56%) or widespread (85.79%), with relatively few rare species (8.63%) and no localised species. This produces a completely different structure to all three of the methods when rarity is defined by local population abundance compared with the structure of the methods when all possible species estimates are included.

[C] Interspecific interactions

Within Chapter 3 there were three key findings. Firstly, the chicken microbiota appears to be dominated by positive interspecific interactions (Figure 3.1). Secondly, the proportional contribution of negative interspecific interactions is greater in rare species than common (regardless of rarity definition, Figure 3.4). Finally, Chapter 3 suggests that interspecific

interactions between rare species appear to be most strongly negative (Figure 3.8). The latter two points are corroborated here.

Looking at the proportional contribution of different interaction types in each rarity group, the same trend can be seen regardless of method type. Specifically, rare focal species experience proportionally more negative interactions than either abundant or moderately abundant species (Appendix Figure 3.1). This trend is persistent even when comparing the relative contribution of species pairs that are not shared between the methods.

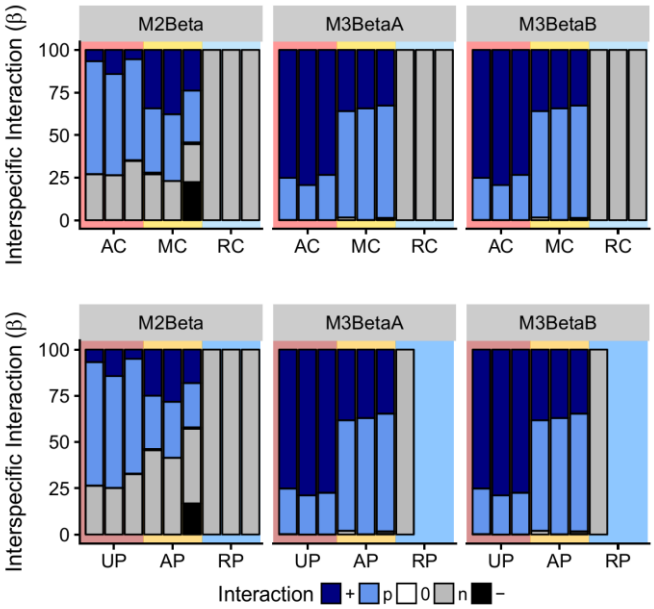


Figure 3.1 – Types of interspecific interactions, for all possible species pairs by each method, experienced within different rarity groups, defined by local population abundance or geographic spread. Method 1 shown in Figure 5.3. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and are defined by the rarity of the focal species. For full explanation of abbreviations see Table 3,2.

Method 3 did not contain neutral interactions in any of the rarity groups. The proportional contribution of positive, moderately positive and negative interactions was significantly different between the rarity groups (regardless of rarity definition) across each of the models.

Considering only the species that were estimated in all three methods, the same pattern is seen within Methods 1 and 2; more positive interspecific interactions in the abundant and moderately abundant groups and more negative in the rare group (Appendix Figure 3.2). Although the generalised linear model used found a difference in the contribution of the positive and moderately positive interactions between the abundant and moderately abundant groups, there were insufficient rare species shared with this group to compare with the more abundant groups statistically. Similarly despite the lack of comparison with the localised group, there was a significant difference between the proportional contribution of positive and moderately positive interactions between the global and widespread species regardless of estimation method used.

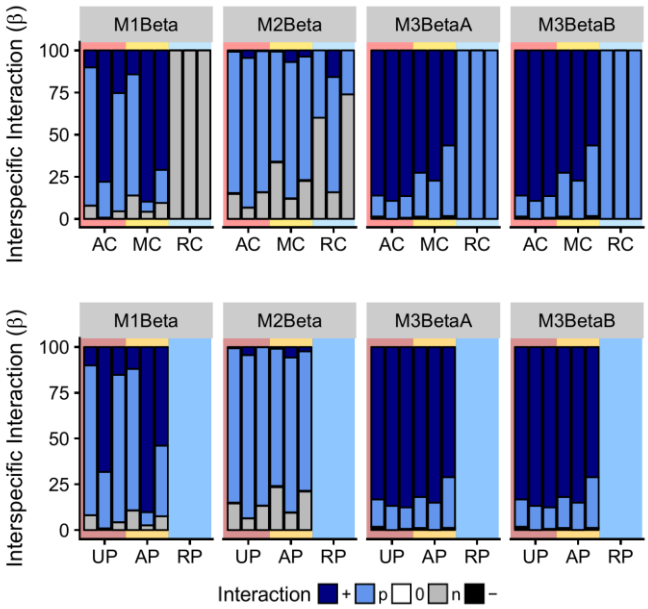


Figure 3.2 – Types of interspecific interactions experienced within different rarity groups, defined by local population abundance or geographic spread, for species pairs able to be estimated by every method. Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative and are defined by the rarity of the focal species. For full explanation of abbreviations see Table 3,2.

[D] Interspecific interactions between rarity groups

As in Method 1 (see Figure 3.9), within each focal species group there was a significant difference between their interspecific interactions depending upon competitor species rarity. Specifically, consistent with Method 1, Method 2's interspecific interactions showed an increased contribution of negative interactions in the moderately abundant/widespread and rare/localised groups. The contribution of negative interactions also increased within each of these groups as the competitor species group became rarer (Appendix Figure 3.3).

Conversely, within Method 3, although the rare focal species groups generally had less positive interactions (consistent with Methods 1 and 2) the effect of competitor species upon the relative contribution of different interactions went the opposite way (Appendix Figure 3.4). That is to say, within each focal species group, the rarer a competitor species group the more likely its interactions with a focal species are to be strongly positive.

Although once again the effects in Appendix Figures 3.3 and 3.4 were present when only species pairs consistently found across all three methods were considered, the effect was weaker (Appendix Figure 3.5).

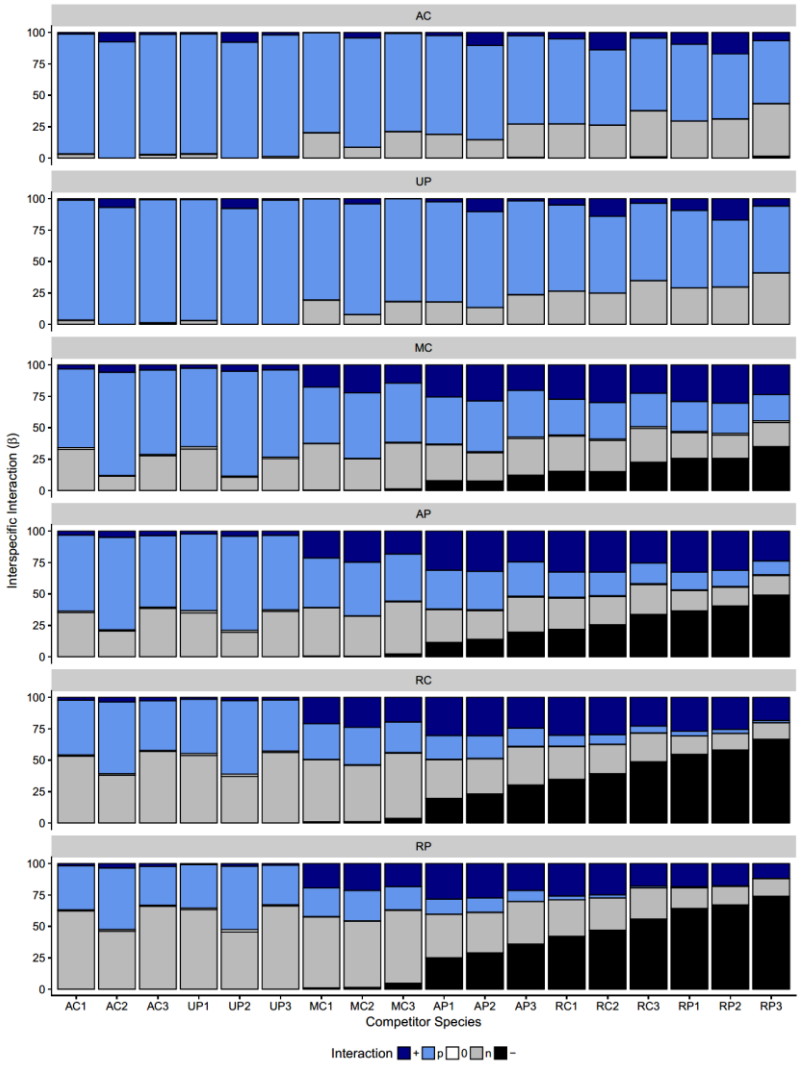


Figure 3.3 – The proportional contribution of different interspecific interaction types as a function of the focal and competitor species rarity groups using Method 2 estimates. Each plot represents a different focal species rarity group. Within each plot the x-axis is a combination of the competitor species plot and the technical replicate number (e.g. competitor group AC within technical replicate 1 is labelled AC1). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

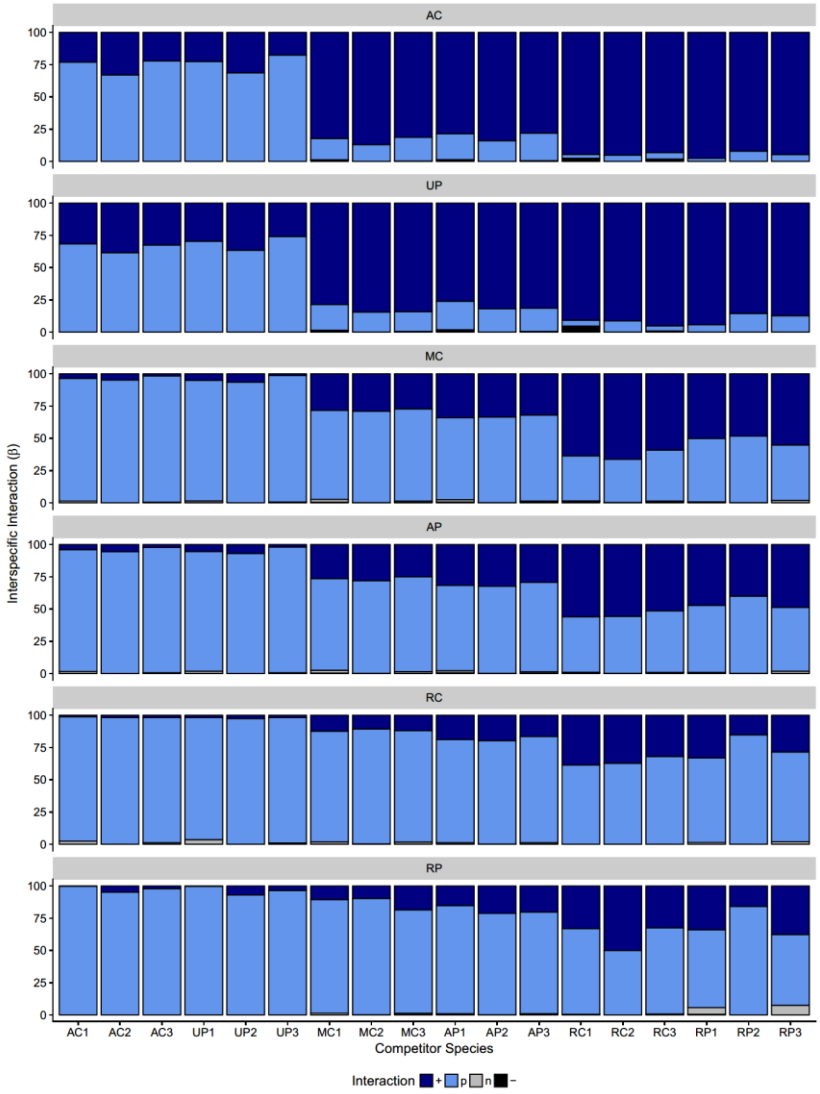


Figure 3.4 – The proportional contribution of different interspecific interaction types as a function of the focal and competitor species rarity groups using Method 3 estimates. Each plot represents a different focal species rarity group. Within each plot the x-axis is a combination of the competitor species plot and the technical replicate number (e.g. competitor group AC within technical replicate 1 is labelled AC1). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

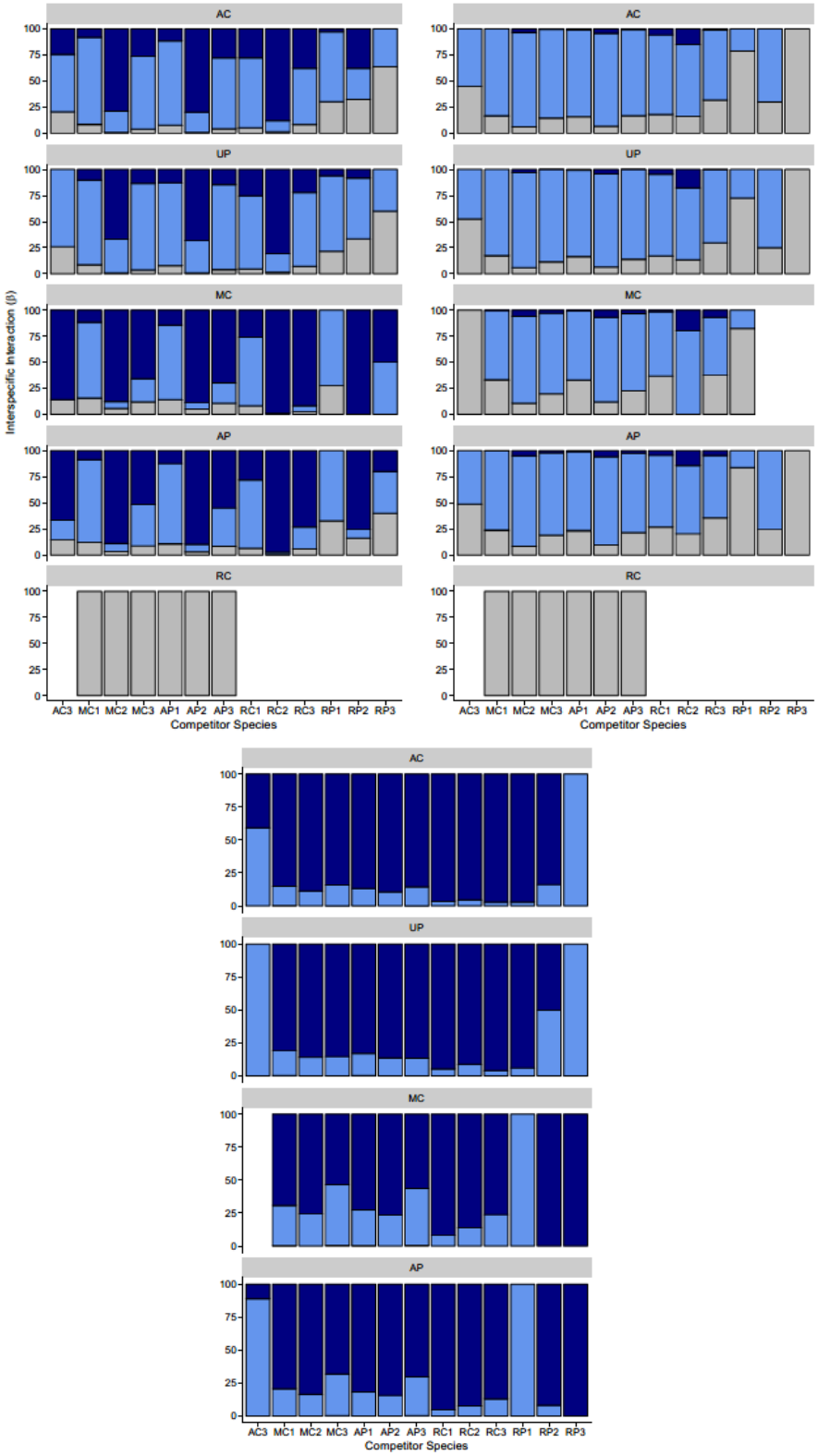


Figure 3.5 – The proportional contribution of different interspecific interaction types as a function of the focal and competitor species rarity groups, looking only at species pairs found consistently using all three estimation methods. Each plot represents a different focal species rarity group. Within each plot the x-axis is a combination of the competitor species plot and the technical replicate number (e.g. competitor group AC within technical replicate 1 is labelled AC1). Rarity group abbreviations are 'AC' abundant, 'MC' moderately abundant, 'RC' rare, 'UP' global, 'AP' widespread, and 'RP' localised. '+' refers to strong positive interspecific interactions, 'p' to weak positive, '0' to neutral, 'n' to weak negative and '-' to strong negative. For full explanation of abbreviations see Table 3.2.

Bibliography

1. Wiens, J. A. Spatial Scaling in Ecology. *Funct. Ecol.* **3,** 385–397 (1989).
2. Coyte, K. Z. & Schluter, J. The ecology of the microbiome: Networks, competition, and stability. *Science (80-.).* **1,** 663–666 (2015).
3. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host microbiome as an ecosystem on a leash. *Nature* **548,** 43–51 (2017).
4. Stein, R. R. *et al.* Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Comput. Biol.* **9,** e1003388 (2013).
5. Chave, J. & Levin, S. Scale and Scaling in Ecological and Economic Systems. *Environ. Resour. Econ.* **26,** 527–557 (2003).
6. Mcgill, B. J. Matters of Scale. *Science (80-.).* **328,** 575–576 (2010).
7. Hanski, I. & Ovaskainen, O. The metapopulation capacity of a fragmented landscape. *Nature* **404,** 755–758 (2000).
8. Menge, B. A. & Sutherland, J. P. Species Diversity Gradients: Synthesis of the Roles of Predation, Competition, and Temporal Heterogeneity. *Am. Nat.* **110,** 351–369 (1976).
9. Whittaker, R. J., Willis, K. J. & Field, R. Scale and species richness : towards a general , theory of species diversity hierarchical. *Diversity* **28,** 453–470 (2001).
10. Cohen, J. M. *et al.* Spatial scale modulates the strength of ecological processes driving disease distributions. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E3359-64 (2016).
11. Ramakrishna, B. S. Role of the gut microbiota in human nutrition and metabolism. *J. Gastroenterol. Hepatol.* **28,** 9–17 (2013).
12. Valdes, A. M., Walter, J., Segal, E. & Spector, T. D. Role of the gut microbiota in nutrition and health. *BMJ* **361,** k2179 (2018).
13. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336,** 1262–1267 (2012).
14. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–18 (2005).
15. Smith, K., McCoy, K. D. & Macpherson, A. J. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin. Immunol.* **19,** 59–69 (2007).
16. Hooper, L. V & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat. Rev. Immunol.* **10,** 159–69 (2010).
17. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses

during health and disease. *Nat. Rev. Immunol.* **9,** 313–23 (2009).

18. Sarkar, A. *et al.* The Microbiome in Psychology and Cognitive Neuroscience. *Trends Cogn. Sci.* **22,** 611–636 (2018).
19. Sampson, T. R. & Mazmanian, S. K. Control of Brain Development, Function, and Behavior by the Microbiome. *Cell Host Microbe* **17,** 565–576 (2015).
20. Stecher, B. & Hardt, W.-D. Mechanisms controlling pathogen colonization of the gut. *Curr. Opin. Microbiol.* **14,** 82–91 (2011).
21. Willing, B. P., Vacharaksa, A., Croxen, M., Thanachayanont, T. & Finlay, B. B. Altering host resistance to infections through microbial transplantation. *PLoS One* **6,** e26988 (2011).
22. Tang, W. H. W., Kitai, T. & Hazen, S. L. Gut Microbiota in Cardiovascular Health and Disease. *Circ. Res.* **120,** 1183–1196 (2017).
23. Garrett, W. S. Cancer and the microbiota. *Science* **348,** 80–6 (2015).
24. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535,** 56–64 (2016).
25. Wells, J. M., Rossi, O., Meijerink, M. & van Baarlen, P. Epithelial crosstalk at the microbiota-mucosal interface. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 4607–14 (2011).
26. Kayama, H. & Takeda, K. Regulation of intestinal homeostasis by innate and adaptive immunity. *Int. Immunol.* **24,** 673–680 (2012).
27. Coombes, J. L. & Powrie, F. Dendritic cells in intestinal immune regulation. *Nat. Rev. Immunol.* **8,** 435–446 (2009).
28. Zhao, Q. & Elson, C. O. Adaptive immune education by gut microbiota antigens. *Immunology* **154,** 28–37 (2018).
29. Cario, E., Gerken, G. & Podolsky, D. K. Toll-Like Receptor 2 Controls Mucosal Inflammation by Regulating Epithelial Barrier Function. *Gastroenterology* **132,** 1359–1374 (2007).
30. Kunisawa, J. & Kiyono, H. Immune regulation and monitoring at the epithelial surface of the intestine. *Drug Discov. Today* **18,** 87–92 (2013).
31. Deplancke, B. & Gaskins, H. R. Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. *Am. J. Clin. Nutr.* **73,** 1131S–1141S (2001).
32. Cash, H. L., Whitham, C. V, Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313,** 1126–30 (2006).
33. Vaishnava, S. *et al.* The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* **334,** 255–8 (2011).
34. Levast, B., Li, Z. & Madrenas, J. The role of IL-10 in microbiome-associated immune modulation and disease tolerance. *Cytokine* **75,** 291–301 (2015).

35. Bauché, D. & Marie, J. C. Transforming growth factor β: a master regulator of the gut microbiota and immune cell interactions. *Clin. Transl. Immunol.* **6,** e136 (2017).

36. Uematsu, S. *et al.* Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. *Nat. Immunol.* **9,** 769–776 (2008).

37. Kawamoto, S. *et al.* Foxp3+ T Cells Regulate Immunoglobulin A Selection and Facilitate Diversification of Bacterial Species Responsible for Immune Homeostasis. *Immunity* **41,** 152–65 (2014).

38. Pabst, O., Cerovic, V. & Hornef, M. Secretory IgA in the Coordination of Establishment and Maintenance of the Microbiota. *Trends Immunol.* **37,** 287–296 (2016).

39. Kubinak, J. L. *et al.* MyD88 Signaling in T Cells Directs IgA-Mediated Control of the Microbiota to Promote Health. *Cell Host Microbe* **17,** 153–163 (2015).

40. Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328,** 1705–9 (2010).

41. Lindner, C. *et al.* Age, microbiota, and T cells shape diverse individual IgA repertoires in the intestine. *J. Exp. Med.* **209,** 365–377 (2012).

42. Jakobsson, H. E. *et al.* Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* **63,** 559–66 (2014).

43. Suzuki, K. *et al.* The sensing of environmental stimuli by follicular dendritic cells promotes immunoglobulin A generation in the gut. *Immunity* **33,** 71–83 (2010).

44. Cerutti, A. The regulation of IgA class switching. *Nat. Rev. Immunol.* **8,** 421–434 (2008).

45. Wlodarska, M., Kostic, A. D. & Xavier, R. J. An Integrative View of Microbiome-Host Interactions in Inflammatory Bowel Diseases. *Cell Host Microbe* **17,** 577–591 (2015).

46. Khosravi, A. & Mazmanian, S. K. Disruption of the gut microbiome as a risk factor for microbial infections. *Curr. Opin. Microbiol.* **16,** 221–7 (2013).

47. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–5 (2008).

48. Stecher, B. *et al.* Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathog.* **6,** e1000711 (2010).

49. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4,** 337–49 (2008).

50. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16731–16736 (2008).

51. Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1

regulates intestinal homeostasis. *Nature* **456,** 507–510 (2008).

52. Stiemsma, L. T. & Michels, K. B. The Role of the Microbiome in the Developmental Origins of Health and Disease. *Pediatrics* **141,** e20172437 (2018).
53. Maynard, C. L., Elson, C. O., Hatton, R. D. & Weaver, C. T. Reciprocal interactions of the intestinal microbiota and immune system. *Nature* **489,** 231–41 (2012).
54. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat. Med.* **22,** 713–722 (2016).
55. Haribhai, D. *et al.* A requisite role for induced regulatory T cells in tolerance based on expanding antigen receptor diversity. *Immunity* **35,** 109–22 (2011).
56. Kuhn, K. A. & Stappenbeck, T. S. Peripheral education of the immune system by the colonic microbiota. *Semin. Immunol.* **25,** 364–369 (2013).
57. Kim, M., Qie, Y., Park, J. & Kim, C. H. Gut Microbial Metabolites Fuel Host Antibody Responses. *Cell Host Microbe* **20,** 202–14 (2016).
58. Cahenzli, J., Köller, Y., Wyss, M., Geuking, M. B. & McCoy, K. D. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* **14,** 559–70 (2013).
59. Nutsch, K. *et al.* Rapid and Efficient Generation of Regulatory T Cells to Commensal Antigens in the Periphery. *Cell Rep.* **17,** 206–220 (2016).
60. Wu, W. *et al.* Commensal A4 bacteria inhibit intestinal Th2-cell responses through induction of dendritic cell TGF-β production. *Eur. J. Immunol.* **46,** 1162–7 (2016).
61. Atarashi, K. *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500,** 232–236 (2013).
62. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous Clostridium species. *Science* **331,** 337–41 (2011).
63. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–98 (2009).
64. Yang, Y. *et al.* Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510,** 152–6 (2014).
65. Atarashi, K. *et al.* Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. *Cell* **163,** 367–80 (2015).
66. Zhao, Q. *et al.* Selective Induction of Homeostatic Th17 Cells in the Murine Intestine by Cholera Toxin Interacting with the Microbiota. *J. Immunol.* **199,** 312–322 (2017).
67. Sczesnak, A. *et al.* The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10,** 260–72 (2011).
68. Fauvart, M. & Michiels, J. Rhizobial secreted proteins as determinants of host specificity

in the rhizobium-legume symbiosis. *FEMS Microbiol. Lett.* **285,** 1–9 (2008).

69. Visick, K. L. & McFall-Ngai, M. J. An exclusive contract: Specificity in the *Vibrio fischeri-Euprymna scolopes* partnership. *J. Bacteriol.* **182,** 1779–1787 (2000).
70. Fraune, S. & Bosch, T. C. G. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13146–51 (2007).
71. Engel, P., Martinson, V. G. & Moran, N. A. Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci.* **109,** 11002–11007 (2012).
72. Rawls, J. F., Mahowald, M. A., Ley, R. E. & Gordon, J. I. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127,** 423–33 (2006).
73. Egerton, S., Culloty, S., Whooley, J., Stanton, C. & Ross, R. P. The gut microbiota of marine fish. *Front. Microbiol.* **9,** 1–17 (2018).
74. Colston, T. J. & Jackson, C. R. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Mol. Ecol.* **25,** 3776–3800 (2016).
75. Stanley, D., Hughes, R. J. & Moore, R. J. Microbiota of the chicken gastrointestinal tract: Influence on health, productivity and disease. *Applied Microbiology and Biotechnology* **98,** 4301–4310 (2014).
76. Wei, S., Morrison, M. & Yu, Z. Bacterial census of poultry intestinal microbiome. *Poult Sci* **92,** 671–683 (2013).
77. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The Human Microbiota in Health and Disease. *Engineering* **3,** 71–82 (2017).
78. Inoue, R., Otsuka, M. & Ushida, K. Development of Intestinal Microbiota in Mice and Its Possible Interaction with the Evolution of Luminal IgA in the Intestine. *Exp. Anim.* **54,** 437–445 (2005).
79. Mcfall-Ngai, M. Care for the community. *Nature* **445,** 153 (2007).
80. Magor, K. E. *et al.* Defense genes missing from the flight division. *Dev. Comp. Immunol.* **41,** 377–88 (2013).
81. Smith, A. L., Powers, C. & Beal, R. K. The Avian Enteric Immune System in Health and Disease. in *Avian Immunology* 112–118 (2013). doi:10.1016/j.neuron.2013.12.034.Reactive
82. Oláh, I., Nagy, N. & Vervelde, L. Structure of the Avian Lymphoid System. in *Avian Immunology* 11–44 (Elsevier, 2014). doi:10.1016/B978-0-12-396965-1.00002-9
83. Lu, J. *et al.* Diversity and succession of the intestinal bacterial community of the maturing broiler chicken. *Appl. Environ. Microbiol.* **69,** 6816–24 (2003).
84. Sergeant, M. J. *et al.* Extensive microbial and functional diversity within the chicken cecal

microbiome. *PLoS One* **9,** e91941 (2014).

85. Choi, J. H., Kim, G. B. & Cha, C. J. Spatial heterogeneity and stability of bacterial community in the gastrointestinal tracts of broiler chickens. *Poult. Sci.* **93,** 1942–1950 (2014).
86. Sarma-Rupavtarm, R. B., Ge, Z., Schauer, D. B., Fox, J. G. & Polz, M. F. Spatial distribution and stability of the eight microbial species of the altered Schaedler flora in the mouse gastrointestinal tract. *Appl. Environ. Microbiol.* **70,** 2791–2800 (2004).
87. Zhang, Z. *et al.* Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* **8,** 881–93 (2014).
88. Videnska, P. *et al.* Succession and replacement of bacterial populations in the caecum of egg laying hens over their whole life. *PLoS One* 1–14 (2014). doi:10.1371/journal.pone.0115142
89. Oakley, B. B. & Kogut, M. H. Spatial and Temporal Changes in the Broiler Chicken Cecal and Fecal Microbiomes and Correlations of Bacterial Taxa with Cytokine Gene Expression. *Front. Vet. Sci.* **3,** 1–12 (2016).
90. Awad, W. A. *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.* **6,** 154 (2016).
91. Tjoelker, L. W. *et al.* Evolutionary conservation of antigen recognition: the chicken T-cell receptor beta chain. *Proc. Natl. Acad. Sci.* **87,** 7856–7860 (1990).
92. Smith, A. L. & Göbel, T. W. Avian T Cells: Antigen Recognition and Lineages. in *Avian Immunology* 91–102 (2014). doi:10.1016/B978-0-12-396965-1.00005-4
93. Su, C., Jakobsen, I., Gu, X. & Nei, M. Diversity and evolution of T-cell receptor variable region genes in mammals and birds. *Immunogenetics* **50,** 301–8 (1999).
94. Parra, D., Takizawa, F. & Sunyer, J. O. Evolution of B Cell Immunity. *Annu. Rev. Anim. Biosci.* **1,** 65–97 (2013).
95. Ratcliffe, M. J. H. & Härtle, S. B Cells, the Bursa of Fabricius and the Generation of Antibody Repertoires. in *Avian Immunology: Second Edition* 65–89 (2013). doi:10.1016/B978-0-12-396965-1.00004-2
96. Parvari, R. *et al.* Chicken immunoglobulin gamma-heavy chains: limited VH gene repertoire, combinatorial diversification by D gene segments and evolution of the heavy chain locus. *EMBO J.* **7,** 739–44 (1988).
97. Yasuda, M., Taura, Y., Yokomizo, Y. & Ekino, S. A comparative study of germinal center: fowls and mammals. *Comp. Immunol. Microbiol. Infect. Dis.* **21,** 179–189 (1998).
98. Yasuda, M. *et al.* Immunobiology of chicken germinal center: I. Changes in surface Ig class expression in the chicken splenic germinal center after antigenic stimulation. *Dev.*

Comp. Immunol. **27,** 159–166 (2003).

99. Mcloughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. Host Selection of Microbiota via Differential Adhesion. (2016). doi:10.1016/j.chom.2016.02.021
100. Cavalier-Smith, T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5,** 174–182 (2000).
101. Berendsen, R. L., Pieterse, C. M. J. & Bakker, P. A. H. M. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17,** 478–486 (2012).
102. Stambler, N. Zooxanthellae: The yellow symbionts inside animals. in *Coral Reefs: An Ecosystem in Transition* 87–106 (Springer Netherlands, 2011). doi:10.1007/978-94-007-0114-4_7
103. Moran, N. A., Plague, G. R., Sandström, J. P. & Wilcox, J. L. A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14543–8 (2003).
104. Kaltenpoth, M. *et al.* Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis. *Proc. Natl. Acad. Sci.* **111,** 6359–6364 (2014).
105. McFall-Ngai, M. J. & Ruby, E. G. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science* **254,** 1491–1494 (1991).
106. Dyall, S. D. & Johnson, P. J. Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr. Opin. Microbiol.* **3,** 404–411 (2000).
107. Pabst, O. *et al.* Adaptation of solitary intestinal lymphoid tissue in response to microbiota and chemokine receptor CCR7 signaling. *J. Immunol.* **177,** 6824–6832 (2006).
108. Kamada, N. & Núñez, G. Regulation of the Immune System by the Resident Intestinal Bacteria. *Gastroenterology* **146,** 1477–1488 (2014).
109. Stanley, D. *et al.* Identification of chicken intestinal microbiota correlated with the efficiency of energy extraction from feed. *Vet. Microbiol.* **164,** 85–92 (2013).
110. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328,** 228–31 (2010).
111. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15718–23 (2004).
112. van der Waaij, D., Berghuis-de Vries, J. M. & Lekkerkerk Lekkerkerk-v. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *J. Hyg. (Lond).* **69,** 405–411 (1971).
113. Kaltenpoth, M. Actinobacteria as mutualists: general healthcare for insects? *Trends Microbiol.* **17,** 529–535 (2009).
114. Koch, H. & Schmid-hempel, P. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. **108,** (2011).

115. Deshmukh, H. S. *et al.* The microbiota regulates neutrophil homeostasis and host resistance to Escherichia coli K1 sepsis in neonatal mice. *Nat. Med.* **20,** 524–30 (2014).

116. Pham, T. A. N. *et al.* Epithelial IL-22RA1-Mediated Fucosylation Promotes Intestinal Colonization Resistance to an Opportunistic Pathogen. *Cell Host Microbe* **16,** 504–516 (2014).

117. Maier, B. R. & Hentges, D. J. Experimental Shigella infections in laboratory animals. I. Antagonism by human normal flora components in gnotobiotic mice. *Infect. Immun.* **6,** 168–73 (1972).

118. O'Mahony, C. *et al.* Commensal-induced regulatory T cells mediate protection against pathogen-stimulated NF-kappaB activation. *PLoS Pathog.* **4,** e1000112 (2008).

119. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16731–6 (2008).

120. Mendes, R. *et al.* Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332,** 1097–1100 (2011).

121. Knights, D., Lassen, K. G. & Xavier, R. J. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut* **62,** 1505–10 (2013).

122. Grigg, J. B. & Sonnenberg, G. F. Host-Microbiota Interactions Shape Local and Systemic Inflammatory Diseases. *J. Immunol.* **198,** 564–571 (2017).

123. Lozupone, C. *et al.* Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts. *Genome Res.* **22,** 1974–1984 (2012).

124. Sellon, R. K. *et al.* Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect. Immun.* **66,** 5224–5231 (1998).

125. Jarry, a *et al.* Subversion of human intestinal mucosa innate immunity by a Crohn's disease-associated E. coli. *Mucosal Immunol.* **8,** 572–581 (2015).

126. Barnich, N. *et al.* CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. *J. Clin. Invest.* **117,** 1566–74 (2007).

127. May, R. M. Will a large complex system be stable? *Nature* **238,** 413–414 (1972).

128. Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483,** 205–208 (2012).

129. Schluter, J. & Foster, K. R. The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol.* **10,** e1001424 (2012).

130. Chu, H. & Mazmanian, S. K. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat. Immunol.* **14,** 668–75 (2013).

131. Suzuki, K. *et al.* Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1981–6 (2004).

132. Doornbos, R. F., Van Loon, L. C. & Bakker, P. a H. M. Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agron. Sustain. Dev.* **32,** 227–243 (2012).

133. Garbeva, P., van Elsas, J. D. & van Veen, J. a. Rhizosphere microbial community and its response to plant species and soil history. *Plant Soil* **302,** 19–32 (2008).

134. Miethling, R., Wieland, G., Backhaus, H. & Tebbe, C. C. Variation of Microbial Rhizosphere Communities in Response to Crop Species, Soil Origin, and Inoculation with Sinorhizobium meliloti L33. *Microb. Ecol.* **40,** 43–56 (2000).

135. Yang, J. W. *et al.* Whitefly infestation of pepper plants elicits defence responses against bacterial pathogens in leaves and roots and changes the below-ground microflora. *J. Ecol.* **99,** 46–56 (2011).

136. Weiland-Bräuer, N. *et al.* Composition of Bacterial Communities Associated with Aurelia aurita Changes with Compartment, Life Stage, and Population. *Appl. Environ. Microbiol.* **81,** 6038–6052 (2015).

137. Seedorf, H. *et al.* Bacteria from Diverse Habitats Colonize and Compete in the Mouse Gut. *Cell* **159,** 253–266 (2014).

138. Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20858–63 (2008).

139. Goto, Y. & Ivanov, I. I. Intestinal epithelial cells as mediators of the commensal-host immune crosstalk. *Immunol. Cell Biol.* **91,** 204–14 (2013).

140. Abreu, M. T. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. *Nat. Rev. Immunol.* **10,** 131–44 (2010).

141. Johansson, M. E. V, Larsson, J. M. H. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 4659–65 (2011).

142. Holm, L. & Phillipson, M. Assessment of Mucus Thickness and Production in situ. *Methods Mol. Biol.* **842,** 217–227 (2012).

143. Atuma, C., Strugala, V., Allen, a & Holm, L. The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280,** G922-9 (2001).

144. Sonnenburg, J. L., Angenent, L. T. & Gordon, J. I. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat. Immunol.* **5,** 569–73 (2004).

145. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).

146. Derrien, M. *et al.* Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* **1,** 254–268 (2010).

147. Koropatkin, N. M., Cameron, E. a. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10,** 323–335 (2012).

148. Meyer-Hoffert, U. *et al.* Secreted enteric antimicrobial activity localises to the mucus surface layer. *Gut* **57,** 764–71 (2008).

149. Salzman, N. H. Paneth cell defensins and the regulation of the microbiome: détente at mucosal surfaces. *Gut Microbes* **1,** 401–6 (2010).

150. He, B. *et al.* Intestinal bacteria trigger T cell-independent immunoglobulin A(2) class switching by inducing epithelial-cell secretion of the cytokine APRIL. *Immunity* **26,** 812–26 (2007).

151. Peterson, D. a, McNulty, N. P., Guruge, J. L. & Gordon, J. I. IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe* **2,** 328–39 (2007).

152. Schluter, J., Nadell, C. D., Bassler, B. L. & Foster, K. R. Adhesion as a weapon in microbial competition. *ISME J.* **9,** 139–149 (2015).

153. Hooper, L. V, Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9833–8 (1999).

154. Pickard, J. M. & Chervonsky, A. V. Intestinal Fucose as a Mediator of Host-Microbe Symbiosis. *J. Immunol.* **194,** 5588–5593 (2015).

155. Weiss, G. a, Chassard, C. & Hennet, T. Selective proliferation of intestinal Barnesiella under fucosyllactose supplementation in mice. *Br. J. Nutr.* **111,** 1602–1610 (2014).

156. Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host–commensal symbiosis in sickness. *Nature* **514,** 638–641 (2014).

157. Kashyap, P. C. *et al.* Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17059–64 (2013).

158. Hartley, C. L., Neumann, C. S. & Richmond, M. H. Adhesion of commensal bacteria to the large intestine wall in humans. *Infect. Immun.* **23,** 128–32 (1979).

159. Fuller, R. & Brooker, B. E. Lactobacilli which attach to the crop epithelium of the fowl. *Am J Clin Nutr* **27,** 1305–1312 (1974).

160. Freter, R. Mechanisms of association of bacteria with mucosal surfaces. *Ciba Found. Symp.* **80,** 36–55 (1981).

161. Schluter, J., Nadell, C. D., Bassler, B. L. & Foster, K. R. Adhesion as a weapon in microbial

competition. *ISME J.* **9,** 139–149 (2015).

162. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6,** 81 (2015).
163. Ann Naughton, J. *et al.* Divergent mechanisms of interaction of helicobacter pylori and campylobacter jejuni with mucus and mucins. *Infect. Immun.* **81,** 2838–2850 (2013).
164. Schroeder, H. W. & Cavacini, L. Structure and Function of immunoglobulins. *J. allergy Clin. Immunol.* **125,** S41-52 (2013).
165. Palm, N. W. *et al.* Immunoglobulin A Coating Identifies Colitogenic Bacteria in Inflammatory Bowel Disease. *Cell* **158,** 1000–1010 (2014).
166. Patrick, C. C., Virella, G., Koistinen, J. & Fudenberg, H. H. Differential binding of IgA proteins of different subclasses and allotypes to Staphylococcal protein A. *Zeitschrift für Immunitätsforschung. Immunobiol.* **153,** 466–9 (1977).
167. Robbe, C., Capon, C., Coddeville, B. & Michalski, J.-C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J.* **384,** 307–16 (2004).
168. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501,** 426–9 (2013).
169. Xavier, J. B. & Foster, K. R. Cooperation and conflict in microbial biofilms. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 876–81 (2007).
170. Nadell, C. D., Foster, K. R. & Xavier, J. B. Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Comput. Biol.* **6,** e1000716 (2010).
171. Kreft, J. U., Picioreanu, C., Wimpenny, J. W. & van Loosdrecht, M. C. Individual-based modelling of biofilms. *Microbiology* **147,** 2897–912 (2001).
172. Xavier, J. B., Picioreanu, C. & van Loosdrecht, M. C. M. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environ. Microbiol.* **7,** 1085–103 (2005).
173. Kim, W., Racimo, F., Schluter, J., Levy, S. B. & Foster, K. R. Importance of positioning for microbial evolution. *Proc. Natl. Acad. Sci.* **111,** E1639–E1647 (2014).
174. Kim, J. & Khan, W. Goblet Cells and Mucins: Role in Innate Defense in Enteric Infections. *Pathogens* **2,** 55–70 (2013).
175. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27,** 379–423 (1948).
176. Rosenberg, E., Koren, O., Reshef, L., Efrony, R. & Zilber-Rosenberg, I. The role of microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* **5,** 355–362 (2007).
177. Guzmàn, C. A., Biavasco, F. & Pruzzo, C. News & notes: adhesiveness of Bacteroides

fragilis strains isolated from feces of healthy donors, abscesses, and blood. *Curr. Microbiol.* **34,** 332–4 (1997).

178. Grubb, S. E. W. *et al.* Adhesion of Candida albicans to endothelial cells under physiological conditions of flow. *Infect. Immun.* **77,** 3872–8 (2009).
179. Nowrouzian, F. L., Friman, V., Adlerberth, I. & Wold, A. E. Different phylogenetic profile and reduced mannose-sensitive adherence capacity characterize commensal Escherichia coli in IgA deficient individuals. *Microb. Pathog.* **61–62,** 62–5 (2013).
180. Lukens, J. R. *et al.* Dietary modulation of the microbiome affects autoinflammatory disease. *Nature* **516,** 246–249 (2014).
181. Tachon, S., Lee, B. & Marco, M. L. Diet alters probiotic Lactobacillus persistence and function in the intestine. *Environ. Microbiol.* **16,** 2915–2926 (2014).
182. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–8 (2011).
183. David, L. a *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–63 (2014).
184. Kohl, K. D., Amaya, J., Passement, C. a., Dearing, M. D. & McCue, M. D. Unique and shared responses of the gut microbiota to prolonged fasting: a comparative study across five classes of vertebrate hosts. *FEMS Microbiol. Ecol.* **90,** 883–894 (2014).
185. Caldara, M. *et al.* Mucin biopolymers prevent bacterial aggregation by retaining cells in the free-swimming state. *Curr. Biol.* **22,** 2325–2330 (2012).
186. Huang, J., Lee, S. & Mazmanian, S. The human commensal Bacteroides fragilis binds intestinal mucin. *Anaerobe* **17,** 137–141 (2011).
187. Kinoshita, H. *et al.* Cell surface glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of Lactobacillus plantarum LA 318 recognizes human A and B blood group antigens. *Res. Microbiol.* **159,** 685–91 (2008).
188. Boshuizen, J. A. *et al.* Homeostasis and function of goblet cells during rotavirus infection in mice. *Virology* **337,** 210–21 (2005).
189. Songhet, P. *et al.* Stromal IFN-γR-signaling modulates goblet cell function during Salmonella Typhimurium infection. *PLoS One* **6,** e22459 (2011).
190. Guilmeau, S. *et al.* Intestinal deletion of Pofut1 in the mouse inactivates notch signaling and causes enterocolitis. *Gastroenterology* **135,** 849–60, 860.e1–6 (2008).
191. Faure, M. *et al.* Mucin production and composition is altered in dextran sulfate sodium-induced colitis in rats. *Dig. Dis. Sci.* **48,** 1366–73 (2003).
192. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions Between the Microbiota and the Immune System. *Science (80-.).* **336,** 1268–1273 (2012).
193. Keeney, K. M., Yurist-Doutsch, S., Arrieta, M.-C. & Finlay, B. B. Effect of Antibiotics on

Human Microbiota and Subsequent Disease. *Annu. Rev. Microbiol.* 217–235 (2014). doi:10.1146/annurev-micro-091313-103456

194. Tracy, A. M., Koren, O., Douglas, N., Weil, E. & Harvell, C. D. Persistent shifts in Caribbean coral microbiota are linked to the 2010 warm thermal anomaly. *Environ. Microbiol. Rep.* **7,** 471–479 (2015).
195. Faber, F. & Bäumler, A. J. The impact of intestinal inflammation on the nutritional environment of the gut microbiota. *Immunol. Lett.* **162,** 48–53 (2014).
196. Moeller, A. H. *et al.* Rapid changes in the gut microbiome during human evolution. *Proc. Natl. Acad. Sci.* **111,** 16431–16435 (2014).
197. Stevenson, T. J., Buck, C. L. & Duddleston, K. N. Temporal dynamics of the cecal gut microbiota of juvenile arctic ground squirrels: a strong litter effect across the first active season. *Appl. Environ. Microbiol.* **80,** 4260–8 (2014).
198. Carey, H. V, Walters, W. A. & Knight, R. Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **304,** R33-42 (2013).
199. Bergstrom, K. S. B. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* **23,** 1026–1037 (2013).
200. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host Microbe* **4,** 447–457 (2008).
201. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14,** 20–32 (2016).
202. Wold, A. E. *et al.* Secretory immunoglobulin A carries oligosaccharide receptors for Escherichia coli type 1 fimbrial lectin. *Infect. Immun.* **58,** 3073–7 (1990).
203. Mantis, N. J., Rol, N. & Corthésy, B. Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol.* **4,** 603–11 (2011).
204. Macpherson, A. J., Hunziker, L., McCoy, K. & Lamarre, A. IgA responses in the intestinal mucosa against pathogenic and non-pathogenic microorganisms. *Microbes Infect.* **3,** 1021–1035 (2001).
205. Kawamoto, S. *et al.* The inhibitory receptor PD-1 regulates IgA selection and bacterial composition in the gut. *Science* **336,** 485–9 (2012).
206. Mathias, A. & Corthesy, B. Recognition of Gram-positive Intestinal Bacteria by Hybridoma- and Colostrum-derived Secretory Immunoglobulin A Is Mediated by Carbohydrates. *J. Biol. Chem.* **286,** 17239–17247 (2011).
207. Moreau, M. C., Ducluzeau, R., Guy-Grand, D. & Muller, M. C. Increase in the population of duodenal immunoglobulin A plasmocytes in axenic mice associated with different

living or dead bacterial strains of intestinal origin. *Infect. Immun.* **21,** 532–9 (1978).

208. Round, J. L., O'Connell, R. M. & Mazmanian, S. K. Coordination of tolerogenic immune responses by the commensal microbiota. *J. Autoimmun.* **34,** J220-5 (2010).
209. van der Waaij, L. a *et al.* Immunoglobulin coating of faecal bacteria in inflammatory bowel disease. *Eur. J. Gastroenterol. Hepatol.* **16,** 669–674 (2004).
210. van der Waaij, L. A., Limburg, P. C., Mesander, G. & van der Waaij, D. In vivo IgA coating of anaerobic bacteria in human faeces. *Gut* **38,** 348–54 (1996).
211. D'Auria, G. *et al.* Active and secreted IgA-coated bacterial fractions from the human gut reveal an under-represented microbiota core. *Sci. Rep.* **3,** 1–9 (2013).
212. Tsuruta, T. *et al.* The amount of secreted IgA may not determine the secretory IgA coating ratio of gastrointestinal bacteria. *FEMS Immunol. Med. Microbiol.* **56,** 185–189 (2009).
213. Olmsted, S. S. *et al.* Diffusion of macromolecules and virus-like particles in human cervical mucus. *Biophys J* **81,** 1930–7 (2001).
214. Biesbrock, A. R., Reddy, M. S. & Levine, M. J. Interaction of a salivary mucin-secretory immunoglobulin A complex with mucosal pathogens. *Infect. Immun.* **59,** 3492–7 (1991).
215. Phalipon, A. *et al.* Secretory Component. *Immunity* **17,** 107–115 (2002).
216. Bollinger, R. R. *et al.* Secretory IgA and mucin-mediated biofilm formation by environmental strains of Escherichia coli: role of type 1 pili. *Mol. Immunol.* **43,** 378–87 (2006).
217. Randal Bollinger, R. *et al.* Human secretory immunoglobulin A may contribute to biofilm formation in the gut. *Immunology* **109,** 580–587 (2003).
218. Friman, V. *et al.* Decreased expression of mannose-specific adhesins by Escherichia coli in the colonic microflora of immunoglobulin A-deficient individuals. *Infect. Immun.* **64,** 2794–8 (1996).
219. Rogier, E. W. *et al.* Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 3074–9 (2014).
220. MacKenzie, D. A., Tailford, L. E., Hemmings, A. M. & Juge, N. Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. *J Biol Chem* **284,** 32444–53 (2009).
221. Forbes, S. J. *et al.* Association of a protective monoclonal IgA with the O antigen of Salmonella enterica serovar Typhimurium impacts type 3 secretion and outer membrane integrity. *Infect. Immun.* **80,** 2454–63 (2012).
222. Boullier, S. *et al.* Secretory IgA-mediated neutralization of Shigella flexneri prevents intestinal tissue destruction by down-regulating inflammatory circuits. *J. Immunol.* **183,**

5879–85 (2009).

223. Lindner, C. *et al.* Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. *Nat. Immunol.* **16,** 880–888 (2015).
224. Mantis, N. J. & Forbes, S. J. Secretory IgA: arresting microbial pathogens at epithelial borders. *Immunol. Invest.* **39,** 383–406 (2010).
225. Kaetzel, C. S. Cooperativity among secretory IgA, the polymeric immunoglobulin receptor, and the gut microbiota promotes host–microbial mutualism. *Immunol. Lett.* **162,** 10–21 (2014).
226. Mirpuri, J. *et al.* Proteobacteria-specific IgA regulates maturation of the intestinal microbiota. *Gut Microbes* **5,** 28–39 (2014).
227. Brandtzaeg, P. Development and basic mechanisms of human gut immunity. *Nutr. Rev.* **56,** S5-18 (1998).
228. Beatty, D. W., Napier, B., Sinclair-Smith, C. C., McCabe, K. & Hughes, E. J. Secretory IgA synthesis in Kwashiorkor. *J. Clin. Lab. Immunol.* **12,** 31–6 (1983).
229. Akhiani, A. A., Stensson, A., Schon, K. & Lycke, N. Y. IgA Antibodies Impair Resistance against Helicobacter pylori Infection: Studies on Immune Evasion in IL-10-Deficient Mice. *J. Immunol.* **174,** 8144–8153 (2005).
230. Keppel, G. *et al.* Refugia: Identifying and understanding safe havens for biodiversity under climate change. *Glob. Ecol. Biogeogr.* **21,** 393–404 (2012).
231. Stewart, J. R., Lister, A. M., Barnes, I. & Dalén, L. Refugia revisited: individualistic responses of species in space and time. *Proc. Biol. Sci.* **277,** 661–671 (2010).
232. Stanley, D., Geier, M. S., Hughes, R. J., Denman, S. E. & Moore, R. J. Highly Variable Microbiota Development in the Chicken Gastrointestinal Tract. *PLoS One* **8,** e84290 (2013).
233. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–7 (2009).
234. Zhou, X. *et al.* Cecal microbiota of Tibetan Chickens from five geographic regions were determined by 16S rRNA sequencing. *Microbiologyopen* **5,** 753–762 (2016).
235. Suzuki, T. A. & Worobey, M. Geographical variation of human gut microbial composition. *Biol. Lett.* **10,** 20131037–20131037 (2014).
236. Meng, H. *et al.* Body weight selection affects quantitative genetic correlated responses in gut microbiota. *PLoS One* **9,** e89862 (2014).
237. Kafsi, H. El, Gorochov, G. & Larsen, M. Host genetics affect microbial ecosystems via host immunity. *Curr. Opin. Allergy Clin. Immunol.* **16,** 413–420 (2016).
238. Franklin, C. L. & Ericsson, A. C. Microbiota and reproducibility of rodent models. **46,** 114–122 (2017).

239. Kubasova, T. *et al.* Housing systems influence gut microbiota composition of sows but not of their piglets. *PLoS One* **12,** 1–10 (2017).

240. Bolnick, D. I. *et al.* Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nat. Commun.* **5,** 4500 (2014).

241. Markle, J. G. M. *et al.* Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339,** 1084–8 (2013).

242. Singh, R. K. *et al.* Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15,** 1–17 (2017).

243. Zhang, H., Sparks, J. B., Karyala, S. V, Settlage, R. & Luo, X. M. Host adaptive immunity alters gut microbiota. *ISME J.* 1–12 (2014). doi:10.1038/ismej.2014.165

244. 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.* **28,** 243–73 (2010).

245. Wesemann, D. R. Microbes and B Cell Development. *Adv. Immunol.* **125,** 155–178 (2015).

246. Rescigno, M. Tfr cells and IgA join forces to diversify the microbiota. *Immunity* **41,** 9–11 (2014).

247. Zhou, Y. *et al.* Biogeography of the ecosystems of the healthy human body. *Genome Biol.* **14,** R1 (2013).

248. Macpherson, A. J. & Uhr, T. Compartmentalization of the mucosal immune responses to commensal intestinal bacteria. *Ann. N. Y. Acad. Sci.* **1029,** 36–43 (2004).

249. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and stability. *Science (80-.).* **350,** 663–666 (2015).

250. Hanski, I. Spatial scale, patchiness and population dynamics on land. *Philos. Trans. R. Soc. (Biological Sci.* **343,** 19–25 (1994).

251. Durrett, R. & Levin, S. Spatial Aspects of Interspecific Competition. *Theor. Popul. Biol.* 30–43 (1998).

252. Davison, F. The Importance of the Avian Immune System and Its Unique Features. in *Avian Immunology* 1–11 (2008). doi:10.1016/B978-0-12-396965-1.00003-0

253. McCormack, W. T. *et al.* Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. *Cell* **56,** 785–91 (1989).

254. Parvari, R., Ziv, E., Lantner, F., Heller, D. & Schechter, I. Somatic diversification of chicken immunoglobulin light chains by point mutations. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 3072–3076 (1990).

255. Oakley, B. B. *et al.* The chicken gastrointestinal microbiome. *FEMS Microbiol. Lett.* **360,**

100–112 (2014).

256. Gómez Del Moral, M. *et al.* Appearance and development of lymphoid cells in the chicken (Gallus gallus) caecal tonsil. *Anat. Rec.* **250,** 182–9 (1998).

257. Seal, B. S., Lillehoj, H. S., Donovan, D. M. & Gay, C. G. Alternatives to antibiotics: a symposium on the challenges and solutions for animal production. *Anim. Heal. Res. Rev.* **14,** 78–87 (2013).

258. Pourabedin, M. & Zhao, X. Prebiotics and gut microbiota in chickens. *FEMS Microbiol. Lett.* **362,** 1–8 (2015).

259. Gaggìa, F., Mattarelli, P. & Biavati, B. Probiotics and prebiotics in animal feeding for safe food production. *International Journal of Food Microbiology* **141,** S15–S28 (2010).

260. Treves, D. S., Xia, B., Zhou, J. & Tiedje, J. M. A Two-Species Test of the Hypothesis That Spatial Isolation Influences Microbial Diversity in Soil. *Microb. Ecol.* **45,** 20–28 (2003).

261. Yutin, N. & Galperin, M. Y. Genomics update A genomic update on clostridial phylogeny : Gram-negative spore formers and other misplaced clostridia. **15,** 2631–2641 (2013).

262. Wang, Y. & Qian, P.-Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* **4,** e7401 (2009).

263. Caporaso, J. G. *et al.* correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Publ. Gr.* **7,** 335–336 (2010).

264. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–7 (2007).

265. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **69,** 330–9 (2007).

266. Bystrykh, L. V. Generalized DNA Barcode Design Based on Hamming Codes. *PLoS One* **7,** e36852 (2012).

267. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–336 (2010).

268. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13,** 581–583 (2016).

269. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–72 (2006).

270. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6,** 610–8 (2012).

271. R Development Core Team. R: A language and environment for statistical computing. (2008).

272. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20,** 289–90 (2004).

273. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3,** 217–223 (2012).

274. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8,** 28–36 (2017).

275. Wickham, H. *ggpot2: Elegant Graphics for Data Analysis*. (Springer New York, 2009). doi:10.1007/978-0-387-98141-3_1

276. Claus O. Wilke. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2' version 0.7.0 from CRAN. (2016). Available at: https://rdrr.io/cran/cowplot/. (Accessed: 30th June 2017)

277. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4.

278. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84,** 45–67 (2014).

279. Dixon, P. VEGAN: a package of R functions for community ecology. *J. Veg. Sci.* **14,** 927–930 (2003).

280. Heiberger, R. M. & Holland, B. *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS.* (2004). doi:10.1007/978-1-4757-4284-8_1

281. Tang, Y., Horikoshi, M. & Li, W. ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages.

282. Simpson, G. L. Analogue Methods in Palaeoecology: Using the analogue Package. *J. Stat. Softw.* **22,** (2007).

283. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–41 (2009).

284. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10,** 996–998 (2013).

285. Schirmer, M. *et al.* Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* **43,** e37 (2015).

286. Edgar, R. C. *et al.* Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31,** 3476–3482 (2015).

287. Chiu, C.-H. & Chao, A. Estimating and comparing microbial diversity in the presence of

sequencing errors. *PeerJ* **4,** e1634 (2016).

288. Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N. & Quince, C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17,** 125 (2016).
289. Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10,** 57–9 (2013).
290. Rosen, M. J., Callahan, B. J., Fisher, D. S. & Holmes, S. P. Denoising PCR-amplified metagenome data. *BMC Bioinformatics* **13,** 1–16 (2013).
291. Chao, A. & Jost, L. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* **93,** 2533–2547 (2012).
292. Kochar, S. C. & Korwar, R. On random sampling without replacement from a finite population. *Annals of the Institute of Statistical Mathematics* **53,** 631–646 (2001).
293. Davis, M. M. T cell receptor gene diversity and selection. *Annu. Rev. Biochem* **59,** 475–96 (1990).
294. Ratcliffe, M. J. H. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev. Comp. Immunol.* **30,** 101–18 (2006).
295. McCormack, W. T. *et al.* Selection for B cells with productive IgL gene rearrangements occurs in the bursa of Fabricius during chicken embryonic development. *Genes Dev.* **3,** 838–47 (1989).
296. Gong, J. *et al.* 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: From crops to ceca. *FEMS Microbiol. Ecol.* **59,** 147–157 (2007).
297. Qu, A. *et al.* Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS One* **3,** e2945 (2008).
298. Zhu, X. Y., Zhong, T., Pandya, Y. & Joerger, R. D. 16S rRNA-based analysis of microbiota from the cecum of broiler chickens. *Appl. Environ. Microbiol.* **68,** 124–37 (2002).
299. Magurran, A. E. *Measuring biological diversity*. (Blackwell Pub, 2004).
300. Pylro, V. S. *et al.* Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *J. Microbiol. Methods* **107,** 30–37 (2014).
301. Mwangi, W. N. *et al.* Regional and global changes in TCRalphabeta T cell repertoires in the gut are dependent upon the complexity of the enteric microflora. *Dev. Comp. Immunol.* **34,** 406–17 (2010).
302. Merkle, H., Cihak, J., Mehmke, S., Stangassinger, M. & Lösch, U. Helper activity of chicken V beta 1+ and V beta 2+ alpha beta T cells for in vitro IgA antibody synthesis. *Immunobiology* **197,** 543–9 (1997).

303. Cihak, J. *et al.* T cells expressing the V beta 1 T-cell receptor are required for IgA production in the chicken. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 10951–5 (1991).

304. Das, S., Nikolaidis, N., Klein, J. & Nei, M. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16647–52 (2008).

305. Reynaud, C. A., Anquez, V. & Weill, J. C. The chicken D locus and its contribution to the immunoglobulin heavy chain repertoire. *Eur. J. Immunol.* **21,** 2661–70 (1991).

306. Mansikka, A. Chicken IgA H chains. Implications concerning the evolution of H chain genes. *J. Immunol.* **149,** 855–861 (1992).

307. Dunon, D., Schwager, J., Dangy, J. P., Cooper, M. D. & Imhof, B. A. T cell migration during development: homing is not related to TCR V beta 1 repertoire selection. *EMBO J.* **13,** 808–15 (1994).

308. Clarke, K. R., Somerfield, P. J. & Chapman, M. G. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J. Exp. Mar. Bio. Ecol.* **330,** 55–80 (2006).

309. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129,** 271–280 (2001).

310. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149,** 1578–93 (2012).

311. Lozupone, C., Stomabaugh, J., Gordon, J., Jansson, J. & Knight, R. Diversity , stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).

312. De Cruz, P. *et al.* Characterization of the Gastrointestinal Microbiota in Health and Inflammatory Bowel Disease. *Inflamm. Bowel Dis.* **18,** 372–390 (2012).

313. McCann, K. S. The diversity-stability debate. *Nature* **405,** 228–233 (2000).

314. Salanitro, J. P., Blake, I. G. & Muirhead, P. A. Studies on the cecal microflora of commercial broiler chickens. *Appl. Microbiol.* **28,** 439–47 (1974).

315. MacArthur Award Lecture Author, R. H., Levin, S. A., Robert Macarthur Award Lecture, T. H., Simon Levin, by A. & Levin MacArthur Award Recipient, S. A. The Problem of Pattern and Scale in Ecology. *Source Ecol. Ecol.* **73,** 1943–1967 (1992).

316. Větrovský, T. & Baldrian, P. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* **8,** e57923 (2013).

317. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14,** 667–685 (2014).

318. Koenig, W. D. Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.* **14,** 22–26 (1999).

319. Dormann, C. F. Effects of incorporating spatial autocorrelation into the analysis of

species distribution data. *Glob. Ecol. Biogeogr.* **16,** 129–138 (2007).

320. Morin, P. J. (Peter J. *Community ecology*. (Wiley, 2011).
321. Grilli, J. *et al.* Feasibility and coexistence of large ecological communities. *Nat. Commun.* **8,** 1–8 (2017).
322. Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483,** 205–208 (2012).
323. Donohue, I. *et al.* Navigating the complexity of ecological stability. *Ecol. Lett.* **19,** 1172–1185 (2016).
324. Faith, J. J. *et al.* The Long-Term Stability of the Human Gut Microbiota. *Science (80-.).* **341,** 1237439–1237439 (2013).
325. Martínez, I., Muller, C. E. & Walter, J. Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. *PLoS One* **8,** e69621 (2013).
326. Huse, S. M. *et al.* Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PLoS Genet.* **4,** e1000255 (2008).
327. Schokker, D. *et al.* Perturbation of microbiota in one-day old broiler chickens with antibiotic for 24 hours negatively affects intestinal immune development. *BMC Genomics* **18,** 241 (2017).
328. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395,** 794–798 (1998).
329. Neutel, A.-M., Heesterbeek, J. A. P. & De Ruiter, P. C. Stability in real food webs: weak links in long loops. *Science* **296,** 1120–3 (2002).
330. Berlow, E. L. Strong effects of weak interactions in ecological communities. *Nature* **398,** 330–334 (1999).
331. Tang, S., Pawar, S. & Allesina, S. Correlation between interaction strengths drives stability in large ecological networks. *Ecol. Lett.* **17,** 1094–1100 (2014).
332. Neutel, A.-M. *et al.* Reconciling complexity with stability in naturally assembling food webs. *Nature* **449,** 599–602 (2007).
333. Bairey, E., Kelsic, E. D. & Kishony, R. High-order species interactions shape ecosystem diversity. *Nat. Commun.* **7,** 12285 (2016).
334. May, R. M. (Robert M. *Stability and complexity in model ecosystems*. (Princeton University Press, 2001).
335. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395,** 794–798 (1998).
336. de Ruiter, P. C., Neutel, A.-M. & Moore, J. C. Energetics, Patterns of Interaction Strengths, and Stability in Real Ecosystems. *Science (80-.).* **269,** 1257–1260 (1995).

337. Jacquet, C. *et al.* No complexity–stability relationship in empirical ecosystems. *Nat. Commun.* **7,** 12573 (2016).

338. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31,** 343–366 (2000).

339. Emmerson, M. & Yearsley, J. M. Weak interactions, omnivory and emergent food-web properties. *Proceedings. Biol. Sci.* **271,** 397–405 (2004).

340. McNaughton, S. J. Ecology of a Grazing Ecosystem: The Serengeti. *Ecol. Monogr.* **55,** 259–294 (1985).

341. Tilman, D. Biodiversity: Population Versus Ecosystem Stability. *Ecology* **77,** 350–363 (1995).

342. Wootton, J. T. & Emmerson, M. Measurement of Interaction Strength in Nature. *Annu. Rev. Ecol. Evol. Syst.* **36,** 419–444 (2005).

343. Clavel, T., Lagkouvardos, I., Blaut, M. & Stecher, B. The mouse gut microbiome revisited: From complex diversity to model ecosystems. *Int. J. Med. Microbiol.* **306,** 316–327 (2016).

344. Simon, K. *et al.* Long-term effects of early life microbiota disturbance on adaptive immunity in laying hens. *Poult. Sci.* **95,** 1543–54 (2016).

345. Videnska, P., Sisak, F., Havlickova, H., Faldynova, M. & Rychlik, I. Influence of Salmonella enterica serovar Enteritidis infection on the composition of chicken cecal microbiota. *BMC Vet. Res.* **9,** 140 (2013).

346. Schwab, C. *et al.* Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J.* **8,** 1101–14 (2014).

347. Dill-McFarland, K. a *et al.* Hibernation alters the diversity and composition of mucosa-associated bacteria while enhancing antimicrobial defence in the gut of 13-lined ground squirrels. *Mol. Ecol.* **23,** 4658–69 (2014).

348. Videnska, P. *et al.* Chicken faecal microbiota and disturbances induced by single or repeated therapy with tetracycline and streptomycin. *BMC Vet. Res.* **9,** 30 (2013).

349. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci.* **108,** 4554–4561 (2010).

350. Macfarlane, S. Antibiotic treatments and microbes in the gut. *Environ. Microbiol.* **16,** 919–924 (2014).

351. Danzeisen, J. L., Kim, H. B., Isaacson, R. E., Tu, Z. J. & Johnson, T. J. Modulations of the Chicken Cecal Microbiome and Metagenome in Response to Anticoccidial and Growth Promoter Treatment. *PLoS One* **6,** e27949 (2011).

352. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12,** R50

(2011).

353. Wickham, H. Reshaping Data with the \pkgreshape Package. *J. Stat. Softw.* **21,** 1–20 (2007).

354. Rabinowitz, D. Seven forms of rarity. in *The biological aspects of rare plant conservation* 205–217 (1981). doi:10.1177/004728702237415

355. Wootton, J. T. & Emmerson, M. Measurement of Interaction Strength in Nature. *Annu. Rev. Ecol. Evol. Syst.* **36,** 419–444 (2005).

356. Yin, Y. *et al.* Exposure of different bacterial inocula to newborn chicken affects gut microbiota development and ileum gene expression. *ISME J.* **4,** 367–376 (2010).

357. Singh, P. *et al.* Influence of penicillin on microbial diversity of the cecal microbiota in broiler chickens. *Poult. Sci.* **92,** 272–276 (2013).

358. Maurice, C. F. *et al.* Marked seasonal variation in the wild mouse gut microbiota. *ISME J.* **9,** 2423–2434 (2015).

359. Imhof, B. A., Dunon, D., Courtois, D., Luhtala, M. & Vainio, O. Intestinal CD8 alpha alpha and CD8 alpha beta intraepithelial lymphocytes are thymus derived and exhibit subtle differences in TCR beta repertoires. *J. Immunol.* **165,** 6716–22 (2000).

360. Schusser, B. *et al.* Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 20170–5 (2013).

361. Garrett, W. S. *et al.* Enterobacteriaceae Act in Concert with the Gut Microbiota to Induce Spontaneous and Maternally Transmitted Colitis. *Cell Host Microbe* **8,** 292–300 (2010).

362. Pilar, A. V. C., Reid-Yu, S. A., Cooper, C. A., Mulder, D. T. & Coombes, B. K. Active modification of host inflammation by Salmonella. *Gut Microbes* **4,** 140–145 (2013).

363. Ott, S. J. *et al.* Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* **53,** 685–93 (2004).

364. Kelly, D. & Mulder, I. E. Microbiome and immunological interactions. *Nutr. Rev.* **70,** S18–S30 (2012).

365. Slack, E., Balmer, M. L. & Macpherson, A. J. B cells as a critical node in the microbiota-host immune system network. *Immunol. Rev.* **260,** 50–66 (2014).

366. Wines, B. D. & Hogarth, P. M. IgA receptors in health and disease. *Tissue Antigens* **68,** 103–14 (2006).

367. Kim, S.-H., Kim, Y. N. & Jang, Y.-S. Cutting Edge: LL-37-Mediated Formyl Peptide Receptor-2 Signaling in Follicular Dendritic Cells Contributes to B Cell Activation in Peyer's Patch Germinal Centers. *J. Immunol.* **198,** 629–633 (2017).

368. Bunker, J. J. *et al.* Innate and Adaptive Humoral Responses Coat Distinct Commensal Bacteria with Immunoglobulin A. *Immunity* **43,** 541–553 (2015).

369. Fagarasan, S. *et al.* Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298,** 1424–7 (2002).

370. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19256–61 (2009).

371. Moor, K. *et al.* High-avidity IgA protects the intestine by enchaining growing bacteria. *Nature* **544,** 498–502 (2017).

372. Johnson, S., Sypura, W. D., Gerding, D. N., Ewing, S. L. & Janoff, E. N. Selective neutralization of a bacterial enterotoxin by serum immunoglobulin A in response to mucosal disease. *Infect. Immun.* **63,** 3166–73 (1995).

373. Tokuhara, D. *et al.* Secretory IgA-mediated protection against V. cholerae and heat-labile enterotoxin-producing enterotoxigenic Escherichia coli by rice-based vaccine. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8794–9 (2010).

374. Mazanec, M. B., KAETZELt, Charlot. S., LAMMt, M. E., FLETCHERt, D. & NEDRUDt, J. G. Intracellular neutralization of virus by immunoglobulin A antibodies (mucosal immunity/host defense). *Immunology* **89,** 6901–6905 (1992).

375. Taminiau, J. a. Protective effect of rotavirus VP-6 specific IgA monoclonal antibodies that lack neutralizing activity. *J. Pediatr. Gastroenterol. Nutr.* **27,** 123 (1998).

376. Lécuyer, E. *et al.* Segmented Filamentous Bacterium Uses Secondary and Tertiary Lymphoid Tissues to Induce Gut IgA and Specific T Helper 17 Cell Responses. *Immunity* **40,** 608–620 (2014).

377. Harris, N. L. *et al.* Mechanisms of Neonatal Mucosal Antibody Protection. *J. Immunol.* **177,** 6256–6262 (2006).

378. Bevins, C. L. & Salzman, N. H. The potter's wheel: the host's role in sculpting its microbiota. *Cell. Mol. Life Sci.* **68,** 3675–85 (2011).

379. Wei, M. *et al.* Mice carrying a knock-in mutation of Aicda resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense. *Nat. Immunol.* **12,** 264–70 (2011).

380. Cunningham-Rundles, C. Physiology of Iga and Iga deficiency. *J. Clin. Immunol.* **21,** 303–309 (2001).

381. Macpherson, A. J. & Uhr, T. Induction of Protective IgA by Dendritic Cells Carrying Commensal Bacteria. *Science (80-.).* **1662,** 1–8 (2004).

382. Fransen, F. *et al.* BALB/c and C57BL/6 Mice Differ in Polyreactive IgA Abundance, which Impacts the Generation of Antigen-Specific IgA and Microbiota Diversity. *Immunity* **43,** 527–540 (2015).

383. Rey, J., Garin, N., Spertini, F. & Corthésy, B. Targeting of secretory IgA to Peyer's patch

dendritic and T cells after transport by intestinal M cells. *J. Immunol.* **172,** 3026–33 (2004).

384. Zhang, Y. *et al.* Germinal center B cells govern their own fate via antibody feedback. *J. Exp. Med.* **210,** 457–64 (2013).
385. Johansen, F. E. *et al.* Absence of epithelial immunoglobulin A transport, with increased mucosal leakiness, in polymeric immunoglobulin receptor/secretory component-deficient mice. *J. Exp. Med.* **190,** 915–22 (1999).
386. Josefowicz, S. Z., Lu, L.-F. & Rudensky, A. Y. Regulatory T Cells: Mechanisms of Differentiation and Function. *Annu. Rev. Immunol.* **30,** 531–564 (2012).
387. Perruzza, L. *et al.* T Follicular Helper Cells Promote a Beneficial Gut Ecosystem for Host Metabolic Homeostasis by Sensing Microbiota-Derived Extracellular ATP. *Cell Rep.* **18,** 2566–2575 (2017).
388. Cervantes-Barragan, L. *et al.* Lactobacillus reuteri induces gut intraepithelial CD4+CD8αα+ T cells. *Science* **357,** 806–810 (2017).
389. Li, M. O., Wan, Y. Y. & Flavell, R. A. T Cell-Produced Transforming Growth Factor-β1 Controls T Cell Tolerance and Regulates Th1- and Th17-Cell Differentiation. *Immunity* **26,** 579–591 (2007).
390. Stavnezer, J. & Kang, J. The Surprising Discovery That TGF Specifically Induces the IgA Class Switch. *J. Immunol.* **182,** 5–7 (2009).
391. Wang, S. *et al.* MyD88 Adaptor-Dependent Microbial Sensing by Regulatory T Cells Promotes Mucosal Tolerance and Enforces Commensalism. *Immunity* **43,** 289–303 (2015).
392. Tsuji, M. *et al.* Preferential generation of follicular B helper T cells from Foxp3+ T cells in gut Peyer's patches. *Science* **323,** 1488–92 (2009).
393. Josefowicz, S. Z. *et al.* Extrathymically generated regulatory T cells control mucosal TH2 inflammation. *Nature* **482,** 395–9 (2012).
394. Cao, A. T. *et al.* Interleukin (IL)-21 promotes intestinal IgA response to microbiota. *Mucosal Immunol.* **8,** 1072–82 (2015).
395. Kohl, K. D. Diversity and function of the avian gut microbiota. *J. Comp. Physiol. B* **182,** 591–602
396. Forder, R. E. A., Howarth, G. S., Tivey, D. R. & Hughes, R. J. Bacterial modulation of small intestinal goblet cells and mucin composition during early posthatch development of poultry. *Poult. Sci.* **86,** 2396–2403 (2007).
397. Kitamura, D., Roes, J., Kühn, R. & Rajewsky, K. A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. *Nature* **350,** 423–426 (1991).

398. Shulzhenko, N., Morgun, A. & Matzinger, P. Spontaneous mutation in the Cd79b gene leads to a block in B-lymphocyte development at the C′ (early pre-B) stage. *Genes Immun.* **10,** 722–726 (2009).

399. Hasan, M., Polic, B., Bralic, M., Jonjic, S. & Rajewsky, K. Incomplete block of B cell development and immunoglobulin production in mice carrying the muMT mutation on the BALB/c background. *Eur. J. Immunol.* **32,** 3463–3471 (2002).

400. Manuscript, A. Mu-deficient mice possess B1 celss and produce IgG and IgE, but not IgA, following systemic sensitization and inhalational challenge in a fungal asthma model. **189,** 1322–1329 (2013).

401. Dobbs, A. K. *et al.* Cutting Edge: A Hypomorphic Mutation in Ig (CD79b) in a Patient with Immunodeficiency and a Leaky Defect in B Cell Development. *J. Immunol.* **179,** 2055–2059 (2007).

402. Cerutti, A., Chen, K. & Chorny, A. Immunoglobulin responses at the mucosal interface. *Annu. Rev. Immunol.* **29,** 273–93 (2011).

403. Macpherson, A. J., McCoy, K. D., Johansen, F.-E. & Brandtzaeg, P. The immune geography of IgA induction and function. *Mucosal Immunol.* **1,** 11–22 (2008).

404. J. Sansonetti, P., Macpherson, A. J. & McCoy, K. D. Stratification and compartmentalisation of immunoglobulin responses to commensal intestinal microbes. *Semin. Immunol.* **25,** 358–363 (2013).

405. Plotkin, J. B. & Dushoff, J. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc. Natl. Acad. Sci.* **100,** 7152–7157 (2003).

406. Gigord, L. D. B., Macnair, M. R. & Smithson, A. Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid Dactylorhiza sambucina(L.) Soo. *Symp. A Q. J. Mod. Foreign Lit.* **98,** 6253–6255 (2001).