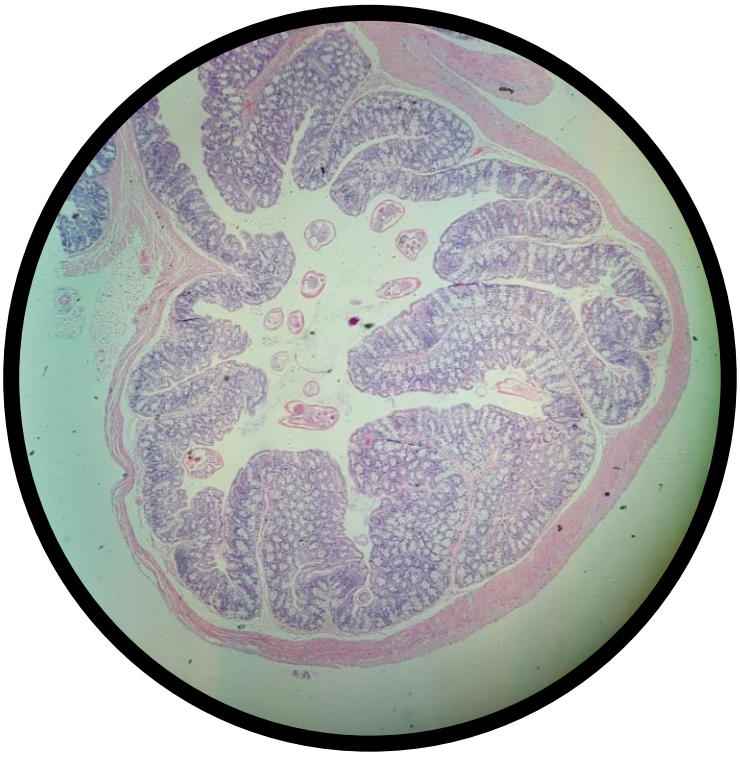
UNIVERSITY OF LIVERPOOL

The gut eukaryome of wild house mice



Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Simon James Hunter-Barnett

April 2023

Abstract

The gut microbiome is an important aspect of mammalian biology, contributing to host phenotype and fitness. To date, much of our understanding has focussed on the bacterial taxa found within the mammalian gut (the bacteriome). However, there is now emerging evidence that the microbial eukaryotes in the gut (the eukaryome) also have an important effect on the biology of the host. The taxonomic composition of the mammalian gut eukaryome is highly varied among hosts, but few studies have explored the drivers behind such variation in wild mammals. In this thesis, I study wild rodents, focussing particularly on house mice (*Mus musculus domesticus*), to extend our knowledge of what factors contribute to variation in gut eukaryome composition. I first use a methodical literature review to describe the gut protozoa of wild rodents and test the hypothesis that host sociality and behaviour affects the prevalence of gut protozoa. I found that rodent species differed in the number of protozoa genera they were host to, and that protozoa prevalence varied among host species. However, contrary to the hypothesis, variation in prevalence was not explained by host sociality and behaviour traits. Next, I investigate the use of flow cytometry to better characterise the gut eukaryome, by quantifying eukaryotic taxa identified by a eukaryotic-specific stain. I tested the accuracy of the proposed quantification method by sequencing the putative eukaryotic cells to assign taxonomy. The findings of this study showed that the method could not accurately identify eukaryotic cells. Finally, I characterise the gut eukaryome of wild house mice from three locations using 18S rRNA amplicon sequencing to determine the host factors associated with eukaryome composition. I observed that the presence of parasitic nematodes and *Eimeria* was associated with eukaryome diversity and composition, as was the host's immune and disease state, specifically gut inflammation and faecal IgA concentration. In sum, my thesis has highlighted some of the potential factors driving the gut eukaryome composition of wild rodents. In doing so, these findings provide a better understanding of the mammalian gut eukaryome, which can be incorporated in future studies of host-gut microbiome interactions.

Acknowledgements

I would first like to thank my supervisor, Mark Viney. His advice and support over the last few years have undoubtedly contributed to my development as a scientist, and I have learned much under his guidance. From day 1, Mark encouraged me to believe in myself and his excitement in our findings was always inspiring. I would also like to thank Jane Hurst for her insightful comments and support over the years. Additionally, thanks to NERC and the ACCE DTP for funding my PhD.

Thank you also to Louise Cheynel. From long hours in the lab pipetting together to sharing our frustrations at yet another day going by without catching any mice, her moral support, scientific advice and friendship made my PhD journey all the better. I would also like to extend this thanks to other members of the Viney lab group for their friendship and scientific discussions over many, many Zoom meetings.

Thanks to those colleagues in Liverpool who have assisted me throughout my PhD. Specifically, Liam Dougherty for his tremendous help and patience with my barrage of questions about “how to do a meta-analysis” and Lukasz Lukomski for his expertise, advice and willingness to traipse through pig poo just to catch a few mice. I would also like to thank Lorenzo Ressel and the pathology team for their help with my histology work. Additionally, thanks to Sandra Cachinho and Chris Law for their help and advice with flow cytometry work.

Thank you to all the other PhD students who have shared this experience with me and have offered both scientific and some definitely-not-scientific discussions. In particular, special thanks to Sophie, Flora, and Georgie – we started this journey together and I doubt I would have made it this far without you. Special thanks also to Jordan. Ever since you showed me how to make fly food all those years ago, you’ve always been there to offer me advice in times of need.

Thanks also to my family. Their constant questions of “what’s the point of it?” and “when are you finishing?” have inspired me to get to this point, and your encouragement and positivity were always welcome. Lastly, I would like to thank Danny. From patiently listening whilst I drone on about fungi and stats to pretty much keeping me fed and hydrated over these last few months, your unwavering positivity and encouragement has kept me sane throughout the highs and lows of this PhD.

Table of Contents

List of figures

List of tables

Chapter 1: General introduction

1.1 The gut microbiome

1.1.1 The microbiome concept

Most organisms play host to a large community of microbes that use their host as their primary niche (Turner *et al.*, 2013; Douglas, 2014; Hammer *et al.*, 2019). These microbial communities are described as the host's microbiota, a term used interchangeably with "microbiome", which represents the microbiota, and their collective genomes and environment (Marchesi and Ravel, 2015). Within a host, multiple sites are colonised by microbes, thus leading to distinct communities at each site (Costello *et al.*, 2009; Turner *et al.*, 2013; Reynoso-García *et al.*, 2022). For example, *Staphylococcus* and *Streptococcus* are the dominant genera in the human skin microbiome, whereas *Clostridium, Bacteroides, and Lactobacillus* are dominant in the gastrointestinal tract (Reynoso-García *et al.*, 2022).

In mammals, the majority of microbes are found within the gastrointestinal tract (gut) (Turnbaugh *et al.*, 2007; Sender *et al.*, 2016). Studies on gut microbiome research were previously limited to microbes that could be cultured *in vitro* from faecal samples (Moore *et al.*, 1969; Wilson and Blitchington, 1996). However, this method could not identify unculturable species, such that the diversity of the gut microbiome composition was previously unknown (Suau *et al.*, 1999; Gupta *et al.*, 2019). However, the development of DNA sequencing technologies revolutionised gut microbiome studies: 16S rRNA gene sequencing of DNA extracted from faecal samples identified novel bacterial species that had not yet been described from culture-based methods (Suau *et al.*, 1999). The development of high-throughput sequencing methods led to a plethora of studies characterising the gut microbiome of humans and other mammals (Eckburg *et al.*, 2005; Qin *et al.*, 2010; Kieser *et al.*, 2022). These studies have established two key concepts in the field of gut microbiomes: i) the interaction between the microbiome and its host is a key component of the host's development and physiology and ii) there is a large amount of intra- and inter-individual heterogeneity in gut microbiome community composition (Lozupone *et al.*, 2013; Sommer and Bäckhed, 2013; Falony *et al.*, 2016).

1.1.2 Importance of the gut microbiome

First, to address the importance of the host-gut microbiome interaction. Often described as the "forgotten organ", the gut microbiome represents a key aspect of a host's biology (O'Hara and Shanahan, 2006; Baquero and Nombela, 2012; Sommer and Bäckhed, 2013). Many host traits are affected by the gut microbiome, and these traits can be grouped into three main categories: metabolism, immunity and disease, and behaviour (Sommer and Bäckhed, 2013). To this end, the following section will review how the microbiome impacts the host for each of

these three groups in turn, drawing examples primarily from studies in humans and laboratory animals, where there has been the most research effort.

The metabolic capabilities of the gut microbiome can provide nutrients to the host, which would otherwise be inaccessible (Rowland *et al.*, 2018). For example, the gut microbiome catabolises plant material into short-chain fatty acids (SCFAs), which are a key energy source to the host (LeBlanc *et al.*, 2017). Microbial-SCFA production is particularly important in ruminants, as a high proportion of energy is derived from microbial fermentation of food (Flint *et al.*, 2008; Williams *et al.*, 2020). Differences in digestive efficiency of mammals can be attributed to differences in the SCFAs produced by the gut microbiome: cows with a lower diversity of bacterial species in the rumen can extract more energy from their feed, compared to cows with a higher bacterial diversity (Shabat *et al.*, 2016). These findings were largely attributed to production of a greater ratio of more nutritionally valuable SCFAs by the lower diversity microbiome, compared to the higher diversity microbiome (Shabat *et al.*, 2016). The production of metabolites by the gut microbiome has also been suggested to promote the development of obesity (Turnbaugh *et al.*, 2006). For example, obesity was associated with a decrease in bacterial diversity in humans, and obesity-enriched genes were from Actinobacteria (Turnbaugh *et al.*, 2009).

The gut microbiome can also be an important source of amino acids and vitamins for the host (Smith *et al.*, 2007; Bergen, 2015). For example, enzymes used in amino acid synthesis were enriched in herbivore microbiomes compared to carnivore microbiomes, suggesting a reliance on the gut microbiome for amino acids (Muegge *et al.*, 2011). Conversely, amino acid degradation pathways were increased in carnivore microbiomes compared to herbivore microbiomes, suggesting their gut microbiome have specialised to degrade proteins as an energy source (Muegge *et al.*, 2011). However, more studies are needed to confirm the gut microbiome is allowing the host to access different diets, rather than the gut microbiome adapting to changes in the host's diet (Dearing and Kohl, 2017). For example, desert woodrats were no longer able to consume the toxic creosote bush, and had a lower body mass, after being treated with antibiotics compared to the controls (Kohl *et al.*, 2014). This suggests that the woodrat's gut microbiome allows it to feed on the creosote toxins (Kohl *et al.*, 2014).

The gut microbiome is also important for the development of the immune system in mammals (Belkaid and Hand, 2014; Gensollen *et al.*, 2016). Gut microbial colonisation during early life develops the immune system, promotes immune homeostasis and is linked to higher disease resistance compared to individuals colonised later in life (Geuking *et al.*, 2011; Mulder *et al.*, 2011; El Aidy *et al.*, 2013). Furthermore, the maturation of the immune system is dependent on host-specific microbes: mice inoculated with microbes from the rat and human gut

microbiome had lower T-cell activation and proliferation compared to mice inoculated with microbes from other mice (Chung *et al.*, 2012). Properties of the immune system modulated by the gut microbiome include the gut mucus layer, the lymphoid structure, immune cell differentiation, and the production of immune mediators (reviewed by Sommer and Bäckhed, 2013).

Further to immune development and modulation, many diseases are now linked to the changes in the gut microbiome composition. For example, humans with chronic diseases such as coeliac disease, inflammatory bowel disease (IBD) and atopic eczema have a different gut microbiome diversity when compared with healthy individuals (Wang *et al.*, 2008; Schippa *et al.*, 2010; De Palma *et al.*, 2017). Transplantation of the faecal microbiome from humans with IBD into mice caused faster gastrointestinal transit time and the development of intestinal barrier dysfunction in the recipient mice (De Palma *et al.*, 2017). The gut microbiome can also contribute to pathogen resistance (Brown and Clarke, 2017; Leung *et al.*, 2018b). This can be indirectly *via* the host immune system or directly *via* resource competition and predation (Stevens *et al.*, 2021). For example, the mouse protozoa *Tritrichomonas musculus* indirectly protects the host against *Salmoenlla* infection by inducing inflammasome-driven IL-18 release, which protects against mucosal infection (Chudnovskiy *et al.*, 2016). In contrast, the gut bacteria *Escherichia coli* and *Bacteroides thetaiotaomicron* can directly protect against gut colonisation of *Citrobacter rodentium,* a pathogenic bacterium in mice, by competing for dietary carbohydrates (Kamada *et al.*, 2012).

The gut microbiome can also impact host behaviour and neurocognition (Bercik *et al.*, 2011; Heijtz *et al.*, 2011; Vuong *et al.*, 2017). Research has shown that the gut microbiome can alter the development of stress responses in mice, with a greater stress response seen in mice raised without a gut microbiome compared to mice with a gut microbiome (Sudo *et al.*, 2004). Other work has shown that mice treated with antimicrobial drugs, altering their gut microbiome composition, had increased exploratory behaviour and hippocampal protein expression, compared to untreated mice (Bercik *et al.*, 2011). Furthermore, sociality is linked to the gut microbiome: offspring of mice fed high-fat diets were less social, displaying fewer reciprocal social interactions compared to offspring of mice fed normal diets (Buffington *et al.*, 2016). This finding was attributed to a lower gut microbiome diversity in the offspring of mice fed high-fat diets (Buffington *et al.*, 2016). Similarly, submissive mice are shown to have a less diverse gut microbiome compared to dominant mice (Agranyoni *et al.*, 2021).

1.1.3 Differences in gut microbiome composition among hosts

Second, to address the question of heterogeneity in gut microbiome community composition. Different host species have different gut microbiomes *i.e.* conspecifics will have a microbiome more similar to each other than to that of other host species (Ley *et al.*, 2008). For mammals, differences in the gut microbiome composition among host species are often associated with the phylogeny of the host (Youngblut *et al.*, 2019; Mallott and Amato, 2021). This is likely because of the continual co-evolution of the gut microbiome and host (Moeller *et al.*, 2016a, but see Groussin *et al.*, 2020). The effect of phylogeny is often attributed to closely-related host species sharing similar ecological niches and diets, but the relative contribution of diet and phylogeny is debated (Youngblut *et al.*, 2019; Mallott and Amato, 2021). In primates for example, the influence of host phylogeny on gut microbiome composition was shown to be more important than diets (Ochman *et al.*, 2010; Amato *et al.*, 2019), and the giant panda's gut microbiome is more similar to that of other (non-herbivorous) bears than it is to other herbivores (Xue *et al.*, 2015). In contrast, distantly related myrmecophagous (ant-feeding) mammals have similar gut microbiome compositions (Delsuc *et al.*, 2014), suggesting that dietary effects on gut microbiome composition are greater than phylogeny.

However, even among individuals within a host species there are differences in the gut microbiome community composition (Linnenbrink *et al.*, 2013; Tung *et al.*, 2015; Weldon *et al.*, 2015; Moeller *et al.*, 2016b; Goertz *et al.*, 2019; Suzuki *et al.*, 2019a). In particular, this is seen when comparing among populations (Linnenbrink *et al.*, 2013; Weldon *et al.*, 2015; Suzuki *et al.*, 2019a) or among laboratory *vs.* captive *vs.* wild animals (Gibson *et al.*, 2019; Rosshart *et al.*, 2019; Bornbusch *et al.*, 2022). Even within populations, there are differences among individuals (Tung *et al.*, 2015; Moeller *et al.*, 2016b; Goertz *et al.*, 2019). Furthermore, the gut microbiome composition can vary over an individual's lifetime. In humans, there are up to 10-fold differences in the abundance of microbial genera between daily faecal samples (Vandeputte *et al.*, 2021), and in meerkats, bacteria were more abundant in the morning, but more diverse in the afternoons (Risely *et al.*, 2021a). Seasonal shifts are also seen in gut microbiome composition in baboons and mice (Björk *et al.*, 2022; Marsh *et al.*, 2022) and the diversity of gut microbial taxa fluctuates throughout the year in lemurs (Murillo *et al.*, 2022). Additionally, changes in the microbiome composition are seen as mammals age. For example, the meerkat microbiome becomes more individualised as the host ages (Risely *et al.*, 2022). In humans, the abundance of bacteria gradually increased following initial gut colonisation at birth, with shifts seen in the dominant microbial taxa as infants aged (Rao *et al.*, 2021).

1.1.4 Drivers of gut microbiome composition

Given that the gut microbiome is an integral part of an animal's biology, knowledge on why the gut microbiome composition differs among and within individual hosts is key for understanding variation in the behaviour and physiology of those hosts (Suzuki, 2017; Henry *et al.*, 2021). Application of metacommunity theory to the gut microbiome indicates that gut community composition is driven by i) the transmission of microbes between hosts, ii) host selection of microbes, and iii) community interactions within the gut (Costello *et al.*, 2012; Scheuring and Yu, 2012; Foster *et al.*, 2017). To this end, the following section will review how the gut microbiome assembly process contributes to heterogeneity in community composition for each of these three processes, with examples from humans, laboratory studies, and wild mammals.

Transmission of microbes is the first process required for the assembly of the gut microbiome (Costello *et al.*, 2012). The initial colonisation of the gut microbiome occurs at birth *via* vertical transmission: mammals are presumed to be sterile *in utero*, but receive a rich community of microbes from their mother during birth (Mueller *et al.*, 2015; Perez-Muñoz *et al.*, 2017). Indeed, the gut microbiome of humans is reflective of their delivery mode: vaginally-born neonates gut microbiomes are more similar to their mother's compared to caesarean-born neonates' (Dominguez-Bello *et al.*, 2010; Wampach *et al.*, 2018). Furthermore, comparison of different bacterial strains in the human infant microbiome has shown that the maternal microbiome is an important source of gut microbes (Korpela *et al.*, 2018; Wampach *et al.*, 2018). In wild gelada baboons, the juvenile gut microbiome composition was more similar to their mothers' microbiome composition than to other adult females, likely mediated through microbial transmission through milk and/or body contact (Baniel *et al.*, 2022). Mother to offspring vertical transmission was shown to be the dominant transmission mode of gut bacteria in mice, with laboratory mice maintaining individual signatures of gut community composition over 10 generations (Moeller *et al.*, 2018).

Horizonal transmission is a further mode of microbial transmission between hosts, through social interactions and shared environments (Moeller *et al.*, 2018; Sarkar *et al.*, 2020). The opportunity for social horizontal transmission increases as hosts become more social, due to increased contact between individuals. Thus, differences in sociality can drive differences in gut microbiome community composition (Sarkar *et al.*, 2020). For example, the gut microbiome of chimpanzees was dictated by their social interactions, with more social individuals having a greater diversity of gut bacteria, and more homogeneity seen within social groups that interact more (Moeller *et al.*, 2016b). Similarly, closely interacting wild horses had more similar gut microbiomes, compared to horses than did not interact as much (Antwis *et al.*, 2018). Social interactions and close associations can be more important than genetic relatedness

and kinship for microbiome community assembly (Tung *et al.*, 2015; Rothschild *et al.*, 2018; Raulo *et al.*, 2021). In humans, there were similarities in the gut microbiome composition of co-habiting, though non-related, individuals (Rothschild *et al.*, 2018). In contrast, similarity in the microbiome community composition of related individuals that did not cohabit was no greater than the similarity of un-related individuals that did not cohabit (Rothschild *et al.*, 2018). However, the importance of sociality can vary among species. For example, repeated interactions over two months between two previously unfamiliar pairs of naked mole rats did not cause their gut microbiome to become more similar (Fitzpatrick *et al.*, 2022).

It is important to note that the mechanisms underlying similarities in the gut microbiome among closely interacting hosts are unclear (Sarkar *et al.,* 2020). It is indeed possible that direct interactions between hosts, *e.g.* physical contact such as grooming, facilitates the transfer of microbes from host to another. However, it is also likely that indirect interactions, such as shared environments and diets, are driving similarities in the microbiome among closely interacting hosts. Non-group-living species can be used to disentangle these effects, as there is less indirect interaction compared to group-living species (Raulo *et al.,* 2021). For example, social interactions in wood mice have been shown to be more important than both kinship and environmentally-acquired microbes in shaping the gut microbiome community (Raulo *et al.*, 2021).

The contribution of horizontal transmission to the microbiome community composition is limited by the chance of transmission occurring between hosts. This is demonstrated by spatially-separated populations: as physical distance increases between hosts, their microbiomes become more dissimilar (Moeller *et al.*, 2017). However, the opportunity for horizontal transmission can also be dependent on the lifestyle and behaviour of the host. For example, there is less opportunity for horizontal transmission in arboreal species compared to terrestrial species, as there is less incidental contact with faeces (Sarkar *et al.*, 2020). This was suggested as a potential mechanism for the greater dissimilarity in gut microbiome composition among arboreal mammals compared to the gut microbiome similarity of terrestrial mammals (Perofsky *et al.*, 2019; Barelli *et al.*, 2020b). However, this could also be driven by differences in substrate use: soil properties were the dominant driver of gut microbiome composition in baboons, potentially from different soils transmitting a different suite of microbes to the host (Grieneisen *et al.*, 2019). Predation can also increase the potential for transmission of gut microbes: predator-prey host-species pairs had more similar gut microbiomes than non-predator-prey host-species pairs (Moeller *et al.*, 2017). Another behaviour important for microbiome transmission is coprophagy. Specifically, coprophagy of the mother's faeces is commonly seen in young herbivores, presumably to encourage beneficial microbial species to colonise the gut microbiome (Soave and Brand, 1991).

Coprophagic behaviour has been shown to maintain diversity in the gut microbiome, potentially allowing re-infection of the host with beneficial microbes (Bo *et al.*, 2020).

Following microbial transmission, the host then plays a role in "selecting" which taxa can survive in the gut microbiome (Foster *et al.*, 2017). In mammals, host-mediated microbiome composition is largely dictated by the immune response of the host (Tanoue *et al.*, 2010; Belkaid and Hand, 2014). In particular, the immune system allows the continual presence of commensal microbes, whilst targeting those microbes that are detrimental to the host (Brugman and Nieuwenhuis, 2010; Tanoue *et al.*, 2010). One theory suggests that the immune system of neonates is primed to recognise commensal bacteria *via* the translocation of bacterial components from the maternal gut to breast milk (Perez *et al.*, 2007). Variation in how the host's immune system responds to microbes, impacted by both ecological and genetic factors, can drive variation in community composition (Costello *et al.*, 2012). Ecological factors that are known to be important for regulating a host's immune state include early life development, body condition, age, sex, and being in a physiologically-demanding state *e.g.* malnutrition, as well as prior exposure to infections (Viney *et al.*, 2005; MacGillivray and Kollmann, 2014; Abolins *et al.*, 2018; Kelly *et al.*, 2018; Clerc *et al.*, 2019a). Alternatively to ecological factors, the host genotype is also associated with how the immune system selects for specific microbial taxa (Benson *et al.*, 2010; McKnite *et al.*, 2012; Blekhman *et al.*, 2015; Suzuki *et al.*, 2019b). For example, the variation in the presence and abundance of Prevotellaceae bacteria was associated with the differential expression of different alleles of the *Tgfb3* gene (McKnite *et al.*, 2012). The *Tgfb3* gene is linked to intestinal barrier function barrier and tolerance to commensal bacteria (McKnite *et al.*, 2012). In wild animals, how the immune state of the host contributes to the gut microbiome composition has largely been overlooked (Viney and Riley, 2017; Kwon and Seong, 2021; Thomson *et al.*, 2022). However, the use of semi-wild enclosures has shown gut microbiome composition changes are associated with both immune state and parasite infection (Leung *et al.*, 2018a; Graham, 2021).

Gut microbiome community composition is not only dependent on the host's immune system, but also on the microbial interactions within the microbiome (Costello *et al.*, 2012; Foster *et al.*, 2017). Early colonisers of the gut microbiome can act as ecosystem engineers and modulate which microbes can subsequently establish there (Bokulich *et al.*, 2016; Coyte *et al.*, 2021). For example, bacteria colonisation at birth contributes to the development of the rumen and shapes the gut to the specific abiotic conditions required for survival of some ciliated protozoa (Michaiowski, 2005; Gilbert, 2020). There are also a number of studies showing how the gut microbiome composition is linked to parasite infection in wild animals, particularly in mice and primates, albeit with the underlying mechanism harder to interpret

(Kreisinger *et al.*, 2015; Weldon *et al.*, 2015; Leung *et al.*, 2018a; Mann *et al.*, 2020; Montero *et al.*, 2021; Kim *et al.*, 2022). This is reviewed in more detail in Chapter 4 (section 4.1.2).

Antagonistic microbial interactions in the gut can also contribute to community composition (Coyte and Rakoff-Nahoum, 2019). In particular, competition over fluctuating resources is theorised to be the dominant driver of an individual's microbiome community composition (Pereira and Berry, 2017; Ho *et al.*, 2022). For example, the closely related Enterobacteriaceae *C. rodentium* and *E. coli,* compete for carbohydrates which can limit *C. rodentium* colonisation, as mentioned above (section 1.1.2, Kamada *et al.*, 2012). Similarly, the protozoa *Tritrichomonas musculus* competes with bacteria for dietary fibre, a resource essential for *T. musculus* colonisation (Wei *et al.*, 2020). To this end, *T musculus* colonisation can be limited by the presence of certain *Bifidobacterium* spp. (Wei *et al.*, 2020). However, the development of vacant niches in the gut can change the dynamics of microbial interactions, leading to fluctuations in community composition (Pereira and Berry, 2017). Changes in diet are largely responsible for these shifting dynamics, by changing the resources available to microbes, thus generating new niches, which lead to rapid changes in microbial community composition (Muegge *et al.*, 2011; David *et al.*, 2014). Indeed, dietary differences in humans, livestock, and wild mammals have been shown to drive gut microbiome community composition (David *et al.*, 2014; Wang *et al.*, 2014; Henderson *et al.*, 2015). Furthermore, seasonal fluctuations in the gut microbiome community of wild mammals, including wood mice and baboons, have been attributed to changes in diet (Amato *et al.*, 2015; Maurice *et al.*, 2015; Baniel *et al.*, 2021; Björk *et al.*, 2022; Marsh *et al.*, 2022).

1.2. The gut eukaryome

1.2.1 Eukaryotic taxa are comparatively less well researched than bacterial taxa

Many of the examples outlined in the preceding sections are drawn from the literature describing the bacteria found in the gut microbiome (*i.e.* the bacteriome). Indeed, the diversity of the bacteriome is now well-characterised for many host species. For example, Firmicutes and Bacteroidetes are the most common bacterial phyla detected in the mammalian gut microbiome (Ley *et al.*, 2008), with many species within these phyla, and other less common phyla, confirmed to be true residents of the gut microbiome (Rajilić-Stojanović and de Vos, 2014). However, whilst other taxa, specifically fungi, protozoa, helminths, viruses, and archaea, are known to be present in the mammalian microbiome, their diversity, residency, and relative importance to the host is comparatively less understood than their bacterial counterparts (Nkamga *et al.*, 2017; Desselberger, 2018; Fiers *et al.*, 2019; del Campo *et al.*, 2020). Indeed, comparison of search results in the Web of Science database indicate the disparity in research effort between the bacteriome and other gut microbes (Table 1.1). These differences have been shown previously (Hooks and O'Malley, 2020; Weiner *et al.*, 2023). The

following section will focus on the eukaryotic gut taxa (fungi, protozoa, and helminths, collectively termed the eukaryome), first addressing the disparity in research effort between the eukaryome and the bacteriome, and then summarising the current knowledge of the gut eukaryome in humans and wild mammals.

Table 1.1. The number of publications returned from the Web of Science database for different microbes found within the gut microbiome. The search term used was "x gut microbiome" where x is each microbe type. The search date was: 19/03/23.

Microbe	Number of publications
Bacteria	12,715
Archaea	386
Parasite	466
Virus	1,568
Eukaryota	167
Fungi	930
Protozoa/Protist	170/92
Helminth	171

Previous views on gut eukaryotes were that they were either strictly parasitic/pathogenic or strictly mutualistic (Parfrey *et al.*, 2011; Huffnagle and Noverr, 2013; del Campo *et al.*, 2020). To this end, studies in mammals are often split into those researching mutualistic eukaryotes *e.g.* protozoa and fungi in ruminated animals (Orpin, 1984) and those investigating gut parasites *via* traditional diagnostic methods (Hamad *et al.*, 2016). Thus, few studies consider the wider eukaryome as a whole, in contrast to bacteriome studies that describe the collective bacteriome, typically irrespective of the potential pathogenicity of taxa. This is analogous of a time when bacteria were also considered to be primarily pathogenic, with little focus on commensal species (del Campo *et al.*, 2020). However, there is now a consensus emerging that the pathogenicity of eukaryotes in the gut microbiome is context-dependent, leading to increased interest in the eukaryome (Andersen *et al.*, 2013; Lukeš *et al.*, 2015; Chabé *et al.*, 2017; Dubik *et al.*, 2022).

Furthermore, methodologies are not yet standardized in gut eukaryome research, which can make the eukaryome more technically difficult to study, compared to the bacteriome (Amaral-Zettler *et al.*, 2009; Hamad *et al.*, 2016; Popovic *et al.*, 2018; Lind and Pollard, 2021; Thielemann *et al.*, 2022; Vaulot *et al.*, 2022). Firstly, the 16S rRNA gene is routinely accepted as the most appropriate target gene for bacteriome studies (Janda and Abbott, 2007; Louca *et al.*, 2018; Johnson *et al.*, 2019). For the eukaryome however, there is no general consensus on which gene – ITS1, ITS2, 18S, and 28S rRNA are commonly used – nor which primer pair

is the most appropriate to capture eukaryome taxa diversity (Popovic *et al.*, 2018; Frau *et al.*, 2019; Vaulot *et al.*, 2022). Secondly, detection of gut eukaryotic taxa *via* sequencing can be affected by a dominance of DNA from sloughed host cells, which is a lesser issue for bacteriome research (del Campo *et al.*, 2019b; Pereira-Marques *et al.*, 2019; Lind and Pollard, 2021). Thirdly, the quantification of absolute cell number in the gut bacteriome has been shown to improve the interpretation of amplicon sequencing data, with many methods being used to do so (Vandeputte *et al.*, 2017; Wang *et al.*, 2021; Yang and Chen, 2022). However, these methods have been little applied and rarely optimised for eukaryotic taxa (Haak *et al.*, 2021; Rao *et al.*, 2021). This is reviewed in more detail in Chapter 3.

1.2.2 Eukaryome composition and impacts on the host

Gut eukaryome studies are becoming increasingly common in humans and laboratory animals. A pioneer study in 2008, using sequencing methods to identify unculturable taxa, found that there were unexpected taxa in the gut, and their role was unknown (Scanlan and Marchesi, 2008). Since then, many studies have explored the human gut eukaryome composition (Rajilić-Stojanović and de Vos, 2014; Hamad *et al.*, 2016; Desselberger, 2018; Hooks and O'Malley, 2020). The human gut eukaryome is vastly smaller than the gut bacteriome: it has a lower diversity, fewer species, and lower abundances of taxa (Nam *et al.*, 2008; Rajilić-Stojanović and de Vos, 2014; Rao *et al.*, 2021). Additionally, less than 0.1% of the microbial genes found in the human gut microbiome originate from the eukaryome (Qin *et al.*, 2010).

To this end, there is a misconception that gut eukaryotes are less important than gut bacteria due to their reduced abundance (Laforest-Lapointe and Arrieta, 2018). However, the relative abundance of taxa is not necessarily proportional to their importance (Laforest-Lapointe and Arrieta, 2018). Indeed, research into how the eukaryome impacts the host has shown that there a wide range of effects on the host development, immunity, and disease state (Lukeš *et al.*, 2015; Laforest-Lapointe and Arrieta, 2018; Weiner *et al.*, 2023). For example, eukaryotes have been shown to beneficially modulate the host's immune state (Broadhurst *et al.*, 2012; Deng *et al.*, 2022, but see Escalante *et al.*, 2016). Specifically, the presence of the protozoa *Blastocystis* activates Th2 immune responses which can improve the host's response to intestinal injury (Deng *et al.*, 2022). Similarly, infection with the nematode whipworm *Trichuris trichiura* was shown to improve clinical symptoms of colitis in monkeys (Broadhurst *et al.*, 2012).

In the examples above, the beneficial effects of colonisation of *Blastocystis* and *T. trichiura* were linked to interactions with the pre-established gut bacteriome. Interactions between the gut eukaryome and bacteriome can also contribute to disease resistance: for example,

Blastocystis can reduce the growth of the *Bacteroides vulgatus*, a pathogenic bacterium, by inducing oxidative stress (Deng and Tan, 2022). Additionally, the protozoa *T. musculus* protects against *Salmonella* infections, as mentioned previously (section 1.1.2, Chudnovskiy *et al.*, 2016). In contrast to their beneficial role, inter-kingdom interactions in the gut microbiome can also negatively affect the host. For example, interactions between the bacteria *Serratia marcescence* and *E. coli,* and the fungus *Candia topicalis,* are hypothesised to contribute to Crohn's disease, potentially through the formation of immunostimulatory biofilms in the gut (Hoarau *et al.*, 2016).

The amount of variation in gut eukaryome community composition among humans, and within individuals over time, is substantially different to the gut bacteriome (Parfrey *et al.*, 2014; Hallen-Adams *et al.*, 2015; Nash *et al.*, 2017). Of the few studies characterising the eukaryome of other mammals, comparison among host species has shown that different host species have vastly different gut eukaryome compositions (Parfrey *et al.*, 2014; Mann *et al.*, 2020). Thus, given the impact of the gut eukaryome on the host's biology, there is need to understand what drives variation in gut eukaryome community composition among hosts (as discussed above, section 1.1.3). However, there is less research into the assembly process of the gut eukaryome compared to the bacteriome, particularly which host-mediated factors contribute to eukaryome composition. Surveys in non-human primates indicated that the impact of host phylogeny on gut eukaryome composition is weaker than its impact on gut bacteriome composition (Mann *et al.*, 2020). Instead, Mann *et al.* (2020) suggested that gut eukaryome composition is driven by individual host behaviour and local ecology, but could not verify the particular mechanisms. In contrast, a study comparing the gut mycobiome of 49 host species found that host phylogeny was correlated with mycobiome composition, albeit less strongly than correlation with the gut bacteriome (Harrison *et al.*, 2021). In support of Mann *et al.* (2020), diet was significantly correlated with gut mycobiome diversity in mammals (Harrison *et al.*, 2021). Factors affecting gut eukaryome composition are reviewed in more detail in Chapter 4 (section 4.1.2).

1.3 Wild rodents as a study system

Whilst much of what we know about what drives gut microbiome composition comes from studying laboratory animals, it is now commonly recognised that laboratory animals do not display the same amount of inter-individual differences that wild animals do (Amato, 2013; Viney, 2019; Kuziel and Rakoff-Nahoum, 2022). Further to this, the laboratory animals' gut microbiome has been shown to be distinct to that of their wild counterparts (Rosshart *et al.*, 2019; Bowerman *et al.*, 2021). An alternative to using laboratory born and bred animals is to use wild-caught animals that are then maintained in captivity. However, this also presents problems because captivity changes the gut microbiome of wild animals, and the amount of

change differs among species (Kohl and Dearing, 2014; Rosshart *et al.*, 2017; Rosshart *et al.*, 2019; Schmidt *et al.*, 2019). Thus, there is a need to study the gut microbiome in a more natural setting to understand the main drivers of microbiome composition. To this end, studies in wild animals have become more common in recent years, with longitudinal sampling over an individual's lifetime to monitor which factors are driving the gut microbiome composition (Björk *et al.*, 2019; Grieneisen *et al.*, 2023).

Rodents (order Rodentia) are a highly speciose order of mammals with species representing a wide diversity of different life-histories and behaviours (Fabre *et al.*, 2012). This makes the Rodentia ideal for investigating how behavioural differences among wild species and individuals contribute to eukaryome composition. Further to this, the house mouse, *Mus musculus domesticus,* offers a unique perspective for wild eukaryome studies: *M. m. domesticus* is a key laboratory animal used in many gut microbiome studies, allowing eukaryome studies of wild house mice to be directly comparable to laboratory studies (Rosshart *et al.*, 2019; Viney, 2019). To date, there has been very limited study of the wild house mouse eukaryome, and none of what drives its composition, with focus generally considering fungi, protozoa, and helminths separately (Viney and Riley, 2017; Rosshart *et al.*, 2019; Bendová *et al.*, 2020). This contrasts to the wild bacteriome, which has become more characterised in recent years (Linnenbrink *et al.*, 2013; Weldon *et al.*, 2015; Goertz *et al.*, 2019; Suzuki *et al.*, 2019b). Additionally, the immune response of wild house mice is significantly different to their laboratory counterparts (Viney and Riley, 2017). Thus, consideration of the immune phenotype in wild mice will bring a new perspective to how the immune system is linked to eukaryome composition (Pedersen and Babayan, 2011; Viney and Riley, 2017).

Although ideal for comparison to laboratory studies, it should be noted that there are certain limitations when using house mice as a wild study system. House mice are reported to be much less common than other wild rodent species, and this can lead to a low trapping rate, and thus low sample sizes (Brown, 1953; Flowerdew *et al.,* 2003). This is further exacerbated by heterogeneity in how house mice respond to trapping, which can lead to individuals with certain behaviours and traits being more likely to be trapped, thus biasing sample sizes towards a particular host trait (Crowcroft and Jeffers, 1961; Hurst and Berreen, 1985). The extent to which population structures change over time is also varied among populations. For example, some house mice populations show little variation, whereas others undergo extreme boom or bust dynamics (Pocock *et al.,* 2004; Andreassen *et al.,*2020). This can make it hard to generalise across the species and limits comparability among populations, as well as limiting tractability within a population. Furthermore, the house mouse is considered an invasive species, originally from India and now found with a global distribution (Boursot *et al.,*

1993; Phifer-Rixey and Nachman, 2015). Their invasiveness is reliant on their ability to quickly adapt to novel and diverse environments (Pocock *et al.,* 2004; Phifer-Rixey and Nachman, 2015). Thus, comparability among wild house mice and other wild rodent species in how the gut microbiome responds to environmental and host traits is limited. A final limitation is the low diversity of gut parasites observed in wild house mice compared to other wild mice species (Ehret *et al.,* 2017). If this extends to other taxa in the gut eukaryome, the factors controlling eukaryome composition in wild house mice may not be representative of such factors in other wild rodent species, and indeed other wild mammals.

1.4 Thesis objectives and outline

1.4.1 Addressing the eukaryome knowledge gap

The literature reviewed above has demonstrated that there is i) large contrast between our understanding of the gut bacteriome and the gut eukaryome and ii) the majority of research in wild mammals has focussed on the gut bacteriome, with few studies addressing the eukaryotic taxa also found within the gut. Additionally, the optimisation of amplicon sequencing methods for bacteriome research have not been translated or applied to eukaryome research. From the literature, there is evidence that environmental and host factors contribute to the eukaryome community composition in wild mammals, but further work is needed to clarify these processes. Thus, this thesis primarily aims to determine if the host factors known to impact the bacteriome also extend to the gut eukaryome of wild mammals, using wild rodents as a study system.

1.4.2 Thesis objectives

This thesis had three objectives to further our understanding of the rodent gut eukaryome:

1. Describe the protozoa found in the gut eukaryome of wild rodents and what host traits contribute to their prevalence. This first objective is addressed in the work presented in Chapter 2: **The presence and prevalence of gut protozoa of wild rodents: a meta-analysis**. Here, I conduct a methodical literature search to identify the protozoa commonly described in the wild rodent gut eukaryome. I then use meta-analyses to identify which host traits explain variation in the prevalence of gut protozoa among host species, focussing particularly on host behaviour and sociality.

2. Design and test a novel method for quantifying the abundance of eukaryotic microbes in the gut eukaryome. This method is addressed in Chapter 3: **Using flow cytometry to quantify and describe the gut eukaryome of house mice.** The chapter describes why quantification of microbial cells in the gut microbiome is important, before testing

a novel use of flow cytometry to identify and quantify eukaryotic taxa in the mammalian gut.

3. Characterise the gut eukaryome of wild mice and identify which host traits are associated with gut eukaryome diversity and composition. In Chapter 4: **The eukaryome of wild house mice and their disease state**, I use amplicon sequencing to describe the eukaryotic gut taxa of wild mice from three locations in England. Then, I investigate how the immune and disease state of the mice is associated with eukaryome composition, before comparing these results to their bacteriome.

Chapter 2: The presence and prevalence of gut protozoa of wild rodents: a meta-analysis

Abstract

Gut protozoa can have both negative and positive impacts on their host. Thus, it is important to understand what drives the variation in protozoan diversity seen among host species, and among con-specific individuals. The gut protozoa of rodents are not well-described as research typically focus on parasitic protozoa, overlooking commensal species. Here, I use a methodical literature search to first describe the gut protozoa found in wild rodents before using meta-analyses to identify drivers of variation in protozoa prevalence. In particular, I investigated how host sociality and behaviour might be contributing to the transmission of protozoa, thus driving variation in prevalence. I found that some rodents species were capable of hosting many different protozoa genera, whereas other were host to only a few. I estimated the prevalence of protozoa in wild rodents to be 24%, and found that prevalence was heterogenous among host species. However, I found no evidence that host sociality and behaviour traits were underlying the heterogeneity of prevalence seen among host species. This synthesis of the previous descriptions of gut protozoa found in wild rodents has provided a better understanding of the host-specificity of gut protozoa.

2.1 Introduction

2.1.1 Gut protozoa impact host biology

Eukaryotic taxa are an integral part of the mammalian gut microbiome (Filyk and Osborne, 2016; del Campo *et al.*, 2020), but are often overlooked in host-microbiome studies in favour of prokaryotic taxa (Laforest-Lapointe and Arrieta, 2018). In particular, the colonisation mechanisms of gut protozoa are only just being clarified, and their role in host health and disease is still debated (Chabé *et al.*, 2017; Dubik *et al.*, 2022). Gut protozoa exist across the entire parasitism-mutualism continuum, and so can range from: disease-causing parasitic species to long-term residents of the gut, providing benefits to their host (Lukeš *et al.*, 2015; Dubik *et al.*, 2022).

Gut protozoa that provide nutritional benefits to the host are typically described as mutualistic, and their role is well-documented in ruminants (Williams *et al.*, 2020; Solomon and Jami, 2021). For example, the protozoa species *Eudiplodinium maggii* and *Polyplastron multivesiculatum* contribute to enzymatic degradation of plant polysaccharides in sheep (Béra-Maillet *et al.*, 2005). Gut protozoa can also positively contribute to host disease resistance (Lukeš *et al.*, 2015; Leung *et al.*, 2018a; Dubik *et al.*, 2022). For example, the mouse protozoa *Tritrichomonas musculus* indirectly protects the host against *Salmoenlla* infection *via* inducing inflammasome-driven IL-18 release which protects against mucosal infection (Chudnovskiy *et al.*, 2016). Furthermore, *Blastocystis* ST4 can directly induce oxidative stress in, and thus decrease the growth of, the pathogenic prokaryote *Bacteroides vulgatus* (Deng and Tan, 2022).

Negative interactions between gut protozoa and the host will typically result in the manifestation of gastrointestinal disease *e.g.* gastroenteritis (Huh *et al.*, 2009). Some gut protozoa, *e.g. Giardia* and *Cryptosporidium,* can directly cause disease in the host, by damaging the gut epithelium and mucosa, and causing inflammation (Savioli *et al.*, 2006). Gut protozoa can also indirectly contribute to the host's poor health and disease state, by changing the wider species composition of the gut microbiome (Burgess *et al.*, 2017). For example, the presence of *Blastocystis* is associated with a decrease in abundance of beneficial bacteria, such as the bacterium *Bifidobacterium* which can limit and prevent infections of pathogens (Russell *et al.*, 2011; Yason *et al.*, 2019; Caudet *et al.*, 2022).

It can be difficult to categorise protozoa species as either beneficial or harmful, as the impact of gut protozoa on the host is context-dependent (Parfrey *et al.*, 2011; Lukeš *et al.*, 2015; Sardinha-Silva *et al.*, 2022). For example, the host diet, age, immune status, microbiome, and genotype, as well as the protozoa genotype, can all influence the strength of the interaction between a protozoan species and its host (Thompson and Monis, 2012; Ryan *et al.*, 2014;

Lepczyńska *et al.*, 2017; Dubik *et al.*, 2022). To this end, there is variation within and among protozoa species in how they impact the host. For example, *Blastocystis* individuals can shift from being mutualistic and/or commensal, to becoming pathogenic when the host immune system is compromised (Scanlan *et al.*, 2014).

2.1.2 Common mammalian gut protozoa

The diversity of protozoa found in the gut eukaryome, and their interactions with the host, can vary substantially among and within mammalian species (Parfrey *et al.*, 2011; Langda *et al.*, 2020; Guzzo *et al.*, 2022). A brief summary of the commonly described protozoa found in the mammalian gut is in Table 2.1. The five meta-groups of protozoa described in the following section are supported by two reviews of eukaryotic taxonomy (Ruggiero *et al.*, 2015; Adl *et al.*, 2019). The term “meta-group” is used here to account for differences in the taxonomic rank of these groups.

Table 2.1. Protozoa commonly described from the mammalian gut. The lists of hosts provided are not exhaustive, and aim only to show the diversity of mammalian hosts for each protozoan meta-group in the literature.

Meta-Group	Amoebozoa	Apicomplexa	Ciliophora	Metamonada	Stramenopiles
Examples	*Entamoeba* *Endolimax* *Iodamoeba*	*Eimeria* *Sarcosystis* *Toxoplasma* *Cryptosporidium*	*Balantidium* *Entodinium*	*Giardia* *Spironucleus muris* *Chilomastix mesnilii* *Retortomonas intestinalis* *Dientamoeba fragilis* *Tritrichomonas muris*	*Blastocystis*
Disease potential	Parasitic and commensal species	Mostly parasitic species	Mostly mutualistic species but can be parasitic	Mostly parasitic species but can be commensal	Considered to be an opportunistic pathogen
Hosts	Domestic livestock Fore-gut fermenting herbivores Rodents Primates	Carnivora Insectivora Lagomorpha Livestock Primates Rodentia	Artiodactyla Perissodactyla Primates Proboscida Rodentia	Canids Felids Livestock Primates Rodentia	Artiodactyla Carnivora Marsupials Perissodactyla Primates Rodentia
References	(Lau *et al.*, 2014; Parfrey *et al.*, 2014; Poulsen and Stensvold, 2016; Ragazzo *et al.*, 2018; Cui *et al.*, 2019; Mann *et al.*, 2020; Ai *et al.*, 2021; Dubik *et al.*, 2022)	(Appelbee *et al.*, 2005; Heitlinger *et al.*, 2017; Dubey and Almeria, 2019; Duszynski, 2021; Solarczyk, 2021; Bangoura *et al.*, 2022; Guardone *et al.*, 2022)	(Dehority, 1986; Schuster and Ramirez-Avila, 2008; Henderson *et al.*, 2015; Ahmed *et al.*, 2020; Murillo *et al.*, 2022)	(Dobell, 1935; Imai and Ogimoto, 1988; Silberman *et al.*, 2002; Appelbee *et al.*, 2005; Keeling and Brugerolle, 2006; Kolisko *et al.*, 2008; Parfrey *et al.*, 2011; Jackson *et al.*, 2013; Escalante *et al.*, 2016; Hamad *et al.*, 2016; Collántes-Fernández *et al.*, 2018; Cacciò, 2018; Zhang *et al.*, 2019; Li *et al.*, 2020; Füssy *et al.*, 2021; Dubik *et al.*, 2022; Yildiz and Erdem Aynur, 2022)	(Boreham and Stenzel, 1993; Alfellani *et al.*, 2013; Parfrey *et al.*, 2014)

2.1.2.1 Amoebozoa

Amoebae protozoa are diverse, but only a few genera have been described in the gut eukaryome of mammals (Cavalier-Smith *et al.*, 2015). *Entamoeba* is the most commonly described genus of gut amoebae, with many species identified from humans (Cui *et al.*, 2019; Dubik *et al.*, 2022). The host-specificity of amoebae is highly varied with differences seen at the protozoa species level. For example, *Ent. bovis* is found in artiodactyls, whereas *Ent. coli* and *Ent. hertmanii* are found in primates (Parfrey *et al.*, 2014). Genetic data also indicates the host-specificity of *Entamoeba* spp. is dependent on the particular species/strain in question (Stensvold *et al.*, 2011). The genetic data for other gut amoeba protozoa is limited *e.g. Endolimax* spp. and *Iodaoemba* spp., and so host-specificity is often unknown and/or disputed (Stensvold *et al.*, 2012; Poulsen and Stensvold, 2016).

2.1.2.2 Apicomplexa

Apicomplexan protozoa are a diverse group that are largely parasitic (Morrison, 2009; Mathur *et al.*, 2021). The apicomplexan protozoa in the mammalian gut eukaryome are generally limited to coccidia and *Cryptosporidium* (del Campo *et al.*, 2019a). The coccidian protozoa known to infect mammals were recently reviewed, but the distinction between those found in the gut protozoa compared to other tissues is unclear (Duszynski, 2021). Genera in the family Eimeriidae known to reside within the mammalian gut include *Cyclospora, Cystoisospora, Eimeria* and *Isospora* (Heitlinger *et al.*, 2017; Dubey and Almeria, 2019; Solarczyk, 2021; Bangoura *et al.*, 2022). *Sarcocystis* and *Toxoplasma* are also known from the gut eukaryome of mammals, although many hosts are often considered to be intermediate or dead-end hosts (Guardone *et al.*, 2022). *Cryptosporidium* is the sole genus within the Cryptogregarinorida (Adl *et al.*, 2019) but has a wide range of mammalian hosts, including human and non-human primates, rodents, carnivores and ruminants (Appelbee *et al.*, 2005).

2.1.2.3 Ciliophora

The ciliated protozoa are a diverse, widely-distributed phylum of protozoa, but only the class Litostomatea are found in the gut eukaryome (Foissner *et al.*, 2008; Parfrey *et al.*, 2011). Known for their nutritional role in degrading plant material, Litostomatea protozoa can represent up to 50% of the biomass of the rumen eukaryome (Newbold *et al.*, 2015). Ciliate species found within herbivorous mammals are often host-specific, with entire families of protozoa being found only in one host species *e.g.* Troglodytellidae in gorillas (Dehority, 1986). The most commonly described ciliate protozoa genus is *Balantidium,* found in humans, pigs, rodents and non-human primates (Schuster and Ramirez-Avila, 2008; Ahmed *et al.*, 2020).

2.1.2.4 Metamonada

Several orders of metamonad protozoa can be found in the mammalian gut eukaryome (Table 2.1, Parfrey *et al.*, 2011; Hamad *et al.*, 2016). Firstly, the diplomonads (Order: Diplomonadida) are a speciose group of protozoa, and are often associated with hosts, either as parasites or commensals (Jørgensen and Sterud, 2007; Xu *et al.*, 2016). Perhaps the most commonly studied mammalian-associated diplomonad genus is *Giardia* (Appelbee *et al.*, 2005). Other diplomonad species known to inhabit the mammalian gut eukaryome include *Octomitus* spp. (sister genus to *Giardia*), *Spironucleus muris* and *Enteromonas* spp. (Dobell, 1935; Keeling and Brugerolle, 2006; Kolisko *et al.*, 2008; Jackson *et al.*, 2013). Secondly, the retortamonads (Order Retortamonadida) are represented by two genera, *Chilomastix* and *Retortomonas,* and all but one species are obligate residents of animal guts (Silberman *et al.*, 2002). The most commonly described species are *Chilomastix mesnilii* and *Retortomonas intestinalis*, owing to their presence in humans (Dubik *et al.*, 2022). Tritrichomonadida is a third order of gut protozoa isolated from mammalian hosts (Rajilić-Stojanović and de Vos, 2014; Dubik *et al.*, 2022). Commonly described species include *Dientamoeba fragilis* and *Tritrichomonas spp.* (Escalante *et al.*, 2016; Collántes-Fernández *et al.*, 2018; Cacciò, 2018; Yildiz and Erdem Aynur, 2022). Other metamonad protozoa identified in the mammalian gut eukaryome include *Pentatrichomonas* spp. and *Tetratrichomonas* (Imai and Ogimoto, 1988; Zhang *et al.*, 2019; Li *et al.*, 2020; Dubik *et al.*, 2022).

2.1.2.5 Stramenopiles

Despite being a highly diverse group, the only stramenopiles found within the mammalian gut eukaryome are *Blastocystis* spp. (Parfrey *et al.*, 2014; Derelle *et al.*, 2016). Multiple, genetically diverse *Blastocystis* subtypes have been described with prevalence and host-specificity varying greatly between the subtypes (Stensvold and Clark, 2020; Deng *et al.*, 2021). *Blastocystis* is found in a wide range of mammalian hosts (Table 2.1, Boreham and Stenzel, 1993; Alfellani *et al.*, 2013) and is considered a normal and important component of the human gut eukaryome (Parfrey *et al.*, 2014; Deng *et al.*, 2021).

2.1.3 Mechanisms driving gut protozoa composition

Protozoa predominately rely on faecal-oral routes of transmission among hosts, typically through coprophagy or faecal contamination of food and/or water (Dehority, 1986; Burgess *et al.*, 2017). Some gut protozoa, *e.g.* ciliates, are dependent on the rapid faecal-oral transmission of infective stages (Michaiowski, 2005). In contrast, some species *e.g. Giardia* and *Cryptosporidium,* can form environmentally-resistant cysts or oocysts*,* which can persist in the environment for long periods of time, thus reducing the dependency on rapid transmission (Dumètre *et al.*, 2012).

At the individual level, various host behaviours contribute to the chance of a host encountering and acquiring the infective stage of a protozoa (Kołodziej-Sobocińska, 2019). Foremost, a more social individual with comparatively greater social interactions will have a greater chance of being exposed to the transmission of protozoa (Ezenwa *et al.*, 2016). For example, a meta-analysis showed that male vertebrates with a higher social status have an overall higher parasite infection risk, compared to those with a lower social status, which can be attributed to increased mating effort (Habig *et al.*, 2018). Similarly, increased parent-offspring interactions will increase the offspring's exposure to the parent's existing protozoan community, and this is observed with ciliated protozoan infections in ruminants (Michaiowski, 2005).

Host population structure can also affect protozoa transmission, because as host density increases, the opportunity for transmission of gut protozoa is increased (Ostfeld and Mills, 2008; Ebert, 2013). Host density can also alter individuals' exposure to protozoa, by driving changes in the social organisation of the population and individuals' normal home range (Bertolino *et al.*, 2003; Brei and Fish, 2003; Sanchez and Hudgens, 2019). Host foraging behaviours also affect exposure; for example, foraging on the ground, compared to arboreal and aerial foraging, can increase exposure to environmentally-transmitted protozoa, as is seen with *Entamoeba* in baboons and *Isospora* in birds (Dolnik *et al.*, 2010; Barelli *et al.*, 2020a). Similarly, coprophagy can drive exposure to protozoan infective stages present in faeces and predation can cause exposure *via* ingestion of infected prey (Ezenwa *et al.*, 2016; Malmberg *et al.*, 2021).

An individual's diet, immune state, and pre-existing microbiome can also influence the chance of a protozoan successfully establishing in the gut (Thursby and Juge, 2017; Kołodziej-Sobocińska, 2019; Coyte *et al.*, 2021). Firstly, variation in the host diet can alter nutrient availability, allowing the establishment of different protozoan communities in the gut (Zhang *et al.*, 2022). For example, the relative abundance of the protozoa *Entodinium* in sheep rumen fluid changed in response to different diets (Henderson *et al.*, 2015; Zhang *et al.*, 2022). Secondly, the immune state of the host is also important for affecting the initial establishment and subsequent persistence of protozoa in the gut (Evering and Weiss, 2006; Sardinha-Silva *et al.*, 2022). Long-term co-evolution of protozoa with host species has allowed many protozoa to develop complex mechanisms to either i) be tolerated by and/or ii) evade the host immune response (Zambrano-Villa *et al.*, 2002; Macpherson *et al.*, 2005; Schmid-Hempel, 2009; Tanoue *et al.*, 2010; Sardinha-Silva *et al.*, 2022). Thus, the immune state of the host at the time of transmission is important in affecting the chance of successful establishment of protozoa. Many factors contribute to an individual's immune state, including early life development, body condition, age, sex, and being in a physiologically-demanding state *e.g.*

malnutrition, as well as prior exposure to infections (Viney *et al.*, 2005; MacGillivray and Kollmann, 2014; Abolins *et al.*, 2018; Kelly *et al.*, 2018).

Thirdly, protozoan establishment within the gut can also depend on the microbial species already present in the gut (Coyte *et al.*, 2021). For example, some ciliated protozoa in the rumen microbiome require a pre-established bacteriome to shape the gut to the specific abiotic conditions required for survival (Michaiowski, 2005). Ciliates can also need other microbial species as a source of nutrition; for example*, Entodinium caudatum* engulfs other microbial cells, which is necessary for its survival and growth (Wang *et al.*, 2019). Furthermore, there is often an obligate succession pattern of ciliate protozoa establishment: *Entodinia* spp. establish first, followed by other ciliates, showing that *Entodinia* is a primary coloniser species (Michaiowski, 2005). A species' reliance on another species to be present reduces the chance of successful establishment compared to those species that are not reliant (Coyte *et al.*, 2021). Negative microbial interactions can also affect protozoa establishment (Leung *et al.*, 2018a). Competition for nutrients and other resources results in the generation of niches within the gut, limiting the diversity of protozoa that can establish (Pereira and Berry, 2017). For example, *Tritrichomonas musculus* competes with the bacteriome for dietary fibre, a resource essential for *T. musculus* colonisation (Wei *et al.*, 2020). Additionally, the gut bacteriome can produce molecules that limit the establishment of protozoa. For example, the bacteria *Lactobacillus reuteri* and *L. acidophilus* produce factors that are capable of inactivating *Cryptosporidium* oocysts (Foster *et al.*, 2003).

2.1.4 The gut protozoa of wild rodents

The majority of the examples described above come from studies in humans, domesticated livestock, and laboratory animals. There are limited studies describing the diversity of gut protozoa in wild mammals, and the drivers of variation in protozoa composition. The gut microbiomes of laboratory and domesticated animals are likely to be quite distinct to that of their wild counterparts (Viney, 2019; Prabhu *et al.*, 2020; Bowerman *et al.*, 2021), so there is a need to study wild animals in greater detail. The Rodentia are a highly speciose order of mammals (Fabre *et al.*, 2012), but their gut protozoa are not well described. As with most mammals, the majority of described gut protozoa in wild rodents are parasitic, rather than commensal/mutualist species (Parfrey *et al.*, 2014). In rodents, this disparity is exacerbated by wild populations being an important group of reservoirs for zoonotic transmission, and so are frequently surveyed for potentially zoonotic parasites (Meerburg *et al.*, 2009; Han *et al.*, 2015). There has been little effort to describe the mutualistic/commensal gut protozoa of wild rodents, except in those species with high digestive efficiency *e.g.* the capybara, *Hydrochoerus hydrochaeris* (Borges *et al.*, 1996).

In order to further our understanding of protozoa in the mammalian gut eukaryome, this work aimed to collate information on which protozoa are found within the gut eukaryome of wild rodent species. This has not, as far as I am aware, been done before. To do this, a methodical literature search was conducted to identify protozoa previously reported in the gut eukaryome of wild rodents, with a description of the host ranges of different protozoa. Here, I hypothesise that the number of protozoa genera found within the gut eukaryome of wild rodents is dependent on surveying effort. Then, I hypothesise that there are differences in how common protozoa are in the rodent gut eukaryome, and aim to address this using a meta-analysis to compare protozoa prevalence among host species. Next, I predict that hosts that have increased opportunities for protozoal transmission between hosts will have a greater prevalence of protozoa in their gut eukaryome. I test this hypothesis using host social and behavioural traits identified from the literature that offer potential variation in the amount of direct and indirect interactions, and thus transmission, among hosts. Finally, I hypothesise that the different methodologies used for identifying protozoa in the gut eukaryome may bias our predictions in how prevalent protozoa are in the wild rodent gut eukaryome.

2.2 Methods

2.2.1 Literature search

The literature search was conducted following PRISMA guidelines (Moher *et al.*, 2009; Page *et al.*, 2021), to identify articles describing gut protozoa in wild rodents in the Web of Science database. All databases within the Web of Science were used. Two independent literature searches were performed. The first was conducted in March 2020, using the following four search terms: infection rodent protozoa gut; gut protozoa rodent; parasite rodent gut; eukaryotic microbiome rodent. Each term was searched for simultaneously using the multi-row option in the Basic Search interface using the “OR” Boolean command and searched within ‘Topic’ (Supplementary Figure 1A); no publication date was specified. The second search was completed in April 2020, performed as above but using the search term: protozoa wild rodent, with an additional seven search terms using the “NOT” Boolean command: wild-type; "wild type"; model; and the four search terms used in the previous search (Supplementary Figure 1B). These additional seven search terms reduced the chance of returning studies using laboratory rodents and excluded any potential duplicate articles from the first search.

2.2.2 Article selection

The titles and abstracts of 6,852 articles were screened to ensure suitability, and 2,018 articles were carried forward for full-text screening, as shown in Figure 2.1 (Identification and Screening). The suitability of articles was based on the report or survey of a naturally occurring protozoa in the digestive system (oesophagus-faecal) of a wild rodent. Articles reporting blood, brain and other tissue parasites were excluded, as were those only focussing on non-protozoa parasites *e.g.* nematodes. Additionally, articles that did not give the location of the wild rodent were excluded, as were those that did not identify the rodent host and gut protozoa to the genus level. Once data were extracted (section 2.2.3) from the selected articles, their reference lists were searched to identify potential articles not identified in the literature search. A further 112 articles were identified and had data extracted in this manner.

2.2.3 Data extraction

Articles were categorised as either i) a report describing the presence of a protozoa or ii) a survey to determine protozoa prevalence. These two article types are henceforth termed presence and prevalence articles respectively. To create data records (Figure 2.1), the following data were extracted from all articles: host species, protozoa species, geographical location sampled, diagnostic technique (DNT), and year sampled. The location of the record was recorded in three ways: continent; country; and latitude and longitude (if provided). Any other location information was also recorded. One article could produce multiple records. For prevalence articles, protozoa prevalence was also recorded, and where necessary, this was

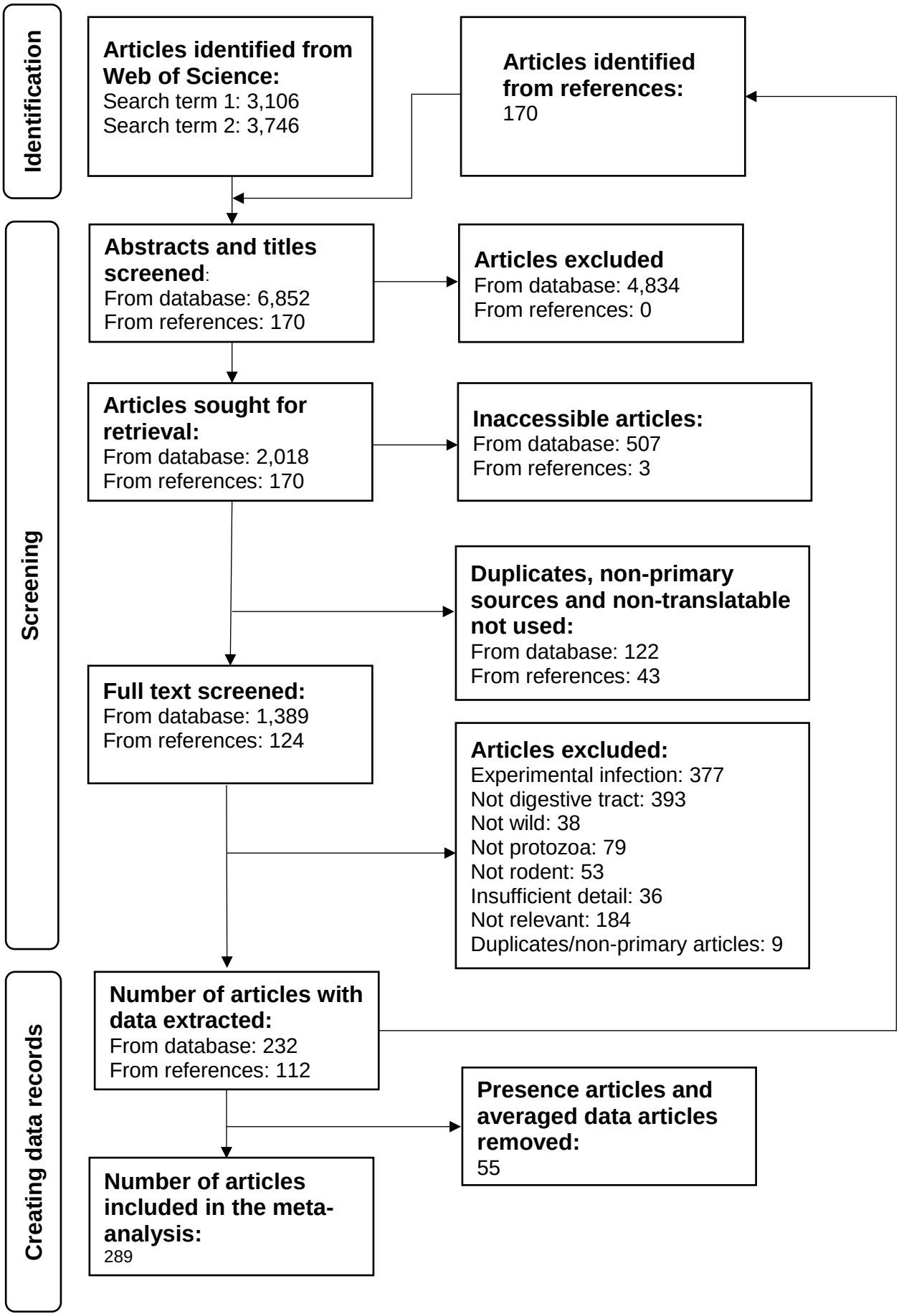


Figure 2.1. PRISMA diagram showing the source of articles and the subsequent screening stages used to generate the data records used in the meta-analysis.

calculated from data presented in the article. Prevalence is defined here as the proportion of hosts that have that particular protozoan in their gut microbiome. If a prevalence range was reported, the median prevalence was used. For articles listing host subspecies and species complexes with different prevalences, the average prevalence was calculated. Additionally, if an article reported multiple prevalences for different con-generic protozoa species, an average protozoa genus prevalence was calculated. Weighted averages were calculated based on the sample size of the individual reports. Articles that used multiple DNTs for the same rodents recorded either i) the combined prevalence stated in the article or ii) the calculated average prevalence, if the overall prevalence was not reported. The DNT for these records were recorded as “Mixed”.

2.2.4 Describing protozoa in the rodent gut eukaryome

Both presence and prevalence articles identified from the literature search were used to generate a meta-table recording the presence of different protozoa in the wild rodent gut eukaryome. Data were recorded at the genus level for protozoa, and species level for host. Rodent host taxonomy was after the Handbook of the Mammals of the World (Wilson *et al.,* 2017). Protozoa genera were assigned to one of five meta-groups: Amoebozoa; Apicomplexa; Ciliophora; Metamonada; and Other (Adl *et al.*, 2019). A generalised linear model (GLM) with a Poisson error distribution (Zuur and Ieno, 2016) was used to determine if the number of protozoa genera identified in a rodent species was dependent on the surveying effort (*i.e.* the number of records) for that rodent species.

2.2.5 Meta-analyses of protozoa prevalence

2.2.5.1 General strategy

Data records from prevalence articles were used to identify if and why there is variation in protozoa prevalence in the gut eukaryome. The records for which an average prevalence was calculated were removed, but the average prevalence record was kept (Figure 2.1). This was to ensure that there was no pseudo-replication of the data. Each data record was assigned an article ID and a unique record number (URN). Within RStudio, the metafor package was used to conduct all meta-analyses (v2.4.0, Viechtbauer, 2010). The general strategy was as follows: i) create a base restricted maximum likelihood estimator (REML) model with only random effects that would be used throughout the following data analyses, ii) assess if there was variation in the prevalence of protozoa in the eukaryome across different rodent host species, iii) identify what variables are contributing to variation in protozoa prevalence and iv) assess potential publication and methodological biases in the dataset.

2.2.5.2 The base model

The base REML model used throughout the meta-analysis listed article ID, URN, DNT, and host phylogeny as random factors. Host phylogeny accounted for potential variation in prevalence due to shared evolutionary history (Koricheva *et al.*, 2013). The phylogeny was created using the Open Tree of Life (OTL) database (Hinchliff *et al.*, 2015) and the rotl R package (v3.0.14, Michonneau *et al.*, 2016) to extract phylogenetic relationships for the rodent taxonomy identified in section 2.2.4. Some species were not identified in the OTL as they had been renamed or reclassified, and so these were manually added to the tree. Grafen's method (Grafen, 1989) was used to compute branch lengths using the ape R package (v5.6.2, Paradis *et al.*, 2004). The final phylogenetic tree is available in Supplementary Figure 2. DNT was listed as a random factor to account for potential variation in prevalence due to the DNT used. In all models, double-arcsine transformed prevalence was the dependent variable (Wang, 2018). Prevalence was transformed in this way to fit the assumptions of normality required for meta-analyses (Barendregt *et al.*, 2013). Given that recent literature has recommended against the use of double-arcsine transformation in meta-analyses (Röver and Friede, 2022; Lin and Xu, 2020), justification for its continued use here is given in Appendix 1.1. The addition of a variable as a fixed effect (henceforth called a moderator) to the base model was used to test if that moderator significantly influenced the overall mean protozoa prevalence in wild rodents.

The overall double-arcsine transformed prevalence of protozoa in the rodent gut eukaryome was calculated from the base model, using the rma.mv function. The result was back-transformed to obtain the summary percentage prevalence and 95% confidence intervals (CI) (Wang, 2018). The number of records included in the model (k) was also recorded. Heterogeneity of prevalences was examined using the I^2 statistic, which is the proportion of variance in effect sizes that is not attributable to sampling (error) variance (Higgins *et al.*, 2003). The proportion of I^2 attributable to differences in article ID, URN, DNT, and host phylogeny was calculated using the i2_ml function in the orchaRrd R package (Nakagawa *et al.*, 2021).

2.2.5.3 Variation in protozoa prevalence

Two meta-regressions were performed to identify any variation in how prevalent gut protozoa were among rodent hosts, with either host family or host species incorporated as the moderator. The variable "protozoa genus" and the subsequent interaction terms with the host family/species were also included in the models. Only the rodent families/species and protozoa genera with ≥ 10 records were included in the model, to account for bias caused by small sample sizes (Lin, 2018). Significant moderators indicated variation in the mean protozoa prevalence. Significance was defined by examining the Q_M statistic. Additionally,

marginal R^2 values were calculated to establish how much heterogeneity in prevalence is described by the moderators, using the r2_ml function in the orchaRd R package (Nakagawa and Schielzeth, 2013; Nakagawa *et al.*, 2021).

Significant protozoa-host interactions were further examined *via* subsetting the host family/species into subgroups and running a separate meta-regression for each subgroup, with protozoa genus as the moderator. Only the host subgroups that had ≥ 2 protozoa genera, with ≥ 10 records per protozoa genera, were tested. Tukey post hoc comparisons were used to conduct pairwise comparisons between protozoa genera if there was a significant effect of protozoa genus identified. This was done by re-running the meta-regression and excluding the intercept, and using the multcomp R package to compare combinations of protozoa genera (v1.4.17, Hothorn *et al.*, 2008). The holm method was used to correct for multiple testing (Holm, 1979). Finally, the average double-arcsine transformed prevalence for each subgroup within each moderator was obtained by using the subset function within the rma.mv model. The result was then back-transformed to obtain the percentage prevalence and 95% CIs. Orchard plots are used to show differences in prevalence among subgroups (Nakagawa *et al.*, 2021). The orchard plots also include 95% CIs and 95% prediction intervals. Prediction intervals here represent the range of prevalence in which the prevalence of a new observation would fall (IntHout *et al.*, 2016). Precision, as the inverse of the standard error for each record, was used in these plots, where a larger precision equates to a larger sample size.

2.2.5.4 Variables contributing to variation prevalence

To ensure geographical differences were not contributing to variation in protozoa prevalence, three geographical moderators were included in a meta-regression: longitude, latitude and continent. Latitude and longitude were converted from degrees, minutes, seconds format to the decimal degrees format using OSMscale (v0.5.1, Boessenkool, 2017) to generate a continuous variable. In this model, the interactions of latitude and longitude with continent were also included as moderators. Additionally, protozoa genus and its interactions with each of the three geographical moderators was also included, to account for variation stemming from different protozoa genera. An orchard plot was used to show differences in protozoa prevalence among continents.

To test if host behaviour was contributing to variation in the prevalence of protozoa in the gut eukaryome, host behaviour moderators were created for each host species. A single resource was used to extract behavioural information (Wilson *et al.,* 2017). From this information, eight moderators were created that were hypothesised to drive differences in the amount of interactions between rodent hosts, and thus differences in protozoa transmission (Ostfeld and Mills, 2008; Sarkar *et al.*, 2020). Average values of density, home range, and dispersal

distance were extracted as quantitative values. Information on the typical social grouping, typical mating system, and development type was recorded as categorical data. Social grouping was split into two variables: binary social groups (solitary *vs.* group-living) and social system (eleven sub-groups). Mating system was defined at the binary level: monogamous or polygamous, and development type was classed as either altricial or precocial (Derrickson, 1992; Wilson *et al.,* 2017). The typical lifestyle of the species was also defined and recorded, according their general behaviour, locomotion and morphology (Wilson *et al.,* 2017). If behavioural information was not available for a species, family characteristics were used but only if this characteristic applied to all species in that family. From these data, eight moderators were tested separately in a meta-regression, each with protozoa genus also included and the relevant interaction term. The eight moderators tested were: dispersal distance, density, home range size, mating system, social system, binary social grouping, development type, and lifestyle. Significant protozoa-host behaviour interactions were further examined as described in 2.2.5.3.

2.2.5.5 Assessing bias

DNT was added as a moderator in a meta-regression to test if different diagnostic methods identified different prevalences of protozoa in the gut eukaryome. This model removed DNT from the random effects. Post hoc tests were completed as described in 2.2.5.3. A second meta-regression was conducted, with precision as a moderator, to determine if sample size affected protozoa prevalence. A funnel plot was used to visualise publication bias: an asymmetrical plot indicates missing effect sizes, potentially from publication bias (Koricheva *et al.*, 2013; Shi and Lin, 2019). Publication bias is defined here as studies being published selectively based on their results (Shi and Lin, 2019). A trim-and-fill test (Duval and Tweedie, 2000) was used to detect missing effect sizes, and predict the average effect size if these were to be included in the analysis.

2.3 Results

2.3.1 Protozoa presence in wild rodents

A total of 344 suitable articles were identified by the literature search, published between 1915-2020 (Supplementary Table 1). From this literature, 2,245 data records of protozoa in the gut eukaryome of rodents were identified, across 69 countries (Figure 2.2). 44 protozoa genera were identified encompassing all five protozoa meta-groups (Table 2.2). The Apicomplexa and Ciliophora meta-groups had the highest number of protozoa genera recorded (13 and 14 respectively) followed by Metamonada protozoa (11 genera). Only four Amoebozoa protozoa were recorded. *Blastocystis* and *Pharyngomonas* were the only genera assigned to the Other meta-group. Apicomplexa and Metamonada protozoa had the most data records, and the most common data records per protozoa genus were from *Eimeria, Cryptosporidium* and *Giardia* (Figure 2.3A). 275 rodent species were identified in the literature search, from 110 genera and 21 families, with large variation in the number of data records generated for each host species. The most common data records per rodent genus were from *Apodemus, Rattus* and *Microtus* (Figure 2.3B).

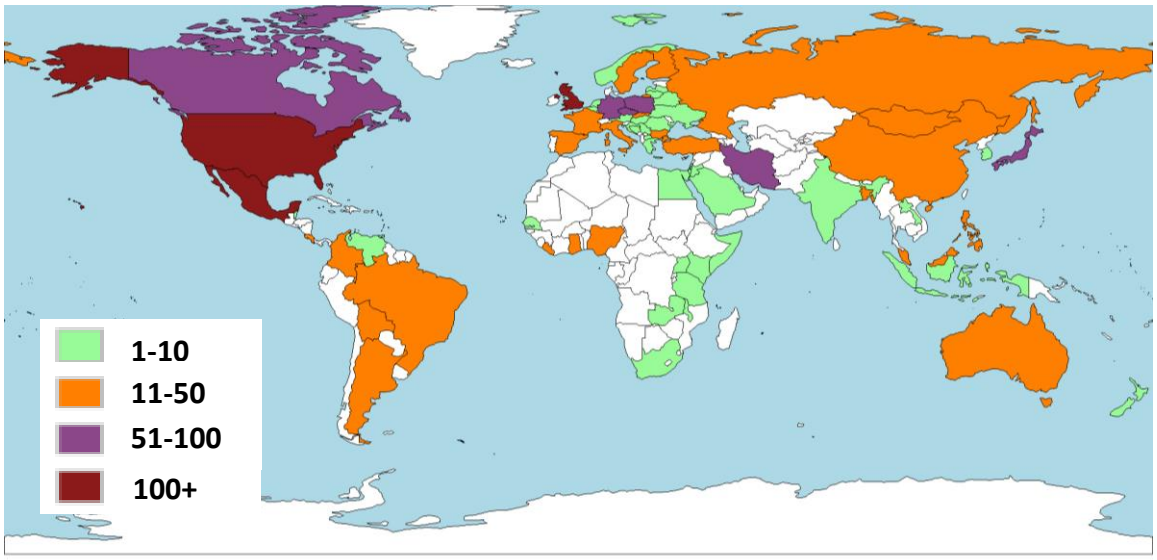


Figure 2.2. The number of records per country that identified or surveyed a gut protozoan in a wild rodent. 2,245 records of protozoa were created from 344 articles.

Table 2.2. The number of records per protozoa meta-group, and the number of protozoa genera assigned to that group.

Protozoa meta-group	Number of records	Number of genera
Amoebozoa	95	4
Apicomplexa	1,725	13
Ciliophora	38	14
Metamonada	368	11
Other	19	2
Total	2,245	44

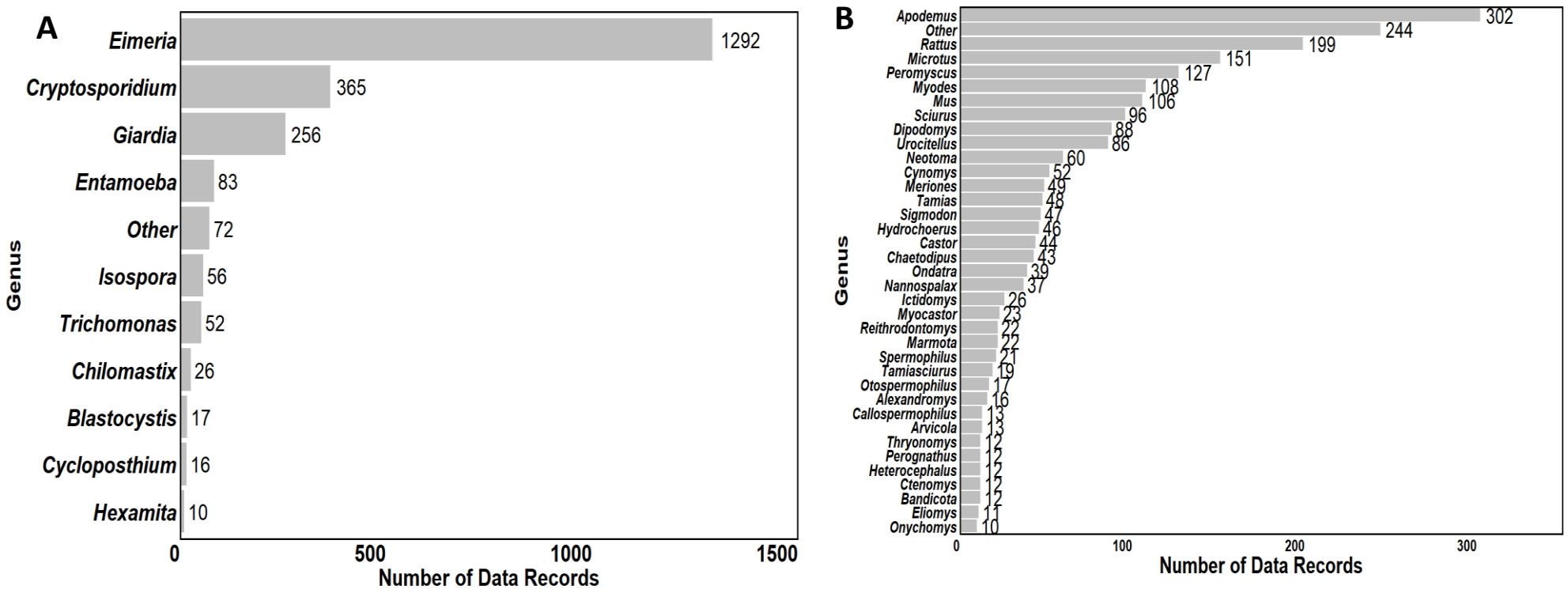


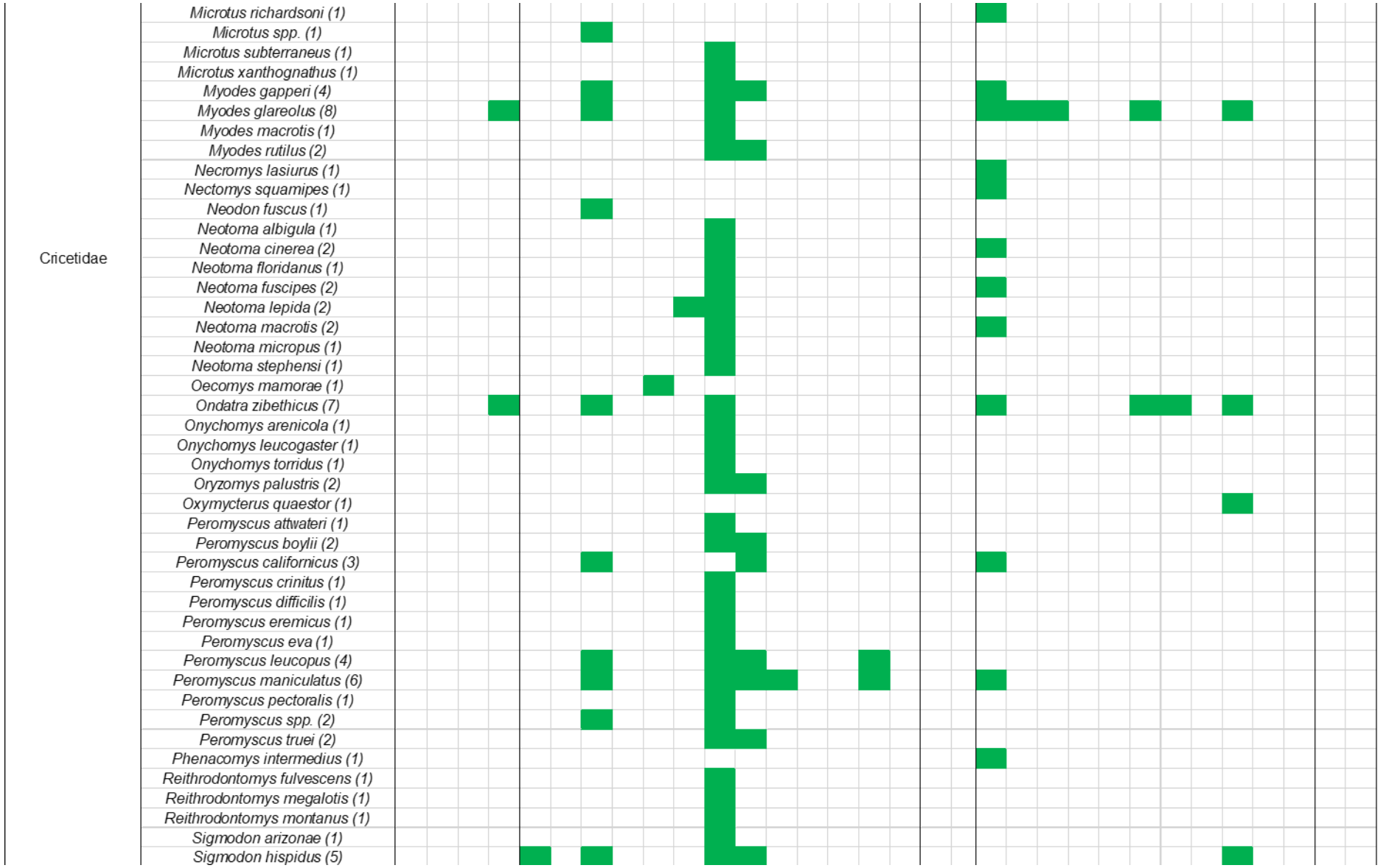
Figure 2.3. The number of data records identified in the literature search. Records are grouped at the A) protozoa genus level and B) host genus level. Only genera with a count of ≥ 10 records are shown, with the Other category combining those with a count of < 10.

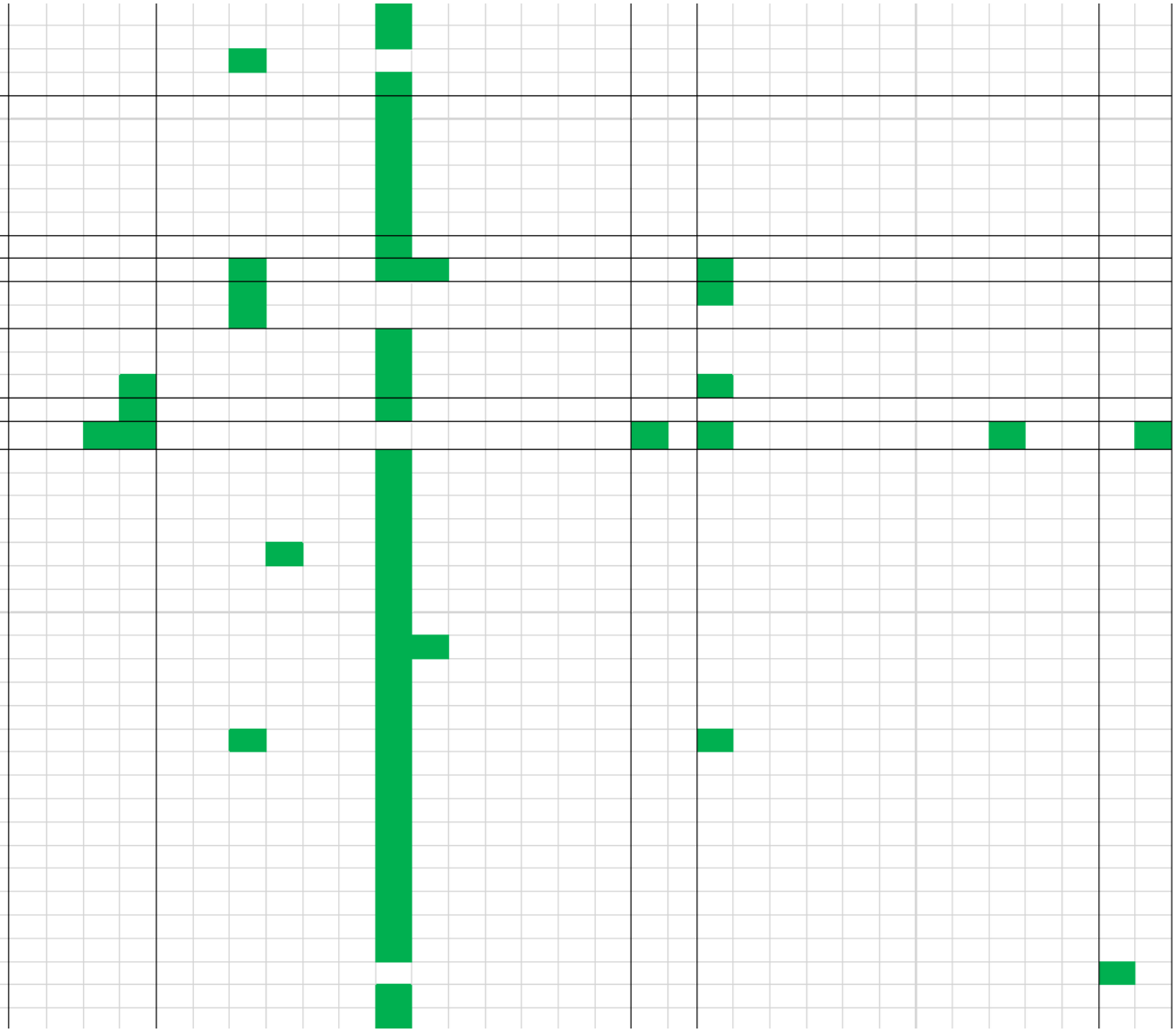
From 2,245 total data records, 1,886 records recorded the presence of gut protozoa in wild rodent hosts. Of the 275 host species identified, 228 had a confirmed protozoan in the gut eukaryome, from both presence and prevalence articles. The protozoa genera found within the gut eukaryome of each rodent host species is summarised in Table 2.3. Protozoa genera are highly varied in the number of host species from which they have been reported. Only seven protozoa genera (from Apicomplexa, Metamonada and Amoebozoa) are found in the gut eukaryome of more than 10 host species. *Eimeria* is the most widely distributed protozoa genus, identified in 194 (of 228) host species. In comparison, 27 protozoa have been reported in the gut eukaryome of only one host species, including 13 (of 14) Ciliophoran genera.

The number of protozoa genera identified in the gut eukaryome of each wild rodent host species is highly variable. 19 (of 228) species had five or more protozoa identified in the literature search, with the majority of the species belonging to the Muridae and Cricetidae families. The greater capybara (*H. hydrochaeris*) had the highest number (17) followed by the brown rat (13, *Rattus norvegicus*) and the black rat (11, *R. rattus*). The majority of rodent species (145 of 228) had just a single protozoan recorded in their gut eukaryome, and these species were from 14 rodent families. The number of different protozoa identified in rodent species is linked to how frequently that host is surveyed: there is a significant, positive relationship between the number of data records for a rodent host and the number of different protozoa identified (Figure 2.4, GLM: $F_{1,226} = 145.5$, $p < 0.001$).

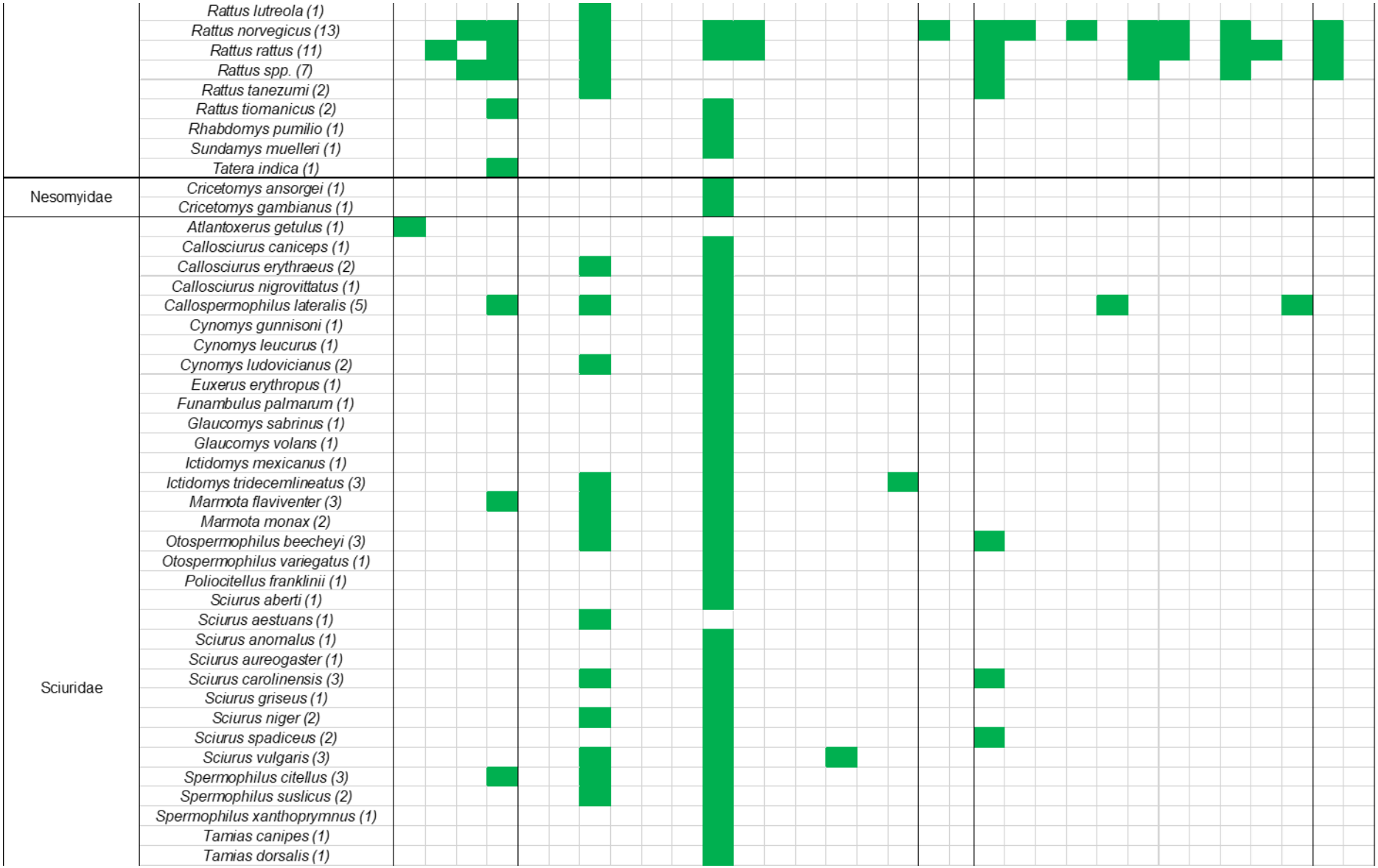
Table 2.3. **Protozoa found in the gut eukaryome of wild rodents.** Rodent species are grouped by family and then ordered alphabetically and the number in brackets indicates the number of protozoa genera found in that host species. The described protozoa are grouped by meta-group, and then alphabetically. The number in brackets indicates the number of host species from which that protozoa has been identified. 'Cilio' are ciliated protozoa. Data were extracted from 344 articles (Supplementary Table 1).

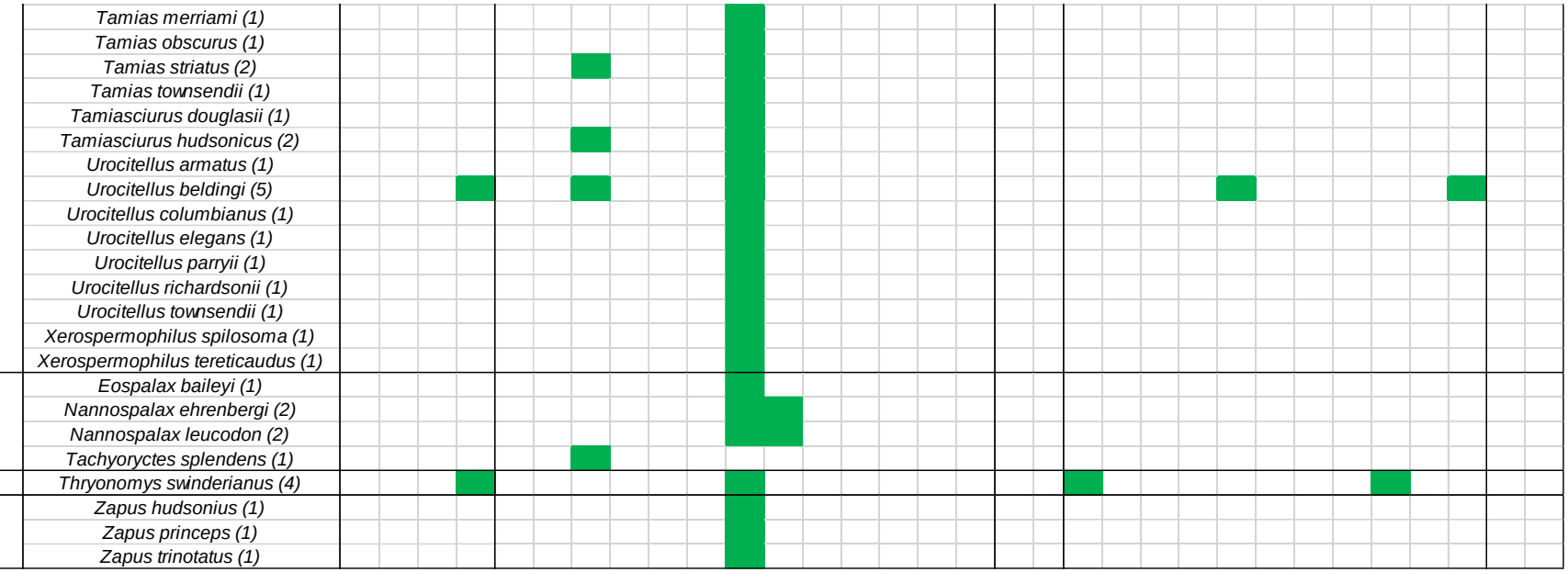
		Amoebozoa				Apicomplexa												Cilio		Metamonada										Other	
Rodent family	**Rodent species**	*Acanthamoeba (1)*	*Amoeba (1)*	*Endolimax (5)*	*Entamoeba (29)*	*Adelina (1)*	*Caryospora (1)*	*Cryptosporidium (59)*	*Cyclospora (1)*	*Cystoisospora (1)*	*Dorisiella (1)*	*Eimeria (194)*	*Isospora (22)*	*Klossia (1)*	*Sarcocystis (1)*	*Toxoplasma (1)*	*Tyzzeria (2)*	*Monocystis (1)*	*Balantidium (2)*	Other (1)	*Giardia (50)*	*Hexamita (4)*	*Octomitus (3)*	*Spironucleus (3)*	*Hexamastix (2)*	*Chilomastix (13)*	*Retortamonas (4)*	*Tetratrichomonas (1)*	*Trichomonas (21)*	*Dientamoeba (1)*	*Trittrichomonas (3)*	*Blastocystis (8)*	*Pharyngomonas (1)*
Aplodontiidae	*Aplodontia rufa (1)*											■																				
Bathyergidae	*Heliophobius argenteocinereus (1)*											■																				
Calomyscidae	*Calomyscus spp. (2)*																											■			■	
Castoridae	*Castor canadensis (3)*							■				■									■											
	Castor fiber (3)							■				■									■											
Caviidae	*Hydrochoerus hydrochaeris (17)*							■				■			■					*	■											
Chinchillidae	*Lagostomus maximus (1)*											■																				
	Akodon montensis (2)							■																					■			
	Alexandromys montebelli (4)											■									■						■			■		
	Alexandromys oeconomus (1)											■																				
	Arvicola amphibius (4)				■			■				■									■											
	Baiomys taylori (1)											■																				
	Craesomys rufocanus (2)											■	■																			
	Cricetulus migratorius (5)				■							■									■					■			■			
	Lemmus lemmus (1)											■																				
	Lemmus trimucronatus (1)											■																				
	Melanomys caliginosus (1)											■																				
	Microtus agrestis (8)				■			■				■									■	■	■			■			■			
	Microtus arvalis (4)							■				■									■		■									
	Microtus californicus (1)											■																				
	Microtus longicaudus (2)											■									■											
	Microtus mexicanus (1)											■																				
	Microtus miurus (1)											■																				
	Microtus montanus (4)				■							■									■			■								
	Microtus mystacinus (2)							■				■																				
	Microtus ochrogaster (2)											■									■											
	Microtus oregoni (1)											■																				
	Microtus pennsylvanicus (5)						■	■				■	■								■											
	Microtus pinetorum (1)							■				■																				


Cricetidae
Microtus richardsoni (1)
Microtus spp. (1)
Microtus subterraneus (1)
Microtus xanthognathus (1)
Myodes gapperi (4)
Myodes glareolus (8)
Myodes macrotis (1)
Myodes rutilus (2)
Necromys lasiurus (1)
Nectomys squamipes (1)
Neodon fuscus (1)
Neotoma albigula (1)
Neotoma cinerea (2)
Neotoma floridanus (1)
Neotoma fuscipes (2)
Neotoma lepida (2)
Neotoma macrotis (2)
Neotoma micropus (1)
Neotoma stephensi (1)
Oecomys mamorae (1)
Ondatra zibethicus (7)
Onychomys arenicola (1)
Onychomys leucogaster (1)
Onychomys torridus (1)
Oryzomys palustris (2)
Oxymycterus quaestor (1)
Peromyscus attwateri (1)
Peromyscus boylii (2)
Peromyscus californicus (3)
Peromyscus crinitus (1)
Peromyscus difficilis (1)
Peromyscus eremicus (1)
Peromyscus eva (1)
Peromyscus leucopus (4)
Peromyscus maniculatus (6)
Peromyscus pectoralis (1)
Peromyscus spp. (2)
Peromyscus truei (2)
Phenacomys intermedius (1)
Reithrodontomys fulvescens (1)
Reithrodontomys megalotis (1)
Reithrodontomys montanus (1)
Sigmodon arizonae (1)
Sigmodon hispidus (5)

	Sigmodon ochrognathus (1)	
	Synaptomys borealis (1)	
	Thaptomys nigrita (1)	
	Zygodontomys brevicauda (1)	
Ctenomyidae	*Ctenomys boliviensis (1)*	
	Ctenomys conoveri (1)	
	Ctenomys frater (1)	
	Ctenomys lewisi (1)	
	Ctenomys opimus (1)	
	Ctenomys talarum (1)	
Dasyproctidae	*Dasyprocta leporina (1)*	
Echimyidae	*Myocastor coypus (4)*	
Erethizontidae	*Coendou spinosus (2)*	
	Erethizon dorsatum (1)	
Geomyidae	*Orthogeomys grandis (1)*	
	Thomomys bottae (1)	
	Thomomys talpoides (3)	
Gliridae	*Eliomys quercinus (2)*	
Heterocephalidae	*Heterocephalus glaber (6)*	
Heteromyidae	*Chaetodipus baileyi (1)*	
	Chaetodipus californicus (1)	
	Chaetodipus fallax (1)	
	Chaetodipus formosus (1)	
	Chaetodipus hispidus (2)	
	Chaetodipus intermedius (1)	
	Chaetodipus penicillatus (1)	
	Chaetodipus spinatus (1)	
	Dipodomys agilis (2)	
	Dipodomys californicus (1)	
	Dipodomys elator (1)	
	Dipodomys gravipes (1)	
	Dipodomys heermanni (3)	
	Dipodomys merriami (1)	
	Dipodomys microps (1)	
	Dipodomys ordii (1)	
	Dipodomys ornatus (1)	
	Dipodomys panamintinus (1)	
	Dipodomys spectabilis (1)	
	Dipodomys venustus (1)	
	Heteromys irroratus (1)	
	Heteromys pictus (1)	
	Heteromys spp. (1)	
	Perognathus flavescens (1)	
	Perognathus flavus (1)	


Muridae
Acomys cahirinus (3)
Acomys dimidiatus (1)
Apodemus agrarius (3)
Apodemus argenteus (3)
Apodemus flavicollis (3)
Apodemus mystacinus (1)
Apodemus speciosus (4)
Apodemus spp. (8)
Apodemus sylvaticus (7)
Apodemus uralensis (1)
Arvicanthis niloticus (1)
Bandicota savilei (1)
Berylmys bowersi (1)
Chiropodomys gliroides (1)
Chrotomys whiteheadi (1)
Conilurus penicillatus (1)
Dasymys incomtus (1)
Dephomys defua (1)
Gerbilliscus guineae (1)
Lemniscomys striatus (3)
Leopoldamys edwardsi (2)
Leopoldamys sabanus (1)
Leporillus conditor (1)
Lophuromys flavopunctatus (1)
Lophuromys sikapusi (1)
Mastomys natalensis (1)
Maxomys surifer (1)
Maxomys whiteheadi (1)
Meriones hurrianae (2)
Meriones meridianus (2)
Meriones persicus (7)
Meriones shawii (2)
Meriones spp. (6)
Meriones unguiculatus (2)
Mus musculoides (1)
Mus musculus (8)
Mus spretus (1)
Niviventer fulvescens (2)
Otomys denti (1)
Otomys irroratus (1)
Praomys tullbergi (1)
Rattus andamanensis (2)
Rattus argentiventer (1)
Rattus exulans (2)


Rattus lutreola (1)
Rattus norvegicus (13)
Rattus rattus (11)
Rattus spp. (7)
Rattus tanezumi (2)
Rattus tiomanicus (2)
Rhabdomys pumilio (1)
Sundamys muelleri (1)
Tatera indica (1)
Nesomyidae
Cricetomys ansorgei (1)
Cricetomys gambianus (1)
Atlantoxerus getulus (1)
Callosciurus caniceps (1)
Callosciurus erythraeus (2)
Callosciurus nigrovittatus (1)
Callospermophilus lateralis (5)
Cynomys gunnisoni (1)
Cynomys leucurus (1)
Cynomys ludovicianus (2)
Euxerus erythropus (1)
Funambulus palmarum (1)
Glaucomys sabrinus (1)
Glaucomys volans (1)
Ictidomys mexicanus (1)
Ictidomys tridecemlineatus (3)
Marmota flaviventer (3)
Marmota monax (2)
Otospermophilus beecheyi (3)
Otospermophilus variegatus (1)
Poliocitellus franklinii (1)
Sciurus aberti (1)
Sciurus aestuans (1)
Sciurus anomalus (1)
Sciurus aureogaster (1)
Sciurus carolinensis (3)
Sciurus griseus (1)
Sciurus niger (2)
Sciurus spadiceus (2)
Sciurus vulgaris (3)
Spermophilus citellus (3)
Spermophilus suslicus (2)
Spermophilus xanthoprymnus (1)
Tamias canipes (1)
Tamias dorsalis (1)
Sciuridae

	Tamias merriami (1)	
	Tamias obscurus (1)	
	Tamias striatus (2)	
	Tamias townsendii (1)	
	Tamiasciurus douglasii (1)	
	Tamiasciurus hudsonicus (2)	
	Urocitellus armatus (1)	
	Urocitellus beldingi (5)	
	Urocitellus columbianus (1)	
	Urocitellus elegans (1)	
	Urocitellus parryii (1)	
	Urocitellus richardsonii (1)	
	Urocitellus townsendii (1)	
	Xerospermophilus spilosoma (1)	
	Xerospermophilus tereticaudus (1)	
Spalacidae	*Eospalax baileyi (1)*	
	Nannospalax ehrenbergi (2)	
	Nannospalax leucodon (2)	
	Tachyoryctes splendens (1)	
Thryonomyidae	*Thryonomys swinderianus (4)*	
Zapodidae	*Zapus hudsonius (1)*	
	Zapus princeps (1)	
	Zapus trinotatus (1)	

*The following 13 ciliated protozoa are only found in *H. hydrochaeris*: *Cycloposthium; Monoposthium; Anacharon; Cunhamunizia; Enterophrya; Eriocharon; Hydrochoerella; Muniziella; Ogimotoa; Ogimotopsis; Paracunhamunizia; Protohallia; Uropogon.*

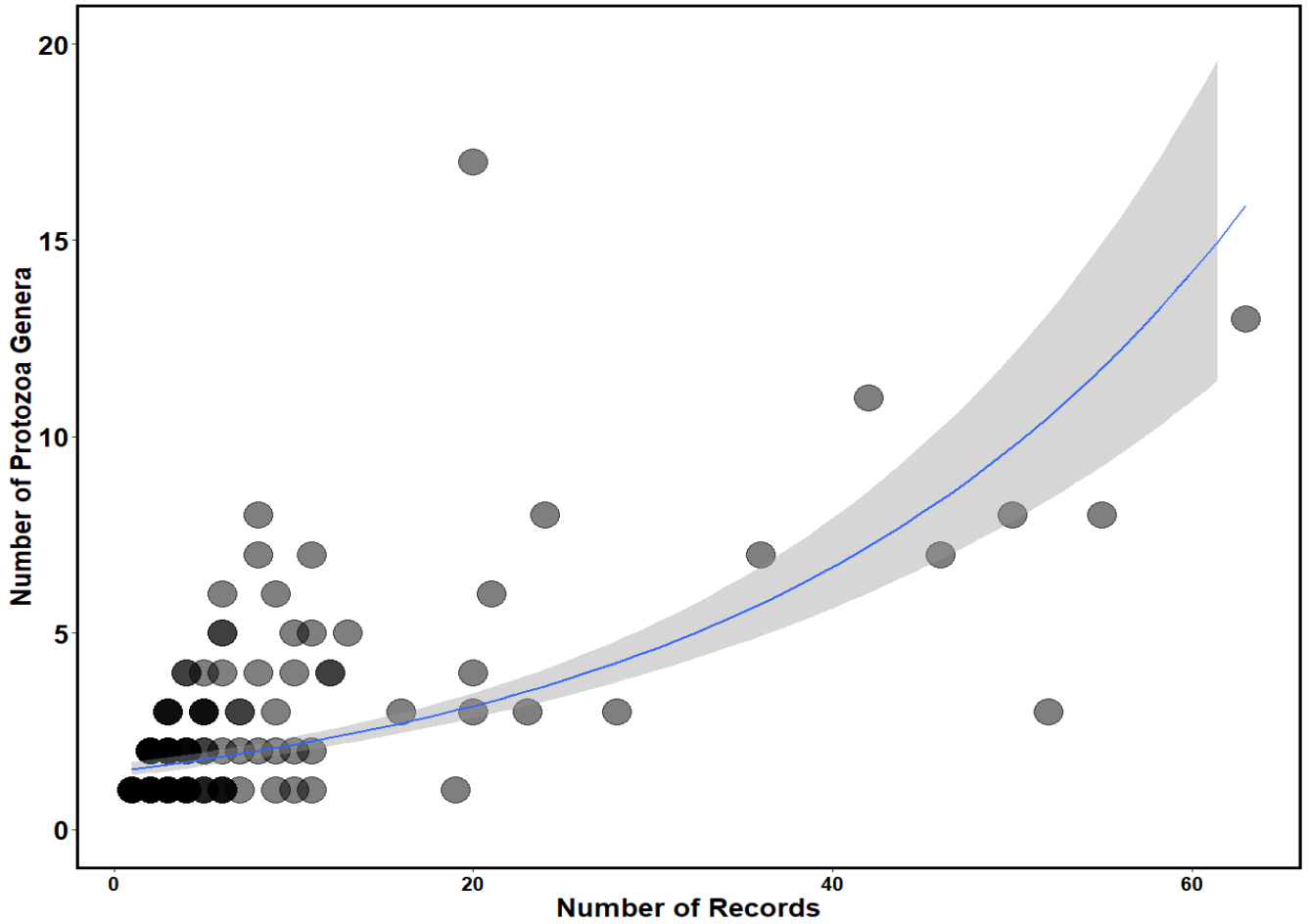


Figure 2.4. The relationship between the number of data records for a host species and number of protozoa genera identified in that host. Only the data records describing the presence of protozoa are included. Each point is a different rodent species. The predicted relationship is represented by the blue line, with the grey shading representing 95% confidence intervals.

2.3.2 Protozoa prevalence

2.3.2.1 Predicted protozoa prevalence

A total of 1,237 (of 2,245) data records (after the removal of pseduoreplicated data records and presence records, section 2.2.5.1) were used to look for variation in the prevalence of protozoa in the rodent gut eukaryome. 255 wild rodent species were surveyed across 289 articles, from 102 genera and 21 families. 36 protozoa genera were used in the meta-analysis.

Across all wild rodents, the average prevalence of any protozoa in the wild rodent gut eukaryome is predicted to be 23.7% (95% CI 4.78-48.5, k = 1,237). There is a high amount of heterogeneity of prevalence in the dataset (I^2 = 97.8%), with the majority of this variation stemming from differences among the individual data records (32.3%) and differences attributed to the article ID of a data record (32.0%). Diagnostic techniques accounted for 6.5% of variation. The shared evolutionary history of the rodent hosts explained 26.9% of the variation in protozoa prevalence.

2.3.2.2 Variation among rodent hosts

The estimated protozoa prevalence of each rodent family is shown in Table 2.4. Only the 7 (of 21) host families with ≥ 10 data records were used in the analysis. There were no significant differences in protozoan prevalence among different rodent families, but there was a significant interaction between protozoa genus and host family (Table 2.5). Thus, the variation among host families in the prevalence of the protozoa found in the gut eukaryome is dependent on the protozoa genus examined. To investigate this interaction further, individual rodent families were analysed separately. The three host families that had ≥ 2 protozoa genera, with ≥ 10 records per protozoa genera were: Cricetidae, Muridae and Sciuridae. Protozoa genus was a significant moderator in all three families tested, showing that for these three families, different protozoa have different prevalences (Table 2.6, Figure 2.5).

Table 2.4. The predicted gut protozoa prevalence for each rodent family. k is the number of data records contributing to the prediction.

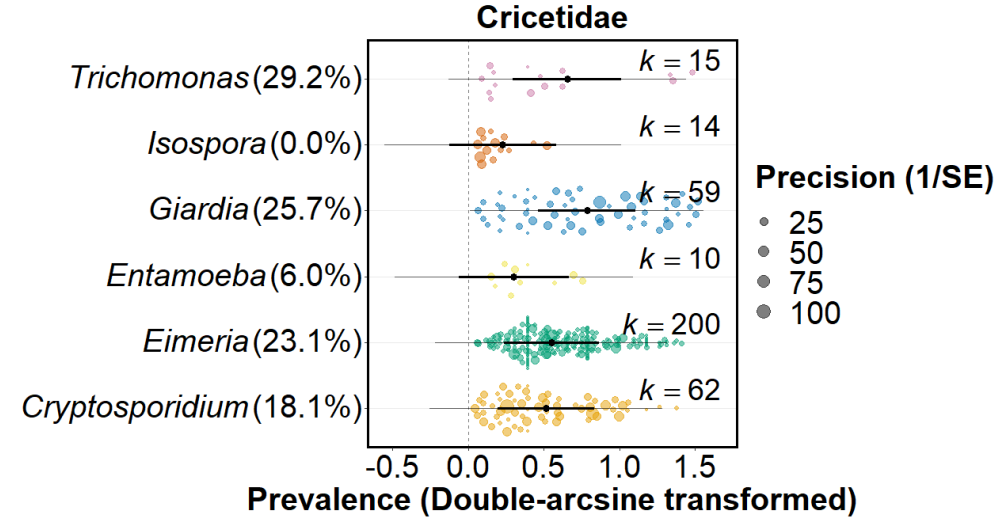
Family	Prevalence (%)	k	95% CIs	
			Lower	Upper
Castoridae	3.6	34	0.0	100.0
Cricetidae	25.4	368	1.1	60.7
Heteromyidae	13.2	93	4.9	23.6
Muridae	12.3	448	0.0	70.3
Sciuridae	36.0	146	0.0	100.0
Spalacidae	42.1	11	0.2	93.3
Zapodidae	1.5	11	0.0	100.0

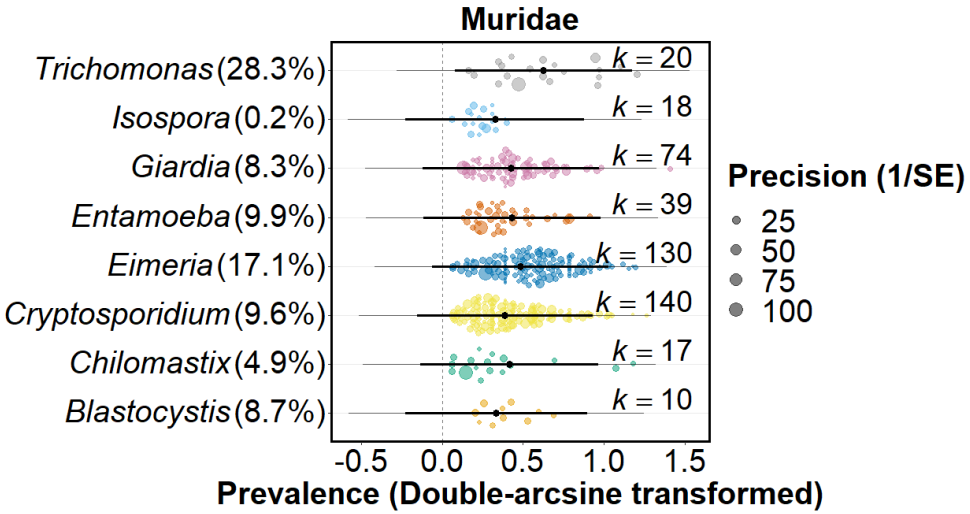
Table 2.5. Meta-regression statistics for two models testing for differences in protozoa prevalence among rodent hosts. REML models were run separately for host family and host species, and each included protozoa genus as an interaction term. k is the number of data records included in the model and marginal R^2 indicates how much variance in prevalence is explained by the model. Q_M is the test statistic for each moderator in the model.

Moderator	k	Marginal R^2	Host moderator		Interaction term	
			Q_M	*p* value	Q_M	*p* value
Host family	1,111	0.146	1.50	0.959	107.6	< 0.001
Host species	538	0.396	41.7	< 0.001	122.4	< 0.001

Table 2.6. Meta-regression statistics testing for the effect of protozoa genus on protozoa prevalence in three different host families. Each host family was tested in a separate model. Only protozoa genera with ≥10 records were included. k is the number of data records included in the model and marginal R^2 indicates how much variance in prevalence is explained by the model. Q_M is the test statistic for the effect of the protozoa genus moderator in the model. *p* values are shown for post hoc comparisons between those protozoa genera with significant differences**.**

Family	Meta regression statistics			Post hoc comparisons	
	Marginal R^2	Q_M	*p* value	Comparison	*p* value
Cricetidae (k = 448)	0.027	33.2	< 0.001	*Cryptosporidium*: *Giardia*	<0.001
				Eimeria: *Giardia*	0.006
				Eimeria: *Isospora*	0.007
				Entamoeba: *Giardia*	<0.001
				Giardia: *Isospora*	<0.001
				Isospora: *Trichomonas*	0.012
Muridae (k = 360)	0.106	46.2	< 0.001	*Blastocystis*: *Trichomonas*	0.012
				Chilomastix: *Trichomonas*	0.010
				Cryptosporidium: *Trichomonas*	<0.001
				Entamoeba: *Trichomonas*	0.004
				Giardia: *Trichomonas*	0.001
				Isospora: *Trichomonas*	<0.001
Sciuridae (k = 142)	0.188	42.0	< 0.001	*Cryptosporidium*: *Eimeria*	<0.001
				Cryptosporidium: *Giardia*	0.050
				Eimeria: *Giardia*	<0.001





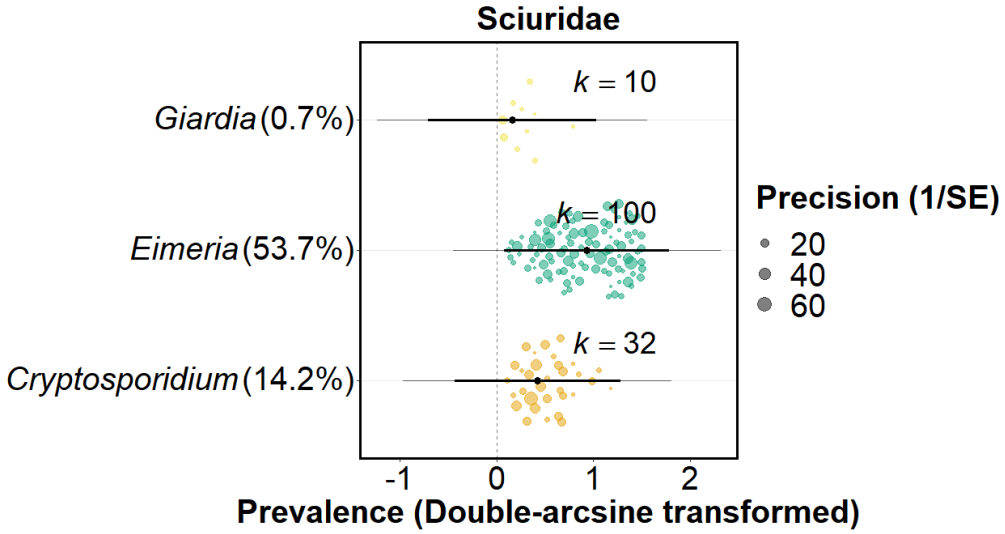


Figure 2.5. The average prevalence of the protozoa identified in the Cricetidae, Muridae and Sciuridae families. Prevalence is double- arcsine transformed and the x-axis scale is family-specific. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that protozoa. The back-transformed predicted prevalence percentage is provided next to the protozoa genus label.

There was a significant difference in the prevalence of protozoa found in the gut eukaryome among rodent host species, which contrasts against the finding of no differences among rodent families (Table 2.5). The overall protozoa prevalence in each host species is shown in Table 2.7. Only 19 (of 255) rodent species were present in this analysis. There was a significant interaction between protozoa genus and host species (Table 2.5). The variation among host species in the prevalence of the protozoa found in the gut eukaryome is dependent on the protozoa genus examined. Seven rodent species were further analysed separately: *Apodemus agrarius, Ap. flavicollis, Ap. sylvaticus, Mus musculus, Myodes glareolus, Ondatra zibethicus* and *R. rattus.* Protozoa genus was not a significant moderator for six of the seven species tested (Figure 2.6, Table 2.8). The muskrat (*O. zibethicus*) is the exception, with *Giardia* prevalence in the gut eukaryome (64.2%) being significantly higher compared to *Cryptosporidium* prevalence (29.2%). Furthermore, testing for differences in *Giardia* prevalence among host species shows that *Giardia* prevalence is significantly higher in the gut eukaryome of *O. zibethicus*, compared to *Giardia* prevalence in *Castor canadensis, M. musculus* and *R. rattus* ($Q_M = 18.8$, $p < 0.001$, k = 65, Figure 2.7).

Table 2.7. The average gut protozoa prevalence for each rodent species. Protozoa richness is the number of protozoa genera found within the gut eukaryome of that host species.

Species	Protozoa richness	Prevalence (%)	Sample size (k)	95% CI	
				Lower	Upper
Apodemus agrarius	3	25.8	29	11.8	42.0
Apodemus flavicollis	3	23.3	54	14.9	32.6
Apodemus speciosus	4	33.1	12	3.1	71.8
Apodemus sylvaticus	7	23.2	43	13.5	34.2
Castor canadensis	3	6.3	27	2.1	11.8
Dipodomys merriami	1	14.1	11	1.4	33.0
Dipodomys ordii	1	5.2	10	0.0	26.7
Microtus agrestis	8	43.4	23	21.0	67.0
Microtus arvalis	4	42.5	22	10.3	77.9
Microtus pennsylvanicus	5	36.2	10	14.3	60.8
Mus musculus	8	7.1	68	3.0	12.4
Myodes glareolus	8	32.5	51	22.1	43.6
Ondatra zibethicus	7	47.8	26	31.5	64.3
Peromyscus	6	5.9	19	0.0	19.2
Rattus norvegicus	13	9.1	55	3.3	16.7
Rattus rattus	11	23.6	41	14.0	34.4
Sciurus carolinensis	3	70.4	15	26.6	99.9
Sciurus vulgaris	3	53.8	10	25.6	81.0
Sigmodon hispidus	5	13.0	12	5.6	22.1

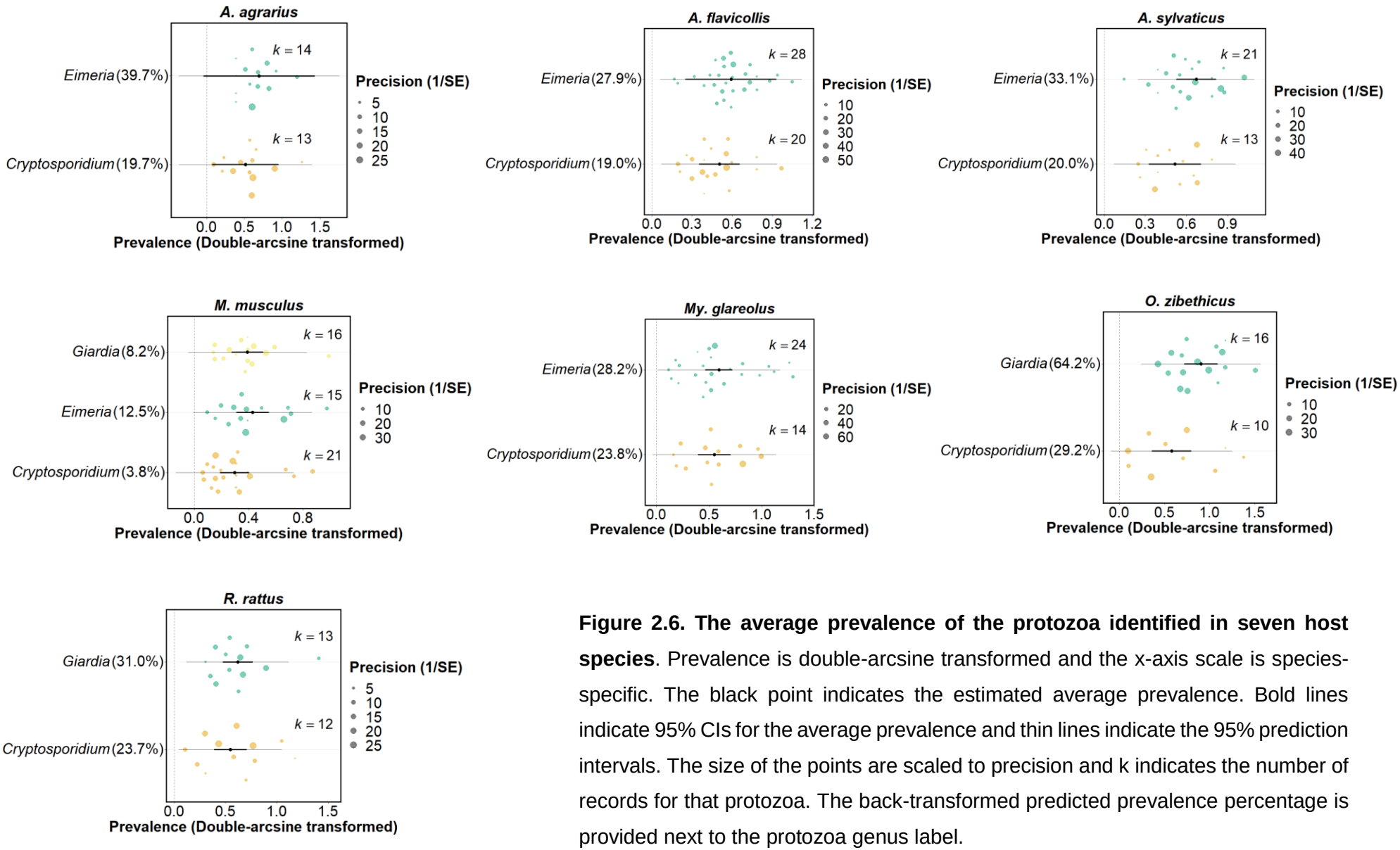


Figure 2.6. The average prevalence of the protozoa identified in seven host species. Prevalence is double-arcsine transformed and the x-axis scale is species-specific. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that protozoa. The back-transformed predicted prevalence percentage is provided next to the protozoa genus label.

Table 2.8. Meta-regression statistics testing for the effect of protozoa genus on protozoa prevalence for seven rodent species. Each host species was tested in a separate model. Only protozoa genera with ≥ 10 records were included. k is the number of data records included in the model and marginal R^2 indicates how much variance in prevalence is explained by the model. Q_M is the test statistic for the effect of the protozoa genus moderator in the model. The host species for which protozoa genus had a significant effect of protozoa prevalence are highlighted in bold.

Species	k	Marginal R^2	Q_M	*p* value
Apodemus agrarius	27	0.053	0.17	0.678
Apodemus flavicollis	48	0.042	0.21	0.648
Apodemus sylvaticus	34	0.124	1.63	0.202
Mus musculus	52	0.068	3.03	0.220
Myodes glareolus	38	0.005	0.18	0.674
Ondatra zibethicus	26	0.200	5.40	0.020
Rattus rattus	25	0.021	0.41	0.522

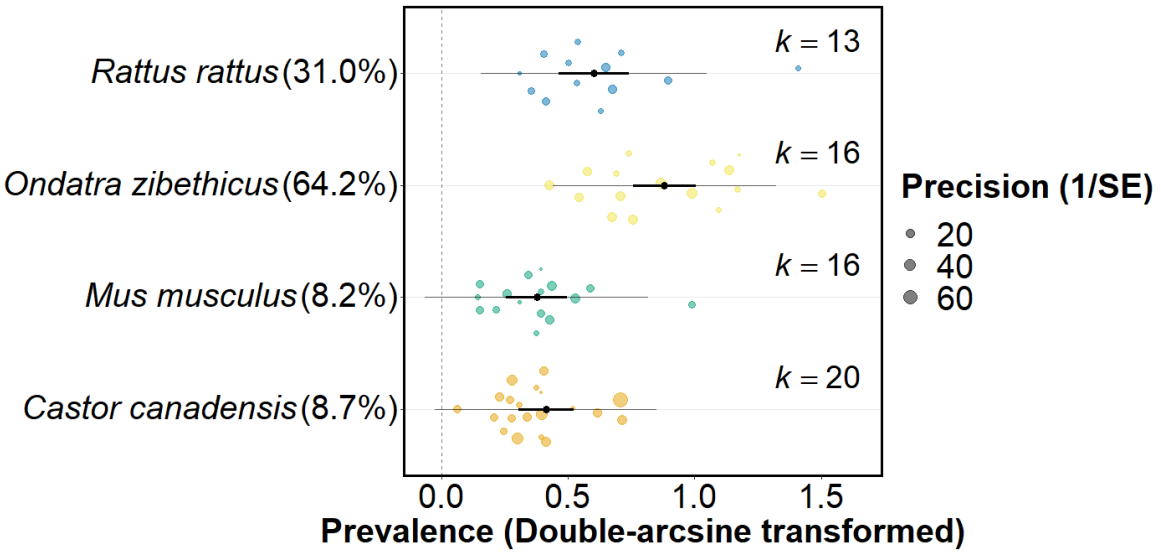


Figure 2.7. The prevalence of *Giardia* in four rodent host species. Prevalence is double-arcsine transformed. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that species. The back-transformed predicted prevalence percentage is provided next to the host species label.

2.3.3 Factors contributing to differences in prevalence

There was no evidence that geographical variation is contributing to differences among rodent species in the prevalence of protozoa in the gut eukaryome (Table 2.9, Figure 2.8). Furthermore, rodent host sociality does not affect protozoa prevalence, as measured by seven variables: home range size; dispersal distance; density; social system; binary social system; development type; and mating system (Figures 2.9 and 2.10, Table 2.10).

Table 2.9. Meta-regression statistics testing for the effects of geographical moderators on protozoa prevalence. All three geographical moderators were tested in the same model. Interactions terms with continent were included for latitude and longitude. Protozoa genus was included as an interaction term for all three geographical moderators. Only protozoa genera with ≥ 10 records were included. 225 data records were included in the model. Q_M is the test statistic for the effect of each moderator in the model.

Moderator	Geographical moderator		Continent interaction term		Protozoa interaction term	
	Q_M	*p* value	Q_M	*p* value	Q_M	*p* value
Continent	4.08	0.537			17.9	0.653
Latitude	1.08	0.299	6.88	0.230	3.38	0.760
Longitude	0.39	0.533	6.78	0.238	3.26	0.353

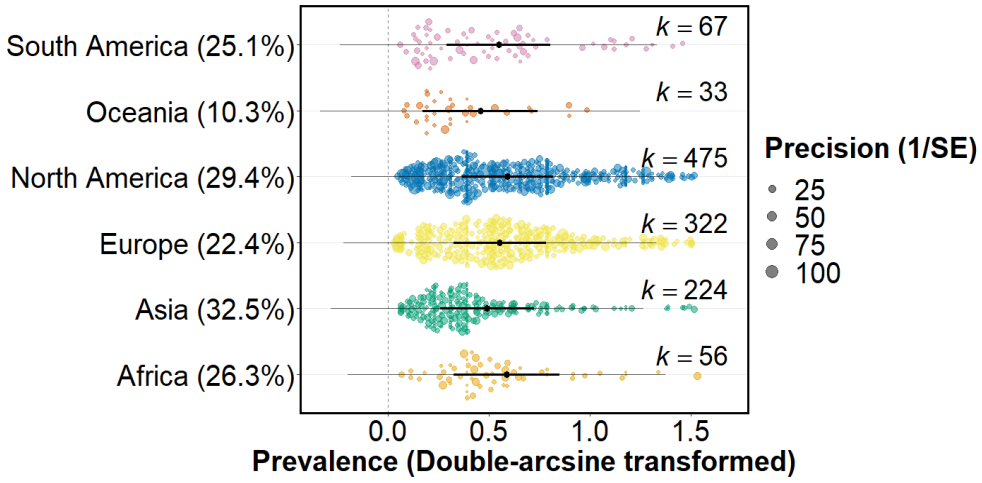


Figure 2.8. The prevalence of protozoa according to continent. Prevalence is double-arcsine transformed. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that continent. The back-transformed predicted prevalence percentage is provided next to the continent label.

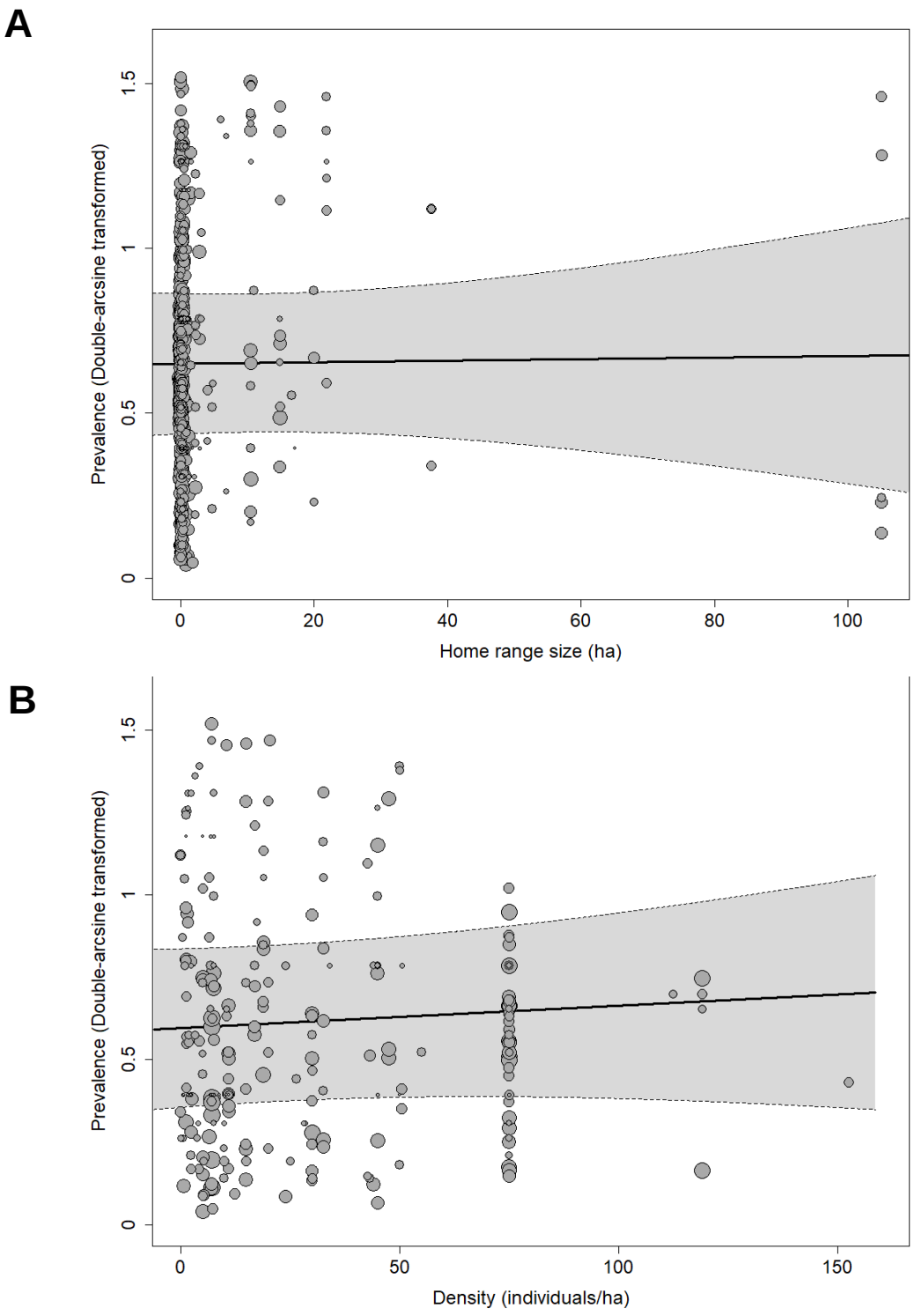


Figure 2.9. Protozoa prevalence according to A) host home range size and B) host density. Points are weighted according to precision. The solid line indicates the predicted prevalence from the meta-regression model, and the shading is 95% CIs.

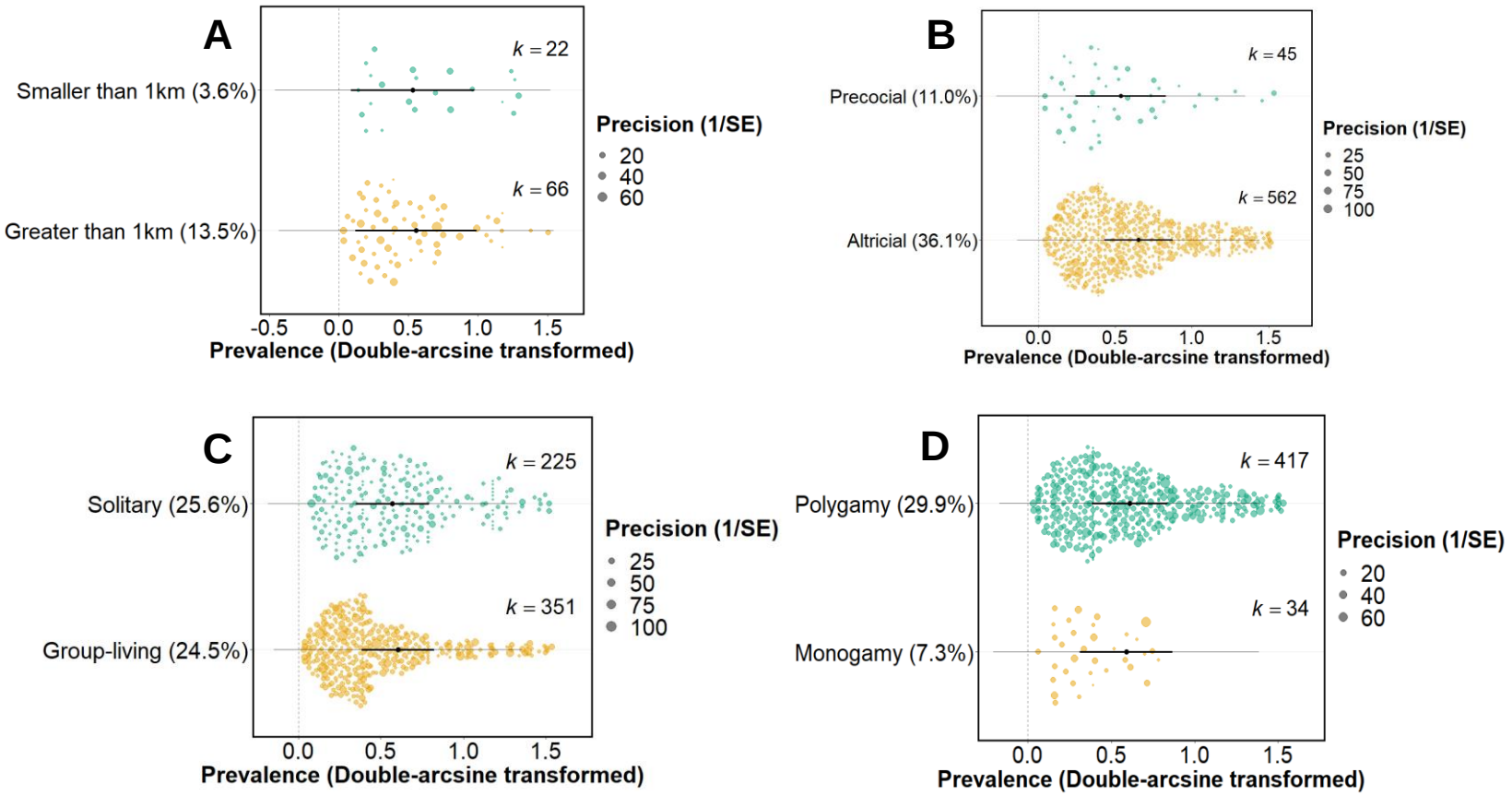


Figure 2.10. The prevalence of protozoa according to four host sociality variables. A) average dispersal distance, B) development type, C) social system (binary), and D) mating system (binary). Prevalence is double-arcsine transformed. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that subgroup. The back-transformed predicted prevalence percentages are provided next to the subgroup labels.

Table 2.10. Meta-regression statistics testing for the effects of host sociality on protozoa prevalence. Each moderator was tested in a separate model. Protozoa genus was included as an interaction term in all models. Only protozoa genera with ≥ 10 records were included. k is the number of data records included in the model and marginal R^2 indicates how much variance in prevalence is explained by the model. Q_M is the test statistic for the effect of each sociality moderator in the model. Moderators with a significant interaction effect are highlighted in bold.

Moderator	**k**	**Marginal R^2**	**Life style trait**		**Interaction term**	
			Q_M	***p* value**	**Q_M**	***p* value**
Home range	510	0.103	1.06	0.303	10.9	0.093
Density	259	0.152	1.70	0.192	3.58	0.733
Dispersal	88	0.102	1.65	0.198	4.89	0.087
Development	607	0.084	0.02	0.880	7.09	0.312
Social system	551	0.140	10.7	0.152	33.4	0.220
Social grouping[a]	576	0.100	0.09	0.765	6.83	0.447
Mating system	451	0.091	< 0.001	0.992	0.68	0.878
Lifestyle	988	0.116	1.06	0.983	**57.3**	**0.003**

a Binary social system: solitary *vs*. group living

Variation in host lifestyle did not significantly affect protozoa prevalence. However, there was a significant interaction between the host lifestyle and protozoa genera (Table 2.10), thus how the prevalence of the protozoa found in the gut eukaryome differs among host lifestyles is dependent on the protozoa genus examined. Five host lifestyles were individually analysed further. In three of five lifestyles, protozoa genus had a significant effect on prevalence (Table 2.11). *Eimeria* had a significantly higher prevalence in the gut eukaryomes of arboreal and fossorial rodents (82.9% and 40.8%) compared to the other protozoa genera present in these rodents (Figure 2.11). *Eimeria* was also significantly more prevalent in the gut eukaryome of terrestrial rodents compared to *Cryptosporidium* (26.8% and 15.0%, respectively). Additionally, *Trichomonas* was significantly more prevalent in the gut eukaryome of terrestrial rodents (28.5%), compared to *Entamoeba* (8.9%) and *Cryptosporidium*. Different protozoa genera did not have a significantly different prevalence in both semi-aquatic and semi-fossorial rodents.

Table 2.11. Meta-regression statistics testing for the effect of protozoa genus on protozoa prevalence for five different host lifestyles. Each lifestyle was tested in a separate model. Only protozoa genera with ≥ 10 records were included. k is the number of data records included in the model and marginal R^2 indicates how much variance in prevalence is explained by the model. Q_M is the test statistic for the effect of the protozoa genus moderator in the model. Lifestyles for which protozoa genus had a significant effect of protozoa prevalence are highlighted in bold. *p* values are shown for post hoc comparisons between those protozoa genera with significant differences.

Lifestyle	**Meta regression statistics**			**Post hoc comparisons**	
	Marginal R^2	**Q_M**	***p* value**	**Comparison**	***p* value**
Arboreal (k = 62)	**0.507**	**33.8**	**< 0.001**	*Cryptosporidium*: *Eimeria*	<0.001
Fossorial (k = 76)	**0.254**	**15.9**	**0.001**	*Cryptosporidium*: *Eimeria*	0.019
				Eimeria: *Entamoeba*	0.023
				Eimeria: *Giardia*	0.001
Semi-aquatic (k = 86)	0.039	3.45	0.178	NA	NA
Semi-fossorial (k = 84)	0.025	1.37	0.242	NA	NA
Terrestrial (k = 547)	**0.070**	**26.3**	**< 0.001**	*Cryptosporidium*: *Eimeria*	0.007
				Cryptosporidium: *Trichomonas*	0.017
				Entamoeba: *Trichomonas*	0.045

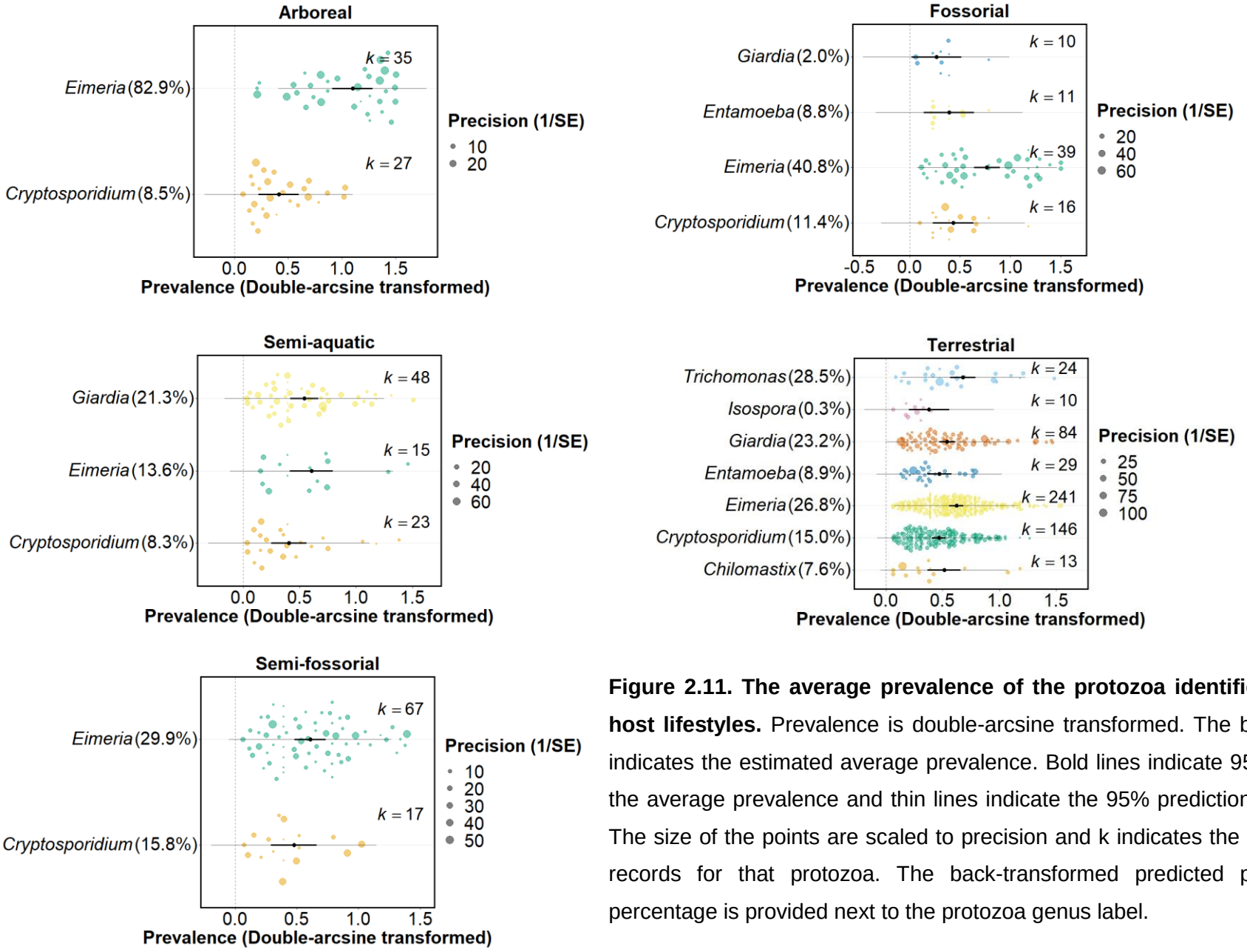


Figure 2.11. The average prevalence of the protozoa identified in five host lifestyles. Prevalence is double-arcsine transformed. The black point indicates the estimated average prevalence. Bold lines indicate 95% CIs for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that protozoa. The back-transformed predicted prevalence percentage is provided next to the protozoa genus label.

2.3.4 Assessing bias in the dataset

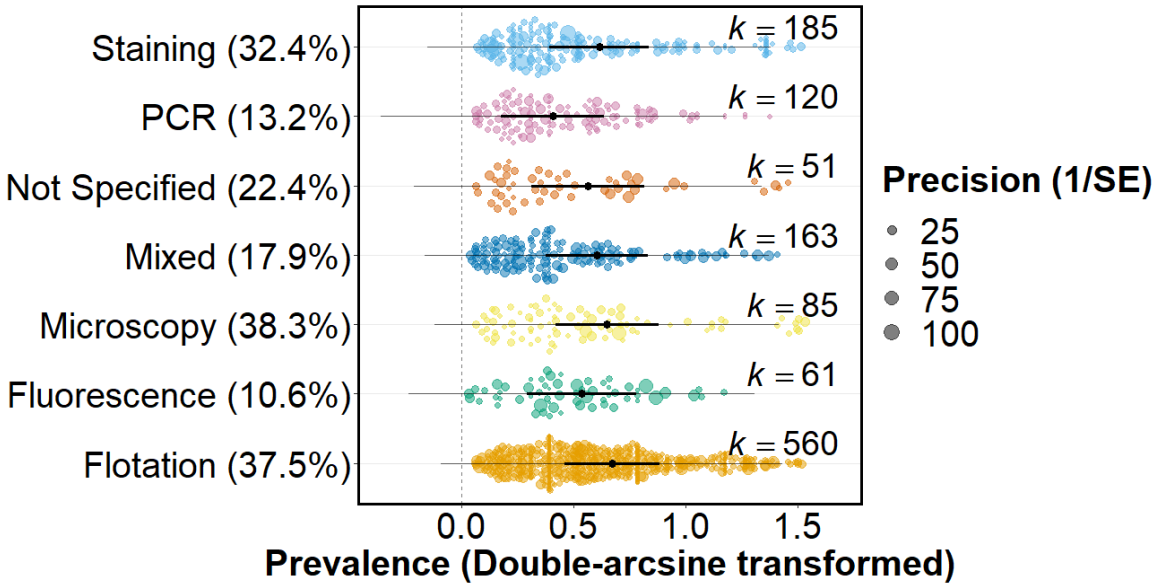
2.3.4.1 Methodological bias

Eight different DNTs were identified in the literature search. The most common were flotation (550 records), staining (185) and PCR (120). 163 records were produced using mixed DNTs. The least common DNTs used were ELISA (5) and sedimentation (7), which were not included in the meta-regression testing for an effect of DNT. There is significant variation in the prevalence of protozoa found in the gut eukaryome of wild rodents according to the diagnostic method used to identify protozoa infections (Q_M = 23.62, p < 0.001, k = 1,225, Figure 2.12). Post hoc comparisons indicated that PCR-based DNTs identified a significantly lower prevalence of protozoa in the gut eukaryome (13.2%) compared to microscopy, flotation, and staining DNTs (38.3%, 37.5% and 32.4% respectively). Using multiple DNTs to identify protozoa in the same hosts did not increase the calculated prevalence in comparison to using just one DNT (except PCR-based DNTs).

A meta-regression did not detect a significant relationship between study precision and protozoa prevalence (Q_M = 0.920, p = 0.338, k = 1,237), indicating that across the whole dataset, larger sample sizes do not reveal a higher prevalence of protozoa.

2.3.4.2 Testing for publication bias

The trim-and-fill test detected asymmetry in the funnel plot, with 187 missing effect sizes being added above the mean (Figure 2.13). Adding these 187 effect sizes adjusted the overall protozoa prevalence from 23.7% (95% CI 4.78-48.5, k= 1,237) to 32.9% (CI 30.6-35.1, k = 1,424). There was no change in prevalence over time (Q_M = 0.023, p = 0.880, k = 1,015).



Post hoc comparisons	*p* value
PCR: Flotation	<0.001
PCR: Microscopy	0.017
PCR: Mixed	0.038
PCR: Staining	0.024

Figure 2.12. The average prevalence of the protozoa according to the diagnostic technique. Prevalence is double-arcsine transformed. The black point indicates the estimated average prevalence. Bold lines indicate 95% confidence intervals for the average prevalence and thin lines indicate the 95% prediction intervals. The size of the points are scaled to precision and k indicates the number of records for that protozoa. The back-transformed predicted prevalence percentage is provided next to the diagnostic technique label. *p* values are shown for post hoc comparisons between those diagnostic techniques with significant differences.

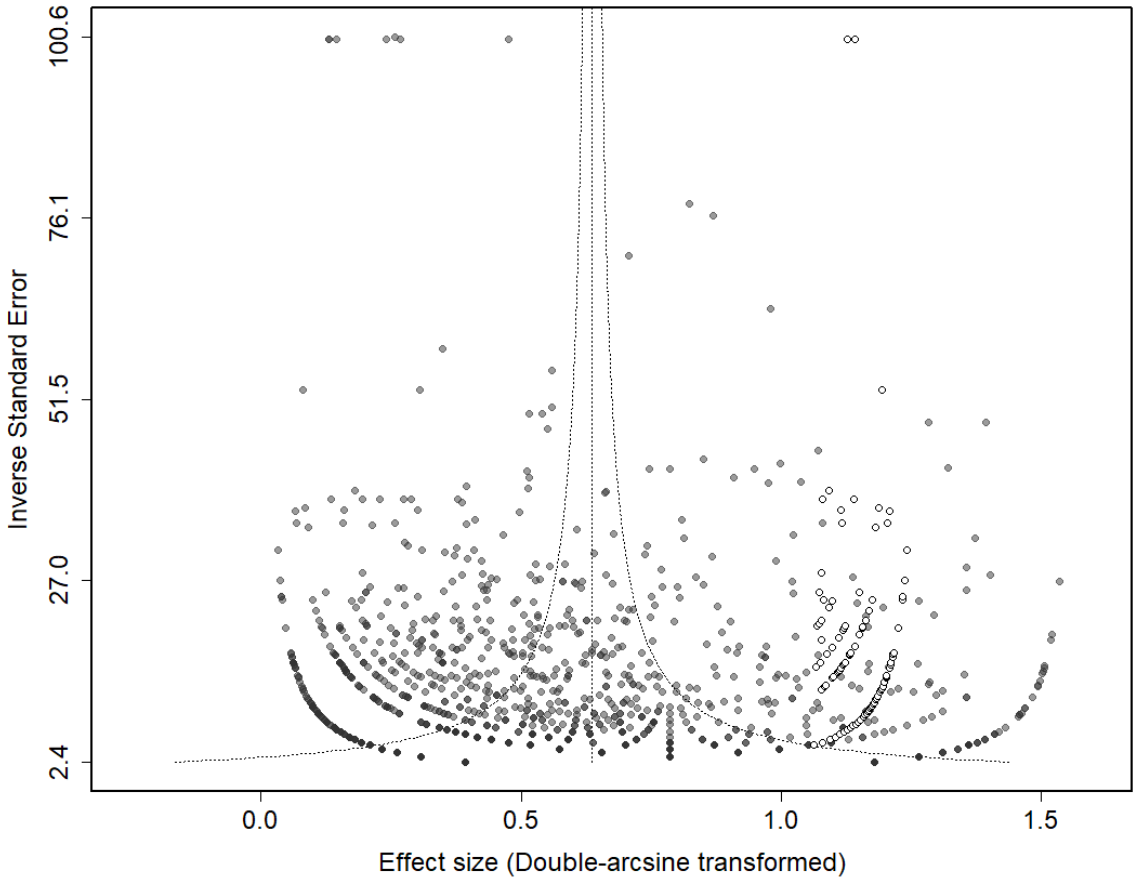


Figure 2.13. Funnel plot indicating the trim-and-fill results. The X-axis is the double-arcsine transformed protozoa prevalence. The inverse of the standard error on the Y-axis represents sample size, where a larger inverse standard error equates to a larger sample size. Filled circles indicate the prevalences extracted from the data whereas open circles indicate missing prevalences. The vertical dotted line indicates the predicted average prevalence when including the missing effect size, and the curved dashed lines are 95% CIs.

2.4 Discussion

2.4.1 The host range of protozoa

44 protozoa genera were identified in the systematic literature search. All five groups of protozoa known to reside within the mammalian gut eukaryome were identified (Parfrey *et al.*, 2011; Parfrey *et al.*, 2014; Hamad *et al.,* 2016). Some genera, *e.g. Cryptosporidium, Eimeria, Entamoeba,* and *Giardia*, occurred in 30 or more host species, indicating it is a common member of the rodent gut eukaryome. This finding is consistent with their wide host range across other vertebrate taxa (Appelbee, Thompson and Olson, 2005; Ryan, Fayer and Xiao, 2014; Duszynski, 2021; Zanetti *et al.*, 2021). *Blastocystis* was found in only eight rodent species in the present study, which contrasts to reports of a wide host range (Alfellani *et al.*, 2013). *Isospora* had a wide host range (22 host species), contrasting to previous suggestions that rodents are not natural hosts of *Isospora* (Trefancová *et al.*, 2019). For protozoa taxa found in the eukaryome of many host species, there has likely been adaptation and diversification to achieve the broad host range identified (Stensvold and Clark, 2016; Seabolt *et al.*, 2021). However, it is also important to note that the protozoa found in the gut eukaryome of multiple host species in the present study are typically considered to be parasitic. Thus, this result may instead reflect a sampling bias towards parasitic protozoa, leading to an underrepresentation of commensal/mutualistic protozoa in the gut eukaryome.

In comparison, some protozoa genera were found in the gut eukaryome of only one or two host species, thus likely having a narrower host range. The ciliate *Balantidium* was found in only two host species, *R. norvegicus* and *Heterocephalus glaber.* This is consistent with reviews suggesting that the mammalian host range of *Balantidium* is generally limited to pigs and primates, with rats acting as potential carriers (Schuster and Ramirez-Avila, 2008). Thirteen Ciliophoran protozoa genera (Class: Litostomatea) identified in the present study were described from a single host species, the capybara. This is consistent with previous reports of the high host specificity of Litostomatea protozoa (Dehority, 1986). The protozoa *Caryospora*, *Toxoplasma* and *Sarcosystis* were each found to be reported from only one host species. Rodents can act as intermediate hosts of these taxa, with the protozoa persisting in muscle tissues, rather than in the gut lumen (Kim *et al.*, 2011; Stacy *et al.*, 2019; Guardone *et al.*, 2022). Their limited host range in the present study suggests that whilst rodents may be considered intermediate hosts, these protozoa are not capable of prolonged colonisation of the gut.

Defining the host range of protozoan taxa needs to consider the matter of accurate protozoa identification. Incorrect identification of protozoa is common, particularly when only using phenotypic traits (Long and Joyner, 1984; Tenter *et al.*, 2002). Furthermore, the idea of a protozoan species only being able to infect one host species can led to the proposal of a new

protozoa species when a protozoan is identified from a previously unknown host species (Wilber *et al.*, 1998; Poulsen and Stensvold, 2016). In turn, this could underestimate the host range of protozoan species. To avoid this complication in the present analysis, protozoa data were collected at the genus level, rather than species. However, even at the protozoa genus level, definitions of host range can be affected by changing protozoa taxonomy. In the present review, *Trichomonas* was reported from 21 rodent species, despite being more commonly associated with the digestive tract of birds or the human vaginal microbiome (Malik *et al.*, 2011). This would suggest that *Trichomonas* has a wide host range among vertebrates. However, some *Trichomonas* spp. are synonymous with *Tritrichomonas* spp. (Burr *et al.*, 2012). *Tritrichomonas* is described in the laboratory rodent gut microbiome (Escalante *et al.*, 2016), but was only reported in one wild rodent species in the present study. Combining the presence records of the synonymous *Trichomonas* and *Tritrichomonas* spp. would provide a more accurate assessment of the rodent host range. Doing this leads to the conclusion that *Tritrichomonas* has a wide rodent host range. Similarly, *Spironucleus muris* is known to colonise the gut of many laboratory rodent species (Jackson *et al.*, 2013) but *Spironucleus* was only reported from three wild rodent species in the current literature survey. However, *Spironucleus* spp. are often misidentified as *Hexamita* spp. and reclassifications are common (Jørgensen and Sterud, 2007; Jackson *et al.*, 2013). *Hexamita,* whilst being better known for infecting fish and birds (Uldal and Buchmann, 1996; Cooper *et al.*, 2004), has records in four rodent species. Combining *Spironueclus* and *Hexamita* presence records leads to the conclusion that it has a wide rodent host range. Clarifying protozoa taxonomy would help improve understanding of the host range of these gut protozoa in wild rodents.

Three of the protozoa genera identified in wild rodents in the present study – *Adelina, Klossia,* and *Monocystis* – are also known to infect arthropods and earthworms (Field and Michiels, 2005; Bekircan and Tosun, 2021; Zeldenrust and Barta, 2021). While these rodent records could be true infections of rodents, it is also possible that their presence reflects the rodent ingestion of arthropods and/or earthworms infected with these protozoa. Furthermore, the amoebae *Acanathomoeba* spp. and *Amoeba* spp. are typically considered to be free-living (Rodríguez-Zaragoza, 1994) but were each identified from one rodent species. These putative rodent infections are more likely transient infections, rather than true members of the rodent gut eukaryome. Similarly, the genus *Pharyngomonas,* originally *Trichomastix* (Park and Simpson, 2015), was described from the naked mole rat, *Heterocephalus glaber,* though it is a halophilic protozoan (Park and Simpson, 2015) and so it unlikely a natural resident of naked mole rat gut.

2.4.2 Protozoa prevalence is heterogenous among host species

This analysis has found that the global protozoan prevalence in wild rodents is 23.7%. This estimate is slightly higher than previous prevalence estimates for individual protozoa genera in wild rodents *e.g.* 18%, 19.8% and 20.1% in *Blastocysti*s, *Cryptosporidium* and *Giardia,* respectively (Li *et al.*, 2017; Zhang *et al.*, 2021; Barati *et al.*, 2022). It is important to note that this global estimate may be an underestimation if considering prevalence irrespective of protozoa genus, as many articles surveyed for a particular genus/species of protozoa, rather than generic protozoa infection.

Whilst there was a significant difference in the protozoa prevalence among rodent host species, there was no difference in the prevalence of different protozoa genera within a host species. This finding, combined with no evidence of geographical effects on protozoan prevalence, suggests that the rodent species-level effect applies to a wide range of different protozoa, perhaps underlined by host species-specific traits or wider demographic effects. The exception to this finding was the muskrat, *Ondatra zibethicus*, which had a significantly higher prevalence of *Giardia* compared to *Cryptosporidium. Giardia* cysts are detected in water more frequently than *Cryptosporidium*, which may explain the comparatively higher prevalence of *Giardia* in the semi-aquatic muskrat (Cacciò *et al.*, 2005; Ganoe *et al.*, 2020). There were no differences in prevalence among different rodent families. For some rodent families – Cricetidae, Muridae and Sciuridae – there were protozoa-level effects, which warrants further investigation into the underlying cause and mechanism.

2.4.3 Host behaviours alone do not explain prevalence variation

The meta-analysis found no effect of host sociality on protozoan prevalence. This is perhaps surprising given that there are rodent species-level effects (above, 2.4.2) and an increasing awareness of the importance of social interactions affecting transmission of gut microbes (Grieneisen *et al.*, 2017; Raulo *et al.*, 2021). In particular, no effect of host population density or home range size was identified, despite prior evidence that both are associated with the chance of incidental transmission of gut microbes in wild mammals (Li *et al.,* 2016; Sarkar *et al..*2020; Wikberg *et al.,*2020). Thus, the findings presented here suggest that other rodent species-level traits are potentially more important than sociality in driving protozoa prevalence variation. Conversely, the majority of data records included in these meta-analyses were from protozoa typically considered to be parasitic. Therefore, it is important to consider that these findings may not be representative of all protozoa in the gut eukaryome, but only those that cause disease. Indeed, when considering parasitic taxa only, other work has shown there is no relationship between rodent sociality and endoparasite load (*e.g.* Bordes *et al.*, 2007; Hillegass *et al.*, 2008), supporting the findings of the present study. Including data on commensal/mutualistic protozoa may highlight how protozoa pathogenicity is an important

consideration as to how host behaviours may be associated with protozoa transmission and thus prevalence in the gut eukaryome.

Other caveats to the results of the meta-analysis presented here are two-fold. Firstly, host sociality was defined at species-level, but intraspecies variation in sociality might be important in parasite transmission variance (*e.g*. Friant *et al.*, 2016) with individual social behaviours varying temporally and spatially (Apfelbach *et al.*, 2005; Previtali *et al.*, 2009; Rémy *et al.*, 2013; Wilson *et al.,* 2017; Hawley *et al.*, 2021). Secondly, this meta-analysis did not account for other aspects of sociality (*e.g.* grooming behaviour, parent-infant feeding) that may impact protozoa transmission and prevalence (Sarkar *et al.*, 2020). Thus, additional data are needed to tease apart which social behaviours, if any, are affecting protozoan transmission and so protozoan prevalence.

There was no effect of host lifestyle on protozoa prevalence. Theory suggests that arboreal and semi-arboreal rodents would have a lower protozoan prevalence (compared to hosts with other lifestyles) because arboreal lifestyles disfavour faecal-oral protozoan transmission (Gilbert, 1997; Barelli *et al.*, 2020a). However, this idea is not supported by the findings reported here. Furthermore, this meta-analysis found no differences in protozoan prevalence between semi-aquatic rodents and those with other lifestyles. This suggests that protozoan transmission is not higher in semi-aquatic species compared with non-semi-aquatic rodents. This contrasts to theory suggesting that parasite and pathogen longevity, and thus opportunities for transmission, in water is increased compared to terrestrial environments (Behringer *et al.*, 2018; Ganoe *et al.*, 2020). These findings may be due to covariation between host lifestyle and other factors that may affect protozoa transmission and thus prevalence. For example, arboreality and group size are negatively correlated, and if each has different effects on protozoan transmission, these host effects would be masked by each other (Sarkar *et al.*, 2020).

For certain lifestyles – arboreal, fossorial, and terrestrial – there were protozoa-level effects. *Eimeria* prevalence, compared to other protozoa genera, was higher in arboreal and fossorial rodents and *Trichomonas* and *Eimeria,* compared to other protozoa, were more prevalent in terrestrial rodents. However, it is important to note that these findings may be driven by protozoa-level effects within the Sciuridae, Muridae, and Cricetidae (above, 2.4.2). Specifically, i) *Eimeria* was more prevalent in the Sciuridae, compared to other protozoa genera, and many Sciuridae species were classed as either arboreal or fossorial and ii) *Trichomonas* and *Eimeria* were more prevalent in the Muridae and Cricetidae, compared to other protozoa genera, and many Muridae and Cricetidae species were classed as terrestrial

rodents. Thus, it is likely that the protozoa-levels effects seen within the arboreal, fossorial, and terrestrial rodents are confounded by rodent family-level taxonomic effects.

The finding of protozoa-level effects throughout these analyses are likely driven by the diversity of traits displayed by the different protozoa genera (Chabé *et al.*, 2017). Thus, protozoa are perhaps too diverse to be generalised into one group for eukaryome research. A better approach may be to consider protozoan species separately to identify what contributes to their presence in the eukaryome. Alternatively, a guild-based approach could be used to group protozoa according to their similarities in how they interact with the host and with other protozoa, as seen with bacteria (Wu *et al.,* 2021).

2.4.4 Considerations of diagnostic method and publication bias

This analysis found that PCR diagnoses resulted in reports of lower protozoa prevalence, compared with other diagnostic methods. This was unexpected because PCR is typically highly sensitive (McHardy *et al.*, 2014; Compton, 2020). However, the lower PCR-reported prevalence may be due to difficulties extracting DNA from protozoan (oo)cycts, whereas (oo)cysts are often readily detected (and diagnosed) by microscopic examination (Hawash, 2014). Furthermore, it is important to note PCR diagnosis (and ELISA) are often very specific, only targeting one protozoan species or genus whereas other diagnostic methods can detect a broader range of taxa (den Hartog *et al.*, 2013; Compton, 2020). To get a better view of the entire protozoa community in the gut eukaryome, metagenomic sequencing (with more rigorous DNA extraction) or multiple diagnostic methods may be more appropriate (Lokmer *et al.*, 2019). This would also help to reduce the potential sampling bias towards parasitic taxa that is discussed above (section 2.4.1). For example, the literature search presented here identified only three protozoa in the gut eukaryome of the striped field mouse (*Apodemus agrarius*), which were identified by targeted surveys for parasitic taxa. In comparison, an 18S amplicon sequencing survey, published after this literature search took place, identified at least 15 protozoa genera in the caecal eukaryome of *A. agrarius*, some of which are considered commensal taxa (Kim *et al.,* 2022).

Publication bias was detected in the dataset, driven by a lack of studies reporting high prevalences. However, publication bias (defined in 2.2.5.4) normally arises from a tendency to not publish studies with less significant results and/or smaller sample sizes (Shi and Lin, 2019). Conversely, one might expect a publication bias in favour of reporting high prevalence of protozoa, and given the comparative rarity of such reports, this suggests that high protozoan prevalence is actually rare. The current study also identified a significant, positive relationship between the surveying effort of a rodent species, and the number of protozoa genera reported from that host species. This emphasises the need to continue to undertake wider surveys of

rodents and other animals, particularly those less frequently studied, to find the full biodiversity of gut protozoa.

2.4.5 Conclusion

This analysis is the first, of which I am aware, synthesis of the description of the gut protozoa of wild rodents, providing an unbiased summary of host records for protozoa genera. The analysis also provides the first estimate of protozoa global prevalence in wild rodents, and points to host species-level effects affecting protozoan prevalence, although the exact cause could not be identified. To further the understanding of the context-specific role of host behavioural traits, better data on a host's individual- and population-level traits will be required. Furthermore, considering the transmission mechanism of gut protozoa would complement studies aiming to investigate host drivers of protozoa prevalence in the eukaryome.

This meta-analysis has identified some key limits within the literature that limit our ability to identify factors affecting protozoa presence in the gut eukaryome. Firstly, the extensive protozoa taxonomic reclassifications and revisions make it hard to define, even at the genus level, which protozoa can colonise the gut of which rodent hosts. To this end, there is a need for a consensus on the taxonomy and identification of different gut protozoa. Secondly, there is a current bias in the literature towards reports of disease-causing protozoa, whilst the commensal taxa are less frequently surveyed. Future studies would benefit from a broader approach that allows the identification of all potential gut protozoa, rather than focusing on those that are already well known. Finally, I make the case that studies need to consider the context of the host and its biology to fully understand how different host traits and behaviours interact to impact the composition of the eukaryome.

Chapter 3: Using flow cytometry to quantify and describe the gut eukaryome of house mice

Abstract

Accurate characterisation of the diversity and composition of the mammalian gut microbiome is critical for understanding host-microbiome interactions. As such, quantifying the absolute abundance of microbial cells in the gut is advantageous compared with the more commonly used measurement of relative abundance. However, absolute quantification methods of gut microbes have rarely been optimised for eukaryotic taxa. In this chapter, I explore the use of flow cytometry to identify and quantify microbial eukaryotes from faecal samples. Specifically, I aimed to identify and quantify eukaryotic cells by fluorescently staining their endoplasmic reticulum, a eukaryotic-specific organelle. Subsequent separation of the putative eukaryotic cells from prokaryotic cells by fluorescent-activated cell sorting (FACS) allows the use of 18S rRNA sequencing to identify, and thus quantify, the taxa in the gut eukaryome. I tested the accuracy of the proposed method by sequencing putative eukaryotic and prokaryotic cells from FACS-processed faecal samples from captive house mice (*Mus musculus domesticus*). I found that the eukaryotic-specific stain was not accurately identifying solely eukaryotic taxa. Additionally, I compared FACS-processed faecal samples to non-FACS-processed faecal samples which showed that FACS-processing was artificially altering the gut eukaryome composition of the house mice. These data show that the proposed method is unsuitable for quantification of taxa in the gut eukaryome without further optimisation. In summary, this study highlights the complexity of quantifying gut eukaryotes and provides the groundwork of a method of quantification that future studies can build on.

3.1 Introduction

3.1.1 Diversity metrics used to understand the gut microbiome

The gut microbiome is a large community of many, diverse microbes, thus an ecological approach is required to understand interactions among these microbes (McDonald *et al.*, 2020; Coyte *et al.*, 2021). Diversity is a widely used quantitative measurement in ecology that has been applied to gut microbiome studies (Johnson and Burnet, 2016; Roswell *et al.*, 2021). In particular, diversity metrics are important for measuring how taxa in the gut microbiome respond to a changing environment, whether this stems from interactions between and among species, or from interactions with the host (Lozupone *et al.*, 2012). Despite the continual use of diversity measures to understand the ecology of the gut microbiome, the methodologies and approaches used in microbiome studies are not standardised (Willis, 2019; Risely *et al.*, 2021b). Here I briefly outline the use of common diversity metrics in gut microbiome studies (see Wagner *et al.* (2018) for an in depth review).

Alpha and beta diversity are two fundamental, commonly used, measures of microbiome diversity (Jost, 2007; Chao and Chiu, 2016; Roswell *et al.*, 2021). Alpha diversity summarises the gut microbiome community, and is calculated using species richness (number of species) and/or species evenness (distribution of the abundance of different taxa) (Willis, 2019). As species richness and species evenness increase, so does alpha diversity. The two most common measures of alpha diversity are the Shannon and Simpson indices. Simpson's index represents the probability that two random microbes within a community belong to different species, whereas Shannon's index is the probability of correctly predicting the identity of a random microbe in the community (Nagendra, 2002; Morris *et al.*, 2014; Roswell *et al.*, 2021). Both Shannon's and Simpson's indices are used in gut microbiome studies: some studies use both to measure alpha diversity (Goertz *et al.*, 2019; Strickland *et al.*, 2021), whereas others use only one (Linnenbrink *et al.*, 2013; Jenkins *et al.*, 2018). Furthermore, some studies use neither, opting for a different measure of alpha diversity *e.g.* Chao 1 richness (Weldon *et al.*, 2015).

Beta diversity, in contrast to alpha diversity, measures the difference in species composition between two gut microbiomes (Koleff *et al.*, 2003; Chao and Chiu, 2016). As two gut microbiomes differ more from one another, because they share less species, beta diversity increases. Beta diversity can be used to measure which factors correlate with changes in the gut microbiome community composition between hosts or within the same host at different time points (Anderson *et al.*, 2011; Lozupone *et al.*, 2012). In gut microbiome studies, common beta diversity measures include Bray-Curtis dissimilarity and UniFrac distance (Linnenbrink *et al.*, 2013; Goertz *et al.*, 2019; Suzuki *et al.*, 2019a; Strickland *et al.*, 2021).

3.1.2 Calculating taxa abundance in gut microbiome studies

Measuring diversity in the gut microbiome requires data on the abundance of the microbial taxa present. Commonly, these abundance data are derived from amplicon sequencing methods. Briefly, amplicon sequencing amplifies and then sequences specific target gene fragments from DNA extracted from microbial cells. Genes commonly amplified and sequenced are those that are conserved among species (*e.g.* the 16S rRNA gene of bacteria), but which still contain sufficient sequence diversity among taxa to identify species (Větrovský and Baldrian, 2013). After sequencing, gene sequences are compared and each unique sequence identified is defined as an amplicon sequence variant (ASV). The number of reads (sequences) per ASV is used as a measure of ASV abundance, thus allowing diversity metrics to be calculated using this ASV abundance data. Consequently, by assigning taxonomy to an ASV, one can then estimate the abundance of that taxon in the sample. For example, if there are two ASVs in a sample, each with an read abundance of 100, the taxon assigned to each of the ASVs are equally abundant in that gut microbiome sample.

Despite its common use in gut microbiome studies, there are issues with using ASV read abundance to calculate diversity metrics (Gloor *et al.*, 2017; Morton *et al.*, 2019). In particular, the total number of reads in a sample (read depth) can vary among samples. Variation in read depth can result from random technical variation in the sequencing of different samples or from differences in the initial quantity of DNA to be sequenced (Gloor *et al.*, 2017). Thus, some ASVs may appear to be more abundant in some samples for non-biological reasons. To account for this, many studies compare the relative abundances of ASVs rather than absolute abundances (Gloor *et al.*, 2017; Zaheer *et al.*, 2018). For example, if a sample has 10 ASVs, each with an absolute abundance of 10, the total read depth is 100 and the relative abundance of each ASV is 10%. If a second sample had the same ASVs, but each with an absolute abundance of 100, the total read depth is increased to 1,000. However, the relative abundance of each ASV would still be 10%, thus accounting for non-biological variance in read depth. Rarefaction is another approach used to account for differences in read depth among samples (Weiss *et al.*, 2017). Rarefaction is done by randomly discarding reads from those samples with a comparatively greater read depth until all samples have the same total read depth. However, the problem with rarefaction is that it can result in the loss of large quantities of data (McMurdie and Holmes, 2014) and so can result in the loss of low abundance ASVs, with consequent effects on estimates of diversity. Studies can also compare the ratios of taxa abundance among gut microbiomes, to determine which taxa have changed the most relative to other taxa (Morton *et al.*, 2019; Barlow *et al.*, 2020). Examples of methods developed using this approach include analysis of composition of microbiomes (ANCOM) and Gneiss (Mandal *et al.*, 2015; Morton *et al.*, 2017).

Further to the absolute abundance of ASVs being dependent on the read depth of a sample, sequencing depth can also impact the number of ASVs identified. Specifically, with deeper sequencing, more ASVs are identified, and these ASVs are usually from low-abundance taxa (Zaheer *et al.*, 2018). Thus, if a lower sequencing depth is used (often because of budgetary limits), low-abundance taxa are less likely to be sequenced which therefore under-represents diversity (Sims *et al.*, 2014; Zaheer *et al.*, 2018). Furthermore, even when very low-abundance ASVs are captured in microbiome studies, they are often excluded from microbiome analyses because they can be considered to be sequencing errors or contamination of the sample (Cao *et al.*, 2021).

The gene copy number (GCN) of the targeted gene when amplicon sequencing can also contribute to issues with using ASV abundance data for calculating diversity metrics. Often, the GCN of commonly targeted genes is not equal across microbial taxa (Stoddard *et al.*, 2015; Lofgren *et al.*, 2019; Salmaso *et al.*, 2020). Taxa with a high GCN will have more amplicons generated per genome during PCR amplification, compared to taxa with a low GCN. Thus, for high GCN taxa, more sequences will be generated per individual cell causing these taxa to appear more abundant than they actually are (Silverman *et al.*, 2021), skewing diversity measures and subsequent analyses. Variation in GCN among taxa can be addressed by incorporating the GCN into calculations of relative abundance. As such, publicly available databases can be used to access GCN information (Stoddard *et al.*, 2015). If GCN is known, correcting for variance can be done by dividing the absolute abundance of taxa, estimated from ASV abundance, by the GCN in their genome (Vandeputte *et al.*, 2017; Jian *et al.*, 2020). However, accounting for GCN does not always improve estimates of taxa abundance, especially if the GCN of many taxa is unknown (Louca *et al.*, 2018; Starke *et al.*, 2021).

3.1.3 Microbial load quantification

Microbial load is defined as the absolute number of microbial cells in a sample, which can vary ten-fold among individuals (Vandeputte *et al.*, 2017). However, ASV abundance data does not provide accurate, quantitative information about microbial load, nor does it identify differences in microbial load among samples. For example, if two samples have the same species evenness, but one has a much higher microbial load, the relative abundances of the taxa will appear similar, despite one sample having many more individual cells of each taxon (Stämmler *et al.*, 2016). Knowing the microbial load of samples allows a more accurate view on microbial interactions within the gut microbiome, because it can better identify the co-variation in taxa abundance that is used to infer interactions among taxa (Vandeputte *et al.*, 2017).

The number of studies now incorporating the quantification of microbial load into their analyses of the gut bacterial microbiome (bacteriome) is increasing, with three common approaches

being used: spiking, quantitative PCR (qPCR) and flow cytometry (FC) (Galazzo *et al.*, 2020; Wang *et al.*, 2021). Firstly, spiking involves adding known quantities of unique bacteria (or DNA) into samples prior to DNA extraction (Stämmler *et al.*, 2016; Tourlousse *et al.*, 2017). The relative abundances of endogenous microbiome taxa can then be compared to the known abundance of the spiked-in bacteria/DNA, so allowing accurate absolute quantification of microbial taxa. Secondly, qPCR has been used post-DNA extraction to calculate absolute abundance of microbial taxa (Jian *et al.*, 2020; Callegari *et al.*, 2021). In gut microbiome studies using the 16S rRNA gene, qPCR quantifies the number of 16S rRNA gene amplicons in the starting faecal DNA template, to estimate the number of bacterial genomes per gram of faeces (Smith and Osborn, 2009; Jian *et al.*, 2020). The estimated number of bacterial genomes per gram of faeces is then used in conjunction with relative abundance data from amplicon sequencing to calculate absolute abundance of each microbial taxon (Jian *et al.*, 2020).

Flow cytometry (FC) has been used widely to quantify microbial taxa obtained from soil, water, and physical substrates (Whiteley *et al.*, 2003; Prest *et al.*, 2013; Heinrichs *et al.*, 2021; Xu *et al.*, 2021). In these studies, microbial cells were stained using fluorescent dyes and FC used to measure the fluorescence of individual particles, thus allowing the fluorescently-stained microbes to be distinguished from background debris, and then quantified. For gut microbiome studies, FC has been used to quantify bacterial cells in faecal samples (Vandeputte *et al.*, 2017). In doing so, microbial load data combined with amplicon sequencing data can both quantify and identify the taxa in the gut microbiome. Stained cells can also be separated from background debris in a process called fluorescence-activated cell sorting (FACS) for further work. There are many criteria by which cells can be FACS-sorted, allowing the quantification and identification of different cell types, depending on the biological question. For example, staining microbial cells for IgA – a host immunoglobulin – allowed the isolation and characterisation of gut microbial cells that elicited a host immune response (Palm *et al.*, 2014). In addition, by staining for the presence of RNA, FACS has also been used to identify metabolically-active cells from the faecal microbiome (Peris-Bondia *et al.*, 2011). Other examples of microbial load quantification using FACS and sequencing include: staining for membrane integrity, allowing the quantification of intact, damaged and dead faecal bacterial cells (Ben-Amor *et al.*, 2005) and staining for nucleic acid content and cell membrane changes to identify the response of faecal microbial cells to xenobiotics (Maurice *et al.*, 2013).

3.1.4 Calculating eukaryome diversity and microbial load

Studies of the eukaryotic taxa in the gut microbiome (eukaryome) commonly sequence amplicons of the 18S rRNA gene to generate taxonomy and abundance data (Scanlan and Marchesi, 2008; Amaral-Zettler *et al.*, 2009; Stoeck *et al.*, 2010; Hadziavdic *et al.*, 2014;

Parfrey *et al.*, 2014; Popovic *et al.*, 2018). However, the 18S rRNA gene is also shared with the mammalian host, and mammalian DNA from host cells can dominate eukaryome DNA. In such cases, the dominating host DNA can block eukaryome template DNA from being amplified at all, severely limiting the ability to detect low abundance taxa, thus compromising measures of eukaryome diversity (Polz and Cavanaugh, 1998; Green and Minz, 2005; Pereira-Marques *et al.*, 2019). To overcome this problem, blocking primers can be used to prevent host DNA from being amplified (Vestheim and Jarman, 2008) but these methods are not fully effective (Thompson *et al.*, 2017; del Campo *et al.*, 2020). Additionally, the GCN of the 18S rRNA gene is much more variable among eukaryotic species compared to its 16S counterpart (Lofgren *et al.*, 2019; Salmaso *et al.*, 2020; Lavrinienko *et al.*, 2021b). Thus, for the eukaryome, there is an even greater challenge in using ASV abundance data to calculate diversity metrics, compared to the bacteriome (Gong and Marchetti, 2019).

To this end, there is a need for an accurate method to quantify the microbial load of eukaryotic cells in the mammalian gut. To date, eukaryome quantification has primarily focussed on gut fungi (mycobiome), using spike-in or qPCR methods (Dollive *et al.*, 2013; Haak *et al.*, 2021; Rao *et al.*, 2021). These studies have found that the mycobiome is highly variable within individuals, and inter-kingdom interactions contribute to the microbial load of fungal taxa (Dollive *et al.*, 2013; Haak *et al.*, 2021; Rao *et al.*, 2021).

The work presented here investigates the use of FC to identify and quantify all microbial eukaryotes in the gut microbiome, using a eukaryotic-specific stain. I hypothesise that there are differences in the microbial load of the gut eukaryome and bacteriome, and test this using the proposed quantification method. Then, I predict that there is variation among individual hosts in their microbial load, and test the hypotheses that differences in faecal microbial load may also arise from host sex and the month of sampling. Sorting of the putative eukaryotic or prokaryotic cells by FACS allows the use of amplicon sequencing to generate diversity metrics and to identify the taxa in the gut eukaryome. I use such amplicon sequencing data to confirm the accuracy of the eukaryotic-specific stain. Then, I aim to check that the proposed method does not artificially alter the gut eukaryome in comparison to traditional methods. To do this, I test the prediction that there are no differences in measures of eukaryome diversity and composition when comparing between samples processed using the proposed method and samples processed using a traditional methodology. If these predictions are met, the proposed method offers an alternative to qPCR and spike-ins for generating eukaryome microbial load data. In doing so, the method could then be used to improve how we calculate diversity in the gut eukaryome and improve understanding of the wider mammalian gut eukaryome. The work focussed on the faecal microbiome of captive house mice, *Mus musculus domesticus*.

3.2 Methods

To test if FACS was achieving the aim of separating eukaryotic and prokaryotic cells from a faecal sample to allow the quantification of eukaryotic cells, a large-scale experiment was conducted (Figure 3.1). The cells at each stage of the FACS process were sequenced and the resulting ASVs compared to see if the aims had been achieved. The development of the methods used here can be found in Appendix 1.2.

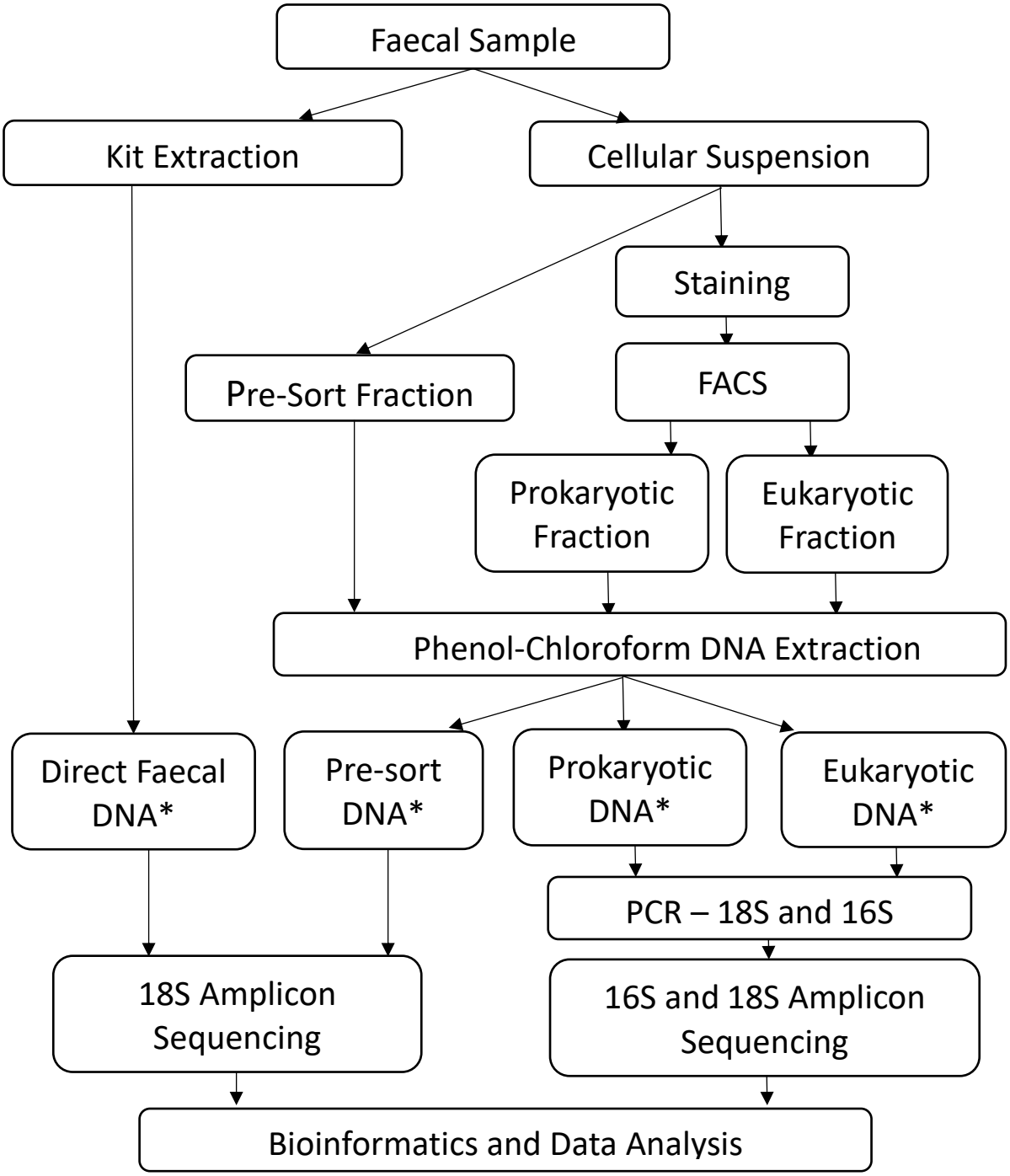


Figure 3.1. The process used to test whether FACS can separate eukaryotic cells from prokaryotic cells. Four stages of the process had their DNA extracted (*), which allowed comparisons between traditional sequencing methods and the proposed method.

3.2.1 Sample preparation and cell counting

3.2.1.1 Sample preparation

The study used three female and three male house mice from a captive colony, which was derived from wild mice ancestors caught from across northwest England. Three females and three males were used to account for potential differences in microbial load between the sexes. For each mouse, faecal samples were collected three times across a three-month period, giving a total of 18 samples. Samples were stored at -80 °C before processing. For each faecal sample, four different fractions were prepared (Figure 3.1): i) Direct Faecal, ii) Pre-Sort, iii) Eukaryotic and iv) Prokaryotic. To prepare the Direct Faecal fraction, 200 mg of faeces was used for DNA extraction using the QIAamp PowerFecal Pro DNA Kit (Qiagen). The DNA was stored at -80 °C until sequencing. This method is the standard microbiome preperation method and so acts as a control for the novel method tested here. To prepare the next three fractions, a cellular suspension was first prepared (Figure 3.1). To do this, 100 mg of faeces was added to 1 mL phosphate-buffered saline (PBS, Sigma-Aldrich) and incubated for 1 hour on ice in a 2 mL microcentrifuge tube with regular vortexing. The mixture was then homogenised using a pellet pestle motor (Kontes) and blue polypropylene pellet pestles (Sigma-Aldrich), for approximately two minutes. A new pestle was used for each sample. This homogenate was then centrifuged at 50 x g for 15 minutes at 4 °C to pellet faecal debris, leaving a supernatant of microbial cells. 100 µL of this supernatant was then transferred into 1 mL of staining buffer (1% w/v bovine serum albumin (BSA, Sigma-Aldrich) in PBS). The sample was then centrifuged at 8,000 x g for 5 minutes at 4 °C to pellet the microbial cells. The supernatant was then discarded and the pellet resuspended in 1 mL fresh staining buffer. The Pre-Sort fraction (Figure 3.1) consisted of 20 µL of this suspension to which 180 µL of fresh staining buffer added, and this was then stored at -80 °C until DNA extraction. DNA was extracted from the Pre-Sort fraction using the method described in Palm *et al.* (2014).

Prior to preparing the samples for FACS, a negative control was prepared, which consisted of 40 µL of the suspended cellular suspension (above), added to 500 µL fresh staining buffer. The remainder of the cellular suspension was then stained for FACS. Two stains were used: SYBR green (SG) and ER-Tracker Red (ERT). SG stains DNA, thereby putatively staining living cells. The ERT specifically stains the sulphonylurea receptors of ATP-sensitive potassium channels, which are found predominantly on the endoplasmic reticulum, a eukaryotic-specific organelle, and so ERT staining should be specific for eukaryotic cells (Appendix 1.2).

Staining was done by adding 1 µL of SG working solution to the remaining cellular suspension, followed by incubation at 37 °C for 20 minutes in the dark. The SG working solution was made by diluting 10 µL of SYBR green I (10,000X concentrate, S7563, Invitrogen™) in 990 µL

dimethyl sulfoxide (DMSO, Sigma-Aldrich). The solution was then filtered through a DMSO-resistant, 0.22 µm pore syringe filter using a 2.5 mL syringe, into a new microcentrifuge tube. This step removed contaminating bacteria from the staining solution. After staining with SG, the samples were washed once with 1 mL staining buffer. Then, samples were stained with ERT. To do this, the samples were centrifuged at 8,000 x g for 5 minutes at 4 °C to pellet the cells, the supernatant removed and the pellet resuspended in 1 mL of Hanks' Balanced Salt Solution (HBSS, Gibco). 1 µL of ERT working solution was added to the sample and then incubated at 37 °C for 30 minutes in the dark. The ERT working solution consisted of 1 µL of ER-Tracker Red (Invitrogen™) diluted in 999 µL of DMSO, filter sterilised as above. Once stained with ERT, the samples were washed once with 1 mL HBSS and then resuspended in 1 mL staining buffer. All staining work was carried out in the dark with solutions covered. Samples were kept on ice, in the dark, until used in FACS.

3.2.1.2 Flow cytometry and gating

Flow cytometry was performed using a BD FACSAria™ III flow cytometer. Fluorescence intensity was collected at 530 nm ± 30 for SG and 610 nm ± 60 for ERT. Additionally, sideward (SSC) and forward (FSC) scatter light data were collected. Data were processed using FACS Diva software and electronic gating was used to identify events that gave positive signals for each stain type (Figure 3.2). Measurements were performed at a pre-set flowrate of 5,000-10,000 cells/second, using a 100 µm nozzle. A threshold value of 600 was used.

Density plots of side scatter and green fluorescence generated from the SG stain allowed for the identification of events that were likely cells (Figure 3.2A) and electronic gating selected for these cells. These selected cells were then secondarily gated using density plots showing side scatter and red fluorescence generated from the ERT stain (Figure 3.2B), separating putative prokaryotic (ERT-) and putative eukaryotic cells (ERT+), thereby generating the Prokaryotic and Eukaryotic fractions, respectively (Figure 3.1).

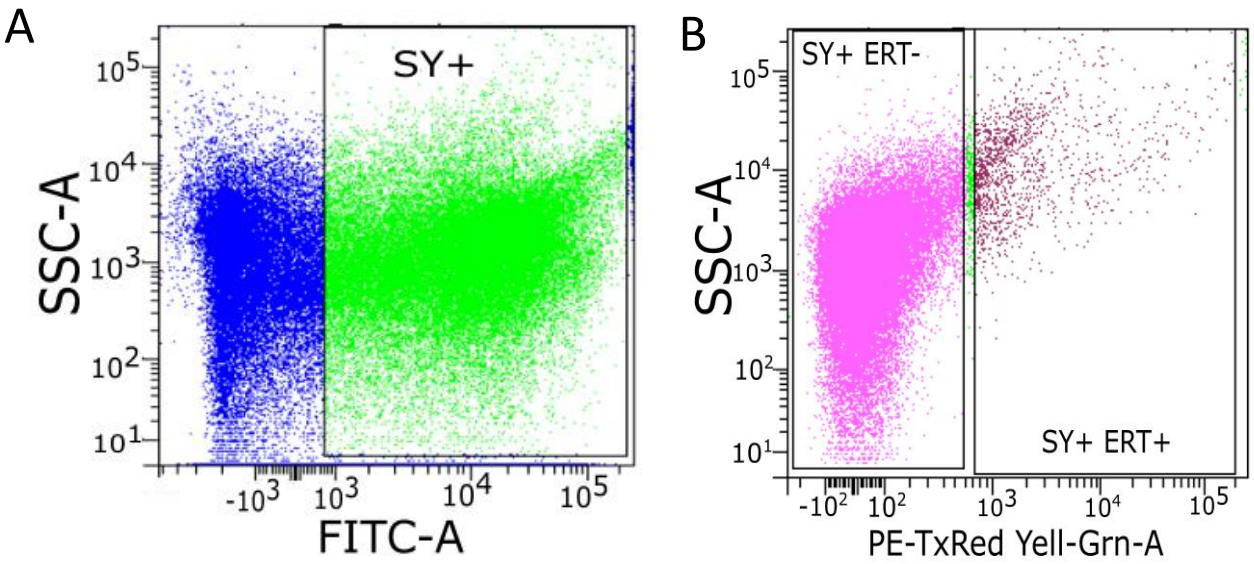


Figure 3.2. Gating strategy used to separate cells according to their staining properties. A) Putative cells are selected based on their ability to be stained with SG (FITC-A x-axis) and their sideward scatter (SSC-A y-axis). Events that fall within the 'SY+' (green) gate are defined as cells. B) Putative cells are separated based on their ERT staining (PE-TxRed Yell-Grn-A x-axis) and SSC-A. The 'SY+ERT-' (pink) gate selects those cells that do not stain with ERT, thus identifying putative prokaryotic cells. Cells in the 'SY+ERT+' (red) gate are classed as putative eukaryotic cells. FACS-sorting of the putative prokaryotic and eukaryotic cells forms the Prokaryotic and Eukaryotic fractions, respectively. Data are from one representative sample.

3.2.1.3 PCR confirmation of prokaryotic *vs*. eukaryotic gating accuracy

PCR was used to determine if the gating strategy accurately separated prokaryotic and eukaryotic cell types. To do this, each fraction was PCR amplified twice, using 16S and 18S rRNA primers separately (Table 3.1), with the expectation that the Eukaryotic fraction would only result in 18S amplicons and the Prokaryotic fraction in 16S amplicons. DNA was extracted from the Eukaryotic and Prokaryotic fractions using the method described in Palm *et al.* (2014). 1.5 µL of DNA template was used in the 15 µL PCR reaction. The cycling conditions were 2 mins at 94 °C followed by 35 cycles of: 45 secs at 94 °C; 45 secs at 54 °C; 1 min at 72 °C, followed by 5 mins at 72 °C for extension. Successful amplification was defined as the presence of a band following gel electrophoresis on a 1% w/v Tris-acetate-EDTA (TAE) agarose gel run at 150 V for 30 mins.

Table 3.1. Primers used to amplify the Eukaryotic and Prokaryotic fractions for confirming Prokaryotic *vs.* Eukaryotic gating accuracy.

Primer Name	Sequence	Region	Reference	Target
N341 (F)	CCTAYGGGRBGCASCAG	V3-V4 16S rRNA gene	Novogene, 2023	Prokaryote
N806 (R)	GGACTACNNGGGTATCTAAT			
N528 (F)	GCGGTAATTCCAGCTCCAA	V4 18S rRNA gene	Novogene, 2023	Eukaryote
N706 (R)	AATCCRAGAATTTCACCTCT			

3.2.2 DNA sequencing and bioinformatics

3.2.2.1 Amplicon sequencing

To identify the eukaryotic taxa at each stage of the FACS process, 18S rRNA amplicon sequencing was carried out for each of the four fractions (Figure 3.1). The V4 region of the 18S rRNA gene was sequenced using the 528F-706R primer pair (Novogene, 2023) *via* 250 bp pair-end Illumina sequencing on a NovaSeq 6000 platform. Library preparation and sequencing was performed by Novogene Co., Ltd. Prior to sequencing, the sorted Eukaryotic and Prokaryotic fractions were PCR amplified as described in 3.2.1.3 to generate sufficient DNA for sequencing. For each sorted fraction, 1.5 µL of template DNA was amplified in five replicates. The amplicons from the five replicates were then pooled, purified using the QIAquick PCR Purification kit (Qiagen), and the DNA quantified *via* Qubit (Thermofisher). This method was repeated for those samples where more DNA was required. A blank extraction was used as a negative control and a mock community as a positive control (ZymoBIOMICS™ D3605). The Eukaryotic and Prokaryotic fractions were also sequenced by targeting the V4-V5 region of the 16S rRNA gene, using the 341F and 806R primer pair (Novogene, 2023), following the same amplification and sequencing protocol as above.

3.2.2.2 Bioinformatics

Sequence data were analysed using QIIME 2 2021.2 (Bolyen *et al.*, 2019). The paired-end reads were merged and the data were quality filtered and denoised using DADA2 (Callahan *et al.*, 2016). Both forward and reverse reads were truncated at 200bp to account for a decrease in sequencing quality. The ASVs produced were then aligned using the mafft programme (Katoh *et al.*, 2002) and a phylogeny based on ASV sequence similarity created using fasttree (Price *et al.*, 2010). The ASVs were then assigned to taxa using a naïve Bayes taxonomy classifier trained against the SILVA 138 99% OTUs reference sequences (Quast *et al.*, 2013; Bokulich *et al.*, 2018). The data were then transferred into R Studio using the qiime2R package (v0.99.6, Bisanz, 2018) for subsequent analysis. All statistical analyses were carried out in RStudio.

3.2.3 Identifying the success of the method

3.2.3.1 Microbial load quantification

Quantification was based on the number of FACS events gated into either the Eukaryotic or Prokaryotic fraction. Events in the Eukaryotic fraction represented the eukaryotic cell count and events in the Prokaryotic fractions represented the prokaryotic cell count. Cell counts are reported as microbial load (cells/g of faeces) to compare to other studies. A Pearson's test of correlation was performed between Eukaryotic and Prokaryotic microbial loads. The microbial load of the two sorted fractions were compared using a t-test. A two-way analysis of variance (ANOVA) was used to assess if the sample collection month or mouse ID affected the microbial load. Differences in microbial load between sexes were compared using a t-test. In all these analyses, the microbial load abundance data were square-root transformed cell counts. To calculate the absolute abundance of individual microbial taxa, the proportionate abundance of each ASV was multiplied by the eukaryotic/prokaryotic microbial load as previously reported (Vandeputte *et al.*, 2017; Jian *et al.*, 2020). In instances where both 16S and 18S amplification occurred for a FACS-sorted fraction, the starting ratio of eukaryotic to prokaryotic cells is not known, and so this method cannot be used to calculate absolute abundance of these taxa.

3.2.3.2 Confirmation of the separation of prokaryotic and eukaryotic cells

The R phyloseq package (v1.38.0, McMurdie and Holmes, 2013) was used to filter out unclassified reads and singletons from the data. The number of uniquely observed 18S and 16S ASVs was calculated for each fraction of each sample, and these were compared to determine how ASV representation changed during FACS processing. An ANOVA was used to compare how the fraction type impacted the number of unique 18S ASVs observed, whilst a t-test was used to compare the number of unique 16S ASVs in the Prokaryotic and Eukaryotic fractions. To compare alpha diversity, Shannon's index was calculated using the

phyloseq package with the number of reads for each 18S ASV used as the measure of abundance. Differences were tested using an ANOVA.

3.2.3.3 Measuring changes in the eukaryome composition

Three measures of change in inter-sample variation were calculated: i) the number of shared ASVs, ii) the variation within fractions, and iii) beta diversity. Firstly, to simplify analyses on the number of shared ASVs between samples, a maximum of 10 18S ASVs per fraction were examined. The limit of 10 was established because for the majority of fractions and samples, the 10 most abundant ASVs accounted for ≥ 70% of reads. The 10 most abundant 18S ASVs among samples were compared to assess which 18S ASVs were shared among samples. This was done for each fraction type. Secondly, the amount of variation within the fraction types was calculated using the betadisper function, with adjustment for sample size bias (vegan package, v2.5.7, Oksanen *et al.,* 2020). The betadisper function generates the median distance of the samples within each fraction to the fraction centroid in multivariate space. The permutest function (vegan package) was used to compare the amount of variation among the different fractions. Thirdly, Bray-Curtis dissimilarity (BC) was used as the measure of beta diversity. BC was calculated using the relative abundance of ASVs, for those with a minimum read abundance of 5 in ≥ 5% of samples (Cao *et al.*, 2021), and a principal coordinate analysis (PCoA) was used to visualise the data. A permutational multivariate analysis of variance (PERMANOVA) was used to test for differences in ASV composition among the fractions, *via* the adonis function (vegan package). Pairwise comparisons were conducted among the four fractions, using the Benjamini-Hochberg (BH) correction for multiple testing (Benjamini and Hochberg, 1995).

The change in proportional abundance of the top 10 most abundant 18S was compared among fractions. Then, each 18S ASV was categorised as likely belonging to one of four groups, based on their taxonomic assignment (described in 3.2.2.2): Food which was identified as belonging to the Phragmoplastophyta, with the majority of 18S ASVs consisting of wheat (*Triticum aestivum*) or corn (*Zea mays*); Fungi, identified via phyla name; Host which was classified as belonging to Vertebrata; Other which consisted of all other 18S ASVs. 18S ASVs grouped on their phyla name were confirmed to be Fungi or Other using the NCBI Taxonomy Browser (Schoch *et al.*, 2020). The representation of 18S ASVs in these four categories through the FACS process was then compared.

3.3 Results

3.3.1 Microbial load

The microbial load of the Prokaryotic and Eukaryotic fractions was determined; it was predicted that the Prokaryotic fraction would have a higher load than Eukaryotic fraction. This prediction was supported: Eukaryotic and Prokaryotic mean microbial load = 3.2×10^5 (SE = 5.4×10^4) and 1.0×10^8 (SE 1.0×10^7) cells/g of faeces, respectively (Table 3.2). These loads were significantly different ($t_{17.35} = -18.8$, $p < 0.001$). Total microbial load ranged from 4.8×10^7 to 1.9×10^8 (average: 1.0×10^8, SE: 1.0×10^7). There is a slight negative correlation between the Prokaryotic and Eukaryotic microbial loads – as Prokaryotic microbial load increases, Eukaryotic microbial load decreases. However, this correlation was not significant ($r = -0.304$, df = 16, $p = 0.220$, Figure 3.3). There was no significant difference in the Prokaryotic microbial load between males and females ($t_{15.99} = -0.62$, $p = 0.542$). Whilst females tended to have a higher Eukaryotic microbial load compared to males, the difference was not significant ($t_{15.95} = 2.01$, $p = 0.062$). Furthermore, the variation in the microbial loads within the sexes was as large as the variation between the sexes. Neither mouse ID nor month of sample collection had a significant effect on the Eukaryotic or Prokaryotic fractions' microbial load (Table 3.3).

The presence of visible bands on gel electrophoresis for both 16S and 18S PCR products indicated that both fractions contained both prokaryotic and eukaryotic cell types. This result was confirmed by amplicon sequencing (section 3.3.2). Because of this, the absolute abundance of taxa cannot be calculated using relative abundance data obtained from amplicon sequencing, as proposed in 3.2.3.1.

Table 3.2. The microbial load (cell counts) for the Prokaryotic and Eukaryotic fractions. Microbial load is expressed as cell counts/g of faeces.

	Month 1		Month 2		Month 3	
Mouse	**Prok**	**Euk**	**Prok**	**Euk**	**Prok**	**Euk**
Male 1	1.62×10^8	1.81×10^5	1.62×10^8	1.92×10^5	4.72×10^6	5.47×10^5
Male 2	9.18×10^6	2.85×10^5	1.07×10^8	4.28×10^4	1.17×10^8	2.04×10^5
Male 3	4.94×10^6	7.19×10^3	1.36×10^8	2.73×10^5	9.09×10^6	2.42×10^5
Female 1	5.58×10^6	3.98×10^5	1.11×10^8	9.42×10^5	1.30×10^8	1.77×10^5
Female 2	7.38×10^6	3.45×10^5	5.18×10^6	6.08×10^5	8.27×10^6	3.51×10^5
Female 3	1.93×10^8	8.31×10^4	7.65×10^6	5.66×10^5	7.63×10^6	2.92×10^5

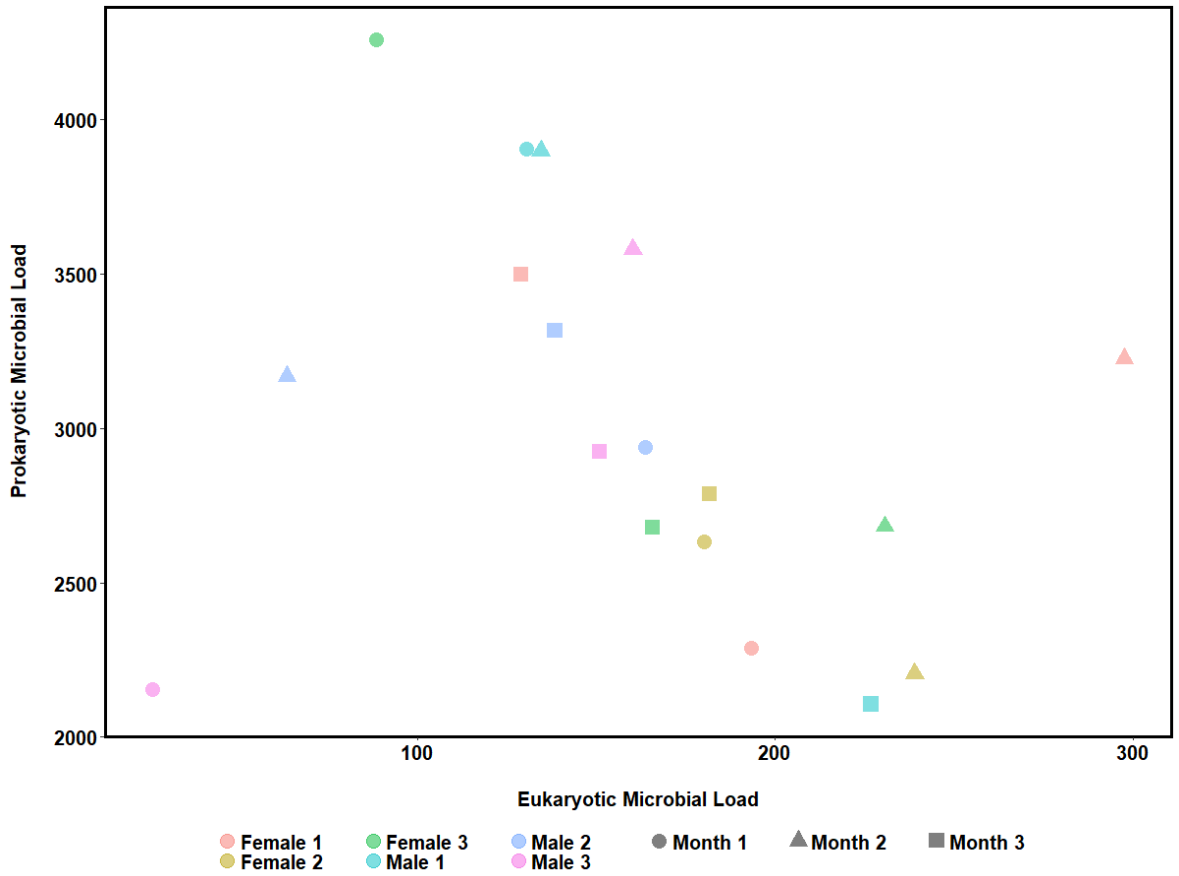


Figure 3.3. The correlation between Eukaryotic and Prokaryotic microbial load, counted using FACS. Six mice were sampled monthly. Data are square-root transformed.

Table 3.3. Two-way ANOVAs testing the effect of mouse ID and month of sample collection on microbial load. A two-way ANOVA was conducted using either Eukaryotic or Prokaryotic microbial load as the dependent variable and mouse ID and month collected as the independent variable. DF is the degrees of freedom.

	Mouse ID			Month		
	F Statistic	**DF**	***p value***	**F Statistic**	**DF**	***p value***
Eukaryotic	1.14	5	0.399	1.26	2	0.324
Prokaryotic	0.40	5	0.841	0.16	2	0.858

3.3.2 Separation of eukaryotic and prokaryotic cells

Whilst all four fractions types were prepared from each of the 18 faecal samples, some Eukaryotic and Prokaryotic fractions could not be sequenced due to insufficient DNA yield (Table 3.4). Only 7 samples were 18S sequenced for both Eukaryotic and Prokaryotic fractions whereas 11 samples were 16S sequenced for both Eukaryotic and Prokaryotic fractions.

Table 3.4. The number of faecal samples that were amplicon sequenced per fraction. Both the 16S and 18S rRNA genes were targeted for the FACS-sorted fractions but 16S rRNA gene sequencing was not attempted for Direct Faecal and Pre-Sort fractions.

Fraction Type	18S	16S
Direct Faecal	18	NA
Pre-Sort	18	NA
Eukaryotic	14	15
Prokaryotic	9	13

Figure 3.4 shows the number of 18S ASVs found in each of the four fractions. If the FACS-sorting of eukaryotic cells worked, it would be expected that i) the same number of 18S ASVs would be present in the Direct Faecal, Pre-Sort and Eukaryotic fractions and ii) the Prokaryotic fraction would have fewer 18S ASVs compared with the Eukaryotic fraction. An ANOVA identified a significant difference in the number of 18S ASVs among the four fractions ($F_{3,55}$ = 3.42, p = 0.023). Post hoc analysis showed a significant difference in the number of unique 18S ASVs observed between the Eukaryotic and Pre-Sort fractions (Table 3.5). Thus, there appears to have been a loss of 18S ASVs from the Pre-Sort fractions (average number of ASVs: 191 SE = 25) when generating the Eukaryotic fractions (103 SE = 6), contrary to the first expectation. However, there was no significant difference in the average number of ASVs between the Direct Faecal (155 SE = 20) and Eukaryotic fractions (Table 3.5). Post hoc comparison of the number of 18S ASVs observed between the Eukaryotic (103 SE = 6) and Prokaryotic (100 SE = 6) fractions identified no significant difference, contrary to the second expectation (Table 3.5). This result indicates that some eukaryotic cells were not stained by ERT and were thus separated into the Prokaryotic fraction.

Table 3.5. Tukey post hoc analysis of an ANOVA of the number of 18S ASVs in different FACS fractions. Significant *p* values are highlighted in bold.

Fraction Comparison	*p* value
Direct Faecal – Pre-Sort	0.707
Eukaryotic – Direct Faecal	0.320
Eukaryotic – Pre-Sort	**0.041**
Prokaryotic – Direct Faecal	0.400
Prokaryotic – Pre-Sort	0.079
Eukaryotic – Prokaryotic	> 0.999

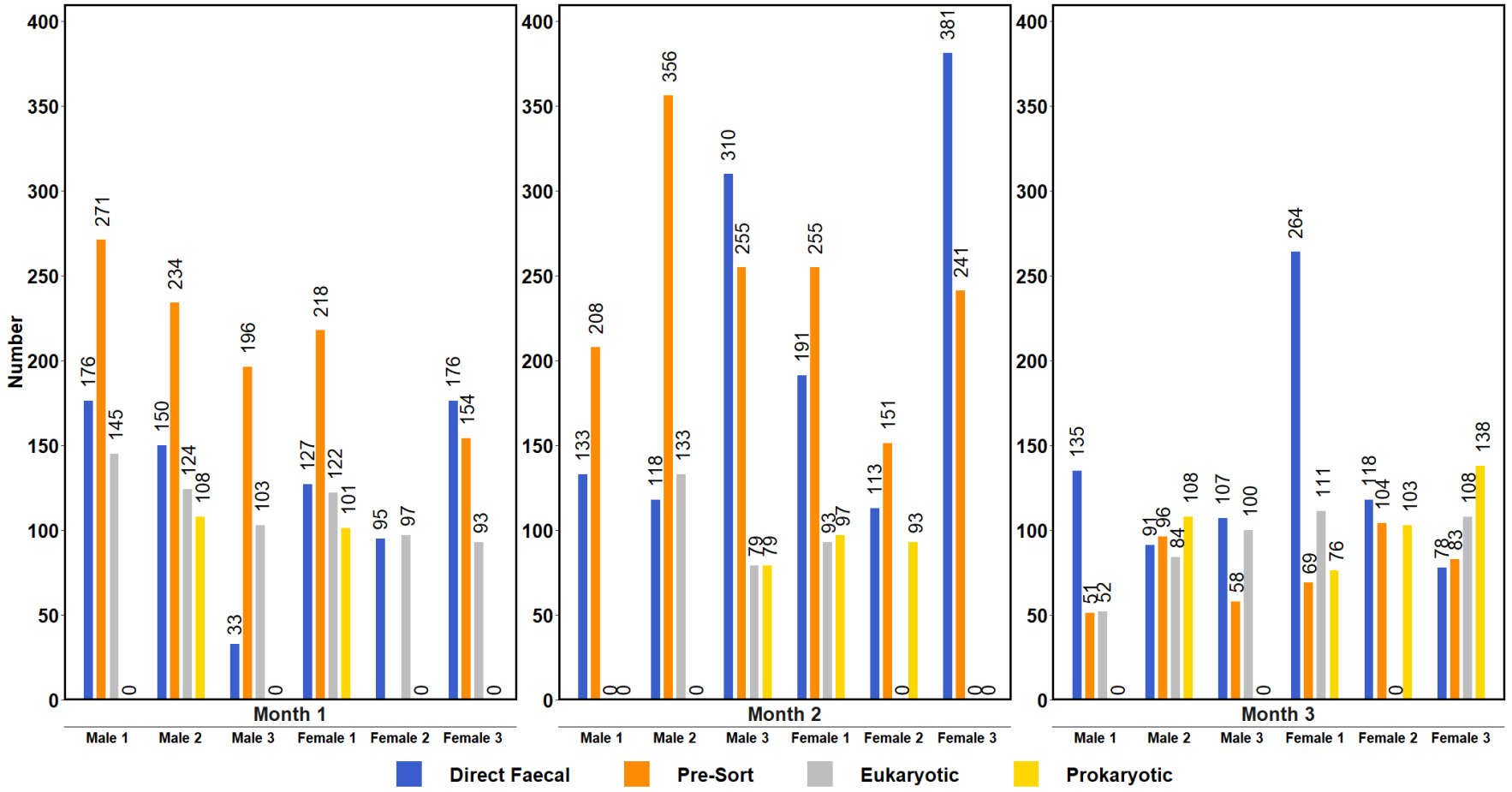


Figure 3.4. The number of unique 18S ASVs in each of the four fractions. The sources of the 18S ASVs are: Direct Faecal, from DNA extracted directly from faecal material; Pre-Sort, from DNA extracted from an aqueous suspension of the faecal cells prior to FACS; Eukaryotic and Prokaryotic, from DNA extracted from FACS-sorted fractions. Samples not sequenced due to insufficient DNA are represented by 0.

Figure 3.5 shows the number of 16S ASVs identified in the Eukaryotic and Prokaryotic fractions. It is expected that the Prokaryotic fraction will have more 16S ASVs compared to the Eukaryotic fraction. Overall, this expectation is supported: The Eukaryotic fractions had a mean number of 444 (SE = 17.5) 16S ASVs whereas the Prokaryotic fractions had a mean number of 510 (SE 12.7) 16S ASVs, and this difference was significant ($t_{31.7}$ = -3.07, p = 0.005). This result suggests that the Prokaryotic fraction is selecting for more prokaryotic species compared to the Eukaryotic fraction.

The Shannon's index measure of alpha diversity was calculated for the seven samples that had all four fractions sequenced (Figure 3.6). It was expected that across the Direct Faecal, Pre-Sort and Eukaryotic fractions, the alpha diversity would not change because the same 18S ASVs should be present and sequenced in each fraction. Additionally, the Prokaryotic fraction was expected to have lower 18S alpha diversity than the Eukaryotic fraction. However, an ANOVA showed that the two expectations were not supported. There was a significant difference in alpha diversity among the fraction types ($F_{3,24}$ = 10.3, p < 0.001) and post hoc analysis showed that the alpha diversity of the Direct Faecal fractions (mean Shannon index: 1.09 SE = 0.18) was significantly lower compared to the Pre-Sort (2.46 SE = 0.38) and Eukaryotic fractions (2.67 SE = 0.06) (Table 3.6). Furthermore, there was no significant difference between the alpha diversity of the Eukaryotic and Prokaryotic fractions (2.95 SE = 0.07), contrary to the expectation (Table 3.6).

Table 3.6. Tukey post hoc analysis of an ANOVA of alpha diversity of different FACS fractions. Significant *p* values in bold.

Fraction Comparison	***p value***
Direct Faecal – Pre-Sort	**0.003**
Eukaryotic – Direct Faecal	**0.002**
Eukaryotic – Pre-Sort	> 0.999
Prokaryotic – Direct Faecal	**< 0.001**
Prokaryotic – Pre-Sort	0.685
Eukaryotic – Prokaryotic	0.715

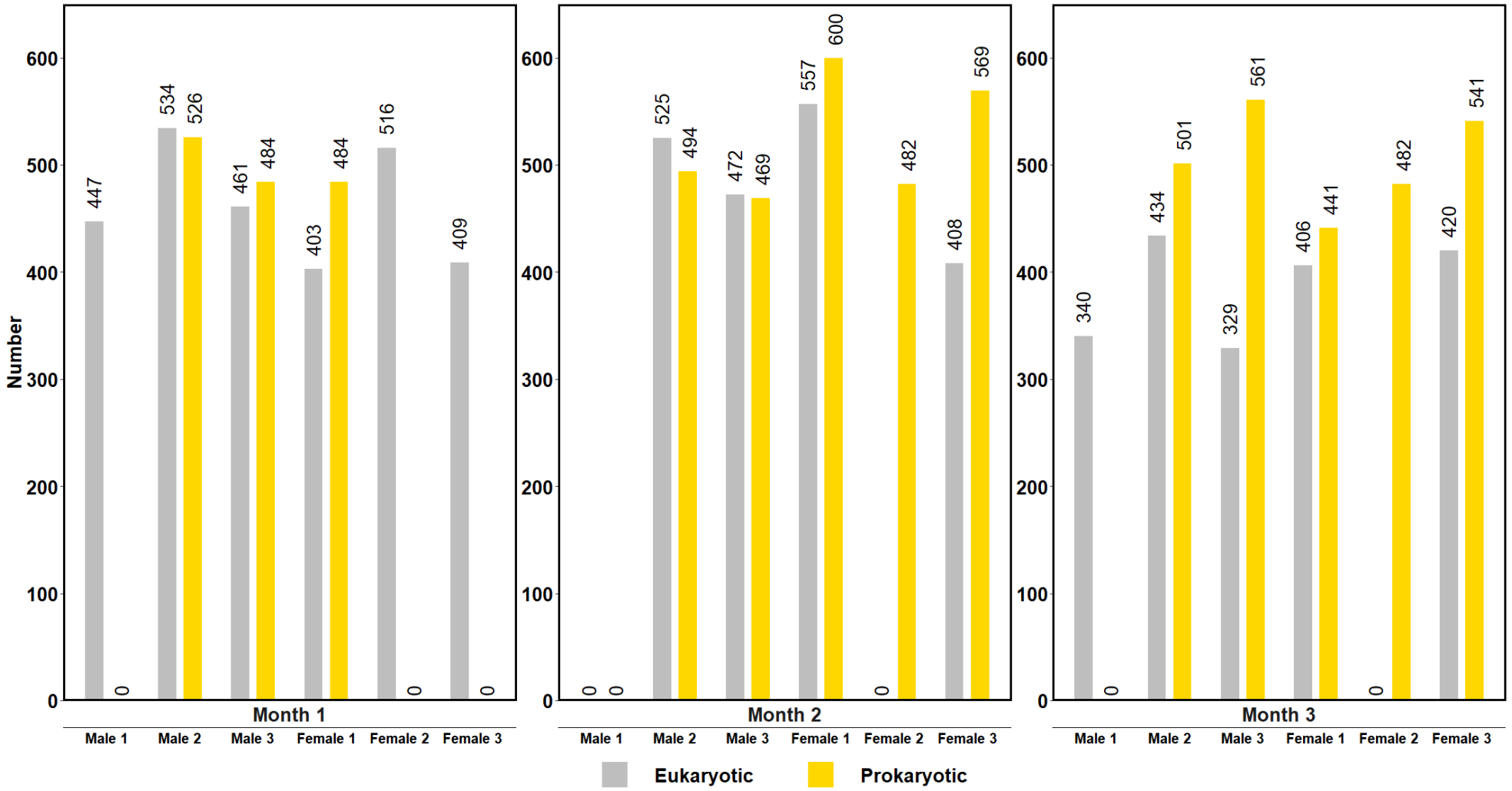


Figure 3.5. The number of unique 16S ASVs in the Eukaryotic and Prokaryotic fractions. The source of the 16S ASVs is from DNA extracted from FACS-sorted fractions. Samples not sequenced due to insufficient DNA are represented by 0.

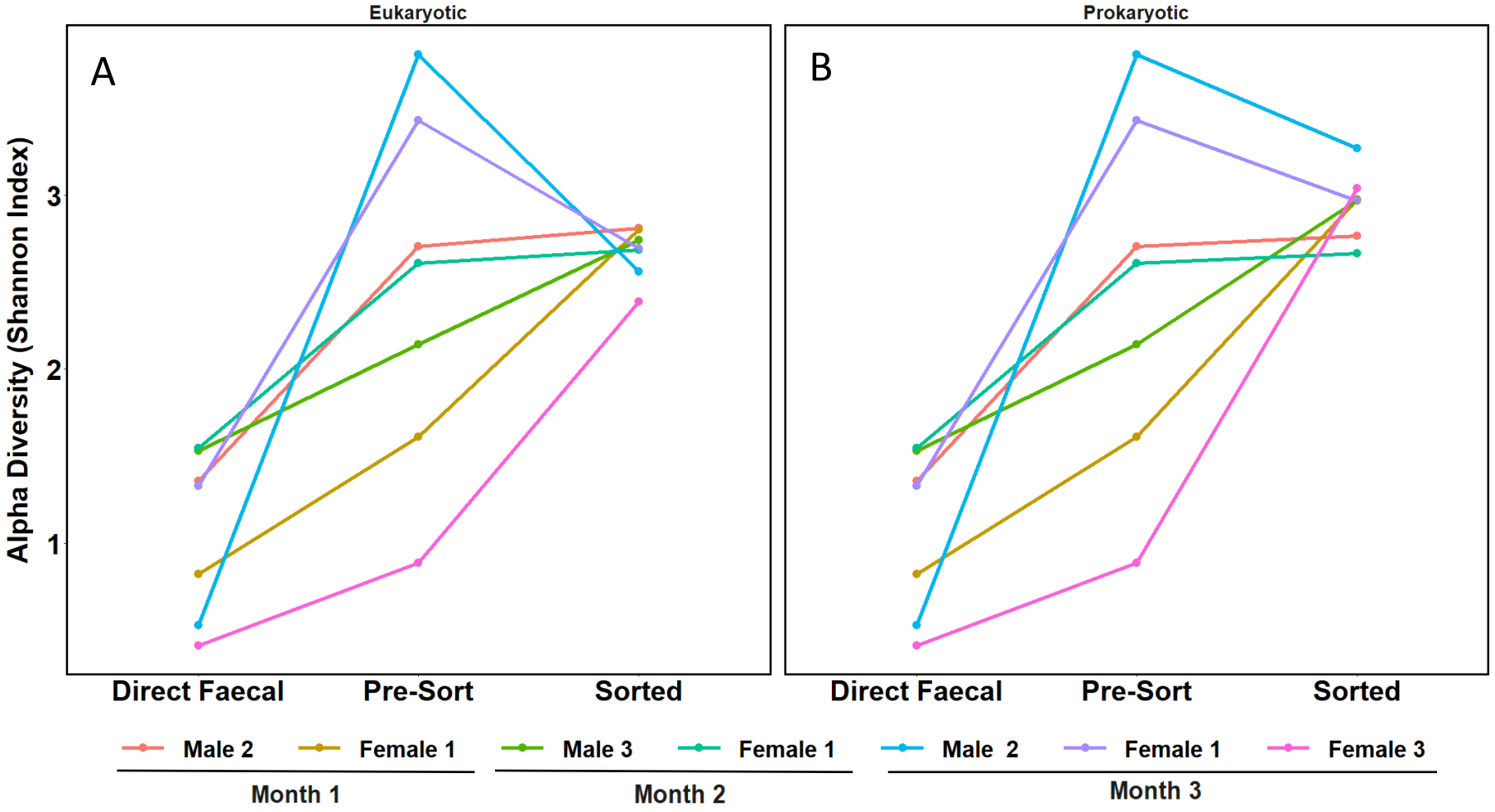


Figure 3.6. Shannon's index measure of alpha diversity for seven samples at each fraction of the FACS process. The sorted Eukaryotic (A) and Prokaryotic (B) fractions are depicted in separate panels because they were both generated from the Pre-Sort; note that the Direct Faecal and Pre-Sort are the same in both panels.

3.3.3 Eukaryome composition changes

The amount of inter-sample variation was expected to remain constant across all fractions because each faecal sample has been processed in the same way. To understand inter-sample variation, the number of 18S ASVs shared among the samples was examined. Across the majority of all fractions and samples, the 10 most abundant ASVs accounted for ≥ 70% of reads, and so only the 10 most abundant 18S ASVs were considered further. First, considering the Direct Faecal fraction: 7 (SE = 0.18) of the 10 18S ASVs present in any one sample only ever occur in at most two other samples. Only 2 (SE = 0) of the 10 18S ASVs can be found across all seven samples. This shows that inter-sample variation is high in the Direct Faecal fraction. Second, considering the Pre-Sort fraction: 5 (SE = 0.75) of the 10 18S ASVs only ever occur in at most two other samples. 2 (SE = 0) of the 10 18S ASVs can be found across all seven samples. This indicates the Pre-Sort samples have a slightly lower inter-sample variation than in the Direct Faecal fraction. Three, considering the Eukaryotic fraction: 2 (SE = 0.49) 18S ASVs only ever occur in at most two other samples and 4 (SE = 0) 18S ASVs are found across all samples. Similarly, the Prokaryotic fraction shows 3 (SE = 0.26) of the 10 18S ASVs only ever occur in at most two other samples. 4 (SE = 0) of the 10 18S ASVs can be found across all seven samples. Therefore, inter-sample variation is low in the Eukaryotic and Prokaryotic fractions. The comparatively low inter-sample variation in the Eukaryotic and Prokaryotic fractions, compared with the Direct Faecal and Pre-Sort, was contrary to the expectation. This shows that FACS causes the samples' Eukaryotic and Prokaryotic fractions to become more similar to each other.

In contrast to considering the number of shared ASVs, betadisper identified that the samples at the Pre-Sort stage had the most inter-sample variation (0.46). Inter-sample variation for the Direct Faecal, Eukaryotic and Prokaryotic fractions was 0.28, 0.26, and 0.28, respectively. A permutest found that this difference in inter-sample variation among the fractions was significant (F_3 = 3.26, p = 0.023). However, pairwise comparisons showed that the only significant difference in inter-sample variation was between samples at the Pre-Sort stage and the Eukaryotic fractions (Table 3.7).

Bray-Curtis dissimilarity was used to visualise beta diversity among the fractions based on their 18S ASVs (Figure 3.7). It was expected that the Direct Faecal and Pre-Sort fractions would cluster together as these samples should contain the same taxa with the same relative abundances. However, these two fraction types cluster separately meaning that their preparation is changing their composition. Additionally, it was expected that the Eukaryotic and Prokaryotic fractions should cluster separately as the Prokaryotic fraction should contain fewer 18S ASVs. However, these two fractions are indistinguishable from each other, again suggesting that the FACS-sorting is not providing the separation that was expected. A

PERMANOVA showed that the clustering of fractions according to Bray-Curtis dissimilarity was significant: there was significant variation in the ASV composition among the different fractions (F_3 = 26.4, p < 0.001). Pairwise comparisons identified that all fractions were significantly different to each other, with the exception of Eukaryotic *vs.* Prokaryotic (Table 3.7), contrary to the prediction.

Table 3.7. *p* values from pairwise comparisons among fraction types for inter-sample variation *via* a permutest and for ASV composition *via* a PERMANOVA. Significant *p* values are in bold.

Pairwise comparisons	permutest	PERMANOVA
Direct Faecal – Pre-Sort	**0.050**	**0.001**
Eukaryotic – Direct Faecal	0.724	**0.001**
Eukaryotic – Pre-Sort	**0.012**	**0.001**
Prokaryotic – Direct Faecal	0.963	**0.001**
Prokaryotic – Pre-Sort	0.084	**0.001**
Eukaryotic – Prokaryotic	0.494	0.300

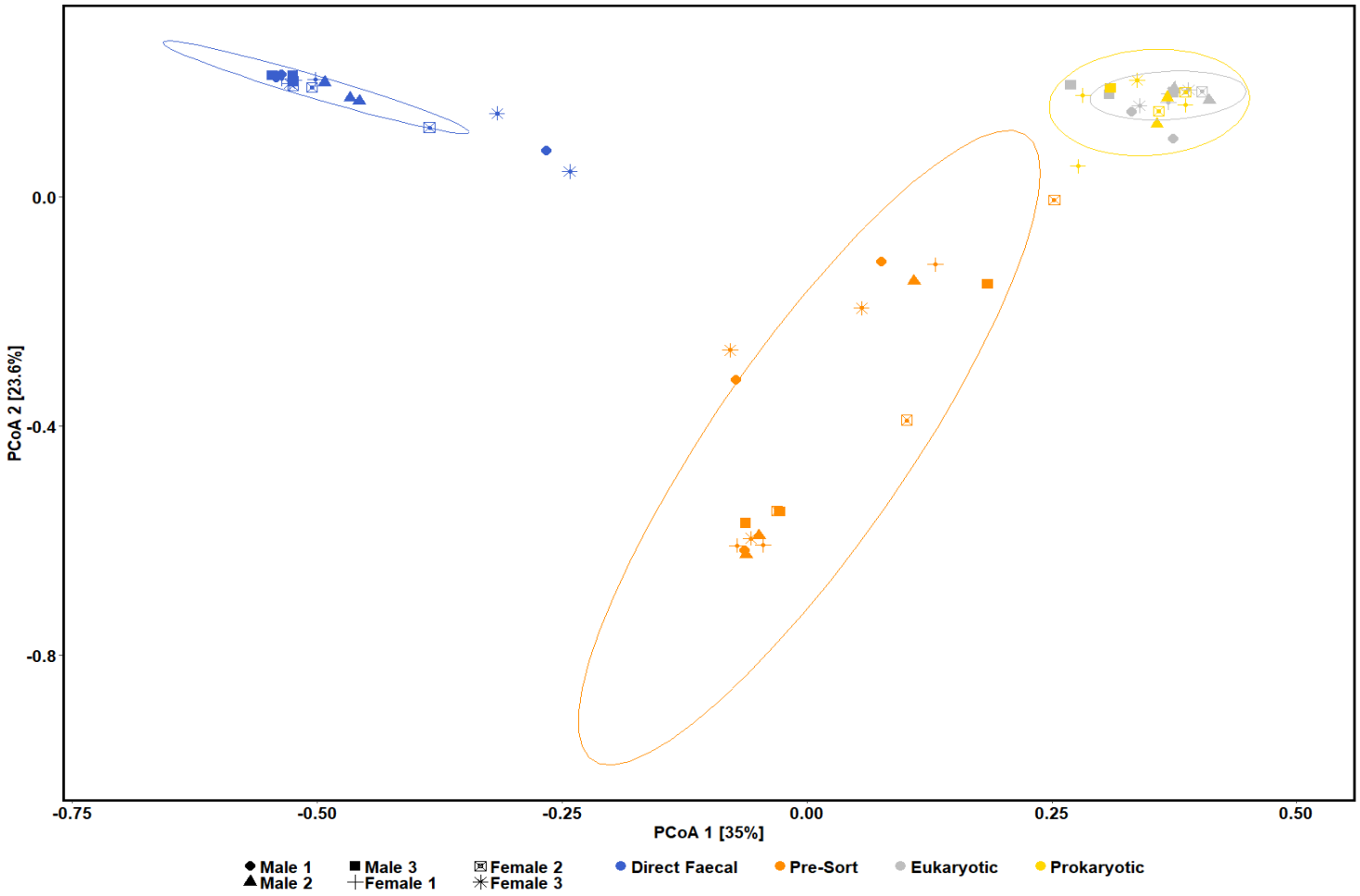


Figure 3.7. Bray-Curtis dissimilarity PCoA plot of the FACS-processed samples. The four fractions are differentiated by colour, with ellipses representing 95% confidence intervals**.** Mice are differentiated by symbols. 18 samples are shown for the Direct Faecal and Pre-Sort fractions, 14 for the Eukaryotic and 9 for the Prokaryotic fractions. PCoA axes 1 and 2 represent a total of 58.6% of the variation in the samples.

3.3.4 Changes in proportional abundance of 18S ASVs

Because the relative abundance of 18S ASVs can change during FACS, the proportionate change of 18S ASVs was examined for each fraction for the seven samples. As above, only the 10 most abundant 18S ASVs were used for this analysis. All Direct Faecal samples contain one high proportionate abundance 18S ASV that is then reduced to a very low proportionate abundance in the FACS-sorted Eukaryotic and Prokaryotic fractions or lost completely). On average, seven 18S ASVs present in the Direct Faecal fractions are entirely lost in both Eukaryotic (SE = 0.2) and Prokaryotic (SE = 0.4) fractions. Additionally, few low proportionate abundance Direct Faecal 18S ASVs become enriched in the sorted fractions.

All seven samples show evidence of low proportionate abundance Pre-Sort 18S ASVs being enriched in the Eukaryotic (4 SE = 0.7) and Prokaryotic fractions (4 SE = 0.7). Furthermore, in comparison to the Direct Faecal analysis, fewer 18S ASVs are lost between the Pre-Sort fraction and the Eukaryotic (5 SE = 0.8) and Prokaryotic (4 SE = 0.8) fractions. The high proportionally abundant 18S ASVs in the Pre-Sort fractions are not as substantially reduced in the Eukaryotic and Prokaryotic, compared to the Direct Faecal to Eukaryotic and Prokaryotic transition, due to the lower initial proportionate abundance. Additionally, more 18S ASVs are enriched when generating the sorted fractions from the Pre-Sort, and fewer 18S ASVs are eliminated, compared to the Direct Faecal to Eukaryotic and Prokaryotic transition.

Comparison of the Eukaryotic and Prokaryotic fractions show that they are similar. Across the seven samples, the difference in the proportionate abundance of the ten 18S ASVs is small and few 18S ASVs are not in both fractions (1 SE = 0.3 and 0.8). Enrichment (or lack of) of 18S ASVs cannot be compared between Eukaryotic and Prokaryotic fractions as one fraction is not produced from the other. The most abundant 18S ASVs in the Eukaryotic and Prokaryotic fractions have a lower proportionate abundance than the most abundant 18S ASVs in the Pre-Sort and Direct Faecal stages

3.3.5 Taxonomic comparison

The 18S ASVs sequenced at each stage in the FACS process were classified into four distinct groups as described in section 3.2.3.3: Food, Fungi, Host, and Other. Each categories' proportionate abundance in each fraction was examined to see if a particular category was enriched during the FACS process (Figure 3.8). All seven samples analysed follow a similar trend. In the Direct Faecal fraction, the majority of 18S ASVs are classified as Host (84.7% SE = 3.75). This category is much smaller in the subsequent Pre-Sort (3.7% SE = 1.07) and FACS-sorted fractions (10.4% SE = 2.02 and 8.8% SE = 1.01, respectively). In the Pre-Sort, the loss of Host 18S ASVs is replaced by an increase of 18S ASVs classified as Other (69.1% SE = 11.5). From the Pre-Sort to the FACS-sorted fraction, there is a loss of the Other

category. In the Eukaryotic fraction, 10.4% (SE = 2.02) of reads were classified into the Other category, closer to that seen in the Direct Faecal (2.0% SE = 0.72). In comparison, the Prokaryotic fraction has less Other 18S ASVs (8.3% SE = 1.44), compared to the Eukaryotic fraction. Food 18S ASVs are consistently low across all fractions. Fungi assigned 18S ASVs occur at low levels in the Direct Faecal fraction (3.8% SE = 1.13), but are proportionally more abundant in the Eukaryotic and Prokaryotic fractions where they are the largest single group of 18S ASV types (79.3% SE 2.59 and 74.0% SE 3.57, respectively). The increase in Fungi replaces the high levels of Host and Other 18S ASVs seen in the Direct Faecal and the Pre-Sort, respectively.

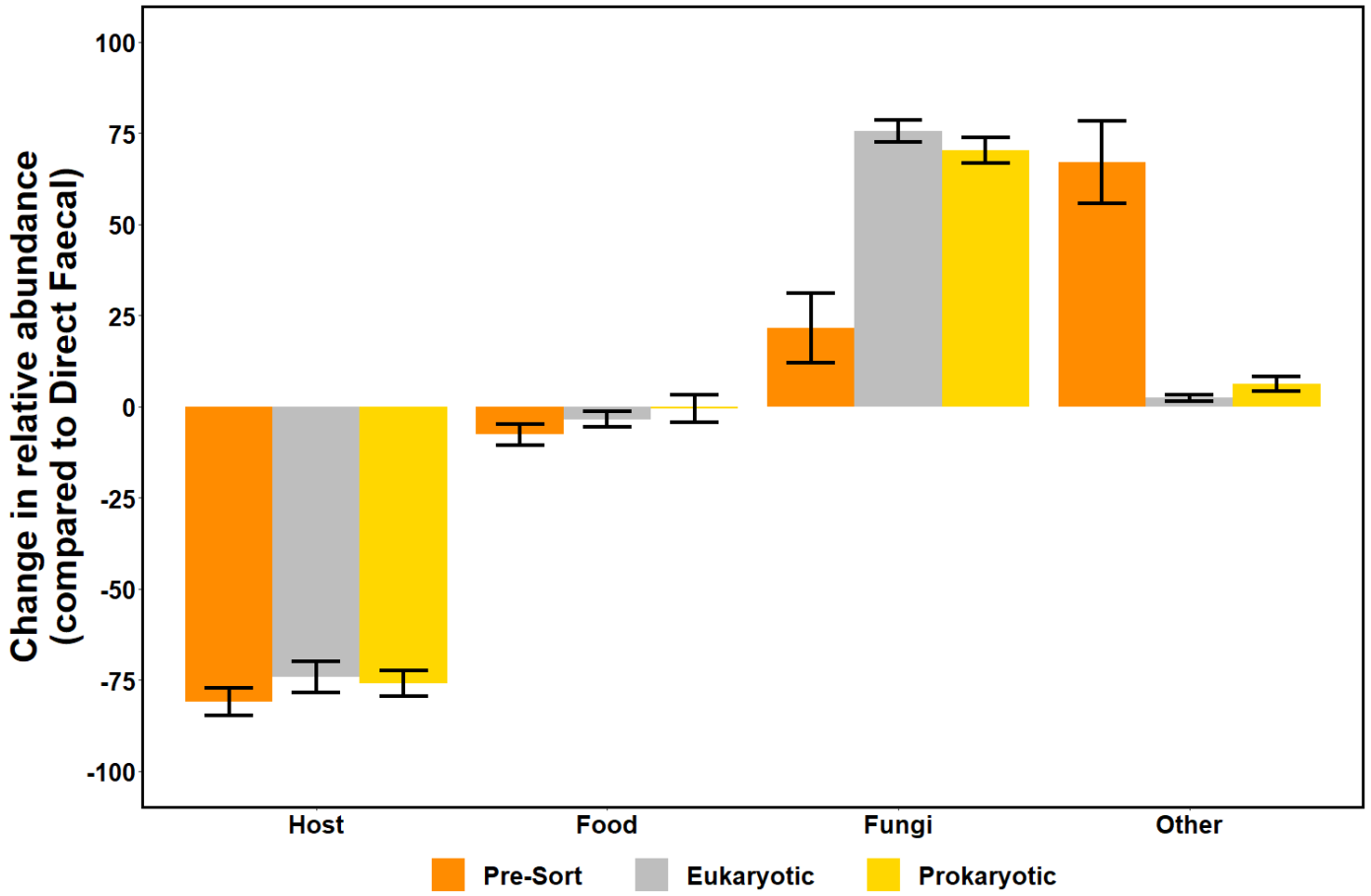


Figure 3.8. The change in the relative abundance of 18S ASVs based on their assignment to one of four categories – Host, Food, Fungi, Other – at each stage of the FACS process. Data are averaged across the seven samples with all four fractions sequenced. Comparisons are between the Direct Faecal fraction and each of the three subsequent fractions. Error bars show the standard error of the mean.

3.4. Discussion

3.4.1 Separation of eukaryotic and prokaryotic taxa

The FACS-based sorting methodology tested here aimed to i) separate putative eukaryotic and prokaryotic taxa of the faecal microbiome by the staining of a eukaryotic-specific organelle, and ii) quantify the microbial load of these taxa. Amplicon sequencing of these FACS-separated fractions showed – based on the similar number of 18S ASVs sequenced, similar alpha diversities, and close Bray-Curtis dissimilarity clustering of the fractions – that there was no clear separation of eukaryotic and prokaryotic taxa. Whilst other studies have successfully used FC to isolate specific eukaryotic taxa from mixtures containing bacteria (Porter *et al.*, 1997; Lepere *et al.*, 2011; Escalante *et al.*, 2016), this is the first attempt, of which I am aware of, to identify and separate all eukaryotic members of a microbial community.

Targeting the entire eukaryome presents a greater challenge compared to focussing on a specific taxon. In particular, differences in cell permeability and transport mechanisms among diverse eukaryotic taxa might have led to variation in the success of i) nucleic acid staining by SG and ii) endoplasmic reticulum staining by ERT (Jackson *et al.*, 2021). The ability of SG to stain phylogenetically distant bacterial and archaeal taxa has been confirmed (Martens-Habbena and Sass, 2006), but comparative studies of SG staining ability for eukaryotic taxa have not been done, of which I am aware. However, SG can successfully specifically identify yeast, protists, and plants, suggesting that a wide range of eukaryotic taxa can be successfully selected by SG-staining (Fortuna *et al.*, 2000; Christaki *et al.*, 2011; Clarindo and Carvalho, 2011). The ability of the ERT stain to identify phylogenetically distant eukaryotic taxa is not known, but endoplasmic reticulum staining has been successfully demonstrated in both protozoa and fungal pathogens of humans (Teixeira and Huston, 2008; Benhamou *et al.*, 2018). The aforementioned two studies, and the presence of 18S ASVs in the FACS-sorted fractions in the study presented here, suggest that eukaryotic gut taxa can be stained by both SG and ERT. Collectively, the results presented here indicate that the staining combination and flow cytometry gating parameters used in this study could not accurately select solely eukaryotic taxa.

3.4.2 Microbial load of eukaryotic taxa

The total microbial load of the faecal microbiome was determined by counting SG-stained cells, as used previously (Vandeputte *et al.*, 2017; Jackson *et al.*, 2021). A tenfold variation of the total microbial load was identified among samples from different mice, which is consistent with previous reports from the human faecal microbiome (Vandeputte *et al.*, 2017; Galazzo *et al.*, 2020). The analysis presented here showed that the number of putative eukaryotic cells was significantly lower than number of putative prokaryotic cells. This result is also consistent with other studies showing bacterial cells in the faecal microbiome are more abundant than

eukaryotic cells (Parfrey *et al.*, 2014; Underhill and Iliev, 2014; Hooks and O'Malley, 2020). No significant correlation was observed between the putative eukaryotic and prokaryotic microbial loads in the study presented here, in contrast to the literature which identifies inter-kingdom interactions between prokaryotic and eukaryotic taxa (Rao *et al.*, 2021; Harrison *et al.*, 2021). However, caution must be used when interpreting these microbial load data: the inaccurate separation of eukaryotic and prokaryotic taxa consequently means the quantification of the eukaryome and bacteriome is also not accurate.

3.4.3 Community composition analyses of FACS-processed samples

Analysis of the 18S sequence-based taxon identification at different stages of the FACS process showed that i) the community composition of the two final FACS-sorted fractions was different to both the Direct-Faecal and Pre-Sort fractions, and ii) the community composition of the two FACS-sorted fractions was similar. These observations mean that FACS consistently selected the same subset of taxa into these two supposedly different fractions, with the two FACS-sorted fractions being more similar to each other, while more dissimilar to the Direct Faecal and Pre-Sort samples. This means FACS was artificially altering the faecal microbiome community composition, in contrast to findings in other faecal microbiome FACS studies (Ben-Amor *et al.*, 2005; Peris-Bondia *et al.*, 2011; Maurice *et al.*, 2013; Palm *et al.*, 2014). For example, Palm *et al.* (2014) found the Pre-Sort fraction had a similar community composition to the FACS-sorted IgA- fraction, whereas the community composition of the FACS-sorted IgA+ fraction was a distinct subcommunity of the Pre-Sort fraction. Thus, the analyses in this present study indicate FACS-sorted fractions may not be representative of the true faecal eukaryome and bacteriome.

Analysis of the taxonomy of the 18S ASVs indicated that the majority of highly abundant ASVs lost during the FACS process belonged to vertebrate taxa, likely from host DNA. The change in relative abundance of these host ASVs may potentially be driven by host extracellular DNA present in the Direct Faecal fraction, which would not be present in the FACS-sorted fractions (Galazzo *et al.*, 2020). The decreased proportion of host DNA in the FACS-sorted fractions, compared to the Direct Faecal fraction, would allow less abundant eukaryotic taxa be detected (Pereira-Marques *et al.*, 2019). This could explain the proportionally greater abundance of fungi and other non-host ASVs in the FACS-sorted fractions, compared to the Direct Faecal, but further work would be needed to confirm this. These results show that FACS is a potential method to remove host-derived 18S ASVs and so increase the ability to detect less abundant eukaryotes. This would provide an alternative, albeit more expensive, method to host-blocking primers, which have their own limitations (see section 31.4, Thompson *et al.*, 2017; del Campo *et al.*, 2020).

3.4.4 Considerations and future perspectives

The method described and tested here is the first attempt, of which I am aware, to use FC to separate, quantify, and sequence the eukaryotic taxa of the faecal microbiome. However, it is evident that the method needs further optimisation by i) determining the optimal FACS gating parameters for sorting diverse eukaryote cells and ii) determining the specificity of the ERT stain (Jackson *et al.*, 2021). The method tested here utilised a broad approach to capture as much heterogeneity in gut eukaryotic taxa as possible, as done with bacterial taxa (Jackson *et al.*, 2021). However, a more fine-scale, multi-step FACS approach may be more suitable. Specifically, each sorting step would target and isolate one eukaryotic taxon group, and the remainder of cells would then be sorted again to target a different taxon group. This approach would alleviate the need to find a stain suitable that works for all gut eukaryotic taxa (see section 3.4.1 above) but would potentially be costly and low-throughput. Further to this, it is important to note that FACS can only process single cells, and so metazoan parasites in the eukaryome, such as helminths, would not be identified (Müller and Nebe-von-Caron, 2010).

The methodology presented here may provide a potential new method of sample preparation, which could be utilised to remove host extracellular DNA from faecal samples. In doing so, less abundant taxa may be more easily sequenced and identified when characterising the gut eukaryome. Thus, future work would benefit from exploring the use of flow cytometry to remove host DNA from faecal samples. Finally, whilst the study shows the proposed method was not successful at separating and quantifying the eukaryome, it has established the groundwork for future studies to build on when attempting to quantify the microbial load of eukaryotic gut taxa.

Chapter 4: The eukaryome of wild house mice and their disease state

Abstract

The gut microbiome can profoundly affect an animal's biology. However, we still don't fully understand the contribution of eukaryotic taxa – the eukaryome – to the wider microbiome, nor what affects the composition of the eukaryome. This contrasts to our extensive knowledge of the bacterial microbiome (bacteriome), particularly in wild animals. Here, I aimed to characterise the eukaryome of wild house mice (*Mus musculus domesticus*) using 18S rRNA sequencing. I combine this with parasitological and histopathological analyses, to determine if disease or immune state contributes to eukaryome composition. Additionally, I used 16S rRNA sequencing data from the same mice to identify if variation in the eukaryome was correlated with variation in the bacteriome. I found that mice from different locations differ in their eukaryome composition, and that these differences are potentially driven by the presence of parasitic nematodes and coccidia (*Eimeria* spp). I also found that both gut inflammation and faecal IgA concentration were associated with variation in gut eukaryome composition. Furthermore, I found that the diversity metrics of the gut eukaryome and bacteriome were not correlated, suggesting different factors drive variation in the two microbiome types, and that inter-kingdom interactions are likely contributing to gut microbiome composition. These data are the first to characterise the gut eukaryome of wild house mice and provide a better understanding of how the eukaryome may be modulated by the host immune and disease state. Additionally, these results have highlighted the importance of considering both eukaryotic and bacterial taxa when researching drivers of gut microbiome composition.

4.1 Introduction

4.1.1 The mammalian eukaryome

The gut microbiome is the diverse community of microbial taxa found within the digestive tract of an individual, and includes bacteria, archaea, and eukaryotes (Eckburg *et al.*, 2005; Parfrey *et al.*, 2014; Wampach *et al.*, 2017). The interaction of these microbes with their host can affect the host's development and physiology (Sommer and Bäckhed, 2013). For example, microbial colonisation of the gut is key for: immune system and behavioural development (Heijtz *et al.*, 2011; Gensollen *et al.*, 2016), nutrition and metabolism (Rowland *et al.*, 2018; Williams *et al.*, 2020), and pathogen resistance (Chudnovskiy *et al.*, 2016; Brown and Clarke, 2017). Thus, to understand variation in the behaviour and physiology of hosts, it is important to understand how and why the gut microbiome composition differs among hosts (Suzuki, 2017; Henry *et al.*, 2021).

To date, the majority of gut microbiome research has focused on the bacterial taxa (bacteriome) and there has been comparatively less study of the eukaryome (fungi, protozoa, and helminths) in its entirety (del Campo *et al.*, 2020). The difference in research effort is largely due to i) eukaryotic gut taxa traditionally being considered parasitic, rather than commensal (Parfrey *et al.*, 2014; del Campo *et al.*, 2020) and ii) eukaryotic gut taxa being less abundant than bacterial taxa (Laforest-Lapointe and Arrieta, 2018). Further to this, characterising the gut eukaryome was largely limited by a lack of accurate methods capable of detecting all three groups of eukaryotes: fungi, protozoa, and helminths (Popovic *et al.*, 2018). Thus, studies often investigated, and still do, each of these three groups independently using different methods. For example, amplicon sequencing of the internal transcribed spacer (ITS) region of DNA is used to identify fungi (Sun *et al.*, 2018; Barelli *et al.*, 2020a; Rao *et al.*, 2021), whereas targeted PCRs and ELISAs are commonly used for protozoa (Blessmann *et al.*, 2002; den Hartog *et al.*, 2013). Traditional parasitological techniques, *e.g.* flotation and sedimentation, are also used for protozoa and helminths (Barelli *et al.*, 2020b). More recently, sequencing amplicons of the 18S rRNA gene to identify taxa within all three eukaryotic groups simultaneously (Amaral-Zettler *et al.*, 2009; Stoeck *et al.*, 2010; Hadziavdic *et al.*, 2014; Popovic *et al.*, 2018) has become more common in gut eukaryome studies (Parfrey *et al.*, 2014; Heitlinger *et al.*, 2017; Li *et al.*, 2018; Mann *et al.*, 2020; Kim *et al.*, 2022).

18S rRNA amplicon sequencing has been used to describe the gut eukaryome for some mammals, including bats, ungulates, hyenas, non-human primates, and rodents, and these studies are summarised in Table 4.1. All of the summarised studies identified all three eukaryotic groups in their study system(s) but focus was often towards just one or two of the three groups. For example, fungi were the main taxa of interest in bats (Li *et al.*, 2018), but were only briefly mentioned for non-human primates (Mann *et al.*, 2020). Whilst summarised

Table 4.1. A summary of gut eukaryome studies in mammals. The eukaryotic taxa identified from these studies are divided into three groups (fungi, helminths and protozoa), with P, C, O, and F representing eukaryotic phyla, classes, orders, and families respectively. Genera are shown in italics. Protozoa are grouped as seen in Chapter 2 (section 2.1.2). A summary of the gut eukaryome diversity and composition is also given. The table is likely not a complete list of studies describing the mammalian eukaryome, and aims only to show the common eukaryotic taxa identified.

Animal	Fungi	Helminths	Protozoa	Diversity and Composition
Bats Li *et al.*, 2018	Ascomycota (P) • *Candida* • *Hanseniaspora* • *Geotrichium* • *Cladosporidium* • *Blasobotrys* • *Kodamea* • *Clavispora* • *Debaryomcyes* Basidiomycota (P) • *Malassezia* Glomeromycota (P) Entomophthoromycota (P)[b]	Nematoda (P) Platyhelminthes (P)	Apicomplexa	• Mycobiome compositional differences were greater between dietary groups compared to differences within dietary groups. • The most dominant fungal taxa were yeasts from Ascomycota and Basidiomycota.
Hyenas[a] Heitlinger *et al.*, 2017	Ascomycota (P) Basidiomycota (P) Zygomycota (P)[c] Microsporidia[d]	Nematoda (P) • *Ancylostoma* • *Ostertagia* • *Haemonchus* • *Trichuris* Platyhelminthes (P) • *Diphyllobothrium* • *Spirometra* • *Dipylidium* • Taeniidae (F)	Apicomplexa • *Eimeria* • *Toxoplasma* • *Besnoitia* • *Isospora* Cercozoa Ciliophora Metamonada Euglenozoa	• A greater diversity in the gut eukaryome was associated with higher ranking individuals. • Ascomycota and Basidiomycota were the most diverse and abundant fungal taxa.

Non-human primates Mann *et al.*, 2020	Ascomycota (P)	Nematoda (P) • *Trypanoxyuris* • *Trichuris* • Rhabditida (O) • Spirurida (O)	Amoebozoa • *Entamoeba* • *Iodamoeba* • *Endolimax* Apicomplexa Stramenopiles • *Blastocystis* Cercozoa • *Cercomonas*	• The eukaryome community composition was different among different primate species, with high variability within species. • There was no dominant fungal species. • The gut eukaryome tended to be dominated with either Ascomycota or Amoebozoa.
Rodents Kim *et al.*, 2022	Ascomycota (P) • *Kazachstania* • *Candida* • *Periconia* • *Cladosporium* Mucoromycota (P) • *Mucor* • *Rhizopus*	Nematoda (P) • *Heligmosomoides* • *Syphacia* • *Strongyloides* • *Oscheius* Platyhelminth (P) • *Hymenolepis* • *Raillietina* • *Plagiorchis* • *Panagrolaimus*	Metamonada • *Tritrichomonas* • *Monocercomonas* • *Giardia* • *Spironucleus* • *Retortomonas* Apicomplexa • *Isospora* • *Cryptosporidium* Stramenopiles • *Blastocystis* Amoebozoa • *Entamoeba*	• Protozoa relative abundance was greater than fungal relative abundance. • *Kazachstania* was the most prevalent fungal species.
Ungulates Parfrey *et al.*, 2014	Neocallimastigomycota (P) • *Neocallimastix*	Nematoda (P)	Stramenopiles • *Blastocystis* Amoebozoa Metamonada Ciliophora • *Entamoeba*	• There was a low diversity of species within eukaryotic phyla. • The presence of taxa was highly varied across host species. • Differences in gut eukaryome composition were most notable between hindgut and foregut fermenters.

a Only the taxa identified as “eukaryome” by the authors are shown
b Since been reclassified into Zoopagomycota
c Since been divided into two phyla, Mucoromycota and Zoopagomycota
d Taxonomy is disputed

in Table 4.1, the protozoa identified from these studies will not be reviewed in detail here (but see Chapter 2, section 2.1.2). Additionally, the presence of helminths in the mammalian gut has been extensively reviewed, given their typically parasitic lifestyle and ubiquity, and so will not be discussed here (Chowdhury and Aguirre, 2001; Brooker, 2010). The fungal taxa in the mammalian eukaryome (mycobiome) are most commonly represented by yeasts from the phyla Ascomycota and Basidiomycota (Table 4.1). In particular, Ascomycota is often reported as the most abundant and diverse fungal phylum in the gut (Li *et al.*, 2018; Barelli *et al.*, 2020a; Mann *et al.*, 2020). Ascomycota genera identified from the gut eukaryome include *Candida, Cladosporium, Geotrichium,* and *Debaryomyces* (Table 4.1). These taxa are also found in studies describing gut fungi without consideration of protozoa or helminths (Scupham *et al.*, 2006; Sun *et al.*, 2018; Barelli *et al.*, 2020a; Rao *et al.*, 2021). Other fungal phyla in the mammalian gut eukaryome include Neocallimastigomycota, found in the rumen of hind-gut fermenting herbivores (Parfrey *et al.*, 2014), Mucoromycota, and Zoopagomycota (Table 4.1).

Many of the fungal taxa summarised in Table 4.1 may not be true residents or colonisers of the gut, but rather are environmental contamination of faecal samples and/or fungi in the host's diet (Hallen-Adams and Suhr, 2017; Lavrinienko *et al.*, 2021a). Only a few fungal genera are considered to be true residents of the mammalian gut, of which the most commonly accepted is *Candida* (Underhill and Iliev, 2014; Hallen-Adams and Suhr, 2017). *Candida* spp. have been described from humans, bats, and mice (Dollive *et al.*, 2013; Li *et al.*, 2018; Rao *et al.*, 2021; Kim *et al.*, 2022). However, the presence of *Candida* spp. in non-human primates was attributed to dietary sources rather than being a gut resident (Mann *et al.*, 2020). Other fungal genera considered capable of colonising the mammalian gut include *Malassezia, Cladosporidium, Galactomyces,* and *Saccharomyces* (Hallen-Adams and Suhr, 2017).

Whilst the gut eukaryome species diversity is lower than that the gut bacteriome (Parfrey *et al.*, 2014; Nash *et al.*, 2017), the studies summarised in Table 4.1 have shown that the diversity and composition of the eukaryome can differ substantially among different host species (Parfrey *et al.*, 2014; Mann *et al.*, 2020). For example, cercopithecine monkeys have a greater eukaryome diversity compared to non-cercopithecine monkeys (Mann *et al.*, 2020) and highly abundant eukaryotic taxa found in one host species can be absent from other host species (Parfrey *et al.*, 2014). Studies that compare only the gut mycobiome have shown that primate species have different mycobiome compositions, and that mice have a higher diversity of fungal taxa compared to humans (Marchesi, 2010; Barelli *et al.*, 2020a). Furthermore, species richness in the gut mycobiome is driven by host phylogeny (Harrison *et al.*, 2021).

4.1.2 Factors affecting to eukaryome composition

Variation in gut microbiome composition among species, and among con-specific individuals, has led to investigation into which factors affect the composition of the microbiome. Factors affecting the bacteriome are well-characterised and can be split into three groups: host-mediated, environmental, or microbiome-mediated (Schmidt *et al.*, 2018). The following section will review how the factors within these groups may also affect mammalian gut eukaryome composition.

Host-mediated factors known to drive variation in gut bacteriome diversity and composition include: demographic traits (such as sex, age, sociality), immunity, and genetics. However, the contribution of these factors to eukaryome diversity and composition is undetermined. Studies investigating the effect of sex on the gut eukaryome are sparse, and show contradictory results: studies have shown the mycobiome community composition is significantly different between males and females in both humans and macaques (Strati *et al.*, 2016; Sun *et al.*, 2018), whereas a third study found no difference in mycobiome composition between male and female humans (Rao *et al.*, 2021). Similarly, comparison of the eukaryome among hosts of different ages have also found contrasting results. Studies in humans and hyenas have shown no clear pattern between an individual's age and the diversity and composition of their eukaryome (Heitlinger *et al.*, 2017; Wampach *et al.*, 2017; Rao *et al.*, 2021), whereas another study showed that the gut mycobiome of 18-month old infants is more diverse, with greater variation in community composition among individuals, compared to 6-month old infants (Turunen *et al.*, 2023). The social behaviour of a host can also contribute to eukaryome community composition (Sarkar *et al.*, 2020). In social species, it has been shown that more con-specific interactions can lead to a greater gut microbiome species richness, and decreased variance in gut community composition among individuals within a population (Moeller *et al.*, 2016b). This mechanism was suggested to potentially explain the higher diversity of gut eukaryotes seen in cercopithecine monkeys, compared to non-cercopithecine monkeys (Mann *et al.*, 2020). Furthermore, the social status of an animal in a population influences eukaryome diversity: higher ranking hyenas have a greater gut eukaryome species richness compared to lower ranking individuals (Heitlinger *et al.*, 2017). Thus, differences in these demographic traits, like population size and structure, drive differences in gut eukaryome diversity and composition between and within different host populations.

The host immune state is important for regulating the eukaryotic taxa in the gut (Clerc *et al.*, 2018; Sardinha-Silva *et al.*, 2022; Swidergall and LeibundGut-Landmann, 2022). For example, immune regulation of the host's intestinal mucus production can directly limit the colonisation of potentially pathogenic eukaryotic taxa (Hasnain *et al.,* 2013). Additionally, intestinal immunoglobulin production can also regulate which eukaryotic taxa can successfully colonise

the gut (Ost *et al.,* 2021). For instance, the colonisation of *Candida albicans* is regulated by the host's intestinal immunoglobulin A (IgA) responses, which suppresses pathogenic morphotypes of *C. albicans,* selecting for the growth of commensal morphotypes (Ost *et al.,* 2021). In wild animals, studies investigating how the immune state of mammals regulates gut eukaryotes have focused on parasitic and pathogenic taxa, rather than the general eukaryome (Clerc *et al.*, 2018; Ferreira *et al.*, 2021). For example, immunoglobulins are important for explaining parasitic community composition in wood mice: infection with *Heligmosomoides polygyrus* and *Eimeria* was affected by the titre of *H.polygyrus*-specific IgG and faecal IgA respectively (Clerc *et al.*, 2018). Additionally, variation in the parasite community of hyenas was explained by faecal IgA concentration (Ferreira *et al.*, 2021). To this end, differences in the immune state of hosts, and thus their regulation of eukaryotic taxa, can also contribute to variation in eukaryome composition.

Differences in the host genotype among individuals, including the major histocompatibility complex (MHC), can contribute to variation in the immune-regulation of eukaryotic gut taxa in wild animals (Radwan *et al.*, 2020; Worsley *et al.*, 2022). Specifically, there is an association between MHC-I diversity and mycobiome composition in the Seychelles warbler (Worsley *et al.*, 2022). This was driven by a greater abundance of *Lasiodiplodia* spp. in individuals with lower MHC-I diversity, and a greater abundance of *Sympodiomyces* spp. in individuals with higher MHC-I diversity (Worsley *et al.*, 2022). Whilst, no study, of which I am aware, has investigated the impact of MHC diversity on the gut eukaryome composition of mammals, it has been shown that host genotype is responsible for variation in the gut mycobiome community composition of laboratory mice (Gupta *et al.*, 2023). Other factors regulating a host's immune state include early life development, body condition, age, sex, and being in a physiologically-demanding state *e.g.* malnutrition (Viney *et al.*, 2005; MacGillivray and Kollmann, 2014; Abolins *et al.*, 2018; Kelly *et al.*, 2018; Clerc *et al.*, 2019a). Additionally, exposure to microbes early in the host's development is important for priming the immune system in order to recognise and regulate future commensal and pathogenic taxa (Gensollen *et al.,* 2016). The presence of parasites and pathogens can also alter the host's innate and adaptive immunity pathways, such as the up-regulation of pro-inflammatory pathways (Leung *et al.,* 2018a). Such inflammation can alter the host's ability to immunomodulate the colonization of other eukaryotes, as seen with chemically-induced inflammation (Jawhara *et al.,* 2008; Leung *et al.,* 2018a). To date, there has been little research in wild mammals linking such host factors to the immune phenotype of the host and the subsequent variation in gut eukaryome composition seen among hosts.

Environmental factors, *e.g.* seasonal variation and the biogeography of the host, are important variables that can contribute to mammalian gut bacteriome composition (Linnenbrink *et al.*,

2013; Goertz *et al.*, 2019; Marsh *et al.*, 2022). The few studies that have investigated the effect of environment on the eukaryome also support these findings. The sampling site of bank voles had a detectable effect of fungal community composition (Antwis *et al.*, 2021), and host habitat differences were associated with different fungal eukaryomes in monkeys (Barelli *et al.*, 2020a). Similarly, humans in either urban or rural locations had different gut mycobiomes (Kabwe *et al.*, 2020). However, environmentally-driven differences in gut microbiome composition can often be attributed to variation in the host diet, which is a major source of variation in gut bacteriome composition (Wang *et al.*, 2014). For the eukaryome, diet was hypothesised to be the major source of differences in gut eukaryome composition between herbivorous and insectivorous bats (Li *et al.*, 2018). This contrasts to non-human primates: differences in eukaryome community composition were greater among different host species compared to differences among different diets (Mann *et al.*, 2020). In the laboratory setting, evidence suggests the gut mycobiome is shaped by diet: the mycobiome composition of laboratory mice was different between mice fed high-fat or low-fat diets (Heisel *et al.*, 2017; Gupta *et al.*, 2023). Whilst there was no difference in the diversity of fungal taxa, mice fed high-fat diets had a greater relative abundance of Basidiomycota species, and lower relative abundance of Ascomycota species, compared to mice with low-fat diets (Heisel *et al.*, 2017; Gupta *et al.*, 2023).

The composition of the eukaryome can also be mediated by direct interactions with the pre-existing gut microbial community (Coyte *et al.*, 2021). For protozoa, colonisation success is often dependent on the presence of other protozoa and bacterial species (see Chapter 2, section 2.1.3). Likewise, there is emerging evidence that the colonisation success of fungi is also dependent on pre-established gut taxa. For example, antibiotic treatment increases colonisation of the gut by *Candida* spp., compared to non-antibiotic treatment, in mice and humans, which suggests the pre-existing gut microbiome normally inhibits *Candida* colonisation (Sam *et al.*, 2017). Further research has shown that the growth of *Candida* spp. in the gut is inhibited by the presence of the bacteria *Staphylococcus* (Rao *et al.*, 2021). In contrast, *Candida* spp. can increase their colonisation success by exploiting pre-established Enterobacteriaceae spp. (Rao *et al.*, 2021). Several studies have also identified co-occurrence between bacterial and fungal taxa in the mammalian gut (Heisel *et al.*, 2017; Barelli *et al.*, 2020a; Mann *et al.*, 2020 but see Harrison *et al.*, 2021). This may be indicative of a co-dependence among the taxa, or the simultaneous response to an independent mechanism (Mann *et al.*, 2020).

There are also associations between the bacteriome community composition and helminth colonisation in wild mammals (Kreisinger *et al.*, 2015; Montero *et al.*, 2021; Kim *et al.*, 2022). For example, the presence of three common helminths – *Hymenolepis* spp., *Syphacia* spp.,

and *H. polygyrus* – is associated with changes in the gut bacteriome composition of yellow-necked mice (Kreisinger *et al.*, 2015). Similarly, *Heligmosomoides* colonisation is associated with differences in gut bacteriome composition in striped field mice, with *Lactobacillus* bacteria being more abundant in *Heligmosomoides*-infected mice compared to *Heligmosomoides*-uninfected mice (Kim *et al.*, 2022). Furthermore, a study in wild lemurs found MHC-I diversity was associated with gut bacteriome composition, which in turn was associated with helminth colonisation (Montero *et al.*, 2021). In laboratory settings, there is evidence that helminth colonisation success is dependent on the pre-existing gut microbiome. For example, *Trichuris muris,* the mouse whipworm, has a decreased hatching success when gut bacterial load is lower (Hayes *et al.*, 2010) and *T. muris*-induced changes in gut microbiome composition limit further *T. muris* colonisation (White *et al.*, 2018). A further study has shown that the hatching success of *T. muris* is dependent on the bacterial species/strain present (Vejzagić *et al.*, 2015). Specifically, the hatching success of *T. muris* was dependent on the incubation of eggs with particular strains of *Escherichia coli,* and incubation with *Lactobacillus reuteri* induced *T. muris* hatching whereas *L. amylvorus* and *L. murinus* did not (Vejzagić *et al.*, 2015). Colonisation success of the rodent intestinal roundworm *H. polygyrus* can also been affected by the composition in the gut bacteriome: mice with *Lactobacillus* bacteria in their gut are more susceptible to *H. polygyrus* colonisation, compared to mice without *Lactobacillus* (Reynolds *et al.*, 2014). However, the driving mechanisms behind these associations are hard to infer in wild study systems, and may be the result of indirect interactions *via* the host immunity as described above, rather than direct microbial interactions.

4.1.3 The wild house mouse eukaryome composition

Laboratory mice are commonly used in gut microbiome studies, despite evidence that their bacteriomes differ from their wild counterparts (*Mus musculus domesticus*) (Rosshart *et al.*, 2019; Viney, 2019; Bowerman *et al.*, 2021). The gut mycobiome also differs between wild house mice and laboratory mice (Rosshart *et al.*, 2019), and the diversity of protozoa and helminths found in laboratory mice, whilst well-characterised, is not representative of the diversity found in wild house mice (Pritchett, 2007; Ehret *et al.*, 2017). Despite attention shifting to wild house mice for bacteriome studies (Linnenbrink *et al.*, 2013; Kreisinger *et al.*, 2014; Weldon *et al.*, 2015; Goertz *et al.*, 2019; Suzuki *et al.*, 2019a), there has been, at best, very limited study of their eukaryome, and none of what drives its composition. Currently, to my knowledge, there are only two studies describing the mycobiome of wild house mice, neither of which characterised the protozoa and helminths found in the gut eukaryome (Rosshart *et al.*, 2019; Bendová *et al.*, 2020).

In this respect, the study presented here is unique, given its three main aims to further our understanding of the gut eukaryome: i) to characterise the eukaryome composition of wild

house mice from three sampling sites in the UK, ii) identify which host factors may be associated with eukaryome composition in wild mice, and iii) investigate potential correlation between gut eukaryome and bacteriome diversity metrics, as well as identifying co-occurrence between gut eukaryotic and bacterial taxa. To achieve the first aim, I used 18S rRNA amplicon sequencing to identify the eukaryotic taxa in both the caecum and faeces of wild house mice, and compared the composition among three sampling sites. Here, I hypothesise that there are differences in the gut eukaryome diversity and composition among the sites. Then, I measured the host's disease and immune state to identify how this was associated with eukaryome diversity and composition. Specifically, I predict that changes to the host's mucosal immunity following parasitic infection, as measured by gut inflammation and the production of IgA and mucin, are associated with changes in the gut eukaryome composition. Finally, I used 16S rRNA amplicon sequencing data to characterise the bacteriome of the mice, before looking for correlation between the eukaryome and bacteriome and any co-occurring taxa. Here, I hypothesise that the gut eukaryome and bacteriome are impacted by the same host factors.

4.2 Methods

4.2.1 Overview of the study

This study aimed to characterise the eukaryome composition of wild house mice and identify any host factors that may be associated with their eukaryome composition. To do so, mice were live-trapped from three sampling sites in England and their disease and immune state measured, as well as other general host characteristics. Then, 18S rRNA amplicon sequencing was used to describe and compare the gut eukaryome composition among the sampling sites. Then, the effect of the host factors on eukaryome diversity and composition was measured. Finally, the gut bacteriome of the mice was described, and potential correlations between the diversity of the gut eukaryome and bacteriome were investigated.

4.2.2 Sample collection and processing

Wild house mice were live-trapped from three sampling sites in England between July and December 2021 (Figure 4.1). Mice were trapped with Longworth traps, which were baited with a mixture of oats, apple, peanut butter, Haribo sweets, and mixed seeds. Traps were set in the late afternoon and checked the following morning. A total of 58 mice were trapped across the three sites.

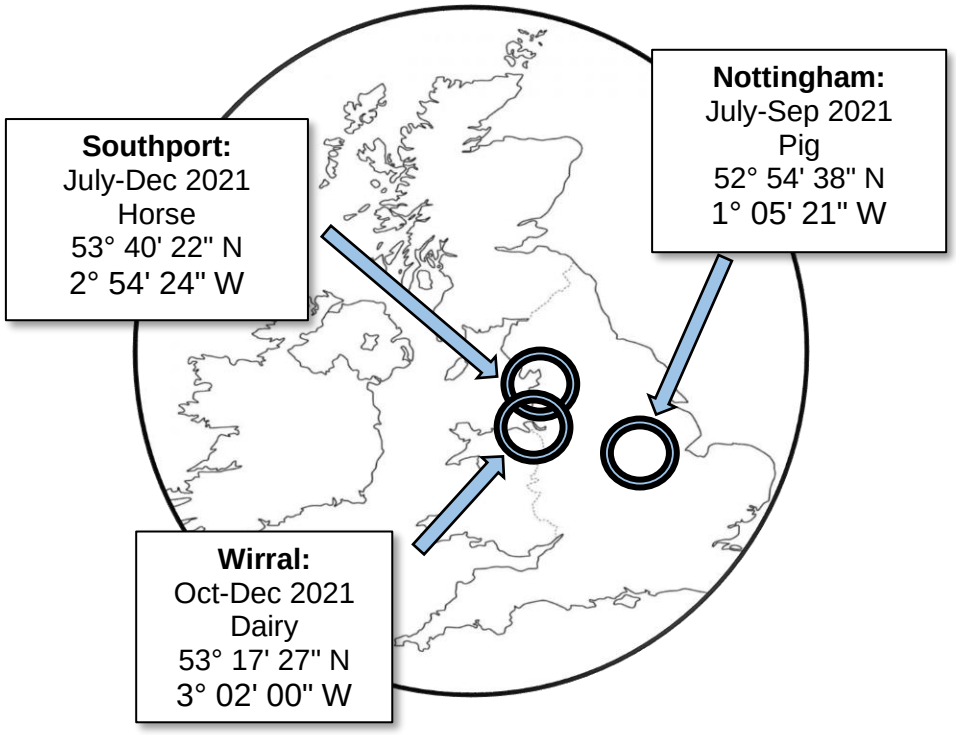


Figure 4.1. The location of the three sampling sites from which wild house mice were caught for characterising their gut eukaryome. The months of sampling, the farm type, and the longitude and latitude are given for each of the three sites.

Mice were culled by cervical dislocation. After culling, the sex, reproductive status, mass (g), and body length (cm) were recorded. Reproductive status was classed as either active or inactive. For males, those with scrotal (descended) testes were classed as active. Pregnant females, or those with enlarged nipples, were classed as active (Goertz *et al.*, 2019). Body length was measured from the tip of the snout to the base of the tail. Body condition was then calculated as log body mass / log body length, which can accurately predict fat mass of both sexes of mice (Labocha *et al.*, 2014). The eyes were dissected out and stored in 2 mL of 10% formal saline (10% w/v paraformaldehyde (Sigma-Aldrich) in phosphate-buffered saline (PBS, Sigma-Aldrich)) and later processed to calculate age as previously described (Rowe *et al.*, 1985). Briefly, lenses were removed from the eyes, oven dried until a constant weight, and the total weight of both lenses used to predict age (weeks) according to the equation provided by Rowe *et al.* (1985). The guts (stomach to distal large intestine) of the mice were exteriorised from the cadaver. The caecum was separated from the small and large intestine, cut open along the greater curvature (Williams *et al.*, 2016) and the contents washed into 2 mL of PBS. The caecal content was stored at -80°C immediately after collection. A pair of wet curved forceps were used to gently massage any pellets out of the distal end of the large intestine (Williams *et al.*, 2016). The pellets were stored at -80°C immediately after collection, along with any faecal pellets found in the trap. The stomach and small intestine, caecum, and large intestine were fixed in 25 mL bijou tubes filled with 10% formal saline.

4.2.3 Host disease and immune state

4.2.3.1 Gut antibody response

The gut intestinal antibody response was measured by assaying the faecal concentration of immunoglobulin A (IgA), the most common antibody in mucosal secretions (Gutzeit *et al.*, 2014). Faecal IgA data were kindly generated by Dr. Louise Cheynel. Briefly, faeces were weighed and dissolved in 50 µl PBS supplemented with a protease inhibitor cocktail at a final concentration of 20% v/v (SIGMAFAST™ Protease Inhibitor cocktail Tablets, EDTA-free, Sigma-Aldrich) to form a slurry that was left at room temperature for 1 h. The samples were then centrifuged at 13,000 x g for 10 min at 4°C, and the supernatant removed. Supernatant samples were serially diluted (in PBS with 1% v/v Tween and 10% w/v bovine serum albumin (BSA, Sigma-Aldrich)), starting from a 1:50 dilution in a doubling series to a 1:1600 dilution. The IgA concentration of the supernatants was then measured using a mouse IgA ELISA kit following the manufacturer's instructions (IgA Mouse Uncoated ELISA Kit with Plates, Invitrogen). Standard curves, made using the mouse IgA supplied with the kit, were used to calculate faecal IgA concentrations, expressed as mg IgA/g of faeces.

4.2.3.2 Faecal mucin

Mucin proteins are an important aspect of mucosal immunity in mammals (McGuckin *et al.*, 2011) that can be quantified by determining the faecal concentration of N-Acetylgalactosamine (GalNAc). GalNAc is a central region within the mucin monomer and so it's concentration can be a proxy for mucin concentration (Ansia and Drackley, 2020). GalNAc concentration was measured using a commercial fluorometric assay kit (Crowther and Wetmore, 1987; COSMO BIO CO, CSR-FFA-MU-K01E). Faecal samples were processed following the manufacturer's instructions and sample fluorescence was measured using a FlexStation 3 microplate reader, set to ex: 336 nm, em: 383 nm. Two adjustments were made to the manufacturer's instructions. Firstly, rather than a single measurement of fluorescence, a double measurement was taken, with the readings averaged, to account for random, technical variation. Secondly, the GalNAc concentration calculated from the standard curve (µg/mL) was converted to mg/g of faeces using the formula: 100/(faecal powder mass*50). This corrected for variation in the starting mass of faecal powder. Faecal mucin concentration is expressed as mg GalNAc/g of faeces. For two mice, there was insufficient volume of faeces to carry out the fluorometric assay. Three laboratory mouse controls were used: i) a male specific-pathogen free (SPF) C57BL/6 (henceforth called "SPF"), ii) a female non-SPF C57BL/6 (henceforth called "standard") and iii) a female from a captive colony derived from wild house mice (henceforth called "wildling"). The controls allowed comparison of the faecal mucin concentration of wild mice to laboratory mice.

4.2.3.3 Gut histology and assessment of gut inflammation

The fixed gut tissue was histologically processed and analysed for evidence of inflammation. To do this, five sections of gut tissue were cut for each mouse: 1 cm length cut of the small intestine (one replicate); 0.5 cm length cut of the large intestine (two replicates); 0.2 cm length cut of the caecum (two replicates). The tissue sections were placed into tissue cassettes and orientated to allow cutting of longitudinal sections of the large and small intestines and transverse sections of the caecum. Sections were then processed for staining by the Veterinary Pathology Diagnostic Service, University of Liverpool. The staining process was: embedding the tissue sections in paraffin wax, cutting at 3-5 µm thickness, and staining using haematoxylin and eosin (H&E).

The H&E-stained slides were used to score inflammation in the small intestine, caecum and large intestine. Two types of inflammation were scored: inflammation mediated by luminal antigens (for all five sections) and chemically-induced inflammation (for the large intestine and caecum). The scoring used methods described in Erben *et al.*, (2014), where a greater score was indicative of greater inflammation. Briefly, chemically-induced colonic inflammation was scored by categorising the i) severity and extent of immune cell infiltrate in the lamina propria

and ii) the amount of erosion and ulceration. Luminal antigen-mediated inflammation was scored by the i) severity and extent of immune cell infiltrate in the lamina propria, ii) the amount of hyperplasia and goblet cell loss, and iii) the presence of ulcerations and crypt loss. For the small intestine, scoring of luminal antigen-mediated inflammation also included the presence of villous blunting and/or atrophy. For all combinations of tissues and inflammation scoring, the entirety of the sectioned tissue was scored. The scores of a subset of samples were independently checked to ensure accuracy. Due to errors in the dissection and cutting process, some tissues sections were not scored. The final sample sizes were 57, 55, and 53 for the small intestine, caecum, and large intestine, respectively. The gut tissue from the control laboratory mice in 4.2.3.2 was also processed and scored in the same way to compare to the gut inflammation of wild mice.

4.2.3.4 Eimeria infection

Eimeria is a common parasite of wild rodents (Duszynski, 2021), and so *Eimeria* infection was recorded to determine if this common parasite was associated with changes in the gut eukaryome. To identify *Eimeri*a infection, DNA from faeces and caecal content was extracted using the QIAamp PowerFecal Pro DNA Kit (Qiagen). For caecal content, the protocol was adjusted to account for their storage in PBS: prior to use in the kit, caecal samples were centrifuged at 20,000 x g and the supernatant discarded. The contents of the PowerBead Pro Tubes and 800 µL of solution CD1 (from the extraction kit) were then added to the pellet of caecal contents. The manufacturer's instructions were then followed as normal. The DNA for both faeces and caecal content was stored at -80 °C. These DNA samples were then used in a PCR to identify the presence of *Eimeria.* The primers used in the PCR, Ap5_Fwd (YAAAGGAATTTGAATCCTCGTTT) and Ap5_Rev (YAGAATTGATGCCTGAGYGGTC), were targeted to a region of the apicoplast genome, the plastid found in Apicomplexan protozoa (Jarquín-Díaz *et al.*, 2019). 1.5 µL of DNA template was used in the 15 µL PCR. The PCR thermocycling protocol was as previously described (Jarquín-Díaz *et al.*, 2019). DNA from oocysts of *E. falciformis* and *E. ferrisi* (kindly provided by Professor Emanuel Heitlinger) were used as positive controls. Successful amplification was defined as the presence of a band of the expected size (~448 bp) on a 1% w/v Tris-acetate-EDTA (TAE) agarose gel, run at 150 V for 30 mins. Amplification of samples was triplicated and mice were categorised as *Eimeria*-positive if all three triplicates showed evidence of successful amplification.

4.2.3.5 Statistical analyses

To compare the host factors of mice among the different sampling sites, analyses were carried out in RStudio. The age, body condition, and faecal IgA concentration of mice were compared among sites using an analysis of variance (ANOVA). Tukey's post hoc tests were used to conduct pairwise comparisons. Faecal IgA concentration and age were log transformed to fit

the assumptions of the ANOVA. For mucin concentration, a generalised linear mixed-effects model (GLMM, Bates *et al.*, 2015) was used to compare faecal GalNAc concentration among the sites. The GLMM incorporated the assay plate number, faecal starting weight and faecal source (trap or intestine) as random effect terms, and used a Gamma error distribution. A Wilcoxon one-sample test was used to compare the faecal mucin concentration of wild mice to the control, laboratory mice. For the inflammation scores of the large intestine and caecum, chemically-induced inflammation and antigen-mediated inflammation were correlated, and so only the antigen-mediated inflammation was analysed further (Supplementary Figure 4). The possible range of scores for antigen-mediated inflammation were: 0-8 (small intestine) and 0-4 (large intestine/caecum). However, the highest scores were 5 and 3 respectively, and so these will be considered as the maximum score for further analyses. For tissues where there were two replicates (large intestine and caecum), the higher (more inflamed) score was used, given the interest in how inflammation is associated with the gut eukaryome. Discordance in the scores of the replicates is suggestive of localised inflammation – mice that had a difference of >1 between tissue replicates were classified as having localised inflammation. To compare the inflammation scores among sites, Fisher's exact test was used – a significant *p* value was indicative of differences in the distribution of inflammation scores among mice from different sampling sites. Chi-squared tests were used to determine if the observed distribution of inflammation scores was different to the expected (equal) distribution, for each tissue type. The inflammation scores of wild mice were compared to the control, laboratory mice using a Wilcoxon one-sample test.

To compare the number of *Eimeria*-infected mice among sites, Fisher's exact test, and the subsequent pairwise tests, were used. The Holm method was used to account for multiple testing (Holm, 1979). Here, a significant *p* value indicated a difference in the proportion of *Eimeria*-infected mice among the sampling sites. Chi-squared tests were used to determine if the proportion of mice infected with *Eimeria* was different to the null (50%), for each sampling site. To compare the infection status of faecal and caecal samples, two tests were used. Firstly, the association between the infection status of faecal and caecal samples was tested using a chi-squared test. Secondly, a McNemar's chi-squared test was used to identify if either the faecal or caecal samples were more likely to be identified as positive for *Eimeria.* To compare the faecal IgA concentration, faecal mucin concentration, and intestinal inflammation between *Eimeria*-positive and *Eimeria*-negative mice, a Wilcoxon two-sample test, GLMM (as above), and Fisher's exact test were used, respectively.

4.2.4 Eukaryome analysis

4.2.4.1 DNA sequencing

To characterise the gut eukaryotic taxa of mice from the three sampling sites, the DNA extracted from faeces and caecal content (see above, 4.2.3.4) was amplicon sequenced. To do this, first-round PCR amplification was completed by targeting the V4 region of the 18S rRNA gene, using the primers 528F (GCGGTAATTCCAGCTCCAA) and 707R (AATCCRAGAATTTCACCTCT) (Novogene, 2023). The thermocycling conditions were 2 mins at 94 °C followed by 25 cycles of: 45 secs at 94 °C; 45 secs at 54 °C; 1 min at 72 °C, followed by 5 mins at 72 °C for extension. 2.5 µL of DNA template was used in the 25 µL PCR. Second-round amplification, library preparation, and sequencing were completed by the Centre for Genomic Research (CGR), University of Liverpool. Sequencing used a single Illumina MiSeq v3 run (2x300bp), generating 10,477,596 reads for the total of 116 samples. A negative and positive control were also sequenced. The negative control was a blank extraction *i.e.* the DNA extraction protocol was completed without any starting material (Kim *et al.*, 2017). The positive control was a microbial community DNA standard, obtained from ZymoBIOMICS™ (D3605). The community contained two species that would be expected to amplify when using 18S amplicon sequencing: *Saccharomyces cerevisiae* and *Cryptococcus neoformans.*

4.2.4.2 Bioinformatics

To generate taxonomy data, the raw sequences provided by the CGR were analysed using QIIME 2 2021.2 (Bolyen *et al.*, 2019). The sequences were trimmed of primers and CGR adapters using cutadapt (Martin, 2011). The paired-end reads were then merged and the data were quality filtered and denoised using DADA2 to produce amplicon sequence variants (ASVs) (Callahan *et al.*, 2016). The reverse reads were truncated at 220 bp to account for a decrease in sequencing quality after this point. The ASVs were then aligned using the mafft programme (Katoh *et al.*, 2002) and a phylogeny based on ASV sequence similarity created using fasttree (Price *et al.*, 2010). Sequences matching the 528F and 707R primers, and their taxonomic classification, were extracted from the SILVA 138 database (Quast *et al.*, 2013) to train a Naïve Bayes classifier using the feature-classifier tool (Bokulich *et al.*, 2018). Taxonomy was then assigned to each ASV using the trained Naïve Bayes classifier. The sequence data and host variable metadata were then transferred into R Studio using the qiime2R package (v0.99.6, Bisanz, 2018) for subsequent analysis. ASVs that failed to match the SILVA database at the phylum level were checked against the NCBI nt database using BLAST (Altschul *et al.*, 1990; Sayers *et al.*, 2021) and taxonomy assigned manually using the R phyloseq package (v1.38.0, McMurdie and Holmes, 2013).

4.2.4.3 Data processing

All eukaryome data processing and analyses were carried out in RStudio. ASVs were agglomerated at the species level to simplify analyses, retaining any unassigned taxa (phyloseq package). ASVs that occurred fewer than ten times across the entire dataset were removed from all samples, following Mann *et al.* (2020). ASVs assigned to five phyla were removed from all samples: Phragmoplastophyta, Chlorophyta, Klebsormidiophyceae (plants); Arthropoda (insects); and Vertebrata (host). The SILVA-defined ASV classification was checked against two recent reviews of eukaryotic taxonomy to ensure accurate comparison to other studies (Ruggiero *et al.*, 2015; Adl *et al.*, 2019). Henceforth, the SILVA-defined phyla Nematozoa and SAR are referred to as Nematoda and Bigyra, respectively. ASVs were then classified as coming from either a gut resident (GR) or a non-gut resident (NGR) as recommended by Lavrinienko *et al.* (2021a). To do this, fungal ASVs were first divided into: microfungi, macrofungi, and lichens (Microfungi Collections Consortium, 2022). All macrofungi and lichens were classed as NGRs (Lavrinienko *et al.*, 2021a), whereas microfungi were further categorised as either GRs or NGRs, according to two sources (Hallen-Adams and Suhr, 2017; Mann *et al.*, 2020). Protozoa taxa were also predicted to be GRs or NGRs using two sources (Mann *et al.*, 2020; Mathison and Sapp, 2021). Multicellular parasites of mammals were classed as GRs, whereas non-parasitic multicellular species were classed as NGRs (predicted using Mathison and Sapp (2021)). Notably, ASVs identified as coming from parasitic nematodes (orders Oxyurida and Trichocephalida) were identified in the dataset, and these were used to identify mice with nematode infections.

4.2.4.4 Comparative analyses

The overall aim was to describe and compare the eukaryome of mice from different sampling sites. To do this, the presence and relative abundance of ASVs, and their taxonomy, were described among the sampling sites for both the caecal and faecal samples, before focussing only on GR taxa. Then, to identify how the eukaryome composition varied between the sample types (faeces *vs.* caecal contents), the alpha and beta diversity of the two sample types were compared. Shannon's index was used as the measure of alpha diversity, calculated from ASV abundance using the microbiome package (v1.19.1, Lahti and Shetty, 2017). A generalised linear model (GLM) with a Gamma error distribution was used to test for differences in Shannon's index between the sample types – an interaction term between sample type and sampling site was also included. To fit the assumptions of the GLM, Shannon's index was first reflected to achieve positively-skewed data, and $\log_{10}$ transformed (Osborne, 2010). The model revealed an interaction between sample type and sample site, and so paired Wilcoxon signed-rank tests were used to compare the Shannon's index of the faecal and caecal eukaryome for each sampling site separately. In these models, Shannon's index was

untransformed. Bray-Curtis dissimilarity (BC) was used as the measure of beta diversity. BC was calculated using the relative abundance of ASVs, for those with a minimum read abundance of 5 in ≥ 5% of samples (Cao *et al.*, 2021), and a principal coordinate analysis (PCoA) was used to visualise the data. A permutational multivariate analysis of variance (PERMANOVA) was used to identity if the sample type explained variation in BC, *via* the adonis function (vegan package, v2.5.7, Oksanen *et al.,* 2020). The interaction term between sample type and sampling site was also included in the model. The interaction was significant so PERMANOVAs were used to compare BC between sample types for each sampling site separately.

To compare the eukaryome among sampling sites, the caecal and faecal samples were analysed separately. Analyses were done at four taxonomic levels: ASV, genus, family, and phylum, with four comparative metrics: alpha diversity, beta diversity, variation within sampling sites, and differential abundance of taxa. Firstly, to compare the alpha diversity of the faecal eukaryome of mice from different sampling sites: Kruskal-Wallis tests compared Shannon's index at the ASV, genus, and family level and an ANOVA compared Shannon's index at the phylum level. In these models, Shannon's index was untransformed. Tukey's post hoc test was used to conduct pairwise comparisons among sites for phylum-level Shannon's index. Secondly, to compare the beta diversity of the faecal eukaryome among the three sampling sites, a PERMANOVA was used to test if the sampling site explained variation in BC. PERMANOVA pairwise comparisons were conducted for the three sites, using the Benjamini-Hochberg (BH) correction for multiple testing (Benjamini and Hochberg, 1995). Thirdly, to identify if the amount of variation in the faecal eukaryome community composition within sampling sites differed among the sites, the betadisper function was used, with adjustment for sample size bias (vegan package). The betadisper function generates the median distance of mice within a sampling site to the centroid in multivariate space, for each site. The permutest function (vegan package) was then used to compare the amount of variation in BC among sampling sites. Finally, the differential abundance of taxa in the faecal eukaryome was compared between sampling sites, at the ASV and phylum level only. To do this, an analysis of composition of microbiomes with bias correction (ANCOM-BC) was used (Lin and Peddada, 2020). ANCOM-BC tests used the in-built conservative approach to account for small sample sizes and used the Holm method to account for multiple testing. The centred-log ratio (CLR) transformation was used to visualise the differences in abundance. To compare the eukaryome of the caecum among sampling sites, the above analyses were repeated for the caecal samples with two differences: i) all alpha diversity comparisons used ANOVAs to compare untransformed Shannon's index and ii) differential abundance of taxa was compared at the phylum level only.

4.2.4.5 Host factors

To investigate how host factors affect eukaryome diversity and composition, faecal and caecal samples were analysed separately. The host factors of interest were: age, body condition, sex, reproductive state (from section 4.2.2), faecal IgA concentration, faecal mucin concentration, gut inflammation, *Eimeria* infection (from section 4.2.3) and helminth infection (from section 4.2.4.3). Three of the host factors were covariates, identified from correlation tests on Pearson's product moment correlation coefficient: log transformed age, log transformed faecal IgA concentration, and body condition (Supplementary Figures 5A-C). To account for this, a principal component analysis (PCA) was carried out. The PCA used the untransformed variables, which were scaled to have unit variance. The first principal component (henceforth called PC1) explained 59% of the variance in these three host factors, and so was used to represent this variation in models (Supplementary Figure 5D). All three variables were negatively correlated with PC1 (Supplementary Figure 5E).

For alpha diversity, linear models (LM) were used to determine which host variables explained variation in Shannon's index. For the faecal eukaryome, Shannon's index was recalculated prior to running the model after removing ASVs assigned to *Eimeria*, Oxyurida and Trichocephalida. The explanatory variables in the initial LM were: PC1, sex, reproductive status, faecal mucin concentration and gut inflammation. *Eimeria* infection status and nematode infection status, and their interactions with the other host factors were also included. Interaction terms were excluded from the final reported model if not significant. For the caecal eukaryome, two LMs were used. In the first LM, the explanatory variables were PC1, sex, reproductive status, faecal mucin concentration and gut inflammation. In the second LM, the explanatory variables were *Eimeria* infection status, Oxyurida infection status and *Trichuris muris* infection status. For this second caecal LM, Shannon's index was recalculated after removing ASVs assigned to *Eimeria*, Oxyurida and Trichocephalida, as done with the faecal eukaryome.

For beta diversity, PERMANOVAs were used to test if a host variable significantly explained variation in BC. Host variables were tested separately, with BH correction to account for multiple testing. BC between samples was recalculated for PERMANOVAs testing for an effect of parasitic status, after removing ASVs assigned to that focal parasite. For differential abundance of taxa, ANCOM-BC was used to identify taxa that had a significantly different abundance between parasitized (*Eimeria,* Oxyurida, and/or *Trichuris*) and unparasitized mice. For alpha and beta diversity, Shannon's index and BC were calculated at the ASV level, whereas differential abundance of taxa was compared at the phylum level.

4.2.5 Eukaryome and bacteriome comparisons

4.2.5.1 Bacteriome analysis

The overall aim was to determine if any patterns identified in the gut eukaryome of mice were also seen in the bacteriome, and then to consider if there was a causal relationship. To do this, the bacteriome was characterised for all 58 mice from 16S rRNA amplicon sequencing data kindly provided by Dr. Louise Cheynel. The sequences were generated from faecal DNA that was amplified and sequenced using primers targeting the 16S rRNA gene V4 region: 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) (Caporaso *et al.*, 2011). DNA extraction was completed by Dr Louise Cheynel and PCR amplification, library preparation, and sequencing were completed by the Centre for Genomic Research (CGR), University of Liverpool. The positive control was a mixture of DNA from eight bacterial species, from a microbial community DNA standard (ZymoBIOMICS^{TM}, D3605). A total of 7,382,623 reads were generated from the 58 faecal samples. The 16S amplicon sequences were then processed to generate taxonomy data following the pipeline used for the 18S sequences (section 4.2.4.2) with the following changes: the reverse reads were truncated at 210 bp; sequences matching the 515F and 806R primers were extracted from the SILVA 138 database; and ASVs unclassified at the phylum level were not checked against the NCBI database. As with the 18S ASVs, 16S ASVs that occurred fewer than ten times across the entire dataset were removed from all samples. ASVs assigned to Archaea were also removed. The presence and relative abundance of 16S ASVs, and their taxonomy, were described among the sampling sites. Henceforth, to ensure clarity, gut microbial taxa identified from 18S and 16S ASVs are referred to as the eukaryome and bacteriome, respectively. The term 'microbiome' is used to refer to both the eukaryome and bacteriome together.

4.2.5.2 Comparisons between the gut eukaryome and bacteriome

To compare the diversity and composition of the eukaryome and bacteriome, diversity metrics were first generated for the bacteriome. The same four metrics used for the eukaryome were generated and compared among sampling sites (section 4.2.4.4): alpha diversity, beta diversity, variation within sampling sites, and differential abundance of taxa. Comparisons were made at the ASV level only, in contrast to the eukaryome analysis. Then, comparisons of faecal eukaryome and faecal bacteriome alpha and beta diversity were made. To do this, differences in alpha diversity between the eukaryome and bacteriome were tested using a paired Wilcoxon signed-rank test, first for all mice, and then for each sampling site separately. Correlations between eukaryome and bacteriome alpha diversity was tested using Spearman's correlation coefficient, for all mice, and for each sampling site separately. In both the Wilcoxon and Spearman's correlation tests, untransformed Shannon's index was the measure of alpha diversity. For beta diversity, Mantel tests (vegan package) were used to

identify correlation of eukaryome and bacteriome BC. Positive correlations identified from Mantel tests of BC would indicate that as mice became more dissimilar in their eukaryome community composition, they also become more dissimilar in the bacteriome community composition. As with alpha diversity, Mantel tests were carried out first for all mice, then for each sampling site separately.

To investigate which host factors affect both eukaryome and bacteriome diversity and composition, the effect of host factors on the bacteriome were investigated (section 4.2.4.5). For alpha diversity, all host factors were tested in one model, including *Eimeria* infection status and nematode infection status. Only one model was used because Shannon's index did not need to be recalculated following removal of reads assigned to *Eimeria*, Oxyurida and Trichocephalida, as seen for the eukaryome.

4.2.5.3 Microbiome network analysis and co-occurrence

To investigate ecological interactions in the gut microbiome, associations between microbial taxa were predicted using the R package SpiecEasi (v1.1.2, Kurtz *et al.*, 2015). SpiecEasi predicts microbial associations using ASV abundance data. To do so, the SpiecEasi pipeline generates an inverse covariance matrix from the abundance data after i) accounting for compositional datasets commonly seen in amplicon sequencing datasets and ii) normalization to make ASVs independent of one another (Kurtz *et al.*, 2015; Tipton *et al.*, 2018). For these analyses, the SpiecEasi neighbourhood selection method (MB) was used, with the number of subsamples set to 50, and all other settings kept as the default. Predictions were first made for the eukaryome and bacteriome separately, before combining both datasets to infer associations within the microbiome. Analyses were completed for each sampling site separately. The input data for the analyses were untransformed ASV read abundances, following aggregation at the family taxonomic rank. Only the families present in ≥ 20% of mice were included in the analysis. For ASVs unclassified at the family level, the lowest known taxonomic rank was used. The output from SpiecEasi was then plotted as networks, using the R package igraph, to make biological inferences from the predicted associations (v1.2.6, Csardi and Nepusz, 2006). For clarity, the taxa in the network are nodes, and the predicted associations between taxa are edges connecting the representative nodes.

Network connectivity was compared among sampling sites using three measures: i) node degree, ii) percentage of singletons, and iii) the number and size of components. Firstly, node degree is the number of edges connected to that node: a network with an average node degree of one indicates that on average, the nodes in that network are connected with only one other taxon. Secondly, the percentage of singletons is defined here as the percentage of nodes with no connected edges. Thus, a high percentage of singletons is indicative of lower

network connectivity. Thirdly, components are groups of nodes connected by edges, with component size being equal to the number of nodes in that component *e.g.* a component of size one is a singleton node. Networks with few, large components have greater connectivity than networks with many, small components. The biological inference from comparing network connectivity is that more taxa are interacting with each other in a network with greater connectivity, and that taxa with a high node degree are likely important in the microbiome ecological network, potentially keystone species (Kurtz *et al.,* 2015; Tipton *et al.,* 2018). To identify how the gut microbiome ecological networks differed among sampling sites, the taxa with a high node degree and the taxa found within large components were compared among sampling sites. Positive edges between nodes infer that these taxa are co-occurring, and so the proportion of positive and negative edges was also compared among sampling sites.

Complementary to the network analyses, correlation tests were used to identify correlations in the abundance of taxa in the gut microbiome. Read depth was used as the measure of abundance, following aggregation of ASVs to the phylum level. Only the eukaryotic and bacterial phyla present in ≥ 20% of mice were included in the analysis. Spearman's correlation coefficient was used to identify negative or positive correlations between phyla abundance, and significance was inferred using *p* values following correction for multiple testing *via* the BH method.

4.3. Results

4.3.1 The host factors of the wild house mice

4.3.1.1 General characteristics

To survey the gut eukaryome of wild house mice and investigate what host factors affect eukaryome composition, 58 mice were sampled across from three sampling sites (Table 4.2). Males were caught more frequently than females. Nottingham-caught mice were significantly older than Southport mice (ANOVA: $F_2 = 8.13$, $p < 0.001$**,** Figure 4.2A). Mice from Southport also had a lower body condition score compared to Nottingham and Wirral mice (ANOVA: $F_2 = 7.93$, $p < 0.001$, Figure 4.2B).

Table 4.2. **Descriptive characteristics of wild mice caught from the three sampling sites.**

Characteristic		Nottingham	Southport	Wirral	Combined
Number Caught		31	15	12	58
Sex (M/F)		20/11	9/6	10/2	39/19
Age (weeks)	Mean (±SE)	10.0 (± 1.3)	4.6 (± 0.6)	7.3 (± 1.0)	8.0 (± 0.8)
	Range	2-41	2-9	4-16	2-41
Reproductively Active (%)		16 (52%)	3 (20%)	9 (75%)	28 (48%)
Body Condition	Mean (±SE)	1.3 (± 0.0)	1.2 (±0.0)	1.3 (±0.0)	1.2 (± 0.0)
	Range	1.1-1.5	1.0-1.3	1.2-1.4	1.0-1.5

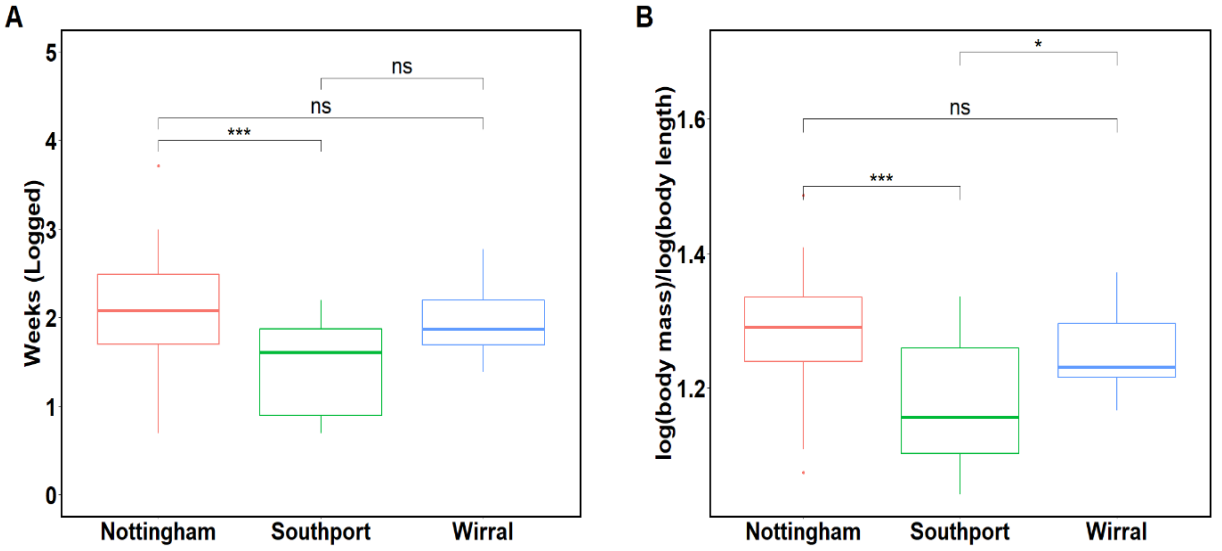


Figure 4.2. The A) age and B) body condition of wild mice from three sampling sites. Age (weeks) is log transformed. Body condition is calculated as log body mass (g) / log body length (cm). Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. *** $p < 0.001$, * $p < 0.05$, ns $p > 0.05$ from Tukey's post hoc tests.

4.3.1.2 Immunological factors

The immune state of mice was measured in three ways: faecal IgA concentration, faecal mucin concentration and intestinal inflammation. The faecal IgA concentration of mice was significantly different among the different sampling sites (ANOVA: $F_2 = 14.8$, $p < 0.001$, Figure 4.3A). The faecal IgA concentration (mean ± SE) was significantly lower in Southport mice (129 mg/g ± 19) compared to Nottingham (851 mg/g ± 142) and Wirral (442 mg/g ± 147) mice (TukeyHSD: $p < 0.001$ and $p = 0.035$, respectively). There was no difference in the faecal GalNAc concentration among sites (GLMM: Wald's $X^2 = 0.093$, $p = 0.955$, Figure 4.3B). However, wild mice had a significantly higher faecal GalNAc concentration on average, compared to all three laboratory mice (Table 4.3). Specifically, the concentration (mean ± SE) was approximately three times greater in wild mice (9.52 ± 1.4) compared to three laboratory mice (3.01 ± 0.32).

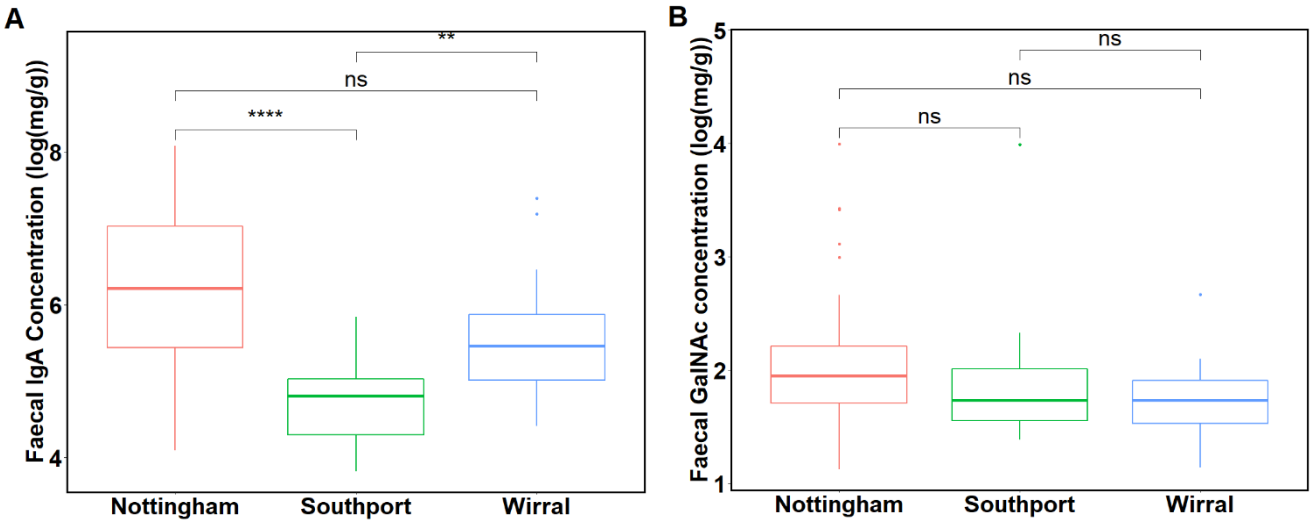


Figure 4.3. The faecal A) IgA and B) GalNAc concentrations of wild mice from three sampling sites. Both concentrations (mg /g of faeces) are log transformed. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. **** $p < 0.0001$, ** $p < 0.01$, ns $p > 0.05$ from A) Tukey's post hoc tests and B) a GLMM.

Table 4.3. The faecal mucin concentration of three laboratory mice, compared to wild mice. GalNAc concentration is measured as a proxy for mucin concentration. The average concentration ± SE is given for wild mice (n = 56). Comparisons were made *via* a Wilcox one-sample test. V is the test statistic.

Mouse	Faecal GalNAc concentration (mg/g of faeces)	V	*p* value
SPF	3.63	1,583	< 0.001
Standard	2.83	1,596	< 0.001
Wilding	2.57	1,596	< 0.001
Wild	9.52 (± 1.4)		

There was no difference in the distribution of inflammation scores for the small intestine (Fisher's exact test, $p = 0.448$) nor the large intestine (Fisher's exact test, $p = 0.071$) among mice from different sampling sites (Figure 4.4). However, there was a significant difference in the distribution of inflammation scores for the caecum among mice from the different sampling sites (Fisher's exact test, $p = 0.007$), with a higher proportion of Nottingham-caught mice having greater inflammation (Figure 4.4B). Chi-squared tests indicated that inflammation scores were significantly higher than expected for the large intestine and caecum for all sampling sites (Supplementary Table 3). In contrast, inflammation scores in the small intestine were equally distributed in mice from Nottingham and Wirral, and significantly lower than expected in mice from Southport (Supplementary Table 3). Localised inflammation in the large intestine was identified in eight mice, whereas only one mouse had localised inflammation in the caecum. There was no common pattern among these mice that may explain the localised inflammation seen (Supplementary Table 4). Wild mice had greater levels of inflammation in the small and large intestine compared to laboratory mice (Table 4.4). In the caecum, wild mice had more inflammation compared to the two C57BL/6 laboratory mice, but lower inflammation compared to the wildling mouse.

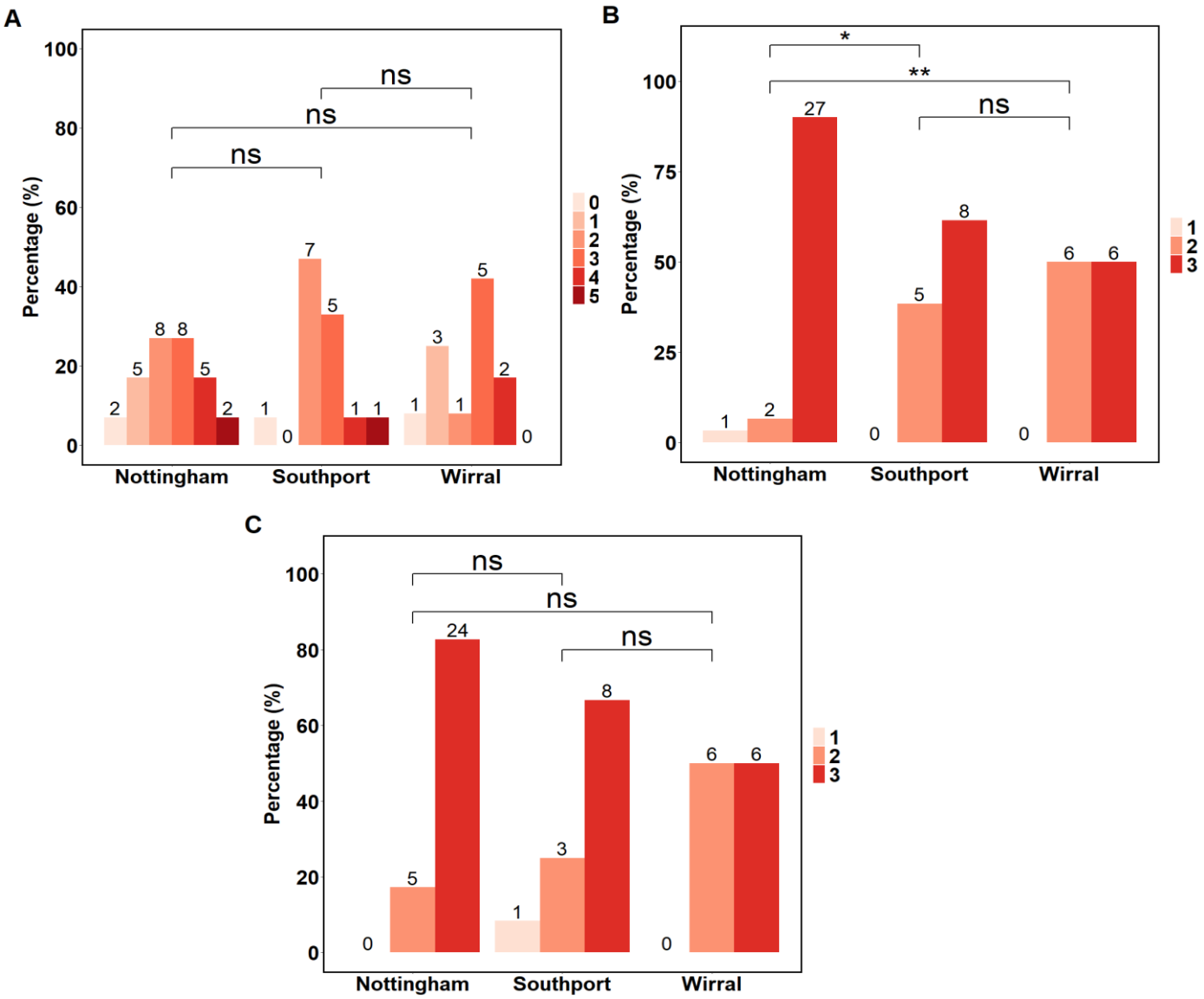


Figure 4.4. Luminal antigen-mediated inflammation in the gut of wild mice caught from three sampling sites. A) Small intestine, n = 57, B) Caecum, n = 55, C) Large intestine, n = 53. The y axes are the percentage of mice for each possible score. A higher score, shown by the darker colour, is indicative of greater inflammation. Counts for each score are shown on top of the columns. The range of scores possible were 0-8 (small intestine) and 0-4 (caecum and large intestine). ** $p < 0.01$, * $p < 0.05$, ns $p > 0.05$ from pairwise Fisher's tests.

Table 4.4. The inflammation scores of three laboratory mice, compared to wild mice. The average inflammation scores ± SE are given for wild mice (n = 55, 55, 53 for the small intestine, caecum, and large intestine, respectively). Comparisons were made using a Wilcox one-sample test. V is the test statistic.

	Small Intestine			Caecum			Large Intestine		
Mouse	**Score**	**V**	***p* value**	**Score**	**V**	***p* value**	**Score**	**V**	***p* value**
SPF	2	623	0.010	2	881.5	< 0.001	2	760	< 0.001
Standard	2	623	0.010	2	881.5	< 0.001	2	760	< 0.001
Wilding	2	623	0.010	3	0	< 0.001	2	760	< 0.001
Wild	2.5 (0.2)			2.7 (0.1)			2.7 (0.1)		

4.3.1.3 *Eimeria* prevalence and impact

PCR diagnostics identified there was a strong, significant association of *Eimeria* infection status between caecal and faecal samples (Pearson's Chi-squared test: $X^2_1 = 33.55$, $p < 0.001$). However, there was a significant difference in the probability of a positive infection status when comparing the two sample types (McNemar's Chi-squared test: $X^2_1 = 4.14$, $p = 0.041$), indicating that mice were more likely to show a positive infection when testing faecal samples compared to caecal samples. Comparison among sampling sites identified significant differences in the proportion of mice infected with *Eimeria*, for both caecal and faecal samples (Fisher's exact test: $p < 0.001$ and $p < 0.001$, Figure 4.5). Pairwise fisher tests indicated that this was driven by Nottingham (proportion infected: caecal: 90%, faecal: 97%), which was significantly different to both Southport (caecal: 13%, $p < 0.001$, faecal: 20%, $p < 0.001$) and Wirral (caecal: 33%, $p < 0.001$, faecal: 58%, $p = 0.008$) (Figure 4.5). There was no difference between the proportion of mice infected with *Eimeria* when comparing mice caught from Wirral and Southport (caecal: $p = 0.357$, faecal: $p = 0.057$). Chi-squared tests for each sampling site separately supported these findings: Nottingham had a significantly greater proportion of *Eimeria*-positive mice than the null (50%), Southport mice had a significantly smaller proportion of *Eimeria*-positive mice than the null, whereas there was no difference to the null in the proportion of *Eimeria*-positive in Wirral mice (Supplementary Table 5).

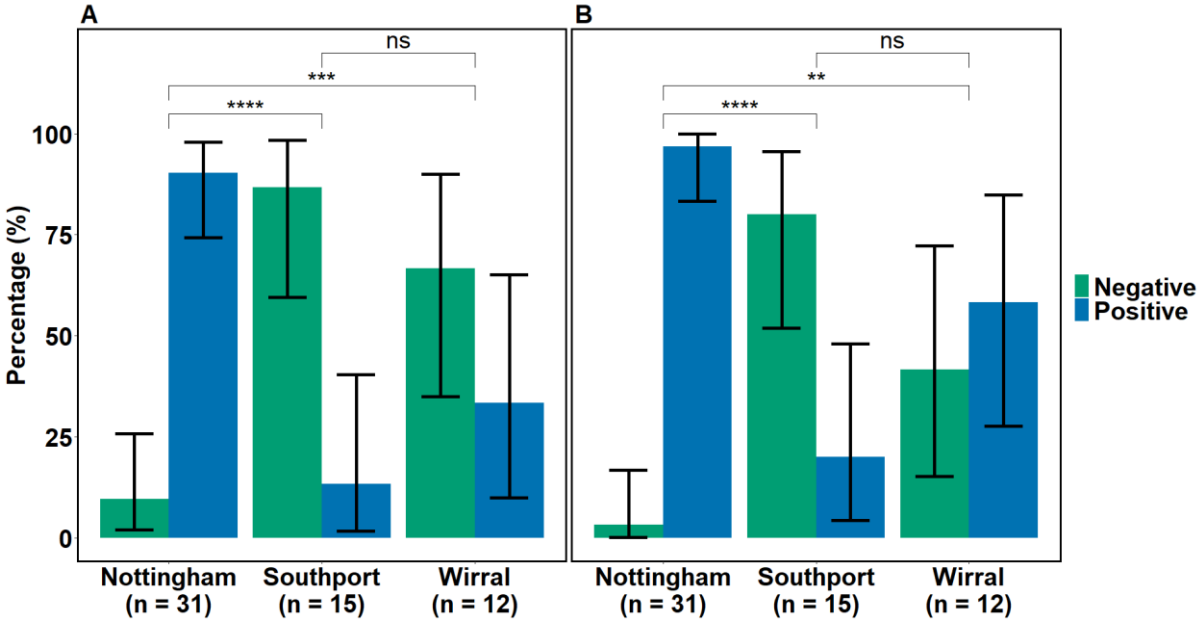


Figure 4.5. The percentage of mice infected with *Eimeria* from three sampling sites. A) Caecum, B) Faeces. Infection status was determined by PCR. Sample size for each site is shown on the x axes. Error bars are 95% binomial confidence intervals. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, ns $p > 0.05$ from pairwise Fisher's tests.

Following the identification of *Eimeria* in mice from all three sampling sites, the potential impact of *Eimeria* on the host immune state was investigated. Firstly, comparison of *Eimeria*-positive and Eimeria-negative mice identified a significant difference in faecal IgA concentration (Wilcox two-sample test: W = 170, p = 0.001) and faecal GalNAc concentration (GLMM: Wald's X^2 = 6.40, p = 0.011). *Eimeria*-positive mice had a greater faecal IgA concentration (mean ± SE: 705 mg/g ± 118) compared to *Eimeria*-negative mice (302 mg/g ± 106, Figure 4.6A). Faecal GalNAc concentrations were also higher in *Eimeria*-positive mice (mean ± SE: 9.6 mg/g ± 1.6) compared to *Eimeria*-negative mice (9.3 mg/g ± 2.7, Figure 4.6B). Finally, there was no difference in the distribution of gut inflammation scores when comparing *Eimeria*-positive and *Eimeria*-negative mice for the small intestine (Fisher's exact test: p = 0.603) or large intestine (p = 0.341). In the caecum, there was a mild difference in the distribution of inflammation scores, but this difference was not significant (p = 0.073) (Figure 4.7). These relationships are suggestive that *Eimeria* presence is associated with changes in the gut immune state of mice, particularly a higher faecal IgA and mucin concentration. However, the directionality of this relationship is unknown.

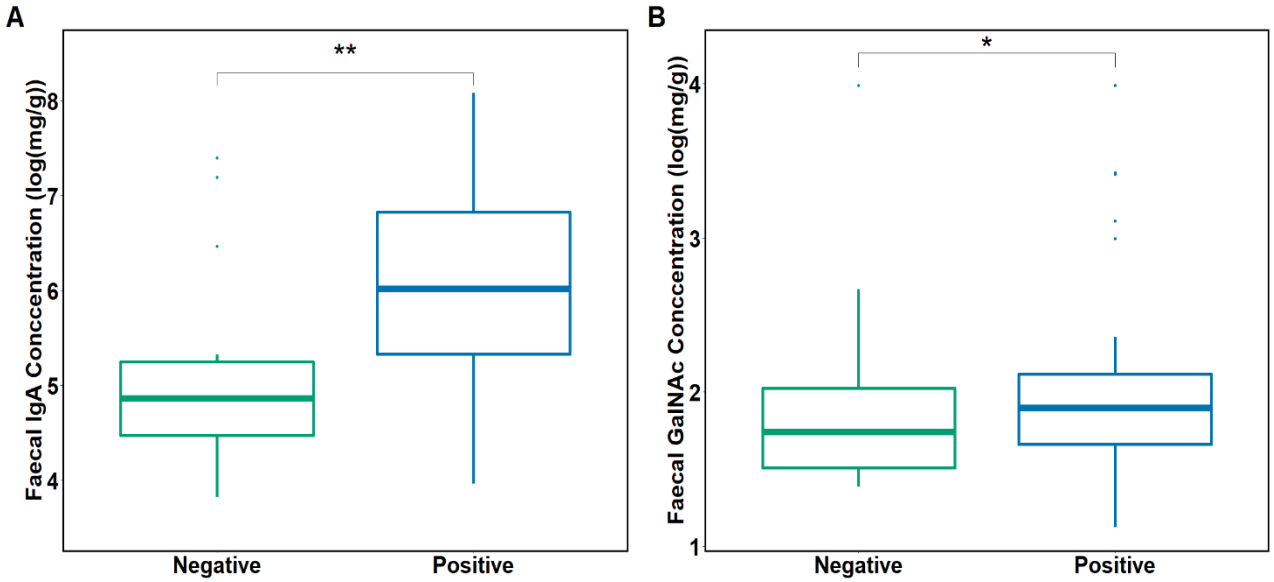


Figure 4.6. The faecal A) IgA and B) mucin concentrations of wild mice by *Eimeria* infection status. Both concentrations (mg/g of faeces) are log transformed. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. ** p < 0.01, *p < 0.05 from A) a Wilcox two-sample test and B) a GLMM.

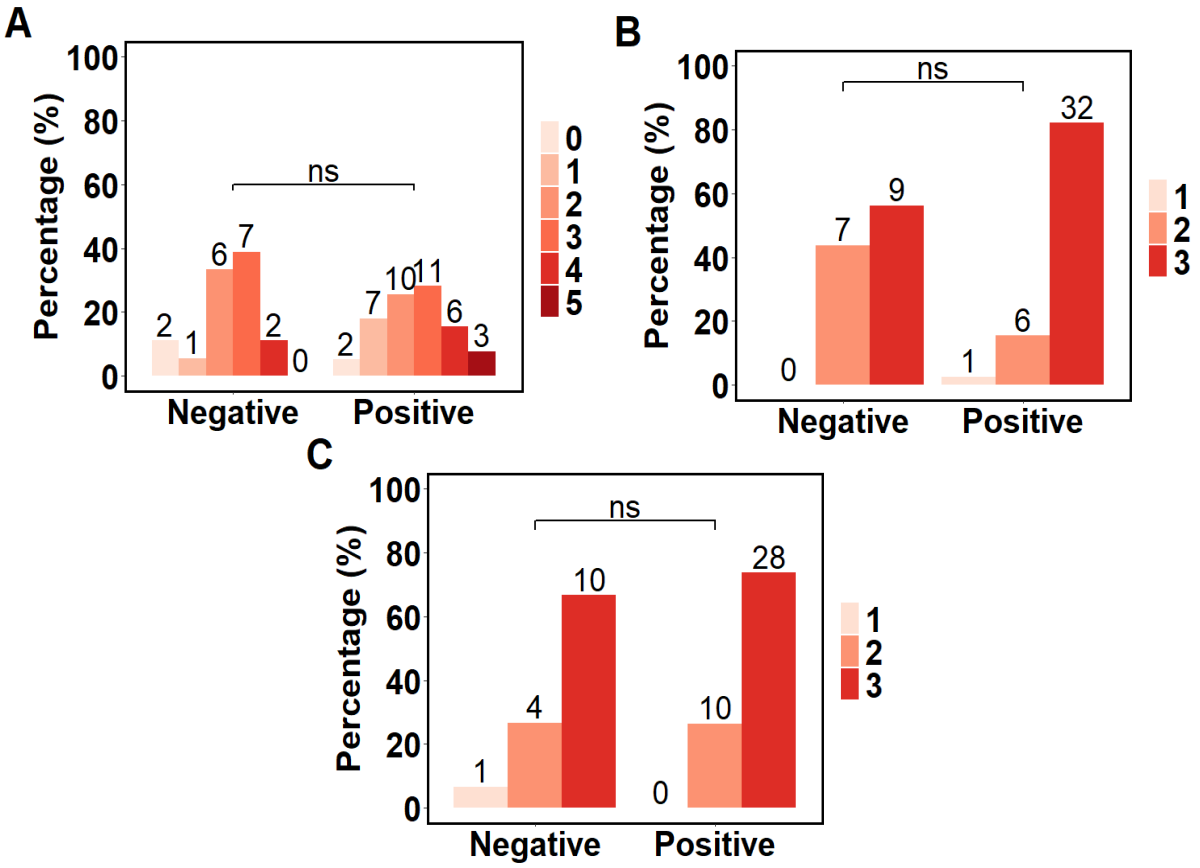


Figure 4.7. Luminal antigen-mediated inflammation in the gut of wild mice by *Eimeria* infection status. A) Small intestine, n = 57, B) Caecum, n = 55, C) Large intestine, n = 53. The y axes are the percentage of mice for each possible score. A higher score, shown by the darker colour, is indicative of greater inflammation. Counts for each score are shown on top of the columns. The range of scores possible were 0-8 (small intestine) and 0-4 (caecum and large intestine). ns $p > 0.05$ from pairwise Fisher's tests.

4.3.2 The eukaryome composition of wild house mice

4.3.2.1 Sequencing summary

18S rRNA amplicon sequencing was used to identify taxa in the gut eukaryome. Prior to analysis, ASVs in the negative control were examined to ensure there was no contamination of samples. Only one ASV was identified in the negative control, which was classified as the fungus *Rhodotorula*, and this ASV was present in 17/116 of the wild mouse samples, with a total abundance of 3,660. This ASV was removed from all samples before further analysis. The five ASVs identified in the positive control represented the two fungal species – *Saccharomyces cerevisiae* and *Cryptococcus neoformans* – present in the mock microbial community, confirming successful amplification of these two taxa. Following the removal of the control samples, a total of 2,666 ASVs were identified in the wild mice. Across both faecal and caecal samples (n = 116), the total read depth was 10,477,695, with an average (± standard error) reads per sample of 90,324 (± 4,193). After agglomeration of ASVs at the species level, removal of low abundance taxa, and removal of taxa classified as vertebrates, arthropods, and plants, the total read depth was 2,458,285 and 2,543,954 in caecal and faecal samples, respectively (Table 4.5).

Table 4.5. Summary of the read depths in caecal and faecal samples before and after filtering. ASVs were agglomerated at the species level before filtering.

	Caecal		Faecal	
	Pre-filter	**Post-filter**	**Pre-filter**	**Post-filter**
Total read depth	5,602,209	2,458,285	4,875,387	2,543,954
Average reads per sample (± SE)	96,589 (± 6,064)	42,384 (± 4,752)	84,058 (± 5,727)	43,861 (± 3,814)
Number of ASVs	1,381	379	2,417	535

4.3.2.2 Eukaryome diversity and composition

The presence of a eukaryotic taxon was indicated by the presence of ASV(s) assigned to that taxon. 32 eukaryotic phyla were identified across all 58 wild mice, with six phyla identified as potential gut residents (Table 4.6, Supplementary Table 6). In the faecal eukaryome, prevalence of these six phyla ranged from 100% (Ascomycota and Basidiomycota) to 14% (Bigyra), and relative abundance ranged from 42.5% (Ascomycota) to 0.3% (Ciliophora). ASVs coming from multicellular parasites (Nematoda) were present in 94% of Nottingham mice, but were absent from Southport mice. In comparison to the faecal eukaryome, fewer phyla (25) were present in the caecal eukaryome across all 58 mice (Table 4.6, Supplementary Table 7). The majority of phyla were more prevalent in the faecal eukaryome compared to the caecal eukaryome (Figure 4.8). Only the phylum Nematoda was more prevalent in the caecum (72%) compared to faeces (60%).

Table 4.6. Phyla identified from the faecal and caecal eukaryome of wild mice. Phyla are ordered by their relative abundance across all 58 wild mice. Only the phyla with a relative abundance of > 0.1% and > 0.05% are shown for the faecal and caecal eukaryome, respectively. The complete list of phyla can be found in Supplementary Tables 6 and 7. Abundance is the number of reads classified for each phylum. Prevalence is the percentage of mice from which that phylum was identified. Phyla in bold are known gut residents. N'ham = Nottingham, S'port = Southport.

	Phylum	Abundance		Prevalence (%)			
		Relative (%)	Total	Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Faecal	**Ascomycota**	42.5	1,080,915	100	100	100	100
	Basidiomycota	32.0	815,080	100	100	100	100
	Mucoromycota	13.4	341,999	100	100	100	100
	Apicomplexa	9.2	233,732	88	100	80	67
	Nematoda	0.7	18,520	60	94	0	50
	Cercozoa	0.6	15,977	88	94	87	75
	Bigyra	0.5	13,671	14	19	0	17
	Ciliophora	0.3	6,749	78	84	67	75
	Rotifera	0.1	3,613	17	26	13	0
	Ochrophyta	0.1	3,585	67	68	67	67
	Unclassified[c]	0.1	3,426	81	87	67	83
Caecal	**Nematoda**	41.1	1,011,293	72	100	27	58
	Basidiomycota	24.5	603,440	100	100	100	100
	Ascomycota	24.2	595,891	100	100	100	100
	Apicomplexa	6.0	147,763	88	100	73	75
	Mucoromycota	3.5	86,044	98	100	93	100
	Bigyra	0.14	3,461	7	13	0	0
	Cercozoa	0.14	3,355	57	52	80	42
	Ciliophora	0.07	1,811	52	52	33	75
	Unclassified[a]	0.07	1,680	48	32	67	67
	Ochrophyta	0.06	1,358	40	35	33	58

a ASVs not classified at the phylum level

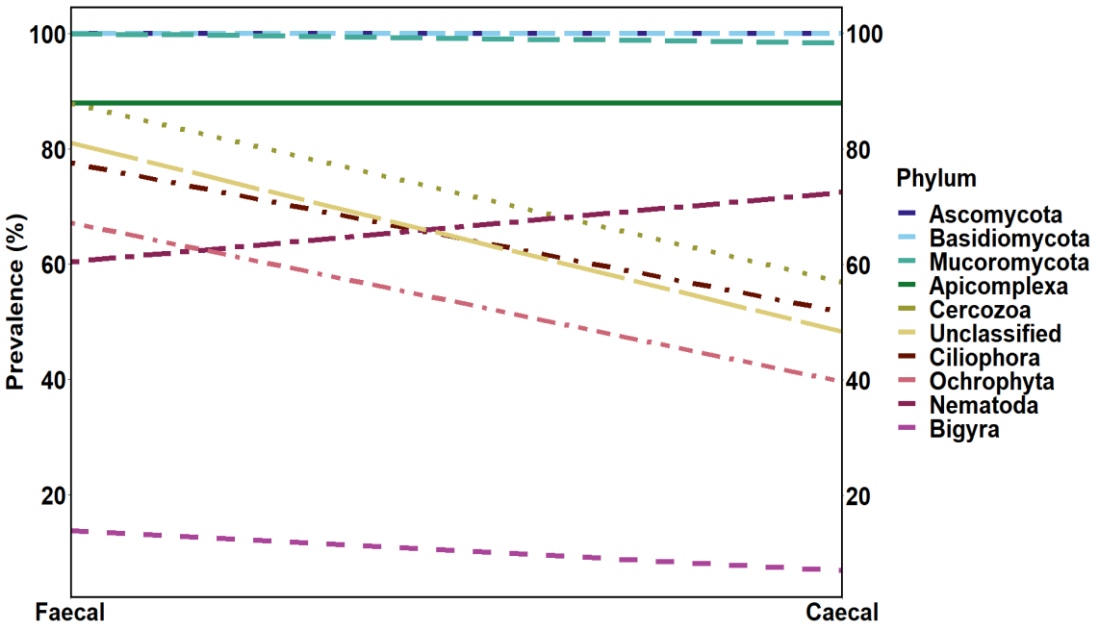


Figure 4.8. The prevalence of eukaryotic phyla identified in the caecal and faecal eukaryome of wild mice. Prevalence is calculated from all 58 wild mice. Only the phyla with a relative abundance of > 0.1% in the faecal eukaryome which also appear in the caecal eukaryome are shown. Phyla are ordered in the legend by prevalence in the faecal eukaryome.

There was variation in the relative abundance of phyla-assigned ASVs among and within sampling sites (Figure 4.9, Table 4.7). In the faecal eukaryome (Figure 4.9A), Ascomycota ASVs had the highest relative abundance in Nottingham mice, whereas Basidiomycota was highest in Southport and Wirral mice. Apicomplexa ASVs had a relative abundance of 17.3% in Nottingham mice, which was higher compared to Southport and Wirral mice. In the caecal eukaryome (Figure 4.9B), Basidiomycota ASVs had the highest relative abundance in Southport and Wirral mice, consistent with the faecal eukaryome. However, Nematoda had the highest relative abundance in Nottingham mice, which contrasts to the finding of Ascomycota in faecal samples.

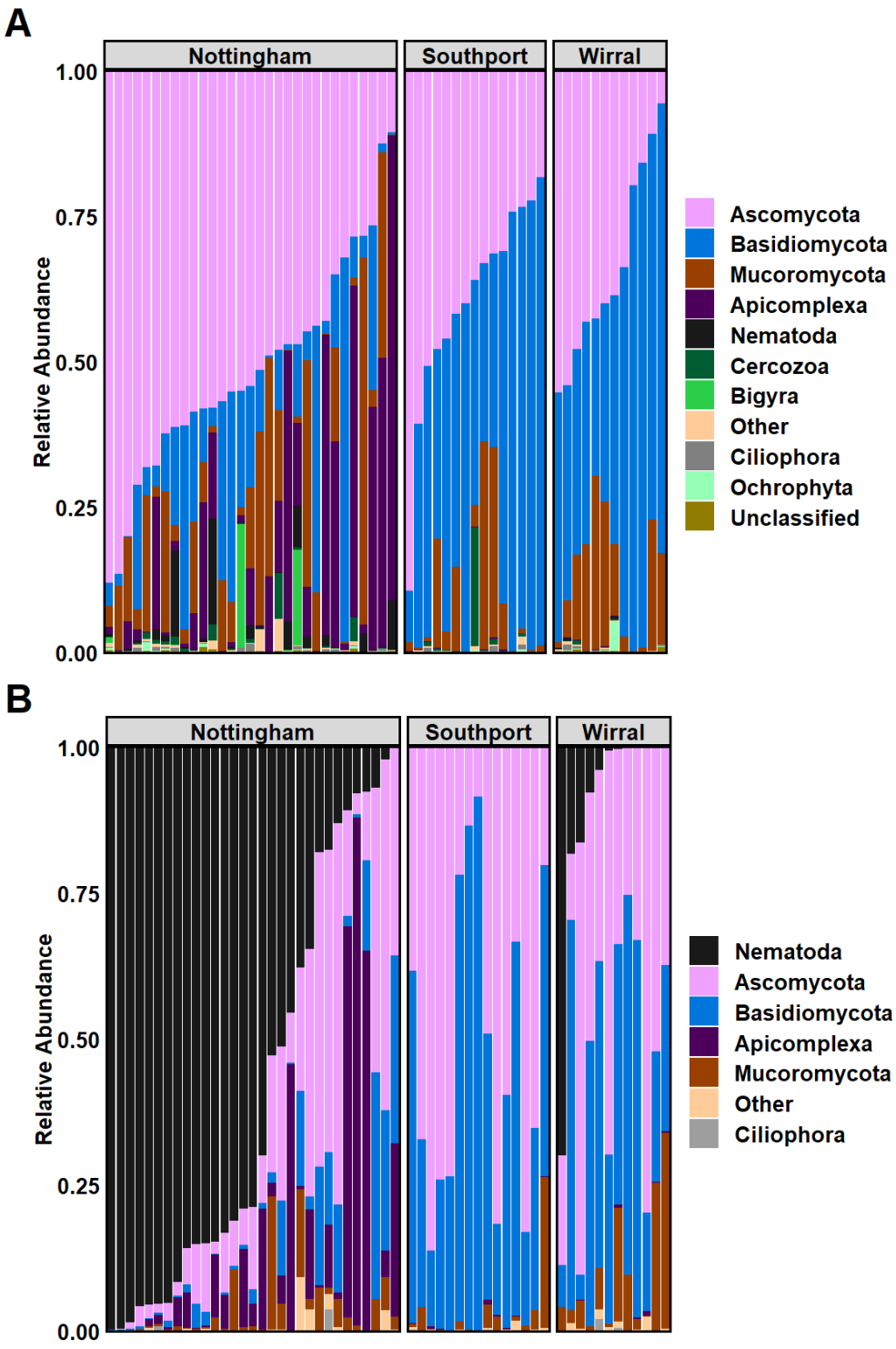


Figure 4.9. Relative abundances of eukaryotic phyla identified in the A) faecal and B) caecal eukaryome of wild mice. Relative abundance is the proportion of reads classified for each phylum. Only the phyla with a minimal read abundance of 200, in ≥ 5% of all mice, are shown. Less abundant phyla are grouped into Other. Phyla are ordered by their relative abundance across all mice.

Table 4.7. Relative abundances of eukaryotic phyla identified from the faecal and caecal eukaryome of wild mice. Relative abundance is the percentage of reads classified for each phylum. Only the phyla with a minimal read abundance of 200, in ≥ 5% of all mice, are shown. Phyla are ordered by their relative abundance across all mice.

	Phylum	Relative Abundance (%)		
		Nottingham	Southport	Wirral
Faecal	Ascomycota	46.9	42.1	34.2
	Basidiomycota	17.3	43.7	47.6
	Mucoromycota	12.0	15.0	17.0
	Apicomplexa	19.2	0.1	0.1
	Nematoda	1.5	NA	0.1
	Cercozoa	0.9	0.5	0.1
	Bigyra	1.1	NA	< 0.1
	Ciliophora	0.2	0.3	0.3
	Ochrophyta	0.1	0.1	0.3
Caecal	Nematoda	64.9	<0.1	7.4
	Ascomycota	16.3	38.4	35.0
	Basidiomycota	6.1	60.3	45.8
	Apicomplexa	9.7	0.1	0.2
	Mucoromycota	2.3	1.0	10.9
	Ciliophora	0.1	<0.1	0.2

To investigate which taxa were contributing to the phyla identified above, comparisons were made at the genus level (Figure 4.10). Of the 348 eukaryotic genera present in the faecal eukaryome, only eight were detected with more than 500 reads in least 20% of all mice (Figure 4.10A). *Wallemia* and *Eimeria* were the dominant genera contributing to the high relative abundance of Basidiomycota and Apicomplexa, respectively. Several genera of Ascomycota contributed to its high relative abundance. In the caecal eukaryome, eight (of 254) genera were present with more than 500 reads in least 20% of all mice (Figure 4.10B). Consistent with the faecal eukaryome, *Wallemia* and *Eimeria* were present in the caecal eukaryome, as well as several Ascomycota genera. The high relative abundance of Nematoda in the caecum was driven by the presence of the nematode order Oxyurida (undefined at the genus level) and the nematode genus *Trichuris*. Despite the high prevalence of Ciliophora and Cercozoa, no genera of these phyla were present in over 20% of mice.

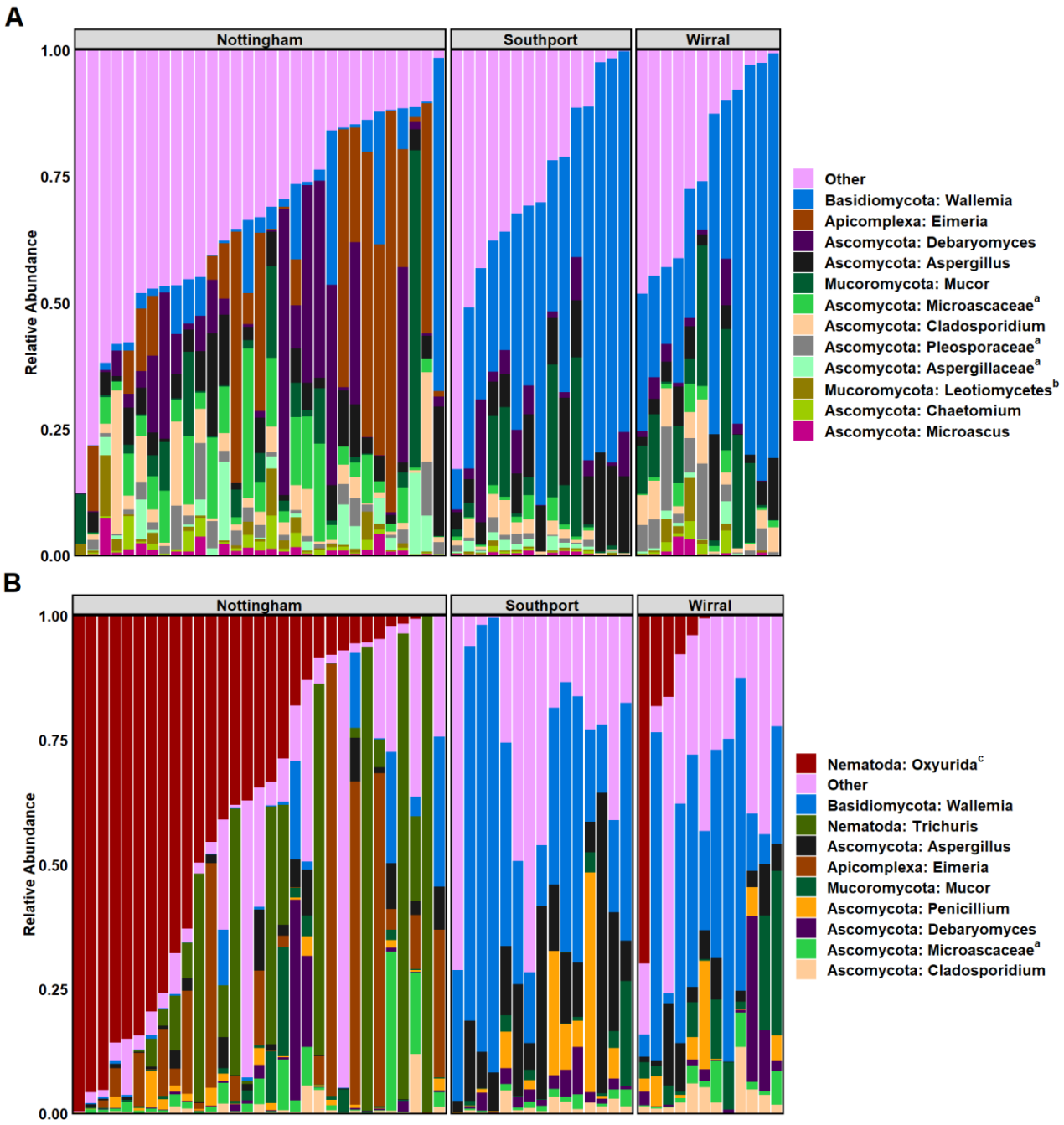


Figure 4.10. Relative abundances of eukaryotic genera identified in the A) faecal and B) caecal eukaryome of wild mice. Relative abundance is the proportion of reads classified for each genus. Only the genera with a minimal read abundance of 500, in ≥ 20% of mice, are shown. Less abundant genera are grouped into Other. Genera are ordered by their relative abundance across all mice. Some abundant ASVs were not classified at the genus level – the given taxonomic rank is [a] Family, [b] Class and [c] Order.

The eukaryotic taxa identified *via* ASV presence were then classified as either gut residents (GR) or non-gut residents (NGR). There were 14 genera, and two orders, of GR taxa in the wild mouse gut eukaryome (Table 4.8). In the faecal eukaryome, *Trichuris* was present only in Nottingham mice, but was present in the caecal eukaryome of mice from all three sampling sites. The nematode order Oxyurida was present in both faecal and caecal samples of Nottingham and Wirral mice, but was absent from Southport mice. The order Ascaridida (Nematoda) was the only taxonomic group present in the caecal eukaryome, but not the faecal eukaryome. *Strongyloides* was identified in one mouse from Nottingham. The fungal GR taxa included *Aspergillus, Penicillium, Candida*, and *Saccharomyces*. For protozoa, *Cryptosporidium* was present in mice from all three sampling sites, whereas *Blastocystis* was present in Nottingham and Wirral mice, but absent in Southport mice. *Neobalantidium* (= *Balantidium*) was identified in only one faecal sample from Nottingham. *Eimeria* was found in both the faecal (Nottingham mice) and caecal eukaryome (Nottingham and Wirral mice). However, the number of mice with *Eimeria* contrasted to the number identified using PCR diagnostics (4.3.1.3). PCR diagnostics identified a greater proportion of mice infected with *Eimeria* in all combinations of sample type (faecal *vs.* caecal) and sampling site (Supplementary Table 8).

The differences in presence and prevalence of eukaryotic taxa among the faecal and caecal eukaryome was investigated further. The alpha diversity (mean Shannon's index ± SE) of the faecal and caecal eukaryome was 2.2 (± 0.1) and 1.6 (± 0.1) respectively, and this difference was significant (GLM: Wald's X^2 = 15.4, p < 0.001, Figure 4.11A). There was a significant interaction between the sample type and sampling site (GLM: Wald's X^2 = 6.9, p = 0.032). Comparison of alpha diversity at each sampling site identified that alpha diversity was higher in the faecal eukaryome compared to the caecal eukaryome for Nottingham and Southport mice (Wilcox signed-rank test: V = 65, p < 0.001; V = 25, p < 0.048, respectively). For beta diversity, the faecal and caecal eukaryome compositions were significantly different, as measured by BC (PERMANOVA: F_1 = 9.26, R^2 = 0.060, p < 0.001, Figure 4.11B). However, there was a significant interaction between the sample type and sampling site (PERMANOVA: F_2 = 2.83, R^2 = 0.037, p < 0.001) and only Nottingham mice had significantly different caecal and faecal eukaryomes (PERMANOVA: F_1 =10.49, R^2 = 0.149, p < 0.001). These results suggest that are differences in eukaryome diversity and composition between caecal contents and faeces, but the amount of differentiation is specific to sampling site.

Table 4.8. The gut resident taxa identified in the dataset and their prevalence in the gut eukaryome of wild mice. Prevalence is the percentage of mice from which ASVs classified as that gut resident were identified. The number in brackets is the total sample size and sample size for each site. N'ham = Nottingham, S'port = Southport.

	Type	Genus	Prevalence (%)			
			Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Faecal	Nematode	Oxyurida[a]	53	90	0	25
		Trichuris	5	10	0	0
		Strongyloides	2	3	0	0
	Protozoa	*Eimeria*	41	77	0	0
		Cryptosporidium	31	42	13	25
		Blastocystis	14	19	0	17
		Neobalantidium	2	3	0	0
	Fungi	*Aspergillus*	98	97	100	100
		Penicillium	66	84	27	67
		Saccharomyces	29	16	40	50
		Glaebosa[b]	16	23	13	0
		Geotrichum	3	6	0	0
		Lodderomyces[b]	3	3	7	0
		Nakaseomyces[b]	3	0	0	17
		Malassezia	2	3	0	0
Caecal	Nematode	Oxyurida[a]	64	100	0	50
		Trichuris	43	65	27	8
		Ascaridida[a]	3	6	0	0
	Protozoa	*Eimeria*	47	77	0	8
		Cryptosporidium	29	32	13	42
		Blastocystis	5	10	0	0
	Fungi	*Aspergillus*	95	90	100	100
		Penicillium	81	84	80	75
		Malassezia	59	52	73	58
		Lodderomyces[b]	10	0	20	25
		Saccharomyces	9	3	13	17
		Glaebosa[b]	5	10	0	0
		Nakaseomyces[b]	3	0	7	8

a Identified to the order level, not genus
b *Candida*-associated clade

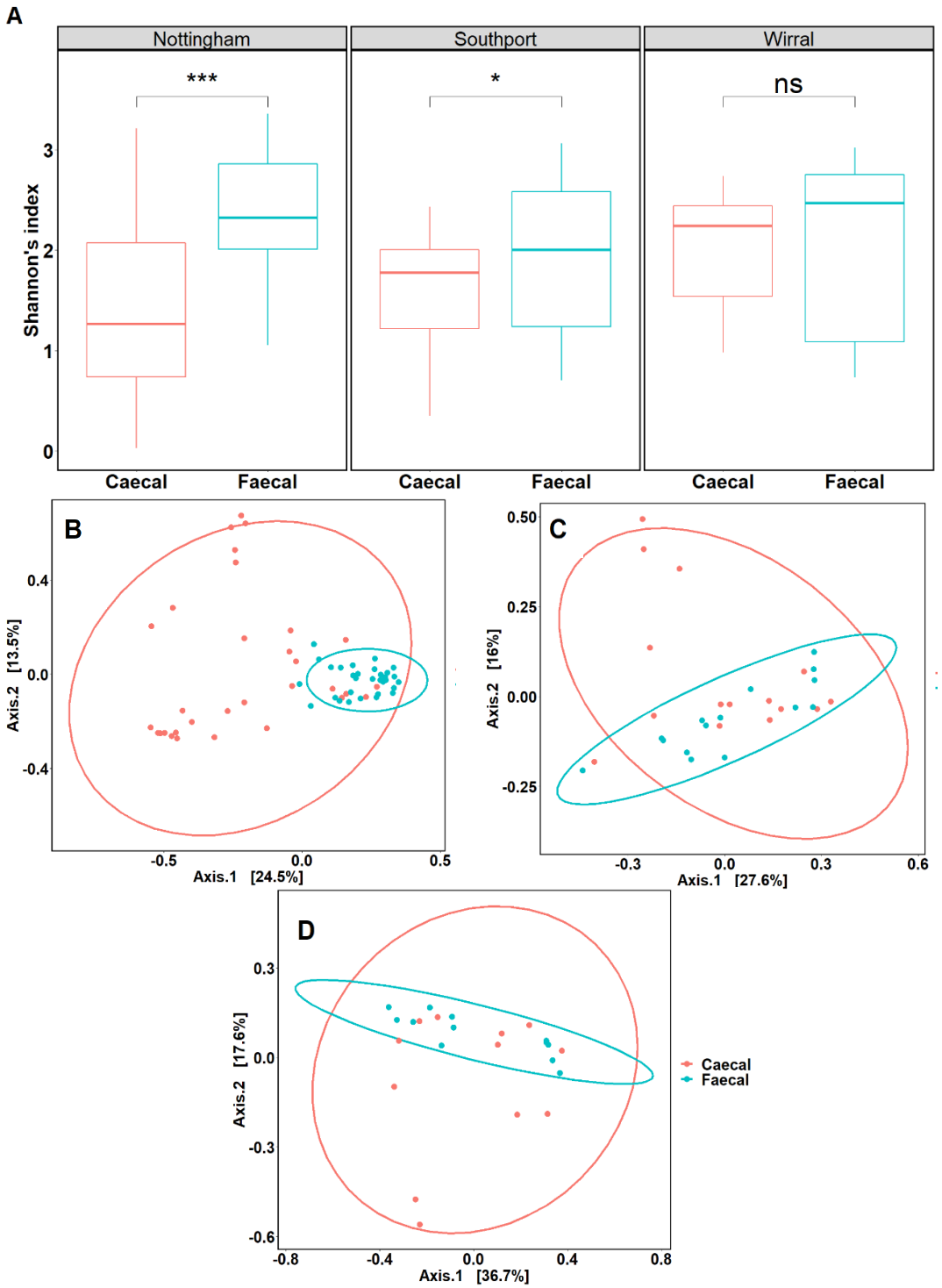


Figure 4.11. Comparison of the eukaryome diversity of wild mice between sample types. A) Comparison of Shannon's index split by sampling site. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. B-C) PCoA of Bray-Curtis dissimilarity between sample types, with the proportion of variance explained by each axis shown. B) Nottingham, C) Southport, D) Wirral. Ellipses are 95% confidence intervals. *** $p < 0.001$, * $p < 0.05$, ns $p > 0.05$ from Wilcox signed-rank tests.

4.3.2.3 Comparison of the gut eukaryome among sampling sites.

Following the identification of different prevalences and relative abundances of taxa among different sampling sites, the eukaryome diversity and composition was compared among the sampling sites. For the faecal eukaryome, there was no difference in ASV-level alpha diversity (mean Shannon's index ± SE) among mice from the three sampling sites (Kruskal-Wallis: X^2_2 = 3.47, p = 0.176): Nottingham (2.4 ± 0.1), Southport (1.9 ± 0.2), and Wirral (2.0 ± 0.3) (Figure 4.12A). This result was consistent at the genus and family level, but phylum-level Shannon's index was significantly higher for Nottingham mice (1.0 ± 0.0) compared to Southport mice (0.8 ± 0.1) (ANOVA: F_2 = 3.33, p = 0.042, Figure 4.12B, Supplementary Table 9). This contrasts to the caecal eukaryome, whose alpha diversity differed significantly among the mice from different sampling sites at the genus (ANOVA: F_2 = 3.40, p = 0.041), and family taxonomic ranks (ANOVA: F_2 = 3.68, p = 0.032, Supplementary Table 9). Post hoc comparisons showed that Wirral mice had a higher caecal alpha diversity, compared to Nottingham mice, for these taxonomic ranks (p = 0.031 and p = 0.024, Figures 4.12C and 4.12D).

For beta diversity, the sampling site of mice accounted for 21.1% of the variation in ASV-level BC of the faecal eukaryome (PERMANOVA: F_2 = 7.37, R^2 = 0.211, p < 0.001, Figure 4.13A). BC pairwise comparisons showed that Nottingham mice had a significantly different eukaryome composition compared to both Southport (p = 0.002) and Wirral mice (p = 0.002, Table 4.9). Comparison of beta diversity in the faecal eukaryome at different taxonomic ranks gave consistent results (Supplementary Table 10). Consistent with the faecal eukaryome, the sampling site accounted for 26.2% of variation in ASV-level BC of the caecal eukaryome (PERMANOVA: F_2 = 9.79, R^2 = 0.262, p < 0.001, Figure 4.13B). However, BC pairwise comparisons showed that caecal community composition was also significantly different between Southport and Wirral mice, in contrast to the faecal eukaryome (Table 4.9). Comparison of beta diversity in the caecal eukaryome at the genus and family taxonomic ranks gave consistent results (Supplementary Table 10). However, at the phylum level, there was no difference in eukaryome composition between Southport and Wirral mice (PERMANOVA: F_1 = 1.47, R^2 = 0.055, p < 0.207, Supplementary Table 10). The faecal eukaryome composition of mice was more varied in Nottingham compared to Southport and Wirral (Supplementary Table 11). This difference was significant at the ASV, genus, and family taxonomic ranks, but not at the phylum level (Supplementary Table 11). In contrast, the variation in the caecal eukaryome composition of mice at each sampling site was not significantly different, except between Nottingham and Southport at the ASV level (Supplementary Table 11).

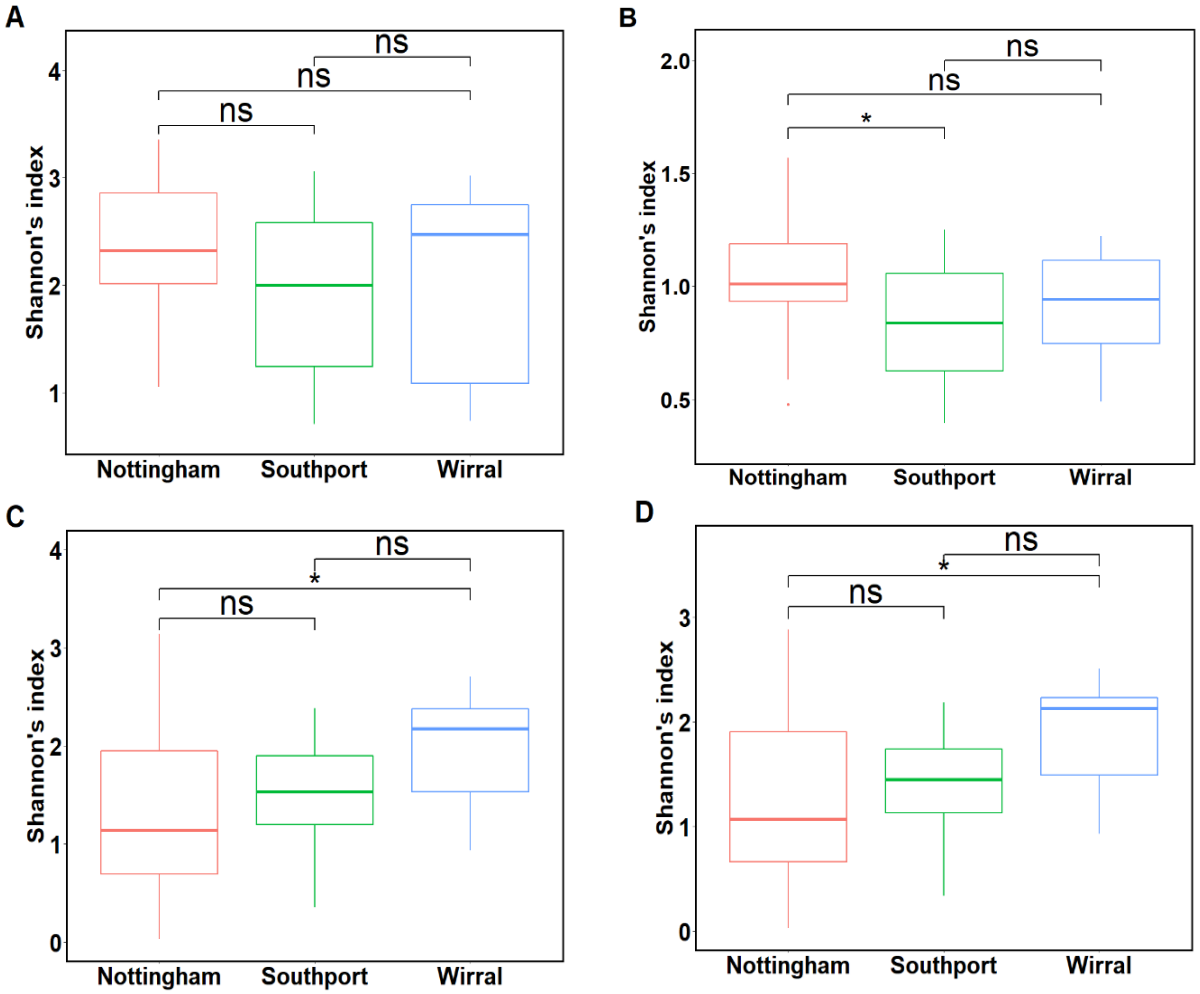


Figure 4.12. Shannon's index of the wild mouse caecal and faecal eukaryome at different taxonomic ranks. A) Faecal, ASV. B) Faecal, phylum. C) Caecal, genus. D) Caecal, family. Boxplots indicate the median v alue, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. The y-axis scale is unique to each figure. * $p < 0.05$, ns $p > 0.05$ from A) a Kruskal-Wallis test and B-D) Tukey's post hoc tests.

Table 4.9. Pairwise comparisons of gut eukaryome Bray-Curtis dissimilarity of wild mice from three sampling sites. Sampling sites were compared using a PERMANOVA, with BH correction for multiple testing. Significant *p* values are highlighted in bold.

	Comparisons	*p value*
Faecal	Nottingham: Southport	**0.002**
	Nottingham: Wirral	**0.002**
	Wirral: Southport	0.065
Caecal	Nottingham: Southport	**0.002**
	Nottingham: Wirral	**0.002**
	Wirral: Southport	**0.032**

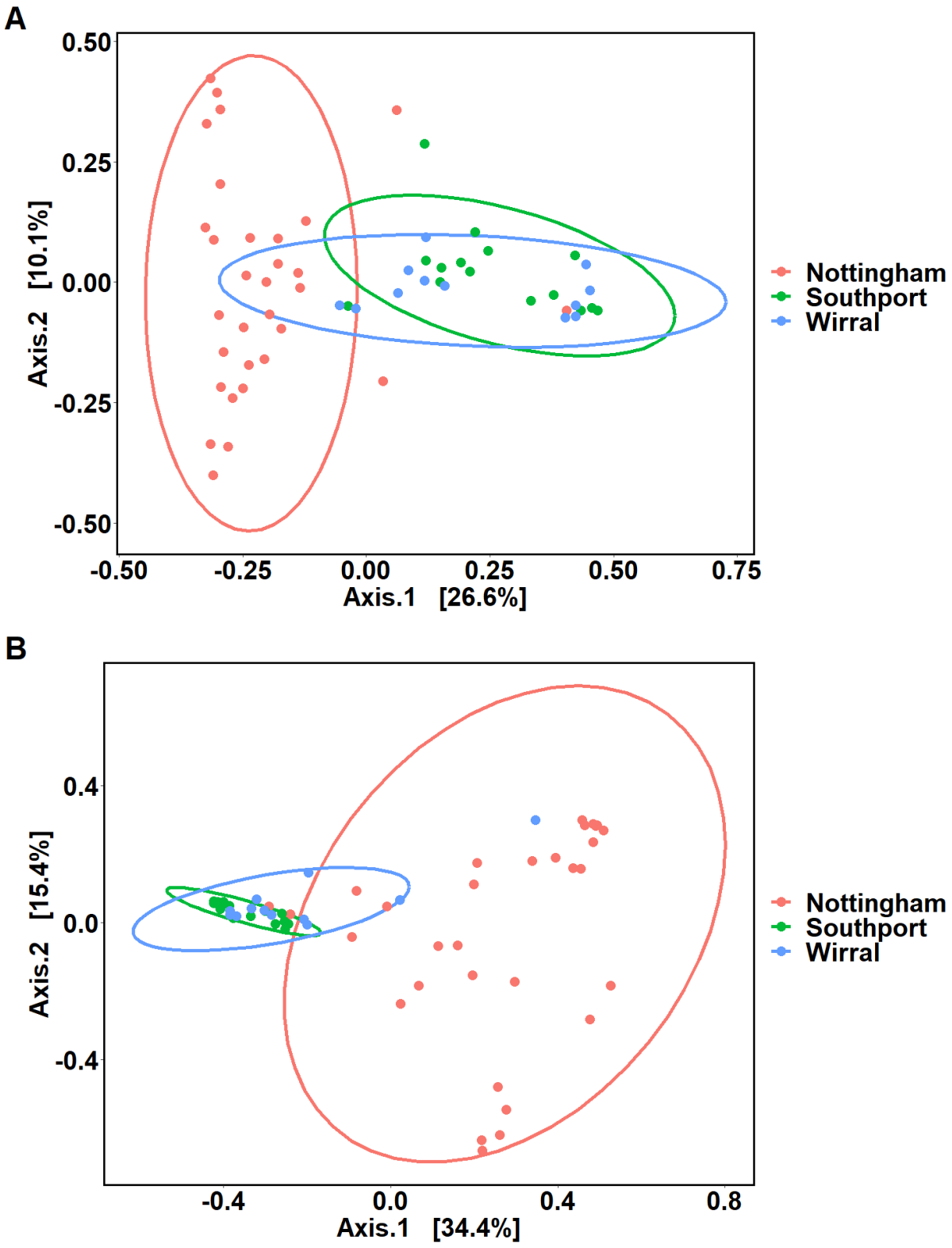


Figure 4.13. Wild mouse eukaryome Bray-Curtis dissimilarity PCoA plots by sampling site. A) Faecal eukaryome and B) caecal eukaryome. Ordination was generated using the ASV-level taxonomic rank. The proportion of variance explained by each axis is shown. Ellipses are 95% confidence intervals.

To account for there being no difference in the faecal community composition between Southport and Wirral mice, these mice were grouped together to compare the differential abundance of taxa against Nottingham mice. There were 14 species in faecal samples with a significantly different abundance when comparing Nottingham mice to Southport and Wirral mice (Figure 4.14). Seven taxa were more abundant in Nottingham, three of which are parasitic species: *Eimeria falciformis, E. telekii* and *Ozolaimus linstowi* (Supplementary Table 12). The species that were more abundant in Southport and Wirral were fungi, including two species of *Wallemia*. At the phylum level, Apicomplexa and Nematoda ASVs were more abundant in faecal samples of Nottingham mice, compared to Southport and Wirral mice (Figure 4.15A, Supplementary Table 13). Basidiomycota species were more abundant in Southport and Wirral mice compared to Nottingham mice. This pattern was also seen in caecal samples, with the addition of Ascomycota ASVs also being more abundant in Southport and Wirral mice (Figure 4.15B, Supplementary Table 13).

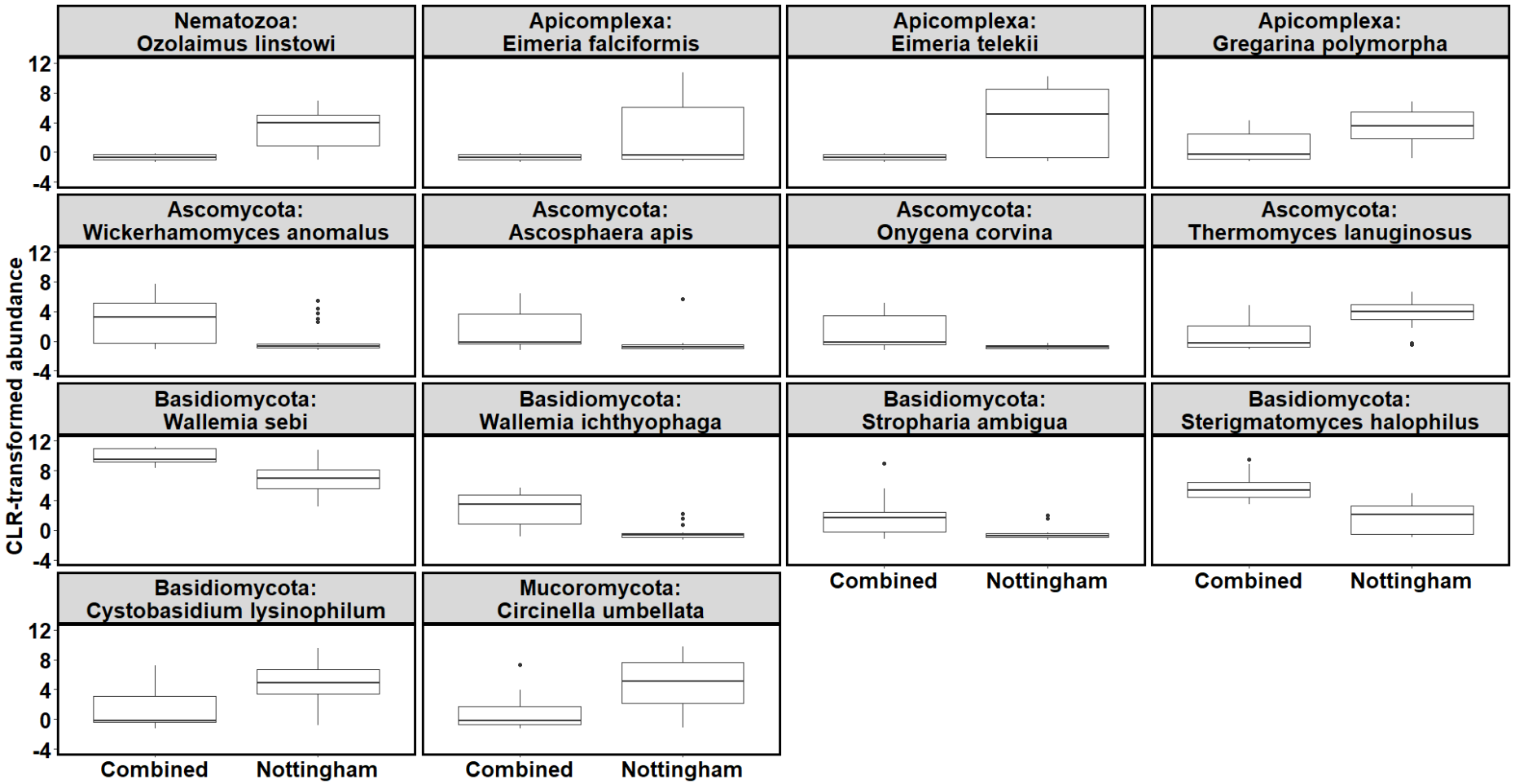


Figure 4.14: Species that were differentially abundant between the faecal samples of Nottingham mice and mice from the two other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance. ANCOM-BC analyses were used to identify the species with a significantly different abundance. Abundance values can be seen in Supplementary Table 12. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range.

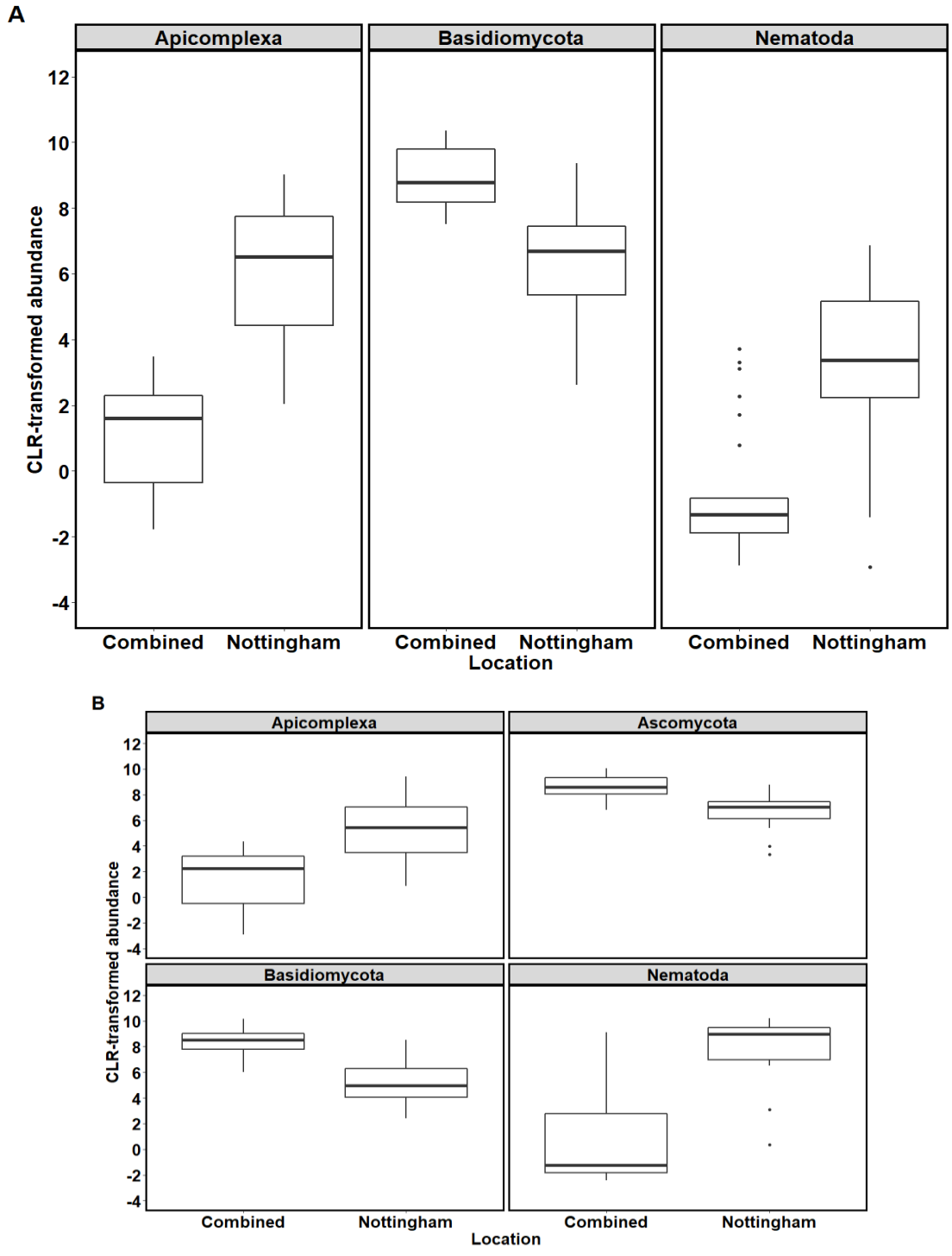


Figure 4.15. Phyla that were differentially abundant between the A) faecal and B) caecal samples of Nottingham mice and mice from other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance. ANCOM-BC analyses were used to identify the phyla with a significantly different abundance. Abundance values can be seen in Supplementary Table 13. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range.

4.3.3 Drivers of eukaryome composition

Having identified significant differences in eukaryome composition among and within sampling sites, the host factors described above (4.3.1) were investigated to determine their contribution to variation in eukaryome diversity and composition. For alpha diversity, linear models were used to identify factors that significantly affected the Shannon's index of the faecal and caecal eukaryome (Table 4.10). Oxyurida infection was the only explanatory factor that was significant in both the faecal and caecal models (LM: $F_1 = 10.7$, $p = 0.002$ and $F_1 = 9.2$, $p = 0.004$, respectively). Oxyurida-negative mice had a significantly lower Shannon's index compared to Oxyurida-positive mice, in both the faecal eukaryome (Welch two-sample t-test: $t_{45.8} = -3.3$, $p = 0.002$, Figure 4.16A) and caecal eukaryome (Welch two-sample t-test: $t_{36.5} = -3.3$, $p = 0.002$, Figure 4.16B). Inflammation of the small intestine was a significant factor in the caecal model (LM: $F_5 = 3.4$, $p = 0.012$). To investigate this further, an ANOVA was used to compare caecal eukaryome Shannon's index among mice with different small intestine inflammation scores, with Tukey's post hoc tests to conduct pairwise comparisons. The ANOVA identified a significant difference among the inflammation scores, consistent with that of the linear model (ANOVA: $F_5 = 3.05$, $p = 0.018$). The general trend observed was higher inflammation scores of the small intestine were associated with a lower caecal eukaryome Shannon's index (Figure 4.17). However, post hoc comparisons showed that the only significant difference in Shannon's index was between mice with small intestine inflammation scores of 1 and 3 (TukeyHSD: $p = 0.047$).

Table 4.10. Statistics for the two models testing for the effects of host factors on gut eukaryome alpha diversity in wild mice. Shannon's index was used as the measure of alpha diversity. Significant host factors are highlighted in bold. For the parasite model, Shannon's index was recalculated after removing ASVs assigned to *Eimeria*, Oxyurida and *T. muris*. *T. muris* infection was not included in the faecal parasite model as all mice infected with *T. muris* were also infected with Oxyurida.

	Model	Host factor	F value	*p* value
Faecal	$F_{15,34}$ = 1.99 *p* = 0.048	PC1	3.10	0.087
		Sex	0.03	0.866
		Reproductive status	1.21	0.279
		Faecal mucin concentration	0.57	0.457
		Small intestine inflammation	1.63	0.180
		Caecum inflammation	2.04	0.146
		Large intestine inflammation	0.62	0.545
		Eimeria	1.95	0.172
		Oxyurida	**10.71**	**0.002**
Caecal	Non-parasite: $F_{13:38}$ = 2.00 *p* = 0.049	PC1	0.42	0.519
		Sex	0.50	0.483
		Reproductive status	1.70	0.200
		Faecal mucin concentration	0.00	0.950
		Small intestine inflammation	**3.40**	**0.012**
		Caecum inflammation	3.23	0.051
		Large intestine inflammation	1.75	0.188
	Parasite: $F_{3:54}$ = 4.31, *p* = 0.008	*Eimeria*	0.09	0.771
		Oxyurida	**9.24**	**0.004**
		Trichuris muris	1.04	0.312

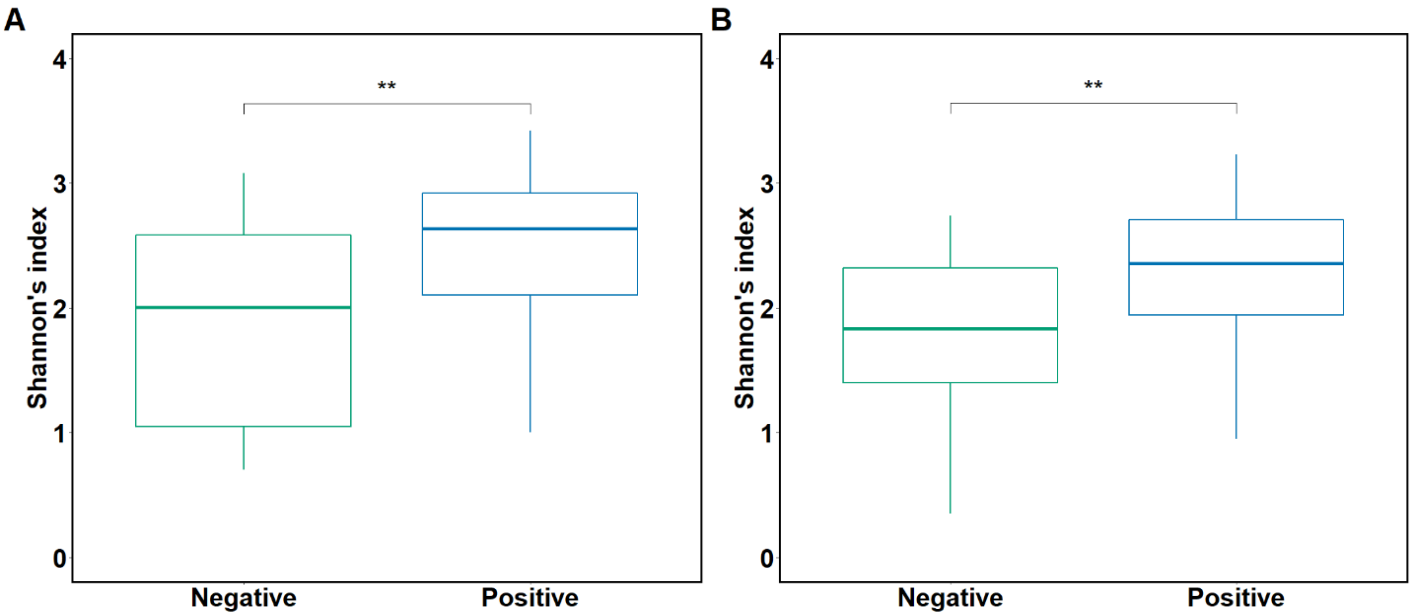


Figure 4.16. Shannon's index of the A) faecal and B) caecal eukaryome by Oxyurida infection status. Oxyurida-positive mice were defined by the presence of ASVs assigned to the order Oxyurida. Shannon's index was recalculated after removing ASVs assigned to *Eimeria*, Oxyurida and *T. muris*. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. ** $p < 0.01$ from Welch two-sample t-tests.

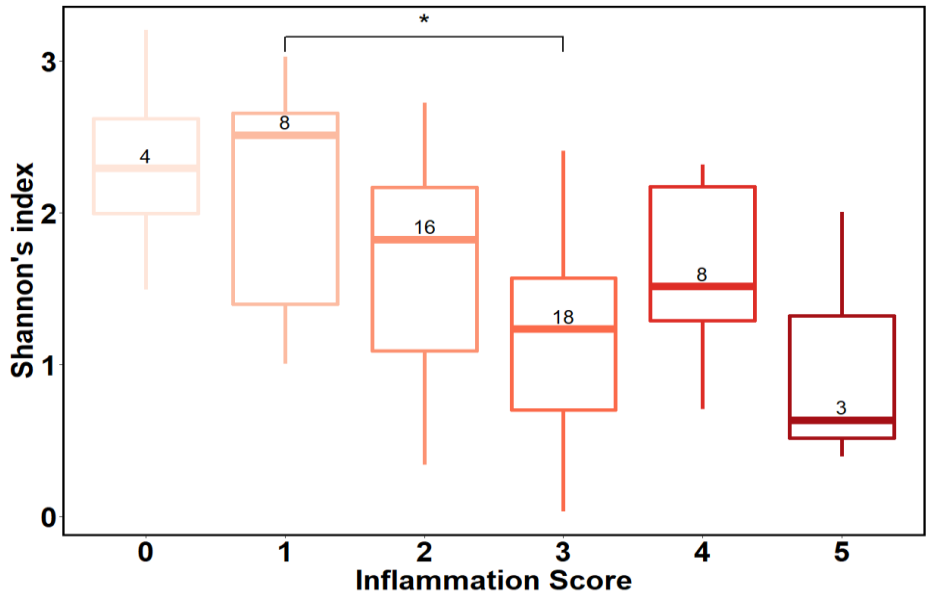


Figure 4.17. Shannon's index of the caecal eukaryome of wild mice by inflammation of the small intestine. A higher score, shown by the darker colour, is indicative of greater inflammation. The number of mice for each score is shown in the boxplots. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. Only scores 1 and 3 were significantly different. * $p < 0.05$ from Tukey's post hoc test.

In the faecal eukaryome, parasitism (infection with any of *Eimeria*, Oxyurida, and *Trichuris*) was a significant explanatory factor of community composition, accounting for 10.8% of the variation in BC (PERMANOVA: $F_1 = 6.78$, $R^2 = 0.108$, $p = 0.003$). When tested independently, *Eimeria* and Oxyurida were significant, accounting for 9.8% and 14.0% of the variation in BC in faecal samples, respectively (Table 4.11, Figure 4.18). These factors were also significant explanatory factors of caecal eukaryome composition (Table 4.11). *T. muris* infection accounted for 5.1% of the variation in BC in the caecal eukaryome, and was a significant explanatory factor, which contrasts to the findings in the faecal eukaryome (Table 4.11). Differential abundance analysis indicated that Basidiomycota ASVs were significantly more abundant in parasitized mice compared to unparasitized mice (Figure 4.19). Variation in PC1 significantly accounted for 5.5% and 6.8% of the variation in BC in the faecal and caecal eukaryome, respectively (Table 4.11). However, the covariation in the factors contributing to PC1 – age, body condition, and faecal IgA concentration – means it is difficult to determine how they are associated with differences in eukaryome composition. There was a mild association between caecal inflammation and gut eukaryome composition, with *p* values approaching the significance threshold (Table 4.11).

Table 4.11. PERMANOVA test statistics for identifying which host factors contribute to variation in gut eukaryome community composition in wild mice. Variation in eukaryome composition was measured by Bray-Curtis dissimilarity. Significant host factors are highlighted in bold. Parasite infection is defined as infection with *Eimeria,* Oxyurida and/or *Trichuris muris*. Distances between samples were recalculated for each parasite model after removing ASVs assigned to that parasite.

	Host factor	**F statistic**	***R²***	***p* value**	**Adjusted *p* value**
Faecal	**PC1**	**3.27**	**0.055**	**< 0.001**	**0.003**
	Sex	0.61	0.011	0.851	0.851
	Reproductive status	1.07	0.019	0.310	0.487
	Faecal mucin concentration	0.67	0.012	0.805	0.851
	Small intestinal inflammation	1.01	0.090	0.438	0.602
	Caecal inflammation	1.60	0.058	0.036	0.066
	Large intestinal inflammation	0.92	0.035	0.570	0.697
	Parasite infection	**6.78**	**0.108**	**< 0.001**	**0.003**
	Eimeria	**6.05**	**0.098**	**< 0.001**	**0.003**
	Oxyurida	**9.09**	**0.140**	**< 0.001**	**0.003**
	Trichuris muris	1.98	0.034	0.032	0.066
Caecal	**PC1**	**4.11**	**0.068**	**< 0.001**	**0.003**
	Sex	0.39	0.007	0.954	0.954
	Reproductive status	1.00	0.017	0.363	0.499
	Faecal mucin concentration	0.80	0.015	0.575	0.703
	Small intestine inflammation	0.76	0.069	0.834	0.917
	Caecum inflammation	1.60	0.058	0.048	0.088
	Large intestine inflammation	1.32	0.050	0.149	0.234
	Parasite infection	**3.52**	**0.059**	**< 0.001**	**0.003**
	Eimeria	**6.65**	**0.106**	**< 0.001**	**0.003**
	Oxyurida	**7.88**	**0.123**	**< 0.001**	**0.003**
	Trichuris muris	**3.03**	**0.051**	**0.016**	**0.035**

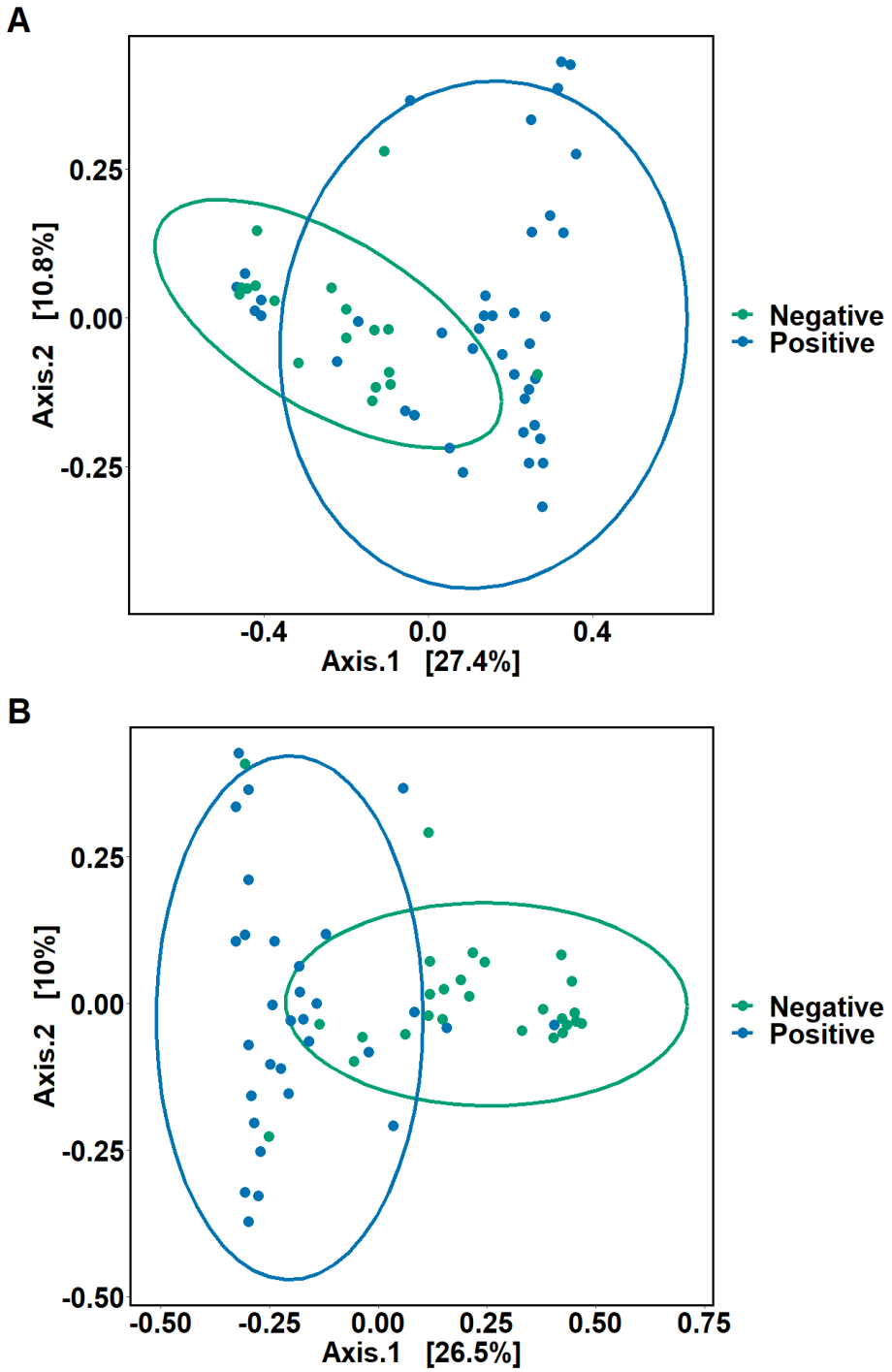


Figure 4.18. Bray-Curtis dissimilarity PCoA plots of the wild mouse faecal eukaryome by A) *Eimeria* and B) Oxyurida infection status. Ordination was generated using ASVs agglomerated at the species level, after removing ASVs assigned to *Eimeria* and Oxyurida, respectively. The proportion of variance explained by each axis is shown. Ellipses are 95% confidence intervals.

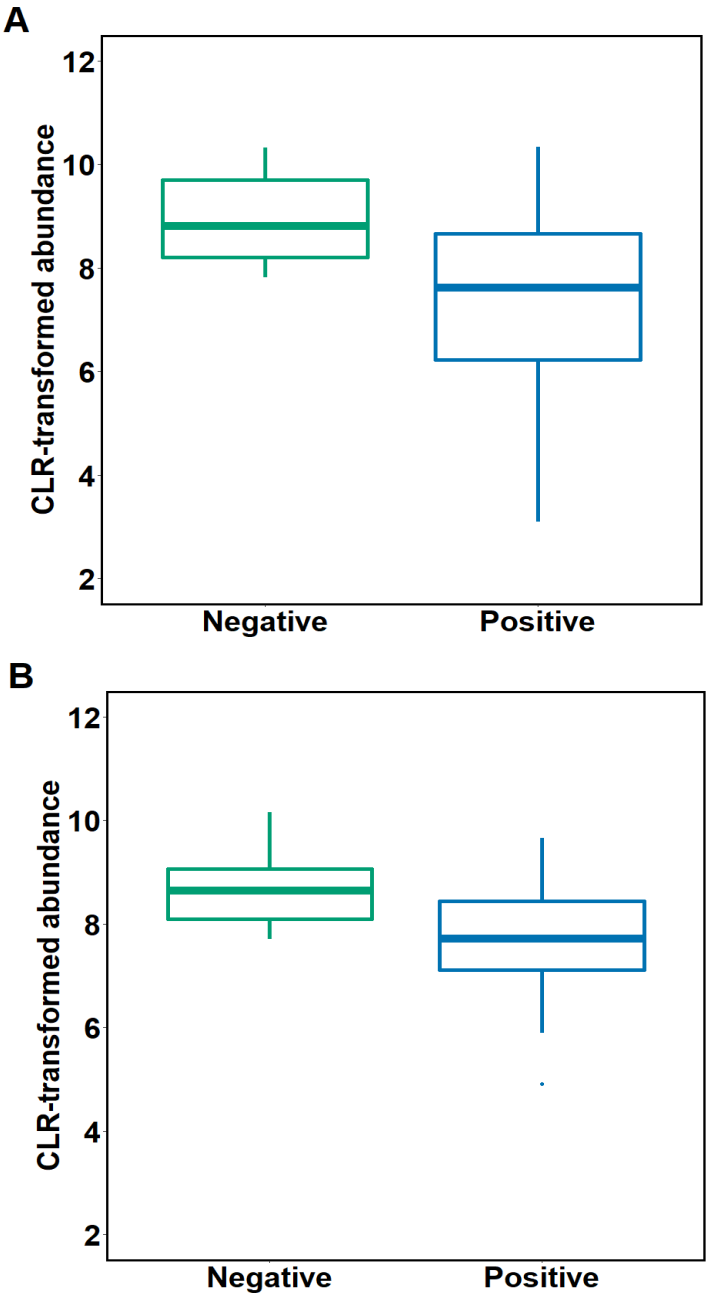


Figure 4.19. Basidiomycota ASV abundance in the A) faecal and B) caecal eukaryome of wild mice by parasite infection status. Parasite infection is defined as infection with any of *Eimeria,* Oxyurida, and *T. muris.* Abundance is based on the CLR transformation of read abundance. ANCOM-BC analyses were used to identify the phyla with a significantly different abundance. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range.

4.3.4 Comparison of the faecal eukaryome and bacteriome

4.3.4.1 The bacteriome composition

Having characterised the faecal eukaryome of wild mice, the faecal bacteriome was also characterised to identify similarities and differences between the two. For the 16S rRNA amplicon sequencing, the 31 ASVs identified in the positive control were correctly assigned to the eight bacterial species present in the positive control mock microbial community. A total of 3,064 ASVs were identified in the wild mice faecal samples (n = 58). The total read depth was 7,382,623, with an average read depth (± standard error) per sample of 127,287 (± 22,224). Following agglomeration to species level, removal of low abundance taxa, and removal of Archaea-classified ASVs, there were 714 bacterial ASVs, with a total read depth of 7,381,509. There was no correlation between the filtered read depths generated from 16S and 18S amplicon sequencing (Spearman's correlation: r = 0.081, *p* = 0.548).

18 bacterial phyla were identified across all 58 wild mice (Table 4.12). Firmicutes, Bacteroidota, Actinobacteriota, Campilobacterota, and Proteobacteria were found in 100% of mice, and Desulfobacterota in 98% of mice. Firmicutes and Bacteroidota had the highest relative abundance, accounting for 61.8% and 29.9% of sequences, respectively (Table 4.12, Figure 4.20A). To investigate which taxa were contributing to the phyla identified, comparisons were made at the family level (Figure 4.20B). Of the 150 families present in the faecal bacteriome, 20 were detected with more than 500 reads in least 20% of all mice. Lachonospiraceae and Lactobacillaceae were the dominant families contributing to the high relative abundance of Firmicutes, whereas Muribaculaceae and Bacteroidaceae had the largest contribution to the high relative abundance of Bacteroidota. The four other highly prevalent phyla were each represented by just one family (Figure 4.20).

Table 4.12. Phyla identified from the faecal bacteriome of wild mice. Phyla are ordered by their relative abundance across all 58 wild mice. Only the phyla with a relative abundance of >0.05% are shown. The complete list of phyla can be found in Supplementary Table 14. Abundance is the number of reads classified for each phylum. Prevalence is the percentage of mice from which that phylum was identified. N'ham = Nottingham, S'port = Southport.

Phylum	Abundance		Prevalence (%)			
	Relative (%)	Total	Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Firmicutes	61.8	4,561,505	100	100	100	100
Bacteroidota	29.9	2,205,342	100	100	100	100
Actinobacteriota	2.2	166,016	100	100	100	100
Campilobacterota	2.0	145,104	100	100	100	100
Desulfobacterota	1.9	138,905	98	100	100	92
Proteobacteria	1.5	112,720	100	100	100	100
Verrucomicrobiota	0.2	17,883	14	16	13	8
Deferribacterota	0.2	17,445	72	94	40	58
Cyanobacteria	0.1	4,898	55	74	40	25
Patescibacteria	0.1	4,043	55	77	20	42
Fusobacteriota	0.1	3,854	3	3	0	8

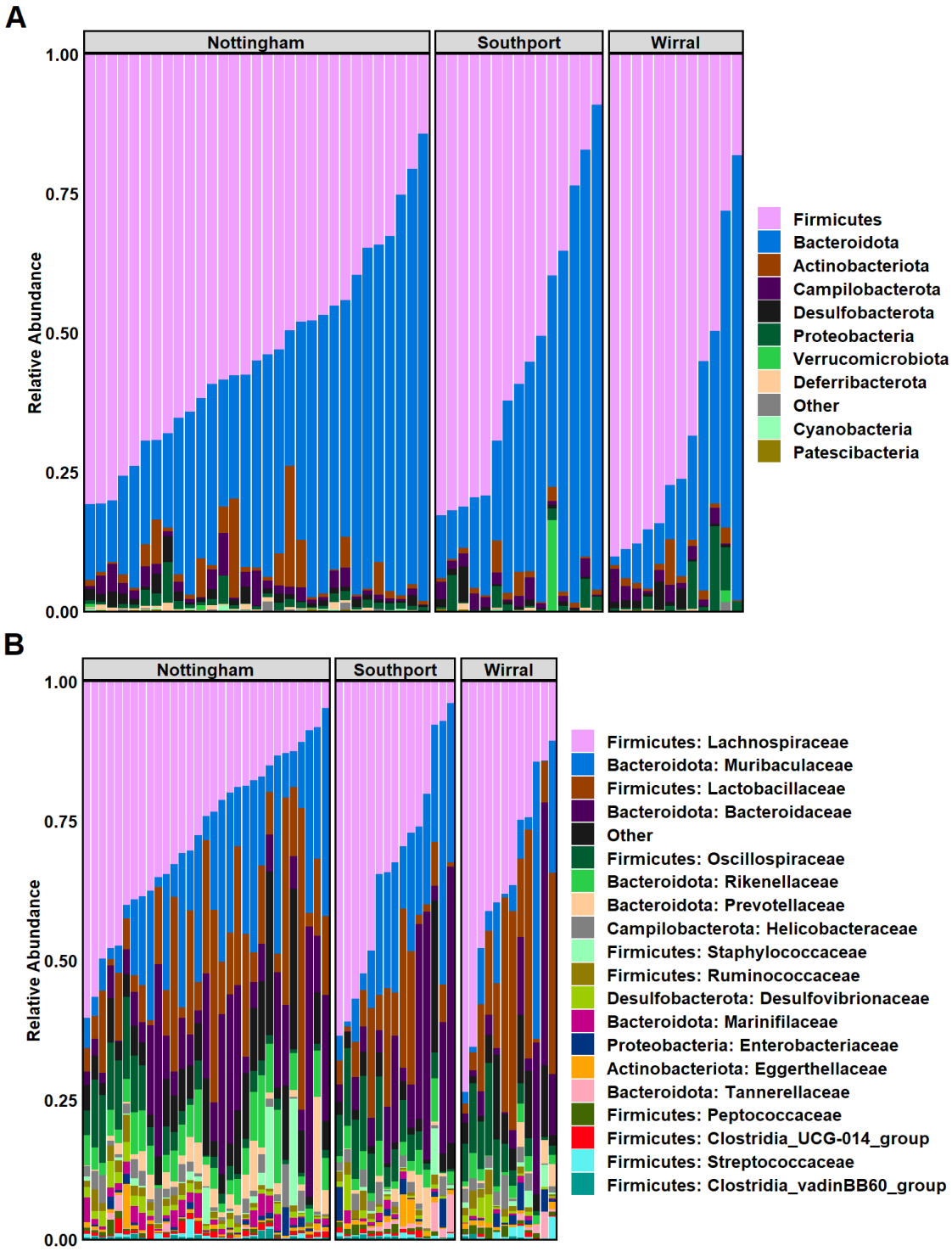


Figure 4.20. Relative abundances of bacterial A) phyla and B) families identified in the faecal bacteriome of wild mice. Relative abundance is the proportion of reads classified for each taxonomic rank. Only the taxa with a minimal read abundance of 200 in ≥ 5% of all mice (phyla) and 500 in ≥ 20% of mice (families) are shown. Less abundant taxa are grouped into Other. Taxa are ordered by their relative abundance across all mice.

4.3.4.2 Comparison of the faecal eukaryome and bacteriome diversity metrics

In contrast to the faecal eukaryome, there was a significant difference in faecal bacteriome ASV-level alpha diversity (mean Shannon's index ± SE) among mice from the three sampling sites (Kruskal-Wallis: $X^2_2 = 6.91$, $p = 0.032$): Nottingham (3.4 ± 0.1), Southport (3.1 ± 0.1), and Wirral (3.0 ± 0.1). Whilst post hoc comparisons showed no significance differences when correcting for multiple testing, *p* values were approaching the significance threshold when comparing Nottingham mice to Southport and Wirral mice (Table 4.13). Across all mice, the average faecal bacteriome alpha diversity (3.2 ± 0.1) was significantly higher than the average faecal eukaryome alpha diversity (2.2 ± 0.1) (Wilcox signed-rank test: V = 1642, $p < 0.001$). This difference was also seen when comparing the alpha diversities for each sampling site separately (Figure 4.21A). There was no correlation between the ASV-level alpha diversities of the eukaryome and bacteriome when considering all mice (Spearman's rank: r = -0.02, $p = 0.894$) and each sampling site separately (Figure 4.21B). These results indicate that the faecal bacteriome is more diverse, with greater variation among sampling sites, compared to the faecal eukaryome.

Table 4.13. Test statistics for comparing the Shannon's index of the wild mouse faecal bacteriome among sampling sites. Comparison among sampling sites used a Kruskal-Wallis test (test statistic: X^2). Post hoc comparisons between sampling sites used a Dunn's test with correction for multiple testing. N'ham = Nottingham, S'port = Southport.

	Shannon's Index (± SE)			Test		Post hoc comparisons	
Rank	**N'ham**	**S'port**	**Wirral**	X^2	***p* value**	**Comparisons**	***p* value**
ASV	3.5 (0.1)	3.1 (0.1)	3.0 (0.1)	6.91	0.032	N'ham: S'port	0.067
						N'ham: S'port	0.067
						S'port: Wirral	0.756

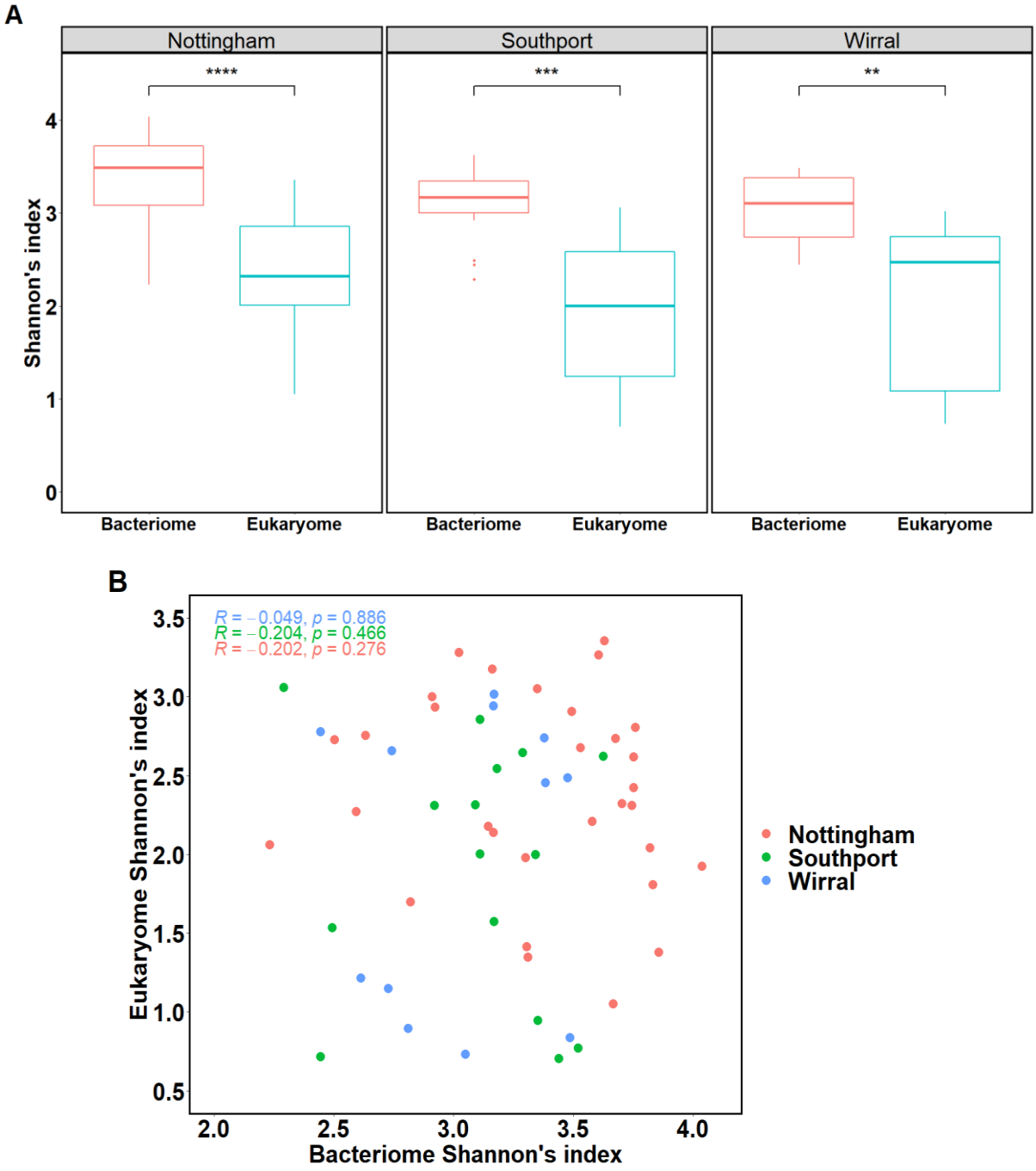


Figure 4.21. Comparison of alpha diversity in the faecal bacteriome and eukaryome of wild mice. A) Shannon's index of the bacteriome and eukaryome for each sampling site. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range. B) Correlation between the Shannon's index of the faecal eukaryome and bacteriome. Spearman's rank correlation coefficient (*R*), and the corresponding *p* value, are shown. Each point is an individual mouse, coloured by sampling site. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, from Wilcox signed-rank tests.

For beta diversity, sampling site accounted for 8.8% of the variation in ASV-level BC of the faecal bacteriome (PERMANOVA: $F_2 = 2.66$, $R^2 = 0.088$, $p < 0.001$, Figure 4.22A). This effect was smaller for the bacteriome compared to the eukaryome, for which sampling site explained 21.1% of the variation (see section 4.3.2.3 above). In support of the findings in the faecal eukaryome, pairwise comparisons of bacteriome BC showed that Nottingham mice had a significantly different bacteriome composition compared to both Southport ($p = 0.003$) and Wirral mice ($p = 0.009$), with no difference between mice from Southport and Wirral ($p = 0.333$). There was no difference in how varied the bacteriome community was within sampling sites. This contrasts to the findings for the eukaryome (see section 4.3.2.3 above). There was a significant correlation between BC in the eukaryome and bacteriome of mice (Mantel test: r = 0.130, $p = 0.024$). This indicates that as the eukaryome community composition becomes more dissimilar between two mice, so does their bacteriome community composition. However, considering each sampling site separately, only Wirral mice had a significant correlation between eukaryome and bacteriome BC (Mantel test: r = 0.296, $p = 0.013$, Figure 4.22B).

To account for there being no difference in the bacteriome community composition between Southport and Wirral mice, these mice were grouped together to compare the differential abundance of taxa against Nottingham mice. There were 4 bacterial species in faecal samples, compared to the 14 eukaryotic species, with a significantly different abundance when comparing Nottingham mice to Southport and Wirral mice (Figure 4.23, Supplementary Table 15). *Lactobacillus reuteri* and *Bacteroides vulgatus* were more abundant in Nottingham mice. Two *Helicobacter* spp. were differentially abundant: *H. apodemus* was more abundant in Nottingham mice whereas *H. mastomyrinus* was more abundant in Southport and Wirral mice.

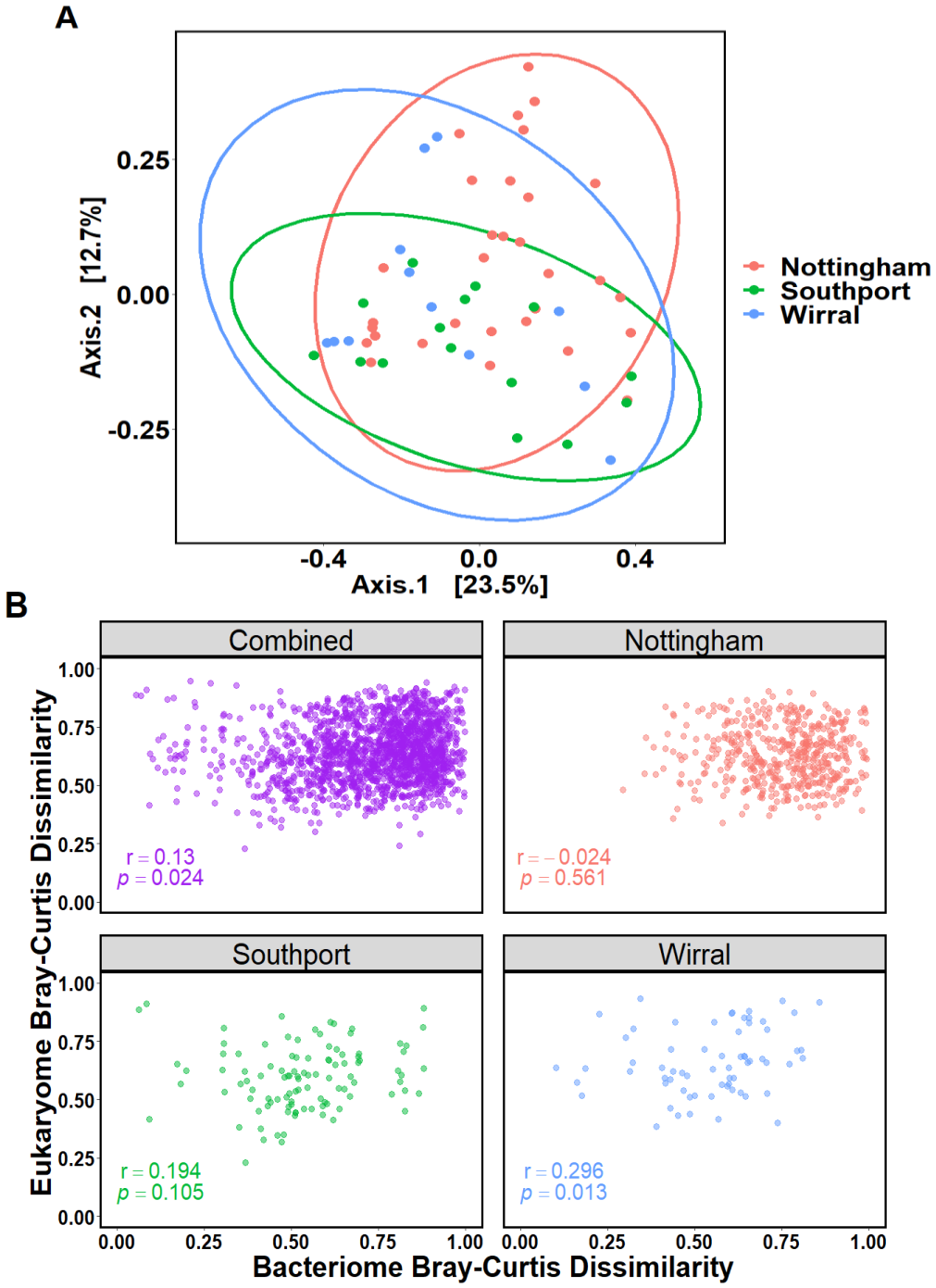


Figure 4.22. Comparison of faecal bacteriome and eukaryome Bray-Curtis dissimilarity among wild mice from three sampling sites. Dissimilarity matrices were generated using the ASV-level taxonomic rank. A) PCoA of faecal bacteriome Bray-Curtis dissimilarity by sampling site. The proportion of variance explained by each axis is shown. Ellipses are 95% confidence intervals. B) Correlation between bacteriome (x-axis) and eukaryome (y-axis) Bray-Curtis dissimilarity. Each point represents a pairwise comparison between two mice for the two microbiomes, where 0 indicates mice share all the same ASVs, and 1 is where no ASVs are shared.

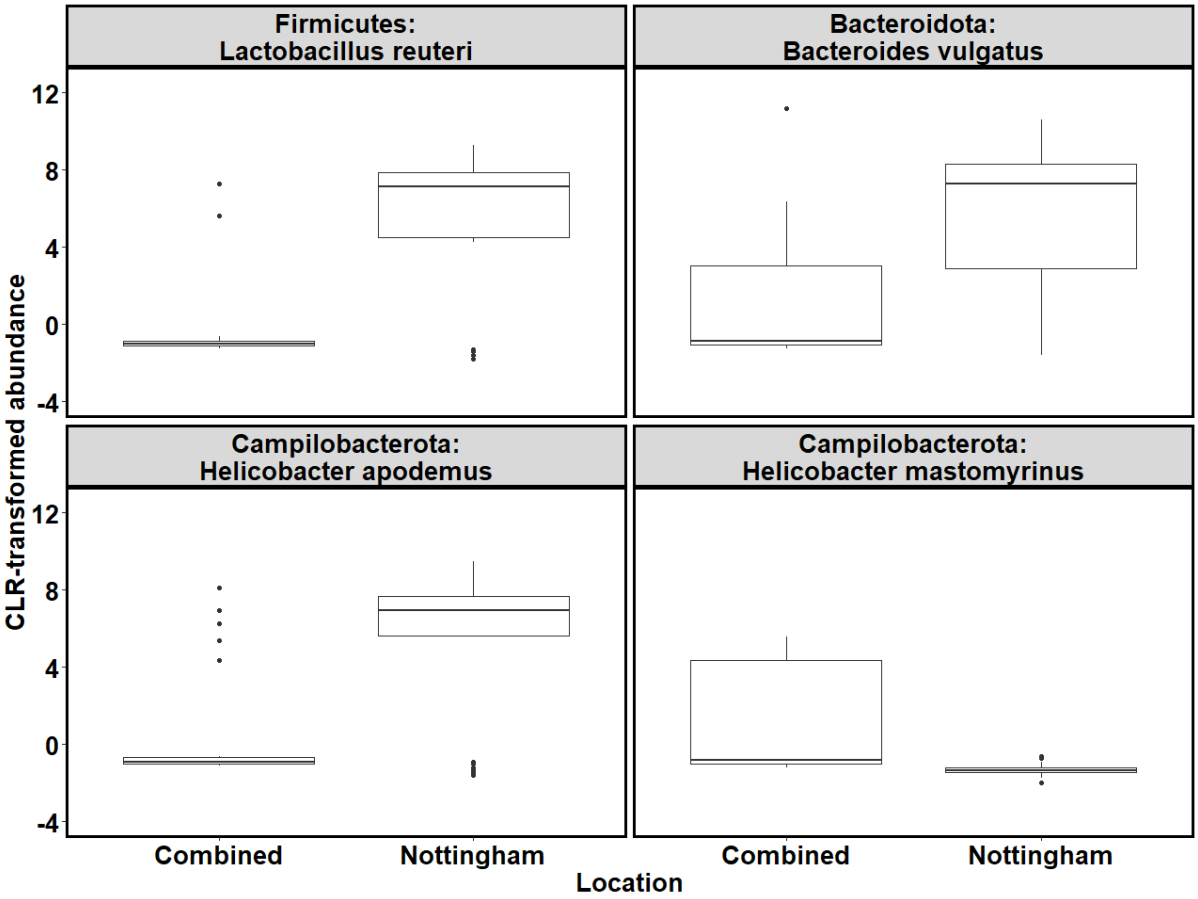


Figure 4.23. Bacterial species that were differentially abundant between the faecal samples of Nottingham mice and mice from the two other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance. ANCOM-BC analyses were used to identify the ASVs with a significantly different abundance. Abundance values can be seen in Supplementary Table 15. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range.

4.3.4.3 Drivers of bacteriome variation

To investigate if faecal bacteriome diversity and composition were affected by the same host factors as the eukaryome, the analysis reported in section 4.3.3 was repeated for the faecal bacteriome. For alpha diversity, all host factors were included in a single model to identify which factors significantly affected Shannon's index (Table 4.14). In contrast to the eukaryome, PC1 was the only significant explanatory factor of bacteriome alpha diversity (LM: $F_1 = 5.41$, $p = 0.026$, Figure 4.24). The Shannon's index of mice decreased as PC1 increased. All three host factors used in the PCA were negatively correlated with PC1 (Supplementary Figure 5E). Thus, as mice get older, have a better condition, and have a greater faecal IgA concentration, alpha diversity in the faecal bacteriome increases.

Table 4.14. Statistics for the model testing for the effects of host factors on faecal bacteriome alpha diversity in wild mice. Shannon's index was used as the measure of alpha diversity. Significant host factors are highlighted in bold. The full model statistics were: $F_{15:34} = 1.06$, $p = 0.422$.

Host factor	F value	*p* value
PC1	**5.40**	**0.026**
Sex	0.02	0.878
Reproductive status	4.06	0.052
Faecal mucin concentration	1.19	0.284
Small intestine inflammation	0.81	0.548
Caecum inflammation	0.77	0.470
Large intestine inflammation	0.15	0.858
Eimeria	0.83	0.369
Oxyurida	0.14	0.706

For beta diversity, infection with *Eimeria,* Oxyurida, and/or *Trichuris* was a significant explanatory factor of bacteriome community composition, accounting for 4.8% of the variation in BC (PERMANOVA: $F_1 = 2.84$, $R^2 = 0.048$, $p = 0.026$, Figure 4.25A). This effect was smaller for the bacteriome compared to the eukaryome, for which parasitism explained 10.8% of the variation (see section 4.3.3 above). Both *Eimeria* and Oxyurida were significant explanatory factors of bacteriome community composition when tested independently, accounting for 4.4% and 5.1% of the variation in BC in faecal samples, respectively (Table 4.15). Differential abundance analysis indicated that two bacterial species – *L. reuteri* and *H. apodemus* – were significantly more abundant in parasitized mice compared to unparasitized mice (Figure 4.25B). In contrast to the faecal eukaryome, PC1 was not a significant explanatory factor of bacteriome community composition (Table 4.15).

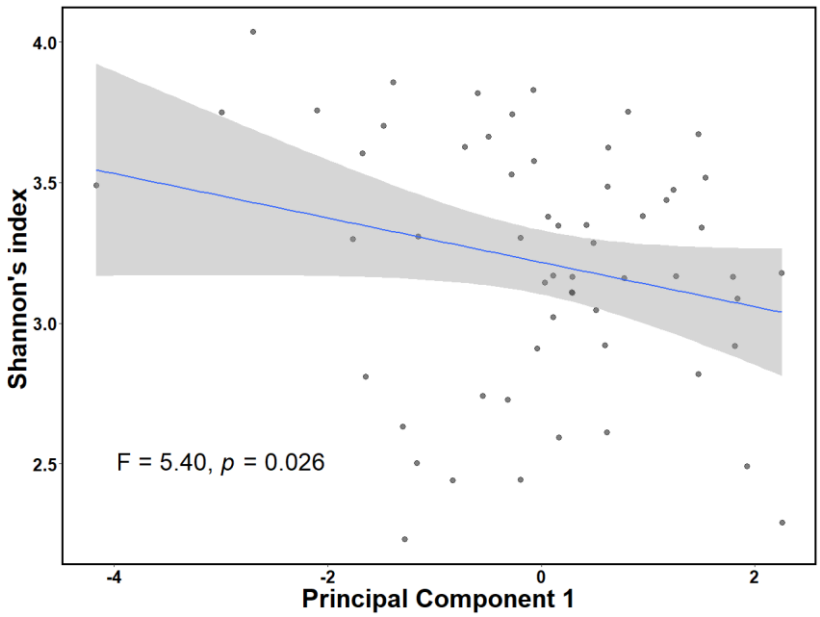


Figure 4.24. The association between the faecal bacteriome alpha diversity and PC1 in wild mice. Shannon's index is the measure of alpha diversity. PC1 was generated from a principal component analysis of three correlated host factors: faecal IgA concentration, age, and body condition. All three factors were negatively correlated with PC1. The blue line is the predicted relationship between PC1 and Shannon's index, with grey shading representing the 95% confidence intervals. The F statistic and *p* value shown are extracted from a linear model which incorporated all host factors.

Table 4.15. PERMANOVA test statistics for identifying which host factors contribute to variation in faecal bacteriome community composition in wild mice. Variation in bacteriome composition was measured by Bray-Curtis dissimilarity. Significant host factors are highlighted in bold. Parasite infection is defined as infection with *Eimeria,* Oxyurida and/or *Trichuris muris*.

Host factor	**F statistic**	**R^2**	***p* value**	**Adjusted *p* value**
PC1	1.50	0.026	0.124	0.195
Sex	0.78	0.014	0.665	0.665
Reproductive status	1.54	0.027	0.113	0.195
Faecal mucin concentration	1.22	0.022	0.218	0.288
Small intestinal inflammation	1.23	0.108	0.118	0.195
Caecal inflammation	1.16	0.043	0.236	0.288
Large intestinal inflammation	1.14	0.044	0.263	0.289
Parasite infection	**2.84**	**0.048**	**0.005**	**0.026**
Eimeria	**2.55**	**0.044**	**0.007**	**0.026**
Oxyurida	**3.00**	**0.051**	**0.002**	**0.022**
Trichuris muris	1.80	0.031	0.047	0.129

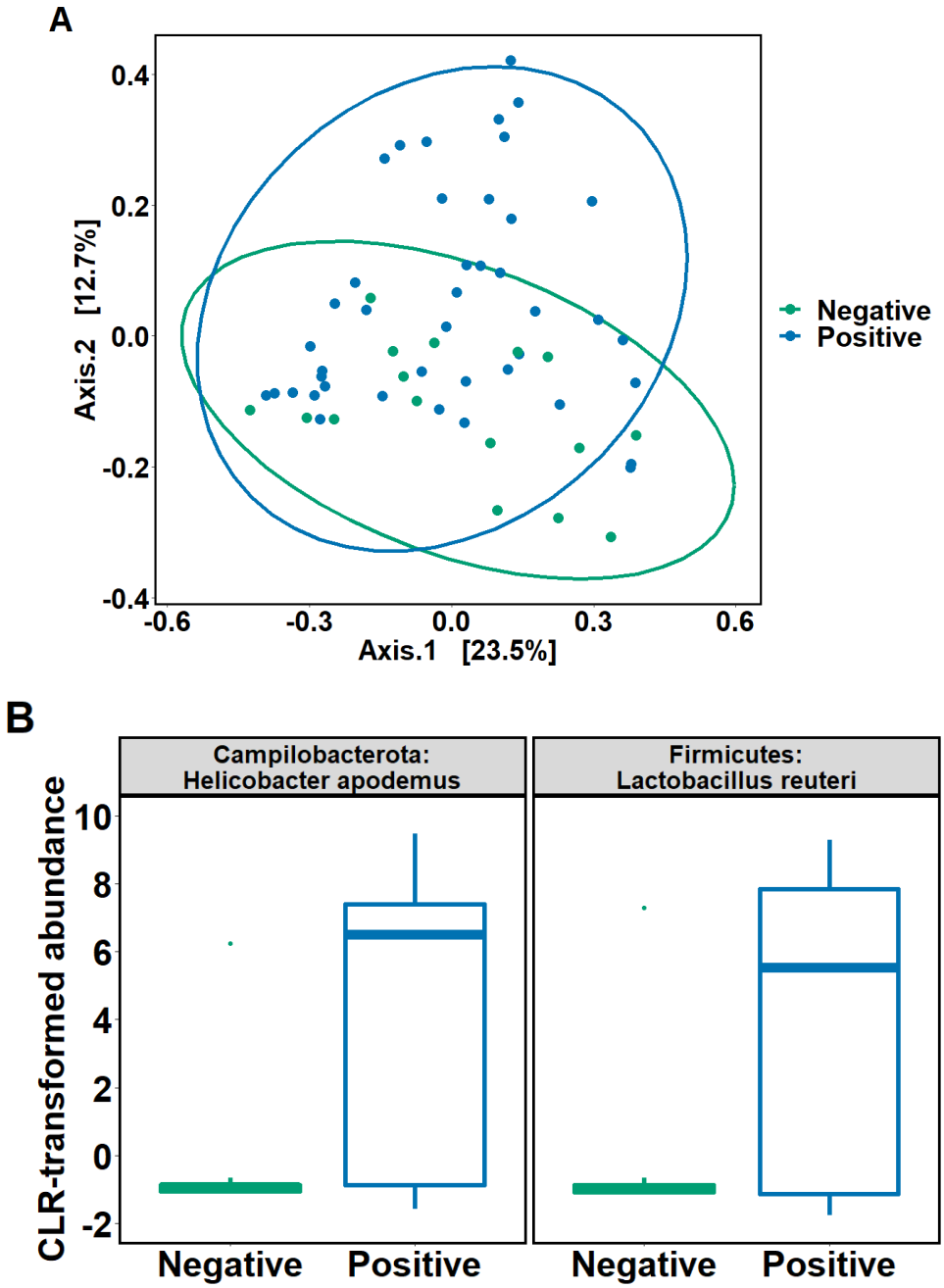


Figure 4.25. Wild mouse faecal bacteriome community composition differences by parasite infection status. Parasite infection is defined as infection with any of *Eimeria,* Oxyurida, and *T. muris.* A) Bray-Curtis dissimilarity PCoA plot**.** Ordination was generated using ASVs agglomerated at the species level. The proportion of variance explained by each axis is shown. Ellipses are 95% confidence intervals. B) Differential abundance of two bacterial species. Abundance is based on the CLR transformation of read abundance. ANCOM-BC analyses were used to identify species with a significantly different abundance. Boxplots indicate the median value, the first and third quartiles and whiskers extend to 1.5 times the inter-quartile range.

4.3.4.4 Co-occurrence of taxa in the faecal microbiome

Having characterised both the faecal eukaryome and bacteriome, associations among microbial taxa were predicted using network analysis (Figure 4.26). Three measures of network connectivity were compared among the sampling sites: node degree, percentage of singletons, and component size. Comparison of the eukaryome network among sampling sites identified differences in the node degree (average number of connected edges per node ± SE) (Table 4.16). For Nottingham and Southport mice, average node degree was low (0.21 ± 0.05 and 0.68 ± 0.08, respectively). For Wirral mice, the average node degree was 1.09 (± 0.09), indicating than on average, nodes were connected to one other node. This pattern is supported when looking at the percentage of singleton nodes (nodes without connected edges): only 27.8% of nodes were not connected in Wirral mice, compared to 82.8% and 50.5% in Nottingham and Southport mice, respectively. These differences are driven by the presence of a large component (a group of connected nodes) in Wirral mice (15 nodes), whereas the largest component in Nottingham and Southport mice was comprised of four and eight nodes, respectively. These results contrasted to that of the bacteriome, where Southport mice had a greater node degree and lower proportion of singleton nodes compared to Nottingham and Wirral mice (Table 4.16). Bacteriome network connectivity was greater than eukaryome network connectivity in Nottingham and Southport mice, as measured by node degree and percentage of singletons. In contrast, eukaryome connectivity was higher than bacteriome connectivity in Wirral mice.

After considering the eukaryome and bacteriome separately, the two taxa groups were combined together for microbiome analysis. Nottingham and Southport mice had a higher node degree (1.74 ± 0.13 and 2.49 ± 0.11, respectively) compared to their respective eukaryome and bacteriome networks (Table 4.16). Additionally, the percentage of singleton nodes was low for both Nottingham and Southport mice (28.2% and 3.1%), driven by the presence of many interacting nodes in the largest components (94 and 140 nodes). This contrasts to Wirral mice, which had a low percentage of singleton nodes (26.7%), but with multiple, smaller components.

Table 4.16. Network connectivity in the faecal microbiome of wild mice from the three sampling sites. The number of nodes is the number of taxa used to generate the network. Node degree is the average number of edges connected to that node ± standard error. The number of singleton nodes (no connected edges) is expressed as a percentage to account for differences in the number of nodes. The number of components (groups of connected nodes) is shown, as well as the number of nodes in the largest component. Taxa were analysed separately for the eukaryome and bacteriome and then combined for the microbiome. Taxa were aggregated at the family taxonomic rank for analyses and networks were predicted using the SpiecEasi pipeline. Only families presented in ≥ 20% of mice for a given sampling site were included in the analysis. N'ham = Nottingham, S'port = Southport.

Sampling site	Type	Number of nodes	Node degree (± SE)	Singletons (%)	Components	
					Total	Largest
N'ham	Eukaryome	87	0.21 (± 0.05)	82.8	78	4
	Bacteriome	62	0.54 (± 0.11)	66.1	47	6
	Microbiome	149	1.74 (± 0.13)	28.2	49	94
S'port	Eukaryome	109	0.68 (± 0.08)	50.5	74	8
	Bacteriome	51	0.78 (± 0.12)	45.1	31	11
	Microbiome	160	2.49 (± 0.11)	3.1	12	140
Wirral	Eukaryome	97	1.09 (± 0.09)	27.8	45	15
	Bacteriome	53	0.23 (± 0.07)	81.1	47	4
	Microbiome	150	1.05 (± 0.07)	26.7	71	9

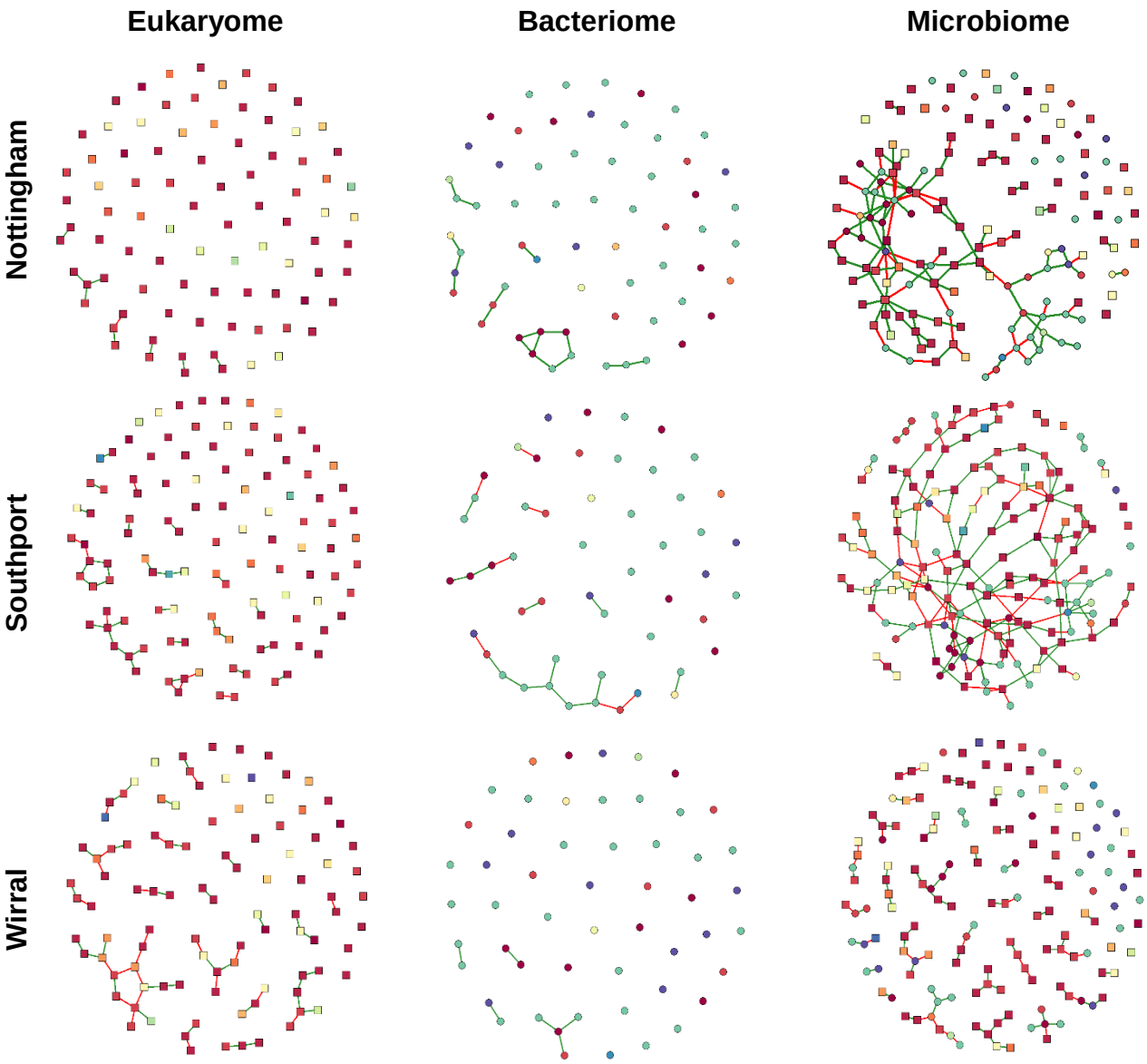


Figure 4.26. Microbial networks of the faecal microbiome of wild mice from three sampling sites. Each node represents a eukaryotic (square) or bacterial (circle) taxon, and are coloured by phylum. Edges represents associations between taxa, coloured by positive (green) and negative (red) associations. Networks were predicted using the SpiecEasi pipeline. Only families present in ≥ 20% of mice for a given sampling site were included in the analysis. Networks were generated for the eukaryome and bacteriome separately, and then together (microbiome), for each of the three sampling sites.

To explore the eukaryome and bacteriome networks further, the taxonomic assignment of nodes, and the number of predicted associations per taxa, was identified (Figure 4.26). In the eukaryome of Nottingham mice, all edges were between fungi (Ascomycota or Basidiomycota), and only two nodes were connected with more than one other node. The family Extremaceae (Ascomycota) had positive edges with Herpotrichiellaceae, Capnodiales (Ascomycota), and Cystobasidiomycetes (Basidiomycota). The family Mrakiaceae (Basidiomycota) had a positive edge with another basidiomycete family (Sporidiobolaceae) but a negative edge with Chaetomiaceae (Ascomycota). In the eukaryome of Southport and Wirral mice, all connected nodes belonged to ten and twelve phyla respectively, of which Ascomycota and Basidiomycota nodes were the most common. The most connected node in Southport mice was Cephalothecaceae (Ascomycota), which had four edges with other Ascomycota and Basidiomycota fungi. The most connected node in Wirral mice was Wallemiaceae (Basidiomycota), connected to four other nodes: two other Basidiomycota, one Mucuromycota, and one Protosteliida. Two Apicomplexan taxa, Cryptosporida and Eugregarinorida, were connected to other nodes in the eukaryome networks of Southport and Wirral mice. Eugregarinorida had a positive edge with Cunninghamellaceae (Mucuromycota) in Wirral mice and two positive edges with fungi in Southport mice. Cryptosporida had a positive edge with Chromulinales (Ochrophyta) in Wirral mice. Oxyurida nematodes had a positive edge with Pucciniaceae (Basidiomycota) in Wirral mice.

For the bacteriome, Nottingham and Southport mice had connected nodes from seven phyla, compared with only four phyla in Wirral mice. In Nottingham mice, five of six multi-node components had nodes from multiple phyla. In the largest component, the families Corynebacteriaceae and Dermabacteraceae (Actinobacteriota) each had positive edges with three other bacterial families. In Southport mice, seven of eight multi-node components had nodes from multiple phyla. The largest component contained 11 nodes, with Oscillospiraceae and Ruminococcaceae (Firmicutes) having the highest node degree. In the bacteriome of Wirral mice, the largest component was comprised of four nodes, of which Eggerthellaceae (Actinobacteriota) was central to the three other families.

In Nottingham and Southport mice, the microbiome network identified more connected phyla compared to the eukaryome and bacteriome networks. In particular, ten more eukaryotic phyla were identified in Nottingham mice as connecting to other phyla, including Apicomplexa and Nematoda. Apicomplexa had positive edges with other eukaryotes (Dothideomycetes (Ascomycota) and Glissomonadida (Cercozoa) and a negative edge with the bacterial family Aerococcaceae (Firmicutes). In contrast to the findings in the Wirral eukaryome, Oxyurida in the Nottingham microbiome had positive edges with the fungi Cladosporiaceae (Ascomycota) and Ustilaginaceae (Basidiomycota). The large component identified in the microbiome of

Nottingham mice comprised both bacterial and eukaryotic taxa: the most connected taxa were bacterial (Aerococcaceae, Corynebacteriaceae, and Moraxellaceae). In Southport mice, the microbiome network additionally identified the phylum Amoebozoa, which had a positive edge with Cryptomycota fungi. The most connected taxa in the microbiome network of Southport mice were fungi (Cordycipitaceae and Sporidiobolaceae) and bacteria (Corynebacteriaceae), which contrasted to Nottingham mice. In Wirral mice, the largest microbiome component consisted of only nine nodes, and the most connected taxon did not change from that seen in the bacteriome network (Eggerthellaceae), which contrasts to Nottingham and Southport mice.

There were fewer inter-kingdom edges (19%, 24% and 25%) than intra-kingdom edges in the microbiome networks (Table 4.17). It is notable that there were more negative than positive inter-kingdom edges, which contrasts to the findings for intra-kingdom edges, as seen by both the separate and combined network analyses (Table 4.17). In sum, the above comparisons indicate that incorporating both eukaryotic and bacterial taxa into network analyses can change network connectivity and predict more microbial associations.

Table 4.17. The proportion of positive and negative edges in microbial networks of the faecal microbiome of wild mice from the three sampling sites. Edges are representative of predictive associations between taxa. For the single-kingdom networks (eukaryome and bacteriome), only intra-kingdom edges can be predicted. For the multi-kingdom network (microbiome), both inter- and intra-kingdom edge can be predicted. The percentage of positive and negative edges, and the total number of edges identified, is given for the three possible kingdom interactions: eukaryote-eukaryote, bacteria-bacteria, and eukaryote-bacteria. Taxa were aggregated at the family taxonomic rank for analyses and networks were predicted using the SpiecEasi pipeline. Only families presented in ≥ 20% of mice for a given sampling site were included in the analysis. N'ham = Nottingham, S'port = Southport.

Sampling site	Kingdom interactions	Single-kingdom networks			Multi-kingdom networks		
		Positive	Negative	Total	Positive	Negative	Total
N'ham	Eukaryote-eukaryote	89% (8)	11% (1)	9	35% (46)	4% (5)	51 (39%)
	Bacteria-bacteria	100% (17)	0% (0)	17	37% (48)	5% (6)	54 (42%)
	Eukaryote-bacteria				0% (0)	19% (25)	25 (19%)
S'port	Eukaryote-eukaryote	78% (29)	22% (8)	37	49% (97)	9% (18)	115 (58%)
	Bacteria-bacteria	65% (13)	35% (7)	20	18% (35)	1% (1)	36 (18%)
	Eukaryote-bacteria				2% (4)	22% (44)	48 (24%)
Wirral	Eukaryote-eukaryote	58% (31)	42% (22)	53	43% (34)	11% (9)	43 (54%)
	Bacteria-bacteria	100% (6)	0% (0)	6	19% (15)	1% (1)	16 (20%)
	Eukaryote-bacteria				6% (5)	19% (15)	20 (25%)

Correlation analyses of ASV abundance, grouped at the phylum level, indicated that 36 phyla were significantly correlated in the faecal microbiome (Figure 4.27). Only three significant negative correlations were identified: Firmicutes and Bacteroidota; Basidiomycota and Apicomplexa; and Basidiomycota and Nematoda. The majority of inter-kingdom correlations identified were negative, but not significant (Figure 4.27). However, two significant positive inter-kingdom correlations were identified: Apicomplexa and Cyanobacteria and Nematoda and Cyanobacteria. Together, network analyses and correlations of abundance suggest that inter-kingdom associations may be typically more antagonistic than positive.

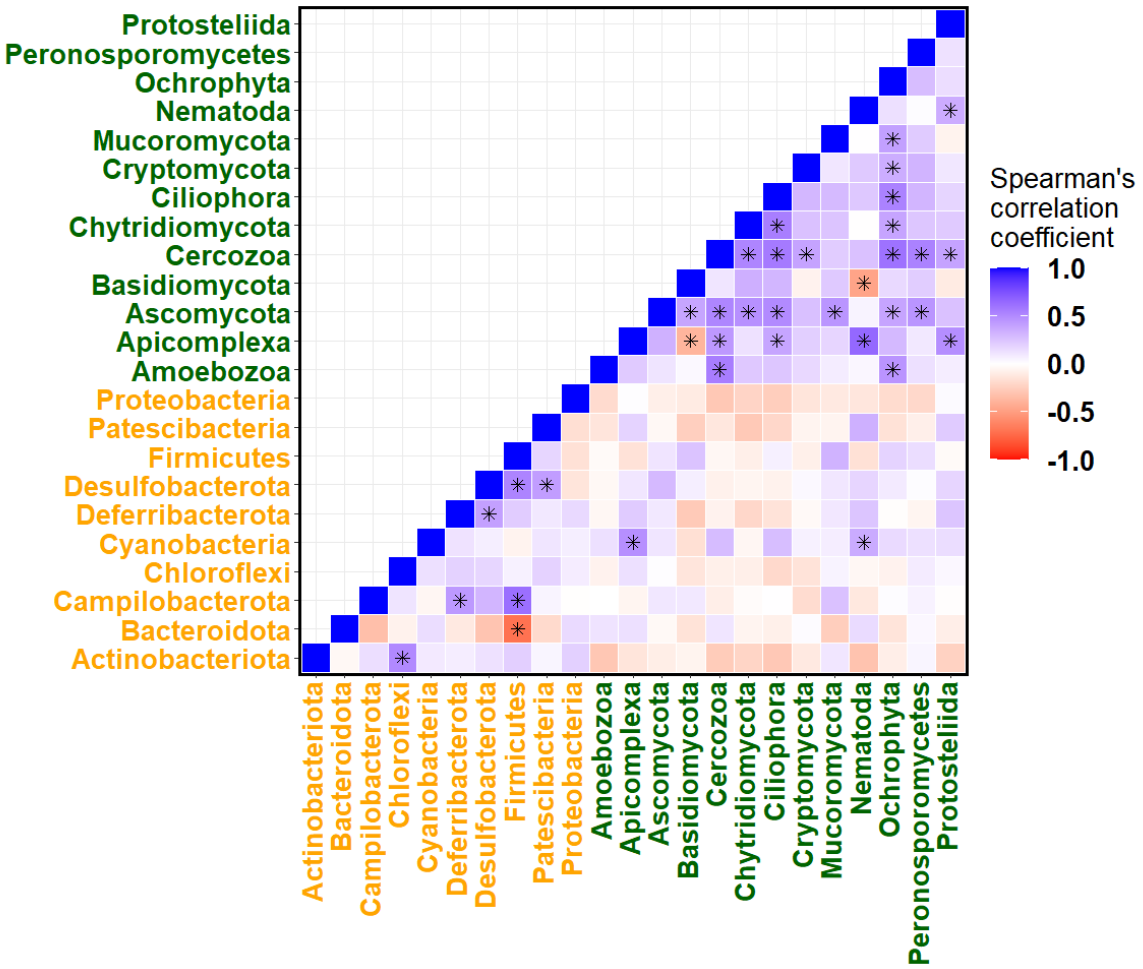


Figure 4.27. Correlations of phyla abundance in the faecal microbiome. Abundance is defined as the number of ASVs assigned to that phylum. Spearman's correlation coefficient was used to identify correlation and significant correlations are indicated by *. Bacterial phyla are shown in orange, and eukaryotic phyla in green. Only phyla present in ≥ 20% of mice are shown.

4.4 Discussion

4.4.1 Characterisation of the gut eukaryome of wild house mice

The study presented here aimed to characterise the gut eukaryome of wild house mice and identify which host factors are associated with eukaryome diversity and composition. 18S rRNA amplicon sequencing identified fungi, protozoa, and helminths in the gut eukaryome, as seen with other wild mammals (Heitlinger *et al.*, 2017; Li *et al.*, 2018; Mann *et al.*, 2020; Kim *et al.*, 2022; Murillo *et al.*, 2022). First, to address the relative abundances of these three taxa groups. In the study presented here, the relative abundance of fungi was greater than helminths and protozoa. This is consistent with findings for bats and hyenas (Heitlinger *et al.*, 2017; Li *et al.*, 2018) but contrasts to findings in another wild rodent species, the striped field mouse, *Apodemus agrarius* (Kim *et al.*, 2022). The comparative difference between the higher relative abundance of fungal taxa in the study presented here, and the higher relative abundance of protozoa in *A. agrarius* is consistent with previous findings that gut eukaryome composition is species-specific (Parfrey *et al.*, 2014; Mann *et al.*, 2020). For the study presented here, I used relative abundance to compare the taxa in the wild house mouse eukaryome, but there are caveats to using relative abundance (reviewed in detail in Chapter 3, section 3.1.2). Specifically, relative abundance estimations of taxa from 18S rRNA amplicon sequencing are biased by both gene copy number variation and taxa-specific amplification biases (Gogarten *et al.*, 2020; Lavrinienko *et al.*, 2021b). For this study, I used 18S primers which were designed against the fungal 18S V4 region, potentially explaining the higher relative abundance of fungal taxa, despite protozoa and helminth sequences also being identified.

Ascomycete fungi with a high relative abundance included *Debaryomcyes, Aspergillus, Cladosporidium, Chaetomium,* and *Microascus* (Ascomycota). These genera were not found in one of the earliest mycobiome studies in laboratory mice (Scupham *et al.*, 2006) but have since been identified in other studies (Dollive *et al.*, 2013; Heisel *et al.*, 2017). The identification of these fungal genera in both laboratory and wild house mice suggests that they are common in the gut eukaryome. In other wild mice, specifically *A. agrarius* and *M. musculus,* the genus *Kazachstania* was highly prevalent and abundant, in contrast to my study (Bendová *et al.*, 2020; Kim *et al.*, 2022). This suggests that fungal taxa of the gut eukaryome can be variable among and within rodent species. Of particular note, the prevalence of *Malassezia* was considerably higher in the caecum than in faeces, suggesting this genus may be true resident of the house mouse gut eukaryome.

The prevalence and relative abundance of *Candida* spp. was low, consistent with findings in other studies of wild mice (Bendová *et al.*, 2020; Kim *et al.*, 2022), but contrasting to findings in laboratory mice where *Candida* spp. are abundant and prevalent (Iliev *et al.*, 2012; Dollive

et al., 2013; Heisel *et al.*, 2017). Thus, *Candida* spp. being dominant in laboratory mice, but not in wild mice, may explain the increased fungal diversity and abundance seen in free-living mice compared to laboratory mice (Rosshart *et al.*, 2019; Yeung *et al.*, 2020). However, it should be noted that many *Candida* spp. have recently been re-classified into other genera (Kidd *et al.*, 2023). Thus, the comparatively low prevalence and abundance of *Candida* spp. in the present study may be confounded by fewer ASVs being classified as *Candida,* following changes to taxonomy in the SILVA database (del Campo *et al.*, 2018). The frequent reclassification of fungi, and other microbial eukaryotes, is recognised as a limiting factor in our understanding of the mammalian gut eukaryome (del Campo *et al.*, 2018).

The high prevalence and relative abundance of *Wallemia* spp. (Basidiomycota) identified here was unexpected as this taxon is hypothesised to be incapable of colonising the mammalian gut (Hallen-Adams and Suhr, 2017). However, *Wallemia* has also been reported from laboratory mice, with colonisation potentially dependent on dietary exposure (Paterson *et al.*, 2017; Mims *et al.*, 2021). Furthermore, it's abundance in the gut can increase following antibiotic and antifungal treatment, with expansion in the gut exacerbating host disease (Wheeler *et al.*, 2016; Skalski *et al.*, 2018). Thus, the high prevalence and abundance of *Wallemia* in wild house mice suggests it has potential to have important interactions with other taxa in the gut microbiome, even if it is not a true gut resident (Hallen-Adams and Suhr, 2017; Jackson *et al.*, 2022). The fungal genus *Mucor* (Mucuromycota) was also highly prevalent in the gut eukaryome in the present study. This is consistent with reports from humans and wild woodrats where *Mucor* was highly prevalent (Mar Rodríguez *et al.*, 2015; Weinstein *et al.*, 2022), but contrasts to studies in wild *A. agrarius,* where *Mucor* was found at a very low prevalence (Kim *et al.*, 2022). Given than colonisation of *Mucor* spp. in the gut can alter the gut bacteriome, and the ubiquity of *Mucor* spp. in the environment, its high prevalence in wild house mice warrants further investigation (Mueller *et al.*, 2019; Weinstein *et al.*, 2022).

The helminth taxa identified in the mice included *Strongyloides, Trichuris,* Oxyurida, and Ascaridida. For *Trichuris,* Oxyurida, and Ascaridida, prevalence was greater in the caecal eukaryome compared to the faecal eukaryome. This is likely to due to faecal detection being dependent on egg production rates, which can vary according to the host's immune state as well as the time of collection (Davey *et al.*, 2021). Both pinworm (Oxyurida) and whipworm (*Trichuris*) are commonly found in the mammalian gut, and have been identified previously in the mammalian gut eukaryome using 18S rRNA sequencing, and so their finding here is not unexpected (Heitlinger *et al.*, 2017; Mann *et al.*, 2020; Kim *et al.*, 2022). Whilst *Strongyloides* spp. are common nematode parasites of vertebrates, infection in wild mice was previously unreported (Viney and Kikuchi, 2017), and so finding evidence of *Strongyloides* DNA in these mice was surprising. However, two *Strongyloides* spp. are known parasites of brown rats

(*Rattus norvegicus*) (Viney and Kikuchi, 2017) and so it is possible that this was a transient infection, transmitted from rats observed at the sampling site. Indeed, *Strongyloides* was only identified in one of 58 mice. Alternatively, this finding could be from incorrect taxonomic assignment of the ASVs (Tanaka *et al.*, 2014) or from laboratory contamination of the samples

18S rRNA sequencing identified protozoa taxa known to colonise the mammalian gut, including *Blastocystis, Balantidium, Eimeria,* and *Cryptosporidium* (Parfrey *et al.*, 2014; Heitlinger *et al.*, 2017; Mann *et al.*, 2020; Kim *et al.*, 2022; Murillo *et al.*, 2022). In wild *A. agrarius,* the protozoa genera *Tritrichomonas, Monocercomonas,* and *Giardia* (Metamonada) were the most prevalent, but these were not observed in the present study (Kim *et al.*, 2022). Furthermore, the prevalent gut protozoa identified in wild primates were also not identified in this present study (Mann *et al.*, 2020; Murillo *et al.*, 2022). These protozoa taxa included *Entamoeba, Endolimax,* and *Iodamoeba* (Amoebozoa)*,* Litostomatea (Ciliophora), and Trichomonadida (Metamonada). The differences between wild house mice, *A. agrarius,* and other wild mammals in which are the prevalent protozoa is consistent with previous reports that the protozoa found in the mammalian gut eukaryome are highly host-species-specific (Parfrey *et al.*, 2014; Mann *et al.*, 2020). A caveat to this finding in the present study is the 18S primers used here may be a poor match for some protozoa taxa. This is evidenced by *Eimeria* in the present study: there was a lower prevalence of *Eimeria* spp. when using 18S rRNA sequencing, compared to PCR assays which used primers designed against the *Eimeria* apicoplast (Jarquín-Díaz *et al.*, 2019). This suggests that using primers that may be a poor match for protozoa, and other eukaryotes, can lead to false negatives, lower prevalences of protozoa, and lower estimates of eukaryome diversity (Parfrey *et al.*, 2014; Kounosu *et al.*, 2019; Vaulot *et al.*, 2022). Future studies would benefit from identifying primers that maximise matches for eukaryotic taxa, or using multiple primer pairs and diagnostic techniques, to avoid this issue (Carta and Li, 2018; Kounosu *et al.*, 2019; Lokmer *et al.*, 2019; del Campo *et al.*, 2020; Jarquín-Díaz *et al.*, 2022; Vaulot *et al.*, 2022).

The eukaryome diversity and composition differed between caecal and faecal samples, consistent with a previous study comparing the caecal and faecal mycobiome composition of wild voles (Antwis *et al.*, 2021). This contrasts to the bacteriome, where other studies have shown diversity and community composition was consistent between faecal and caecal samples (Weldon *et al.*, 2015; Suzuki and Nachman, 2016). The difference in eukaryome composition between the two sample types in the present study suggests that the caecum and colon may be colonised by different eukaryotic taxa. Differences in the gut environment, such as pH, oxygen concentration, mucus production, and gut transit time, may be selecting for the survival and colonisation of different eukaryotic taxa (Donaldson *et al.*, 2016). However, a comparison of the ileum, caecum, and colon of wild house found the mycobiome community

composition was homogenous along the digestive tract (Bendová *et al.*, 2020). To investigate why the eukaryome composition was comparatively different between sample types in the present study, the survival of gut eukaryotes could be tested under the different conditions seen in the different regions on the gut. Differences in eukaryome composition between faecal and caecal samples could also result from post-depositional changes to the faecal eukaryome, such as the growth of fast-growing moulds, which would not be seen for the caecal eukaryome (Tedersoo *et al.*, 2022).

4.4.2 Drivers of gut eukaryome composition in wild house mice

The study presented here found an effect of sampling site on gut eukaryome composition in wild house mice. This is consistent with findings in another rodent species, which found the sampling site of wild bank voles had a significant effect on gut mycobiome composition (Antwis *et al.*, 2021). Studies in humans and wild primates have also shown that differences in the host's environment is associated with mycobiome composition (Barelli *et al.*, 2020b; Kabwe *et al.*, 2020). For the present study, mice from Nottingham had a lower relative abundance of Basidiomycota ASVs, and a greater relative abundance of Apicomplexa and Nematoda ASVs, compared to Southport and Wirral mice. The differences among sampling sites are likely shaped by either sampling-site-specific factors or factors that may differentiate between mice from Nottingham and mice from Southport and Wirral (Weldon *et al.*, 2015).

First, to address the sampling-site-specific factors. Different animal feeds available to the house mice at each of three sampling sites could be a driving factor behind differences in the eukaryome composition, given that dietary differences have been shown to drive differences in eukaryome composition in laboratory mice (Heisel *et al.*, 2017; Gupta *et al.*, 2023). Alternatively, seasonal effects may be contributing to the differences seen in eukaryome composition among the sampling sites. Nottingham mice were trapped earlier in the year than Southport and Wirral mice, and so may have had access to different diets. This was hypothesised to be the driver behind seasonal effects on the gut bacteriome in wild mice (Maurice *et al.*, 2015; Marsh *et al.*, 2022). Other sampling-site-specific factors could include population density and soil microbiome composition, as seen for the bacteriome (Grieneisen *et al.*, 2019; Sarkar *et al.*, 2020).

Second, to address the host factors that may differentiate between mice from Nottingham and mice from Southport and Wirral. Nottingham mice were found to have a higher prevalence of *Eimeria* spp. compared to Southport and Wirral mice, suggesting *Eimeria* infection may be driving the differences in eukaryome composition seen among sampling sites. Comparison between *Eimeria*-positive and *Eimeria*-negative mice confirmed that mice parasitized with *Eimeria* had a different eukaryome composition. This is consistent with *Eimeria* infection in

chickens, where the gut mycobiome community was comparably different between pre- and post-infection of *Eimeria* (Hume *et al.*, 2012). Additionally, *Eimeria* infection has been shown to change the gut bacteriome composition in mice, further supporting the evidence presented in this current study that *Eimeria* changes the eukaryome composition (Huang *et al.*, 2018). In the present study, the causes of differences in gut eukaryome composition between *Eimeria*-positive and *Eimeria*-negative mice is unknown. However, one hypothesis is that *Eimeria* infection alters the gut eukaryome composition by changing the host's immune state, which was also associated with gut eukaryome composition, specifically IgA production and intestinal inflammation.

In the present study, *Eimeria*-positive mice had a greater faecal IgA concentration compared to *Eimeria*-negative mice. This suggests that *Eimeria* infection increases the host's intestinal IgA production, consistent with findings in laboratory rats (Smith *et al.*, 1995). Host IgA responses allow the colonisation of commensal fungi whilst limiting colonisation of pathogenic fungi (Ost *et al.*, 2021). Thus, if *Eimeria* infection increases the host's IgA production, the host's modulation of other gut eukaryotes by IgA may be altered, changing the eukaryome composition. In the present study, I also found a near-significant effect of *Eimeria* infection on caecal inflammation. This is consistent with a previous study in laboratory mice, which found mice had elevated inflammatory pathways in the gut following *Eimeria* infection (Schmid *et al.*, 2014). Caecal inflammation had a near-significant effect on the caecal eukaryome composition, which suggests *Eimeria*-induced inflammation of the caecum is disrupting the immune regulation of caecal eukaryotic taxa. Furthermore, it is conceivable that the decreased caecal eukaryome diversity seen in mice with greater inflammation in the small intestine is caused by *Eimeria* infection. *Eimeria* develops in the villi of the small intestine of laboratory mice (Nowell and Higgs, 1989; Schmid *et al.*, 2014), and so it's presence may cause physical changes to the small intestine which may promote or impede the growth of other intestinal eukaryotes (Lu *et al.*, 2021).

The hypothesis outlined above focuses on how *Eimeria* infection may change the gut eukaryome composition by changing the host's immune state and subsequent modulation of gut eukaryotes. However, the direction of causality is unknown and the differences in the host immune state seen among sampling sites are likely driven by other taxa in the gut microbiome, as well as *Eimeria* (Iliev *et al.*, 2012; Chudnovskiy *et al.*, 2016; Yang *et al.*, 2020; Ost *et al.*, 2021). Of particular note are the parasitic nematodes identified by 18S rRNA sequencing. In the present study, Oxyurida-positive mice had a significantly greater gut eukaryome alpha diversity compared to Oxyurida-negative mice. Thus, it is likely that the presence of Oxyurida nematodes is more important than *Eimeria* for driving eukaryome diversity composition. Indeed, parasitic nematodes play a large role in the rate of host IgA production, although this

was not formally tested in the present study (Ramos *et al.*, 2022). Furthermore, the dynamic interactions seen between *Eimeria* and parasitic nematode infections are likely to further drive variation in host IgA production and how this subsequently modulates gut eukaryome composition (Rausch *et al.*, 2010; Knowles *et al.*, 2013; Clerc *et al.*, 2019b).

To further investigate and test the hypothesised association between eukaryome composition, host immune state, and *Eimeria* and/or nematode parasitism, more studies are needed. Longitudinal sampling of the gut eukaryome would provide more insight into how parasitism impacts gut eukaryome composition (Telfer *et al.*, 2008; Murillo *et al.*, 2022). Specifically, monitoring the changes in abundance of eukaryotic taxa over time, and how changes were associated with the intensity of *Eimeria* and nematode infection, would allow better predictions of which gut taxa are affected by the presence of these parasites. Furthermore, perturbation experiments could be used to determine which parasites were the driving factors behind differences in eukaryome composition seen among sampling sites. For example, a laboratory system could be used to monitor how the eukaryome composition changed following infection with *Eimeria* (Huang *et al.*, 2018). Additionally, anti-helminthic treatment could be used in a wild system to investigate how parasitic nematodes, and their interaction with *Eimeria*, contributes to both the host immune state and gut eukaryome composition (Knowles *et al.*, 2013). Finally, to determine how the host's immune response mediates interactions between taxa in the eukaryome, quantifying species-specific IgA would provide context as to which taxa are inducing host IgA (Ost *et al.*, 2021). Alternatively, studies using IgA knockout mouse strains would determine the relative contribution of IgA production to eukaryome composition (Jackson *et al.*, 2021).

In addition to parasitism and immune state, other host factors tested for an effect on eukaryome diversity and composition were age, body condition, sex, and reproductive status. In the present study, an effect of age and body condition differentiated Nottingham mice from Wirral and Southport mice and these factors had an effect on eukaryome composition. However, the correlation between age, body condition, and faecal IgA concentration makes it hard to unpick the relative contribution of each covariate to eukaryome composition. Studies have shown mixed age-related effects on eukaryome composition (Heitlinger *et al.*, 2017; Wampach *et al.*, 2017; Rao *et al.*, 2021; Turunen *et al.*, 2023) but the effect of faecal IgA concentration on eukaryome composition has not been researched in wild mammals. To this end, future studies should aim to clarify age-related changes to eukaryome composition, and how this may interact with IgA-related intestinal immunity. There was no effect of sex on the eukaryome composition, consistent with findings in another wild rodent species (Antwis *et al.*, 2021). Other studies have shown there are mixed effects of host sex on the gut eukaryome in

both humans and wild mammals including baboons, macaques, and hyenas (Strati *et al.*, 2016; Heitlinger *et al.*, 2017; Sun *et al.*, 2018; Rao *et al.*, 2021; Murillo *et al.*, 2022).

4.4.3 Comparisons and interactions between eukaryotic and bacterial taxa

This study aimed to compare the diversity of eukaryotic and bacterial taxa in the gut microbiome of wild house mice and identify co-occurrence between the two taxa types. In contrast to the eukaryome, the taxa in the bacteriome were similar to that seen in other wild mice and in laboratory mice. Specifically, Firmicutes and Bacteroidota were the predominant phyla in the faecal bacteriome, as shown previously (Linnenbrink *et al.*, 2013; Weldon *et al.*, 2015; Goertz *et al.*, 2019; Rosshart *et al.*, 2019). In the present study, despite fewer bacterial ASVs and phyla being identified compared to eukaryotic ASVs, bacteriome diversity was always higher than eukaryome diversity. This is consistent with previous studies in humans and other mammals (Parfrey *et al.*, 2014; Nash *et al.*, 2017). This finding is likely due to few highly abundant taxa in the eukaryome – *e.g. Wallemia, Eimeria*, and Oxyurida – and many low abundance taxa resulting in low measures of diversity.

There was no association between eukaryome and bacteriome alpha diversity. This contrasts to a previous study in primates, which found that the alpha diversity of the two taxa types were significantly correlated (Mann *et al.*, 2020). This suggests that different processes drive the diversity of these two groups of taxa in wild house mice. Indeed, the study presented here identified differences in bacteriome alpha diversity among the three sampling sites, and a correlation between alpha diversity and age, body condition, and IgA concentration, none of which were seen for the faecal eukaryome. Additionally, in contrast to faecal eukaryome diversity, Oxyurida infection status was not a significant explanatory factor of faecal bacteriome alpha diversity. Previous studies have also shown the rodent gut bacteriome alpha diversity is not impacted by helminth and *Eimeria* presence (Kreisinger *et al.*, 2015; Maurice *et al.*, 2015; Kim *et al.*, 2022, but see Weldon *et al.*, 2015; Bouilloud *et al.*, 2023). Thus, further work is needed to understand why eukaryome alpha diversity, but not bacteriome alpha diversity, changes in response to Oxyurida infection.

Having identified that parasitism changes the faecal eukaryome composition (above, section 4.4.2), there was also evidence that parasitism changed the faecal bacteriome composition, in support of other studies in wild mammals (Kreisinger *et al.*, 2015; Mann *et al.*, 2020; Montero *et al.*, 2021; Kim *et al.*, 2022). In particular, parasitized mice had a higher relative abundance of *Lactobacillus reuteri* and *Helicobacter apodemus* compared to unparasitized mice. Previous studies in laboratory and wild mice have identified a higher abundance of both *Lactobacillus* spp. and *Helicobacter* spp. when co-infecting with helminths (Fox *et al.*, 2000; Reynolds *et al.*, 2014; Kreisinger *et al.*, 2015; Kim *et al.*, 2022). In contrast, *Eimeria* infection in laboratory mice

has been shown to increase *Helicobacter* spp. abundance, but decrease *Lactobacillus* spp. abundance (Huang *et al.*, 2018). This suggests that the greater abundance of *L. reuteri* and *H. apodemus* in the faecal bacteriome of parasitized mice compared to unparasitized mice in the present study is driven by the presence of nematodes rather than *Eimeria.*

Differential abundance and network analyses gave different answers when identifying microbial interactions in the gut. As discussed above, differential abundance analysis identified that parasitic nematodes, and potentially *Eimeria,* are associated with *L. reuteri* and *H. apodemus*. However, network analyses did not predict these inter-kingdom interactions. Furthermore, network analyses did not predict that *Eimeria* and parasitic nematodes were negatively interacting with Basidiomycota fungi, despite differential abundance analysis identifying that parasitized mice had lower Basidiomycota ASVs compared to unparasitized mice. Why different methods of data analysis have revealed different associations is unclear and warrants further investigation. Possibilities include i) network analyses considering correlations of abundance among all ASVs, whereas differential abundance analyses compared abundance of ASVs based on the presence or absence of parasites and ii) considering each sampling site separately for network analyses, whereas differential abundance considered all mice simultaneously.

Analysis of the different faecal microbiome networks in the present study identified the degree of connectivity, and the taxa predicted to be interacting, differed among mice from different sampling sites. Given that Wirral mice were dissimilar to Southport and Nottingham mice in network connectivity, whereas Nottingham mice were dissimilar to Southport and Wirral mice in faecal microbiome community composition, it is likely that variation in faecal microbiome networks is driven by sampling-site-specific or host-specific factors, as well as faecal microbiome community composition. Incorporating sampling-site-specific or host-specific factors, such as diet and immune state, into microbial network analyses would help identify which factors are important for influencing microbial interactions (Matchado *et al.*, 2021).

Further to comparisons among sampling sites, I also identified that networks had greater connectivity when both eukaryotic and bacterial taxa were in the network together. This is consistent with a previous study in the lung and skin microbiome of humans, which found incorporating fungi into bacterial networks increased connectivity among the taxa in the microbiomes (Tipton *et al.*, 2018). In the present study, most interactions in the multi-kingdom networks were intra-kingdom, rather than inter-kingdom, and some of these intra-kingdom associations were not previously seen when considering the eukaryome and bacteriome separately. Thus, these findings show that considering multi-kingdom taxa may improve predictions of which taxa are interacting, compared to considering bacteria or eukaryotes

separately (Tipton *et al.*, 2018). Network analyses, along with correlations of ASV abundance, also found the proportion of negative inter-kingdom associations was greater than the proportion of negative intra-kingdom interactions. This is again consistent with the findings in the human lung and skin microbiome, where including fungal taxa into bacterial networks increased the percentage of negative associations (Tipton *et al.*, 2018). Negative interactions in the gut microbiome are theorised to be more common than positive (Coyte and Rakoff-Nahoum, 2019), but the network analyses in both the present study and Tipton *et al.* (2018) show that this applies only to inter-kingdom interactions.

It is important to note that the associations among microbial taxa identified from microbial networks and differential abundance analyses are predictions, rather than evidence of interactions (Matchado *et al.*, 2021). In order to test and validate these predicted microbial associations, co-culture *in vitro* and *in vivo* methods could be used (Tipton *et al.*, 2018; Rao *et al.*, 2021). Further to validating predicted associations, co-culturing could be used to identify why associations are sampling-site and/or host-specific and why inter-kingdom associations are more likely to be negative, by manipulation of factors known to modulate gut microbial interactions *e.g.* diet and host immunity (Hoffmann *et al.*, 2013; Paterson *et al.*, 2017).

4.4.4 The immune state of wild house mice

This study collected data on faecal IgA concentration, faecal mucin concentration, and intestinal inflammation to determine how the host immune state was linked to the eukaryome composition of wild house mice. In doing so, this work contributes to currently limited characterisation of the immune state of wild house mice (Abolins *et al.*, 2018). The faecal IgA concentrations of wild house mice varied among the different sampling sites, consistent with previous comparisons among populations of wild house mice (Abolins *et al.*, 2018). Whilst Abolins *et al.* (2018) did not find a correlation between IgA concentration, age, and body condition as seen in the study presented here, age and faecal IgA have been shown to correlate in wild wood mice and hyenas (Abolins *et al.*, 2018; Clerc *et al.*, 2019a; Ferreira *et al.*, 2021).

To my knowledge, this is the first study to measure mucin concentration in wild house mice (Abolins *et al.*, 2018; Ferreira *et al.*, 2021). In the present study, faecal mucin concentration was higher in *Eimeria*-positive mice than *Eimeria*-negative mice. This suggests a mucogenic response to *Eimeria* infection in wild mice that is not seen in laboratory mice, but has been shown in chickens (Collier *et al.*, 2008; Linh *et al.*, 2009). However, the directionality of this association is unknown, and other taxa in the gut, such as parasitic nematodes or bacteria, may be driving this difference in faecal mucin concentration (Leung *et al.*, 2018a; Leung *et al.,* 2018b; Ferreira *et al.*, 2021). Also of note, the average faecal mucin concentration of wild mice

was higher than each of the three laboratory controls. Given that i) wild mice are likely exposed to greater antigenic challenge than laboratory mice (Abolins *et al.*, 2017) and ii) mucin production contributes to mucosa formation, an important barrier that protects against pathogen and parasite invasion (McGuckin *et al.*, 2011), the comparative difference in faecal mucin concentration between wild and laboratory mice suggests that mucin production is an important and continuous aspect of pathogen defence of wild house mice. This supports previous research comparing mucus composition between laboratory and wild mice, which found that microbial penetration of the mucus layer was lower in wild mice (Jakobsson *et al.,* 2015).

The study presented here is also the first, of which I am aware, to histologically assess gut inflammation in wild house mice (Rosshart *et al.*, 2017; Abolins *et al.*, 2018; Rynkiewicz *et al.*, 2019). Comparison of the distribution of inflammation scores for each tissue type showed that gut inflammation was varied along the digestive tract. In particular, the large intestine and caecum were typically more inflamed than expected, whereas the small intestine showed equal distribution of inflammation scores, if not lower than expected. These results suggest that the drivers of gut inflammation in wild house mice are different for the small and large intestine. This is consistent with a review in laboratory mice that stated that the small and large intestine of mice should be considered as two separate immunological sites (Bowcutt *et al.*, 2014). The differences in inflammation seen along the digestive tract in the level of inflammation are potentially driven by the differences in microbial load and microbiome community composition seen between the small and large intestine (Suzuki and Nachman, 2016; Barlow *et al.*, 2020).

Further to variation of inflammation along the digestive tract, the present study also found variation among sampling sites in caecal inflammation. This is consistent with comparisons of other measures of immune state among populations of wild house mice, such as the number and proportion of different splenocytes (Abolins *et al.*, 2018). A caveat to the description of gut inflammation in wild house mice in the present study is that inflammation scores were based on methods developed for laboratory mice (Erben *et al.*, 2014). It is unknown how well these methods translate to wild mice, as the baseline inflammation levels of wild house mice are likely very different to laboratory mice (Abolins *et al.*, 2017). Thus, high inflammation in laboratory mice may not be equivalent to high inflammation in wild mice. Additionally, it is important to note that cutting open the caecum to extract its contents likely disrupted mucosal integrity, which may have biased measures of inflammation.

4.4.5 Conclusion

The study presented here is the first, of which I am aware, to simultaneously characterise the fungi, helminths, and protozoa found in the gut eukaryome of wild house mice. In doing so, the study has shown that the house mouse eukaryome is distinct to that of both its laboratory counterparts and to that of other rodents. However, it should be noted that comparability among studies is limited by the use of different methodologies, such as the 18S primers used. Thus, finding a methodology that is capable of consistently capturing the diversity of taxa in the eukaryome should be a primary focus for future studies. In this study, I have also demonstrated that the faecal eukaryome and caecal eukaryome differ, and considered why this might be. This means that faecal samples may not be sufficient to fully characterise wild animals' gut eukaryome, Additionally, the high relative abundance and prevalence of *Wallemia* identified in this study highlights that taxa considered incapable of colonisation are, as a minimum, likely to be interacting with other taxa in the eukaryome, if not capable of colonising the gut. Thus, taxa considered dietary and transient should be investigated further rather than disregarded from analyses.

In addition to characterising the eukaryome of wild house mice, this study has identified that parasitism is likely a key driver of eukaryome composition, with this effect potentially modulated by host immune state, specifically IgA and intestinal inflammation. Whilst this is an observational study, and the directionality of the association is unknown, it generates a hypothesis that future studies can investigate, either by using longitudinal studies or perturbation experiments. I have also shown that the diversity and composition of the gut eukaryome and bacteriome of wild mice can respond to different host factors and that these effects are sampling-site-specific. Thus, future studies should not assume that factors modulating the gut bacteriome also apply to the gut eukaryome. Finally, I have identified that inter-kingdom interactions may be more antagonistic than positive, which requires further work to identify why this may be.

Chapter 5: General discussion

5.1 Summary of findings

The gut microbiome is now widely recognised as a key aspect of mammalian biology, owing to its role in developing the host's immune system, the provision of key metabolites needed for host nutrition, and its contribution to host health (O'Hara and Shanahan, 2006; Sommer and Bäckhed, 2013). To this end, there has been extensive research into characterising how changes in the gut microbiome can contribute to host fitness, and the wider impacts this has on the host's ecology and evolution (Suzuki, 2017; Henry *et al.*, 2021). However, lagging behind is our understanding of what factors drive changes in the gut microbiome composition and how this could change host fitness (Costello *et al.*, 2012; Foster *et al.*, 2017; Coyte *et al.*, 2021). The variation in eukaryome composition over an individual's lifetime, and among individual hosts, is substantially different to the bacteriome but the causes of this difference are relatively unexplored (Parfrey *et al.*, 2014; Nash *et al.*, 2017). To this end, the work of my thesis aimed to address this knowledge gap by focussing on which host factors contribute to gut eukaryome composition in wild rodents. In doing so, this research would provide knowledge that could be applied to future studies to determine how the gut eukaryome contributes to the host's biology. In the following section, I briefly summarise the reasoning behind each of my objectives and the main findings.

The first results presented in my thesis (Chapter 2) focused on the protozoal component of the gut eukaryome. To date, mammalian gut protozoa research has tended to focus on parasitic species and there is less focus on the general diversity of protozoa found in mammals (Parfrey *et al.*, 2014; del Campo *et al.*, 2020). To this end, I conducted a methodical literature search to compile a comprehensive list of the protozoa found with the eukaryome of wild rodents. From this review, I presented evidence that some protozoa had an extremely wide host range, whereas others were only identified in a few host species. Similarly, I found evidence that some rodent species were host to many protozoa genera, whereas other rodent species host to only a few. I also highlighted that environmentally-transmitted protozoa, and the ingestion of protozoa from dietary sources, must be considered when describing the diversity of gut protozoa in rodents. Finally, I discussed that taxonomic confusion from reclassifications of protozoa could underestimate the diversity of protozoa capable of colonising wild rodents. It is well-established that host sociality is a key driver in the transmission and maintenance of gut microbial taxa, with more opportunities for horizontal transmission of gut microbes as hosts become more social (Moeller *et al.*, 2016b; Sarkar *et al.*, 2020). Thus, I conducted a meta-analysis to test the hypothesis that host sociality and behaviour affected the prevalence of protozoa in wild rodents. The results showed that the

prevalence of protozoa was heterogenous among rodent species. However, I found no evidence that sociality and behaviour was driving the heterogeneity of prevalence among host species, contrary to the hypothesis.

For my second objective (Chapter 3), I focussed on improving methods of quantifying eukaryotic gut taxa. Currently, amplicon sequencing methods are the standard approach to characterise the gut microbiome composition (Franzosa *et al.*, 2015; Gupta *et al.*, 2019). Obtaining quantitative data on microbial load in the gut can improve analyses of amplicon sequencing data (Vandeputte *et al.*, 2017; Rao *et al.*, 2021). However, optimisation of quantification methods has largely focussed on the bacteriome, and have yet to be translated to eukaryome research. Thus, I designed and tested a novel method for quantifying eukaryotic taxa in the gut, based on previously published methods for bacteriome quantification (Vandeputte *et al.*, 2017). The method used flow cytometry to separate eukaryotic cells from prokaryotic cells by using a eukaryotic-specific stain. In doing so, eukaryotic cells could be quantified before sequencing to better determine eukaryome community composition. The results from the pilot study conducted showed that the method was inaccurate in identifying eukaryotic cells in faecal samples. I concluded that the inaccuracy may be due to unexpected low-specificity of the eukaryotic-specific stain or an incorrect gating strategy.

The results presented in Chapter 4 focussed on the biggest objective of my thesis: characterising the gut eukaryome of wild house mice, and identifying factors associated with eukaryome composition. The laboratory mouse is the primary model organism used for gut microbiome research (Rosshart *et al.*, 2019; Viney, 2019). However, research has shown that it's gut bacteriome is different to that of its wild counterparts (Viney, 2019; Bowerman *et al.*, 2021). Furthermore, there has been limited study of the eukaryome, nor what drives its composition, in wild house mice. To address this, I used 18S rRNA amplicon sequencing to identify the common taxa in the gut eukaryome, alongside measuring host factors hypothesised to drive eukaryome composition. From this observational study, I found that the eukaryome of wild house mice was primarily composed of fungal taxa, and that the diversity of protozoa was low. Comparison of the eukaryome composition of wild mice to studies in laboratory mice and other wild rodents showed that different taxa dominated the gut eukaryome in different studies. I also discovered that when present, *Eimeria* and nematode parasites had a high relative abundance that was negatively correlated with Basidiomycota relative abundances, suggesting that parasitic infection changes the gut eukaryome composition. I presented data that allowed me to hypothesise that this change in eukaryome composition may be driven by host immune responses that are induced following parasitic infection, and discussed how future studies could test this hypothesis. Finally, it was concluded that the gut eukaryome and gut bacteriome can change in response to different host factors,

and that considering both taxa types together could improve our understanding of microbial interactions in the gut.

5.2 Synthesis of findings

Having previously discussed the findings of my objectives in their respective chapters, I will now synthesise the results of my thesis. In doing so, I aim to highlight some common themes and contrasting results in my research. In Chapter 2, I presented a comprehensive list of the gut protozoa found in wild rodents. From this table, eight protozoa genera were described from wild house mice. Five of these genera (*Chilomastix*, *Entamoeba, Giardia, Isospora,* and *Trichomonas*) were not subsequently identified in the wild mice in my own eukaryome sequencing study (Chapter 4). This suggests that previous reports of these protozoa are not showing members of the common core house mice eukaryome but rather represent incidental transmission from other host species (Risely, 2020). Alternatively, it is possible that these genera are common members of the wild house mice eukaryome, but were not detected in my eukaryome sequencing study due to spatial gradients in the distributions of these protozoa or false negatives (Parfrey *et al.*, 2014; Risely, 2020). False negatives could have arisen from the 18S primers used in my eukaryome sequencing study being a poor match to these protozoa or from inefficient DNA extraction (Parfrey *et al.*, 2014; Vaulot *et al.*, 2022).

The three other protozoa genera described from wild house mice in Chapter 2 were *Cryptosporidium, Eimeria,* and *Blastocystis,* and these were subsequently identified in Chapter 4. This provides good evidence that these three protozoa should be considered members of the common core gut eukaryome, although it should be noted that their prevalence was varied among sampling sites. My eukaryome sequencing study also described the presence of *Balantidium* in one mouse from Nottingham (Chapter 4), a protozoa genus that had not previously been described from wild house mice (Chapter 2). It is likely that the presence of *Balantidium* was from horizontal transmission from pigs found on the farm, which are the natural host of *Balantidium* (Schuster and Ramirez-Avila, 2008). However, it is unknown if its presence in the mouse was a transient infection following spill-over from the pigs, or if *Balantidium* is capable of colonising the gut given sufficient opportunities of infection.

An emergent theme from my research is the presence of transient eukaryotes in the gut of wild rodents. Transient eukaryotes, rather than gut residents, are those taxa that are considered incapable of surviving in the mammalian gut, such as macrofungi and free-living protozoa (Hallen-Adams and Suhr, 2017; Heitlinger *et al.*, 2017; Mann *et al.*, 2020; Lavrinienko *et al.*, 2021a; Murillo *et al.*, 2022). Transient taxa are often removed from characterisations and analyses of the gut eukaryome in order to focus on taxa that are considered true gut

residents (Mann *et al.*, 2020; Lavrinienko *et al.*, 2021a; Murillo *et al.*, 2022). However, transient taxa may still have functional roles in the gut microbiome and so should not be overlooked in gut eukaryome research (Zhang *et al.*, 2016; Hallen-Adams and Suhr, 2017; Lavrinienko *et al.*, 2021a; Jackson *et al.*, 2022). This is demonstrated by the findings of *Acanathomoeba* and *Wallemia* in wild rodents in Chapters 2 and 4 respectively. In Chapter 2, I concluded that the protozoan genera *Acanathomoeba* was likely to be a transient infection due to its free-living lifestyle (Rodríguez-Zaragoza, 1994). However, *Acanathomoeba* spp. can act as carriers of both pathogenic bacteria and parasitic protozoa, thus their ingestion could have a large impact on the host health (Vaerewijck *et al.*, 2014). The fungal genus *Wallemia* was found to be highly prevalent in the gut eukaryome of wild house mice in Chapter 4. *Wallemia* spp. are considered to be incapable of long-term colonisation and residency of the gut due to requiring low water content for growth, although its presence in the gut can exacerbate host disease (Zalar *et al.*, 2005; Wheeler *et al.*, 2016; Hallen-Adams and Suhr, 2017; Skalski *et al.*, 2018). The findings of both *Acanathomoeba* and *Wallemia* in wild rodents highlights the need to consider how transient infections may impact the host and the wider gut microbiome. Further to their impact on the host, it is important to note that overlooking supposedly transient taxa may limit the identification of novel resident taxa in the gut (Lavrinienko *et al.*, 2021a).

Another common theme identified in my research is that frequent reclassifications of microbial eukaryotes, as well as taxonomic confusion, can limit our understanding of gut eukaryome composition. For example, I discussed in Chapter 2 that the host range of *Tritrichomonas* is likely underestimated due to some species being synonymous with *Trichomonas* (Burr *et al.*, 2012). This is exacerbated by the frequent and contradictory reclassifications of eukaryotic taxa following the advance of high throughput sequencing (Tenter *et al.*, 2002; Ruggiero *et al.*, 2015; Adl *et al.*, 2019; Stensvold and Clark, 2020). Rapidly changing eukaryotic taxonomy can lead to databases not reflecting the latest taxonomic changes, as reviewed in del Campo *et al.* (2018). This presents a challenge to characterising the taxa found in the gut eukaryome, as taxonomic assignment of amplicon sequencing data from different databases gives different taxonomic information (del Campo *et al.*, 2018; Kataoka and Kondo, 2019; Gogarten *et al.*, 2020). This problem is highlighted by my own eukaryome sequencing study, where I found that taxonomic assignment of ASVs from *Candida* spp. was complicated by the reclassification of many *Candida* spp. into other genera (Kidd *et al.*, 2023). For example, ASVs from *Nakaseomyces glabrata,* formerly *Candida glabrata*, were assigned as *Nakaseomyces-Candida*-clade at the genus rank but *Candida glabrata* at the species rank (Kidd *et al.*, 2023). Thus, there is a requirement for the thorough and systematic checking of eukaryote taxonomy before aggregating ASVs at different taxonomic ranks for eukaryome characterisation and analysis.

5.3 Future directions

The three objectives of my thesis have given a much better understanding of the eukaryome composition of wild rodents, as well as the methodologies needed to optimise eukaryome research. In the following section, I will first address how further work could build on the findings in my thesis. Then, I will discuss how my research has highlighted key areas and considerations that should be applied to other future studies.

5.3.1 Opportunities to build on the findings of this thesis

First, to address my research on the gut protozoa of wild rodents. My meta-analyses focussed particularly on how host sociality and lifestyle could impact the prevalence of gut protozoa in wild rodents, as different host behaviours can impact transmission rates of gut microbes (Ezenwa, 2004; Moeller *et al.*, 2016b; Barelli *et al.*, 2020b; Sarkar *et al.*, 2020). However, transmission is also dependent on survivability of the protozoa outside of the host (King and Monis, 2007; Costello *et al.*, 2012; Dumètre *et al.*, 2012). Incorporating this factor of protozoa transmission and infection into analyses would determine the relative contribution of host and protozoa traits in driving the variation in rodent gut protozoa composition. For example, a study of the different transmission mechanisms of parasites in wild primates found that host-specificity of protozoa was largely dictated by transmission strategy (Pedersen *et al.*, 2005). It would also be interesting to test if protozoa transmission strategy is associated with pathogenicity in rodents, as seen in bacterial taxa in mice (Moeller *et al.*, 2018). In addition to considering protozoan traits, it would beneficial to consider host behaviour at the individual level, rather than species-level (Hawley *et al.*, 2021). An approach to do this is to conduct meta-analyses on a smaller subset of studies that measure the sociality of individual hosts as well as describing their gut protozoa. Alternatively, semi-wild enclosures could be used to monitor the transmission of laboratory-infected protozoa, or indeed other microbial taxa, whilst measuring social associations of individual mice (Raulo *et al.*, 2021).

Second, to address the quantification of gut eukaryotes. The identification and quantification of eukaryotic cells using the method developed in this thesis was shown to be inaccurate, thus there is opportunity to refine and develop these methods for potential future use. In particular, spike-in experiments of specific eukaryotic taxa, and the use of germ-free mice, could be used to better identify the most suitable FACS gating parameters and the specificity of the stains used (Jackson *et al.*, 2021). Further to refining the proposed method, it would be interesting to apply other flow cytometry methods to the eukaryome. For example, IgA-staining alongside eukaryotic-specific stains could be used to identify which taxa in the eukaryome are modulated by the host immune system, rather than focussing on a specific eukaryotic taxon (Palm *et al.*, 2014; Jackson *et al.*, 2021; Ost *et al.*, 2021). Finally, the FACS work presented in my thesis identified a potential method for removing host DNA from faecal samples. DNA extracted from

FACS-processed fractions had proportionally fewer ASVs belonging to Vertebrate taxa, compared to DNA extracted directly from faecal samples. Thus, FACS processing could be used in gut eukaryome studies to allow less abundant taxa to be detected, improving diversity estimates (Pereira-Marques *et al.*, 2019). To develop and optimise this further, future studies could identify at what stage of the FACS process the host DNA was removed.

Thirdly, to address the eukaryome composition of wild house mice. A major finding from this work was that mice parasitized with *Eimeria* and nematodes had a different eukaryome composition compared to unparasitized mice. Determining that this is causal and the direction of that causality would be the first step to build on these results. For example, using anthelminthic or anticoccidial drugs to remove *Eimeria* and nematodes from wild mice with longitudinal sampling would determine how the eukaryome composition changed in response to these parasites (Knowles *et al.*, 2013). Alternatively, infecting laboratory mice with *Eimeria* and nematodes would provide a more controlled environment to test, and so infer causality of, the effect of these parasites on the gut eukaryome composition of mice (Huang *et al.*, 2018). It would be interesting to compare the results from these proposed laboratory and wild experiments to identify how the natural variation in the eukaryome composition seen in wild mice, and indeed other environmental and host factors, might change the dynamics of parasite-eukaryome interactions (Viney and Riley, 2017; Leung *et al.*, 2018b; Viney, 2019).

A further development of my work could focus on defining which of the eukaryotes identified in my eukaryome sequencing survey are true gut residents and which are transient infections. For example, infecting laboratory mice with taxa of interest, *e.g. Wallemia,* and monitoring it's abundance over time would provide a better understanding of how persistent these taxa are in the gut (Skalski *et al.*, 2018). Additionally, immune-deficient mice could be used to determine if transient taxa are unable to persist in the gut due to immunomodulation by the host or by other factors. Furthermore, using RNA transcriptomics alongside amplicon sequencing could be used to identify taxa that are actively transcribing genes, suggesting survivability in the gut, compared to those that are not actively transcribing genes (Suhr and Hallen-Adams, 2015; Lind and Pollard, 2021).

5.3.2 Key considerations for future research

The primary aim of my thesis was to determine which host factors contribute to the gut eukaryome composition in wild rodents. In doing so, I have identified some key considerations for future eukaryome studies. Firstly, my research has contributed to our understanding that the gut eukaryome of wild house mice is distinct to that of laboratory mice (Ehret *et al.*, 2017; Rosshart *et al.*, 2019; Viney, 2019; Bendová *et al.*, 2020). To this end, it is important for future eukaryome studies to consider using a wild system when investigating interactions between

the host and the gut eukaryome, as seen with the bacteriome (Amato, 2013; Viney, 2019). Secondly, my meta-analysis demonstrated the importance of considering host's individual- and population-level traits, rather than species-specific traits, to understand the context-dependent role of host behaviour on microbial transmission. Thirdly, I have also shown that gut eukaryome diversity does not always respond to the same host factors as the bacteriome, and so this should be not be assumed in future studies. Fourthly, I have demonstrated that the diversity and composition of the gut eukaryome of wild house mice were associated with the host immune state. Thus, it is recommended that future studies incorporate some measure of host immune state when investigating host-eukaryome interactions in wild mammals (Rynkiewicz *et al.*, 2019; Ferreira *et al.*, 2021). Finally, I have demonstrated that the relative abundances of fungal taxa in the gut are different when hosts are infected with both *Eimeria* and parasitic nematodes. Therefore, future studies considering only the mycobiome need to consider that the presence of protozoa and/or helminths may be driving changes in eukaryome composition.

My research also found evidence for inter-kingdom interactions in the gut microbiome of wild house mice. This highlights that future studies using wild mammalian systems for gut microbiome research need to consider inter-kingdom interactions, as such interactions can have a large impact on both the gut microbiome composition and the host (Hoffmann *et al.*, 2013; Burgess *et al.*, 2017; Paterson *et al.*, 2017; Haak *et al.*, 2021; Rao *et al.*, 2021). In my work, inter-kingdom interactions were identified between microbial eukaryotes and bacteria. However, it is likely that inter-kingdom interactions also extend to archaea and viruses (Minot *et al.*, 2011; Hoffmann *et al.*, 2013; Nkamga *et al.*, 2017; Vemuri *et al.*, 2020). Thus, the presence of these taxa, and their interactions with the host and other microbes, should also be considered when researching the gut microbiome in wild animals.

Finally, the work presented in this thesis has provided two valuable resources for further studies: i) a comprehensive review of the gut protozoa identified in wild rodents and ii) the groundwork of a method for quantifying eukaryotic gut taxa using flow cytometry.

5.4 Final conclusion

My work has used a multi-disciplinary approach to expand our knowledge of the gut eukaryome of wild mammals including: meta-analyses, cell biology, genomics, immunology, and histology. In doing so, I have demonstrated that the prevalence of gut protozoa is species-specific in wild rodents, and this is not driven solely by host behaviour. Furthermore, I have characterised the eukaryotic gut taxa of wild mice, as well as presented evidence that the disease and immune state of the host may underlie how the gut eukaryome is assembled in wild animals. Finally, I have addressed the importance of considering multi-kingdom

interactions in gut microbiome research. Taken together, these findings offer novel insights into the mammalian gut eukaryome which could be applied to future studies investigating host-microbiome interactions.

References

Abolins, S., King, E.C., Lazarou, L., Weldon, L., Hughes, L., *et al.* (2017). The comparative immunology of wild and laboratory mice, Mus musculus domesticus. *Nature Communications*, 8, p. 14811.

Abolins, S., Lazarou, L., Weldon, L., Hughes, L., King, E.C., *et al.* (2018). The ecology of immune state in a wild mammal, Mus musculus domesticus. *PLOS Biology*, 16(4), p. e2003538.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., *et al.* (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*, 66(1), pp. 4–119.

Agranyoni, O., Meninger-Mordechay, S., Uzan, A., Ziv, O., Salmon-Divon, M., *et al.* (2021). Gut microbiota determines the social behavior of mice and induces metabolic and inflammatory changes in their adipose tissue. *npj Biofilms and Microbiomes*, 7, p. 28.

Ahmed, A., Ijaz, M., Ayyub, R.M., Ghaffar, A., Ghauri, H.N., *et al.* (2020). Balantidium coli in domestic animals: An emerging protozoan pathogen of zoonotic significance. *Acta Tropica*, 203, p. 105298.

Ai, S., Zhang, Z., Wang, X., Zhang, Q., Yin, W., *et al.* (2021). The first survey and molecular identification of Entamoeba spp. in farm animals on Qinghai-Tibetan Plateau of China. *Comparative Immunology, Microbiology and Infectious Diseases*, 75, p. 101607.

Alfellani, M.A., Taner-Mulla, D., Jacob, A.S., Imeede, C.A., Yoshikawa, H., *et al.* (2013). Genetic Diversity of Blastocystis in Livestock and Zoo Animals. *Protist*, 164(4), pp. 497–509.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp. 403–410.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. and Huse, S.M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLOS ONE*, 4(7), p. e6372.

Amato, K.R. (2013). Co-evolution in context: The importance of studying gut microbiomes in wild animals. *Microbiome Science and Medicine*, 1(1), pp. 10–29.

Amato, K.R., Leigh, S.R., Kent, A., Mackie, R.I., Yeoman, C.J., *et al.* (2015). The Gut Microbiota Appears to Compensate for Seasonal Diet Variation in the Wild Black Howler Monkey (Alouatta pigra). *Microbial Ecology*, 69(2), pp. 434–443.

Amato, K.R., G. Sanders, J., Song, S.J., Nute, M., Metcalf, J.L., *et al.* (2019). Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *The ISME Journal*, 13(3), pp. 576–587.

Andersen, L.O., Vedel Nielsen, H. and Stensvold, C.R. (2013). Waiting for the human intestinal Eukaryotome. *The ISME Journal*, 7(7), pp. 1253–1255.

Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., *et al.* (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, 14(1), pp. 19–28.

Andreassen, H.P., Sundell, J., Ecke, F., Halle, S., Haapakoski, M., *et al.* (2020). Population cycles and outbreaks of small rodent: ten essential questions we still need to solve. *Oecologia,* 195, 601-622.

Ansia, I. and Drackley, J.K. (2020). Technical note: Evaluation of 3 methods to determine mucin protein concentration in ileal digesta of young preweaning calves. *Journal of Dairy Science*, 103(7), pp. 6250–6257.

Antwis, R.E., Lea, J.M.D., Unwin, B. and Shultz, S. (2018). Gut microbiome composition is associated with spatial structuring and social interactions in semi-feral Welsh Mountain ponies. *Microbiome*, 6, p. 207.

Antwis, R.E., Beresford, N.A., Jackson, J.A., Fawkes, R., Barnett, C.L., *et al.* (2021). Impacts of radiation exposure on the bacterial and fungal microbiome of small mammals in the Chernobyl Exclusion Zone. *The Journal of Animal Ecology*, 90(9), pp. 2172–2187.

Apfelbach, R., Blanchard, C.D., Blanchard, R.J., Hayes, R.A. and McGregor, I.S. (2005). The effects of predator odors in mammalian prey species: A review of field and laboratory studies. *Neuroscience & Biobehavioral Reviews*, 29(8), pp. 1123–1144.

Appelbee, A.J., Thompson, R.C.A. and Olson, M.E. (2005). Giardia and Cryptosporidium in mammalian wildlife – current status and future needs. *Trends in Parasitology*, 21(8), pp. 370–376.

Bangoura, B., Bhuiya, M.A.I. and Kilpatrick, M. (2022). Eimeria infections in domestic and wild ruminants with reference to control options in domestic ruminants. *Parasitology Research*, 121(8), pp. 2207–2232.

Baniel, A., Amato, K.R., Beehner, J.C., Bergman, T.J., Mercer, A., *et al.* (2021). Seasonal shifts in the gut microbiome indicate plastic responses to diet in wild geladas. *Microbiome*, 9, p. 26.

Baniel, A., Petrullo, L., Mercer, A., Reitsema, L., Sams, S., *et al.* (2022). Maternal effects on early-life gut microbiota maturation in a wild nonhuman primate. *Current Biology*, 32(20), pp. 4508–4520.

Baquero, F. and Nombela, C. (2012). The microbiome as a human organ. *Clinical Microbiology and Infection*, 18(Suppl. 4), pp. 2–4.

Barati, M., KarimiPourSaryazdi, A., Rahmanian, V., Bahadory, S., Abdoli, A., *et al.* (2022). Global prevalence and subtype distribution of Blastocystis sp. in rodents, birds, and water supplies: A systematic review and meta-analysis. *Preventive Veterinary Medicine*, 208, p. 105770.

Barelli, C., Pafčo, B., Manica, M., Rovero, F., Rosà, R., *et al.* (2020a). Loss of protozoan and metazoan intestinal symbiont biodiversity in wild primates living in unprotected forests. *Scientific Reports*, 10, p. 10917.

Barelli, C., Albanese, D., Stumpf, R.M., Asangba, A., Donati, C., *et al.* (2020b). The Gut Microbiota Communities of Wild Arboreal and Ground-Feeding Tropical Primates Are Affected Differently by Habitat Disturbance. *mSystems*, 5(3), pp. e00061-20.

Barendregt, J.J., Doi, S.A., Lee, Y.Y., Norman, R.E. and Vos, T. (2013). Meta-analysis of prevalence. *J Epidemiol Community Health*, 67(11), pp. 974–978.

Barlow, J.T., Bogatyrev, S.R. and Ismagilov, R.F. (2020). A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. *Nature Communications*, 11, p. 2590.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), pp. 1–48.

Behringer, D.C., Karvonen, A. and Bojko, J. (2018). Parasite avoidance behaviours in aquatic environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, p. 20170202.

Bekircan, Ç. and Tosun, O. (2021). First record and distribution of Adelina sp. (Coccidia: Adeleidae) in populations of the Anisoplia segetum Herbst (Coleoptera: Scarabaeidae) in Turkey. *Microbial Pathogenesis*, 154, p. 104848.

Belkaid, Y. and Hand, T. (2014). Role of the Microbiota in Immunity and inflammation. *Cell*, 157(1), pp. 121–141.

Ben-Amor, K., Heilig, H., Smidt, H., Vaughan, E.E., Abee, T., *et al.* (2005). Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Applied and Environmental Microbiology*, 71(8), pp. 4679–4689.

Bendová, B., Piálek, J., Ďureje, Ľ., Schmiedová, L., Čížková, D., *et al.* (2020). How being synanthropic affects the gut bacteriome and mycobiome: comparison of two mouse species with contrasting ecologies. *BMC Microbiology*, 20, p. 194.

Benhamou, R.I., Jaber, Q.Z., Herzog, I.M., Roichman, Y. and Fridman, M. (2018). Fluorescent Tracking of the Endoplasmic Reticulum in Live Pathogenic Fungal Cells. *ACS Chemical Biology*, 13(12), pp. 3325–3332.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., *et al.* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences*, 107(44), pp. 18933–18938.

Béra-Maillet, C., Devillard, E., Cezette, M., Jouany, J.-P. and Forano, E. (2005). Xylanases and carboxymethylcellulases of the rumen protozoa Polyplastron multivesiculatum Eudiplodinium maggii and Entodinium sp. *FEMS Microbiology Letters*, 244(1), pp. 149–156.

Bercik, P., Denou, E., Collins, J., Jackson, W., Lu, J., *et al.* (2011). The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. *Gastroenterology*, 141(2), pp. 599–609.

Bergen, W.G. (2015). Small-intestinal or colonic microbiota as a potential amino acid source in animals. *Amino Acids*, 47(2), pp. 251–258.

Bertolino, S., Wauters, L.A., De Bruyn, L. and Canestri-Trotti, G. (2003). Prevalence of coccidia parasites (Protozoa) in red squirrels (Sciurus vulgaris): effects of host phenotype and environmental factors. *Oecologia*, 137(2), pp. 286–295.

Bisanz, J. E.. (2018). Importing QIIME2 artifacts and associated data into R sessions. R package version 0.99.6. https://github.com/jbisanz/qiime2R.

Björk, J.R., Dasari, M., Grieneisen, L. and Archie, E.A. (2019). Primate microbiomes over time: Longitudinal answers to standing questions in microbiome research. *American Journal of Primatology*, 81, p. e22970.

Björk, J.R., Dasari, M.R., Roche, K., Grieneisen, L., Gould, T.J., *et al.* (2022). Synchrony and idiosyncrasy in the gut microbiome of wild baboons. *Nature Ecology & Evolution*, 6(7), pp. 955–964.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., *et al.* (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, 16, p. 191.

Blessmann, J., Buss, H., Nu, P.A.T., Dinh, B.T., Ngo, Q.T.V., *et al.* (2002). Real-Time PCR for Detection and Differentiation of Entamoeba histolytica and Entamoeba dispar in Fecal Samples. *Journal of Clinical Microbiology*, 40(12), pp. 4413–4417.

Boessenkool, B. (2017). OSMscale: Add a Scale Bar to 'OpenStreetMap' Plots. R package version 0.5.1. https://CRAN.R-project.org/package=OSMscale

Bo, T.-B., Zhang, X.-Y., Kohl, K.D., Wen, J., Tian, S.-J., *et al.* (2020). Coprophagy prevention alters microbiome, metabolism, neurochemistry, and cognitive behavior in a small mammal. *The ISME Journal*, 14(10), pp. 2625–2645.

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., *et al.* (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science Translational Medicine*, 8(343), p. 343ra82.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., *et al.* (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, 6, p. 90.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., *et al.* (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), pp. 852–857.

Bordes, F., Blumstein, D.T. and Morand, S. (2007). Rodent sociality and parasite diversity. *Biology Letters*, 3, pp. 692–694.

Boreham, P.F.L. and Stenzel, D.J. (1993). Blastocystis in Humans and Animals: Morphology, Biology, and Epizootiology. in J.R. Baker and R. Muller (eds) *Advances in Parasitology*. Academic Press, pp. 1–70.

Borges, P.A., Dominguez-Bello, M.G. and Herrera, E.A. (1996). Digestive physiology of wild capybara. *Journal of Comparative Physiology B*, 166(1), pp. 55–60.

Bornbusch, S.L., Greene, L.K., Rahobilalaina, S., Calkins, S., Rothman, R.S., *et al.* (2022). Gut microbiota of ring-tailed lemurs (Lemur catta) vary across natural and captive populations and correlate with environmental microbiota. *Animal Microbiome*, 4, p. 29.

Bouilloud, M., Galan, M., Dubois, A., Diagne, C., Marianneau, P., *et al.* (2023). Three-way relationships between gut microbiota, helminth assemblages and bacterial infections in wild rodent populations. *Peer Community Journal*, 3, p. e18.

Boursot, P., Auffray, J.-C., Britton-Davidian, J. and Bonhomme, F. (1993). The evolution of house mice. *Annual Review of Ecology and Systematics,* 24, pp. 119-152.

Bowcutt, R., Forman, R., Glymenaki, M., Carding, S.R., Else, K.J., *et al.* (2014). Heterogeneity across the murine small and large intestine. *World Journal of Gastroenterology*, 20(41), pp. 15216–15232.

Bowerman, K.L., Knowles, S.C.L., Bradley, J.E., Baltrūnaitė, L., Lynch, M.D.J., *et al.* (2021). Effects of laboratory domestication on the rodent gut microbiome. *ISME Communications*, 1, p. 49.

Brei, B. and Fish, D. (2003). Comment on 'Parasites as a Viability Cost of Sexual Selection in Natural Populations of Mammals'. *Science*, 300(5616), pp. 55–55.

Broadhurst, M.J., Ardeshir, A., Kanwar, B., Mirpuri, J., Gundra, U.M., *et al.* (2012). Therapeutic Helminth Infection of Macaques with Idiopathic Chronic Diarrhea Alters the Inflammatory Signature and Mucosal Microbiota of the Colon. *PLOS Pathogens*, 8(11), p. e1003000.

Brooker, S. (2010). Estimating the global distribution and disease burden of intestinal nematode infections: Adding up the numbers – A review. *International Journal for Parasitology*, 40(10), pp. 1137–1144.

Brown, R.Z. (1953). Social behaviour, reproduction, and population changes in the house mouse (Mus musculus L.). *Ecological Monographs,* 23(3), pp. 217-240.

Brown, R.L. and Clarke, T.B. (2017). The regulation of host defences to infection by the microbiota. *Immunology*, 150(1), pp. 1–6.

Brugman, S. and Nieuwenhuis, E.E.S. (2010). Mucosal control of the intestinal microbial community. *Journal of Molecular Medicine*, 88(9), pp. 881–888.

Buffington, S.A., Di Prisco, G.V., Auchtung, T.A., Ajami, N.J., Petrosino, J.F., *et al.* (2016). Microbial Reconstitution Reverses Maternal Diet-Induced Social and Synaptic Deficits in Offspring. *Cell*, 165(7), pp. 1762–1775.

Burgess, S.L., Gilchrist, C.A., Lynn, T.C. and Petri, W.A. (2017). Parasitic Protozoa and Interactions with the Host Intestinal Microbiota. *Infection and Immunity*, 85(8), pp. e00101-17.

Burke, M.A., Mutharasan, R.K. and Ardehali, H. (2008). The sulfonylurea receptor, an atypical ATP-binding cassette protein, and its regulation of the KATP channel. *Circulation Research*, 102(2), pp. 164–176.

Burr, H.N., Paluch, L.-R., Roble, G.S. and Lipman, N.S. (2012). Chapter 32 - Parasitic Diseases. in M.A. Suckow, K.A. Stevens, and R.P. Wilson (eds) *The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents*. Boston: Academic Press (American College of Laboratory Animal Medicine), pp. 839–866.

Cacciò, S.M., Thompson, R.C.A., McLauchlin, J. and Smith, H.V. (2005). Unravelling Cryptosporidium and Giardia epidemiology. *Trends in Parasitology*, 21(9), pp. 430–437.

Cacciò, S.M. (2018). Molecular epidemiology of Dientamoeba fragilis. *Acta Tropica*, 184, pp. 73–77.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., *et al.* (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), pp. 581–583.

Callegari, M., Crotti, E., Fusi, M., Marasco, R., Gonella, E., *et al.* (2021). Compartmentalization of bacterial and fungal microbiomes in the gut of adult honeybees. *npj Biofilms and Microbiomes*, 7, p. 42.

Crowcroft, P. and Jeffers, J.N.R. (1961). Variability in the behaviour of wild house mice (*Mus musculus* L.) towards live traps. *Proceedings of the Zoological Society of London,* 13(4), pp. 573-582.

Csardi G, and Nepusz, T. (2006). The igraph software package for complex network research. R package version 1.2.6. https://igraph.org

del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., *et al.* (2018). EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLOS Biology*, 16(9), p. e2005849.

del Campo, J., Heger, T.J., Rodríguez-Martínez, R., Worden, A.Z., Richards, T.A., *et al.* (2019a). Assessing the Diversity and Distribution of Apicomplexans in Host and Free-Living Environments Using High-Throughput Amplicon Data and a Phylogenetically Informed Reference Framework. *Frontiers in Microbiology*, 10, p. 2373.

del Campo, J., Pons, M.J., Herranz, M., Wakeman, K.C., del Valle, J., *et al.* (2019b). Validation of a universal set of primers to study animal-associated microeukaryotic communities. *Environmental Microbiology*, 21(10), pp. 3855–3861.

del Campo, J., Bass, D. and Keeling, P.J. (2020). The eukaryome: Diversity and role of microeukaryotic organisms associated with animal hosts. *Functional Ecology*, 34(10), pp. 2045–2054.

Cao, Q., Sun, X., Rajesh, K., Chalasani, N., Gelow, K., *et al.* (2021). Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. *Frontiers in Microbiology*, 11, p. 607325.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(Suppl. 1), pp. 4516–4522.

Carta, L.K. and Li, S. (2018). Improved 18S small subunit rDNA primers for problematic nematode amplification. *Journal of Nematology*, 50(4), pp. 533–542.

Caudet, J., Trelis, M., Cifre, S., Soriano, J.M., Rico, H., *et al.* (2022). Interplay between Intestinal Bacterial Communities and Unicellular Parasites in a Morbidly Obese Population: A Neglected Trinomial. *Nutrients*, 14(15), p. 3211.

Cavalier-Smith, T., Fiore-Donno, A.M., Chao, E., Kudryavtsev, A., Berney, C., *et al.* (2015). Multigene phylogeny resolves deep branching of Amoebozoa. *Molecular Phylogenetics and Evolution*, 83, pp. 293–304.

Chabé, M., Lokmer, A. and Ségurel, L. (2017). Gut Protozoa: Friends or Foes of the Human Gut Microbiota? *Trends in Parasitology*, 33(12), pp. 925–934.

Chao, A. and Chiu, C.-H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution*, 7(8), pp. 919–928.

Chowdhury, N. and Aguirre, A.A. (2001). *Helminths of wildlife.* Science Publishers, Inc.

Christaki, U., Courties, C., Massana, R., Catala, P., Lebaron, P., *et al.* (2011). Optimized routine flow cytometric enumeration of heterotrophic flagellates using SYBR Green I. *Limnology and Oceanography: Methods*, 9(8), pp. 329–339.

Chudnovskiy, A., Mortha, A., Kana, V., Kennard, A., Ramirez, J.D., *et al.* (2016). Host-Protozoan Interactions Protect from Mucosal Infections through Activation of the Inflammasome. *Cell*, 167(2), pp. 444–456.

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., *et al.* (2012). Gut Immune Maturation Depends on Colonization with a Host-Specific Microbiota. *Cell*, 149(7), pp. 1578–1593.

Clerc, M., Devevey, G., Fenton, A. and Pedersen, A.B. (2018). Antibodies and coinfection drive variation in nematode burdens in wild mice. *International Journal for Parasitology*, 48(9), pp. 785–792.

Clerc, M., Babayan, S.A., Fenton, A. and Pedersen, A.B. (2019a). Age affects antibody levels and anthelmintic treatment efficacy in a wild rodent. *International Journal for Parasitology: Parasites and Wildlife*, 8, pp. 240–247.

Clerc, M., Fenton, A., Babayan, S.A. and Pedersen, A.B. (2019b). Parasitic nematodes simultaneously suppress and benefit from coccidian coinfection in their natural mouse host. *Parasitology*, 146(8), pp. 1096–1106.

Collántes-Fernández, E., Fort, M.C., Ortega-Mora, L.M. and Schares, G. (2018). Trichomonas. in M. Florin-Christensen and L. Schnittger (eds) *Parasitic Protozoa of Farm Animals and Pets*. Cham: Springer International Publishing, pp. 313–388.

Collier, C.T., Hofacre, C.L., Payne, A.M., Anderson, D.B., Kaiser, P., *et al.* (2008). Coccidia-induced mucogenesis promotes the onset of necrotic enteritis by supporting Clostridium perfringens growth. *Veterinary Immunology and Immunopathology*, 122(1), pp. 104–115.

Compton, S.R. (2020). PCR and RT-PCR in the Diagnosis of Laboratory Animal Infections and in Health Monitoring. *Journal of the American Association for Laboratory Animal Science*, 59(5), pp. 458–468.

Cooper, G.L., Charlton, B.R., Bickford, A.A. and Nordhausen, R. (2004). Hexamita meleagridis (Spironucleus meleagridis) Infection in Chukar Partridges Associated with High Mortality and Intracellular Trophozoites. *Avian Diseases*, 48(3), pp. 706–710.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., *et al.* (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science*, 326(5960), pp. 1694–1697.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J.M. and Relman, D.A. (2012). The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science*, 336(6086), pp. 1255–1262.

Coyte, K.Z. and Rakoff-Nahoum, S. (2019). Understanding Competition and Cooperation within the Mammalian Gut Microbiome. *Current Biology*, 29(11), pp. R538–R544.

Coyte, K.Z., Rao, C., Rakoff-Nahoum, S. and Foster, K.R. (2021). Ecological rules for the assembly of microbiome communities. *PLOS Biology*, 19(2), p. e3001116.

Crowther, R.S. and Wetmore, R.F. (1987). Fluorometric assay of O-linked glycoproteins by reaction with 2-cyanoacetamide. *Analytical Biochemistry*, 163(1), pp. 170–174.

Cui, Z., Li, J., Chen, Y. and Zhang, L. (2019). Molecular epidemiology, evolution, and phylogeny of Entamoeba spp. *Infection, Genetics and Evolution*, 75, p. 104018.

Davey, M.L., Utaaker, K.S. and Fossøy, F. (2021). Characterizing parasitic nematode faunas in faeces and soil using DNA metabarcoding. *Parasites & Vectors*, 14, p. 422.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., *et al.* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484), pp. 559–563.

De Palma, G., Lynch, M.D.J., Lu, J., Dang, V.T., Deng, Y., *et al.* (2017). Transplantation of fecal microbiota from patients with irritable bowel syndrome alters gut function and behavior in recipient mice. *Science Translational Medicine*, 9(379), p. eaaf6397.

Dearing, M.D. and Kohl, K.D. (2017). Beyond Fermentation: Other Important Services Provided to Endothermic Herbivores by their Gut Microbiota. *Integrative and Comparative Biology*, 57(4), pp. 723–731.

Dehority, B.A. (1986). Protozoa of the digestive tract of herbivorous mammals. *International Journal of Tropical Insect Science*, 7(Special Issue 3), pp. 279–296.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., *et al.* (2014). Convergence of gut microbiomes in myrmecophagous mammals. *Molecular Ecology*, 23(6), pp. 1301–1317.

Deng, L., Wojciech, L., Gascoigne, N.R.J., Peng, G. and Tan, K.S.W. (2021). New insights into the interactions between Blastocystis, the gut microbiota, and host immunity. *PLoS Pathogens*, 17(2), p. e1009253.

Deng, L., Wojciech, L., Png, C.W., Koh, E.Y., Aung, T.T., *et al.* (2022). Experimental colonization with Blastocystis ST4 is associated with protective immune responses and modulation of gut microbiome in a DSS-induced colitis mouse model. *Cellular and Molecular Life Sciences*, 79, p. 245.

Deng, L. and Tan, K.S.W. (2022). Interactions between Blastocystis subtype ST4 and gut microbiota in vitro. *Parasites & Vectors*, 15, p. 80.

Derelle, R., López-García, P., Timpano, H. and Moreira, D. (2016). A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Molecular biology and evolution*, 33(11), pp. 2890–2898.

Derrickson, E.M. (1992). Comparative Reproductive Strategies of Altricial and Precocial Eutherian Mammals. *Functional Ecology*, 6(1), pp. 57–65.

Desselberger, U. (2018). The Mammalian Intestinal Microbiome: Composition, Interaction with the Immune System, Significance for Vaccine Efficacy, and Potential for Disease Therapy. *Pathogens*, 7(3), p. 57.

Dobell, C. (1935). Researches on the Intestinal Protozoa of Monkeys and Man: VII. On The Enteromonas of Macaques and Embadomonas Intestinalis. *Parasitology*, 27(4), pp. 564–592.

Dollive, S., Chen, Y.-Y., Grunberg, S., Bittinger, K., Hoffmann, C., *et al.* (2013). Fungi of the Murine Gut: Episodic Variation and Proliferation during Antibiotic Treatment. *PLOS ONE*, 8(8), p. e71806.

Dolnik, O.V., Dolnik, V.R. and Bairlein, F. (2010). The Effect of Host Foraging Ecology on the Prevalence and Intensity of Coccidian Infection in Wild Passerine Birds. *Ardea*, 98(1), pp. 97–103.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., *et al.* (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107(26), pp. 11971–11975.

Donaldson, G.P., Lee, S.M. and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, 14(1), pp. 20–32.

Douglas, A.E. (2014). Symbiosis as a General Principle in Eukaryotic Evolution. *Cold Spring Harbor Perspectives in Biology*, 6(2), p. a016113.

Dubey, J.P. and Almeria, S. (2019). Cystoisospora belli infections in humans: the past 100 years. *Parasitology*, 146(12), pp. 1490–1527.

Dubik, M., Pilecki, B. and Moeller, J.B. (2022). Commensal Intestinal Protozoa—Underestimated Members of the Gut Microbial Community. *Biology*, 11(12), p. 1742.

Dumètre, A., Aubert, D., Puech, P.-H., Hohweyer, J., Azas, N., *et al.* (2012). Interaction Forces Drive the Environmental Transmission of Pathogenic Protozoa. *Applied and Environmental Microbiology*, 78(4), pp. 905–912.

Duszynski, D.W. (2021). Biodiversity of the Coccidia (Apicomplexa: Conoidasida) in vertebrates: what we know, what we do not know, and what needs to be done. *Folia Parasitologica*, 68, p. 2021.001.

Duval, S. and Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), pp. 455–463.

Ebert, D. (2013). The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), pp. 623–643.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., *et al.* (2005). Diversity of the Human Intestinal Microbial Flora. *Science*, 308(5728), pp. 1635–1638.

Ehret, T., Torelli, F., Klotz, C., Pedersen, A.B. and Seeber, F. (2017). Translational Rodent Models for Research on Parasitic Protozoa—A Review of Confounders and Possibilities. *Frontiers in Cellular and Infection Microbiology*, 7, p. 238.

El Aidy, S., Hooiveld, G., Tremaroli, V., Bäckhed, F. and Kleerebezem, M. (2013). The gut microbiota and mucosal homeostasis. *Gut Microbes*, 4(2), pp. 118–124.

Erben, U., Loddenkemper, C., Doerfel, K., Spieckermann, S., Haller, D., *et al.* (2014). A guide to histomorphological evaluation of intestinal inflammation in mouse models. *International Journal of Clinical and Experimental Pathology*, 7(8), pp. 4557–4576.

Escalante, N.K., Lemire, P., Cruz Tleugabulova, M., Prescott, D., Mortha, A., *et al.* (2016). The common mouse protozoa Tritrichomonas muris alters mucosal T cell homeostasis and colitis susceptibility. *The Journal of Experimental Medicine*, 213(13), pp. 2841–2850.

Evering, T. and Weiss, L.M. (2006). The immunology of parasite infections in immunocompromised hosts. *Parasite immunology*, 28(11), pp. 549–565.

Ezenwa, V.O. (2004). Host social behavior and parasitic infection: a multifactorial approach. *Behavioral Ecology*, 15(3), pp. 446–454.

Ezenwa, V.O., Archie, E.A., Craft, M.E., Hawley, D.M., Martin, L.B., *et al.* (2016). Host behaviour–parasite feedback: an essential link between animal behaviour and disease ecology. *Proceedings of the Royal Society B: Biological Sciences*, 283(1828), p. 20153078.

Fabre, P.-H., Hautier, L., Dimitrov, D. and P Douzery, E.J. (2012). A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology*, 12, p. 88.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., *et al.* (2016). Population-level analysis of gut microbiome variation. *Science*, 352(6285), pp. 560–564.

Ferreira, S.C.M., Veiga, M.M., Hofer, H., East, M.L. and Czirják, G.Á. (2021). Noninvasively measured immune responses reflect current parasite infections in a wild carnivore and are linked to longevity. *Ecology and Evolution*, 11(12), pp. 7685–7699.

Field, S.G. and Michiels, N.K. (2005). Parasitism and growth in the earthworm Lumbricus terrestris: fitness costs of the gregarine parasite Monocystis sp. *Parasitology*, 130(4), pp. 397–403.

Fiers, W.D., Gao, I.H. and Iliev, I.D. (2019). Gut Mycobiota Under Scrutiny: Fungal Symbionts or Environmental Transients? *Current Opinion in Microbiology*, 50, pp. 79–86.

Filyk, H.A. and Osborne, L.C. (2016). The Multibiome: The Intestinal Ecosystem's Influence on Immune Homeostasis, Health, and Disease. *eBioMedicine*, 13, pp. 46–54.

Fitzpatrick, C.R., Toor, I. and Holmes, M.M. (2022). Colony but not social phenotype or status structures the gut bacteria of a eusocial mammal. *Behavioral Ecology and Sociobiology*, 76(8), p. 117.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R. and White, B.A. (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology*, 6(2), pp. 121–131.

Flowerdew, J.R., Shore, R.F., Poulton, S.M.C. and Sparks, T.H. (2003). Live trapping to monitor small mammals in Britain. *Mammal Review*, 34(1), pp. 31-50.

Foissner, W., Chao, A. and Katz, L.A. (2008). Diversity and geographic distribution of ciliates (Protista: Ciliophora). *Biodiversity and Conservation*, 17(2), pp. 345–363.

Fortuna, M., João Sousa, M., Côrte-Real, M., Leão, C., Salvador, A., *et al.* (2000). Cell Cycle Analysis of Yeasts. *Current Protocols in Cytometry*, 13(1), p. 11.13.1-11.13.9.

Foster, J.C., Glass, M.D., Courtney, P.D. and Ward, L.A. (2003). Effect of Lactobacillus and Bifidobacterium on Cryptosporidium parvum oocyst viability. *Food Microbiology*, 20(3), pp. 351–357.

Foster, K.R., Schluter, J., Coyte, K.Z. and Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. *Nature*, 548(7665), pp. 43–51.

Fox, J.G., Beck, P., Dangler, C.A., Whary, M.T., Wang, T.C., *et al.* (2000). Concurrent enteric helminth infection modulates inflammation and gastric immune responses and reduces helicobacter-induced gastric atrophy. *Nature Medicine*, 6(5), pp. 536–542.

Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., *et al.* (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology*, 13(6), pp. 360–372.

Frau, A., Kenny, J.G., Lenzi, L., Campbell, B.J., Ijaz, U.Z., *et al.* (2019). DNA extraction and amplicon production strategies deeply inf luence the outcome of gut mycobiome studies. *Scientific Reports*, 9, p. 9328.

Friant, S., Ziegler, T.E. and Goldberg, T.L. (2016). Primate reinfection with gastrointestinal parasites: behavioural and physiological predictors of parasite acquisition. *Animal behaviour*, 117, pp. 105–113.

Füssy, Z., Vinopalová, M., Treitli, S.C., Pánek, T., Smejkalová, P., *et al.* (2021). Retortamonads from vertebrate hosts share features of anaerobic metabolism and pre-adaptations to parasitism with diplomonads. *Parasitology International*, 82, p. 102308.

Galazzo, G., van Best, N., Benedikter, B.J., Janssen, K., Bervoets, L., *et al.* (2020). How to Count Our Microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. *Frontiers in Cellular and Infection Microbiology*, 10, p. 403.

Ganoe, L.S., Brown, J.D., Yabsley, M.J., Lovallo, M.J. and Walter, W.D. (2020). A Review of Pathogens, Diseases, and Contaminants of Muskrats (Ondatra zibethicus) in North America. *Frontiers in Veterinary Science*, 7.

Gensollen, T., Iyer, S.S., Kasper, D.L. and Blumberg, R.S. (2016). How colonization by microbiota in early life shapes the immune system. *Science*, 352(6285), pp. 539–544.

Geuking, M.B., Cahenzli, J., Lawson, M.A.E., Ng, D.C.K., Slack, E., *et al.* (2011). Intestinal Bacterial Colonization Induces Mutualistic Regulatory T Cell Responses. *Immunity*, 34(5), pp. 794–806.

Gibson, K.M., Nguyen, B.N., Neumann, L.M., Miller, M., Buss, P., *et al.* (2019). Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health. *Scientific Reports*, 9, p. 7570.

Gilbert, K.A. (1997). Red howling monkey use of specific defecation sites as a parasite avoidance strategy. *Animal Behaviour*, 54(2), pp. 451–455.

Gilbert, S.F. (2020). Developmental symbiosis facilitates the multiple origins of herbivory. *Evolution & Development*, 22(1–2), pp. 154–164.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. and Egozcue, J.J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, 8, p. 2224.

Goertz, S., Menezes, A.B. de, Birtles, R.J., Fenn, J., Lowe, A.E., *et al.* (2019). Geographical location influences the composition of the gut microbiota in wild house mice (Mus musculus domesticus) at a fine spatial scale. *PLOS ONE*, 14(9), p. e0222501.

Gogarten, J.F., Calvignac-Spencer, S., Nunn, C.L., Ulrich, M., Saiepour, N., *et al.* (2020). Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Molecular Ecology Resources*, 20(1), pp. 204–215.

Gong, W. and Marchetti, A. (2019). Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. *Frontiers in Marine Science*, 6, p. 219.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 326(1233), pp. 119–157.

Graham, A.L. (2021). Naturalizing mouse models for immunology. *Nature Immunology*, 22(2), pp. 111–117.

Green, S.J. and Minz, D. (2005). Suicide Polymerase Endonuclease Restriction, a Novel Technique for Enhancing PCR Amplification of Minor DNA Templates. *Applied and Environmental Microbiology*, 71(8), pp. 4721–4727.

Grieneisen, L., Blekhman, R. and Archie, E. (2023). How longitudinal data can contribute to our understanding of host genetic effects on the gut microbiome. *Gut Microbes*, 15(1), p. 2178797.

Grieneisen, L.E., Livermore, J., Alberts, S., Tung, J. and Archie, E.A. (2017). Group Living and Male Dispersal Predict the Core Gut Microbiome in Wild Baboons. *Integrative and Comparative Biology*, 57(4), pp. 770–785.

Grieneisen, L.E., Charpentier, M.J.E., Alberts, S.C., Blekhman, R., Bradburd, G., *et al.* (2019). Genes, geology and germs: gut microbiota across a primate hybrid zone are explained by site soil properties, not host species. *Proceedings of the Royal Society B: Biological Sciences*, 286(1901), p. 20190431.

Groussin, M., Mazel, F. and Alm, E.J. (2020). Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host & Microbe*, 28(1), pp. 12–22.

Guardone, L., Armani, A., Mancianti, F. and Ferroglio, E. (2022). A Review on Alaria alata, Toxoplasma gondii and Sarcocystis spp. in Mammalian Game Meat Consumed in Europe: Epidemiology, Risk Management and Future Directions. *Animals*, 12(3), p. 263.

Gupta, S., Mortensen, M.S., Schjørring, S., Trivedi, U., Vestergaard, G., *et al.* (2019). Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology*, 2, p. 291.

Gupta, Y., Ernst, A.L., Vorobyev, A., Beltsiou, F., Zillikens, D., *et al.* (2023). Impact of diet and host genetics on the murine intestinal mycobiome. *Nature Communications*, 14, p. 834.

Gutzeit, C., Magri, G. and Cerutti, A. (2014). Intestinal IgA production and its role in host-microbe interaction. *Immunological reviews*, 260(1), pp. 76–85.

Guzzo, G.L., Andrews, J.M. and Weyrich, L.S. (2022). The Neglected Gut Microbiome: Fungi, Protozoa, and Bacteriophages in Inflammatory Bowel Disease. *Inflammatory Bowel Diseases*, 28(7), pp. 1112–1122.

Haak, B.W., Argelaguet, R., Kinsella, C.M., Kullberg, R.F.J., Lankelma, J.M., *et al.* (2021). Integrative Transkingdom Analysis of the Gut Microbiome in Antibiotic Perturbation and Critical Illness. *mSystems*, 6(2), pp. e01148-20.

Habig, B., Doellman, M.M., Woods, K., Olansen, J. and Archie, E.A. (2018). Social status and parasitism in male and female vertebrates: a meta-analysis. *Scientific Reports*, 8, p. 3629.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., *et al.* (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLOS ONE*, 9(2), p. e87624.

Hallen-Adams, H.E., Kachman, S.D., Kim, J., Legge, R.M. and Martínez, I. (2015). Fungi inhabiting the healthy human gastrointestinal tract: a diverse and dynamic community. *Fungal Ecology*, 15, pp. 9–17.

Hallen-Adams, H.E. and Suhr, M.J. (2017). Fungi in the healthy human gastrointestinal tract. *Virulence*, 8(3), pp. 352–358.

Hamad, I., Raoult, D. and Bittar, F. (2016). Repertory of eukaryotes (eukaryome) in the human gastrointestinal tract: taxonomy and detection methods. *Parasite Immunology*, 38(1), pp. 12–36.

Hammer, T.J., Sanders, J.G. and Fierer, N. (2019). Not all animals need a microbiome. *FEMS Microbiology Letters*, 366(10), p. fnz117.

Han, B.A., Schmidt, J.P., Bowden, S.E. and Drake, J.M. (2015). Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences*, 112(22), pp. 7039–7044.

Harrison, X.A., McDevitt, A.D., Dunn, J.C., Griffiths, S.M., Benvenuto, C., *et al.* (2021). Fungal microbiomes are determined by host phylogeny and exhibit widespread associations with the bacterial microbiome. *Proceedings of the Royal Society B: Biological Sciences*, 288(1957), p. 20210552.

den Hartog, J., Rosenbaum, L., Wood, Z., Burt, D. and Petri, W.A. (2013). Diagnosis of Multiple Enteric Protozoan Infections by Enzyme-Linked Immunosorbent Assay in the Guatemalan Highlands. *The American Journal of Tropical Medicine and Hygiene*, 88(1), pp. 167–171.

Hasnain, S.Z., Gallagher, A.L., Grencis, R.K. and Thornton, D.J. (2013). A new role for mucins in immunity: Insights from gastrointestinal nematode infection. *The International Journal of Biochemistry & Cell Biology*, 45(2), pp. 364-374.

Hawash, Y. (2014). DNA Extraction from Protozoan Oocysts/Cysts in Feces for Diagnostic PCR. *The Korean Journal of Parasitology*, 52(3), pp. 263–271.

Hawley, D.M., Gibson, A.K., Townsend, A.K., Craft, M.E. and Stephenson, J.F. (2021). Bidirectional interactions between host social behaviour and parasites arise through ecological and evolutionary processes. *Parasitology*, 148(3), pp. 274–288.

Hayes, K.S., Bancroft, A.J., Goldrick, M., Portsmouth, C., Roberts, I.S., *et al.* (2010). Exploitation of the Intestinal Microflora by the Parasitic Nematode Trichuris muris. *Science*, 328(5984), pp. 1391–1394.

Heijtz, R.D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., *et al.* (2011). Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences*, 108(7), pp. 3047–3052.

Heinrichs, M.E., De Corte, D., Engelen, B. and Pan, D. (2021). An Advanced Protocol for the Quantification of Marine Sediment Viruses via Flow Cytometry. *Viruses*, 13(1), p. 102.

Heisel, T., Montassier, E., Johnson, A., Al-Ghalith, G., Lin, Y.-W., *et al.* (2017). High-Fat Diet Changes Fungal Microbiomes and Interkingdom Relationships in the Murine Gut. *mSphere*, 2(5), pp. e00351-17.

Heitlinger, E., Ferreira, S.C.M., Thierer, D., Hofer, H. and East, M.L. (2017). The Intestinal Eukaryotic and Bacterial Biome of Spotted Hyenas: The Impact of Social Status and Age on Diversity and Composition. *Frontiers in Cellular and Infection Microbiology*, 7, p. 262.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., *et al.* (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports*, 5, p. 14567.

Henry, L.P., Bruijning, M., Forsberg, S.K.G. and Ayroles, J.F. (2021). The microbiome extends host evolutionary potential. *Nature Communications*, 12, p. 5141.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.)*, 327(7414), pp. 557–560.

Hillegass, M.A., Waterman, J.M. and Roth, J.D. (2008). The influence of sex and sociality on parasite loads in an African ground squirrel. *Behavioral Ecology*, 19(5), pp. 1006–1011.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., *et al.* (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, 112(41), pp. 12764–12769.

Ho, P.-Y., Good, B.H. and Huang, K.C. (2022). Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *eLife*, 11, p. e75168.

Hoarau, G., Mukherjee, P.K., Gower-Rousseau, C., Hager, C., Chandra, J., *et al.* (2016). Bacteriome and Mycobiome Interactions Underscore Microbial Dysbiosis in Familial Crohn's Disease. *mBio*, 7(5), pp. e01250-16.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., *et al.* (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLOS ONE*, 8(6), p. e66019.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), pp. 65–70.

Hooks, K.B. and O'Malley, M.A. (2020). Contrasting Strategies: Human Eukaryotic Versus Bacterial Microbiome Research. *Journal of Eukaryotic Microbiology*, 67(2), pp. 279–295.

Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift*, 50(3), pp. 346–363.

Huang, G., Zhang, S., Zhou, C., Tang, X., Li, C., *et al.* (2018). Influence of Eimeria falciformis Infection on Gut Microbiota and Metabolic Pathways in Mice. *Infection and Immunity*, 86(5), pp. e00073-18.

Huffnagle, G.B. and Noverr, M.C. (2013). The emerging world of the fungal microbiome. *Trends in Microbiology*, 21(7), pp. 334–341.

Huh, J.-W., Moon, S.-G. and Lim, Y.-H. (2009). A Survey of Intestinal Protozoan Infections among Gastroenteritis Patients during a 3-Year Period (2004-2006) in Gyeonggi-do (Province), South Korea. *The Korean Journal of Parasitology*, 47(3), pp. 303–305.

Hume, M.E., Hernandez, C.A., Barbosa, N.A., Sakomura, N.K., Dowd, S.E., *et al.* (2012). Molecular Identification and Characterization of Ileal and Cecal Fungus Communities in Broilers Given Probiotics,

Specific Essential Oil Blends, and Under Mixed Eimeria Infection. *Foodborne Pathogens and Disease*, 9(9), pp. 853–860.

Hurst, J.L. and Berreen, J. (1985). Observations on the trap-response of wild house mice, *Mus domesticus* Rutty, in poultry houses*, Journal of Zoology,* 207(4), pp. 619-622.

Iliev, I.D., Funari, V.A., Taylor, K.D., Nguyen, Q., Reyes, C.N., *et al.* (2012). Interactions Between Commensal Fungi and the C-Type Lectin Receptor Dectin-1 Influence Colitis. *Science*, 336(6086), pp. 1314–1317.

Imai, S. and Ogimoto, K. (1988). Flagellate protozoa in the digestive tract of the Japanese field vole, Microtus montebelli. *Japanese Journal of Zootechnical Science*, 59(4), pp. 351–356.

IntHout, J., Ioannidis, J.P.A., Rovers, M.M. and Goeman, J.J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7), p. e010247.

Jackson, G.A., Livingston, R.S., Riley, L.K., Livingston, B.A. and Franklin, C.L. (2013). Development of a PCR Assay for the Detection of Spironucleus muris. *Journal of the American Association for Laboratory Animal Science*, 52(2), pp. 165–170.

Jackson, J.A., Antwis, R.E., Beresford, N.A. and Wood, M.D. (2022). Some observations on meaningful and objective inference in radioecological field studies. *Journal of Animal Ecology*, 91(7), pp. 1546–1553.

Jackson, M.A., Pearson, C., Ilott, N.E., Huus, K.E., Hegazy, A.N., *et al.* (2021). Accurate identification and quantification of commensal microbiota bound by host immunoglobulins. *Microbiome*, 9, p. 33.

Jakobsson, H.E., Rodríguez-Piñeiro, A.M., Schütte, A., Ermund, A., Boysen, P., *et al.* (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Reports,* 16, pp. 164-177.

Janda, J.M. and Abbott, S.L. (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology*, 45(9), pp. 2761–2764.

Jarquín-Díaz, V.H., Balard, A., Jost, J., Kraft, J., Dikmen, M.N., *et al.* (2019). Detection and quantification of house mouse Eimeria at the species level – Challenges and solutions for the assessment of coccidia in wildlife. *International Journal for Parasitology: Parasites and Wildlife*, 10, pp. 29–40.

Jarquín-Díaz, V.H., Balard, A., Ferreira, S.C.M., Mittné, V., Murata, J.M., *et al.* (2022). DNA-based quantification and counting of transmission stages provides different but complementary parasite load estimates: an example from rodent coccidia (Eimeria). *Parasites & Vectors*, 15, p. 45.

Jawhara, S., Thuru, X., Standaert-Vitse, A., Jouault, T., Mordon, S., *et al.* (2008). Colonization of mice by *Candida albicans* is promoted by chemically induced colitis and augments inflammatory responses through galectin-3. *The Journal of Infectious Diseases,* 197(7), pp. 972-980.

Jenkins, T.P., Formenti, F., Castro, C., Piubelli, C., Perandin, F., *et al.* (2018). A comprehensive analysis of the faecal microbiome and metabolome of Strongyloides stercoralis infected volunteers from a non-endemic area. *Scientific Reports*, 8, p. 15651.

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A. and Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLOS ONE*, 15(1), p. e0227285.

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., *et al.* (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10, p. 5029.

Johnson, K.V.-A. and Burnet, P.W.J. (2016). Microbiome: Should we diversify from diversity? *Gut Microbes*, 7(6), pp. 455–458.

Jørgensen, A. and Sterud, E. (2007). Phylogeny of Spironucleus (Eopharyngia: Diplomonadida: Hexamitinae). *Protist*, 158(2), pp. 247–254.

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), pp. 2427–2439.

Kabwe, M.H., Vikram, S., Mulaudzi, K., Jansson, J.K. and Makhalanyane, T.P. (2020). The gut mycobiota of rural and urban individuals is shaped by geography. *BMC Microbiology*, 20, p. 257.

Kamada, N., Kim, Y.-G., Sham, H.P., Vallance, B.A., Puente, J.L., *et al.* (2012). Regulated Virulence Controls the Ability of a Pathogen to Compete with the Gut Microbiota. *Science*, 336(6086), pp. 1325–1329.

Kataoka, T. and Kondo, R. (2019). Protistan community composition in anoxic sediments from three salinity-disparate Japanese lakes. *Estuarine, Coastal and Shelf Science*, 224, pp. 34–42.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), pp. 3059–3066.

Keeling, P.J. and Brugerolle, G. (2006). Evidence from SSU rRNA phylogeny that Octomitus is a sister lineage to Giardia. *Protist*, 157(2), pp. 205–212.

Kelly, C.D., Stoehr, A.M., Nunn, C., Smyth, K.N. and Prokop, Z.M. (2018). Sexual dimorphism in immunity across animals: a meta-analysis. *Ecology Letters*, 21(12), pp. 1885–1894.

Kidd, S.E., Abdolrasouli, A. and Hagen, F. (2023). Fungal Nomenclature: Managing Change is the Name of the Game. *Open Forum Infectious Diseases*, 10(1), p. ofac559.

Kieser, S., Zdobnov, E.M. and Trajkovski, M. (2022). Comprehensive mouse microbiota genome catalog reveals major difference to its human counterpart. *PLOS Computational Biology*, 18(3), p. e1009947.

Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., *et al.* (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*, 5(1), p. 52.

Kim, S.L., Choi, J.H., Yi, M., Lee, S., Kim, M., *et al.* (2022). Metabarcoding of bacteria and parasites in the gut of Apodemus agrarius. *Parasites & Vectors*, 15, p. 486.

Kim, T.-H., Han, J.-H., Chang, S.-N., Kim, D.-S., Abdelkader, T.S., *et al.* (2011). Detection of sarcocystic infection in a wild rodent (Apodemus agrarius chejuensis) captured on Jeju island. *Laboratory Animal Research*, 27(4), pp. 357–359.

King, B.J. and Monis, P.T. (2007). Critical processes affecting Cryptosporidium oocyst survival in the environment. *Parasitology*, 134(3), pp. 309–323.

Knowles, S.C.L., Fenton, A., Petchey, O.L., Jones, T.R., Barber, R., *et al.* (2013). Stability of within-host–parasite communities in a wild mammal system. *Proceedings of the Royal Society B: Biological Sciences*, 280(1762), p. 20130598.

Kohl, K.D., Weiss, R.B., Cox, J., Dale, C. and Denise Dearing, M. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecology Letters*, 17(10), pp. 1238–1246.

Kohl, K.D. and Dearing, M.D. (2014). Wild-caught rodents retain a majority of their natural gut microbiota upon entrance into captivity. *Environmental Microbiology Reports*, 6(2), pp. 191–195.

Koleff, P., Gaston, K.J. and Lennon, J.J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, 72(3), pp. 367–382.

Kolisko, M., Cepicka, I., Hampl, V., Leigh, J., Roger, A.J., *et al.* (2008). Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: Implications for the evolutionary history of the double karyomastigont of diplomonads. *BMC Evolutionary Biology*, 8, p. 205.

Kołodziej-Sobocińska, M. (2019). Factors affecting the spread of parasites in populations of wild European terrestrial mammals. *Mammal Research*, 64(3), pp. 301–318.

Koricheva, J., Gurevitch, J. and Mengersen, K. (2013). *Handbook of Meta-analysis in Ecology and Evolution*. Princeton University Press.

Korpela, K., Costea, P., Coelho, L.P., Kandels-Lewis, S., Willemsen, G., *et al.* (2018). Selective maternal seeding and environment shape the human gut microbiome. *Genome Research*, 28(4), pp. 561–568.

Kounosu, A., Murase, K., Yoshida, A., Maruyama, H. and Kikuchi, T. (2019). Improved 18S and 28S rDNA primer sets for NGS-based parasite detection. *Scientific Reports*, 9, p. 15789.

Kreisinger, J., Čížková, D., Vohánka, J. and Piálek, J. (2014). Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Molecular Ecology*, 23(20), pp. 5048–5060.

Kreisinger, J., Bastien, G., Hauffe, H.C., Marchesi, J. and Perkins, S.E. (2015). Interactions between multiple helminths and the gut microbiota in wild rodents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1675), p. 20140295.

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., *et al.* (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Computational Biology*, 11(5), p. e1004226.

Kuziel, G.A. and Rakoff-Nahoum, S. (2022). The gut microbiome. *Current Biology*, 32(6), pp. R257–R264.

Kwon, H.-K. and Seong, J.K. (2021). New insights into the microbiota of wild mice. *Mammalian Genome*, 32(4), pp. 311–318.

Labocha, M.K., Schutz, H. and Hayes, J.P. (2014). Which body condition index is best? *Oikos*, 123(1), pp. 111–119.

Laforest-Lapointe, I. and Arrieta, M.-C. (2018). Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies. *mSystems*, 3(2), pp. e00201-17.

Lahti, L., and Shetty, S. (2017). Tools for microbiome analysis in R. R package version 1.19.1. https://microbiome.github.io/tutorials/

Langda, S., Zhang, C., Zhang, K., Gui, B., Ji, D., *et al.* (2020). Diversity and Composition of Rumen Bacteria, Fungi, and Protozoa in Goats and Sheep Living in the Same High-Altitude Pasture. *Animals*, 10(2), p. 186.

Lau, Y.L., Jamaiah, I., Rohela, M., Fong, M.Y., Siti, C.O.S., *et al.* (2014). Molecular detection of Entamoeba histolytica and Entamoeba dispar infection among wild rats in Kuala Lumpur, Malaysia. *Tropical Biomedicine*, 31(4), pp. 721–727.

Lavrinienko, A., Scholier, T., Bates, S.T., Miller, A.N. and Watts, P.C. (2021a). Defining gut mycobiota for wild animals: a need for caution in assigning authentic resident fungal taxa. *Animal Microbiome*, 3, p. 75.

Lavrinienko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M. and Watts, P.C. (2021b). Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? *Trends in Microbiology*, 29(1), pp. 19–27.

LeBlanc, J.G., Chain, F., Martín, R., Bermúdez-Humarán, L.G., Courau, S., *et al.* (2017). Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microbial Cell Factories*, 16(1), p. 79.

Lepczyńska, M., Białkowska, J., Dzika, E., Piskorz-Ogórek, K. and Korycińska, J. (2017). Blastocystis: how do specific diets and human gut microbiota affect its development and pathogenicity? *European Journal of Clinical Microbiology & Infectious Diseases*, 36(9), pp. 1531–1540.

Lepere, C., Demura, M., Kawachi, M., Romac, S., Probert, I., *et al.* (2011). Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiology Ecology*, 76(3), pp. 513–523.

Leung, J.M., Graham, A.L. and Knowles, S.C.L. (2018a). Parasite-Microbiota Interactions With the Vertebrate Gut: Synthesis Through an Ecological Lens. *Frontiers in Microbiology*, 9, p. 843.

Leung, J.M., Budischak, S.A., The, H.C., Hansen, C., Bowcutt, R., *et al.* (2018b). Rapid environmental effects on gut nematode susceptibility in rewilded mice. *PLOS Biology*, 16(3), p. e2004108.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., *et al.* (2008). Evolution of Mammals and Their Gut Microbes. *Science*, 320(5883), pp. 1647–1651.

Li, J., Wang, H., Wang, R. and Zhang, L. (2017). Giardia duodenalis Infections in Humans and Other Animals in China. *Frontiers in Microbiology*, 8, p. 2004.

Li, J., Li, L., Jiang, H., Yuan, L., Zhang, L., *et al.* (2018). Fecal Bacteriome and Mycobiome in Bats with Diverse Diets in South China. *Current Microbiology*, 75(10), pp. 1352–1361.

Li, W.-C., Huang, J., Fang, Z., Ren, Q., Tang, L., *et al.* (2020). Prevalence of Tetratrichomonas buttreyi and Pentatrichomonas hominis in yellow cattle, dairy cattle, and water buffalo in China. *Parasitology Research*, 119(2), pp. 637–647.

Li, H., Qu, J., Li, T., Li, J., Lin, Q., *et al.* (2016). Pika population density is associated with composition and diversity of gut microbiota. *Frontiers in Microbiology,* 7, p. 758.

Lin, H. and Peddada, S.D. (2020). Analysis of compositions of microbiomes with bias correction. *Nature Communications*, 11, p. 3514.

Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, 13(9), p. e0204056.

Lin, L. and Xu, C. (2020). Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. *Health Science Reports*, 3(3), p. e178.

Lind, A.L. and Pollard, K.S. (2021). Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome*, 9, p. 58.

Linh, B.K., Hayashi, T. and Horii, Y. (2009). Eimeria vermiformis infection reduces goblet cells by multiplication in the crypt cells of the small intestine of C57BL/6 mice. *Parasitology Research*, 104(4), pp. 789–794.

Linnenbrink, M., Wang, J., Hardouin, E.A., Künzel, S., Metzler, D., *et al.* (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, 22(7), pp. 1904–1916.

Lofgren, L.A., Uehling, J.K., Branco, S., Bruns, T.D., Martin, F., *et al.* (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology*, 28(4), pp. 721–730.

Lokmer, A., Cian, A., Froment, A., Gantois, N., Viscogliosi, E., *et al.* (2019). Use of shotgun metagenomics for the identification of protozoa in the gut microbiota of healthy individuals from worldwide populations with various industrialization levels. *PLOS ONE*, 14(2), p. e0211139.

Long, P.L. and Joyner, L.P. (1984). Problems in the Identification of Species of Eimeria. *The Journal of Protozoology*, 31(4), pp. 535–541.

Louca, S., Doebeli, M. and Parfrey, L.W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, 6, p. 41.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K. and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415), pp. 220–230.

Lozupone, C.A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., *et al.* (2013). Meta-analyses of studies of the human microbiota. *Genome Research*, 23(10), pp. 1704–1714.

Lu, C., Yan, Y., Jian, F. and Ning, C. (2021). Coccidia-Microbiota Interactions and Their Effects on the Host. *Frontiers in Cellular and Infection Microbiology*, 11, p. 751481.

Lukeš, J., Stensvold, C.R., Jirků-Pomajbíková, K. and Parfrey, L.W. (2015). Are Human Intestinal Eukaryotes Beneficial or Commensals? *PLOS Pathogens*, 11(8), p. e1005039.

MacGillivray, D.M. and Kollmann, T.R. (2014). The role of environmental factors in modulating immune responses in early life. *Frontiers in Immunology*, 5, p. 434.

Macpherson, A.J., Geuking, M.B. and McCoy, K.D. (2005). Immune responses that adapt the intestinal mucosa to commensal intestinal bacteria. *Immunology*, 115(2), pp. 153–162.

Malik, S.-B., Brochu, C.D., Bilic, I., Yuan, J., Hess, M., *et al.* (2011). Phylogeny of Parasitic Parabasalia and Free-Living Relatives Inferred from Conventional Markers vs. Rpb1, a Single-Copy Gene. *PLOS ONE*, 6(6), p. e20774.

Mallott, E.K. and Amato, K.R. (2021). Host specificity of the gut microbiome. *Nature Reviews Microbiology*, 19(10), pp. 639–653.

Malmberg, J.L., White, L.A. and VandeWoude, S. (2021). Bioaccumulation of Pathogen Exposure in Top Predators. *Trends in Ecology & Evolution*, 36(5), pp. 411–420.

Mandal, S., Van Treuren, W., White, R.A., Eggesbø, M., Knight, R., *et al.* (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, 26, p. 27663.

Mann, A.E., Mazel, F., Lemay, M.A., Morien, E., Billy, V., *et al.* (2020). Biodiversity of protists and nematodes in the wild nonhuman primate gut. *The ISME Journal*, 14(2), pp. 609–622.

Mar Rodríguez, M., Pérez, D., Javier Chaves, F., Esteve, E., Marin-Garcia, P., *et al.* (2015). Obesity changes the human gut mycobiome. *Scientific Reports*, 5, p. 14600.

Marchesi, J.R. (2010). Chapter 2 - Prokaryotic and Eukaryotic Diversity of the Human Gut. in A.I. Laskin, S. Sariaslani, and G.M. Gadd (eds) *Advances in Applied Microbiology*. Academic Press, pp. 43–62.

Marchesi, J.R. and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, 3, p. 31.

Marsh, K.J., Raulo, A.M., Brouard, M., Troitsky, T., English, H.M., *et al.* (2022). Synchronous Seasonality in the Gut Microbiota of Wild Mouse Populations. *Frontiers in Microbiology*, 13, p. 809735.

Martens-Habbena, W. and Sass, H. (2006). Sensitive Determination of Microbial Growth by Nucleic Acid Staining in Aqueous Suspension. *Applied and Environmental Microbiology*, 72(1), pp. 87–95.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), pp. 10–12.

Matchado, M.S., Lauber, M., Reitmeier, S., Kacprowski, T., Baumbach, J., *et al.* (2021). Network analysis methods for studying microbial communities: A mini review. *Computational and Structural Biotechnology Journal*, 19, pp. 2687–2698.

Mathison, B.A. and Sapp, S.G.H. (2021). An annotated checklist of the eukaryotic parasites of humans, exclusive of fungi and algae. *ZooKeys*, 1069, pp. 1–313.

Mathur, V., Kwong, W.K., Husnik, F., Irwin, N.A.T., Kristmundsson, Á., *et al.* (2021). Phylogenomics Identifies a New Major Subgroup of Apicomplexans, Marosporida class nov., with Extreme Apicoplast Genome Reduction. *Genome Biology and Evolution*, 13(2), p. evaa244.

Maurice, C.F., Haiser, H.J. and Turnbaugh, P.J. (2013). Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*, 152(1–2), pp. 39–50.

Maurice, C.F., CL Knowles, S., Ladau, J., Pollard, K.S., Fenton, A., *et al.* (2015). Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal*, 9(11), pp. 2423–2434.

McDonald, J.E., Marchesi, J.R. and Koskella, B. (2020). Application of ecological and evolutionary theory to microbiome community dynamics across systems. *Proceedings of the Royal Society B: Biological Sciences*, 287(1941), p. 20202886.

McGuckin, M.A., Lindén, S.K., Sutton, P. and Florin, T.H. (2011). Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology*, 9(4), pp. 265–278.

McHardy, I.H., Wu, M., Shimizu-Cohen, R., Couturier, M.R. and Humphries, R.M. (2014). Detection of Intestinal Protozoa in the Clinical Laboratory. *Journal of Clinical Microbiology*, 52(3), pp. 712–720.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., *et al.* (2012). Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits. *PLOS ONE*, 7(6), p. e39191.

McMurdie, P.J. and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8(4), p. e61217.

McMurdie, P.J. and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, 10(4), p. e1003531.

Meerburg, B.G., Singleton, G.R. and Kijlstra, A. (2009). Rodent-borne diseases and their risks for public health. *Critical Reviews in Microbiology*, 35(3), pp. 221–270.

Michaiowski, T. (2005). Chapter 3 Rumen protozoa in the growing domestic ruminant. in W.H. Holzapfel, P.J. Naughton, S.G. Pierzynowski, R. Zabielski, and E. Salek (eds) *Biology of Growing Animals*. Elsevier (Microbial Ecology in Growing Animals), pp. 54–74.

Michonneau, F., Brown, J.W. and Winter, D.J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), pp. 1476–1481.

Microfungi Collections Consortium. (2022). What are Microfungi? Available at: https://www.microfungi.org/table1 (Accessed: 16 April 2023).

Mims, T.S., Abdallah, Q.A., Stewart, J.D., Watts, S.P., White, C.T., *et al.* (2021). The gut mycobiome of healthy mice is shaped by the environment and correlates with metabolic outcomes in response to diet. *Communications Biology*, 4, p. 281.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., *et al.* (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research*, 21(10), pp. 1616–1625.

Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., *et al.* (2016a). Cospeciation of gut microbiota with hominids. *Science*, 353(6297), pp. 380–382.

Moeller, A.H., Foerster, S., Wilson, M.L., Pusey, A.E., Hahn, B.H., *et al.* (2016b). Social behavior shapes the chimpanzee pan-microbiome. *Science Advances*, 2(1), p. e1500997.

Moeller, A.H., Suzuki, T.A., Lin, D., Lacey, E.A., Wasser, S.K., *et al.* (2017). Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proceedings of the National Academy of Sciences*, 114(52), pp. 13768–13773.

Moeller, A.H., Suzuki, T.A., Phifer-Rixey, M. and Nachman, M.W. (2018). Transmission modes of the mammalian gut microbiota. *Science*, 362(6413), pp. 453–457.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., and PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), p. e1000097.

Montero, B.K., Wasimuddin, Schwensow, N., Gillingham, M.A.F., Ratovonamana, Y.R., *et al.* (2021). Evidence of MHC class I and II influencing viral and helminth infection via the microbiome in a non-human primate. *PLOS Pathogens*, 17(11), p. e1009675.

Moore, W.E.C., Cato, E.P. and Holdeman, L.V. (1969). Anaerobic Bacteria of the Gastrointestinal Flora and Their Occurrence in Clinical Infections. *The Journal of Infectious Diseases*, 119(6), pp. 641–649.

Morris, E.K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., *et al.* (2014). Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, 4(18), pp. 3514–3524.

Morrison, D.A. (2009). Evolution of the Apicomplexa: where are we now? *Trends in Parasitology*, 25(8), pp. 375–382.

Morton, J.T., Sanders, J., Quinn, R.A., McDonald, D., Gonzalez, A., *et al.* (2017). Balance Trees Reveal Microbial Niche Differentiation. *mSystems*, 2(1), pp. e00162-16.

Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., *et al.* (2019). Establishing microbial composition measurement standards with reference frames. *Nature Communications*, 10, p. 2719.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., *et al.* (2011). Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. *Science*, 332(6032), pp. 970–974.

Mueller, K.D., Zhang, H., Serrano, C.R., Billmyre, R.B., Huh, E.Y., *et al.* (2019). Gastrointestinal microbiota alteration induced by Mucor circinelloides in a murine model. *Journal of Microbiology*, 57(6), pp. 509–520.

Mueller, N.T., Bakacs, E., Combellick, J., Grigoryan, Z. and Dominguez-Bello, M.G. (2015). The infant microbiome development: mom matters. *Trends in molecular medicine*, 21(2), pp. 109–117.

Mulder, I.E., Schmidt, B., Lewis, M., Delday, M., Stokes, C.R., *et al.* (2011). Restricting Microbial Exposure in Early Life Negates the Immune Benefits Associated with Gut Colonization in Environments of High Microbial Diversity. *PLOS ONE*, 6(12), p. e28279.

Müller, S. and Nebe-von-Caron, G. (2010). Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiology Reviews*, 34(4), pp. 554–587.

Murillo, T., Schneider, D., Heistermann, M., Daniel, R. and Fichtel, C. (2022). Assessing the drivers of gut microbiome composition in wild redfronted lemurs via longitudinal metacommunity analysis. *Scientific Reports*, 12, p. 21462.

Nagendra, H. (2002). Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, 22(2), pp. 175–186.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), pp. 133–142.

Nakagawa, S., Lagisz, M., O'Dea, R.E., Rutkowska, J., Yang, Y., *et al.* (2021). The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods*, 12(1), pp. 4–12.

Nam, Y.-D., Chang, H.-W., Kim, K.-H., Roh, S.W., Kim, M.-S., *et al.* (2008). Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. *The Journal of Microbiology*, 46(5), pp. 491–501.

Nash, A.K., Auchtung, T.A., Wong, M.C., Smith, D.P., Gesell, J.R., *et al.* (2017). The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome*, 5, p. 153.

Newbold, C.J., de la Fuente, G., Belanche, A., Ramos-Morales, E. and McEwan, N.R. (2015). The Role of Ciliate Protozoa in the Rumen. *Frontiers in Microbiology*, 6, p. 1313.

Nkamga, V.D., Henrissat, B. and Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. *Human Microbiome Journal*, 3, pp. 1–8.

Novogene (2023). 16S/18S/ITS Amplicon Metagenomic Sequencing. Available at: https://www.novogene.com/eu-en/services/research-services/metagenome-sequencing/16s-18s-its-amplicon-metagenomic-sequencing/ (Accessed: 16 April 2023).

Nowell, F. and Higgs, S. (1989). Eimeria species infecting wood mice (genus Apodemus) and the transfer of two species to Mus musculus. *Parasitology*, 98(3), pp. 329–336.

Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B.N., Peeters, M., *et al.* (2010). Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *PLOS Biology*, 8(11), p. e1000546.

O'Hara, A.M. and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7), pp. 688–693.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., *et al.* (2020). vegan: Community Ecology Package. R package version 2.5.7. https://CRAN.R-project.org/package=vegan

Orpin, C.G. (1984). The role of ciliate protozoa and fungi in the rumen digestion of plant cell walls. *Animal Feed Science and Technology*, 10(2), pp. 121–143.

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15(12).

Ost, K.S., O'Meara, T.R., Stephens, W.Z., Chiaro, T., Zhou, H., *et al.* (2021). Adaptive immunity induces mutualism between commensal eukaryotes. *Nature*, 596(7870), pp. 114–118.

Ostfeld, R.S. and Mills, J.N. (2008). Chapter 41. Social Behavior, Demography, and Rodent-Borne Pathogens. in *Chapter 41. Social Behavior, Demography, and Rodent-Borne Pathogens*. University of Chicago Press, pp. 478–486.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., *et al.* (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, p. n71.

Palm, N.W., de Zoete, M.R., Cullen, T.W., Barry, N.A., Stefanowski, J., *et al.* (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*, 158(5), pp. 1000–1010.

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2), pp. 289–290.

Parfrey, L.W., Walters, W. and Knight, R. (2011). Microbial Eukaryotes in the Human Microbiome: Ecology, Evolution, and Future Directions. *Frontiers in Microbiology*, 2.

Parfrey, L.W., Walters, W.A., Lauber, C.L., Clemente, J.C., Berg-Lyons, D., *et al.* (2014). Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Frontiers in Microbiology*, 5.

Park, J.S. and Simpson, A.G.B. (2015). Characterization of a Deep-Branching Heterolobosean, Pharyngomonas turkanaensis n. sp., Isolated from a Non-Hypersaline Habitat, and Ultrastructural Comparison of Cysts and Amoebae Among Pharyngomonas Strains. *Journal of Eukaryotic Microbiology*, 63(1), pp. 100–111.

Paterson, M.J., Oh, S. and Underhill, D.M. (2017). Host–microbe interactions: commensal fungi in the gut. *Current Opinion in Microbiology*, 40, pp. 131–137.

Pedersen, A.B., Altizer, S., Poss, M., Cunningham, A.A. and Nunn, C.L. (2005). Patterns of host specificity and transmission among parasites of wild primates. *International Journal for Parasitology*, 35(6), pp. 647–657.

Pedersen, A.B. and Babayan, S.A. (2011). Wild immunology. *Molecular Ecology*, 20(5), pp. 872–880.

Pereira, F.C. and Berry, D. (2017). Microbial nutrient niches in the gut. *Environmental Microbiology*, 19(4), pp. 1366–1378.

Pereira-Marques, J., Hout, A., Ferreira, R.M., Weber, M., Pinto-Ribeiro, I., *et al.* (2019). Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. *Frontiers in Microbiology*, 10, p. 1277.

Perez, P.F., Doré, J., Leclerc, M., Levenez, F., Benyacoub, J., *et al.* (2007). Bacterial Imprinting of the Neonatal Immune System: Lessons From Maternal Cells? *Pediatrics*, 119(3), pp. e724–e732.

Perez-Muñoz, M.E., Arrieta, M.-C., Ramer-Tait, A.E. and Walter, J. (2017). A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome*, 5, p. 48.

Peris-Bondia, F., Latorre, A., Artacho, A., Moya, A. and D'Auria, G. (2011). The Active Human Gut Microbiota Differs from the Total Microbiota. *PLOS ONE*, 6(7), p. e22448.

Perofsky, A.C., Lewis, R.J. and Meyers, L.A. (2019). Terrestriality and bacterial transfer: a comparative study of gut microbiomes in sympatric Malagasy mammals. *The ISME Journal*, 13(1), pp. 50–63.

Phifer-Rixey, M. and Nachman, M.W. (2015). Insights into mammalian biology from the wild house mouse *Mus musculus. eLife,* 4, p. e05959.

Pocock, M.J.O., Searle, J.B. and White, P.C.L. (2004). Adaptations of animals to commensal habitats: population dynamics of house mice *Mus musculus domesticus* on farms. *Journal of Animal Ecology,* 73(5), pp. 878-888.

Polz, M.F. and Cavanaugh, C.M. (1998). Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology*, 64(10), pp. 3724–3730.

Popovic, A., Bourdon, C., Wang, P.W., Guttman, D.S., Voskuijl, W., *et al.* (2018). Design and application of a novel two-amplicon approach for defining eukaryotic microbiota. *Microbiome*, 6, p. 228.

Porter, J., Deere, D., Hardman, M., Edwards, C. and Pickup, R. (1997). Go with the flow – use of flow cytometry in environmental microbiology. *FEMS Microbiology Ecology*, 24(2), pp. 93–101.

Poulsen, C.S. and Stensvold, C.R. (2016). Systematic review on Endolimax nana: A less well studied intestinal ameba. *Tropical Parasitology*, 6(1), pp. 8–29.

Prabhu, V.R., Wasimuddin, Kamalakkannan, R., Arjun, M.S. and Nagarajan, M. (2020). Consequences of Domestication on Gut Microbiome: A Comparative Study Between Wild Gaur and Domestic Mithun. *Frontiers in Microbiology*, 11, p. 133.

Prest, E.I., Hammes, F., Kötzsch, S., van Loosdrecht, M.C.M. and Vrouwenvelder, J.S. (2013). Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Research*, 47(19), pp. 7131–7142.

Previtali, M.A., Lima, M., Meserve, P.L., Kelt, D.A. and Gutiérrez, J.R. (2009). Population dynamics of two sympatric rodents in a variable environment: rainfall, resource availability, and predation. *Ecology*, 90(7), pp. 1996–2006.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, 5(3), p. e9490.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), pp. 59–65.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue), pp. D590-596.

R. Pritchett, K. (2007). Chapter 22 - Helminth Parasites of Laboratory Mice. in J.G. Fox, M.T. Davisson, F.W. Quimby, S.W. Barthold, C.E. Newcomer, et al. (eds) *The Mouse in Biomedical Research (Second Edition)*. Burlington: Academic Press (American College of Laboratory Animal Medicine), pp. 551–564.

Radwan, J., Babik, W., Kaufman, J., Lenz, T.L. and Winternitz, J. (2020). Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends in genetics*, 36(4), pp. 298–311.

Ragazzo, L.J., Zohdy, S., Velonabison, M., Herrera, J., Wright, P.C., *et al.* (2018). Entamoeba histolytica infection in wild lemurs associated with proximity to humans. *Veterinary Parasitology*, 249, pp. 98–101.

Rajilić-Stojanović, M. and de Vos, W.M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews*, 38(5), pp. 996–1047.

Ramos, A.C.S., Oliveira, L.M., Santos, Y.L.D.C.O., Dantas, M.C.S., Walker, C.I.B., *et al.* (2022). The role of IgA in gastrointestinal helminthiasis: A systematic review. *Immunology Letters*, 249, pp. 12–22.

Rao, C., Coyte, K.Z., Bainter, W., Geha, R.S., Martin, C.R., *et al.* (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*, 591(7851), pp. 633–638.

Raulo, A., Allen, B.E., Troitsky, T., Husby, A., Firth, J.A., *et al.* (2021). Social networks strongly predict the gut microbiota of wild mice. *The ISME Journal*, 15(9), pp. 2601–2613.

Rausch, S., Held, J., Stange, J., Lendner, M., Hepworth, M.R., *et al.* (2010). A matter of timing: Early, not chronic phase intestinal nematode infection restrains control of a concurrent enteric protozoan infection. *European Journal of Immunology*, 40(10), pp. 2804–2815.

Rémy, A., Odden, M., Richard, M., Stene, M.T., Le Galliard, J.-F., *et al.* (2013). Food distribution influences social organization and population growth in a small rodent. *Behavioral Ecology*, 24(4), pp. 832–841.

Reynolds, L.A., Smith, K.A., Filbey, K.J., Harcus, Y., Hewitson, J.P., *et al.* (2014). Commensal-pathogen interactions in the intestinal tract. *Gut Microbes*, 5(4), pp. 522–532.

Reynoso-García, J., Miranda-Santiago, A.E., Meléndez-Vázquez, N.M., Acosta-Pagán, K., Sánchez-Rosado, M., *et al.* (2022). A complete guide to human microbiomes: Body niches, transmission, development, dysbiosis, and restoration. *Frontiers in Systems Biology*, 2, p. 951403.

Risely, A. (2020). Applying the core microbiome to understand host–microbe systems. *Journal of Animal Ecology*, 89(7), pp. 1549–1558.

Risely, A., Wilhelm, K., Clutton-Brock, T., Manser, M.B. and Sommer, S. (2021a). Diurnal oscillations in gut bacterial load and composition eclipse seasonal and lifetime dynamics in wild meerkats. *Nature Communications*, 12, p. 6017.

Risely, A., Gillingham, M.A.F., Béchet, A., Brändel, S., Heni, A.C., *et al.* (2021b). Phylogeny- and Abundance-Based Metrics Allow for the Consistent Comparison of Core Gut Microbiome Diversity Indices Across Host Species. *Frontiers in Microbiology*, 12, p. 659918.

Risely, A., Schmid, D.W., Müller-Klein, N., Wilhelm, K., Clutton-Brock, T.H., *et al.* (2022). Gut microbiota individuality is contingent on temporal scale and age in wild meerkats. *Proceedings of the Royal Society B: Biological Sciences*, 289(1981), p. 20220609.

Rodríguez-Zaragoza, S. (1994). Ecology of Free-Living Amoebae. *Critical Reviews in Microbiology*, 20(3), pp. 225–241.

Clarindo, W.R. and Carvalho, C.R. (2011). Flow cytometric analysis using SYBR Green I for genome size estimation in coffee. *Acta Histochemica*, 113(2), pp. 221–225.

Rosshart, S.P., Vassallo, B.G., Angeletti, D., Hutchinson, D.S., Morgan, A.P., *et al.* (2017). Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance. *Cell*, 171(5), pp. 1015–1028.

Rosshart, S.P., Herz, J., Vassallo, B.G., Hunter, A., Wall, M.K., *et al.* (2019). Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science*, 365(6452), p. eaaw4361.

Roswell, M., Dushoff, J. and Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, 130(3), pp. 321–338.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., *et al.* (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, 555(7695), pp. 210–215.

Röver, C. and Friede, T. (2022). Double arcsine transform not appropriate for meta-analysis. *Research Synthesis Methods,* 13(5), pp. 547-600.

Rowe, F.P., Bradfield, A., Quy, R.J. and Swinney, T. (1985). Relationship Between Eye Lens Weight and Age in the Wild House Mouse (Mus musculus). *Journal of Applied Ecology*, 22(1), pp. 55–61.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., *et al.* (2018). Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition*, 57(1), pp. 1–24.

Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., *et al.* (2015). A Higher Level Classification of All Living Organisms. *PLoS ONE*, 10(4), p. e0119248.

Russell, D.A., Ross, R.P., Fitzgerald, G.F. and Stanton, C. (2011). Metabolic activities and probiotic potential of bifidobacteria. *International Journal of Food Microbiology*, 149(1), pp. 88–105.

Ryan, U., Fayer, R. and Xiao, L. (2014). Cryptosporidium species in humans and animals: current understanding and research needs. *Parasitology*, 141(13), pp. 1667–1685.

Rynkiewicz, E.C., Clerc, M., Babayan, S.A. and Pedersen, A.B. (2019). Variation in Local and Systemic Pro-Inflammatory Immune Markers of Wild Wood Mice after Anthelmintic Treatment. *Integrative and Comparative Biology*, 59(5), pp. 1190–1202.

Salmaso, N., Boscaini, A. and Pindo, M. (2020). Unraveling the Diversity of Eukaryotic Microplankton in a Large and Deep Perialpine Lake Using a High Throughput Sequencing Approach. *Frontiers in Microbiology*, 11, p. 789.

Sam, Q.H., Chang, M.W. and Chai, L.Y.A. (2017). The Fungal Mycobiome and Its Interaction with Gut Bacteria in the Host. *International Journal of Molecular Sciences*, 18(2), p. 330.

Sanchez, J.N. and Hudgens, B.R. (2019). Impacts of Heterogeneous Host Densities and Contact Rates on Pathogen Transmission in the Channel Island Fox (Urocyon littoralis). *Biological conservation*, 236, pp. 593–603.

Sardinha-Silva, A., Alves-Ferreira, E.V.C. and Grigg, M.E. (2022). Intestinal immune responses to commensal and pathogenic protozoa. *Frontiers in Immunology*, 13, p. 963723.

Sarkar, A., Harty, S., Johnson, K.V.-A., Moeller, A.H., Archie, E.A., *et al.* (2020). Microbial transmission in animal social networks and the social microbiome. *Nature Ecology & Evolution*, 4(8), pp. 1020–1035.

Savioli, L., Smith, H. and Thompson, A. (2006). Giardia and Cryptosporidium join the 'Neglected Diseases Initiative'. *Trends in Parasitology*, 22(5), pp. 203–208.

Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., *et al.* (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 50(Database issue), pp. D20–D26.

Scanlan, P.D. and Marchesi, J.R. (2008). Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. *The ISME journal*, 2(12), pp. 1183–1193.

Scanlan, P.D., Stensvold, C.R., Rajilić-Stojanović, M., Heilig, H.G.H.J., De Vos, W.M., *et al.* (2014). The microbial eukaryote Blastocystis is a prevalent and diverse member of the healthy human gut microbiota. *FEMS Microbiology Ecology*, 90(1), pp. 326–330.

Scheuring, I. and Yu, D.W. (2012). How to assemble a beneficial microbiome in three easy steps. *Ecology Letters*, 15(11), pp. 1300–1307.

Schippa, S., Iebba, V., Barbato, M., Di Nardo, G., Totino, V., *et al.* (2010). A distinctive 'microbial signature' in celiac pediatric patients. *BMC Microbiology*, 10, p. 175.

Schmid, M., Heitlinger, E., Spork, S., Mollenkopf, H.-J., Lucius, R., *et al.* (2014). Eimeria falciformis infection of the mouse caecum identifies opposing roles of IFNγ-regulated host pathways for the parasite development. *Mucosal Immunology*, 7(4), pp. 969–982.

Schmid-Hempel, P. (2009). Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1513), pp. 85–98.

Schmidt, E., Mykytczuk, N. and Schulte-Hostedde, A.I. (2019). Effects of the captive and wild environment on diversity of the gut microbiome of deer mice (Peromyscus maniculatus). *The ISME Journal*, 13(5), pp. 1293–1305.

Schmidt, T.S.B., Raes, J. and Bork, P. (2018). The Human Gut Microbiome: From Association to Modulation. *Cell*, 172(6), pp. 1198–1215.

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., *et al.* (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020, p. baaa062.

Schuster, F.L. and Ramirez-Avila, L. (2008). Current World Status of Balantidium coli. *Clinical Microbiology Reviews*, 21(4), pp. 626–638.

Scupham, A.J., Presley, L.L., Wei, B., Bent, E., Griffith, N., *et al.* (2006). Abundant and Diverse Fungal Microbiota in the Murine Intestine. *Applied and Environmental Microbiology*, 72(1), pp. 793–801.

Seabolt, M.H., Konstantinidis, K.T. and Roellig, D.M. (2021). Hidden Diversity within Common Protozoan Parasites as Revealed by a Novel Genomotyping Scheme. *Applied and Environmental Microbiology*, 87(6), pp. e02275-20.

Sender, R., Fuchs, S. and Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, 164(3), pp. 337–340.

Shabat, S.K.B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., *et al.* (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *The ISME Journal*, 10(12), pp. 2958–2972.

Shi, L. and Lin, L. (2019). The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, 98(23), p. e15987.

Silberman, J.D., Simpson, A.G.B., Kulda, J., Cepicka, I., Hampl, V., *et al.* (2002). Retortamonad flagellates are closely related to diplomonads--implications for the history of mitochondrial function in eukaryote evolution. *Molecular Biology and Evolution*, 19(5), pp. 777–786.

Silverman, J.D., Bloom, R.J., Jiang, S., Durand, H.K., Dallow, E., *et al.* (2021). Measuring and mitigating PCR bias in microbiota datasets. *PLOS Computational Biology*, 17(7), p. e1009113.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A. and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), pp. 121–132.

Skalski, J.H., Limon, J.J., Sharma, P., Gargus, M.D., Nguyen, C., *et al.* (2018). Expansion of commensal fungus Wallemia mellicola in the gastrointestinal mycobiota enhances the severity of allergic airway disease in mice. *PLOS Pathogens*, 14(9), p. e1007260.

Smith, C.J. and Osborn, A.M. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiology Ecology*, 67(1), pp. 6–20.

Smith, K., McCoy, K.D. and Macpherson, A.J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Seminars in Immunology*, 19(2), pp. 59–69.

Smith, N.C., Ovington, K.S., Deplazes, P. and Eckert, J. (1995). Cytokine and immunoglobulin subclass responses of rats to infection with Eimeria nieschulzi. *Parasitology*, 111(1), pp. 51–57.

Soave, O. and Brand, C.D. (1991). Coprophagy in animals: a review. *The Cornell Veterinarian*, 81(4), pp. 357–364.

Solarczyk, P. (2021). Host Range of Cyclospora Species: Zoonotic Implication. *Acta Protozoologica*, 60, pp. 13–20.

Solomon, R. and Jami, E. (2021). Rumen protozoa: from background actors to featured role in microbiome research. *Environmental Microbiology Reports*, 13(1), pp. 45–49.

Sommer, F. and Bäckhed, F. (2013). The gut microbiota — masters of host development and physiology. *Nature Reviews Microbiology*, 11(4), pp. 227–238.

Stacy, B.A., Chapman, P.A., Stockdale-Walden, H., Work, T.M., Dagenais, J., *et al.* (2019). Caryospora-Like Coccidia Infecting Green Turtles (Chelonia mydas): An Emerging Disease With Evidence of Interoceanic Dissemination. *Frontiers in Veterinary Science*, 6.

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., *et al.* (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome*, 4, p. 28.

Starke, R., Pylro, V.S. and Morais, D.K. (2021). 16S rRNA Gene Copy Number Normalization Does Not Provide More Reliable Conclusions in Metataxonomic Surveys. *Microbial Ecology*, 81(2), pp. 535–539.

Stensvold, C.R., Lebbad, M., Victory, E.L., Verweij, J.J., Tannich, E., *et al.* (2011). Increased Sampling Reveals Novel Lineages of Entamoeba: Consequences of Genetic Diversity and Host Specificity for Taxonomy and Molecular Detection. *Protist*, 162(3), pp. 525–541.

Stensvold, C.R., Lebbad, M. and Clark, C.G. (2012). Last of the Human Protists: The Phylogeny and Genetic Diversity of Iodamoeba. *Molecular Biology and Evolution*, 29(1), pp. 39–42.

Stensvold, C.R. and Clark, C.G. (2016). Current status of Blastocystis: A personal view. *Parasitology International*, 65, pp. 763–771.

Stensvold, C.R. and Clark, C.G. (2020). Pre-empting Pandora's Box: Blastocystis Subtypes Revisited. *Trends in Parasitology*, 36(3), pp. 229–232.

Stevens, E.J., Bates, K.A. and King, K.C. (2021). Host microbiota can facilitate pathogen infection. *PLOS Pathogens*, 17(5), p. e1009514.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K. and Schmidt, T.M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, 43(Database issue), pp. D593-598.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19(Suppl. 1), pp. 21–31.

Strati, F., Di Paola, M., Stefanini, I., Albanese, D., Rizzetto, L., *et al.* (2016). Age and Gender Affect the Composition of Fungal Population of the Human Gastrointestinal Tract. *Frontiers in Microbiology*, 7.

Strickland, B.A., Patel, M.C., Shilts, M.H., Boone, H.H., Kamali, A., *et al.* (2021). Microbial community structure and composition is associated with host species and sex in Sigmodon cotton rats. *Animal Microbiome*, 3, p. 29.

Suau, A., Bonnet, R., Sutren, M., Godon, J.-J., Gibson, G.R., *et al.* (1999). Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. *Applied and Environmental Microbiology*, 65(11), pp. 4799–4807.

Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., *et al.* (2004). Postnatal microbial colonization programs the hypothalamic–pituitary–adrenal system for stress response in mice. *The Journal of Physiology*, 558(1), pp. 263–275.

Suhr, M.J. and Hallen-Adams, H.E. (2015). The human gut mycobiome: pitfalls and potentials—a mycologist's perspective. *Mycologia*, 107(6), pp. 1057–1073.

Sun, B., Gu, Z., Wang, X., Huffman, M.A., Garber, P.A., *et al.* (2018). Season, age, and sex affect the fecal mycobiota of free-ranging Tibetan macaques (Macaca thibetana). *American Journal of Primatology*, 80, p. e22880.

Suzuki, T.A. and Nachman, M.W. (2016). Spatial Heterogeneity of Gut Microbial Composition along the Gastrointestinal Tract in Natural Populations of House Mice. *PLoS ONE*, 11(9), p. e0163720.

Suzuki, T.A. (2017). Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. *Integrative and Comparative Biology*, 57(4), pp. 756–769.

Suzuki, T.A., Martins, F.M. and Nachman, M.W. (2019a). Altitudinal variation of the gut microbiota in wild house mice. *Molecular Ecology*, 28(9), pp. 2378–2390.

Suzuki, T.A., Phifer-Rixey, M., Mack, K.L., Sheehan, M.J., Lin, D., *et al.* (2019b). Host genetic determinants of the gut microbiota of wild mice. *Molecular Ecology*, 28(13), pp. 3197–3207.

Swidergall, M. and LeibundGut-Landmann, S. (2022). Immunosurveillance of Candida albicans commensalism by the adaptive immune system. *Mucosal Immunology*, 15(5), pp. 829–836.

Tanaka, R., Hino, A., Tsai, I.J., Palomares-Rius, J.E., Yoshida, A., *et al.* (2014). Assessment of Helminth Biodiversity in Wild Rats Using 18S rDNA Based Metagenomics. *PLOS ONE*, 9(10), p. e110769.

Tanoue, T., Umesaki, Y. and Honda, K. (2010). Immune responses to gut microbiota-commensals and pathogens. *Gut Microbes*, 1(4), pp. 224–233.

Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., *et al.* (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology*, 31(10), pp. 2769–2795.

Teixeira, J.E. and Huston, C.D. (2008). Evidence of a Continuous Endoplasmic Reticulum in the Protozoan Parasite Entamoeba histolytica. *Eukaryotic Cell*, 7(7), pp. 1222–1226.

Telfer, S., Birtles, R., Bennett, M., Lambin, X., Paterson, S., *et al.* (2008). Parasite interactions in natural populations: insights from longitudinal data. *Parasitology*, 135(7), pp. 767–781.

Tenter, A.M., Barta, J.R., Beveridge, I., Duszynski, D.W., Mehlhorn, H., *et al.* (2002). The conceptual basis for a new classification of the coccidia. *International Journal for Parasitology*, 32(5), pp. 595–616.

Thielemann, N., Herz, M., Kurzai, O. and Martin, R. (2022). Analyzing the human gut mycobiome – A short guide for beginners. *Computational and Structural Biotechnology Journal*, 20, pp. 608–614.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., *et al.* (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681), pp. 457–463.

Thompson, R.C.A. and Monis, P. (2012). Chapter 2 - Giardia—From Genome to Proteome. in D. Rollinson and S.I. Hay (eds) *Advances in Parasitology*. Academic Press, pp. 57–95.

Thomson, C.A., Morgan, S.C., Ohland, C. and McCoy, K.D. (2022). From germ-free to wild: modulating microbiome complexity to understand mucosal immunology. *Mucosal Immunology*, 15(6), pp. 1085–1094.

Thursby, E. and Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, 474(11), pp. 1823–1836.

Tipton, L., Müller, C.L., Kurtz, Z.D., Huang, L., Kleerup, E., *et al.* (2018). Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*, 6, p. 12.

Tourlousse, D.M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., *et al.* (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research*, 45(4), p. e23.

Trefancová, A., Mácová, A. and Kvičerová, J. (2019). Isosporan Oocysts in the Faeces of Bank Voles (Myodes glareolus; Arvicolinae, Rodentia): Real Parasites, or Pseudoparasites? *Protist*, 170(1), pp. 104–120.

Tung, J., Barreiro, L.B., Burns, M.B., Grenier, J.-C., Lynch, J., *et al.* (2015). Social networks predict gut microbiome composition in wild baboons. *eLife*, 4, p. e05224.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., *et al.* (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122), pp. 1027–1031.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., *et al.* (2007). The Human Microbiome Project. *Nature*, 449(7164), pp. 804–810.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., *et al.* (2009). A core gut microbiome in obese and lean twins. *Nature*, 457(7228), pp. 480–484.

Turner, T.R., James, E.K. and Poole, P.S. (2013). The plant microbiome. *Genome Biology*, 14(6), p. 209.

Turunen, J., Paalanne, N., Reunanen, J., Tapiainen, T. and Tejesvi, M.V. (2023). Development of gut mycobiome in infants and young children: a prospective cohort study. *Pediatric Research*, pp. 1–9.

Uldal, A. and Buchmann, K. (1996). Parasite host relations:Hexamita salmonis in rainbow trout Oncorhynchus mykiss. *Diseases of Aquatic Organisms*, 25, pp. 229–231.

Underhill, D.M. and Iliev, I.D. (2014). The mycobiota: interactions between commensal fungi and the host immune system. *Nature Reviews Immunology*, 14(6), pp. 405–416.

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(Database issue), pp. D480–D489.

Vaerewijck, M.J.M., Baré, J., Lambrecht, E., Sabbe, K. and Houf, K. (2014). Interactions of Foodborne Pathogens with Free-living Protozoa: Potential Consequences for Food Safety. *Comprehensive Reviews in Food Science and Food Safety*, 13(5), pp. 924–944.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., *et al.* (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, 551(7681), pp. 507–511.

Vandeputte, D., De Commer, L., Tito, R.Y., Kathagen, G., Sabino, J., *et al.* (2021). Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nature Communications*, 12, p. 6740.

Vaulot, D., Geisen, S., Mahé, F. and Bass, D. (2022). pr2-primers: An 18S rRNA primer database for protists. *Molecular Ecology Resources*, 22(1), pp. 168–179.

Vejzagić, N., Adelfio, R., Keiser, J., Kringel, H., Thamsborg, S.M., *et al.* (2015). Bacteria-induced egg hatching differs for Trichuris muris and Trichuris suis. *Parasites & Vectors*, 8, p. 371.

Vemuri, R., Shankar, E.M., Chieppa, M., Eri, R. and Kavanagh, K. (2020). Beyond Just Bacteria: Functional Biomes in the Gut Ecosystem Including Virome, Mycobiome, Archaeome and Helminths. *Microorganisms*, 8(4), p. 483.

Vestheim, H. and Jarman, S.N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, 5, p. 12.

Větrovský, T. and Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE*, 8(2), p. e57923.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36, pp. 1–48.

Viney, M. and Kikuchi, T. (2017). Strongyloides ratti and S. venezuelensis – rodent models of Strongyloides infection. *Parasitology*, 144(3), pp. 285–294.

Viney, M. and Riley, E.M. (2017). The Immunology of Wild Rodents: Current Status and Future Prospects. *Frontiers in Immunology*, 8, p. 1481.

Viney, M. (2019). The gut microbiota of wild rodents: Challenges and opportunities. *Laboratory Animals*, 53(3), pp. 252–258.

Viney, M.E., Riley, E.M. and Buchanan, K.L. (2005). Optimal immune responses: immunocompetence revisited. *Trends in Ecology & Evolution*, 20(12), pp. 665–669.

Vuong, H.E., Yano, J.M., Fung, T.C. and Hsiao, E.Y. (2017). The Microbiome and Host Behavior. *Annual Review of Neuroscience*, 40(1), pp. 21–49.

Wagner, B.D., Grunwald, G.K., Zerbe, G.O., Mikulich-Gilbertson, S.K., Robertson, C.E., *et al.* (2018). On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Frontiers in Microbiology*, 9, p. 1037.

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E.E.L., Narayanasamy, S., *et al.* (2017). Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Frontiers in Microbiology*, 8, p. 738.

Wampach, L., Heintz-Buschart, A., Fritz, J.V., Ramiro-Garcia, J., Habier, J., *et al.* (2018). Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nature Communications*, 9, p. 5091.

Wang, N. (2018). How to Conduct a Meta-Analysis of Proportions in R: A Comprehensive Tutorial. Preprint on ResearchGate. doi: 10.13140/RG.2.2.27199.00161

Wang, J., Linnenbrink, M., Künzel, S., Fernandes, R., Nadeau, M.-J., *et al.* (2014). Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild mice. *Proceedings of the National Academy of Sciences*, 111(26), pp. E2703–E2710.

Wang, L., Abu-Doleh, A., Plank, J., Catalyurek, U.V., Firkins, J.L., *et al.* (2019). The transcriptome of the rumen ciliate Entodinium caudatum reveals some of its metabolic features. *BMC Genomics*, 20, p. 1008.

Wang, M., Karlsson, C., Olsson, C., Adlerberth, I., Wold, A.E., *et al.* (2008). Reduced diversity in the early fecal microbiota of infants with atopic eczema. *The Journal of Allergy and Clinical Immunology*, 121(1), pp. 129–134.

Wang, X., Howe, S., Deng, F. and Zhao, J. (2021). Current Applications of Absolute Bacterial Quantification in Microbiome Studies and Decision-Making Regarding Different Biological Questions. *Microorganisms*, 9(9), p. 1797.

Wei, Y., Gao, J., Kou, Y., Meng, L., Zheng, X., *et al.* (2020). Commensal Bacteria Impact a Protozoan's Integration into the Murine Gut Microbiota in a Dietary Nutrient-Dependent Manner. *Applied and Environmental Microbiology*, 86(11), pp. e00303-20.

Weiner, A., Turjeman, S. and Koren, O. (2023). Gut microbes and host behavior: The forgotten members of the gut-microbiome. *Neuropharmacology*, 227, p. 109453.

Weinstein, S.B., Stephens, W.Z., Greenhalgh, R., Round, J.L. and Dearing, M.D. (2022). Wild herbivorous mammals (genus Neotoma) host a diverse but transient assemblage of fungi. *Symbiosis*, 87(1), pp. 45–58.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., *et al.* (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5, p. 27.

Weldon, L., Abolins, S., Lenzi, L., Bourne, C., Riley, E.M., *et al.* (2015). The Gut Microbiota of Wild Mice. *PLOS ONE*, 10(8), p. e0134643.

Wheeler, M.L., Limon, J.J., Bar, A.S., Leal, C.A., Gargus, M., *et al.* (2016). Immunological Consequences of Intestinal Fungal Dysbiosis. *Cell Host & Microbe*, 19(6), pp. 865–873.

White, E.C., Houlden, A., Bancroft, A.J., Hayes, K.S., Goldrick, M., *et al.* (2018). Manipulation of host and parasite microbiotas: Survival strategies during chronic nematode infection. *Science Advances*, 4(3), p. eaap7399.

Whiteley, A.S., Griffiths, R.I. and Bailey, M.J. (2003). Analysis of the microbial functional diversity within water-stressed soil communities by flow cytometric analysis and CTC+ cell sorting. *Journal of Microbiological Methods*, 54(2), pp. 257–267.

Wikberg, E.C., Christie, D., Sicotte, P. and Ting, N. (2020). Interactions between social groups of colobus monkeys (*Colobus vellerosus*) explain similarities in their gut microbiomes. *Animal Behaviour,* 163, pp. 17-31.

Wilber, P.G., Duszynski, D.W., Upton, S.J., Seville, R.S. and Corliss, J.O. (1998). A revision of the taxonomy and nomenclature of the Eimeria spp. (Apicomplexa: Eimeriidae) from rodents in the Tribe Marmotini (Sciuridae). *Systematic Parasitology*, 39(2), pp. 113–135.

Williams, C.L., Thomas, B.J., McEwan, N.R., Rees Stevens, P., Creevey, C.J., *et al.* (2020). Rumen Protozoa Play a Significant Role in Fungal Predation and Plant Carbohydrate Breakdown. *Frontiers in Microbiology*, 11, p. 720.

Williams, J.M., Duckworth, C.A., Vowell, K., Burkitt, M.D. and Pritchard, D.M. (2016). Intestinal Preparation Techniques for Histological Analysis in the Mouse. *Current Protocols in Mouse Biology*, 6(2), pp. 148–168.

Willis, A.D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology*, 10, p. 2407.

Wilson, K.H. and Blitchington, R.B. (1996). Human colonic biota studied by ribosomal DNA sequence analysis. *Applied and Environmental Microbiology*, 62(7), pp. 2273–2278.

Wilson, D.E., Mittermeier, R.A. and Lacher, T.E. (2017). Handbook of the Mammals of the World (Volumes 6 and 7). Lynx Edicions.

Worsley, S.F., Davies, C.S., Mannarelli, M.-E., Komdeur, J., Dugdale, H.L., *et al.* (2022). Assessing the causes and consequences of gut mycobiome variation in a wild population of the Seychelles warbler. *Microbiome*, 10, p. 242.

Wu, G., Zhao, N., Zhang, C., Lam, Y.Y. and Zhao, L. (2021). Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Medicine*, 13, p. 22.

Xu, F., Jerlström-Hultqvist, J., Kolisko, M., Simpson, A.G.B., Roger, A.J., *et al.* (2016). On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad Trepomonas sp. PC1. *BMC Biology*, 14, p. 62.

Xu, Y., Tandon, R., Ancheta, C., Arroyo, P., Gilbert, J.A., *et al.* (2021). Quantitative profiling of built environment bacterial and fungal communities reveals dynamic material dependent growth patterns and microbial interactions. *Indoor Air*, 31(1), pp. 188–205.

Xue, Z., Zhang, W., Wang, L., Hou, R., Zhang, M., *et al.* (2015). The Bamboo-Eating Giant Panda Harbors a Carnivore-Like Gut Microbiota, with Excessive Seasonal Variations. *mBio*, 6(3), pp. e00022-15.

Yang, C., Mogno, I., Contijoch, E.J., Borgerding, J.N., Aggarwala, V., *et al.* (2020). Fecal IgA Levels Are Determined by Strain-Level Differences in Bacteroides ovatus and Are Modifiable by Gut Microbiota Manipulation. *Cell Host & Microbe*, 27(3), pp. 467–475.

Yang, L. and Chen, J. (2022). A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. *Microbiome*, 10, p. 130.

Yason, J.A., Liang, Y.R., Png, C.W., Zhang, Y. and Tan, K.S.W. (2019). Interactions between a pathogenic Blastocystis subtype and gut microbiota: in vitro and in vivo studies. *Microbiome*, 7, p. 30.

Yeung, F., Chen, Y.-H., Lin, J.-D., Leung, J.M., McCauley, C., *et al.* (2020). Altered Immunity of Laboratory Mice in the Natural Environment Is Associated with Fungal Colonization. *Cell Host & Microbe*, 27(5), pp. 809–822.

Yildiz, İ. and Erdem Aynur, Z. (2022). First detection and molecular characterization of Dientamoeba fragilis in cattle. *Zoonoses and Public Health*, 69(8), pp. 897–903.

Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., *et al.* (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nature Communications*, 10, p. 2200.

Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S.R., Marinier, E., *et al.* (2018). Impact of sequencing depth on the characterization of the microbiome and resistome. *Scientific Reports*, 8, p. 5890.

Zalar, P., Sybren de Hoog, G., Schroers, H.-J., Frank, J.M. and Gunde-Cimerman, N. (2005). Taxonomy and phylogeny of the xerophilic genus Wallemia (Wallemiomycetes and Wallemiales, cl. et ord. nov.). *Antonie van Leeuwenhoek*, 87(4), pp. 311–328.

Zambrano-Villa, S., Rosales-Borjas, D., Carrero, J.C. and Ortiz-Ortiz, L. (2002). How protozoan parasites evade the immune response. *Trends in Parasitology*, 18(6), pp. 272–278.

Zanetti, A. dos S., Malheiros, A.F., de Matos, T.A., dos Santos, C., Battaglini, P.F., *et al.* (2021). Diversity, geographical distribution, and prevalence of Entamoeba spp. in Brazil: a systematic review and meta-analysis. *Parasite*, 28, p. 17.

Zeldenrust, E.G. and Barta, J.R. (2021). Description of the First Klossia Species (Apicomplexa: Eucoccidiorida: Adeleorina: Adeleidae) Infecting a Pulmonate Land Snail, Triodopsis hopetonensis (Mollusca: Polygyridae), in North America. *Journal of Parasitology*, 107(3), pp. 421–429.

Zhang, C., Derrien, M., Levenez, F., Brazeilles, R., Ballal, S.A., *et al.* (2016). Ecological robustness of the gut microbiota in response to ingestion of transient food-borne microbes. *The ISME Journal*, 10(9), pp. 2235–2245.

Zhang, K., Fu, Y., Li, J. and Zhang, L. (2021). Public health and ecological significance of rodents in Cryptosporidium infections. *One Health*, 14, p. 100364.

Zhang, N., Zhang, H., Yu, Y., Gong, P., Li, J., *et al.* (2019). High prevalence of Pentatrichomonas hominis infection in gastrointestinal cancer patients. *Parasites & Vectors*, 12, p. 423.

Zhang, Z., Wei, W., Yang, S., Huang, Z., Li, C., *et al.* (2022). Regulation of Dietary Protein Solubility Improves Ruminal Nitrogen Metabolism In Vitro: Role of Bacteria–Protozoa Interactions. *Nutrients*, 14(14), p. 2972.

Zuur, A.F. and Ieno, E.N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7(6), pp. 636–645.

Appendix

Appendix 1.1: Chapter 2 supplementary material

Justification of the continued use of the double-arcsine transformation

The Freeman-Tukey double-arcsine transformation has become increasingly popular in meta-analyses of prevalence in order to avoid the constraints associated with using proportion data that is bound between 0 and 1 (Barendregt *et al.*, 2013; Lin and Xu, 2020). However, recent literature has shown that the double-arcsine transformation may be inappropriate for meta-analyses given some of its limitations (Schwarzer *et al.,* 2019; Lin and Xu, 2020; Röver and Friede, 2022). In particular, the transformation: i) is not intuitive to the reader, ii) can violate the assumptions of the meta-analyses models, and iii) can have multiple methods for back-transforming the data to predict prevalence which give different values (Lin and Xu, 2020). To this end, it has been suggested that the single-arcsine transformation may be more appropriate for meta-analyses of proportions, despite the double-arcsine stabilizing variances better in general (Röver and Friede, 2022; Lin and Xu, 2020).

In order to decide which transformation was appropriate for my meta-analyses, I conducted analyses using both the double-arcsine and single-arcsine transformations. For all models, the results and conclusions drawn were identical. I also assessed if the data from each transformation fit the assumptions of the meta-analyses: the data were more normally distributed when using the double-arcsine transformation. Furthermore, whilst the main goal of the analyses is to identify factors that may affect prevalence, estimates of prevalence have been calculated. Thus, I also compared prevalence estimates from the two transformations. As above, predicted prevalences were similar. Most prevalences varied by 1-2% and the greatest differences was 4.01%. To this end, there is justification for using the double-arcsine transformation for this analysis as it i) better fits the model assumptions and ii) gives the same results and conclusions as single-arcsine transformation (Doi and Xu, 2021).

References

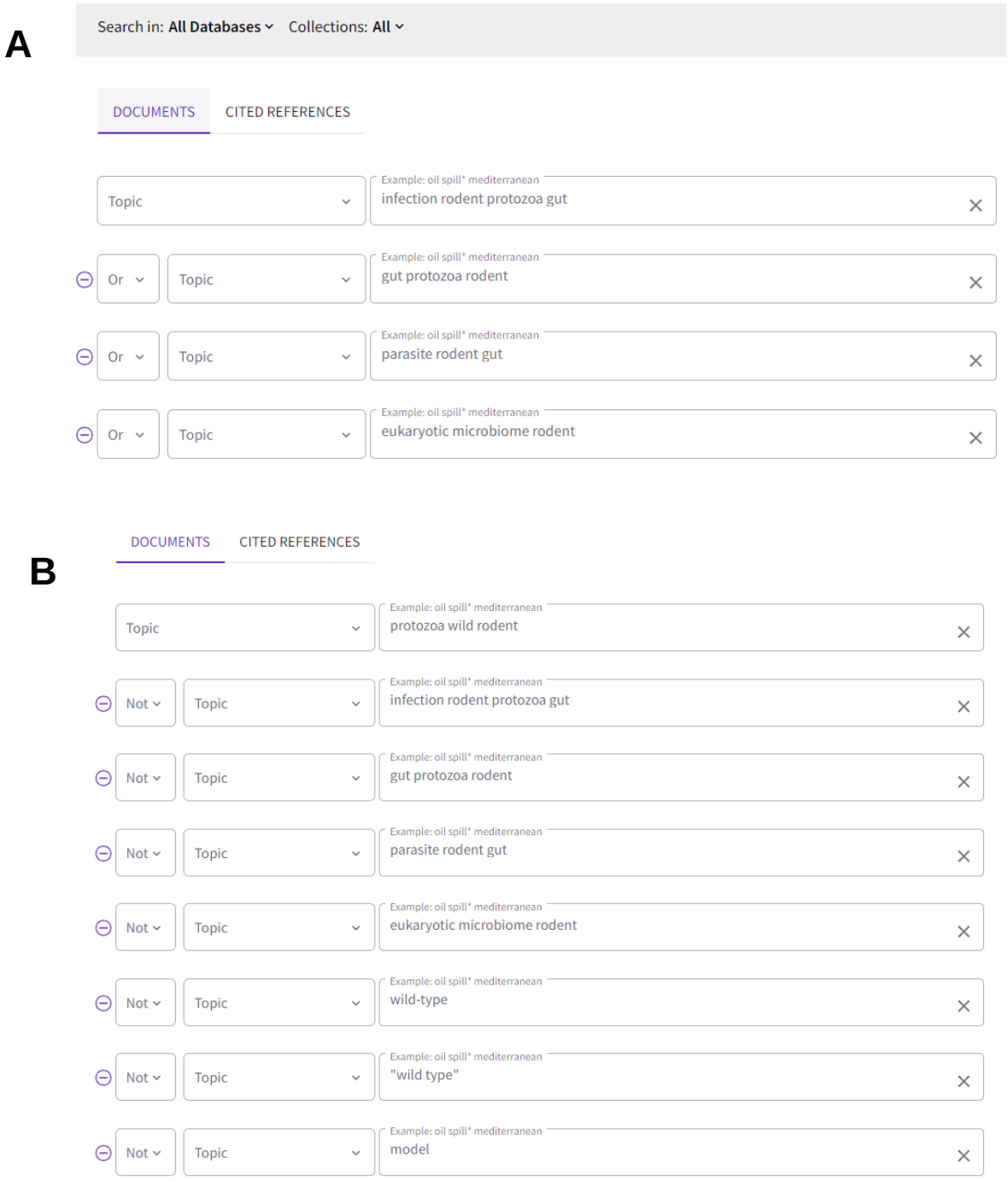
Barendregt, J.J., Doi, S.A., Lee, Y.Y., Norman, R.E. and Vos, T. (2013). Meta-analysis of prevalence. *J Epidemiol Community Health*, 67(11), pp. 974–978

Doi, S.A. and Xu, C. (2021). The Freeman_Tukey double arcsine transformation for the meta-analysis of proportions: Recent critiscims were seriously misleading. *Journal of Evidence-Based Medicine,* 14(4), pp. 259-261.

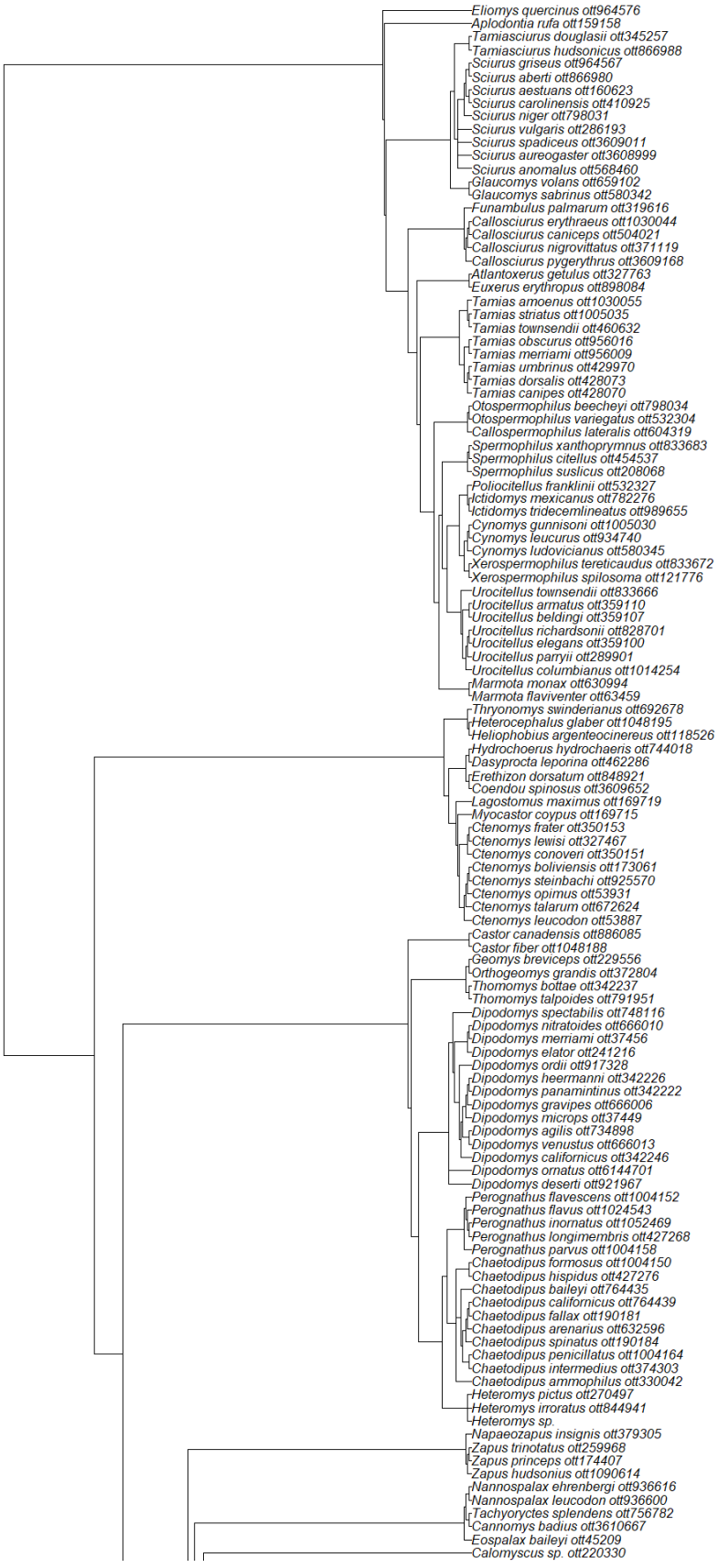
Lin, L. and Xu, C. (2020). Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. *Health Science Reports*, 3(3), p. e178

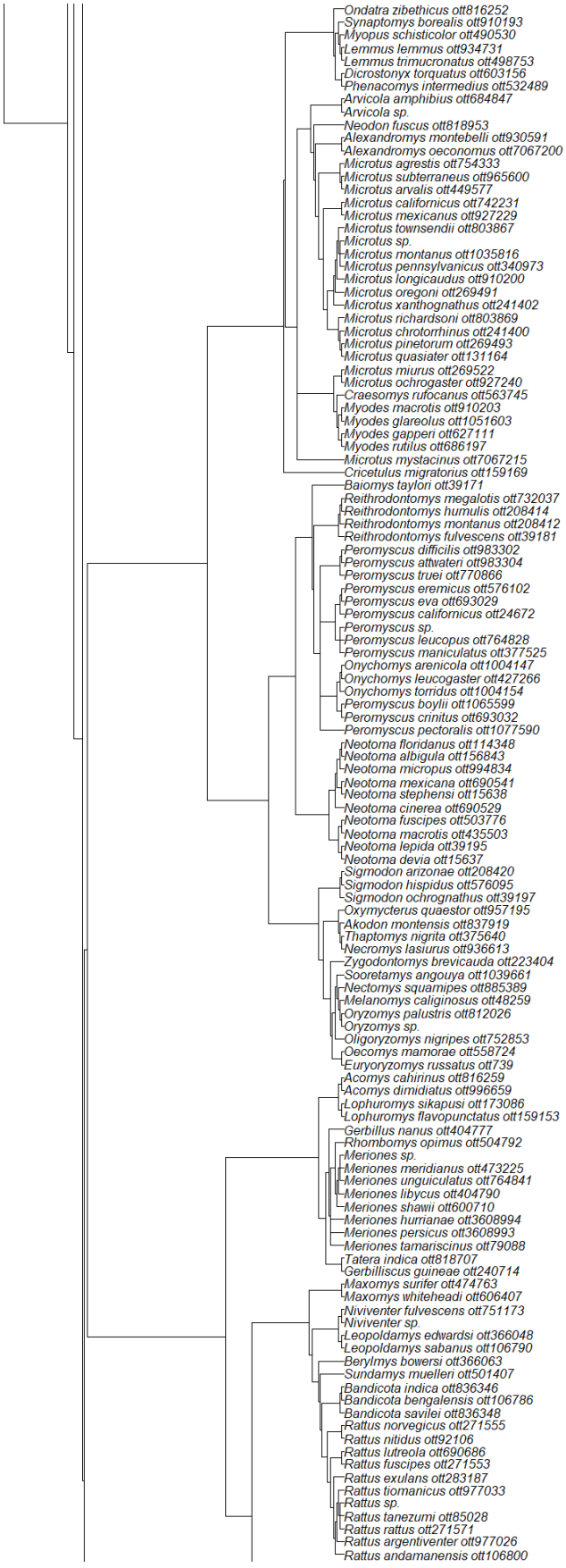
Schwarzer, G., Chemaitelly, H., Abu-Raddad, L.J. and Rücker, G. (2019). Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Research Synthesis Methods,* 10(3), pp. 476-483.

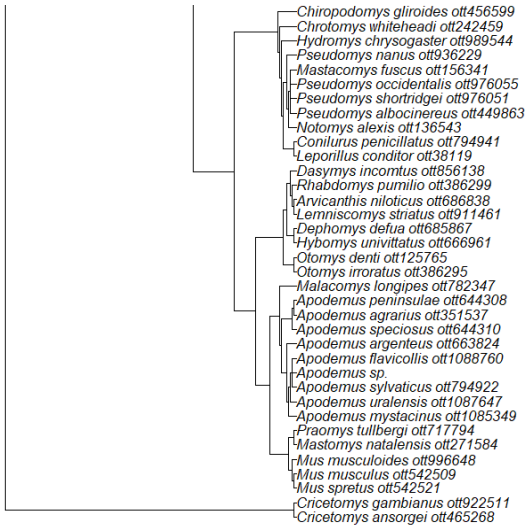
Röver, C. and Friede, T. (2022). Double arcsine transform not appropriate for meta-analysis. *Research Synthesis Methods,* 13(5), pp. 547-600.



Supplementary Figure 1. The search terms used within Web of Science to identify protozoa in wild rodents. A) Search 1 used four different search terms simultaneously and searched within all Web of Science databases. B) Search 2 used one search term but specified not to search for the four terms included in A) and also included three search terms to eliminate laboratory infections.


Eliomys quercinus ott964576
Aplodontia rufa ott159158
Tamiasciurus douglasii ott345257
Tamiasciurus hudsonicus ott866988
Sciurus griseus ott964567
Sciurus aberti ott866980
Sciurus aestuans ott160623
Sciurus carolinensis ott410925
Sciurus niger ott798031
Sciurus vulgaris ott286193
Sciurus spadiceus ott3609011
Sciurus aureogaster ott3608999
Sciurus anomalus ott568460
Glaucomys volans ott659102
Glaucomys sabrinus ott580342
Funambulus palmarum ott319616
Callosciurus erythraeus ott1030044
Callosciurus caniceps ott504021
Callosciurus nigrovittatus ott371119
Callosciurus pygerythrus ott3609168
Atlantoxerus getulus ott327763
Euxerus erythropus ott898084
Tamias amoenus ott1030055
Tamias striatus ott1005035
Tamias townsendii ott460632
Tamias obscurus ott956016
Tamias merriami ott956009
Tamias umbrinus ott429970
Tamias dorsalis ott428073
Tamias canipes ott428070
Otospermophilus beecheyi ott798034
Otospermophilus variegatus ott532304
Callospermophilus lateralis ott604319
Spermophilus xanthoprymnus ott833683
Spermophilus citellus ott454537
Spermophilus suslicus ott208068
Poliocitellus franklinii ott532327
Ictidomys mexicanus ott782276
Ictidomys tridecemlineatus ott989655
Cynomys gunnisoni ott1005030
Cynomys leucurus ott934740
Cynomys ludovicianus ott580345
Xerospermophilus tereticaudus ott833672
Xerospermophilus spilosoma ott121776
Urocitellus townsendii ott833666
Urocitellus armatus ott359110
Urocitellus beldingi ott359107
Urocitellus richardsonii ott828701
Urocitellus elegans ott359100
Urocitellus parryii ott289901
Urocitellus columbianus ott1014254
Marmota monax ott630994
Marmota flaviventer ott63459
Thryonomys swinderianus ott692678
Heterocephalus glaber ott1048195
Heliophobius argenteocinereus ott118526
Hydrochoerus hydrochaeris ott744018
Dasyprocta leporina ott462286
Erethizon dorsatum ott848921
Coendou spinosus ott3609652
Lagostomus maximus ott169719
Myocastor coypus ott169715
Ctenomys frater ott350153
Ctenomys lewisi ott327467
Ctenomys conoveri ott350151
Ctenomys boliviensis ott173061
Ctenomys steinbachi ott925570
Ctenomys opimus ott53931
Ctenomys talarum ott672624
Ctenomys leucodon ott53887
Castor canadensis ott886085
Castor fiber ott1048188
Geomys breviceps ott229556
Orthogeomys grandis ott372804
Thomomys bottae ott342237
Thomomys talpoides ott791951
Dipodomys spectabilis ott748116
Dipodomys nitratoides ott666010
Dipodomys merriami ott37456
Dipodomys elator ott241216
Dipodomys ordii ott917328
Dipodomys heermanni ott342226
Dipodomys panamintinus ott342222
Dipodomys gravipes ott666006
Dipodomys microps ott37449
Dipodomys agilis ott734898
Dipodomys venustus ott666013
Dipodomys californicus ott342246
Dipodomys ornatus ott6144701
Dipodomys deserti ott921967
Perognathus flavescens ott1004152
Perognathus flavus ott1024543
Perognathus inornatus ott1052469
Perognathus longimembris ott427268
Perognathus parvus ott1004158
Chaetodipus formosus ott1004150
Chaetodipus hispidus ott427276
Chaetodipus baileyi ott764435
Chaetodipus californicus ott764439
Chaetodipus fallax ott190181
Chaetodipus arenarius ott632596
Chaetodipus spinatus ott190184
Chaetodipus penicillatus ott1004164
Chaetodipus intermedius ott374303
Chaetodipus ammophilus ott330042
Heteromys pictus ott270497
Heteromys irroratus ott844941
Heteromys sp.
Napaeozapus insignis ott379305
Zapus trinotatus ott259968
Zapus princeps ott174407
Zapus hudsonius ott1090614
Nannospalax ehrenbergi ott936616
Nannospalax leucodon ott936600
Tachyoryctes splendens ott756782
Cannomys badius ott3610667
Eospalax baileyi ott45209
Calomyscus sp. ott220330


Ondatra zibethicus ott816252
Synaptomys borealis ott910193
Myopus schisticolor ott490530
Lemmus lemmus ott934731
Lemmus trimucronatus ott498753
Dicrostonyx torquatus ott603156
Phenacomys intermedius ott532489
Arvicola amphibius ott684847
Arvicola sp.
Neodon fuscus ott818953
Alexandromys montebelli ott930591
Alexandromys oeconomus ott7067200
Microtus agrestis ott754333
Microtus subterraneus ott965600
Microtus arvalis ott449577
Microtus californicus ott742231
Microtus mexicanus ott927229
Microtus townsendii ott803867
Microtus sp.
Microtus montanus ott1035816
Microtus pennsylvanicus ott340973
Microtus longicaudus ott910200
Microtus oregoni ott269491
Microtus xanthognathus ott241402
Microtus richardsoni ott803869
Microtus chrotorrhinus ott241400
Microtus pinetorum ott269493
Microtus quasiater ott131164
Microtus miurus ott269522
Microtus ochrogaster ott927240
Craesomys rufocanus ott563745
Myodes macrotis ott910203
Myodes glareolus ott1051603
Myodes gapperi ott627111
Myodes rutilus ott686197
Microtus mystacinus ott7067215
Cricetulus migratorius ott159169
Baiomys taylori ott39171
Reithrodontomys megalotis ott732037
Reithrodontomys humulis ott208414
Reithrodontomys montanus ott208412
Reithrodontomys fulvescens ott39181
Peromyscus difficilis ott983302
Peromyscus attwateri ott983304
Peromyscus truei ott770866
Peromyscus eremicus ott576102
Peromyscus eva ott693029
Peromyscus californicus ott24672
Peromyscus sp.
Peromyscus leucopus ott764828
Peromyscus maniculatus ott377525
Onychomys arenicola ott1004147
Onychomys leucogaster ott427266
Onychomys torridus ott1004154
Peromyscus boylii ott1065599
Peromyscus crinitus ott693032
Peromyscus pectoralis ott1077590
Neotoma floridanus ott114348
Neotoma albigula ott156843
Neotoma micropus ott994834
Neotoma mexicana ott690541
Neotoma stephensi ott15638
Neotoma cinerea ott690529
Neotoma fuscipes ott503776
Neotoma macrotis ott435503
Neotoma lepida ott39195
Neotoma devia ott15637
Sigmodon arizonae ott208420
Sigmodon hispidus ott576095
Sigmodon ochrognathus ott39197
Oxymycterus quaestor ott957195
Akodon montensis ott837919
Thaptomys nigrita ott375640
Necromys lasiurus ott936613
Zygodontomys brevicauda ott223404
Sooretamys angouya ott1039661
Nectomys squamipes ott885389
Melanomys caliginosus ott48259
Oryzomys palustris ott812026
Oryzomys sp.
Oligoryzomys nigripes ott752853
Oecomys mamorae ott558724
Euryoryzomys russatus ott739
Acomys cahirinus ott816259
Acomys dimidiatus ott996659
Lophuromys sikapusi ott173086
Lophuromys flavopunctatus ott159153
Gerbillus nanus ott404777
Rhombomys opimus ott504792
Meriones sp.
Meriones meridianus ott473225
Meriones unguiculatus ott764841
Meriones libycus ott404790
Meriones shawii ott600710
Meriones hurrianae ott3608994
Meriones persicus ott3608993
Meriones tamariscinus ott79088
Tatera indica ott818707
Gerbilliscus guineae ott240714
Maxomys surifer ott474763
Maxomys whiteheadi ott606407
Niviventer fulvescens ott751173
Niviventer sp.
Leopoldamys edwardsi ott366048
Leopoldamys sabanus ott106790
Berylmys bowersi ott366063
Sundamys muelleri ott501407
Bandicota indica ott836346
Bandicota bengalensis ott106786
Bandicota savilei ott836348
Rattus norvegicus ott271555
Rattus nitidus ott92106
Rattus lutreola ott690686
Rattus fuscipes ott271553
Rattus exulans ott283187
Rattus tiomanicus ott977033
Rattus sp.
Rattus tanezumi ott85028
Rattus rattus ott271571
Rattus argentiventer ott977026
Rattus andamanensis ott106800



Supplementary Figure 2. The rodent phylogeny used as a random factor in all REML models. The phylogeny was generated using the Open Tree of Life (OTL) database. Reclassified or renamed taxa not in OTL were added manually. The ott number found after species is the node ID from the OTL reference taxonomy, version 3.3.

Supplementary Table 1. The digital object identifiers (DOIs) of the 344 articles identified in the methodical literature search. Articles are listed by publication year. If the DOI was not available for an article, the PubMed unique identifier (PMID) or JSTOR accession number is provided instead. If all three are not available, the article reference is given.

DOI	Year	DOI	Year
10.1016/j.ijppaw.2020.01.008	2020	10.1007/s00436-019-06530-4	2020
10.1186/s13071-019-3763-6	2019	10.1093/icb/icz136	2019
10.1016/j.ijppaw.2019.07.004	2019	10.1016/j.ijppaw.2019.03.017	2019
10.1016/j.heliyon.2019.e02382	2019	10.1016/j.ejop.2019.02.005	2019
10.1007/s00436-019-06502-8	2019	PMID: 29922616	2018
10.3347/kjp.2018.56.1.93	2018	10.25225/fozo.v67.i2.a1.2018	2018
10.14202/vetworld.2018.293-296	2018	10.1186/s13071-018-3106-z	2018
10.1186/s13071-018-2892-7	2018	10.1186/s13071-018-2802-z	2018
10.1017/S0031182018001142	2018	10.1016/j.ympev.2018.05.009	2018
10.1016/j.ijppaw.2018.01.004	2018	10.1016/j.ijpara.2018.04.003	2018
10.1016/j.ejop.2018.02.001	2018	10.1016/j.ejop.2017.12.006	2018
10.1007/s11230-018-9788-y	2018	10.1007/s11230-017-9771-z	2018
10.1007/s00436-018-5973-9	2018	10.1007/s00436-018-5827-5	2018
PMID: 28979348	2017	PMID: 28935004	2017
10.1017/S0031182017001524	2017	10.1016/j.vetpar.2017.04.007	2017
10.1016/j.ijppaw.2017.05.003	2017	10.1016/j.ejop.2017.09.007	2017
10.1016/j.actatropica.2017.04.013	2017	Rodríguez-Durán et al. 2015. Zootecnia Trop. 33(3):261-268	2016
PMID: 28127340	2016	PMID: 27244955	2016
10.7589/2015-01-010	2016	10.1894/0038-4909-61.4.331	2016
10.1654/1525-2647-83.1.122	2016	10.1638/2015-0055.1	2016
10.1371/journal.pone.0147090	2016	10.1186/s13071-016-1607-1	2016
10.1155/2016/6834206	2016	10.1155/2016/3263868	2016
10.1111/jeu.12249	2016	10.1016/j.vetpar.2015.10.017	2016
10.1016/j.parint.2016.03.010	2016	10.1016/j.meegid.2016.07.014	2016
10.1016/j.ejop.2016.04.008	2016	10.1007/s12639-015-0720-y	2016
10.1002/jez.2003	2016	PMID: 26114139	2015
10.7589/2014-04-099	2015	10.5604/12321966.1141359	2015
10.3347/kjp.2015.53.6.737	2015	10.3201/eid2112.141711	2015
10.1654/4689.1	2015	10.1080/00222933.2013.825025	2015
10.1017/S0031182014001929	2015	10.1016/j.meegid.2015.10.002	2015
10.1016/j.meegid.2015.03.003	2015	10.1016/j.ijppaw.2015.02.004	2015
PMID: 25776597	2014	PMID: 25642271	2014
10.2478/s11686-014-0304-5	2014	10.2478/bvip-2014-0033	2014
10.1890/13-2381.1	2014	10.1080/00222933.2013.867376	2014
10.1017/S003118201300139X	2014	10.1016/j.meegid.2013.07.020	2014
10.1007/s11230-013-9466-z	2014	10.7589/2013-02-028	2013
10.1645/GE-3144.1	2013	10.1645/12-94.1	2013

10.12816/0006382	2013	10.1186/1746-6148-9-229	2013
10.1128/AEM.01503-13	2013	10.1111/jzo.12076	2013
10.1098/rspb.2013.0598	2013	10.1098/rsbl.2013.0205	2013
10.1017/S0950268812002609	2013	10.1016/j.vetpar.2013.02.011	2013
10.1016/j.meegid.2013.01.011	2013	10.1016/j.ijpara.2013.04.007	2013
10.1016/j.exppara.2013.09.003	2013	Kozerski et al. 2012. Arq. Ciênc. Vet. Zool. UNIPAR, Umuarama. 15(2):133-136.	2012
10.5897/ajmr11.1356	2012	10.5812/jjm.3580	2012
10.2478/s11687-012-0033-y	2012	10.2478/s11686-012-0016-7	2012
10.2478/s11686-012-0006-9	2012	10.1590/S1519-69842012000300019	2012
10.1139/Z2012-083	2012	10.1016/j.exppara.2012.05.009	2012
10.1007/s10344-011-0584-0	2012	10.1007/s00508-012-0237-7	2012
10.2478/s11687-011-0025-3	2011	10.1645/GE-2535.1	2011
10.1017/S0031182011001107	2011	Futagbi et al. 2010. West Afr. J. Appl. Ecol. 17(1):81-87.	2010
10.9775/kvfd.2009.1269	2010	10.7589/0090-3558-46.1.146	2010
10.1501/Vetfak_0000002378	2010	10.1501/Vetfak_0000002326	2010
10.1128/AEM.00687-10	2010	10.1111/j.1469-7998.2010.00734.x	2010
10.1017/S003118200999179X	2010	10.1016/j.vetpar.2010.02.012	2010
Ozmen et al. 2009. Turkiye Parazitol. Derg. 33(3):245-247.	2009	PMID: 19887025	2009
10.2307/26327900	2009	10.1645/GE-1653.1	2009
10.1590/S0103-84782009005000085	2009	10.1128/AEM.01386-09	2009
10.1111/j.1365-2028.2008.01020.x	2009	10.1590/S0103-84782008000200043	2008
10.14411/fp.2008.012	2008	10.1111/j.1749-4877.2008.00069.x	2008
10.1017/S0031182008004952	2008	10.1007/s11230-008-9150-x	2008
PMID: 18247463	2007	10.1654/4269.1	2007
10.1590/S0074-02762007005000008	2007	10.1128/AEM.01034-07	2007
10.1128/AEM.00848-07	2007	10.1017/S0031182006001120	2007
10.1016/j.vetpar.2007.03.034	2007	10.1016/j.vetpar.2007.03.024	2007
10.1016/j.exppara.2007.03.014	2007	10.1007/s00436-007-0488-9	2007
10.1007/s00436-006-0276-y	2007	10.1007/s00436-006-0251-7	2007
10.1002/hyp.6278	2007	10.1638/06-013.1	2006
10.1139/Z06-183	2006	10.1111/j.1600-0587.1998.tb00667.x	2006
10.1111/j.1469-7998.2006.00089.x	2006	10.7589/0090-3558-41.2.442	2005
10.1645/ge-564r.1	2005	10.1645/ge-466r.1	2005
10.1645/ge-3468.1	2005	10.1128/AEM.71.10.5929-5934.2005	2005
10.1051/parasite/2005124305	2005	10.1007/s00436-005-1459-7	2005
10.1007/s00436-005-1353-3	2005	Drózdz et al. 2004. Helminthologia. 41(2):99-101.	2004
10.1590/S0074-02762004000600008	2004	10.1292/jvms.66.983	2004
10.1128/AEM.70.12.7574-7577.2004	2004	10.1128/AEM.70.11.6748-6752.2004	2004
10.1016/j.vetpar.2004.04.005	2004	10.7589/0090-3558-39.4.762	2003

10.3201/eid0911.030084	2003	10.1292/jvms.65.1151	2003
10.1007/s00442-003-1345-x	2003	McAllister and Kessler 2002. J. Ark. Acad. Sci. 56(35):235-238	2002
10.5070/v420110115	2002	10.2307/3285505	2002
10.1654/1525-2647(2002)069[0211:ELAEOF]2.0.CO;2	2002	10.1139/w02-047	2002
10.1079/JOH2001109	2002	10.1017/S0031182002001865	2002
10.1016/S0304-4017(02)00181-4	2002	10.2307/3285195	2001
10.2307/3285191	2001	10.2307/3285044	2001
10.1674/0003-0031(2001)145[0409:POEAEI]2.0.CO;2	2001	10.1186/1751-0147-42-479	2001
10.1128/AEM.67.6.2840-2843.2001	2001	10.1128/AEM.67.3.1154-1162.2001	2001
10.1080/15627020.2001.11657112	2001	10.1017/S0952836901000504	2001
10.1017/S0031182001007259	2001	10.14411/fp.2000.020	2000
10.1017/S0031182099005545	2000	10.1016/S0932-4739(00)80037-1	2000
10.1016/s0932-4739(00)80023-1	2000	10.1016/S0304-4017(00)00331-9	2000
O'Callaghan and O'Donoghue 1999. Trans. R. Soc. S. Aust. 123(4):133-135.	1999	10.7589/0090-3558-35.4.660	1999
10.4067/S0716-07201999000300006	1999	10.2307/3285824	1999
10.1016/S0304-4017(98)00226-X	1999	Bomfim and Lopes 1998. Rev. Bras. Parasitol. Vet. 7(2):129-136.	1998
10.15517/rbt.v46i2.19534	1998	10.2307/3284675	1998
10.2307/3284485	1998	10.1080/09603129873660	1998
10.1017/S0952836998001137	1998	Bajer et al. 1997. Acta Parasitol. 42(4):192-194.	1997
10.2307/3284455	1997	10.2307/3284412	1997
10.2307/3284341	1997	10.2307/2426760	1997
10.1051/parasite/1997044369	1997	10.1016/S0022-1759(96)00239-6	1997
10.1007/s004360050283	1997	10.2307/3545876	1996
10.2307/3284150	1996	10.2307/2265712	1996
10.1016/S0934-8840(96)80106-X	1996	10.1016/S0035-9203(96)90235-7	1996
McKnown et al. 1995. J. Helminthol. Soc. Wash. 62(1):89-93.	1995	10.5962/bhl.title.156442	1995
10.2307/3283946	1995	10.2307/3283928	1995
10.1590/S0074-02761995000600002	1995	10.1017/S0031182000081804	1995
Thomas and Stanton 1994. J. Helminthol. Soc. Wash. 61(1):17-21.	1994	10.7589/0090-3558-30.3.450	1994
10.7589/0090-3558-30.3.359	1994	10.7589/0090-3558-30.1.110	1994
10.2307/3283755	1994	10.1139/z94-032	1994
10.1016/S0932-4739(11)80025-8	1994	PMID: 8013931	1993
McAllister et al. 1993. J. Helminthol. Soc. Wash. 60(2):256-258.	1993	10.7589/0090-3558-29.1.161	1993
10.2307/3283744	1993	10.2307/3283743	1993
10.2307/3283505	1993	10.1080/09603129309356785	1993
JSTOR: 41712736	1992	10.2307/3283636	1992

10.2307/3283482	1992	10.2307/3283322	1992
10.1111/j.1550-7408.1992.tb04866.x	1992	10.1080/00480169.1992.35693	1992
PMID: 1937279	1991	10.2307/3282760	1991
10.2307/3226742	1991	10.1016/S0003-9365(11)80187-9	1991
Miyaji et al. 1989. Jpn. J. Parasitol. 38:368-372.	1990	Yamaura et al. 1990. Jpn. J. Parasitol. 39:439-444.	1990
JSTOR: 41712612	1990	10.2307/3282798	1990
10.2307/3282659	1990	10.1128/aem.56.1.31-36.1990	1990
10.1111/j.1550-7408.1990.tb01261.x	1990	10.1111/j.1550-7408.1990.tb01115.x	1990
Seville and Williams 1989. Proc. Helminthol. Soc. Wash. 56(2):204-206.	1989	10.2307/3282919	1989
10.1111/j.1550-7408.1989.tb01068.x	1989	10.1017/S0031182000061394	1989
10.2508/chikusan.59.351	1988	10.2307/3282497	1988
10.2307/3282226	1988	10.2307/3226510	1988
10.1258/002367788780746665	1988	10.1128/aem.54.11.2777-2785.1988	1988
JSTOR: 44539406	1987	10.2307/3282103	1987
Kietzmann and Kietzmann 1987. Proc. Iowa Acad. Sci. 94(1):22-23.	1987	10.7589/0090-3558-23.4.576	1987
10.1128/aem.53.8.1790-1792.1987	1987	10.1128/aem.53.7.1574-1579.1987	1987
10.7589/0090-3558-22.1.115	1986	10.1139/m86-171	1986
10.1111/j.1550-7408.1986.tb05607.x	1986	Upton et al. 1985. Proc. Helminthol. Soc. Wash. 55(1):60-63.	1985
10.2307/3282011	1985	10.2307/3281719	1985
10.2307/3281445	1985	10.2307/3281432	1985
10.1128/aem.50.1.177-178.1985	1985	10.1111/j.1550-7408.1985.tb04059.x	1985
10.7589/0090-3558-20.4.279	1984	10.2307/3281661	1984
10.1111/j.1550-7408.1984.tb02976.x	1984	10.1111/j.1469-7998.1984.tb05089.x	1984
10.2307/3281301	1983	10.1111/j.1365-2907.1983.tb00276.x	1983
10.1007/BF00009160	1983	10.2307/3281108	1982
10.1111/j.1550-7408.1982.tb04006.x	1982	10.1017/S0031182000056225	1982
JSTOR: 44536562	1980	10.2307/3280827	1980
10.1111/j.1550-7408.1980.tb05376.x	1980	10.1093/oxfordjournals.aje.a113019	1980
Ryšavý and Černá 1979. Folia Parasitol. 26(1):65-68.	1979	Fleming et al. 1979. Proc. Helminthol. Soc. Wash. 46(1):115-127.	1979
Levine and Husar 1979. Proc. Helminthol. Soc. Wash. 46(1):135-137.	1979	10.1111/j.1550-7408.1979.tb04189.x	1979
10.2307/3279841	1977	10.1111/j.1550-7408.1977.tb00966.x	1977
10.1258/002367776781035314	1976	10.2307/3278353	1974
10.1111/j.1550-7408.1972.tb03453.x	1972	10.1111/j.1550-7408.1972.tb03429.x	1972

10.2307/3278801	1971	10.2307/3277983	1971
10.2307/1378687	1971	10.1139/z71-109	1971
10.1111/j.1550-7408.1971.tb03383.x	1971	10.1111/j.1550-7408.1971.tb03379.x	1971
10.1111/j.1550-7408.1971.tb03311.x	1971	de Vos and Dobson 1970. Onderstepoort J. Vet. Res. 37(4):185-190.	1970
10.7589/0090-3558-6.2.107	1970	10.2307/3277446	1970
10.1139/z70-109	1970	10.1111/j.1550-7408.1980.tb05377.x	1970
10.1111/j.1550-7408.1970.tb02368.x	1970	10.1139/z69-123	1969
10.1111/j.1550-7408.1970.tb04713.x	1969	10.1016/0014-4894(69)90107-6	1969
10.2307/3277107	1968	10.1111/j.1550-7408.1968.tb02203.x	1968
10.1111/j.1550-7408.1968.tb02152.x	1968	10.1111/j.1550-7408.1968.tb02129.x	1968
10.1111/j.1550-7408.1968.tb02121.x	1968	10.1111/j.1550-7408.1968.tb02083.x	1968
10.1080/00034983.1968.11686557	1968	10.1111/j.1550-7408.1967.tb02044.x	1967
10.1111/j.1550-7408.1967.tb02040.x	1967	10.1111/j.1550-7408.1967.tb01980.x	1967
10.1111/j.1550-7408.1966.tb01961.x	1966	10.1111/j.1550-7408.1965.tb03245.x	1965
10.1111/j.1550-7408.1965.tb03236.x	1965	10.1111/j.1550-7408.1964.tb01724.x	1964
10.2307/3275782	1963	10.1111/j.1550-7408.1962.tb02615.x	1962
10.2307/3275485	1961	10.2307/3274986	1961
10.2307/3275173	1960	10.1111/j.1550-7408.1960.tb00731.x	1960
10.1111/j.1550-7408.1960.tb00721.x	1960	10.1111/j.1550-7408.1960.tb00709.x	1960
10.1080/00034983.1960.11685992	1960	10.1111/j.1550-7408.1959.tb04360.x	1959
10.1111/j.1469-7998.1959.tb05528.x	1959	10.1111/j.1550-7408.1958.tb02530.x	1958
10.1080/00034983.1958.11685840	1958	Levine et al. 1957. Trans. Ill. Acad. Sci. 50:291-299.	1957
10.1111/j.1550-7408.1957.tb02491.x	1957	10.1111/j.1096-3642.1957.tb00272.x	1957
10.2307/3274144	1955	10.1017/S0022172400036123	1948
10.1017/S0031182000010763	1938	10.1093/oxfordjournals.aje.a118249	1936
10.1017/S0031182000023866	1934	10.1093/oxfordjournals.aje.a117882	1932
10.1111/j.1096-3642.1931.tb01037.x	1931	10.1001/jama.1915.02580260026009	1915

Appendix 1.2: Chapter 3 supplementary material

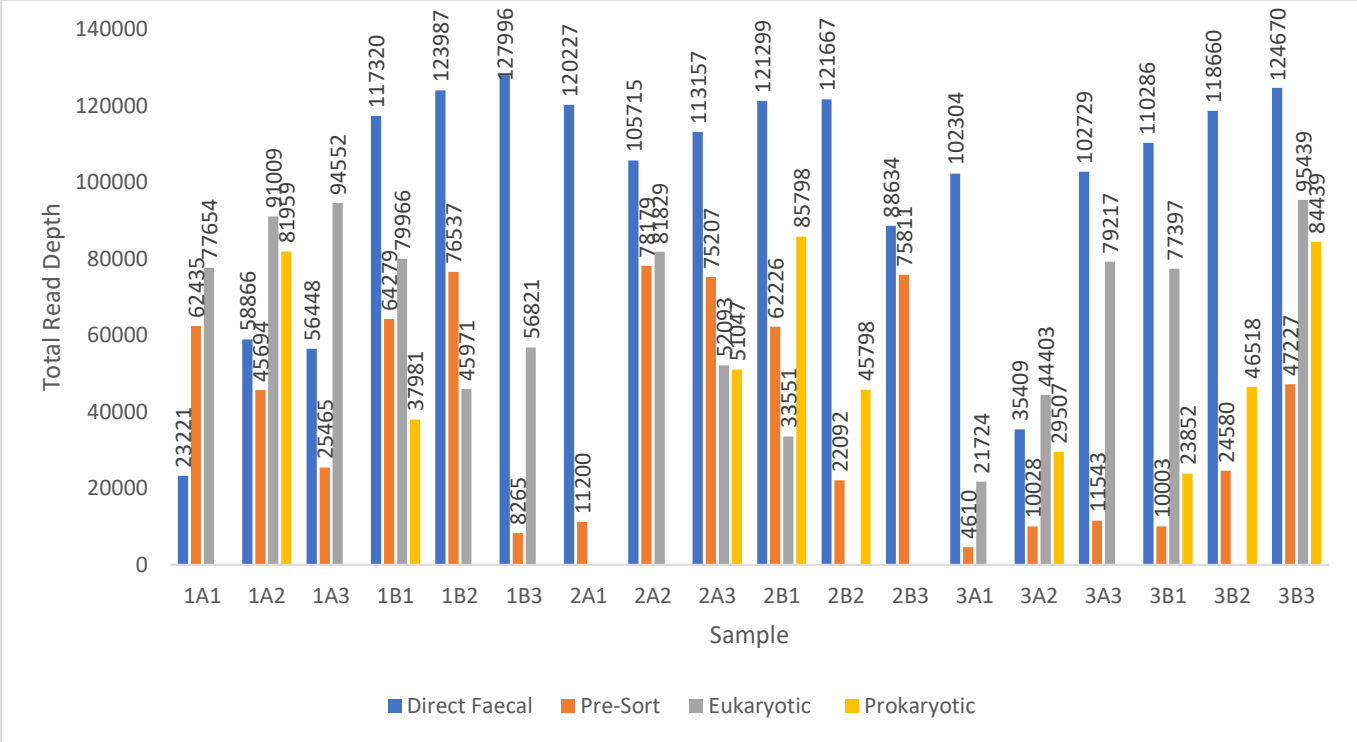
Methods development for sorting eukaryotic and prokaryotic cells by FACS

The specificity of the eukaryotic-specific stain (ERT, ER-Tracker Red (Invitrogen^{TM})) was first confirmed by an *in-silico* study. The ERT stain targets the sulphonylurea receptors (SURs) of ATP-sensitive potassium channels, which are common on the endoplasmic reticulum, a eukaryotic-specific organelle (Invitrogen, 2020). SUR proteins are also unique to eukaryotes, and so should not be found in prokaryotic taxa (Burke *et al.*, 2008). Thus, a BLASTP search was conducted to confirm the absence of SUR protein sequences in prokaryotes (Altschul *et al.*, 1990). The query sequences were for the proteins SUR1 and SUR2 (Burke *et al.*, 2008; UniProt Consortium, 2021). Searches for the query sequences were restricted to Bacteria. The BLASTP search found that the highest percent identity score for SUR1 and SUR2 was 33.9% and 33.8% respectively, indicating no similar protein is found in bacteria. Thus, ERT was considered capable of staining eukaryotic cells, but not prokaryotic cells.

Next, the specificity of the ERT stain was tested using *Tetrahymena pyriformis* and *Escherichia coli* as positive and negative controls, respectively. *T. pyriformis* is a single-celled eukaryotic species and *E. coli* is single-celled prokaryotic species. Cultures of each control were stained and processed using flow cytometry, as described in Chapter 3 (section 3.2.1). As the eukaryotic, positive control, *T. pyriformis* should be identified as ERT-positive, whereas the prokaryotic *E. coli* cells should be identified as ERT-negative. 92% of 5,025 cells in the *E. coli* culture were stained with SYBR-green (SG) but not ERT (Supplementary Table 2). This suggested that the ERT could not generally stain *E. coli.* However, 7.3% of cells in the *E. coli* culture were ERT-positive, suggesting either contamination with eukaryotic cells, or ERT-staining of a small group of potentially contaminating prokaryotic taxa. For *T. pyriformis,* 566 cells were counted. The lower cell count was expected in comparison to *E. coli* culutres*,* as the cultures of *T. pyriformis* were less dense. 99.1% of cells were stained with ERT, suggesting the ERT could stain these eukaryotic cells, suggesting a high staining-specificity of the eukaryotic, positive control.

Supplementary Table 2. The cell counts for *E. coli* and *T. pyriformis* using FACS, when stained using SG and ERT. Any event in FACS that stained with SG is classed as a cell. *E. coli* and *T. pyriformis* cultures were processed and counted separately.

	Escherichia coli	*Tetrahymena pyriformis*
Number of SG-positive cells	5025	566
Number of SG-positive, ERT-negative cells	4624	4
Number of SG-positive, ERT-positive cells	366	561
SG-positive, ERT-positive %	7.28%	99.12%



Supplementary Figure 3. The total read depth for each sample, per fraction type.

References

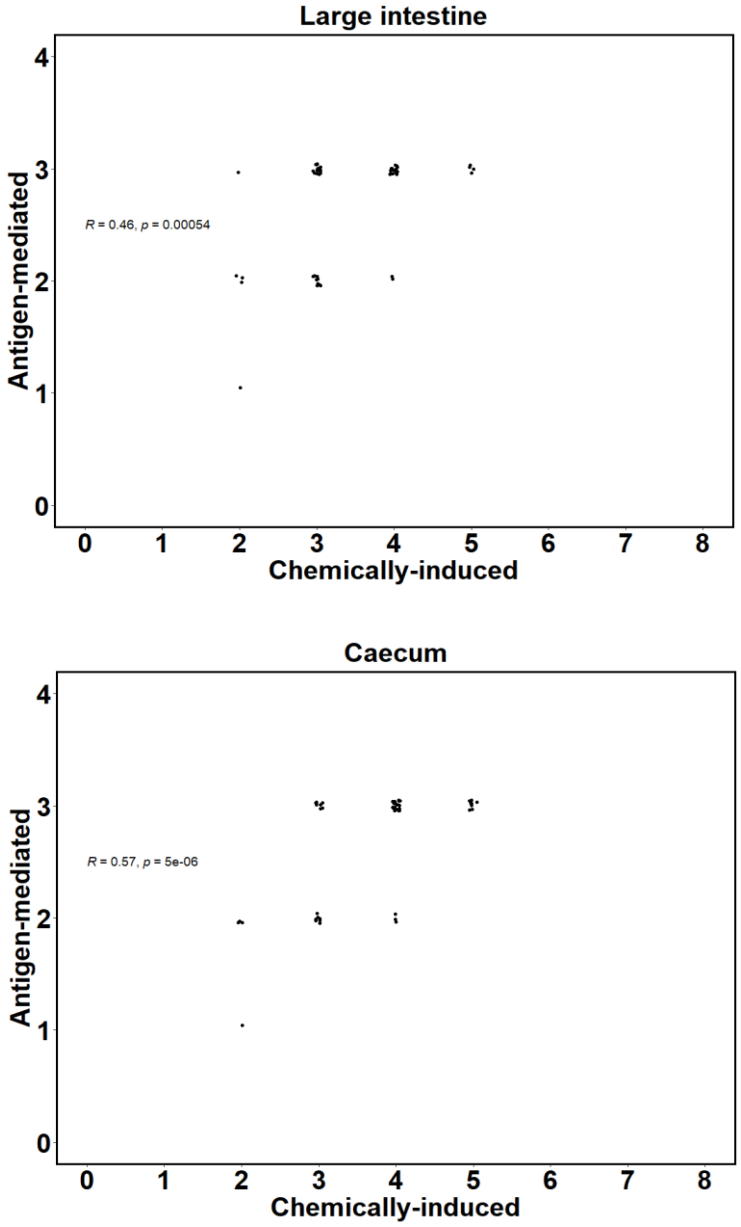
Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp. 403–410.

Burke, M. A., Mutharasan, R. K. and Ardehali, H. (2008). The sulfonylurea receptor, an atypical ATP-binding cassette protein, and its regulation of the KATP channel. *Circulation Research,* 102(2), pp. 164-176.

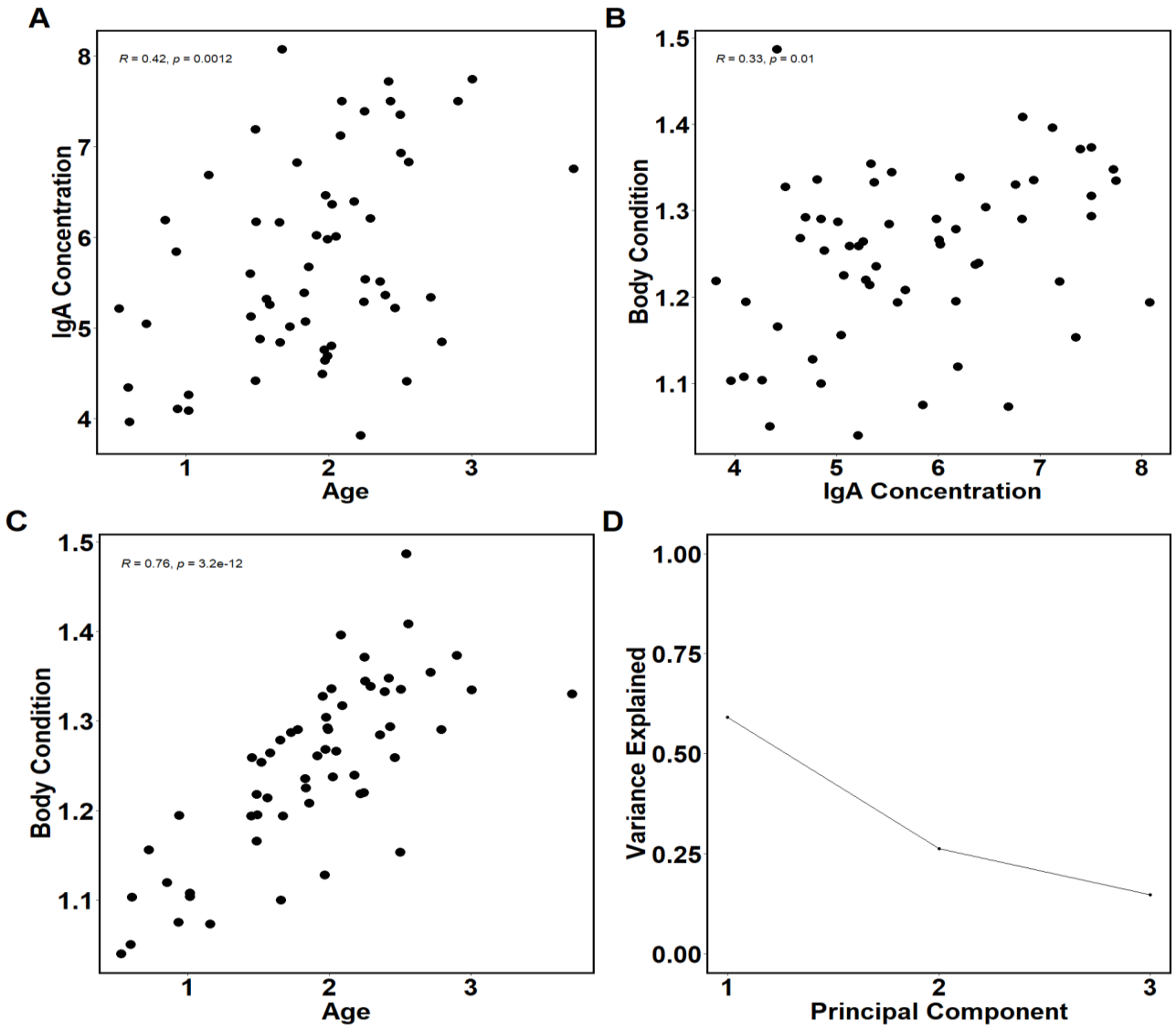
Invitrogen (2023). ER-Tracker™ Red (BODIPY™ TR Glibenclamide), for live-cell imaging. Available at: https://www.thermofisher.com/order/catalog/product/E34250?SID=srch-srp-E34250 (Accessed: 16 April 2023).

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), pp. D480–D489.

Appendix 1.3: Chapter 4 supplementary material



Supplementary Figure 4. Correlation between chemically-induced inflammation and luminal antigen-mediated inflammation in the A) large intestine and B) caecum was identified. Both correlation tests used Spearman's *rho* statistic, shown as R on each plot.



E

Variable	PC1	PC2	PC3
Faecal IgA concentration	-0.467012	0.881863	-0.06494
Age	-0.63185	-0.28143	0.722191
Body condition	-0.6186	-0.3783	-0.68864

Supplementary Figure 5. The correlation for three host factors (A-C) and the results of the subsequent principal component analysis (D-E). A) Faecal IgA concentration and age, B) Faecal IgA concentration and body condition, C) age and body condition. All three correlation tests used Pearson's product moment correlation coefficient. Body condition is log(body mass)/log(body length), faecal IgA concentration is log(mg/g) and age is log(weeks). D) The amount of variance in the three host factors explained by each principal component. E) The loadings of each host factors onto the principal components. A positive (negative) loading indicates a positive (negative) correlation between a variable and a principal component.

Supplementary Table 3. Chi-squared comparisons of gut inflammation scores for each tissue type for the three sampling sites. N is the sample size for each combination of sampling site and tissue type.

	Nottingham			Southport			Wirral		
	N	***X²***	***p value***	**N**	***X²***	***p value***	**N**	***X²***	***p value***
Small Intestine	30	7.2	0.206	15	15.8	0.007	12	8.0	0.156
Caecum	30	43.4	< 0.001	13	7.5	0.023	12	6.0	0.050
Large Intestine	29	33.2	< 0.001	12	6.5	0.039	12	6.0	0.050

Supplementary Table 4. Host factors for mice identified as having localised intestinal inflammation in the large intestine and caecum. Localised inflammation was defined as a difference of >1 between tissue replicates. Eight mice had localised inflammation in the large intestine and one mouse had localised inflammation in the caecum.

Sampling site	Sex	Reproductive status	Eimeria infection	Nematode infection
Nottingham	Male	Inactive	Positive	Negative
Nottingham	Male	Active	Positive	Positive
Southport	Male	Inactive	Negative	Negative
Southport	Male	Active	Negative	Negative
Wirral	Female	Inactive	Positive	Negative
Wirral	Male	Active	Positive	Negative
Wirral	Female	Inactive	Negative	Negative
Wirral	Male	Inactive	Positive	Negative
Southport[a]	Male	Inactive	Negative	Negative

a Localised inflammation in the caecum

Supplementary Table 5. Chi-squared comparisons of *Eimeria* infection status for the three sampling sites. The number in brackets is the sample size for each site.

	Nottingham (31)		Southport (15)		Wirral (12)	
	X²	***p value***	***X²***	***p value***	***X²***	***p value***
Faecal	27.1	< 0.001	5.4	0.020	0.3	0.564
Caecal	20.1	< 0.001	8.1	0.005	1.3	0.248

Supplementary Table 6. Phyla identified from the faecal eukaryome of wild mice. Phyla are ordered by their relative abundance across the entire dataset. Total abundance is the number of sequences classified for each phylum. Prevalence is the percentage of mice the phyla were identified from. Bold phyla are those with species known to inhabit the mammalian gut microbiome. N'ham = Nottingham, S'port = Southport.

Phylum	Abundance		Prevalence (%)			
	Relative (%)	Total	Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Ascomycota	42.5	1080915	100	100	100	100
Basidiomycota	32.0	815080	100	100	100	100
Mucoromycota	13.4	341999	100	100	100	100
Apicomplexa	9.2	233732	88	100	80	67
Nematoda[a]	0.7	18520	60	94	0	50
Cercozoa	0.6	15977	88	94	87	75
Bigyra[b]	0.5	13671	14	19	0	17
Ciliophora	0.3	6749	78	84	67	75
Rotifera	0.1	3613	17	26	13	0
Ochrophyta	0.1	3585	67	68	67	67
Unclassified[c]	0.1	3426	81	87	67	83
Amoebozoa	0.1	2342	31	32	40	17
Chytridiomycota	0.1	1354	62	58	60	75
Peronosporomycetes	<0.05	874	41	35	60	33
Cryptomycota	<0.05	784	33	35	27	33
Holozoa	<0.05	266	19	10	13	50
Protosteliida	<0.05	191	28	39	7	25
Zoopagomycota	<0.05	142	19	6	47	17
LKM15	<0.05	102	16	16	20	8
Blastocladiomycota	<0.05	99	3	0	7	8
Aphelidea	<0.05	94	9	10	7	8
Centrohelida	<0.05	76	17	26	7	8
Schizoplasmodiida	<0.05	72	9	13	0	8
Dinoflagellata	<0.05	70	5	3	0	17
Protalveolata	<0.05	53	9	0	13	25
Gracilipodida	<0.05	43	9	10	7	8
NucFont[d]	<0.05	36	12	13	13	8
Dictyostelia	<0.05	30	2	0	0	8
Protosporangiida	<0.05	26	5	10	0	0
IC_Nephridiophaga	<0.05	13	2	3	0	0
Labyrinthulomycetes	<0.05	10	5	3	13	0
Bicosoecida	<0.05	10	3	0	0	17

a Nematozoa (SILVA-defined)
b SAR (SILVA-defined)
c ASVs not classified at the phylum level
d Nucleariidae and Fonticula group

Supplementary Table 7. Phyla identified from the caecal eukaryome of wild mice. Phyla are ordered by their relative abundance across the entire dataset. Total abundance is the number of sequences classified for each phylum. Prevalence is the percentage of mice the phyla were identified from. Bold phyla are those with species known to inhabit the mammalian gut microbiome. N'ham = Nottingham, S'port = Southport

Phylum	Abundance		Prevalence (%)			
	Relative (%)	Total	Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Nematoda[a]	41.1	1011293	72	100	27	58
Basidiomycota	24.5	603440	100	100	100	100
Ascomycota	24.2	595891	100	100	100	100
Apicomplexa	6.0	147763	88	100	73	75
Mucoromycota	3.5	86044	98	100	93	100
Bigyra[b]	0.14	3461	7	13	0	0
Cercozoa	0.14	3355	57	52	80	42
Ciliophora	0.07	1811	52	52	33	75
Unclassified[c]	0.07	1680	48	32	67	67
Ochrophyta	0.06	1358	40	35	33	58
Chytridiomycota	<0.05	867	31	32	27	33
Peronosporomycetes	<0.05	216	14	16	7	17
Cryptomycota	<0.05	199	12	16	0	17
Aphelidea	<0.05	162	7	13	0	0
Holozoa	<0.05	145	9	3	0	33
Dinoflagellata	<0.05	132	3	6	0	0
Protosteliida	<0.05	129	9	16	0	0
Amoebozoa	<0.05	103	5	10	0	0
Zoopagomycota	<0.05	75	3	3	7	0
Protalveolata	<0.05	65	3	0	0	17
Bicosoecida	<0.05	29	2	0	0	8
Schizoplasmodiida	<0.05	23	2	3	0	0
Centrohelida	<0.05	22	3	6	0	0
NucFont[c]	<0.05	12	2	3	0	0
Labyrinthulomycetes	<0.05	10	2	3	0	0

a Nematozoa (SILVA-defined)
b SAR (SILVA-defined)
c ASVs not classified at the phylum level
d Nucleariidae and Fonticula group

Supplementary Table 8. The percentage of *Eimeria*-positive mice when comparing between PCR diagnostics and 18S sequencing. The percentage of mice positive for *Eimeria* is shown for both faecal and caecal samples. The number in brackets is the sample size for each site.

	Combined (58)		Nottingham (31)		Southport (15)		Wirral (12)	
Sample Type	**PCR**	**18S**	**PCR**	**18S**	**PCR**	**18S**	**PCR**	**18S**
Faecal	69	41	97	77	20	0	58	0
Caecal	59	43	90	77	13	0	33	8

Supplementary Table 9. Test statistics for comparing eukaryome Shannon's index among sampling sites at different taxonomic ranks. Comparison among faecal samples at the ASV, genus and family level used a Kruskal-Wallis test (test statistic: X^2) whereas all other comparison used an ANOVA (test statistic: F). Significant *p* values are highlighted in bold. Post hoc comparisons are shown for those taxonomic levels with a significant difference in alpha diversity between sampling sites. N'ham = Nottingham, S'port = Southport.

		Shannon's Index (± SE)			Test		Post hoc	
	Rank	**N'ham**	**S'port**	**Wirral**	**Test Statistic**	***p* value**	**Comparisons**	***p* value**
Faecal	ASV	2.4 (0.1)	1.91 (0.2)	2.0 (0.3)	3.47	0.176		
	Genus	2.2 (0.1)	1.8 (0.2)	1.9 (0.3)	3.24	0.200		
	Family	2.1 (0.1)	1.8 (0.2)	1.9 (0.3)	2.10	0.350		
	Phylum	1.0 (0.0)	0.8 (0.1)	0.9 (0.1)	3.34	**0.042**	N'ham: S'port	0.044
Caecal	ASV	1.4 (0.2)	1.6 (0.2)	2.0 (0.2)	2.62	0.082		
	Genus	1.3 (0.2)	1.5 (0.2)	2.0 (0.2)	3.40	**0.041**	N'ham: Wirral	0.031
	Family	1.2 (0.1)	1.4 (0.1)	1.9 (0.1)	3.68	**0.032**	N'ham: Wirral	0.024
	Phylum	0.7 (0.1)	0.7 (0.1)	0.9 (0.1)	2.22	0.119		

Supplementary Table 10. Test statistics for comparing eukaryome composition among sampling sites at different taxonomic ranks. A PERMANOVA was used to compare among the sampling sites for each taxonomic rank before conducting pairwise comparisons between the sampling sites. Significant pairwise comparisons are highlighted in bold.

	Rank	F value	R^2	*p value*	Pairwise comparison	*p value*
Faecal	ASV	7.37	0.211	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	0.065
	Genus	8.05	0.226	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	0.084
	Family	8.42	0.235	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	0.077
	Phylum	12.14	0.306	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	0.483
Caecal	ASV	9.79	0.262	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	**0.032**
	Genus	10.99	0.286	< 0.001	**Nottingham: Southport**	**0.002**
					Nottingham: Wirral	**0.002**
					Wirral: Southport	**0.036**
	Family	11.80	0.3000	< 0.001	**Nottingham: Southport**	**0.001**
					Nottingham: Wirral	**0.001**
					Wirral: Southport	**0.030**
	Phylum	21.04	0.433	< 0.001	**Nottingham: Southport**	**0.001**
					Nottingham: Wirral	**0.001**
					Wirral: Southport	0.207

Supplementary Table 11. Test statistics for comparing the amount of variation of eukaryome composition within sampling sites at different taxonomic ranks. A permutest was used to compare among the sampling sites for each taxonomic rank before conducting pairwise comparisons between the sampling sites. Significant pairwise comparisons are highlighted in bold.

	Rank	F value	*p value*	Pairwise comparisons	*p value*
Faecal	ASV	15.52	< 0.001	**Nottingham: Southport**	**0.001**
				Nottingham Wirral	**0.001**
				Wirral: Southport	0.676
	Genus	11.61	< 0.001	**Nottingham: Southport**	**0.002**
				Nottingham Wirral	**0.001**
				Wirral: Southport	0.665
	Family	8.90	< 0.001	**Nottingham: Southport**	**0.001**
				Nottingham Wirral	**0.002**
				Wirral: Southport	0.544
	Phylum	3.08	0.052	Nottingham: Southport	0.050
				Nottingham Wirral	0.081
				Wirral: Southport	0.867
Caecal	ASV	2.84	0.062	**Nottingham: Southport**	**0.025**
				Nottingham Wirral	0.257
				Wirral: Southport	0.326
	Genus	1.21	0.292	Nottingham: Southport	0.133
				Nottingham Wirral	0.464
				Wirral: Southport	0.500
	Family	2.28	0.110	Nottingham: Southport	0.053
				Nottingham Wirral	0.303
				Wirral: Southport	0.392
	Phylum	0.989	0.370	Nottingham: Southport	0.222
				Nottingham Wirral	0.534
				Wirral: Southport	0.458

Supplementary Table 12. Species that were differentially abundant between the faecal samples of Nottingham mice and mice from the two other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance, with the difference in abundance between Nottingham mice and Combined mice (Wirral and Southport) provided. Species highlighted in bold had a greater abundance in Nottingham mice compared to Southport and Wirral mice. The *p* values used to identify significantly different species were generated using an ANCOM-BC analysis.

Phylum	Species	Abundance			*p* value
		Nottingham	Combined	Difference	
Nematoda	***Ozolaimus linstowi***	3.16 (± 0.5)	-0.69 (± 0.1)	3.85	<0.001
Apicomplexa	***Eimeria falciformis***	2.46 (± 0.8)	-0.69 (± 0.1)	3.15	0.02
	Eimeria telekii	4.29 (± 0.8)	-0.69 (± 0.1)	4.99	<0.001
	Gregarina polymorpha	3.2 (± 0.4)	0.56 (± 0.4)	2.64	0.003
Ascoymocta	*Wickerhamomyces anomalus*	0.01 (± 0.3)	2.74 (± 0.6)	2.72	0.002
	Ascosphaera apis	-0.52 (± 0.2)	1.61 (± 0.4)	2.13	0.004
	Onygena corvina	-0.73 (± 0)	1.32 (± 0.4)	2.05	0.007
	Thermomyces lanuginosus	3.51 (± 0.4)	0.6 (± 0.3)	2.90	<0.001
Basidiomycota	*Wallemia sebi*	6.9 (± 0.3)	9.91 (± 0.2)	3.00	<0.001
	Wallemia ichthyophaga	-0.5 (± 0.1)	2.94 (± 0.4)	3.44	<0.001
	Stropharia ambigua	-0.56 (± 0.1)	1.51 (± 0.4)	2.07	0.006
	Sterigmatomyces halophilus	1.48 (± 0.4)	5.73 (± 0.3)	4.25	<0.001
	Cystobasidium lysinophilum	4.51 (± 0.5)	1.25 (± 0.5)	3.27	<0.001
Mucoromycota	***Circinella umbellata***	4.53 (± 0.6)	0.71 (± 0.4)	3.82	0.001

Supplementary Table 13. Phyla that were differentially abundant between the faecal and caecal samples of Nottingham mice and mice from the two other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance, with the difference in abundance between Nottingham mice and Combined mice (Wirral and Southport) provided. Phyla highlighted in bold had a greater abundance in Nottingham mice compared to Southport and Wirral mice. The *p* values used to identify significantly different phyla were generated using an ANCOM-BC analysis.

	Phylum	**Abundance**			***p* value**
		Nottingham	**Combined**	**Difference**	
Faecal	**Apicomplexa**	0.7 (± 0.3)	-1.98 (± 0.3)	2.7	<0.001
	Basidiomycota	1.19 (± 0.3)	5.47 (± 0.2)	4.3	<0.001
	Nematoda	-1.89 (± 0.3)	-3.49 (± 0.3)	1.6	<0.001
Caecal	**Apicomplexa**	5.26 (± 0.4)	1.67 (± 0.4)	3.6	<0.001
	Ascomycota	6.75 (± 0.2)	8.58 (± 0.2)	1.8	0.03
	Basidiomycota	5.17 (± 0.3)	8.4 (± 0.2)	3.2	<0.001
	Nematoda	8.17 (± 0.4)	0.78 (± 0.7)	7.4	<0.001

Supplementary Table 14. Phyla identified from the faecal bacteriome of wild mice. Phyla are ordered by their relative abundance across the entire dataset. Total abundance is the number of sequences classified for each phylum. Prevalence is the percentage of mice the phyla were identified from. N'ham = Nottingham, S'port = Southport.

Phylum	Abundance		Prevalence (%)			
	Relative (%)	Total	Combined (58)	N'ham (31)	S'port (15)	Wirral (12)
Firmicutes	61.8	4561505	100	100	100	100
Bacteroidota	29.9	2205342	100	100	100	100
Actinobacteriota	2.2	166016	100	100	100	100
Campilobacterota	2.0	145104	100	100	100	100
Desulfobacterota	1.9	138905	98	100	100	92
Proteobacteria	1.5	112720	100	100	100	100
Verrucomicrobiota	0.2	17883	14	16	13	8
Deferribacterota	0.2	17445	72	94	40	58
Cyanobacteria	0.1	4898	55	74	40	25
Patescibacteria	0.1	4043	55	77	20	42
Fusobacteriota	0.1	3854	3	3	0	8
Spirochaetota	< 0.05	1904	5	10	0	0
Chloroflexi	< 0.05	1307	31	45	13	17
Deinococcota	< 0.05	414	10	13	0	17
Acidobacteriota	< 0.05	91	2	0	7	0
Planctomycetota	< 0.05	53	2	3	0	0
Bdellovibrionota	< 0.05	14	2	3	0	0
Myxococcota	< 0.05	11	2	0	7	0

Supplementary Table 15. Bacterial species that were differentially abundant between the faecal samples of Nottingham mice and mice from the two other sampling sites ('Combined'). Abundance is based on the CLR transformation of read abundance, with the difference in abundance between Nottingham mice and Combined mice (Wirral and Southport) provided. Species highlighted in bold had a greater abundance in Nottingham mice compared to Southport and Wirral mice. The *p* values used to identify significantly different species were generated using an ANCOM-BC analysis.

Phylum	Species	Abundance			*p* value
		Nottingham	Combined	Difference	
Firmicutes	***Lactobacillus reuteri***	5.25 (± 0.7)	-0.43 (± 0.4)	5.69	<0.001
Bacteroidota	***Bacteroides vulgatus***	5.63 (± 0.7)	0.86 (± 0.6)	4.76	0.004
Campilobacterota	***Helicobacter apodemus***	5.33 (± 0.7)	0.39 (± 0.6)	4.95	<0.001
	Helicobacter mastomyrinus	1.31 (± 0.1)	1.36 (± 0.5)	2.67	<0.001