The taxonomic diversity of the Ju|'hoansi hunter-gatherer intestinal microbiome in Tsumkwe, Namibia

by

Mia Truter

Submitted in partial fulfilment for the degree Magister Scientiae Bioinformatics

in the

Faculty of Natural and Agricultural Sciences,

University of Pretoria

Pretoria

October 2021

Supervisor: Dr Riaan Francois Rifkin

Co-Supervisor: Dr Jessica Edith Koopman

Declaration of Authorship

I, Mia Truter, declare that this dissertation, which I hereby submit for the degree Magister Scientiae Bioinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signed: 

Date: 27 October 2021

Nothing in life is to be feared, only understood. Now is the time to understand more, so that we may fear less

~Marie Curie

Summary

Faculty of Natural and Agricultural Sciences

Magister Scientiae Bioinformatics

By Mia Truter

The intestinal microbiome (IM) comprises all the microorganisms in the human gastro-intestinal tract, including bacteria, fungi, viruses, archaea and protozoans. Over millennia, these microorganisms have formed intimate relationships with the human host, providing the host with benefits such as optimal nutrient absorption and the priming of the immune system. The relationship between host and microbe has developed over at least 200,000 years of human evolutionary history, providing humanity with a means of faster, more flexible adaptation than their own genomes would allow. The advent of Westernisation has brought about many lifestyle changes, some of which are causing changes in IM community structure and function, which may have unforeseen and adverse consequences for human health. The IMs of traditional societies have been studied as examples of the "pre-Westernised" human IM, so that IM changes in response to Westernisation, and the consequences thereof, can be elucidated.

In **Chapter 1**, some of the factors that have been found to impact the composition and function of the human IM, are discussed, as well as how various lifestyle factors differ between traditional and Western societies, leading to the differences in IM structure observed between these two societies. The bioinformatic and statistical approaches employed in microbiome studies are summarised and critically compared.

Studies investigating changes in IM composition have either investigated the IM differences between traditional and Western populations or examined how IM composition changes in response to certain lifestyle factors, largely in a Western context. Furthermore, no IM data has been collected from traditional societies within southern Africa.

Chapter 2 aims to fill this knowledge gap, through the analysis of 40 faecal samples derived from the Ju|'hoansi San/Bushman hunter-gatherers who reside in north-eastern Namibia. This community is in the process of Westernisation, presenting an opportunity to study the evolution of the IM from a traditional to a Western lifestyle. IM composition within the Ju|'hoansi community is also analysed with respect to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant. The Ju|'hoansi IM was comparable to those of other hunter-gatherer societies, being enriched for *Prevotella, Blautia, Faecalibacterium, Succinivibrio* and *Treponema*. No significant differences were found in terms of any lifestyle

factors, except for village of primary residence. **Chapter 2** aims to serve as a baseline characterisation of the Ju|'hoansi IM from which to conduct future research pertaining to the evolution of the IM during the process of Westernisation. **Chapter 2** is written in the format of a draft article intended for publication in **Cell Press Community Review**.

Finally, Chapter 3 provides a brief summary of the work carried out in this dissertation, addresses the limitations of the study, and offers insight into future research concerning the Ju|'hoansi IM.

Acknowledgements

Completing one's MSc is an important milestone for any aspiring academic. Such an achievement cannot be accomplished without a sizeable support structure, especially during such trying times as these. First and foremost, thank you to my supervisors, Riaan Rifkin, and Jessica Koopman, for their continuous guidance and input throughout my academic journey. Thank you also for a memorable research trip!

To my family, thank you for being there for me during the good and the bad times. From listening to my fun facts about the intestinal microbiome, to being sympathetic about my having to complete yet another dissertation draft. I would not be who I am today without you.

Thank you also to Jonathan Venter, for supporting every one of my dreams, and for believing in me even during the times when I don't. Thanks to my fellow bioinformaticist and close friend, Odette le Grange, mostly for your friendship, but also for understanding scientific jargon at dinner parties.

The Bioinformatics Department deserves a massive thank you for the inclusion of all its students into a community that is welcoming and supportive. The biggest regret of my postgraduate career thus far is not being able to spend more time with my friends and colleagues at the Bioinformatics Department. Hopefully the labs will reopen soon. Lastly, I would like to thank the Benjamin R. Oppenheimer Trust for their financial provision. Finances represent a considerable hurdle for anyone trying to complete an MSc, and the financial security provided by the Oppenheimer Trust enabled me to focus entirely on the completion of my dissertation.

Table of Contents

List of Figures

Chapter 1: The intestinal microbiome in traditional and Western populations: Implications for human health

Abstract

The intestinal microbiome (IM) has been implicated in various human physiological processes. Dysbiosis of the IM is also noted in many diseases, highlighting a possible role of the IM in health and disease. Many studies have been conducted to determine how the IM changes in healthy and diseased cohorts, as well as in traditional and Westernised human populations. These studies indicate that a Western lifestyle has a profound effect on the IM, and possibly on human health. This review will discuss the literature concerning how certain factors shape the composition of the IM, how these factors contribute to observed differences in Western and traditional IMs, and the possible implications for health in traditional and Western societies, alike. The considerations regarding bioinformatics and statistics for microbiome studies will also be briefly discussed. Finally, this review will propose future research focus areas aimed at further exploration of the traditional IM, particularly in southern Africa.

Introduction

The intestinal microbiome (IM) (Lederberg and McCray, 2001) comprises all the micro-organisms living in the gastrointestinal tract (GIT). The density of microorganisms residing in the GIT increases longitudinally from mouth to rectum, and cross-sectionally from mucosa to lumen (Tropini *et al.*, 2017). The IM comprises over 200 000 prokaryotic genomes, which encode about 171 million protein sequences (Almeida *et al.*, 2020). The number of eukaryotic genomes in the human IM is less certain, partly because IM research is mainly focused on bacteria (Hernández-Santos and Klein, 2017). Innovative development of new algorithms to detect eukaryotes in microbial environments will hopefully mitigate this issue (Lind and Pollard, 2021). Given the substantial metabolic potential that so many encoded genes represent, the influence of the IM on its host becomes apparent. This includes performing a vast array of functions ranging from digestion and utilisation of nutrients (Oliphant and Allen-Vercoe, 2019), to playing a role in tissue homeostasis (Domingues and Hepworth, 2020), as well as maturation of the immune system in infants (KE *et al.*, 2016; Sanidad and Zeng, 2020).

It is intuitive that a change in taxonomic composition of the IM could impact the human host by altering the metabolic profile and functionality of the IM. This adaptability of the human IM in response to changing food availability, geographic location, and climate, is thought to have aided human survival historically (Amato *et al.*, 2019). The Neolithic revolution represents such a change in lifestyle and diet that probably led to alterations of human IM

composition and functionality. While the Neolithic revolution, and the subsequent Westernisation of societies, undoubtedly had positive implications for humanity such as the discovery of modern medicine and provision of a constant food supply, the prevalence of "diseases of civilisation" (Gupta, Paul and Dutta, 2017) is on the rise in the Western world. These diseases, including obesity (Davis, 2016), diabetes (Dunne *et al.*, 2014), and anxiety (Martin *et al.*, 2018), have been linked to dysbiosis, or an imbalance, of the IM. Therefore, it is of interest to elucidate precisely how Westernisation impacts IM composition and host health, since this is still largely unknown (Blaser and Falkow, 2009).

Studies aimed at understanding the effect of the Neolithic revolution and Westernisation on the human IM focus on comparing the IMs of traditional, non-industrialised societies to those of Western, industrialised societies, both in terms of IM composition and functionality (Schnorr *et al.*, 2014; Martínez *et al.*, 2015; Dubois *et al.*, 2017; Vangay *et al.*, 2018). These studies show that the IM adapts in response to different lifestyles (Schnorr *et al.*, 2014), which could possibly be associated with the different diseases seen among different societies.

While traditional societies in present day are not entirely representative of a pre-Neolithic IM, they represent the most accurate way of studying the effect of Westernisation on the IM (Crittenden and Schnorr, 2017). Moreover, an implication of the Westernisation of traditional societies, is that they are at a risk of Westernising into a low socioeconomic setting, thus monitoring their IMs and health in general could lead to improved healthcare in these populations as they modernise. Indeed, comparative studies between Western and traditional societies should include informing better healthcare practice in traditional societies as a study goal.

It is the aim of this review to discuss the factors that shape the bacterial and fungal composition of the IM, how these factors differ between traditional and modern societies, and highlight challenges and future perspectives in this field. This dissertation will focus specifically on the bacterial and fungal components of the microbiome.

Intestinal microbes and humans

Microbial composition of the GIT

The abundance and types of microbes present in different niches of the GIT is dependent on the physiological factors within that niche, such as morphology and pH (Tropini *et al.*, 2017). The human small intestine is characterised by a lower abundance of microbial diversity, largely owing to its very low pH and fast luminal flow. This is in contrast with the large intestine, where the microbial load is much higher due to slower peristalsis and a more neutral pH (Donaldson, Lee and Mazmanian, 2016). The lack of diversity in the small intestine is thought to prevent microbes from outcompeting the host for substrate. In other words, it is in the host's

best interest to prevent microbes from colonising the stomach and small intestine so that the host gets the first opportunity to absorb nutrients from food. More microbes can then colonise the large intestine, where the host benefits from microbial fermentation products that it would otherwise not be able to utilise (Walter and Ley, 2011).

In terms of the bacterial taxonomic composition of the IM, 4 phyla are dominant: *Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria*. Other phyla in the IM include *Chlamydiae, Cyanobacteria, Deferribacteres, Deinococcus–Thermus, Fusobacteria, Spirochaetes* and *Verrucomicrobia* (Dethlefsen, McFall-Ngai and Relman, 2007). However, when analysing the taxonomic composition of the IM at a deeper level, one finds a vast array of species and strains, and large inter-individual variability that some have likened to a genomic fingerprint (Franzosa *et al.*, 2015). This inter-individual variability seems to be a Western trait, however, and in traditional societies, the IM at species level seems to be more cohesive across the population (Martínez *et al.*, 2015; Vangay *et al.*, 2018), possibly owing to more shared communal space in traditional societies that facilitate exposure to common microbes. Another possibility is that the study design when analysing Western or traditional societies often differ. In traditional studies, cohorts often come from the same village or family whereas in Western studies, cohorts are chosen more randomly. This might artificially affect the differences in inter-individual variability seen between these two populations.

The fungal component of the IM (the mycobiome) has not been studied as extensively as the bacterial component for several reasons. Fungi have mainly been studied in culture-dependent ways and since many fungi cannot be cultured (Hallen-Adams and Suhr, 2016), insight has been limited in terms of the fungal diversity in many ecosystems. This limitation has been somewhat mitigated with the advent of novel high-throughput sequencing technologies that enable fungal classification without culturing. Furthermore, (Qin *et al.*, 2010) stated that bacteria comprise around 99.1% of the IM, with archaea comprising most of the remainder, and viruses and eukaryotes only accounting for 0.1% of the IM. This estimate has led to an incorrect general assumption that non-bacterial microorganisms play an insignificant role in human health and homeostasis (Chin *et al.*, 2020). However, the mycobiome has more recently been implicated in diseases like IBD (Sokol *et al.*, 2017) and alcoholic hepatitis (Lang *et al.*, 2020), and also possibly acts as a reservoir for disease in immune-compromised patients (Chen *et al.*, 2011; Polvi *et al.*, 2015). Moreover, the mycobiome could actually comprise more of the IM than previously thought, since the estimation from Qin *et al.*, 2010 was obtained by using pre-existing annotated reference databases, in which fungi are largely underrepresented (Underhill and Iliev, 2014). It is expected that a better understanding of the mycobiome will follow in the next few years.

While some studies have investigated the composition of the intestinal mycobiome in diseased cohorts, attention is now also shifting to include the composition of the intestinal mycobiome in healthy cohorts. Such studies report the dominant phyla to be *Basidiomycota* and *Ascomycota* (Hoffmann *et al.*, 2013; Sam, Chang and Chai, 2017), and the dominant species to be *Saccharomyces, Malassezia,* and *Candida* (Nash *et al.*, 2017; Chin *et al.*, 2020). However, in comparison to intestinal bacterial communities, intestinal fungal communities seem to be less stable and less diverse over time (Hallen-Adams *et al.*, 2015; Sam, Chang and Chai, 2017), and are subject to more inter-individual variability and variability over time (Hallen-Adams *et al.*, 2015; Nash *et al.*, 2017).

There is much speculation as to how host-microbe relationships arose over time

While the host and its microbiome are clearly intimately associated, how this relationship arose over time is a subject still under investigation. Several new terms have been coined to explain the relationship between host and microbiome. One such term is "holobiont", in which host and microbiome genomes comprise a "hologenome" (Bordenstein and Theis, 2015). Under this perspective, the host and microbiome genomes are seen as a single biological unit on which natural selection acts. The hologenome concept thus encourages consideration of interactions within the hologenome as genotype/genotype interactions, rather than genotype/environment interactions. However, this theory implies that microbiome and host have imposed selective pressures on one another over thousands of years, resulting in evolutionary changes in each lineage; a term coined coevolution. Coevolution is very difficult to prove and is probably unlikely in several scenarios (Moran and Sloan, 2015). For example, host and microbiome could be subjected to similar environmental pressures, and as a result, follow a similar evolutionary trajectory without necessarily imposing any selective pressures on one another. Or one entity, for example the host, could act selectively on the other entity, for example the microbiome, without reciprocal selection taking place. The relationship between host and various microbes within the microbiome could also be different, so considering the microbiome as a single entity instead of individual microbes is problematic.

Phylosymbiosis (Lim and Bordenstein, 2020), in which the phylogeny of the host is reflected in the phylogeny of the microbiome, is evident in many host lineages. In effect, the microbiome is more similar in hosts that are phylogenetically more related. This is thought to occur because the microbiome offers functional traits to the host, and similar hosts require similar functional traits. Phylogenetically related hosts could also harbour similar microbiomes because they provide similar ecological niches for microbial colonisation. Phylosymbiosis is not intended to imply anything about the way in which host-microbiome associations arose but is simply an observation regarding the current nature of host-microbiome interactions. As such, the term provides a more accurate framework from which to study hosts and their microbiomes.

Environmental variables like soil type and subsequent food availability, which are often shared between phylogenetically related species, could also influence the phylosymbiotic signal seen between microbiomes of different host species (Grieneisen *et al.*, 2019). More recently, certain bacteria such as *Prevotella copri* and *Eubacterium rectale* were found to exhibit parallel evolutionary histories with humans and have been preserved in the human IM for thousands of generations via vertical transmission (Suzuki *et al.*, 2021). Persistence of certain strains within the IMs of different human populations could thus contribute to population-specific health benefits.

The mycobiome's evolutionary relationship with the host has not been investigated as closely for evidence of phylosymbiosis. However, (Harrison *et al.*, 2021) showed that host-specific covariation exists between fungi and phylosymbiotic bacteria in many species, potentially hinting at a phylosymbiotic signal between host and fungi. However, some fungal species are present in the IM as a result of diet, passing through the GIT but never actually colonising it (Raimondi *et al.*, 2019). The factors that drive intestinal mycobiome composition present an interesting avenue for future research.

Regardless of how host-microbe interactions arose over time, humanity's spread across the globe likely caused changes in host and microbial environments that led to adaptation in both entities (Amato *et al.*, 2019). It therefore follows that geographical location and degree of modernisation in a society could influence the microbial structure in the IM. Indeed, different human populations in various geographic locations harbour different sets of microbes (Gupta, Paul and Dutta, 2017), which perform different functions suited to the needs of the host population. For example, *Prevotella* is an important genus that produces short chain fatty acids (SCFAs) which are beneficial to the host (Sivaprakasam, Prasad and Singh, 2016). The *Prevotella* clade, which is conserved among non-human hominid species, is less evident in traditional human populations, and almost absent in Western populations (Gaulke *et al.*, 2018). Similarly, the divergence of humans from chimpanzees led to increased consumption of animal proteins and fats which was accompanied by a drastic reduction in IM diversity and fibre-degrading enzymes like *Fibrobacter* (Moeller *et al.*, 2014), a trend which is exacerbated in the IMs of Western cultures (Moeller, 2017).

Studying the IMs of people from various geographic locations, as well as from non-human hominid species, could shed light on how the microbiota has changed throughout human history, and what the impact of those changes are for human health.

Studying traditional societies

Studying the IMs of populations across the globe is an endeavour that requires expertise in many different scientific fields. Not only is scientific competence in terms of sampling, sequencing and data analysis required, but also a thorough understanding of the anthropological factors that from the basis of the differences seen in IM composition across different populations. This necessitates the involvement of people from different scientific fields, including biology, anthropology, nutrition and evolution (Crittenden and Schnorr, 2017). While that complicates the logistics of such research, it is necessary for several reasons.

The transition from a hunter-gatherer lifestyle, through the Neolithic revolution, to a modern lifestyle: What did it mean for human health?

The Neolithic revolution, which occurred in the Fertile Crescent around 10 000 years ago and subsequently spread to other parts of the world (Zeder, 2011), had many implications on human health, diet, and social dynamics; some positive and some adverse. The Neolithic revolution increased fertility rates and population size (Bentley, Goldberg and Jasielqska, 2010), mitigated food scarcity and enabled scientific and creative advancement. It also paved the way for future human progress such as the innovations of the Industrial revolution (Weisdorf, 2005). However, increased population size and living in closer proximity also created an environment very well-suited for higher pathogenic loads and infectious diseases (Larsen *et al.*, 2019). The advent of agriculture also led to a higher incidence of dental caries (Eshed, Gopher and Hershkovitz, 2006) due to increased consumption of carbohydrates (Masood, 2020), as well as increased musculoskeletal stress associated with the physical demands of farming (Eshed *et al.*, 2004).

Prior to the Neolithic revolution and the advent of agriculture, humanity subsisted by way of hunting and gathering. The composition of the hunter-gatherer diet depended largely on where the specific population lived, and the climate of that area. For example, palaeolithic populations in the arctic would have subsisted on a diet far higher in animal fat (around 80-90%) than those from Africa or Australia (less than 50%) (Isaac, 2010). As such, there is no single diet that all pre-agricultural populations adhered to; the success of humanity can largely be attributed to the ability to adapt to and thrive in different environments and climates (Brand-Miller, Mann and Cordain, 2018).

The hunter-gatherer period of human evolution, and all its associated dietary and behavioural norms, led to the bigger brain size and increased intelligence that is typical of the human species today (Milton, 1999, 2000; Robson and Kaplan, 2003). Then, upon the advent of the Neolithic revolution, foods were introduced into the human diet that had never been consumed before. These included non-human milk and dairy products, refined cereals,

refined sugars, refined vegetable oils and fatty meats (Cordain, 2006). While humanity had ample time to evolve to suit the conditions of the pre-Neolithic, it is highly unlikely that any genetic changes occurred during the Neolithic revolution that allowed humanity to thrive in this new environment (Carrera-Bastos *et al.*, 2011). It is thus thought that the Neolithic revolution, and the subsequent Westernisation of humanity, has led to disparity between human environment and human genetics, which is likely contributing to the high incidence of certain diseases seen in Western societies such as obesity and diabetes (Carrera-Bastos *et al.*, 2011). While a palaeolithic diet and total conversion to our ancestral hunter-gatherer way of life is not feasible and in many ways undesirable, consumption of more fresh produce and fibre in the Western diet might help mitigate some of the diseases associated with it.

The IM is of interest in this context, since evidence suggest that the microbes therein have a long history with humanity. How did the IM contribute to human fitness and survival? How did the IM change in response to Westernisation? How is dysbiosis of the IM linked to diseases of civilisation? Are there therapeutic avenues to mitigate these diseases by altering the IM? These questions can be answered by studying the IMs of traditional societies as a proxy for the IMs of pre-Neolithic civilisations.

Although traditional societies are not completely untouched by Westernisation (Crittenden and Schnorr, 2017), they are the closest extant examples of pre-Neolithic humans and therefore offer a feasible way of attempting to understand how geographical location, diet and lifestyle factors influence IM composition and human health.

Different lifestyles lead to different IM composition and functionality

Thus far, many studies have been conducted investigating the IM differences between Western and traditional populations. These studies have only elucidated the composition and functionality of the bacterial component of the IM, and studies including other components of the IM would be of interest. Nonetheless, several differences are highlighted between the bacterial component of the IMs of traditional and Western populations.

The diversity of the IM in Western populations is drastically less than the diversity observed in traditional populations, which is highlighted in nearly all of the studies comparing traditional and Western IMs (Gupta, Paul and Dutta, 2017). For example, in a study comparing the IMs of urban and rural Nigerian Bassa populations (Ayeni *et al.*, 2018), it was found that the abundance of fibre-degrading bacteria was progressively lost with urbanisation. Another example is the study by (De Filippo *et al.*, 2017), in which children from Burkina Faso exhibited greater intestinal bacterial diversity than did children from Italy.

The lack of bacterial diversity in Western IMs is thought to be linked to diversity in diet, since the consumption of a more varied diet results in a greater variety of substrates, which in turn

creates more environmental niches in the GIT that can support increased microbial diversity (Walter and Ley, 2011). A decrease in bacterial diversity of the human IM is associated with the onset of non-communicable diseases, probably due to disruption of immunoregulatory networks, as well as increased inflammatory responses (West *et al.*, 2015).

Diversity in the mycobiome has so far only been studied in Western populations. Fungal diversity of the IM was found to be increased in patients with Chron's disease but decreased in patients with diabetes (Mar Rodríguez *et al.*, 2015). While these studies hint at a possible fungal role in diseases, the mechanisms through which this occurs is largely understudied, as are the effects of geographic location and lifestyle habits on fungal diversity.

Moreover, traditional societies allow the study of adaptability of the IM in response to environmental changes. A remarkable example of this is the evolution of xenobiotic degradation in the human IM. Evidence from an ancient coprolite shows that xenobiotic degradation is a feature absent in ancient human IMs and that this trait probably developed in response to modern pollutants (Rifkin *et al.*, 2020). The same can be seen when comparing extant traditional and Western populations; modern Bantu individuals have a higher abundance of xenobiotic degrading pathways than do BaAka hunter-gatherers, enabling them to better degrade food additives common in industrialised societies (Gomez *et al.*, 2016a). The adaptability of the IM in response to new environmental pressures may have enabled human survival across the ages (Amato *et al.*, 2019).

How can traditional societies benefit from these studies?

There are many benefits for Western societies in studying traditional ones, including elucidating the effect of Westernisation on human health, and possible therapeutic intervention for diseases in which the IM plays a role. However, what are the benefits of these studies for traditional populations? Traditional populations represent the minority of today's society and are therefore more vulnerable to exploitation and marginalisation than people in the Western world. ~~A study by~~ (Anderson *et al.*, 2016) showed that indigenous people have poorer health and social outcomes than do non-indigenous people. The reasons for this are complex and numerous and include social determinants such as the use of health care facilities and the "socio-economic policies that shape the conditions of daily living". Indigenous societies are also often excluded from the decision-making process with regards to these socio-economic policies. The inequities between populations of different socio-economic status is highlighted in a study by (Carson *et al.*, 2019)), in which BaAka and Bantu individuals from Cameroon were interviewed regarding their perception of health challenges in the country. The BaAka live a largely traditional way of life and are typically excluded from participation in economic activity. The result is that in comparison to the Bantu, the BaAka are unable to afford health care and education, are subjected to poorer living circumstances, have less frequent access to

clean water, and have a generally lower socio-economic status. This complicated geo-political and socio-economic situation is not unique to the BaAka, but is seen in many indigenous societies across the world (Anderson *et al.*, 2016; Jackson, 2021). Furthermore, a lower socioeconomic status results in more frequent exposure to pollutants (Clark, Millet and Marshall, 2017; Jiao, Xu and Liu, 2018), consumption of a lower quality diet (Zagorsky and Smith, 2017), and increased psychosocial stress (Chen and Miller, 2013). Interplay between the abovementioned factors, IM composition and health have been well documented, and probably result in health inequities seen between individuals of different socioeconomic groups (Amato *et al.*, 2021). Care should therefore be taken that the study of IMs in traditional, low socioeconomic populations also benefit the subjects themselves, and that research of this nature is not contributing to the exploitation of already subjugated societies.

Factors that contribute to IM variation between Western and traditional populations

The factors that shape the composition of the IM are highly complex. It is therefore difficult to assess the contribution of an individual factor in determining the composition of the IM. Despite these difficulties, several studies have been conducted to understand what factors contribute to IM composition, and what role these play in human health and disease. Many factors have been identified thus far, including host genotype (Kurilshikov *et al.*, 2021), diet (Kovatcheva-Datchary *et al.*, 2015), geographic location (Kabwe *et al.*, 2020), socioeconomic status (Miller *et al.*, 2016), use of antibiotics (Pérez-Cobas *et al.*, 2013), age (Yatsunenko *et al.*, 2012), and biological sex (Haro *et al.*, 2016).

Host Diet

Host diet is arguably the most influential determinant in terms of the composition of the bacterial component of the IM and confounds other IM determining factors such as host genetics, environment, geographic location and culture (Wilson *et al.*, 2020). The relationship between host diet, the IM and human fitness has probably changed pre- and post-Neolithic revolution. For example, an IM that enables efficient nutrient extraction from diet would have been beneficial to the pre-Neolithic human host and could possibly have provided a fitness advantage. However, because food is far more readily available in modern settings, the ability to extract extra nutrients from food might be disadvantageous (Walter and Ley, 2011). This IM trait, in combination with increased sedentarism in the modern era, is likely contributing to the obesity epidemic facing modern society (Turnbaugh *et al.*, 2006; John and Mullin, 2016).

Among the most important dietary transitions that took place from our hunter-gatherer past, through the more recent agro-pastoralist era, to the Westernised world today, is a gradual

decrease in fibre intake (Eisenstein, 2020). Western societies have exchanged the traditional high carbohydrate, high fibre diets, for high fat and animal protein diets (Thorburn, Macia and Mackay, 2014); a trait that is reflected in the composition and functionality of the Western IM.

For example, the Hadza hunter-gatherers from Tanzania consume about 75-100g of fibre per day (de Vrieze, 2014), which is very little in comparison to the Western recommendation of around 30g of fibre per day (Lupton, 2002). Upon comparison to Italians, the Hadza harbour higher abundances of key fibre fermenters such as *Prevotella, Treponema, Clostridiales* and *Bacteroidota*, and an absence of common Western IM members such as *Bifidobacterium* (Schnorr *et al.*, 2014). Furthermore, the Hadza IM is functionally enriched for genes involved in complex carbohydrate metabolism and bioconversion of complex plant polysaccharides from plant dietary sources. In contrast, the Italian cohort is functionally enriched for the digestion of simple sugars such as those derived from pasta and bread (Rampelli *et al.*, 2015). The by-products of complex plant polysaccharide fermentation, propionate, butyrate, and acetate, have anti-inflammatory and anti-carcinogenic properties (Sivaprakasam, Prasad and Singh, 2016) and are thought to play a protective role in the development of colorectal cancer (Shuwen *et al.*, 2019), obesity, and asthma (Wood, 2017).

The lack of SCFA production in the Western IM, as a result of a decreased intake of dietary fibre, is therefore thought to be a major contributor to the high incidence of these diseases of civilisation (Carrera-Bastos *et al.*, 2011; Sonnenburg and Sonnenburg, 2014). Indeed, in a diet-switch study between rural Africans and African Americans (O'Keefe *et al.*, 2015), in which rural Africans consumed a high fat and protein diet and African Americans consumed a high fibre and carbohydrate diet, reciprocal changes in the IM were observed. After consuming a high fat, low fibre diet, IM changes in the rural African cohort included lower SCFA production, higher colonic secondary bile acids and higher mucosal proliferative biomarkers which are associated with an increased risk for colorectal cancer. The converse occurred in the African American cohort – IM biomarkers for colorectal cancers were reduced, suggesting that diet has a great impact on cancer risk. Similarly, the period from 1949-1992 in China was marked by a transition from a high fibre, high carbohydrate diet to a higher intake of meat and fat, which has been implicated in the rise of incidence of non-communicable diseases in China during this time (Du *et al.*, 2014). Moreover, Thai immigrants who immigrated to the United States experience IM changes that increase their risk of non-communicable diseases in comparison to Thai citizens who stay in their country of origin (Vangay *et al.*, 2018). Evidence thus suggests that transition from a high fibre, high carbohydrate diet to a Western diet high in saturated fats and animal protein exacerbates the risk of non-communicable diseases (Wilson *et al.*, 2020).

The mycobiome is understudied in comparison the bacterial IM. Nonetheless, it is thought that many fungi are present in the GIT as foodborne microbes (Kong and Morris, 2017). (Hoffmann *et al.*, 2013), found that *Candida* abundance was positively correlated to carbohydrate consumption and negatively correlated to a diet high in amino acids, protein and fatty acids. Interestingly, *Candida* abundance was also positively associated with the abundance of *Methanobrevibacter*. While the study draws no conclusions about the functional significance of these associations, it does suggest that such a significance might exist. Moreover, the fungal landscape in the GIT differs between obese and non-obese individuals, with the abundance of *Mucor racemosus* and *M. fuscus* being significantly higher in non-obese individuals. Curiously, the decreased abundance of these fungal species in obese subjects was reversible upon weight loss, suggesting that *Mucor* might play a role in obesity (Mar Rodríguez *et al.*, 2015).

The effects of a dietary transition from traditional to modern, and the ensuing changes in IM composition and function, can be investigated in the form of longitudinal studies in the IMs of traditional cohorts. Studies of this nature would elucidate the effects of an introduction of Western foods into a largely traditional diet, *in situ.* However, it is important that the health of the traditional community in question is included as a study priority since transition to a Western lifestyle increases the risk of Western diseases.

Host genotype

Evidence suggests that microbiome composition is influenced by host genotype. For example, in most human populations, the ability to digest lactose is quickly lost after weaning due to reduced expression of the enzyme lactase-phlorizin hydrolase (LPH) (Walter and Ley, 2011). However, in societies where milk consumption is practiced into adulthood, LPH persists due to variation in the genes controlling for its expression (Tishkoff *et al.*, 2007). In several studies investigating the corelations between host genotype and bacterial abundance, *Bifidobacterium* abundance was found to be positively correlated to lactase-persistence, presumably because the ability to digest lactose leads to a higher consumption of dairy products, which in turn promotes the growth of lactose-digesting bacteria (Blekhman *et al.*, 2015; Kurilshikov *et al.*, 2021).

Another interesting example of interplay between host genetic selection and microbiome composition is brought to light by (Walter and Ley, 2011)). This article posits that competition between microbiome and host was brought about by the advent of agriculture, during which time the consumption of easily digestible starch became more frequent. Since starch-digesting microbes residing in the small intestine would divert some of the energy derived from starch for their own growth, it is more energetically viable for humans to digest the starch directly. This is suggested to have led to two developments. Firstly, an expulsion of microbes from the

small intestine, and secondly, selection for oral digestion of starch. Both would ensure maximum host absorption of starch-derived nutrients. Indeed, AMY1, the gene encoding for salivary amylase, is more prevalent among historically agricultural societies than those from circum-arctic hunter-gatherer backgrounds (Perry *et al.*, 2007). Interestingly, in modern times type 2 diabetes mellitus is treated by administration of acarbose, an inhibitor of salivary amylase (Zhao *et al.*, 2018). Acarbose thus allows starch to pass to the colon undigested, where it is fermented by bacteria into SCFAs; SCFAs alleviate the symptoms of T2DM. This highlights how the relationship between microbiome and host changes over time, and what was potentially beneficial to humanity in ancient times, could be detrimental in a modern era (Walter and Ley, 2011).

In a recent human genome-wide association study (GWAS), 31 loci were identified as having an effect on the IM (Kurilshikov *et al.*, 2021). One such locus is FUT2-FUT21. FUT2 is responsible for the production of alpha-1,2-fucosyltransferase, which secretes fucosylated mucus glycans in the gastrointestinal mucosa. It was observed in this study that individuals homozygous for the secretor allele had an increased abundance of *Ruminococcus.* It was also found that the secretor allele was negatively associated with risk for cholelithiasis and Crohn's disease, and positively associated with fish intake, highlighting the complex interplay between dietary preferences, host genetics, IM composition and disease susceptibility. Other genes found to be correlated with the IM in this study were involved in innate and adaptive immunity, as well as genes expressed in the brain, which adds to the supposition of the existence of the gut-brain axis (Kushak and Winter, 2018; Martin *et al.*, 2018).

Comparatively, far less studies have been conducted on the interaction between host genotype and intestinal fungi. However, one interesting study showed an association between different variants of Dectin-1 and ulcerative colitis susceptibility (Iliev *et al.*, 2012). Dectin-1 and Dectin-2 are part of a family of C-type lectin receptors (CTLRs) which are known for their ability to recognise fungal β-glycans, and play a role in immunity against fungi, including *Candida*, *Aspergillus* and *Pneumocystis.* Dectin-1 also plays a role in intestinal homeostasis, autoimmunity and allergy (Dambuza and Brown, 2015). These studies hint at a fungal role in human health and disease susceptibility,

While many comparative studies between traditional and Westernised populations control for genetic variation by ensuring the cohort is genetically similar (Ayeni *et al.*, 2018; Keohane *et al.*, 2020), very few studies have actually investigated the link between host genotype and IM composition in a traditional setting. This warrants further investigation for many reasons. Some of these populations, like the Ju|'hoansi hunter-gatherers in Namibia, have not been exposed to much admixture throughout the years, due to their geographic isolation (Owers *et al.*, 2017). Populations with such genetic makeups could provide novel insights into the link between

genotype and IM composition and should be included in such studies for more informative power.

Antibiotics

The use of antibiotics is one of the biggest aspects of modern society that alters the microbiome. While antibiotics have undoubtedly changed quality of life for people across the world, there are also negative aspects of this medication that should be noted, particularly in terms of the effect on the IM. Most antibiotics do not only kill the pathogenic bacteria, but also commensals that reside in the GIT and offer benefits to the host. The loss of commensals results in a loss of colonisation resistance. Colonisation resistance is mediated by commensal microbes in two ways. Directly, through outcompeting pathogenic microbes for space and resources, and indirectly, through training and modulating the immune system (Casals-Pascual, Vergara and Vila, 2018). An example of the consequences of a loss of colonisation resistance is the increased prevalence of antibiotic-associated infections with *Clostridium difficile* in hospitals (Rupnik, Wilcox and Gerding, 2009; Johanesen *et al.*, 2015). Moreover, metronidazole has been shown to decrease expression of Muc2, which is the major component of the mucous layer (Wlodarska *et al.*, 2011). A thinned intestinal mucous layer could lead to inflammation and a higher risk of invasion by pathogenic microbes (Francino, 2016). There is also an increased prevalence of Crohn's disease in adults who frequently used antibiotics in childhood, presumably as a result of long term changes in IM structure (Hildebrand *et al.*, 2008).

Interestingly, the fungal composition of the IM also plays a role in the effectiveness of *C. difficile* treatment (Zuo *et al.*, 2018). Fungi benefit from the use of antibiotics due to the increased availability of space and nutrients, which could result in opportunistic fungal infections (Cottier and Pavelka, 2012). However, it has also been posited that beneficial fungi take over the functional role of commensal bacteria post-antibiotic use (Jiang *et al.*, 2017).

Furthermore, the IM is regarded as the epicentre for antimicrobial resistance (Carlet, 2012). This is because the Western GIT is more frequently exposed to antibiotics in medicine and via consumption of foods that are routinely treated with antibiotics (Francino, 2016), which then encourages more rapid development of antibiotic resistance mechanisms that can be spread to other microbes via horizontal gene transfer (McInnes *et al.*, 2020). The acquisition of resistance mechanisms in the GIT could then influence the efficacy of antibiotic treatment, and possibly contribute to the antibiotic resistance crisis (Schaik, 2015).

Of particular interest is the effect of tuberculosis (TB) on the IM. TB treatment requires six months of intensive antibiotic treatment, and even longer if the patient has drug resistant TB. This, in combination with the high incidence of TB globally (around 10 million infections in 2019) (Geneva: World Health Organization, 2020), and especially within indigenous communities (Tollefson *et al.*, 2013), probably significantly affects IM composition. Indeed,

there is some evidence for an association between TB susceptibility, TB recurrence, and IM composition (Eribo *et al.*, 2019). However, a distinction must be made between the effects of TB itself on the IM, and effects of anti-TB treatment. In a recent study, TB itself was shown not to significantly affect α-diversity or the abundance of individual taxa. However, upon anti-TB treatment, a significant difference in terms of α-diversity and individual taxon abundance was observed. Patients undergoing anti-TB treatment showed a decrease in abundance of *Clostridiales* and an enrichment of *Bacteroides* (Hu *et al.*, 2019). The effect of TB on the indigenous IM has not yet been elucidated but would perhaps make for an interesting study. Furthermore, the role of fungi in anti-TB pharmacokinetics, TB susceptibility and recurrent TB infection is yet to be studied.

Age

It is now well-established that the IM is colonised at birth, at which point the IM is subject to high levels of inter-individual variability. Throughout the first year of life, the IMs of neonates increase in diversity and eventually converge to resemble an adult IM. This increase in diversity can continue up to around 3 years of age, when most IMs stabilise (Yatsunenko *et al.*, 2012). More recently, studies have been conducted investigating the influence of certain factors, such as feeding regime and mode of delivery, on neonatal IM colonisation (Bager, Wohlfahrt and Westergaard, 2008; Cong *et al.*, 2016). Such studies are generally in agreement that natural birth and breast feeding result in increased microbial diversity (Wong *et al.*, 2021) and support the growth of beneficial taxa such as *Bifidobacterium* and *Lactobacillus* (Kumbhare *et al.*, 2019). The health of the mother has also been implicated in neonate health. Breast milk from obese mothers harbour a less diverse microbial community (Cabrera-Rubio *et al.*, 2012), which is posited to affect neonate IM composition, although the latter was not explicitly tested. Prepartum maternal diets are also posited to induce epigenetic changes in the neonate, which could then lead to alterations in the neonate IM that could predispose to the onset of late-life obesity (Li, 2018). Although there is reason to believe that the health of the mother could influence neonatal IM colonisation and neonate health, many of these suppositions need to be explicitly tested before any conclusions can be reached. Studies pertaining to the influence of a traditional lifestyle on neonate health are scarce, although one such study finds that the rural Nigerian infant and adult microbiome are more similar than the urban Nigerian infant and adult microbiome. This is thought to be due to the introduction of adult foods earlier in life, as well as the earlier exposure of infants to the environment (Ayeni *et al.*, 2018). The study of the neonatal IM is promising in terms of therapeutic intervention, but many questions are as yet unanswered. For example, how is the neonatal mycobiome colonised? Moreover, how is the covid-19 pandemic affecting the IM health of neonates born during this time-period, considering the increased sanitation and the lack of exposure to environmental microbes?

Some studies have noted the effect of old age on IM composition, for example that the elderly IM harbours decreased levels of Bacteroides and *Faecalibacterium* (Nuria *et al.*, 2013). It is thought that bodily changes brought about by old age, including intestinal senescence and an impaired ability to absorb nutrients, could, along with changes in the IM, contribute to the pro-inflammatory state often seen in the elderly (Nagpal *et al.*, 2018). These findings suggest that the IM could contribute to inflammatory disease pathophysiology in the elderly, as well as providing potential for dietary intervention to counteract the effects of ageing (Salazar *et al.*, 2014). It is also suggested that an intake of pre- and pro-biotics could significantly improve metabolism in the IM and increase SCFA production (Roberfroid *et al.*, 2010). However, few studies have been conducted on the direct links between ageing, inflammation, and IM composition (Buford, 2017). As a result, there is no clear consensus on the exact age at which IM composition starts to change in response to ageing, or enough data on the effects of ageing on the IM.

Biological sex

Biological sex is also suspected to influence the composition of the IM and in doing so, partially contribute to the sex-bias seen in certain diseases. For example, IBD is more frequent in women than in men (Klem *et al.*, 2017), and the IM is also implicated in the pathophysiology of this disease (Vich Vila *et al.*, 2018). Auto-immune diseases, such as rheumatoid arthritis, are also more common in women (Intriago *et al.*, 2019). Rheumatoid arthritis is a multifactorial disease that is caused when immune cells incorrectly attack healthy tissue (Hassanzadeh and Gholamnezhad, 2020). Since the IM is known to prime the immune system, teaching it to discern between harmful and healthy cells (Sommer and Bäckhed, 2013), it is suspected that incorrect priming of the immune system as a result of dysbiosis in the IM could be implicated in the onset of autoimmune diseases such as rheumatoid arthritis. This, in combination with sex hormones, could potentially cause the sex-bias seen in such auto-immune diseases (Gomez, Luckey and Taneja, 2015). Other examples of sex-biased diseases in which the IM may be involved are cardiovascular disorders. The risk factors for cardiovascular disorders, glucose regulation, dyslipidaemia, hypertension, and obesity, exhibit sexual dimorphism due to varied excretion of sex hormones between the sexes. Since sexual dimorphisms also exist in the IM, it is posited that in combination with sex hormones, the IM could contribute to the pathophysiology of cardiovascular diseases (Razavi *et al.*, 2019). However, whether the IM, in context of sex, plays a clinically significant role in cardiovascular disease onset is still unclear (Cross, Kasahara and Rey, 2018).

In context of traditional vs Western societies, the role of sex in IM composition is largely uncertain. Gender roles are more conservative in traditional societies in comparison to urban ones (Evans, 2019), such that tasks like work and child-tending are more often shared between

men and women in urban settings. This could result in a more homogeneous IM composition between the sexes in urban areas in comparison to traditional areas, where men frequently hunt while women gather, tend to children, and cook. However, since men and women live in close association with one another in traditional settings, the effect of gender tasks on IM composition may be mitigated.

Travel

Some studies have documented the effect of IM composition on the susceptibility of diseases during travel (O' Donovan *et al.*, 2020; Youmans *et al*., 2015), and it is thought that certain members of the IM, particularly SCFA producing bacteria, could offer protective benefits while traveling (Riddle and Connor, 2016). Traveling between continents can also influence the abundance of antibiotic resistance genes in the IM (Bengtsson-Palme *et al*., 2015). Thus, traveling possibly has an effect on IM composition, but conclusive evidence from studies employing robust sample sizes is lacking. A possible reason for the IM fluctuations seen after traveling could be increased exposure to different foods, people, and microbes. Traditional societies, such as the Ju|'hoansi from southern Africa, are relatively isolated and do not travel frequently (Owers *et al.*, 2017), thus come into contact with foreign microbes less often. The effects of travel on traditional IMs are unknown at present.

While many studies have been conducted to determine the impact of certain lifestyle factors on IM composition and functionality, the way in which these studies are performed is noteworthy. Since the field of metagenomic analysis is still in its infancy in comparison to other scientific fields, there is a need for protocol standardisation in order to ensure reproducible and reliable results.

Bioinformatic considerations surrounding microbiome studies

The general workflow of a microbiome analysis study should start with proper planning. Experimental design must encompass determination of the number of samples, inclusion of proper controls, establishment of what metadata is appropriate to collect, as well as a clear, testable hypothesis (Bharti and Grimm, 2021). The choice of sequencing technique is also largely dependent on the needs of the project and on resource availability (Galloway-Peña and Hanson, 2020). The resulting samples should then undergo DNA extraction, sequencing, and quality control. The data can then be used for statistical analysis to prove or disprove the hypothesis, often in correlation with the collected metadata. Finally, functionality can be inferred to gain insight into the metabolic potential of the microbiome.

Sampling, storage, and DNA extraction protocols

The most common way of investigating IM composition is via the collection of faecal samples, from which resident microbiota are extracted and sequenced. Using faeces as a

proxy for IM composition is non-invasive and cost effective (Tang *et al.*, 2020). However, studies show that the colonic microbial community differs from the faecal microbial community, suggesting that perhaps faecal samples are not the best method of analysing the IM (Zoetendal *et al.*, 2002; Rangel *et al.*, 2015). Moreover, the microbial composition between various parts of the GIT also differs (Tropini *et al.*, 2017), so referring to the colonic microbes as the collective "IM" is also somewhat misleading. Comparatively, very little is known about how certain lifestyle factors affect other parts of the GIT such as the stomach.

The way in which faecal samples are stored are known to affect the DNA yield and microbial profile of samples, thereby influencing the outcome of microbiome studies (Ezzy *et al.*, 2019). While immediate storage of faecal samples in -20 °C is the golden standard, this is often impractical in the field. Because research of this nature is often conducted in rural areas, samples may be subjected to temperature fluxes and several weeks of transportation. To investigate the effect of sample storage conditions on IM community composition, (Song *et al.*, 2016) elucidated the changes in community composition of human and dog faecal samples using a variety of storage methods, subjected to a variety of temperature changes. They found that immediate freezing of a sample, in combination with a preservative, resulted in the least amount of community changes. Several sample preservation methods were deemed sufficient when samples were frozen, including 95% ethanol, FTA cards, OMNIgene Gut, and RNAlater. However, RNAlater resulted in greater change in taxon abundance at ambient temperatures than did the other methods, suggesting that RNAlater may not be the ideal solution if sample freezing cannot be guaranteed. Moreover, 70% ethanol was cautioned against, owing to its lack of preservation power.

Furthermore, fungal DNA extraction from faeces is complicated by the fact that DNA extraction protocols have largely been developed around bacteria, as well as the comparatively low yield of fungus in faecal samples (Angebault *et al.*, 2018). Fungal cell walls are complex structures that cell impede lysis and nucleic acid recovery (Fredricks, Smith and Meier, 2005). Different methods of fungal DNA extraction also work optimally on different species, so finding a DNA extraction method that works well for all fungal species in a sample is challenging (Fiedorová *et al.*, 2019). Sample storage and DNA extraction of both fungal and bacterial microbes is an ongoing field of research.

Sequencing

Next generation sequencing (NGS) has greatly sped up the sequencing process by allowing millions of small fragments of DNA to be sequenced in parallel (Behjati and Tarpey, 2013). NGS, in combination with techniques like PCR, have circumvented the need for culturing prior to microbial sequencing, thereby enabling deeper exploration of novel microbial environments (Malla *et al.*, 2019). It must be noted, however, that NGS does not completely mitigate the need

for culturing. While NGS provides the genomic sequence, linking a phenotype to specific microbial strains and investigating interactions between microbes is best done via microbial cultivation (Sommer, 2015).

There are two common ways of sequencing microbial genes. One could either perform whole genome shotgun sequencing (WGS), which targets the genomes of all microbial species in an environment. WGS has several advantages, including enabling better taxonomic resolution, increased diversity detection, and better functional prediction. However, WGS is very costly and is not feasible for researchers with limited budgets (Ranjan *et al.*, 2016). Alternatively, sequencing certain household genes, such as the commonly used 16S rRNA gene or the Internal Transcribed Spacer (ITS) region for bacteria and fungi respectively, yields satisfactory results. Such marker genes contain both conserved regions and hypervariable regions, rendering them useful for taxonomic classification, since the conserved regions are ubiquitous enough to be used for primer design, while the hypervariable regions differ enough for adequate distinguishment between microbial communities (Tringe and Hugenholtz, 2008; Schoch *et al.*, 2012). Sequencing these marker genes, otherwise known as amplicon gene sequencing, is more cost effective than WGS and is adequate for taxonomic classification (Mizrahi-Man, Davenport and Gilad, 2013). Moreover, amplicon gene sequencing requires less extensive data analysis and therefore is not as bioinformatics intensive (Ranjan *et al.*, 2016).

However, certain limitations exist for amplicon gene sequencing. In some cases, the 16S gene is too homologous to identify taxa up to a species level, resulting in low taxonomic resolution. Over-estimation of diversity can also occur using 16S sequencing due to presence of multiple copies of this gene in one bacterial cell (Bailén *et al.*, 2020). Functional prediction is also compromised due to absence of the entire genome (Galloway-Peña and Hanson, 2020), and genome reconstruction is impossible.

The use of controls

While improved sensitivity has been brought about by the use of new sequencing techniques, this has also led to increased amplification of contaminant DNA that could skew research results (Eisenhofer *et al.*, 2019). The lack of standardised controls in microbiome studies could lead to overzealous conclusions that misinform knowledge about certain ecosystems. This is well-illustrated in a study by (Salter *et al.*, 2014), in which the theory that the infant nasopharyngeal microbiome shows age-related clustering, is disproved. Once contamination from the DNA extraction kit was removed, there was no age-related clustering between the nasopharyngeal samples. Contamination can arise from many sources throughout the process, including via the use of unsterile sampling tools (Weiss *et al.*, 2014), or as a result of cross-contamination between samples (Cando-Dumancela *et al.*, 2021).

The use of controls is thus a necessity in microbiome studies. Positive controls can comprise defined mock communities, although because they are comparatively less complex than actual microbial environments, their usefulness in microbiome studies is still uncertain (Hornung, Zwittink and Kuijper, 2019). Negative controls should include a sampling container without any sample, to detect contamination from the container. DNA extraction kits have been implicated as causative of contamination (Salter *et al.*, 2014) and should thus also be included as a negative control.

Although the importance of the use of controls is widely accepted, there is no common practice in terms of how contamination is dealt with in microbiome studies, and a result, handling contamination is widely ignored. Because contamination can come from actual biological sources through cross-contamination of samples, removing all reads found in controls from samples could result in the loss of biological signal. It is therefore recommended that a less conservative approach be taken to only remove contamination that is statistically probable (Eisenhofer *et al.*, 2019; Minich *et al.*, 2019). To this end, an R package has been developed (Davis *et al.*, 2017) called decontam. Decontam works by either assessing the frequencies of reads as a function of DNA concentration, where contaminant reads are classified by their inversely proportional relationship with DNA concentration, or the prevalence of reads between controls and samples, in which the probability of a read being a contaminant is presented as a p-value. To improve the reproducibility, and validity of findings in the microbiome field, it is imperative to employ controls and remove contamination.

Quality control

Quality control is another crucial step in the process and can account for the largest variability in a study (Sinha *et al.*, 2015). Quality control is conducted on the raw DNA sequences and includes steps like primer removal, trimming and filtering low quality reads, as well as the removal of chimeras. During quality control, reads can either clustered into operational taxonomic units (OTUs) at a user-define sequence similarity (often 97% for species-level resolution), or kept as amplicon sequence variants (ASVs), which are single base resolution OTUs. The latter is more accurate (Callahan, McMurdie and Holmes, 2017).

Many programs exist that perform quality control and user choice dictates which one is used. All available software programs have their advantages and disadvantages. For example, DADA2 (Callahan *et al.*, 2016) is one of the most sensitive software programs available for quality control, but is less specific than other programs such as USEARCH-UNOISE3 (Edgar, 2010, 2016) and Qiime2-Deblur (Amir *et al.*, 2017; Bolyen *et al.*, 2019; Prodan *et al.*, 2020).

Normalisation

Variations in sample collection, library preparation, and sequencing can result in biases that are not a true representation of actual biological variation in the microbial environment (Weiss *et al.*, 2017). As such, microbiome data are unequal in library size across samples, are sparse (meaning the data contains many zeros), and are compositional (Gloor *et al.*, 2017). Microbiome data therefore contravene the assumptions of many standard biological tests (Xia, 2020), such as the assumptions of normality and homoscedasticity (Odintsova, Tyakht and Alexeev, 2017). Normalisation, the process of transforming the data to enable statistical comparison across groups, is thus a necessity for any microbiome study (McKnight *et al.*, 2019).

Rarefying is the most common normalisation technique, and comprises user selection of a minimum library size, discarding samples with a read abundance less than the chosen minimum, and random sub-sampling of the remaining samples without replacement until they all have an equal library size (Hughes and Hellmann, 2005). However, this discards valid biological data, resulting in decreased sensitivity, overdispersion, and the addition of artificial uncertainty. These consequences are particularly relevant to small or uneven datasets, the prevalence of which is quite high in microbiome studies, owing to the cost of sequencing (McMurdie and Holmes, 2014).

Other methods of data transformation include mathematically simple transformations like changing absolute abundances to proportions, or using the total read count as a factor to estimate library size (Total Sum Scaling) (Xia, 2020). Mathematically more complex normalisation methods have been developed from Aitchison's log ratio (Aitchison, 1982), such as the centred-log ratio (CLR). By log-transforming data, a variance-stabilising transformation is applied, thereby meeting the assumptions for downstream statistical analysis (Odintsova, Tyakht and Alexeev, 2017). The use of log-transformation presents a problem in that microbiome datasets contain many zeros and the log of zero is undefined. To circumvent this problem, a small pseudo-count is added to the data before log-transformation. However, the use optimal value of a pseudo-count has not been established (Weiss *et al.*, 2017), and has been shown to affect statistical outcome (Kaul *et al.*, 2017). More sophisticated normalisation techniques have also been adopted from RNA-seq analyses, and packages like DESeq2 (Love, Huber and Anders, 2014) and EdgeR (Robinson, McCarthy and Smyth, 2010) are widely used in microbiome studies. More recently, normalisation methods have been developed specifically for microbiome studies (Xia, 2020), including Cumulative Sum Scaling (Paulson *et al.*, 2013), which is designed to account for under-sampling, and Geometric Mean of Pairwise Ratios (Chen *et al.*, 2018), which deals with zero-inflation (an excess of zeros). Once the data are normalised, various statistical methods can be used to answer the research question.

Statistical tests

Ultimately, the goal of a microbiome research project is to test the hypothesis in question by statistically analysing the correlations between collected metadata and the microbial community structure. This could comprise several tests to elucidate the overall community composition between different groups. For example, to test whether IM α-diversity (the diversity within groups) is the same between neonates of different sexes (Cong *et al.*, 2016). Alternatively, one could test whether the β-diversity (the diversity between groups) is different between various human microbiomes (Huttenhower *et al.*, 2012).

Researchers can also test whether individual taxa abundances are statistically significantly different between certain groups. This is particularly useful in the identification of microbes that could play a functional role. For example, *Akkermansia muciniphila* has been identified as a potential mediator of obesity and type 2 diabetes (Dao *et al.*, 2016). Moreover, correlations between members of the IM from different kingdoms, from example bacteria and fungi, can be elucidated in context of the collected metadata.

Functional prediction

While taxonomic assignments attempt to determine "who is there", functional prediction attempts to find out "what they are doing". The efficacy of functional prediction is largely determined by the type of sequencing that was conducted. WGS results in far more accurate functional predictions, since it enables identification of reads within the entire metagenome, which can then be compared to databases of known genes, proteins, and protein families to infer functionality of the metagenome (Sharpton, 2014). In contrast, the use of amplicon data limits the study in that only a specific gene can be used for functional prediction, hence amplicon functional prediction relies on taxonomic classification to infer functionality rather than comparison with known genes and proteins.

All of the abovementioned steps in the microbiome workflow can influence the results of the study at hand. Caution is thus advised when drawing conclusions about the nature of the microbiome, especially since protocols in this field are not yet standardised.

A word of caution

While the IM is implicated in the onset of various diseases, care should be taken not to perceive and promote the IM as a single solution to all Western health-related problems. Correlation does not imply causation, and studies in which causation of the IM in disease onset and progression is directly tested, are largely lacking. This constitutes a major problem, especially from the consumer's point of view. Consumers are not scientists, and advertising pre- and

probiotics, and “functional foods”, as a cure to all diseases for the sake of making money is unethical and dangerous (Slashinski *et al.*, 2012).

Furthermore, many microbiome studies have been conducted on mice, which are genetically and physiologically different from humans. Identifying possible interplay between the IM and certain medications or diseases in a mouse-model could help inform human-focused studies. However, it should not be assumed that the same effect will be seen in humans until the hypothesis was explicitly tested on humans (Bik, 2016).

The IM offers promising opportunities for therapy, and this is mentioned in many articles (West *et al.*, 2015; Kho and Lal, 2018). However, there are many factors that contribute to the risk of contracting diseases, particularly non-communicable diseases. Thus, while the microbiome should be included in the treatment of these diseases, and IM-targeted therapy warrants further investigation, it should be noted that altering the IM by itself may not have long-lasting effects. For example, IM-therapy may improve the treatment of obesity, but without a change in diet and exercise regime, it is doubtful whether long-term improvement will be seen. After all, a diverse diet supports a diverse array of microbes in the GIT (Walter and Ley, 2011), so without long term consumption of diverse foods, the GIT might not support the growth of beneficial microbes for long enough to mitigate obesity. Vaginal swabbing as a way of transmitting microbes to babies, especially in cases where the mother underwent c-section as opposed to natural birth, has become quite popular amongst women. This practice is risky, especially when not performed by a qualified doctor, as one could also pass harmful microbes to the baby that could lead to severe infection (Ma *et al.*, 2018). Much more research is needed to determine the underpinnings of the IM in human health and disease, and until research is conclusive enough to inform the healthcare sector, caution is advised when considering IM-therapy.

Future perspectives

While the study of the human microbiome presents a promising avenue for therapy and could help shed light on the commonality of certain diseases in Western populations, much research is still needed before this can be put into practice. Furthermore, standardisation of protocols, including sampling, DNA extraction, sequencing, and analysis protocols, is a necessity if this field is to move forward.

Comparative analyses between the IMs of traditional and Western societies have highlighted compositional and functional differences between them. Analyses of how certain factors, like

antibiotic use, age, and genotype, influence IM composition in Western contexts have also been fruitful. However, to gain a deeper understanding of how those factors influence the IM in light of Westernisation, studies investigating the influence of those factors in traditional, pre-Western settings must be undertaken. Such studies should be performed with the intention of bettering the lives of traditional communities as well and not just with the aim of improving life for Western populations. Additionally, investigation of the IM in traditional populations from southern Africa has not been conducted.

The Ju|'hoansi people from the Nyae Nyae conservancy (NNC) in north-eastern Namibia present an opportunity to study the traditional IM in a southern African context. The Ju|'hoansi lived a largely traditional lifestyle before the 1960s, subsisting by way of hunting and gathering. However, due to the influences of colonisation in Africa, the Ju|'hoansi adopted a more sedentary way of life inclusive of agricultural practices (Hitchcock, 2020). They hunt and gather what they can, but also farm, and increasingly more frequently, consume Western foods available from a small shop in Tsumkwe. Tsumkwe also has a small clinic, where the people obtain antibiotics for tuberculosis, as well as malaria medication. Occasionally, the Ju|'hoansi will make use of paracetamol or ibuprofen for stomach ailments. However, there is limited availability of medicine. As such, the Ju|'hoansi are in the process of Westernising. They live in traditional environments and have access to bush food and traditional medicine, but also have increasingly more frequent access to Western commodities. Studying the IMs of traditional people in relation to factors known to influence the IMs of Westerners, such as age, use of modern medication, biological sex, and residential environment, could shed light on how the IM functions in a traditional context, as well as what the consequences of Westernisation are on IM health.

Conclusion

Fluctuations in IM composition and functionality has been associated with many aspects of human lifestyle, from drug use to age. Although a causal relationship has not been established in many cases in terms of disease, there is ample evidence to justify the notion that such a relationship does exist. Much research is currently being conducted with the specific aim being to find evidence of causality, and the underpinning mechanisms thereof in disease onset and progression, which will hopefully bring us closer to practical application of microbiome research.

The potential applications for microbiome research include informing lifestyle choices, such as dietary choices, frequency of exercise, and use of medication. The effect of a Western

lifestyle on the IM and human health in general should be investigated rigorously to ensure that maximum benefit can be extracted from Westernised living, without its accompanying disadvantages. Hopefully, considerations regarding the microbiome will also be included to inform policies, from the food given to children at school, to the amount of antibiotics prescribed by doctors.

Traditional societies can be used to supplement this research, as their IMs are less affected by Westernisation, and could thus aid the understanding of the effect of Westernisation on the IM. Longitudinal investigations of their IM composition as traditional individuals Westernise could also aid understanding of the evolution of the IM from a traditional to a Western state. Additionally, these societies are at risk of modern-day diseases, without necessarily having the financial means to seek healthcare. Therefore, traditional IM studies, as well as monitoring the health of traditional populations, should be conducted not only with the aim of better informing health in a Western context, but to ensure the health of traditional populations as they enter the Western world.

Finally, more research is needed regarding the bioinformatic aspects of microbiome research. Because the field is relatively new, methods and packages to analyse this data generate highly variable outcomes, which complicates any attempt to reach solid conclusions. To generate reproducible, comparable results in the microbiome field, standardisation of procedures is necessary.

Chapter 2: The diversity of the Ju|'hoansi intestinal microbiome

Mia Truter1*, Jessica E. Koopman1*, Leon ≠Oma Tsamkxao3, Simon J. Underdown1,2, Riaan Rifkin1

1 Center for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, South Africa

2 Department of Anthropology and Geography, Human Origins and Palaeoenvironmental Research Group, Oxford Brookes University, Oxford, UK

3 Ju|'hoan Traditional Authority (JUTA), Tsumkwe, Otjozondjupa Region, Namibia

* Contributed equally to this research.

Abstract

To elucidate the taxonomic composition of a traditional southern African hunter-gatherer intestinal microbiome (IM), we analyse 40 faecal samples derived from the Ju|'hoansi San/Bushmen of north-eastern Namibia. This population subsists largely by hunting and gathering but is increasingly exposed to Western commodities. The Ju|'hoansi therefore presents an opportunity to study the evolution of the IM during transition from a hunter-gatherer to a Western lifestyle, *in situ*. We analyse their IM composition in relation 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant. The Ju|'hoansi IM resembles a typical hunter-gatherer IM, being enriched for *Prevotella, Blautia, Faecalibacterium, Succinivibrio* and *Treponema*. No significant differences were found in IM composition in relation to any of the factors tested, except between subjects of different primary villages of residence. This is expected to change over time, as exposure to Western dietary resources and medication becomes more frequent. This draft article is prepared for publication in **Cell Press Community Review.**

Introduction

The human gastrointestinal tract (GIT) harbours a dynamic population of bacteria, archaea, fungi, protozoa and viruses, *i.e.*, the intestinal microbiota. The human intestinal microbiome (IM) (Lederberg and McCray, 2001) performs critical functions in digestion, development, and immunity (Thursby and Juge, 2017). Modifications of IM composition (dysbiosis) have been associated with the pathogenesis of inflammatory and auto-immune diseases, including allergies (Hansen, Gerasimidis and Turner, 2019), obesity (Mar Rodríguez *et al.*, 2015; Davis, 2016), diabetes (Dunne *et al.*, 2014), and inflammatory bowel disease (Ott *et al.*, 2009; Vich Vila *et al.*, 2018). Prior to the advent of agriculture, following the beginning of the Neolithic *c*. 10,000 years ago, humans subsisted solely by hunting and gathering. The lifestyle changes associated with the Neolithic are stated to have significantly impacted our IM taxonomic composition, and metabolic capacity. But precisely how our IMs changed, is

largely unresolved (Blaser and Falkow, 2009; Adler *et al.*, 2013). Currently, traditional populations provide the only available examples of a 'pre-Neolithic' human IM, owing largely to their comparatively limited exposure to Western lifestyle factors, including various novel sources of food, medication and toxic pollutants. However, even these communities are subject to the influences of Westernisation (Crittenden and Schnorr, 2017) and as such, represent a window of opportunity to study the evolution of the IM during transition from a pre-Western, non-industrialised, to a Western, industrialised lifestyle.

Differences in IM adaptations to diverse lifestyles are prevalent between industrialised Western societies from Europe and North America, and non-industrialised rural populations from Africa and South America (Schnorr *et al.*, 2014). Several socio-economic differences exist between traditional and Western populations, many of which may exert an impact on human IM composition. Traditional populations typically adhere to a high-fibre, low-fat and low-sugar diet, and generally have limited access to contemporary medicine. They also live in closer association with one another, with their pets, livestock and wildlife, and with environmental microbes. In contrast, Western diets tend to comprise processed, high-fat, low-fibre foods, combined with increased sedentarism and easier access to modern medication. Westerners also tend to experience less exposure to the natural environment and associated environmental microbes (Rook, Raison and Lowry, 2014; Thorburn, Macia and Mackay, 2014). These factors are consequently thought to drive the compositional differences seen between the IMs of traditional and Western populations. Traditional populations tend to harbour a more diverse IM which contains a higher abundance of short chain fatty acid (SCFA) producing bacteria, such as *Prevotella*, *Succinivibrio* and *Treponema* (De Filippo *et al.*, 2010; Gupta, Paul and Dutta, 2017). The shift in taxonomic composition, including the disappearance of 'cornerstone' IM members and changes in structural composition, are suspected to partially contribute to the higher prevalence of inflammatory diseases commonly seen in Western populations (Yatsunenko *et al.*, 2012; Sonnenburg and Sonnenburg, 2014; Amato *et al.*, 2019).

Studies concerning the Tanzanian Hadza hunter-gatherers (Schnorr *et al.*, 2014), Venezuelan Yanomami Amerindians (Clemente *et al.*, 2015), the BaAka in the Central African Republic (Gomez *et al.*, 2016b) and the Arctic Inuit (Dubois *et al.*, 2017) have provided insight into the IM composition of 'traditional' societies. To date, comparable research has not been conducted in southern Africa. Moreover, whereas multiple comparative studies between traditional and Western IMs have been carried out in other parts of the world, few studies explicitly investigate the effect of certain lifestyle factors (such as medical history and residential mobility) within traditional communities in terms of their effect on IM variability.

The Ju|'hoansi (pronounced 'zhu-t-wasi') San/Bushmen hunter-gatherers inhabit the Nyae Nyae Conservancy (NNC) in north-eastern Namibia, which was established in 1998 and covers 8,992 square kilometres (**Figure 1**). It is home to around 2,300 Ju|'hoansi, and also to Bantu-speaking Herero agro-pastoralists. Prior to the 1970s, the Ju|'hoansi traditionally subsisted by way of hunting and gathering no less than 85 species of wild plant, including mongongo nuts (*Schinziophyton rautanenii*) (Lee, 2017). Following the onset of the summer rains in December, the Ju|'hoansi diet mostly comprise 'bush-food', including bush-potatoes, water-carrots, various species of geophytes termed 'wild onions' and also *Grewia* sp. and baobab (*Adansonia digitata*) fruits and honey. Hunting and trapping focused on game such as kori bustard (*Ardeotis kori*), helmeted guineafowl (*Numida meleagris*), steenbok (*Raphicerus campestris*), springhare (*Pedetes capensis*) and porcupine (*Hystrix africaeaustralis*) throughout the year. By July, foraging was less important as natural resources became less abundant, although certain rhizomes and *Acacia* tree resins were still collected (Imamura-Hayaki, 1996).

In the 1970s, several small shops, a liquor store and a clinic were introduced to the NNC. This exposed the Ju|'hoansi to Western commodities like sugar, canned foods, coffee, tea, and mielie meel. Charity organisations in the area provide food packages once a month, which contain the abovementioned commodities. In the 1980s, several foundations assisted the Ju|'hoansi to plant gardens and raise livestock such as cattle (Gargallo, 2020). Such agricultural initiatives also included the planting of papaya (*Carica papaya*), beetroot (*Beta vulgaris*), carrots (*Daucus carota*), onions (*Allium cepa*) and tomatoes (*Solanum lycopersicum*) in several Ju|'hoansi villages.

The Ju|'hoansi are thus reliant on a complicated mixture of subsistence strategies (Denker, Thompson and Jarvis, 2012). The Western food packages provided by charity organisations, as well as vegetables grown in their gardens and gathered from the bush, form the basis of the Ju|'hoansi diet. Each village has one or two men that hunt opportunistically, providing game meat for the whole village as often as they can, which ranges anywhere between once a week and once a month.

To determine the taxonomic composition and metabolic functionality of both the bacterial and fungal IMs in a traditional, southern African context, we analysed faecal samples derived from 40 Ju|'hoansi community members inhabiting four villages, namely Duinpos, Den/ui, Mountain Pos and !Om!o!o. The faecal samples were collected in July 2019, during the dry season (winter). Our study focused on how the taxonomic and metabolic variations of the Ju|'hoansi IM might relate to eight biological and abiotic environmental variables, namely; 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis,

3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant.

Figure 1: **Lifestyle of the Namibian Ju|'hoansi.**

(A) The Ju|'hoansi live in the Nyae Nyae Conservancy in north-eastern Namibia.

(B, C) Historically, they subsist by way of hunting and gathering, but also have increasingly more frequent access to Western commodities.

(D) To elucidate the taxonomic composition and metabolic functionality of their bacterial and fungal intestinal microbiomes, faecal samples were collected for analysis. During ethnographic interviews, questionnaires were completed and with the assistance of Leon ≠Oma Tsamkxao, translated from Ju|'hoansi to English, to determine how certain lifestyle factors influence the composition of the IM.

Results

Characterising the Ju|'hoansi IM by 16S rRNA and ITS sequencing

Forty faecal samples and two controls, namely KIT-CTRL (kit control – buffers of the used extraction kit, no sample) and CON-CTRL (sampling container control) were analysed in this study. A total of 4,679,902 16S forward and reverse reads were imported into QIIME2 and merged, resulting in a mean read count of 38,031 reads per sample. A total of 5,938,170 ITS forward reads were imported into QIIME2, resulting in a mean read count of 88,116.4 reads per sample. Reverse ITS reads were not included in the analysis, since merging forward and reverse reads only resulted in a mean read count of 5,446.4 reads per sample. Quality control with DADA2 yielded 4,184 and 1,271 ASVs (Amplicon Sequence Variants) for 16S and ITS data, respectively. As initial ITS taxonomic classification of the ASVs resulted in identification of very few taxa, ITS reads were first clustered at 98% sequence similarity and then re-classified, which resulted in 167 OTUs (Operational Taxonomic Units) (**SI Table 2**).

Following taxonomic classification, the 16S ASV table was filtered to only include reads that appeared in more than two samples, which resulted in the loss of CON-CTRL. The ITS OTU table was left unfiltered to preserve as many OTUs as possible, in light of the comparatively lower OTU count.

ASV/OTU and taxonomy tables were then imported into R and contaminant reads were identified using the "prevalence" contamination identification method of *decontam* (Davis *et al.*, 2017). This method evaluates the probability of each ASV/OTU being a contaminant by comparing its prevalence between true samples and negative controls using the chi-square statistic, under the assumption that contaminants are more prevalent in negative controls than true samples. Twelve reads from three bacterial taxa were identified as contaminants using this method, namely *Streptococcus salivarius, Parabacteroides merdae* and the *Eubacterium coprostanoligenes* group. Following the removal of these three species, *Firmicutes* (26.6%) and *Bacteroidota* (8.5%) emerged as the dominant bacterial phyla, resulting in a *Firmicutes: Bacteroidota* ratio of 3.1. Other phyla present included *Proteobacteria* (3.16%), *Spirochaetota* (0.28%) and *Actinobacteria* (0.24%). In total, 131 bacterial genera were identified, with the top 5 taxa comprising *Prevotella* (6.6%), *Blautia* (3.3%), *Faecalibacterium* (2.0%), *Succinivibrio* (1.8%) and *Christensenellaceae R-7 group* (1.5%). *Treponema* was also present at an abundance of 0.3% **(SI Table 3.1-3.6)**.

Fungal contaminant identification (using the same method as above) yielded 12 contaminant reads from four species, namely *Malassezia globosa*, *Pleosporales sp.*, *Saccharomycetales sp.* and *Candida albicans*, which were subsequently removed from the samples. The two most abundant phyla were *Ascomycota* (19.4%) and *Basidiomycota* (18.6%), with

Chytridiomycota (7.4%) and *Mucoromycota* (5.8%) comprising the remainder. In total, 81 fungal genera were identified, with the top three genera comprising *Malassezia* (7.4%), *Candida* (6.4%) and *Naganishia* (4.3%) (**SI Table 4.1-4.6**).

Core bacterial and fungal taxa of the Ju|'hoansi IM

Core IM taxa are shared across human populations (Martínez *et al.*, 2015) and are thought to perform important metabolic functions (Turnbaugh *et al.*, 2007). One aim of this study was to determine whether a core IM was shared among individuals inhabiting the NNC. Core IM members were defined as those taxa that were present in 60% and 50% of all bacterial and fungal datasets, respectively, at a relative abundance of at least 0.008%. These parameters were chosen using a computational method developed by (Salonen *et al.*, 2012), in which multiple prevalence and detection thresholds are investigated in terms of their effect on core size. The optimal parameters are then chosen such that they have the least effect on core size (**SI Figures 1A and B**). There are currently no standardised prevalence and detection thresholds for elucidating the core microbiome. Prevalence thresholds of 50% (Martínez *et al.*, 2015) and 90% (Mancabelli *et al.*, 2017) have been used, with no mention of detection threshold. The method employed by (Salonen *et al.*, 2012) is the recommended method as per the Microbiome R-package documentation (Lathi and Shetty, 2017). The Ju|'hoansi bacterial and fungal core microbiomes comprised 54 and 8 genera, respectively (**Figure *2***). A large proportion of both the bacterial (5%) and fungal (17%) core microbiomes consisted of unknown taxa.

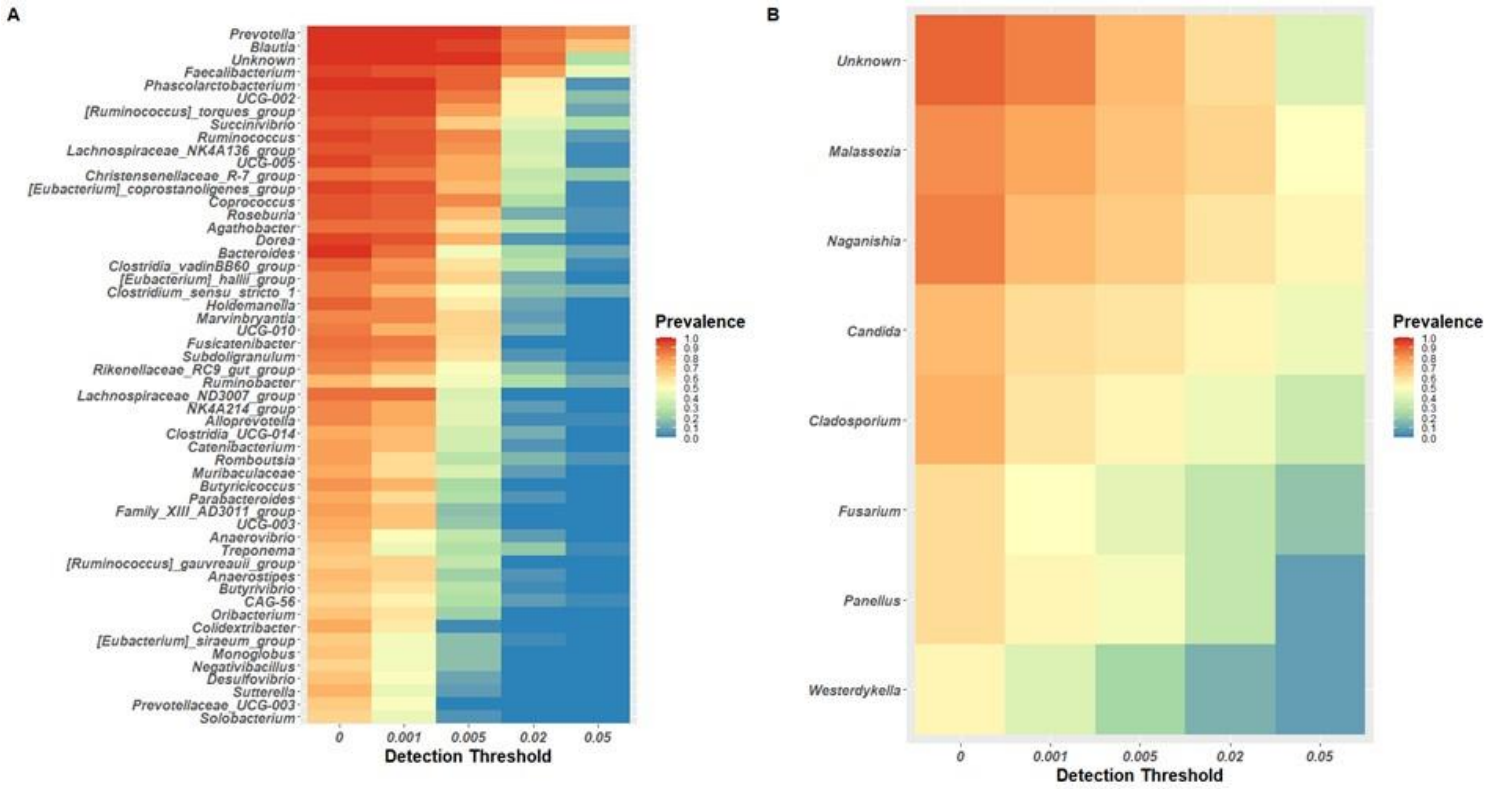


Figure 2: The core bacterial and fungal microbiome of the Ju|'hoansi hunter-gatherers.

(A) The bacterial core microbiome

(B) The fungal core microbiome

Community composition and differentially abundant taxa of the Ju|'hoansi IM

The Ju|'hoansi IM exhibited an even and diverse bacterial landscape, however, the fungal landscape varied widely in terms of evenness and diversity between cohorts. The controls had very low α-diversity in comparison to the true samples (**Figure 3A** (16S) and **Figure 4A** (ITS)). No statistically significant differences in α-diversity were detected between groups for any of the factors tested (**SI Table 5.1-5.3, SI Table 6.1-6.3**).

Whereas the weighted UniFrac metric showed statistically significant (p = 0.001) bacterial β-diversity differences between IM populations in residents from different villages (*i.e.*, Duinpos, Den/ui, Mountain Pos and !Om!o!o.), the Jensen-Shannon metric did not (p = 0.086) (**SI Table 7.1-7.2, Figure 3B and C**). Fungal IM populations of residents from different villages exhibited differences in β-diversity that were statistically significant using both distance metrics (Bray-Curtis p = 0.003, Jensen-Shannon p = 0.001) (**SI Table 8.1-8.2, Figure 4B and C**). No other groups presented statistically significant differences in either bacterial or fungal community structures.

Further analysis of average pairwise Weighted UniFrac distances between specific villages of residence exhibited no statistically significant differences between bacterial community composition, while Jensen-Shannon distance indicated that four out of six village

comparisons were significant in terms of bacterial community composition (**Figure 3D and E**). Analysis of fungal community composition between specific villages showed one (Bray-Curtis distance) and five (Jensen-Shannon distance) out of six significant differences between villages (**Figure 4D and E**).

Due to uneven sample sizes between groups and subsequent statistical inaccuracy, 1) the ages of research participants, 2) their former use of antibiotic treatment for antibiotics, 6) their former or current use of malaria indication and 7) their exposure to local, regional and international travel were excluded from differential abundance analysis.

ANCOM-BC was used to identify genera that exhibited statistically significant differences in abundance between groups. A few genera were statistically significantly different between subjects of different biological sex, those who have and have not used antibiotics, and those who have and have not experienced diarrhoea and intestinal infection (**SI Tables 9.1-9.3 (16S) and 10.1-10.3 (ITS))**. Owing to the significant differences in β-diversity between subjects from different villages however, the different abundance genera between villages was of more interest (**Figure 3F and Figure 4F)**. The bacterial genera included *Eubacterium ruminantium group, Anaerostipes, Butyrivibrio, Desulfovibrio, Ruminococcus* and *Treponema.* The fungal genera included *Aspergillus, Candida, Mycoacia, Panellus, Porodisculus, Sacchararomyces, Schizophyllum* and *Stagonospora*. The functional significance underlying these differences would make for an interesting study.

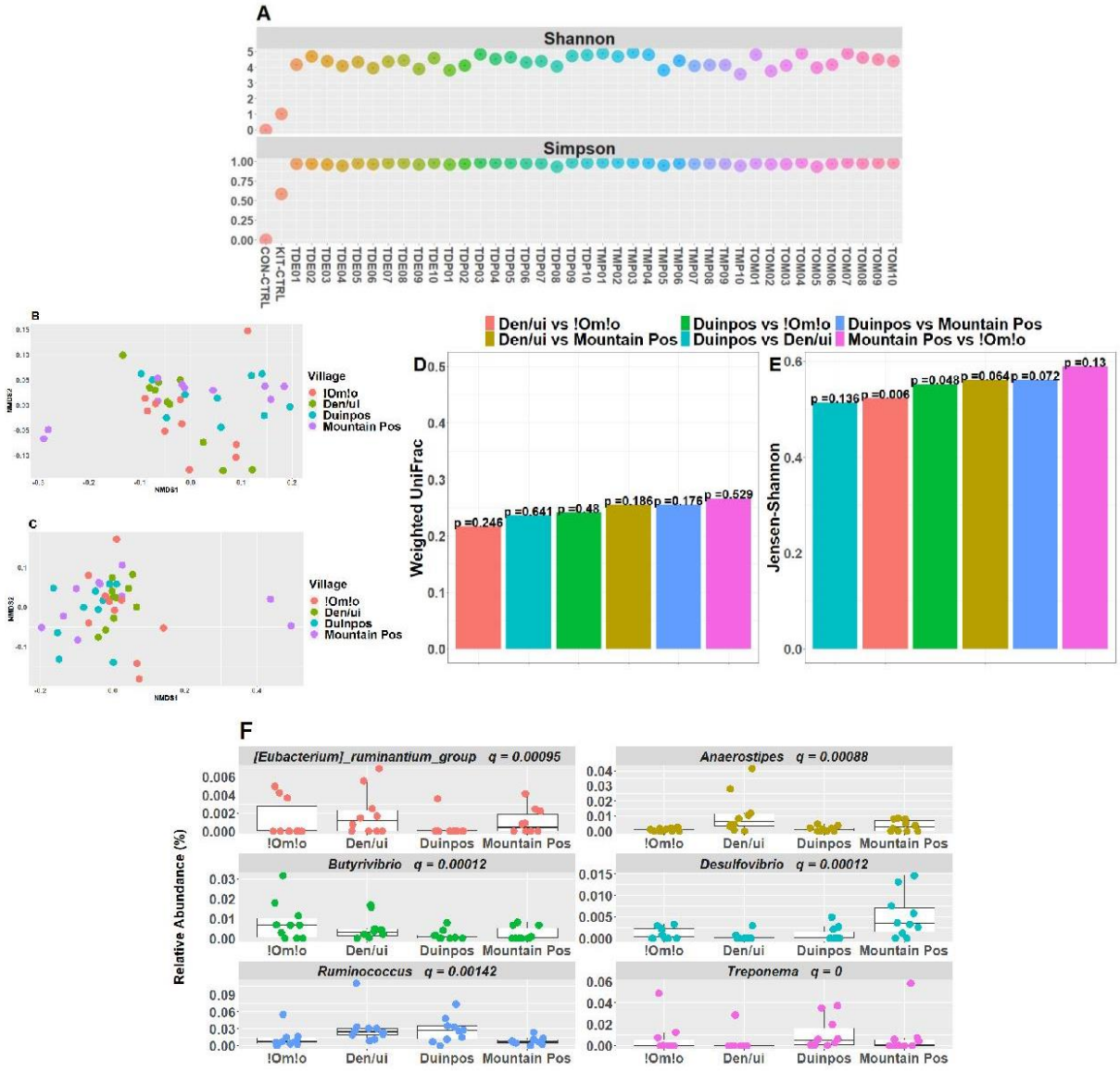


Figure 3: Bacterial community composition of the Ju|'hoansi IM.

(A) Per-sample α-diversity computed with Shannon and Simpson indices.

(B) Weighted UniFrac and (C) Jensen-Shannon NMDS plots for village of primary residence, the only factor that showed significant community structure differences with PERMANOVA.

Average pairwise distances between villages using D) Weighted UniFrac and E) Jensen-Shannon distance metrics. False Discovery Rate (FDR) corrected p-values included on top of each bar. FDR corrected *p*-values < 0.05 were considered significant.

(F) Differentially abundant genera between villages as identified by ANCOM-BC. Genera were considered significant if Benjamin-Hochberg adjusted *p*-values (*q*-values) were <0.05.

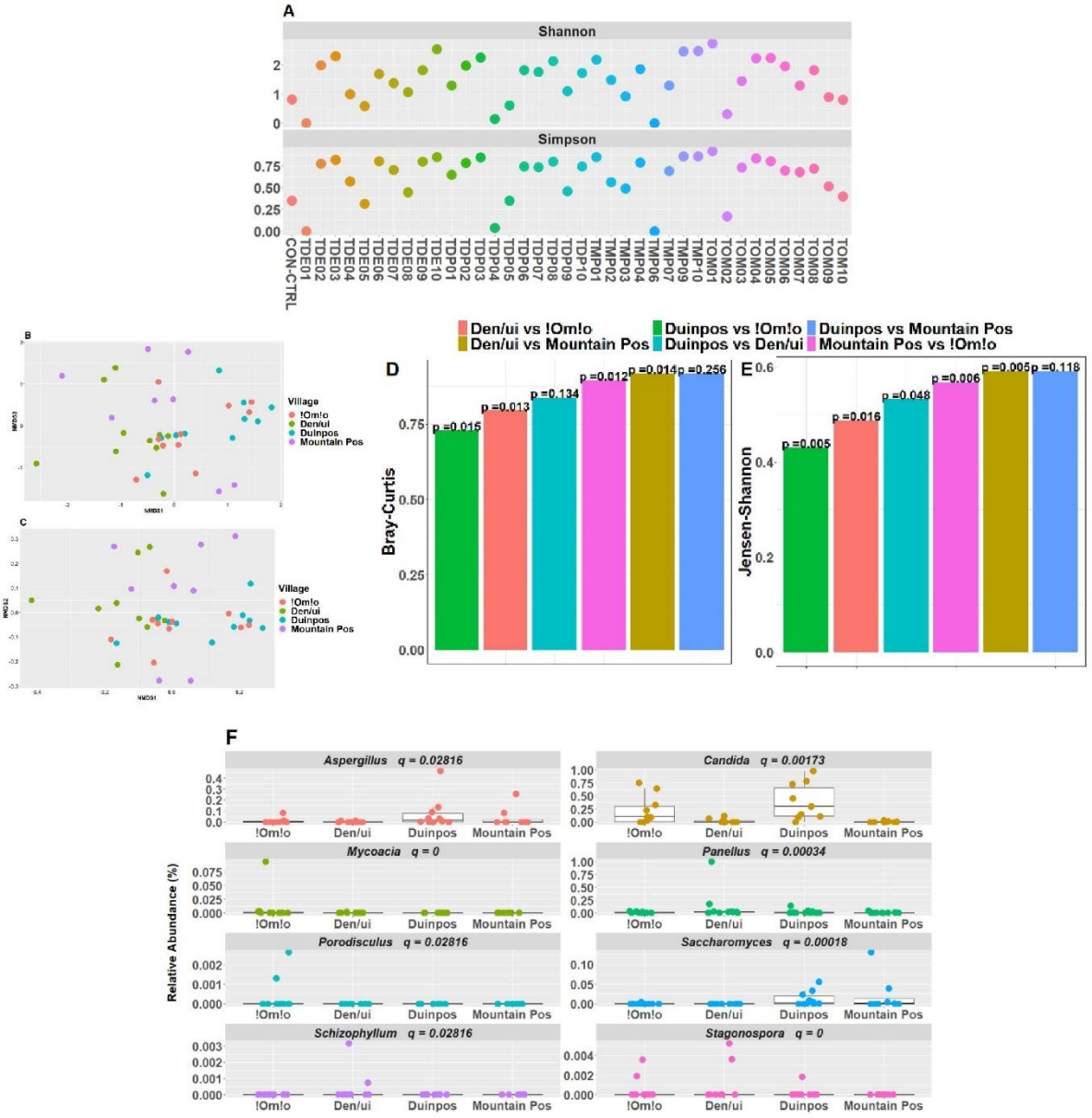


Figure 4: Fungal community composition community composition of the Ju|'hoansi IM.

(A) Per-sample α-diversity computed with Shannon and Simpson indices.

(B) Bray-Curtis and C) Jensen-Shannon NMDS plots for village of primary residence. The only factor that showed significant community structure differences with PERMANOVA.

Average pairwise distances between villages using D) Bray-Curtis and E) Jensen-Shannon distance metrics. False Discovery Rate (FDR) corrected p-values included on top of each bar. FDR corrected *p*-values < 0.05 were considered significant.

(F) Differentially abundant genera between villages as identified by ANCOM-BC. Genera were considered significant if Benjamin-Hochberg adjusted *p*-values (*q*-values) were <0.05.

Discussion

To determine the bacterial and fungal taxonomic composition of the Ju|'hoansi IM, 40 faecal samples were sequenced and analysed. We observe that 1) Ju|'hoansi IM is enriched for bacteria commonly found in other traditional populations, and 2) the only factor for which taxonomic IM differences were statistically significant between subjects, was village of primary residence.

The Ju|'hoansi IM harbours a high abundance of microbes that ferment fibre and plant polysaccharides, including *Prevotella, Blautia, Faecalibacterium, Succinivibrio* and *Treponema.* These bacteria convert fibre into metabolically advantageous SCFAs, namely propionate, acetate and butyrate, which have anti-carcinogenic and anti-inflammatory properties (Cordain *et al.*, 2005; Sivaprakasam, Prasad and Singh, 2016). The high abundance of fibre fermenting bacteria in the Ju|'hoansi IM is likely a reflection of their fibre-rich diet, including food items such mongongo nuts, which have around 3.5 g and 2.7g of fibre per 100g in the flesh and kernel respectively(Lee, 2010). Additionally, the presence of *Treponema* in the Ju|'hoansi IM is arguably indicative of a traditional lifestyle, as this organism is thought to occur in the IMs of traditional societies as a result of cross-contamination from termites or swine and is rare in the IMs of Western populations (Angelakis *et al.*, 2019).

The high abundance of fibre fermenters in the Ju|'hoansi IM is similar to what is found in the IMs of other traditional societies that adhere to a similar lifestyle. Children from Burkina Faso harbour high abundances of the same bacteria (De Filippo *et al.*, 2017), as do individuals from rural Nigeria (Ayeni *et al.*, 2018), as well as the Hadza hunter-gatherers from Tanzania (Schnorr *et al.*, 2014). Interestingly, these populations adhere to a similar diet as the Ju|'hoansi – a diet high in carbohydrates and fibre, and relatively low in animal fats and proteins. The Tanzanian Hadza gatherers, for example, are documented as consuming tubers, berries, baobab fruit and occasionally game, which is very similar to the Ju|'hoansi diet (Marlowe, 2002). The similarities in diet across these rural populations could explain the similarity in IM composition.

Thus far, the majority of studies investigating the mycobiome have been in the context of healthy vs. diseased patients. For example, elucidating mycobiome composition between patients with and without Crohn's disease (Li *et al.*, 2014), or between obese and healthy subjects (Mar Rodríguez *et al.*, 2015). The inclusion of the mycobiome in a study investigating the IM of a traditional population is novel, and as such, comparison of our results to existing literature is challenging. Nonetheless, *Candida* and *Malassezia* were the most abundant fungal species in the Ju|'hoansi IM. Unlike the stark differences between

bacterial residents of the GIT in Western and traditional populations, the fungal residents are somewhat similar. Both *Candida* and *Malassezia* have been identified as common inhabitants of the Western IM (Nash *et al.*, 2017). Interestingly, Nash *et al.*, 2017 also identifies *Saccharomyces* as part of a healthy Western IM, however, this organism does not form part of the core Ju|'hoansi microbiome, although it is present at a low abundance (0.3%). The significance of the fungal landscape in the Ju|'hoansi in terms of its relation to diet, geographic location and culture is somewhat unclear, and will hopefully be elucidated upon further research.

The statistically significant difference in β-diversity observed between residents from different villages could be a result of several factors, including 1) socio-economic status, 2) different vegetation surrounding each village and 3) different water sources at each village. The villages (Duinpos, Den/ui, Mountain Pos and !Om!o!o) have varying degrees of affluence, with some possessing commodities like gardens and cattle, while others do not. Socio-economic status is known to affect IM composition (Bowyer *et al.*, 2019), since socio-economic status determines factors such as the type of food that is accessible to the individual and the level of psychosocial stress that the individual experiences (Amato *et al.*, 2021). Furthermore, the vegetation type surrounding each village is different leading to slight changes in the type of food that is consumed most frequently. Similarly, each village has its own borehole that could be supporting different types of microbial growth, which could in turn affect which microbes the village residents are exposed to. It would be of interest to elucidate what underlying factors drive IM differences between villages of primary residence.

Contrary to expectation, other factors that were investigated, namely 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, did not significantly alter IM composition. This suggests that factors such as diet and culture may have a greater effect on IM composition than do medical history, age, biological sex and travel.

As the Ju|'hoansi Westernise, gaining more frequent access to Western foods and medical care, their IMs are expected to adapt to resemble a Western IM more closely. A longitudinal study in which the Ju|'hoansi IM is analysed in context of gradual Westernisation, presents an interesting opportunity to investigate the effects of Westernisation on a traditional IM, *in situ.* This could shed light on the changes that occurred in the IM as a response to Westernisation during the Neolithic revolution and could aid the understanding of the IM's role in the onset of Western diseases such as obesity, diabetes, cardiovascular disorders and cancer.

Additionally, monitoring the Ju|'hoansi IM during this period of transition to a Western lifestyle could help prevent the Ju|'hoansi from losing their IM diversity as Westerners have, hopefully mitigating against some of the diseases of civilisation that inevitably accompanies Westernisation.

Conclusion

To gain insight into the taxonomic composition and metabolic capacity of a southern African IM transitioning from a hunter-gatherer to a Western lifestyle, faecal samples from 40 Ju|'hoansi participants were sequenced. To analyse how their IMs might differ in response to lifestyle factors within the community, interviews were also conducted to determine to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant. The Ju|'hoansi harboured bacteria associated with plant fibre fermentation, namely *Prevotella, Blautia, Faecalibacterium, Succinivibrio* and *Treponema.* These microbes are present in traditional societies that follow a similar lifestyle to the Ju|'hoansi, indicating that diet could be a significant determinant of IM composition. The Ju|'hoansi IM also contained a high abundance of *Candida* and *Malassezia,* both of which are common in Western IMs. However, the Ju|'hoansi IM lacked the common Western fungi, *Saccharomyces.* The functional significance of this is uncertain.

The statistically significant difference in β-diversity between subjects from different villages is curious and could be a result of 1) socio-economic status, 2) different vegetation surrounding each village and 3) different water sources at each village. Contrary to expectation, the use of antibiotics, biological sex, age and exposure to travel did not significantly affect β-diversity between subjects. This suggests that factors such as diet and culture may have a greater effect on IM composition than do medical history, age, biological sex and travel.

The intention of this study was to function as a baseline characterisation of the Ju|'hoansi IM from which a longitudinal analysis, documenting the evolution of the IM from a traditional to a Western state, can be conducted.

Materials and Methods

Study design

This study was approved by the Research Ethics Committee, Faculty of Health Sciences at the University of Pretoria, South Africa (Protocol number TEMP 2017-01469, Reference Number NAS032/2021). The research permit for this study was approved by the Namibian National Commission on Research and Technology (NCRST) (RPIV00692019). The Ju|'hoansi Traditional Authority (JUTA) provided consent for the enrolment of project participants from the Nyae Nyae Conservancy. Research participants were recruited with the assistance of our Tsumkwe San co-researcher, research facilitator and interpreter, Leon ≠Oma Tsamkxao, who is fluent in Ju|'hoansi, Afrikaans and English, and written informed consent was obtained from all participants. All participants provided consent for publication of study results of the collected biomaterials, agreeing that all information required for the study (i.e., their location, gender, age, and medical history), except for their names, could be disclosed in this study. All the research methods occurred in accordance with the Helsinki Declaration.

Along with these samples, metadata was also collected in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant. Although it was specifically asked whether diarrhoea occurred upon consumption of bush or shop-bought food, we combined this information into 'Yes' or 'No' for whether diarrhoea occurs regardless of food type, to enable statistical analysis with a large enough sample size.

SI Table 1: Metadata collected for this study.

Data availability: NGS data will be made available after publishing, or upon request.

DNA extraction and sequencing

The samples were stored in faecal collection tubes containing 9 ml DNA/RNA ShieldTM (Zymo Research Corp, Irvine, CA, USA). After homogenizing the samples through vortexing, ~1 ml was transferred to a clean 2 ml tube, spun for 5 min at 10.000 x g and the supernatant removed. The average weight of the resulting pellets was 125 mg, which was subsequently resuspended in 750 µl bead solution from the DNeasy® PowerLyzer® PowerSoil® Kit (Qiagen GmbH, Hilden, Germany). The DNA isolation was performed according to the manufacturers protocol, with the following adaptations: two rounds of beadbeating (1 min and 4000 rpm, PowerLyzerTM, Mo Bio Laboratories, Inc., Carlsbad, Ca, USA) followed by 5 min

incubation on ice, the beadbeating tubes were spun for 5 min. After addition of Solution C6 (elution buffer) the spin columns were incubation at RT for 5 min before centrifugation. Paired-end (2 x 300bp) sequencing of the isolated DNA (V3-V4 16S rRNA for bacteria, and ITS1 and 2 for fungi) was performed at Applied Biological Materials Inc., Richmond, B.C. Canada using the MiSeq platform (Illumina, San Diego, CA, USA). (**SI Table 9)**

Two controls were used in this study. CON-CTRL contained DNA/RNA ShieldTM (Zymo Research Corp, Irvine, CA, USA) used to preserve the samples, while KIT-CTRL comprised the contents of the DNeasy® PowerLyzer® PowerSoil® Kit (Qiagen GmbH, Hilden, Germany).

Bioinformatic analysis

Raw paired-end 16S and forward ITS reads were imported into QIIME2-2021.2 (Bolyen *et al*., 2019). Reads underwent quality control using DADA2 (Callahan *et al.*, 2016), including denoising, dereplication, and filtering of chimeras. The 16S reads were also merged during the quality control process. The 3' ends of the 16S forward reads were truncated to a length 292 bp, and 25 bp were trimmed from the 5' end. The 3' ends of the 16S reverse reads were truncated to a length of 250 bp, and 25 bp were trimmed from the 5' end. ITS forward reads were truncated to a length of 297 bp at their 3' ends, and 26 bp were trimmed from the 5' end. The rest of the parameters were left at default. The ITS reads were subsequently clustered using closed-reference clustering at 98% similarity using qiime vsearch (Rognes *et al*., 2016). 16S taxonomic classification was performed by extracting V3-V4 regions from the SILVA-138-99 database (Quast *et al*., 2013) using q2 feature-classifier extract-reads, based on the primer sequences used to amplify the 16S data. A naïve-Bayes classifier was then trained on the extracted SILVA sequences, and full-length UNITE version 8 (Nilsson *et al*., 2019) dynamic sequences for 16S and ITS data respectively. The classifiers were then used to taxonomically classify the respective datasets using the qiime fit-classifier naïve-Bayes plug-in (Pedregosa *et al*., 2011). Post-classification, the 16S feature table was filtered to only include reads that appeared in more than two samples, resulting in the loss of CON-CTRL. The ASV table was then imported into R-4.1.0 (Team, 2013), along with the unfiltered ITS OTU table. Feature tables were normalised to relative abundance and contaminants were then removed using decontam in R (Davis *et al.*, 2017) at a prevalence threshold of 0.1. Decontam works by determining the likelihood of a read being a contaminant based on the prevalence of the read between controls and true samples. The identified contaminant reads were then removed from the phyloseq objects for downstream analysis. Exploring taxonomic abundance in R was done using dplyr (Wickham *et al*., 2018) and plotrix (J, 2006) in R. The core microbiome was elucidated using the microbiome package in R (Lahti and Shetty, n.d.).

α- and β-diversity was visualised in R using the phyloseq (McMurdie and Holmes, 2013), microbiome, tidyverse (Wickham *et al*., 2019), vegan (Oksanen *et al*., 2020), pairwise adonis (Martinez Arbizu, 2017), and mctoolsr (https://github.com/leffj/mctoolsr/) packages. Statistical significance for α-diversity was computed using Kruskal Wallace (two groups) and Dunn's tests (more than two groups). Where applicable, p-values were corrected using the Benjamini Hochberg method. Differential abundance testing was performed with ANCOM-BC (Lin and Peddada, 2020). All R-code is provided as supplementary material (**Supplementary R-Code**). All visualisations were created using ggplot2 (Wickham, 2016) and RColorBrewer (Neuwirth, 2014) in R. All **SI Tables** were created in Microsoft Excel (Microsoft Corporation, n.d.).

Author contributions

Conceptualisation, ethics approval, sample collection, interview conduction, RFR, LOT, SU

DNA extraction, JEK

Bioinformatics analysis, MT

Writing, MT, JEK, RFR

Figure creation, MT

Supervision, JEK, RFR

Conflicts of interest

None of the authors had any conflicts of interest.

Acknowledgements

We thank all our study participants and the Ju|'hoansi Traditional Authority (JUTA) for providing consent for the enrolment of project participants in the study. We thank Loide Uahengo and Edgar Mowa (Namibian National Commission on Research and Technology) and Benetus Nangombe (Namibian Ministry of Health and Social Services) for issuing the relevant research permit. We thank Helvi Elago (National Heritage Council of Namibia) and Alma Nankela (Archaeology Unit, National Heritage Council of Namibia) for providing guidance and supporting documentation concerning our permit application. We thank Manda Smith and members of the Research Ethics Committee, Faculty of Health Sciences at the University of Pretoria, South Africa, for providing clearance to conduct the research. We would also like to thank the Benjamin R. Oppenheimer Trust for their financial contribution towards this research.

Chapter 3: Summary

The intestinal microbiome is thought to have aided human survival through the ages by allowing fast, flexible mechanisms of adaptation. For example, the IM's ability to extract extra nutrients from food sources could provide a fitness advantage to the host. As such, hosts and many of the microbes in the IM have likely formed intricate relationships over millennia, providing benefits to one another in a reciprocal fashion.

A pre-Neolithic lifestyle commonly included hunting, gathering, the use of traditional medicine, nomadism, and frequent exposure to environmental microbes. Transition to a Western lifestyle brought about many changes to humanity, such as increased sedentarism, the consumption of processed foods, access to modern medication, and more sanitary living environments. It is expected that this lifestyle shift will be accompanied by a corresponding shift in the taxonomic structure and functionality of the IM.

The consequences of the advent of Westernisation on the IM is studied by comparing the IMs of traditional and modern societies. Traditional societies, such as those residing in parts of Western Africa, and South America, typically harbour microbes associated with the degradation of fibrous plant material and the production of short chain fatty acids (SCFAs). Their IMs are also more diverse than the IMs of Westerners. In contrast, Western IMs are abundant in microbes associated with the degradation of fats and animal products and have the capability to degrade xenobiotics. The lack of diversity and the loss of some microbes in the Western IM is thought to contribute to the high incidence of certain diseases in Western countries like the United States of America, and Italy. These diseases include obesity, asthma, cancer, and irritable bowel disease. While structural IM changes have been observed following the onset of such diseases, investigation of the biological mechanisms that underpin the correlation between disease and IM composition is yet to be elucidated.

Studies investigating traditional IM structure have largely been conducted in comparison to Western IMs, yet certain factors that are known to influence IM composition in Western contexts have not been investigated in traditional contexts. Furthermore, research on traditional IMs from southern Africa is lacking.

To determine the taxonomic composition and metabolic functionality of both the bacterial and fungal IMs in a traditional, southern African context, 40 faecal samples from the Ju|'hoansi community in north-eastern Namibia were analysed. The Ju|'hoansi largely adhere to a traditional way of life, subsisting by way of hunting and gathering, but are also more frequently exposed to Western commodities like processed foods and modern medication. As such, the Ju|'hoansi are in the process of Westernising and present an opportunity to study the evolution of the IM in a southern African context.

The taxonomic structure of the IM was elucidated, and correlations between IM composition and 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant, was also investigated. Functionality of the Ju|'hoansi IM was predicted by considering the known roles of fungi and bacteria that were classified in this study. This dissertation aims to be a preliminary investigation into the taxonomic structure and function of the Ju|'hoansi IM, so as to provide a platform from which to study its evolution in response to future Westernisation.

The Ju|'hoansi harboured an IM that resembles typical traditional IMs, such as those of the Hadza hunter-gatherers and children from Burkina Faso, and was enriched for *Prevotella*, *Blautia*, *Faecalibacterium*, *Succinivibrio*, and *Treponema*. The Ju|'hoansi lacked common Western microbes such as *Bifidobacterium.* Contrary to expectation, the taxonomic structure of the Ju|'hoansi IM was not significantly different between subjects in terms of in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant. However, subjects from different villages harboured IMs with significantly different structural diversity, although the effect size was small. Some of the bacterial genera responsible for these structural differences included *Eubacterium, Treponema, Anaerostipes, Butyrivibrio, Ruminococcus,* and *Desulfovibrio.* Some of the fungal genera that were differentially abundant between subjects from different villages included *Aspergillus, Candida, Panellus, Mycoacia,* and *Saccharomyces.* The reasons that underpin the differences seen in IM composition between the different villages is yet to be discovered, but could include socio-economic status, the use of different boreholes, consumption of different types of bushfoods, or varying ease of transport to and from Tsumkwe. The predicted functional capabilities included chemoheterotrophy and fermentation in terms of bacteria, while most of the fungi identified in the Ju|'hoansi IM were plant- and animal-associated commensals and pathogens.

Due to the covid-19 pandemic, the samples were left in storage from July 2019 to the end of 2020, which may have resulted in the extraction of less microbial DNA. Furthermore, clustering the ITS reads by various methods (open- and closed-reference clustering) and various similarities (90% and 98%), led to differing success in terms of taxonomic

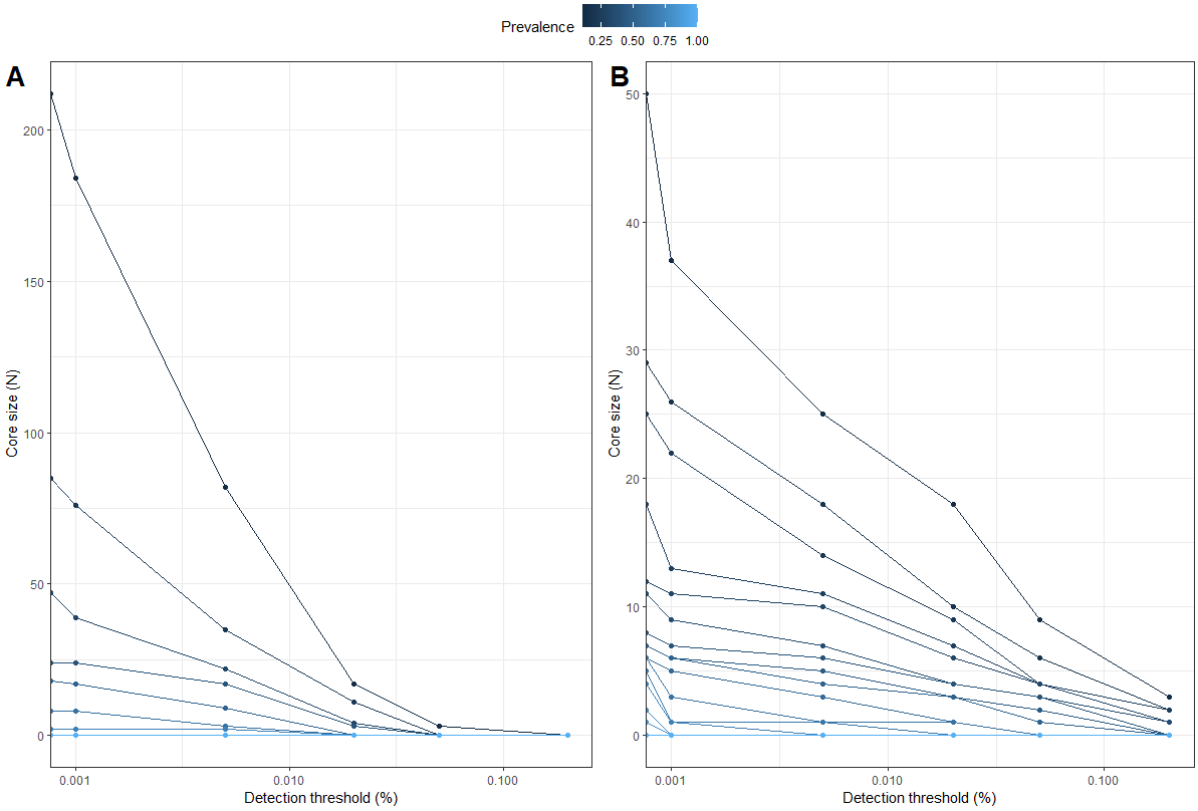
classification. Standardised methods for quality control and pre-processing of microbial data should be prioritised, to ensure reproducibility and reliability of such investigations.

Moreover, the sample sizes between some of the groups were uneven, rendering statistical analysis difficult. For example, when analysing how IM composition changes with age, the "early adult" group comprised only three out of the 40 participants, which affected the number of differentially abundant microbes identified by ANCOM-BC. This can be seen by removing the "early adult" group, since the number of differentially abundant microbes decreased drastically. The "travel" and "malaria medication" groups also had uneven sample sizes, so it is expected that the statistical analyses pertaining to those groups are also somewhat inaccurate.

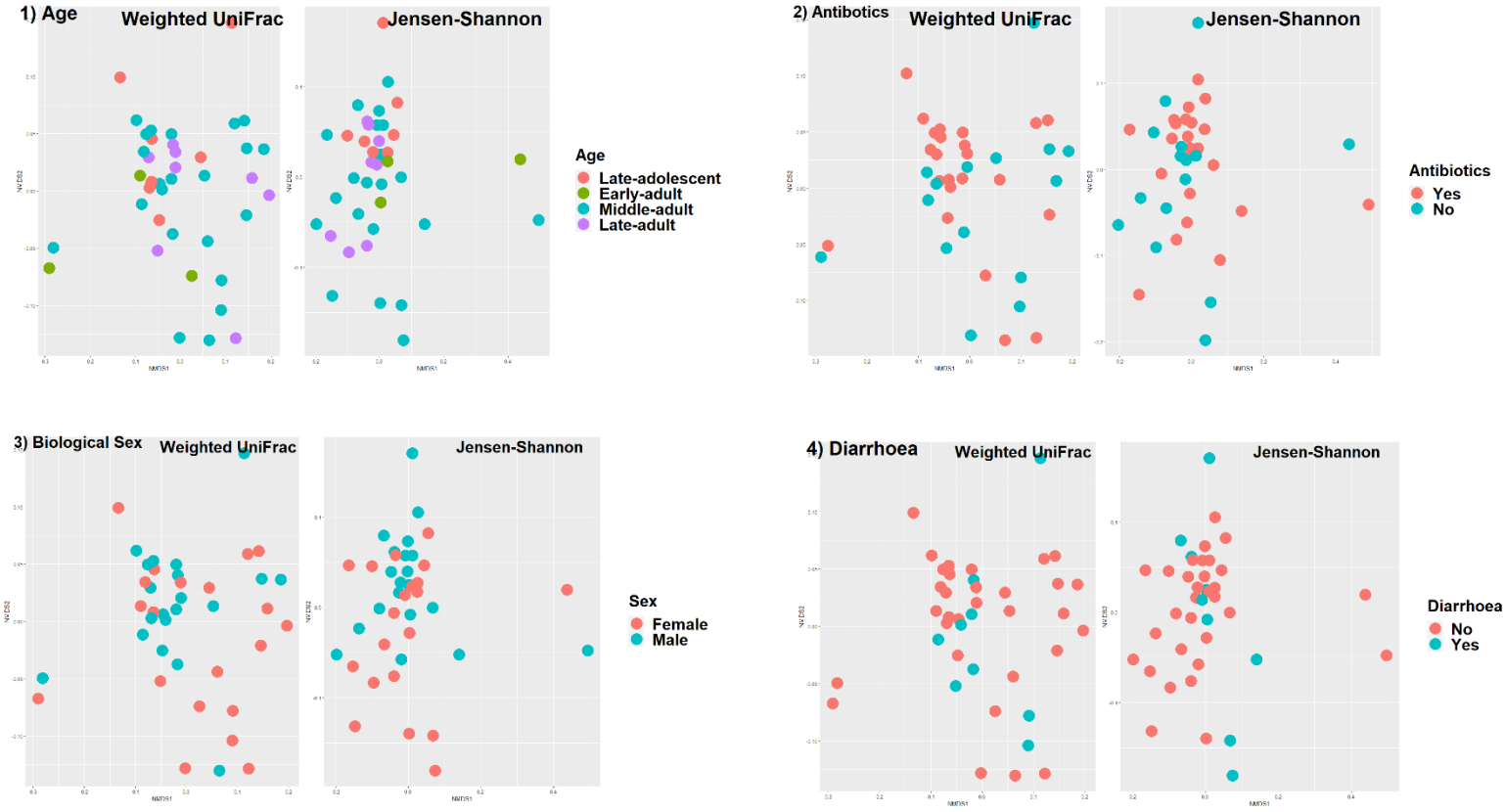
The use of controls and the removal of contaminant reads in this study is in line with recommended best practices for microbial studies. While many research projects make use of controls, few actually remove the identified contamination prior to statistical analysis. The removal of controls based on the statistical probability of a read being a contaminant is strongly advised, as contamination is a big source of confoundment for microbial research investigations.

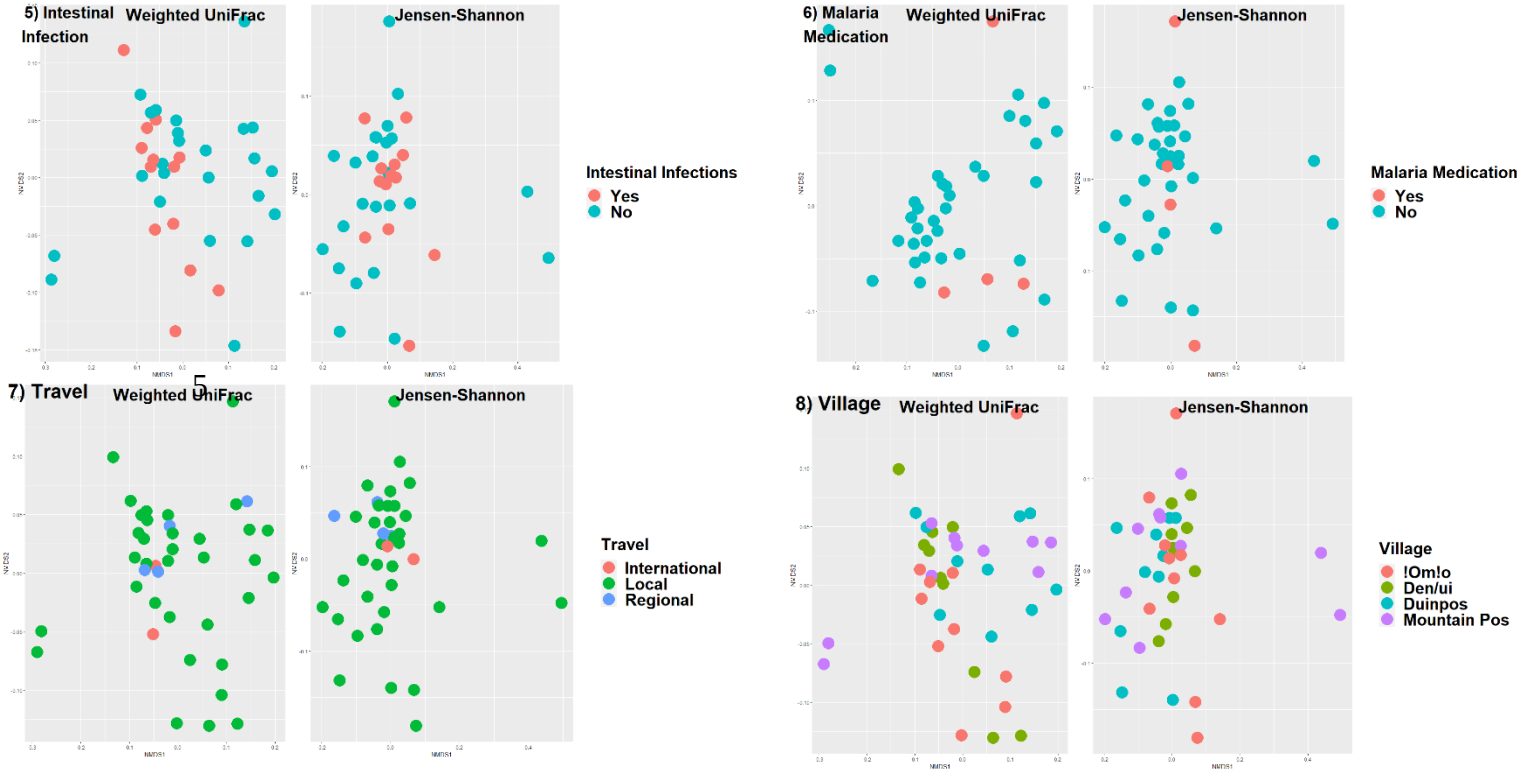
In future, the evolution of the Ju|'hoansi IM during transition from a traditional to a Western lifestyle will be documented. This is predicted to include a taxonomic shift in response to a shift in diet, as well as the disappearance of some microbes associated with traditional IMs, such as *Treponema*. The effect 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant will continue to be studied and will hopefully elucidate the effect of Westernisation on IM composition in terms of the abovementioned factors.

Supplementary Information

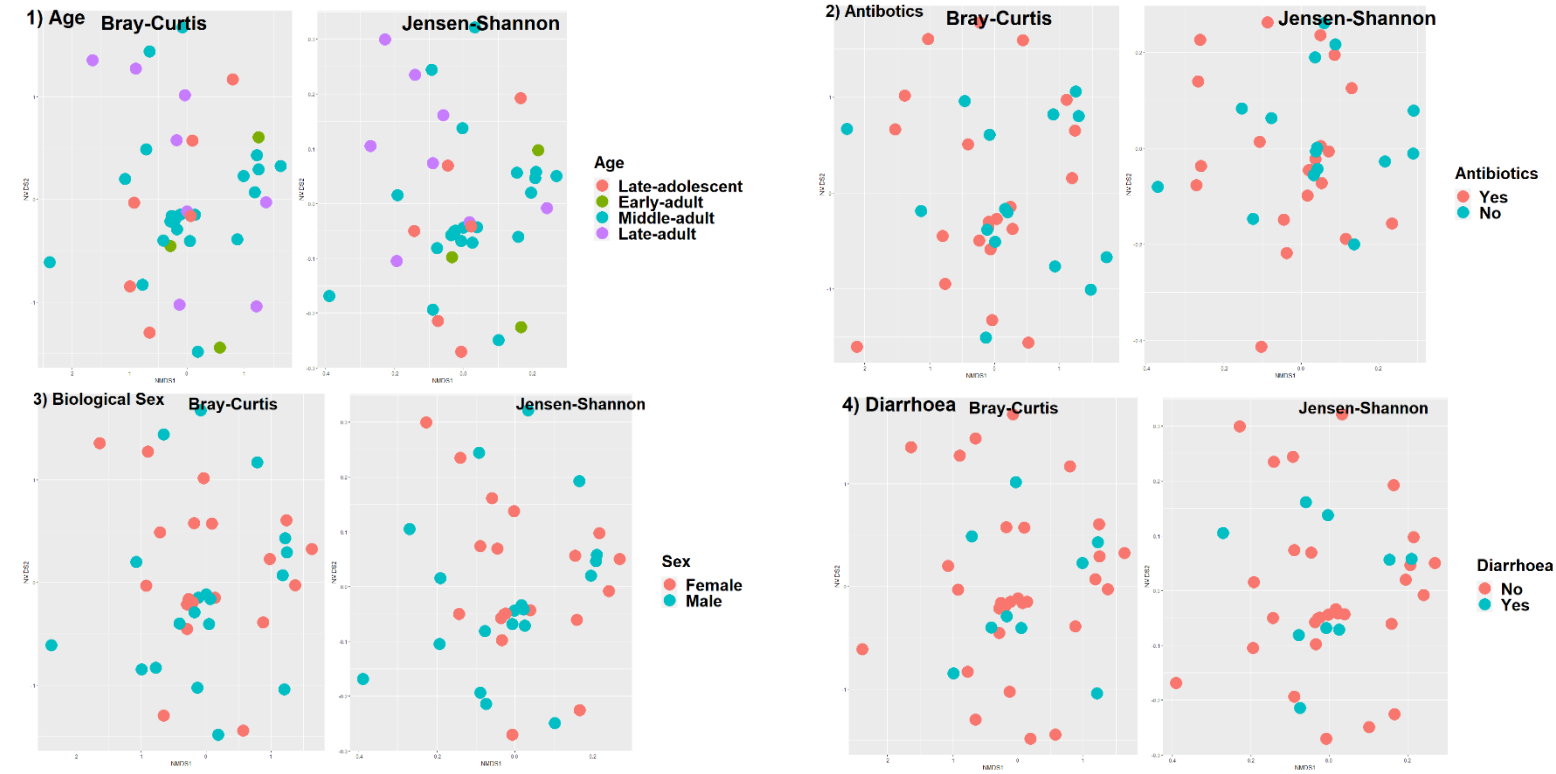


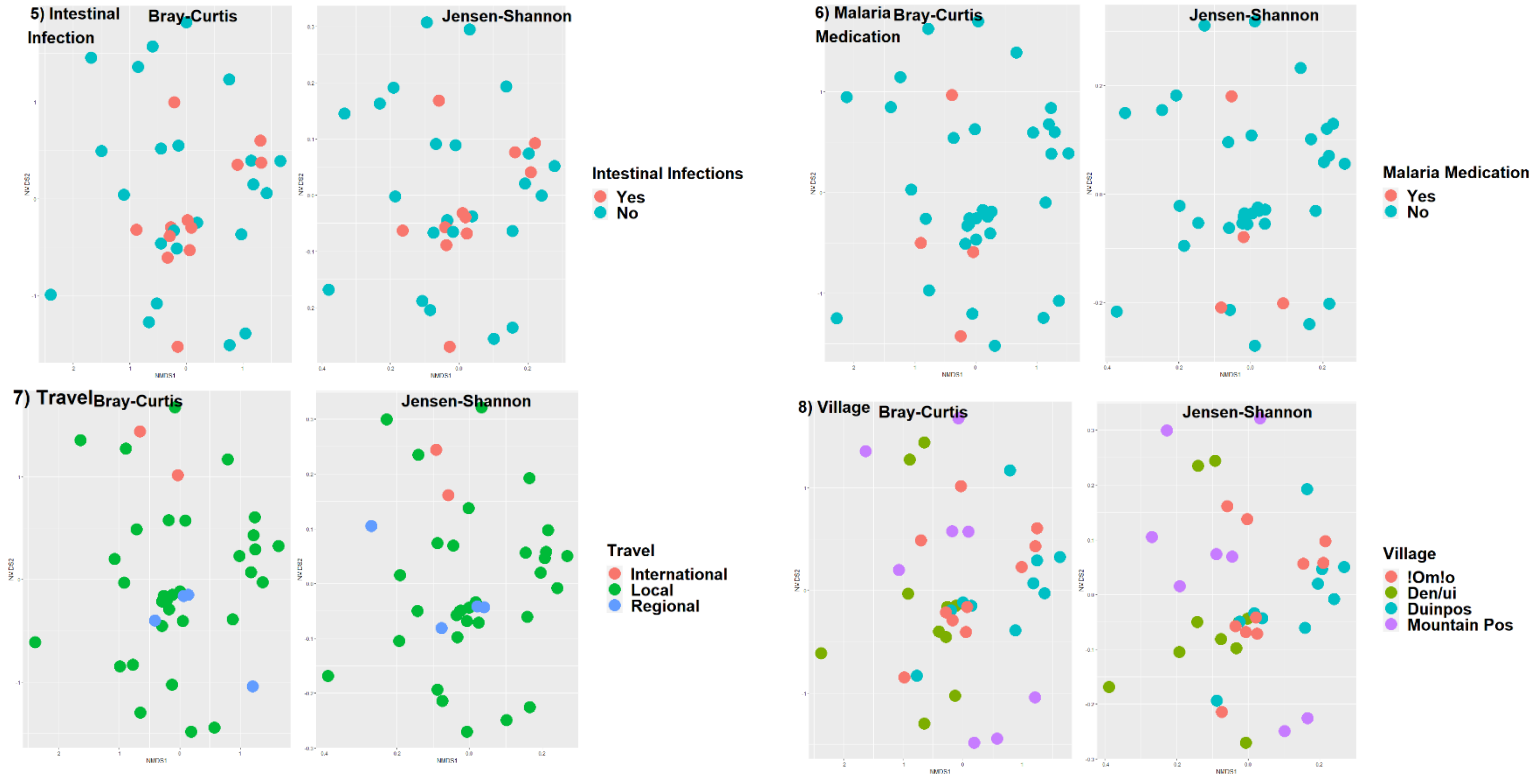
SI Figure 1: A) *Bacterial* ***and B)*** *fungal core IM size at various prevalence and detection thresholds. Thresholds were chosen such that they had the least effect on core size. Areas of the graph where the lines are parallel were thus considered. The bacterial core microbiome was therefore elucidated at 50% prevalence, and the fungal core microbiome at 60% prevalence. Both were elucidated at a detection threshold of 0.008%.*





SI Figure 2: *Bacterial NMDS plots showing β-diversity between subjects in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant.*





SI Figure 3: *Fungal NMDS plots showing β-diversity between subjects in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant*

SI Table 1: *Metadata collected for this study.*

Sampleid	Sampling date	Village	Sex	Age	Travel	Diarrhoea	Intestinal-Infections	Malaria	Malaria-Medication	TB	Antibiotics	Sample_or_Control
TDP01	22/07/2019	Duinpos	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDP02	22/07/2019	Duinpos	Female	Middle-adult	Regional	No	No	No	No	Yes	Yes	True Sample
TDP03	22/07/2019	Duinpos	Male	Late-adult	Local	No	Yes	No	No	No	No	True Sample
TDP04	22/07/2019	Duinpos	Female	Middle-adult	Local	No	No	No	Uncertain	No	Uncertain	True Sample
TDP05	22/07/2019	Duinpos	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDP06	22/07/2019	Duinpos	Female	Late-adult	Local	No	No	No	No	No	Uncertain	True Sample
TDP07	22/07/2019	Duinpos	Male	Late-adolescent	Local	No	No	No	No	Yes	Yes	True Sample
TDP08	22/07/2019	Duinpos	Female	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDP09	22/07/2019	Duinpos	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDP10	22/07/2019	Duinpos	Female	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDE01	23/07/2019	Den/ui	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDE02	23/07/2019	Den/ui	Female	Early-adult	Local	No	Yes	Yes	Yes	Yes	Yes	True Sample
TDE03	23/07/2019	Den/ui	Male	Middle-adult	Local	No	Uncertain	No	No	Yes	Yes	True Sample
TDE04	23/07/2019	Den/ui	Female	Late-adult	Local	No	No	No	No	Yes	Yes	True Sample
TDE05	23/07/2019	Den/ui	Male	Middle-adult	International	No	No	No	No	Yes	Yes	True Sample
TDE06	23/07/2019	Den/ui	Female	Late-adolescent	Local	No	Yes	No	No	Yes	Yes	True Sample
TDE07	23/07/2019	Den/ui	Male	Middle-adult	Regional	Yes	No	No	No	Yes	Yes	True Sample
TDE08	23/07/2019	Den/ui	Male	Late-adult	Local	No	Uncertain	No	No	Yes	Yes	True Sample
TDE09	23/07/2019	Den/ui	Female	Late-adolescent	Local	No	Yes	No	No	Yes	Yes	True Sample
TDE10	23/07/2019	Den/ui	Female	Middle-adult	Local	No	Yes	No	No	Yes	Yes	True Sample
TMP01	24/07/2019	Mountain Pos	Male	Middle-adult	Local	No	No	No	No	No	No	True Sample
TMP02	24/07/2019	Mountain Pos	Female	Late-adolescent	Local	No	No	No	No	No	No	True Sample
TMP03	24/07/2019	Mountain Pos	Male	Late-adult	Regional	Yes	No	No	No	Yes	Yes	True Sample
TMP04	24/07/2019	Mountain Pos	Female	Late-adult	Local	No	No	No	No	Yes	Yes	True Sample
TMP05	24/07/2019	Mountain Pos	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TMP06	24/07/2019	Mountain Pos	Female	Late-adult	Local	No	No	No	No	No	No	True Sample
TMP07	24/07/2019	Mountain Pos	Male	Middle-adult	Local	No	No	No	No	Yes	Yes	True Sample
TMP08	24/07/2019	Mountain Pos	Female	Late-adolescent	Local	No	Yes	No	No	Yes	Yes	True Sample
TMP09	24/07/2019	Mountain Pos	Male	Middle-adult	Local	No	No	No	No	No	No	True Sample
TMP10	24/07/2019	Mountain Pos	Female	Early-adult	Local	No	No	No	No	No	No	True Sample
TOM01	25/07/2019	!Om!o	Female	Middle-adult	Local	Yes	Uncertain	Yes	Yes	No	No	True Sample
TOM02	25/07/2019	!Om!o	Male	Late-adolescent	Local	Yes	No	Yes	Yes	No	No	True Sample
TOM03	25/07/2019	!Om!o	Female	Late-adult	International	Yes	Yes	Yes	Yes	No	No	True Sample
TOM04	25/07/2019	!Om!o	Male	Middle-adult	Local	Yes	Yes	No	No	Yes	Yes	True Sample
TOM05	25/07/2019	!Om!o	Female	Middle-adult	Local	No	Yes	No	No	No	No	True Sample
TOM06	25/07/2019	!Om!o	Male	Middle-adult	Local	Yes	No	No	No	No	No	True Sample
TOM07	25/07/2019	!Om!o	Female	Middle-adult	Local	Yes	Yes	No	No	No	No	True Sample
TOM08	25/07/2019	!Om!o	Male	Late-adolescent	Regional	No	Yes	No	No	No	No	True Sample
TOM09	25/07/2019	!Om!o	Female	Early-adult	Local	No	Yes	No	No	No	No	True Sample
TOM10	25/07/2019	!Om!o	Male	Middle-adult	Local	Yes	Yes	No	No	No	No	True Sample
CON-CTRL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Control Sample
KIT-CTRL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Control Sample

SI Table 2: *Clustering was performed on the fungal ITS reads at 90% and 98% similarity using open- or closed-reference clustering with QIIME's vsearch plug-in. Open-reference clustering clusters reads against a reference database at a user-defined similarity threshold. Reads that do not match are then clustered de novo. Closed-reference clustering clusters reads against a user-defined similarity threshold. Reads that do not match are then discarded. Taxonomic classification was compared across clustering groups. 98% closed-reference clustering allowed the identification of the most reads.*

Taxonomic level	Amplicon sequence variants	98% closed-reference clustering	98% open-reference clustering	90% closed-reference clustering	90% open-reference clustering
Order	56%	96%	47%	94%	52%
Family	55%	92%	43%	87%	47%
Genus	50%	83%	39%	77%	43%
Species	42%	66%	31%	62%	34%

SI Table 3.1*: The Ju|'hoansi IM comprised 13 identified bacterial phyla, of which Firmicutes and Bacteroidota represent the majority.*

Phylum	Relative Abundance
Firmicutes	26.62028
Bacteroidota	8.45785
Proteobacteria	3.16558
Spirochaetota	0.28313
Actinobacteriota	0.24214
Desulfobacterota	0.0836
Elusimicrobiota	0.05528
Verrucomicrobiota	0.04495
Fusobacteriota	0.0144
WPS-2	0.01042
Campilobacterota	0.0099
Cyanobacteria	0.00683
Synergistota	0.00563

***SI Table 3.2**: The Ju|'hoansi IM harboured 18 bacterial classes.*

Class	**Relative Abundance**
Clostridia	23.14241
Bacteroidia	8.45785
Gammaproteobacteria	2.89622
Bacilli	2.16553
Negativicutes	1.31234
Spirochaetia	0.28313
Alphaproteobacteria	0.26937
Actinobacteria	0.15465
Coriobacteriia	0.0875
Desulfovibrionia	0.0836
Elusimicrobia	0.05528
Lentisphaeria	0.03668
Fusobacteriia	0.0144
WPS-2	0.01042
Campylobacteria	0.0099
Verrucomicrobiae	0.00827
Vampirivibrionia	0.00683
Synergistia	0.00563

SI Table 3.3: *Thirty-seven bacterial orders were classified in the Ju|'hoansi IM.*

Order	**Relative Abundance**
Lachnospirales	11.2506
Bacteroidales	8.45785
Oscillospirales	7.88274
Aeromonadales	2.78278
Christensenellales	1.5066
Lactobacillales	1.24688
Clostridiales	1.05027
Acidaminococcales	1.022
Erysipelotrichales	0.86541
Peptostreptococcales-Tissierellales	0.71767
Clostridia_vadinBB60_group	0.41722
Veillonellales-Selenomonadales	0.29034
Spirochaetales	0.28313
Rhizobiales	0.25115
Clostridia_UCG-014	0.2057
Micrococcales	0.15465
Coriobacteriales	0.0875
Desulfovibrionales	0.0836
Monoglobales	0.07082
Burkholderiales	0.06133
Elusimicrobiales	0.05528
Enterobacterales	0.05211
Peptococcales	0.04079
Oligosphaerales	0.02538
Izemoplasmatales	0.02038
Rhodospirillales	0.01822
RF39	0.01587
Staphylococcales	0.01462
Fusobacteriales	0.0144
Victivallales	0.0113
WPS-2	0.01042
Campylobacterales	0.0099
Gastranaerophilales	0.00683
Opitutales	0.00577
Synergistales	0.00563
Verrucomicrobiales	0.00249
Acholeplasmatales	0.00236

SI Table 3.4: *The Ju|'hoansi IM harboured 61 bacterial families.*

Family	Relative Abundance
Lachnospiraceae	11.2506
Prevotellaceae	7.09186
Ruminococcaceae	4.47111
Succinivibrionaceae	2.78278
Oscillospiraceae	2.48284
Christensenellaceae	1.5066
Streptococcaceae	1.14909
Clostridiaceae	1.05027
Acidaminococcaceae	1.022
Erysipelotrichaceae	0.64852
Peptostreptococcaceae	0.58953
Bacteroidaceae	0.54071
[Eubacterium]_coprostanoligenes_group	0.52825
Clostridia_vadinBB60_group	0.41722
Rikenellaceae	0.38075
Spirochaetaceae	0.28313
UCG-010	0.26985
Muribaculaceae	0.25487
Xanthobacteraceae	0.25115
Selenomonadaceae	0.2328
Erysipelatoclostridiaceae	0.21689
Clostridia_UCG-014	0.2057
Micrococcaceae	0.15465
Butyricicoccaceae	0.13069
Anaerovoracaceae	0.1269
p-2534-18B5_gut_group	0.12239
Desulfovibrionaceae	0.0836
Monoglobaceae	0.07082
Lactobacillaceae	0.06534
Veillonellaceae	0.05755
Elusimicrobiaceae	0.05528
Enterobacteriaceae	0.05211
Coriobacteriaceae	0.05093
Peptococcaceae	0.04079
Sutterellaceae	0.03602
Porphyromonadaceae	0.03036
Eggerthellaceae	0.0276
Oligosphaeraceae	0.02538
Neisseriaceae	0.02531
Carnobacteriaceae	0.02237
Izemoplasmatales	0.02038
uncultured	0.01986
Tannerellaceae	0.01965
RF39	0.01587
Gemellaceae	0.01462
Fusobacteriaceae	0.0144
vadinBE97	0.0113
WPS-2	0.01042
Enterococcaceae	0.01008
Campylobacteraceae	0.0099
Marinifilaceae	0.00753
Gastranaerophilales	0.00683
Atopobiaceae	0.00614
Puniceicoccaceae	0.00577
Synergistaceae	0.00563
Barnesiellaceae	0.00525
Bacteroidales_RF16_group	0.00285
Coriobacteriales_Incertae_Sedis	0.00283
Akkermansiaceae	0.00249
Acholeplasmataceae	0.00236
Peptostreptococcales-Tissierellales	0.00124

SI Table 3.5: *The Ju|'hoansi IM contained 131 bacterial genera.*

Genus	Relative Abundance	Genus cont.	Relative Abundance cont.
Prevotella	6.61396	*CAG-873*	0.05912
Blautia	3.46243	*Elusimicrobium*	0.05528
Faecalibacterium	2.05266	*Veillonella*	0.05485
Succinivibrio	1.75435	*Escherichia-Shigella*	0.05211
Christensenellaceae_R-7_group	1.5066	*Collinsella*	0.05093
UCG-002	1.24138	*Lachnospiraceae_UCG-010*	0.04523
Streptococcus	1.14909	*Paeniclostridium*	0.04471
Ruminobacter	1.02844	*Colidextribacter*	0.04369
Phascolarctobacterium	1.022	*[Eubacterium]_ruminantium_group*	0.04204
[Ruminococcus]_torques_group	1.01007	*Sutterella*	0.03602
Clostridium_sensu_stricto_1	0.84804	*Prevotellaceae_UCG-003*	0.03512
Ruminococcus	0.7926	*Candidatus_Soleaferrea*	0.03433
Lachnospiraceae_NK4A136_group	0.78222	*[Eubacterium]_xylanophilum_group*	0.03234
Agathobacter	0.75194	*Lachnospiraceae_FCS020_group*	0.03079
UCG-005	0.6757	*Prevotellaceae_NK3B31_group*	0.03076
Coprococcus	0.63911	*Lachnospiraceae_NK4B4_group*	0.03047
Bacteroides	0.54071	*Porphyromonas*	0.03036
[Eubacterium]_coprostanoligenes_group	0.52825	*Peptococcus*	0.02938
Romboutsia	0.5018	*Ruminococcaceae*	0.02605
uncultured	0.49498	*Z20*	0.02538
Dorea	0.4645	*Neisseria*	0.02531
Roseburia	0.4402	*Faecalitalea*	0.02441
Clostridia_vadinBB60_group	0.41722	*Granulicatella*	0.02237
[Eubacterium]_hallii_group	0.39507	*Fournierella*	0.02209
Rikenellaceae_RC9_gut_group	0.37034	*Intestinibacter*	0.02149
Marvinbryantia	0.3509	*Incertae_Sedis*	0.02131
Holdemanella	0.35047	*Erysipelotrichaceae_UCG-003*	0.02117
Fusicatenibacter	0.3211	*Hungatella*	0.02096
Subdoligranulum	0.30717	*Izemoplasmatales*	0.02038
Treponema	0.28313	*Parabacteroides*	0.01965
Alloprevotella	0.27727	*Terrisporobacter*	0.01895
NK4A214_group	0.27432	*Lachnospiraceae_AC2044_group*	0.01776
UCG-010	0.26985	*Lachnospira*	0.01768
CAG-56	0.23935	*Lachnospiraceae_NK3A20_group*	0.01602
Anaerovibrio	0.2328	*RF39*	0.01587
Lachnospiraceae_ND3007_group	0.20922	*Paraprevotella*	0.01498
Clostridia_UCG-014	0.2057	*Gemella*	0.01462
Muribaculaceae	0.19576	*Senegalimassilia*	0.01449
Catenibacterium	0.19478	*Fusobacterium*	0.0144
Sarcina	0.18876	*Clostridium_sensu_stricto_6*	0.01347
Butyrivibrio	0.17334	*[Eubacterium]_ventriosum_group*	0.01296
Anaerostipes	0.16943	*Family_XIII_UCG-001*	0.01289
Turicibacter	0.16027	*Mogibacterium*	0.01256
CAG-352	0.15552	*vadinBE97*	0.0113
Rothia	0.15465	*WPS-2*	0.01042
Negativibacillus	0.14	*Alistipes*	0.01041
[Ruminococcus]_gauvreauii_group	0.13336	*Enterococcus*	0.01008
UCG-003	0.13183	*Campylobacter*	0.0099
p-2534-18B5_gut_group	0.12239	*UBA1819*	0.00769
Lachnospiraceae_UCG-007	0.12001	*Odoribacter*	0.00753
[Eubacterium]_siraeum_group	0.11442	*Gastranaerophilales*	0.00683
Lachnoclostridium	0.11128	*Coriobacteriaceae_UCG-003*	0.00614
[Ruminococcus]_gnavus_group	0.10992	*Cerasicoccus*	0.00577
Oribacterium	0.10037	*Cloacibacillus*	0.00563
Family_XIII_AD3011_group	0.09794	*Enterorhabdus*	0.00529
Butyricicoccus	0.08459	*Barnesiella*	0.00525
Desulfovibrio	0.07337	*[Eubacterium]_nodatum_group*	0.00352
Monoglobus	0.07082	*[Eubacterium]_eligens_group*	0.00311
Tyzzerella	0.06993	*Bacteroidales_RF16_group*	0.00285
Prevotellaceae_UCG-001	0.06716	*Eisenbergiella*	0.00261
Solobacterium	0.06543	*Peptostreptococcus*	0.00258
Lactobacillus	0.06534	*Akkermansia*	0.00249
Bradyrhizobium	0.06009	*Anaeroplasma*	0.00236
		Slackia	0.00181
		Dialister	0.00135
		Allisonella	0.00134
		Finegoldia	0.00124
		UCG-004	0.00093

***SI Table 3.6**: Seventy-seven bacterial species were present in the Ju|'hoansi IM.*

Species	**Relative Abundance**	**Species cont.**	**Relative Abundance cont.**
uncultured_bacterium	4.31731	*Bacteroides_plebeius*	0.03309
gut_metagenome	2.22252	*uncultured_Blautia*	0.03297
uncultured_organism	1.5766	*uncultured_Roseburia*	0.03291
metagenome	1.44977	*Bacteroides_intestinalis*	0.03223
Prevotella_copri	0.70065	*uncultured_Eubacterium*	0.03076
Streptococcus_salivarius	0.61974	*Ruminococcus_flavefaciens*	0.03027
human_gut	0.42827	*Lactobacillus_ruminis*	0.03014
Prevotellaceae_bacterium	0.35476	*Bacteroides_massiliensis*	0.02767
Prevotella_melaninogenica	0.32349	*Roseburia_sp.*	0.02393
uncultured_rumen	0.28455	*Prevotella_histicola*	0.02371
uncultured_spirochete	0.25562	*Roseburia_hominis*	0.02351
Ruminococcus_champanellensis	0.19577	*bacterium_YE57*	0.02292
uncultured_Porphyromonadaceae	0.18417	*Bacteroides_dorei*	0.02077
uncultured_Ruminococcaceae	0.15067	*uncultured_Faecalibacterium*	0.01826
Dorea_formicigenerans	0.14641	*Eubacterium_coprostanoligenes*	0.01558
Treponema_succinifaciens	0.13291	*Bacteroides_uniformis*	0.01525
Butyrivibrio_crossotus	0.1236	*Fusobacterium_mortiferum*	0.0144
uncultured_prokaryote	0.12013	*Bacteroides_thetaiotaomicron*	0.01398
uncultured_Wautersiella	0.10746	*Clostridium_bornimense*	0.01347
[Eubacterium]_siraeum	0.10345	*unidentified*	0.01207
Bacteroides_vulgatus	0.09648	*uncultured_Desulfovibrionaceae*	0.01023
Prevotella_stercorea	0.08877	*Campylobacter_upsaliensis*	0.0099
Desulfovibrio_piger	0.07337	*Ruminococcus_sp.*	0.00896
uncultured_Lachnospiraceae	0.07294	*Catenibacterium_mitsuokai*	0.00592
Ruminococcus_bicirculans	0.07209	*bacterium_enrichment*	0.00556
Ruminococcus_torques	0.07159	*Odoribacter_splanchnicus*	0.00487
Bacteroides_ovatus	0.0698	*Treponema_berlinense*	0.00383
Ruminococcus_lactaris	0.06721	*Bacteroides_cellulosilyticus*	0.00375
Bacteroidaceae_bacterium	0.06402	*Eubacterium_sulci*	0.00352
Coprococcus_eutactus	0.05808	*uncultured_Clostridiales*	0.00327
Massiliprevotella_massiliensis	0.05613	*Parabacteroides_merdae*	0.00314
Clostridium_perfringens	0.05582	*uncultured_Ruminococcus*	0.00269
Bacteroides_sp.	0.0557	*Akkermansia_muciniphila*	0.00249
Ruminococcus_callidus	0.05384	*Slackia_isoflavoniconvertens*	0.00181
Bacteroides_fragilis	0.04495	*Sutterella_wadsworthensis*	0.00154
uncultured_Clostridium	0.04319	*Alistipes_shahii*	0.00076
uncultured_Clostridia	0.03858	*uncultured_marine*	0.00052
Eubacterium_ramulus	0.03639	*uncultured_beta*	0.00029
Lactobacillus_mucosae	0.03521	*Parabacteroides_distasonis*	0.00025

SI Table 4.1: *Four fungal phyla inhabited the Ju|'hoansi IM.*

Phylum	**Relative Abundance**
Ascomycota	20.28184
Basidiomycota	18.70786
Chytridiomycota	0.0074
Mucoromycota	0.00058

SI Table 4.2: *Twelve fungal classes were identified in the Ju|'hoansi IM.*

Class	**Relative Abundance**
Malasseziomycetes	9.52235
Saccharomycetes	9.01533
Dothideomycetes	6.53822
Tremellomycetes	4.83069
Agaricomycetes	3.88144
Eurotiomycetes	2.31786
Sordariomycetes	2.24014
Pezizomycetes	0.16889
Agaricostilbomycetes	0.01571
unidentified	0.00493
Cystobasidiomycetes	0.00088
Mucoromycetes	0.00058

SI Table 4.3: *There were 26 fungal orders in the Ju|'hoansi IM.*

Order	**Relative Abundance**
Malasseziales	9.52235
Saccharomycetales	9.01533
Filobasidiales	4.28792
Pleosporales	3.66531
Capnodiales	2.81038
Eurotiales	2.24091
Agaricales	2.11903
Hypocreales	1.48718
Russulales	0.79681
Polyporales	0.55373
Tremellales	0.54277
Xylariales	0.30466
Cantharellales	0.27254
Microascales	0.17029
Pezizales	0.16889
Diaporthales	0.14991
Hymenochaetales	0.13582
Chaetothyriales	0.07694
Sordariales	0.0602
Dothideales	0.05792
Agaricostilbales	0.01571
unidentified	0.00493
Trechisporales	0.00351
Cystobasidiomycetes_ord_Incertae_sedis	0.00088
Mucorales	0.00058
Botryosphaeriales	0.00035

SI Table 4.4: *There were 63 fungal families in the Ju|'hoansi IM.*

Family	Relative Abundance	Family cont.	Relative Abundance cont.
Malasseziaceae	9.52235	Didymosphaeriaceae	0.03365
Saccharomycetales_fam_Incertae_sedis	7.14983	Zopfiaceae	0.02935
Filobasidiaceae	4.28792	Dipodascaceae	0.02717
Cladosporiaceae	2.81038	Chaetomiaceae	0.02605
Didymellaceae	2.64559	Lasiosphaeriaceae	0.01738
Aspergillaceae	2.12285	Sordariaceae	0.01677
Tricholomataceae	2.10318	Massarinaceae	0.01622
Saccharomycetaceae	1.63478	Bionectriaceae	0.01613
Nectriaceae	1.45071	Agaricostilbaceae	0.01571
Stereaceae	0.79681	Pleosporales_fam_Incertae_sedis	0.01453
Bulleribasidiaceae	0.43965	Cordycipitaceae	0.01429
Sporormiaceae	0.39553	Trichomeriaceae	0.00698
Fomitopsidaceae	0.31471	Halosphaeriaceae	0.00569
Pleosporaceae	0.28202	Pichiaceae	0.00527
Botryobasidiaceae	0.27254	Psathyrellaceae	0.00497
Xylariaceae	0.26484	unidentified	0.00493
Phaeosphaeriaceae	0.18516	Hypocreaceae	0.00475
Pyronemataceae	0.16889	Dothideales_fam_Incertae_sedis	0.00401
Microascaceae	0.1646	Fistulinaceae	0.00394
Valsaceae	0.14991	Schizophyllaceae	0.00393
Schizoporaceae	0.1349	Hydnodontaceae	0.00351
Debaryomycetaceae	0.12529	Strophariaceae	0.00301
Trichocomaceae	0.11806	Periconiaceae	0.0021
Meruliaceae	0.10409	Irpicaceae	0.00201
Polyporaceae	0.08001	Stachybotryaceae	0.00129
Phaffomycetaceae	0.07265	Hymenochaetales_fam_Incertae_sedis	0.00092
Herpotrichiellaceae	0.06996	Symmetrosporaceae	0.00088
Trimorphomycetaceae	0.06341	Microdochiaceae	0.00084
Phanerochaetaceae	0.05291	Rhizopodaceae	0.00058
Aureobasidiaceae	0.04384	Botryosphaeriaceae	0.00035
Sporocadaceae	0.03844	Metschnikowiaceae	0.00035
		Xylariales_fam_Incertae_sedis	1.00E-04

SI Table 4.5: *Eighty-one fungal genera inhabited the Ju|'hoansi IM.*

Genus	**Relative Abundance**	**Genus cont.**	**Relative Abundance cont.**
Malassezia	7.45879	*Chaetomium*	0.02605
Candida	7.14983	*Sclerostagonospora*	0.02452
Naganishia	4.28792	*Exserohilum*	0.02282
Cladosporium	2.81038	*Preussia*	0.01917
unidentified	2.06849	*Wojnowiciella*	0.01768
Panellus	1.66334	*Stagonospora*	0.01622
Issatchenkia	1.33052	*Clonostachys*	0.01613
Aspergillus	1.24564	*Sterigmatomyces*	0.01571
Fusarium	1.16107	*Triangularia*	0.01483
Penicillium	0.87721	*Leptobacillium*	0.01229
Stereum	0.79681	*Veronaea*	0.01046
Mycena	0.43984	*Pseudorobillarda*	0.00972
Vishniacozyma	0.43965	*Cumuliphoma*	0.00722
Westerdykella	0.37636	*Knufia*	0.00698
Neoascochyta	0.36568	*Alternaria*	0.0068
Amyloporia	0.31471	*Cirrenalia*	0.00569
Saccharomyces	0.30425	*Pichia*	0.00527
Botryobasidium	0.27254	*Gibberella*	0.00511
Curvularia	0.25241	*Coprinellus*	0.00497
Epicoccum	0.24503	*Pleiochaeta*	0.00482
Hypoxylon	0.18694	*Trichoderma*	0.00475
Lasiobolidium	0.16889	*Hortaea*	0.00401
Petriellopsis	0.1646	*Neurospora*	0.00396
Cryptosphaeria	0.14991	*Porodisculus*	0.00394
Didymella	0.14978	*Schizophyllum*	0.00393
Phaeosphaeria	0.13703	*Subulicystidium*	0.00351
Hyphodontia	0.1349	*Hypholoma*	0.00301
Meyerozyma	0.12529	*Phlebia*	0.00264
Talaromyces	0.11806	*Podospora*	0.00256
Mycoacia	0.10144	*Periconia*	0.0021
Trametes	0.08001	*Irpex*	0.00201
Cyberlindnera	0.07265	*Lecanicillium*	0.00201
Hypocopra	0.06386	*Dipodascus*	0.00146
Saitozyma	0.06341	*Striaticonidium*	0.00129
Exophiala	0.0595	*Trichaptum*	0.00092
Phanerochaete	0.05291	*Symmetrospora*	0.00088
Hymenopleella	0.03844	*Idriella*	0.00084
Paraconiothyrium	0.03365	*Rhizopus*	0.00058
Didymocrea	0.02935	*Macrophomina*	0.00035
Aureobasidium	0.02661	*Kodamaea*	0.00035

SI Table 4.6: *There were 87 fungal species residing in the Ju|'hoansi IM.*

Species	**Relative Abundance**	**Species cont.**	**Relative Abundance cont.**
Malassezia_restricta	7.45025	*Sterigmatomyces_halophilus*	0.01571
Candida_albicans	6.83224	*Westerdykella_nigra*	0.01537
Cladosporium_delicatulum	2.56632	*Triangularia_mangenotii*	0.01483
unidentified	2.14754	*Aspergillus_penicillioides*	0.01446
Panellus_serotinus	1.66334	*Westerdykella_ornata*	0.01303
Issatchenkia_orientalis	1.33052	*Leptobacillium_leptobactrum*	0.01229
Penicillium_penicillioides	0.74403	*Veronaea_compacta*	0.01046
Fusarium_oxysporum	0.67325	*Pseudorobillarda_phragmitis*	0.00972
Mycena_renati	0.43984	*Phaeosphaeria_caricis*	0.00939
Vishniacozyma_globispora	0.43965	*Fusarium_delphinoides*	0.00786
Westerdykella_centenaria	0.34758	*Malassezia_globosa*	0.00749
Aspergillus_conicus	0.32281	*Cumuliphoma_indica*	0.00722
Candida_tropicalis	0.31759	*Alternaria_tenuissima*	0.0068
Amyloporia_sinuosa	0.31471	*Epicoccum_dendrobii*	0.00564
Saccharomyces_cerevisiae	0.30411	*Pichia_mandshurica*	0.00527
Botryobasidium_candicans	0.27254	*Gibberella_intricans*	0.00511
Curvularia_hawaiiensis	0.25241	*Coprinellus_curtus*	0.00497
Cladosporium_langeronii	0.24143	*Pleiochaeta_carotae*	0.00482
Lasiobolidium_orbiculoides	0.16889	*Hortaea_werneckii*	0.00401
Petriellopsis_africana	0.1646	*Porodisculus_pendulus*	0.00394
Cryptosphaeria_subcutanea	0.14991	*Schizophyllum_commune*	0.00393
Hyphodontia_pallidula	0.1349	*Aspergillus_wentii*	0.00337
Penicillium_bialowiezense	0.13253	*Hypholoma_fasciculare*	0.00301
Meyerozyma_guilliermondii	0.12529	*Preussia_terricola*	0.00299
Hypoxylon_griseobrunneum	0.11033	*Phlebia_acerina*	0.00264
Mycoacia_fuscoatra	0.10144	*Podospora_prethopodalis*	0.00256
Hypoxylon_macrocarpum	0.07661	*Irpex_lacteus*	0.00201
Trametes_versicolor	0.07534	*Dipodascus_geotrichum*	0.00146
Cyberlindnera_fabianii	0.07265	*Striaticonidium_synnematum*	0.00116
Fusarium_brachygibbosum	0.06401	*Malassezia_arunalokei*	0.00105
Hypocopra_rostrata	0.06386	*Trichaptum_biforme*	0.00092
Saitozyma_paraflava	0.06341	*Symmetrospora_vermiculata*	0.00088
Phanerochaete_stereoides	0.05291	*Periconia_neobrittanica*	0.00083
Epicoccum_thailandicum	0.05226	*Aspergillus_lanosus*	0.00075
Hymenopleella_subcylindrica	0.03785	*Penicillium_kongii*	0.00065
Paraconiothyrium_archidendri	0.03365	*Periconia_macrospinosa*	0.00063
Didymocrea_sadasivanii	0.02935	*Hymenopleella_austroafricana*	0.00059
Aureobasidium_namibiae	0.02661	*Westerdykella_cylindrica*	0.00038
Chaetomium_atrobrunneum	0.02605	*Macrophomina_phaseolina*	0.00035
Aspergillus_flavus	0.0247	*Kodamaea_ohmeri*	0.00035
Sclerostagonospora_rosae	0.02452	*Rhizopus_caespitosus*	0.00029
Exserohilum_gedarefense	0.02282	*Rhizopus_microsporus*	0.00029
Stagonospora_pseudovitensis	0.01622	*Neoidriella_desertorum*	1.00E-04
Clonostachys_miodochialis	0.01613		

***SI Figure 5.1**: Kruskal Wallis tests for bacterial α-diversity shows no significant differences between subjects who have and have not used antibiotics, biological sex, subjects who do and do not experience diarrhoea, and those who do and do not experience intestinal infections*

	p-value
Antibiotics	0.8797
Biological sex	0.6652
Diarrhoea	0.2502
Intestinal Infection	0.475

***SI Figure 5.2:** Dunn's tests for bacterial α-diversity shows no significant differences between subjects of different ages. Benjamini Hochberg corrected p-values < 0.05 were considered significant.*

	Early adult	Late adolescent	Late adult
Late adolescent	0.6881		
Late adult	0.9415	0.4177	
Middle adult	0.3752	0.6901	0.7178

***Figure 5.3:** Dunn's tests for bacterial α-diversity shows no significant differences between subjects from different villages of primary residence. Benjamini Hochberg corrected p-values < 0.05 were considered significant.*

	!Om!o	Den/ui	Duinpos
Den/ui	1		
Duinpos	0.4771	1	
Mountain Pos	1	1	0.8185

***Figure 6.1**: Kruskal Wallis tests for fungal α-diversity shows no significant differences between subjects who have and have not used antibiotics, biological sex, subjects who do and do not experience diarrhoea, and those who do and do not experience intestinal infections*

	p-value
Antibiotics	0.4792
Biological sex	0.2672
Diarrhoea	0.6678
Intestinal Infection	0.1246

***Figure 6.2**: Dunn's tests for fungal α-diversity shows no significant differences between subjects of different ages. Benjamini Hochberg corrected p-values < 0.05 were considered significant.*

	Early adult	Late adolescent	Late adult
Late adolescent	0.9157		
Late adult	0.8742	0.3629	
Middle adult	0.9221	0.6435	0.8283

***Figure 6.3:** Dunn's tests for bacterial α-diversity shows no significant differences between subjects from different villages of primary residence. Benjamini Hochberg corrected p-values < 0.05 were considered significant*

	!Om!o	Den/ui	Duinpos
Den/ui	1		
Duinpos	1	0.488	
Mountain Pos	0.87	1	1

SI Table 7.1: *Weighted UniFrac PERMANOVA results regarding differences in bacterial community structure in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant*

Group	PERMANOVA F-statistic	PERMANOVA R-squared	PERMANOVA p-value	Dispersion F-statistic	Dispersion p-value	Significance?
1) Age	1.0563	0.0809	0.389	0.7907	0.507	Not significant
2) Antibiotics	1.2821	0.03439	0.237	1.7288	0.1969	Not significant
3) Biological Sex	1.5955	0.04029	0.143	2.9342	0.09487	Not significant
4) Diarrhoea	0.62772	0.01625	0.671	1.6422	0.2078	Not significant
5) Intestinal Infection	2.9554	0.07786	0.015*	6.1108	0.01844*	Significant, but assumptions not met
6) Malaria Medication	1.2583	0.03289	0.253	0.2199	0.6419	Not significant
7) Travel	0.41905	0.02215	0.958	2.7485	0.07711	Not significant
8) Village	1.7417	0.12675	0.001***	1.9091	0.00553	Significant

SI Table 7.2: *Jensen-Shannon PERMANOVA results regarding differences in bacterial community structure in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant*

Group	PERMANOVA F-stat	PERMANOVA R-squared	PERMANOVA P-value	Dispersion F-stat	Dispersion P-value	Significance?
1) Age	1.0649	0.08151	0.297	2.6823	0.06125	Not significant
2) Antibiotics	1.3034	0.03494	0.117	1.0708	0.3077	Not significant
3) Biological Sex	1.0886	0.02785	0.311	0.8382	0.3657	Not significant
4) Diarrhoea	1.2493	0.03183	0.168	0	0.9877	Not significant
5) Intestinal Infection	1.8691	0.05069	0.015*	5.4567	0.02535*	Significant, but assumptions not met
6) Malaria Medication	0.97883	0.02577	0.43	0.2447	0.6237	Not significant
7) Travel	0.76244	0.03958	0.914	5.0123	0.01185*	Not significant
8) Village	1.6015	0.11775	0.086	0.7054	0.5551	Not significant

SI Table 8.1: *Bray-Curtis PERMANOVA results regarding differences in fungal community structure in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant*

Group	PERMANOVA F-stat	PERMANOVA R-squared	PERMANOVA P-value	DISPERSION F-stat	DISPERSION P-value	Significance?
1) Age	0.95413	0.07765	0.551	0.0866	0.9669	Not significant
2) Antibiotics	0.85623	0.02456	0.571	0.0233	0.8795	Not significant
3) Biological Sex	0.7157	0.01949	0.752	0.0057	0.9403	Not significant
4) Diarrhoea	0.7872	0.0214	0.667	0.3184	0.5761	Not significant
5) Intestinal infection	1.3781	0.04009	0.16	7.2912	0.01085	Not significant
6) Malaria Medication	0.08662	0.02416	0.592	0.3725	0.5456	Not significant
7) Travel	1.3682	0.07252	0.094	4.8059	0.01432	Not significant
8) Village	2.268	0.16675	0.003***	0.7313	0.5405	Significant

SI Table 8.2: *Jensen-Shannon PERMANOVA results regarding differences in fungal community structure in relation to 1) the ages of research participants, 2) their former use of antibiotic treatment for tuberculosis, 3) their biological sex, 4) whether diarrhoea is or had been experienced following the consumption of certain foods, 5) whether participants have ever experienced an intestinal infection, 6) their former or current use of malaria medication, 7) their exposure to local, regional and international travel, and 8) the villages of primary residency of each research participant*

Group	PERMANOVA F-stat	PERMANOVA R-squared	PERMANOVA P-value	DISPERSION F-stat	DISPERSION P-value	Significance?
1) Age	1.0218	0.0827	0.439	0.1309	0.9411	Not significant
2) Antibiotics	0.9559	0.02735	0.474	0.0022	0.963	Not significant
3) Biological Sex	0.46131	0.01265	0.903	0.0033	0.9547	Not significant
4) Diarrhoea	0.93084	0.02521	0.492	0.2418	0.6259	Not significant
5) Intestinal infection	1.3235	0.03856	0.23	7.7461	0.008837	Not significant
6) Malaria Medication	1.0222	0.02838	0.388	0.2061	0.6562	Not significant
7) Travel	1.468	0.0774	0.129	3.0674	0.05923	Not significant
8) Village	3.0808	0.21374	0.001***	0.8737	0.4643	Significant

SI Table 9.1: *Differentially abundant bacterial genera between subjects who have and have not used antibiotics.*

Genus	**q-value**
Butyrivibrio	q = 0.00043
Escherichia-Shigella	q = 0.01566
Izemoplasmatales	q = 0.042
Lachnospiraceae_UCG-010	q = 0.04281
Treponema	q = 1e-05

SI Table 9.2: *Only one bacterial genus was differentially abundant between subjects who have and have not experienced intestinal infection.*

Genus	**q-value**
Alloprevotella	q = 0.01898

SI Table 9.3: *Differentially abundant bacterial genera between subjects from different villages*

Genus	**q_value**
[Eubacterium]_ruminantium_group	q = 0.00095
Alistipes	q = 0.01204
Anaerostipes	q = 0.00088
Barnesiella	q = 0.00691
Butyrivibrio	q = 0.00012
CAG-352	q = 0.00263
Collinsella	q = 0.04686
Desulfovibrio	q = 0.00012
Family_XIII_UCG-001	q = 0.00689
Intestinibacter	q = 0.00746
Izemoplasmatales	q = 0.00249
Lachnospira	q = 0.00088
Lachnospiraceae_NK4A136_group	q = 0.03858
Lachnospiraceae_NK4B4_group	q = 0.04686
Lachnospiraceae_UCG-010	q = 0.01469
Marvinbryantia	q = 0.00082
Monoglobus	q = 0.00077
Negativibacillus	q = 0.00681
Odoribacter	q = 0.03817
Paraprevotella	q = 0.00691
Peptococcus	q = 0.03119
Ruminococcus	q = 0.00142
Treponema	q = 0
UCG-004	q = 0.04623
Z20	q = 0.03119

SI Table 10.1: *Differentially abundant fungal genera between male and female subjects.*

Genus	**q_value**
Aureobasidium	q = 0.0383
Westerdykella	q = 0.0045

SI Table 10.2*: Only one fungal genus was differentially abundant between subjects who have and have not experienced intestinal infections.*

Genus	**q_val**
Aureobasidium	q = 2e-05

SI Table 10.3: *Differentially abundant fungal genera between subjects from different villages.*

Genus	**q_value**
Aspergillus	q = 0.02816
Candida	q = 0.00173
Mycoacia	q = 0
Panellus	q = 0.00034
Porodisculus	q = 0.02816
Saccharomyces	q = 0.00018
Schizophyllum	q = 0.02816
Stagonospora	q = 0

SI Table 11: *Primers used for the amplification of V3-V4 16S rRNA and ITS1 and 2.*

Region amplified	**Primer name**	**Primer sequence**
V1-V3 (16S rRNA)	337F	5-GACTCCTACGGGAGGCWGCAG-3
V1-V3 (16S rRNA)	805R	5-GACTACHVGGGTATCTAATCC-3
ITS1-F (ITS)		5-CTTGGTCATTTAGAGGAAGTAA-3
ITS4 (ITS)		5-TCCTCCGCTTATTGATATGC-3

Supplementary R-code

```
library(tidyverse)
library(microbiome)
library(phyloseq)
library(decontam)
library(tibble)
library(metagMisc)
library(dplyr)
library(plotrix)
library(ggplot2)
library(RColorBrewer)
library(ggpubr)
library(reshape2)
library(ggpubr)
library(mctoolsr)
library(gridExtra)
library(grid)
library(ANCOMBC)
library(vegan)
library(patchwork)
library(agricolae)
library(FSA)
library(rcompanion)
```

Reading in 16S data (using unfiltered feature table)

```
#create phyloseq object from unfiltered feature table
setwd("<path/to/wd")
biom_path_alpha <- file.path("<file-name> ")
tree_path <- file.path("<file-name>")
map_path <- file.path("<file-name>")
tree <- read_tree(tree_path)
table <- import_biom(BIOMfilename = biom_path_alpha,parallel = T)
sample_map <- import_qiime_sample_data(map_path)
pseq.unfiltered <- merge_phyloseq(table,sample_map,tree)
colnames(tax_table(pseq.unfiltered)) <- c("Domain", "Phylum", "Class", "Order", "Family",  "Genus","Species")

#change metadata fields to factors
pseq.unfiltered@sam_data$Sampleid <- as.factor(pseq.unfiltered@sam_data$Sampleid)
pseq.unfiltered@sam_data$Village<- as.factor(pseq.unfiltered@sam_data$Village)
pseq.unfiltered@sam_data$Sex<- as.factor(pseq.unfiltered@sam_data$Sex)
pseq.unfiltered@sam_data$Travel<- as.factor (pseq.unfiltered@sam_data$Travel)
pseq.unfiltered@sam_data$Diarrhoea <- as.factor (pseq.unfiltered@sam_data$Diarrhoea)
pseq.unfiltered@sam_data$Intestinal.Infections<- as.factor(pseq.unfiltered@sam_data$Intestinal.Infections)
pseq.unfiltered@sam_data$Malaria<- as.factor(pseq.unfiltered@sam_data$Malaria)
pseq.unfiltered@sam_data$Malaria.Medication<- as.factor (pseq.unfiltered@sam_data$Malaria.Medication)
pseq.unfiltered@sam_data$TB<- as.factor(pseq.unfiltered@sam_data$TB)
pseq.unfiltered@sam_data$Antibiotics<- as.factor(pseq.unfiltered@sam_data$Antibiotics)
pseq.unfiltered@sam_data$Age <- as.factor(pseq.unfiltered@sam_data$Age)
pseq.unfiltered@sam_data$Sample_or_Control <- as.factor(pseq.unfiltered@sam_data$Sample_or_Control)

#inspect contaminant using prevalence (presence/absence) at p < 0.1 (default threshold)
sample_data(pseq.unfiltered)$is.neg <- sample_data(pseq.unfiltered)$Sample_or_Control == "Control Sample"
contamdf.prev <- isContaminant(pseq.unfiltered, method="prevalence", neg="is.neg")
table(contamdf.prev$contaminant) #2 contaminants
```

```
#get list of contaminant names
contamdf.prev <- rownames_to_column(contamdf.prev, var="Taxon")
contam.df <- subset(contamdf.prev, contamdf.prev$contaminant =="TRUE")
contam.list <- contam.df$Taxon

#remove contaminant from pseq
allTaxa = taxa_names(pseq.unfiltered)
allTaxa <- allTaxa [!(allTaxa %in% contam.list)]
pseq.unfiltered.clean = prune_taxa(allTaxa, pseq.unfiltered)
```

Reading in 16S data (using filtered feature table)

```
#create phyloseq object from feature table pre-filtered with QIIME to only include reads that appeared in more than two samples
setwd("<path/to/wd>")
biom_path <- file.path("<file-name>")
tree_path <- file.path("<file-name>")
map_path <- file.path("<file-name>")
tree <- read_tree(tree_path)
table <- import_biom(BIOMfilename = biom_path,
                     parallel = T, taxa_are_rows=T)
sample_map <- import_qiime_sample_data(map_path)
pseq <- merge_phyloseq(table,sample_map,tree, taxa_are_rows=T)
colnames(tax_table(pseq)) <- c("Domain", "Phylum", "Class", "Order", "Family",  "Genus","Species")

#change metadata fields to factos
pseq@sam_data$Sampleid <- as.factor(pseq@sam_data$Sampleid)
pseq@sam_data$Village<- as.factor(pseq@sam_data$Village)
pseq@sam_data$Sex<- as.factor(pseq@sam_data$Sex)
pseq@sam_data$Travel<- as.factor (pseq@sam_data$Travel)
pseq@sam_data$Diarrhoea <- as.factor (pseq@sam_data$Diarrhoea)
pseq@sam_data$Intestinal.Infections<- as.factor(pseq@sam_data$Intestinal.Infections)
pseq@sam_data$Malaria<- as.factor(pseq@sam_data$Malaria)
pseq@sam_data$Malaria.Medication<- as.factor (pseq@sam_data$Malaria.Medication)
pseq@sam_data$TB<- as.factor(pseq@sam_data$TB)
pseq@sam_data$Antibiotics<- as.factor(pseq@sam_data$Antibiotics)
pseq@sam_data$Age <- as.factor(pseq@sam_data$Age)
pseq@sam_data$Sample_or_Control <- as.factor(pseq@sam_data$Sample_or_Control)

#inspect contaminant using prevalence (presence/absence) at p < 0.1 (default threshold)
sample_data(pseq)$is.neg <- sample_data(pseq)$Sample_or_Control == "Control Sample"
contamdf.prev <- isContaminant(pseq, method="prevalence", neg="is.neg")
table(contamdf.prev$contaminant) #3 contaminants

#get list of contaminant names
contamdf.prev <- rownames_to_column(contamdf.prev, var="Taxon")
contam.df <- subset(contamdf.prev, contamdf.prev$contaminant =="TRUE")
contam.list <- contam.df$Taxon

#remove contaminant from pseq
allTaxa = taxa_names(pseq)
allTaxa <- allTaxa [!(allTaxa %in% contam.list)]
pseq.clean = prune_taxa(allTaxa, pseq)
```

Read in ITS data

```
#create phyloseq object using ITS feature table
setwd("<path/to/wd>")
biom_path <- file.path("<file-name>")
map_path <- file.path("<file-name>")
table <- import_biom(BIOMfilename = biom_path,
                     parallel = T)
sample_map <- import_qiime_sample_data(map_path)
pseq <- merge_phyloseq(table,sample_map)
colnames(tax_table(pseq)) <- c("Domain", "Phylum", "Class", "Order", "Family",  "Genus","Species")

#change metadata fields to factos
pseq@sam_data$Sampleid <- as.factor(pseq@sam_data$Sampleid)
pseq@sam_data$Village<- as.factor(pseq@sam_data$Village)
pseq@sam_data$Sex<- as.factor(pseq@sam_data$Sex)
pseq@sam_data$Travel<- as.factor (pseq@sam_data$Travel)
pseq@sam_data$Diarrhoea <- as.factor (pseq@sam_data$Diarrhoea)
pseq@sam_data$Intestinal.Infections<- as.factor(pseq@sam_data$Intestinal.Infections)
pseq@sam_data$Malaria<- as.factor(pseq@sam_data$Malaria)
pseq@sam_data$Malaria.Medication<- as.factor (pseq@sam_data$Malaria.Medication)
pseq@sam_data$TB<- as.factor(pseq@sam_data$TB)
pseq@sam_data$Antibiotics<- as.factor(pseq@sam_data$Antibiotics)
pseq@sam_data$Age <- as.factor(pseq@sam_data$Age)
pseq@sam_data$Sample_or_Control <- as.factor(pseq@sam_data$Sample_or_Control)

#ITS contamination
#inspect contaminant using prevalence (presence/absence) at p < 0.1 (default threshold)
sample_data(pseq)$is.neg <- sample_data(pseq)$Sample_or_Control == "Control Sample"
contamdf.prev <- isContaminant(pseq, method="prevalence", neg="is.neg")
table(contamdf.prev$contaminant) #4 contaminants

#get list of contaminant names
contamdf.prev <- rownames_to_column(contamdf.prev, var="Taxon")
contam.df <- subset(contamdf.prev, contamdf.prev$contaminant =="TRUE")
contam.list <- contam.df$Taxon

#remove contaminant from pseq
allTaxa = taxa_names(pseq)
allTaxa <- allTaxa [!(allTaxa %in% contam.list)]
pseq.clean = prune_taxa(allTaxa, pseq)

#write decontaminated feature table to a tsv file for later use with FUNGuild
otu <- as.data.frame(pseq.clean@otu_table)
otu <- rownames_to_column(otu, var = "OTU ID")
tax <- as.data.frame(pseq.clean@tax_table)
tax <- rownames_to_column(tax, var="taxonomy")
write_delim(otu, file ="<path>/decontam_feature_table_ITS.tsv", delim ="\t")
write_delim(tax, file ="<path>/decontam_tax_ITS.tsv", delim ="\t")
```

Function Round_df

```
#round numeric dataframe columns to a specified number of digits
round_df <- function(x, digits) {
  # round all numeric variables
  # x: data frame
```

```
  # digits: number of digits to round
  numeric_columns <- sapply(x, mode) == 'numeric'
  x[numeric_columns] <-  round(x[numeric_columns], digits)
  x
}
```

16S Relative abundance

```
#Get relative abundance
pseq.rel<-microbiome::transform(pseq.clean,"compositional")

#Turn tax and otu tables into dataframes and combine
tax<- as.data.frame(pseq.rel@tax_table)
tax<-rownames_to_column(tax,var = "Taxon")
otu<-as.data.frame(pseq.rel@otu_table)
otu<-rownames_to_column(otu,var = "Taxon")
combined_table<-merge(x = tax, y = otu, by = "Taxon", all = TRUE)

#Add total relative abundance column
combined_table$RelAb <- rowSums( combined_table[,9:48] )

#Aggregate to see amount of taxa
#run round-df first
```

SI Table 3.1

```
#phylum
grouped_tbl_phylum <- aggregate (RelAb ~Phylum, combined_table, sum)
grouped_tbl_phylum<- grouped_tbl_phylum[order(-grouped_tbl_phylum$RelAb),]
grouped_tbl_phylum$rank <- 1:13
grouped_tbl_phylum$Phylum <- gsub("p__","",grouped_tbl_phylum$Phylum)
grouped_tbl_phylum$RelAb<-round_df(grouped_tbl_phylum$RelAb,5)
write.csv(grouped_tbl_phylum,file="<file/path>/phylum_list.csv")
#get FB ratio
fb_ratio <- 26.620283064/8.457853914
fb_ratio #3.117
```

SI Table 3.2

```
#class
grouped_tbl_class <- aggregate (RelAb ~Class, combined_table, sum)
grouped_tbl_class<- grouped_tbl_class[order(-grouped_tbl_class$RelAb),]
grouped_tbl_class$rank <- 1:18
grouped_tbl_class$Class<-gsub("c__","",grouped_tbl_class$Class)
grouped_tbl_class$RelAb<-round_df(grouped_tbl_class$RelAb,5)
write.csv(grouped_tbl_class,file="<file/path>/class_list.csv")
```

SI Table 3.3

```
#order
grouped_tbl_order <- aggregate (RelAb ~Order, combined_table, sum)
grouped_tbl_order<- grouped_tbl_order[order(-grouped_tbl_order$RelAb),]
grouped_tbl_order$rank <- 1:37
grouped_tbl_order$Order<- gsub("o__","",grouped_tbl_order$Order)
grouped_tbl_order$RelAb<-round_df(grouped_tbl_order$RelAb,5)
write.csv(grouped_tbl_order,file="<file/path>/order_list.csv")
```

SI Table 3.4

```
#family
grouped_tbl_fam <- aggregate (RelAb ~Family, combined_table, sum)
grouped_tbl_fam<- grouped_tbl_fam[fam(-grouped_tbl_fam$RelAb),]
```

```
grouped_tbl_fam$rank <- 1:61
grouped_tbl_fam$Family<-gsub("f__","",grouped_tbl_fam$Family)
grouped_tbl_fam$RelAb<--round_df(grouped_tbl_fam$RelAb,5)
write.csv(grouped_tbl_fam,file="<file/path>/fam_list.csv")
```

SI Table 3.5

```
#genus
grouped_tbl_genus <- aggregate(RelAb ~Genus, combined_table, sum)
grouped_tbl_genus<- grouped_tbl_genus[order(-grouped_tbl_genus$RelAb),]
grouped_tbl_genus$rank <- 1:131
grouped_tbl_genus$Genus<-gsub("g__","",grouped_tbl_genus$Genus)
grouped_tbl_genus$RelAb<-round_df(grouped_tbl_genus$RelAb,5)
write.csv(grouped_tbl_genus,file="<file/path>/genus_list.csv")
```

SI Table 3.6

```
#species
grouped_tbl_species <- aggregate (RelAb ~Species, combined_table, sum)
grouped_tbl_species<- grouped_tbl_species[order(-grouped_tbl_species$RelAb),]
grouped_tbl_species$rank <- 1:78
grouped_tbl_species$Species<- gsub("s__","",grouped_tbl_species$Species)
grouped_tbl_species$RelAb<-round_df(grouped_tbl_species$RelAb,5)
write.csv(grouped_tbl_species,file="<file/path>/species_list.csv")

#create relative abundance bargraphs per sample
```

Figure 2A

```
#raw unfiltered data use pseq.unfiltered----
#phylum-level - aggregate to phylum
pseq.phylum.raw <- aggregate_taxa(pseq.unfiltered,"Phylum")
pseq.phylum.rel.raw <- microbiome::transform(pseq.phylum.raw,"compositional")
#remove unknown phyla
pseq.phylum.rel.raw= subset_taxa(pseq.phylum.rel.raw, Phylum!="Unknown")

dat.raw <- psmelt(pseq.phylum.rel.raw)
dat.raw$Phylum <- gsub("p__","",dat.raw$Phylum)
dat.raw$Phylum <- with(dat.raw, reorder(Phylum,-Abundance))
p1 <- ggplot(data=dat.raw, aes(x=Sample, y=Abundance, fill=Phylum))
p1<-p1 + geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = c("darkblue", "darkgoldenrod1", "darkseagreen", "darkorchid", "darkolivegreen1",
"lightskyblue", "darkgreen", "deeppink", "khaki2", "firebrick", "brown1", "darkorange1", "cyan1", "royalblue4",
"darksalmon", "darkblue"))+
  theme(legend.position="right")+
  theme(legend.title = element_blank(),text = element_text(size=15,face="bold"),axis.text.x =
element_text(angle=90, hjust=1,size=15,face="bold"),legend.text=element_text(size=20, face="italic"))+
  labs(y = "Relative Abundance (%)", x = "")+
  guides(fill=guide_legend(ncol=2))
p1
```

Figure 2B

```
#bar plot using filtered data, without contamination removal --> use pseq ----
#phylum-level - aggregate to phylum
pseq.phylum.contaminated <- aggregate_taxa(pseq,"Phylum")
pseq.phylum.rel.contaminated <- microbiome::transform(pseq.phylum.contaminated,"compositional")
#remove unknown phyla
pseq.phylum.rel.contaminated = subset_taxa(pseq.phylum.rel.contaminated, Phylum!="Unknown")
dat.contaminated <- psmelt(pseq.phylum.rel.contaminated)
dat.contaminated$Phylum <- gsub("p__","",dat.contaminated$Phylum)
```

```
dat.contaminated$Phylum <- with(dat.contaminated, reorder(Phylum,-Abundance))

p2 <- ggplot(data=dat.contaminated, aes(x=Sample, y=Abundance, fill=Phylum))
p2<-p2 + geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = c("darkblue", "darkgoldenrod1", "darkseagreen", "darkorchid", "darkolivegreen1",
"lightskyblue", "darkgreen", "deeppink", "khaki2", "firebrick", "brown1", "darkorange1", "cyan1", "royalblue4",
"darksalmon", "darkblue"))+
  theme(legend.position="none") +
  theme(text = element_text(size=15,face="bold"),axis.text.x = element_text(angle=90,
hjust=1,size=15,face="bold"),legend.text=element_text(size=20, face="italic"))+
  labs(y = "Relative Abundance (%)", x = "")+
  guides(fill=guide_legend(nrow=4,byrow=TRUE))
p2
```

Figure 2C

```
#bar plot with contamination removal --> use pseq.clean ----
#phylum-level - aggregate to phylum
pseq.phylum.uncontaminated <- aggregate_taxa(pseq.clean,"Phylum")
pseq.phylum.rel.uncontaminated <- microbiome::transform(pseq.phylum.uncontaminated,"compositional")
#remove unknown phyla
pseq.phylum.rel.uncontaminated = subset_taxa(pseq.phylum.rel.uncontaminated, Phylum!="Unknown")
dat.uncontaminated <- psmelt(pseq.phylum.rel.uncontaminated)
dat.uncontaminated$Phylum <- gsub("p__","",dat.uncontaminated$Phylum)
dat.uncontaminated$Phylum <- with(dat.uncontaminated, reorder(Phylum,-Abundance))

p3 <- ggplot(data=dat.uncontaminated, aes(x=Sample, y=Abundance, fill=Phylum))
p3<-p3 + geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = c("darkblue", "darkgoldenrod1", "darkseagreen", "darkorchid", "darkolivegreen1",
"lightskyblue", "darkgreen", "deeppink", "khaki2", "firebrick", "brown1", "darkorange1", "cyan1", "royalblue4",
"darksalmon", "darkblue"))+
  theme(legend.position="none") +
  theme(text = element_text(size=15,face="bold"),axis.text.x = element_text(angle=90,
hjust=1,size=15,face="bold"),legend.text=element_text(size=20, face="italic"))+
  labs(y = "Relative Abundance (%)", x = "")+
  guides(fill=guide_legend(nrow=4,byrow=TRUE))
p3
leg <- get_legend(p1)
p1 <- p1 + theme(legend.position = "none")
ggarrange(p1, p2, p3, leg, nrow = 2, ncol=2, labels = c("A","B","C"), font.label = list(size=35))
```

ITS Relative abundance

```
#Get relative abundance
pseq.rel<-microbiome::transform(pseq.clean,"compositional")

#Turn tax and otu tables into dataframes and combine
tax<- as.data.frame(pseq.rel@tax_table)
tax<-rownames_to_column(tax,var = "Taxon")
otu<-as.data.frame(pseq.rel@otu_table)
otu<-rownames_to_column(otu,var = "Taxon")
combined_table<-merge(x = tax, y = otu, by = "Taxon", all = TRUE)

#Add total relative abundance column
combined_table$RelAb <- rowSums( combined_table[,9:47] )

#Aggregate to see amount of taxa
#run round-df first
```

SI Table 4.1

```
#phylum
grouped_tbl_phylum <- aggregate (RelAb ~Phylum, combined_table, sum)
grouped_tbl_phylum<- grouped_tbl_phylum[order(-grouped_tbl_phylum$RelAb),]
grouped_tbl_phylum$rank <- 1:4
grouped_tbl_phylum$Phylum <- gsub("p__","",grouped_tbl_phylum$Phylum)
grouped_tbl_phylum$RelAb<-round_df(grouped_tbl_phylum$RelAb,5)
write.csv(grouped_tbl_phylum,file="<file/path>/phylum_list.csv")
```

SI Table 4.2

```
#class
grouped_tbl_class <- aggregate (RelAb ~Class, combined_table, sum)
grouped_tbl_class<- grouped_tbl_class[order(-grouped_tbl_class$RelAb),]
grouped_tbl_class$rank <- 1:12
grouped_tbl_class$Class<-gsub("c__","",grouped_tbl_class$Class)
grouped_tbl_class$RelAb<-round_df(grouped_tbl_class$RelAb,5)
write.csv(grouped_tbl_class,file="<file/path>/class_list.csv")
```

SI Table 4.3

```
#order
grouped_tbl_order <- aggregate (RelAb ~Order, combined_table, sum)
grouped_tbl_order<- grouped_tbl_order[order(-grouped_tbl_order$RelAb),]
grouped_tbl_order$rank <- 1:26
grouped_tbl_order$Order<- gsub("o__","",grouped_tbl_order$Order)
grouped_tbl_order$RelAb<-round_df(grouped_tbl_order$RelAb,5)
write.csv(grouped_tbl_order,file="<file/path>/order_list.csv")
```

SI Table 4.4

```
#family
grouped_tbl_fam <- aggregate (RelAb ~Family, combined_table, sum)
grouped_tbl_fam<- grouped_tbl_fam[order(-grouped_tbl_fam$RelAb),]
grouped_tbl_fam$rank <- 1:63
grouped_tbl_fam$Family<-gsub("f__","",grouped_tbl_fam$Family)
grouped_tbl_fam$RelAb<-round_df(grouped_tbl_fam$RelAb,5)
write.csv(grouped_tbl_fam,file="<file/path>/fam_list.csv")
```

SI Table 4.5

```
#genus
grouped_tbl_genus <- aggregate(RelAb ~Genus, combined_table, sum)
grouped_tbl_genus<- grouped_tbl_genus[order(-grouped_tbl_genus$RelAb),]
grouped_tbl_genus$rank <- 1:81
grouped_tbl_genus$Genus<-gsub("g__","",grouped_tbl_genus$Genus)
grouped_tbl_genus$RelAb<-round_df(grouped_tbl_genus$RelAb,5)
write.csv(grouped_tbl_genus,file="<file/path>/genus_list.csv")
```

SI Table 4.6

```
#species
grouped_tbl_species <- aggregate (RelAb ~Species, combined_table, sum)
grouped_tbl_species<- grouped_tbl_species[order(-grouped_tbl_species$RelAb),]
grouped_tbl_species$rank <- 1:87
grouped_tbl_species$Species<- gsub("s__","",grouped_tbl_species$Species)
grouped_tbl_species$RelAb<-round_df(grouped_tbl_species$RelAb,5)
write.csv(grouped_tbl_species,file="<file/path>/species_list.csv")

#create relative abundance bargraphs per sample
```

SI Table 5.1-5.3

```
alphaObserved = estimate_richness(pseq.unfiltered.clean, measures="Observed")
alphaSimpson = estimate_richness(pseq.unfiltered.clean, measures="Shannon")
alphaChao = estimate_richness(pseq.unfiltered.clean, measures="Chao1")

alpha.stats <- cbind(alphaObserved, sample_data(pseq.unfiltered.clean))
alpha.stats2 <- cbind(alpha.stats, alphaSimpson)
alpha.stats3 <- cbind(alpha.stats2, alphaChao)

#for two factors
kruskal.test(Shannon~Antibiotics, data = alpha.stats3)

#for more than 2 factors
dunn.test(alpha.stats3$Shannon, alpha.stats3$Travel, method="hochberg")
```

Figure 3A

```
#before contamination removal --> use pseq ----
#phylum-level - aggregate to phylum
pseq.phylum.contaminated <- aggregate_taxa(pseq,"Phylum")
pseq.phylum.rel.contaminated <- microbiome::transform(pseq.phylum.contaminated,"compositional")
#remove unknown phyla
pseq.phylum.rel.contaminated = subset_taxa(pseq.phylum.rel.contaminated, Phylum!="Unknown")
dat.contaminated <- psmelt(pseq.phylum.rel.contaminated)
dat.contaminated$Phylum <- gsub("p__","",dat.contaminated$Phylum)

p1 <- ggplot(data=dat.contaminated, aes(x=Sample, y=Abundance, fill=Phylum))
p1<-p1 + geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = c("darkblue", "darkgoldenrod1", "darkseagreen", "darkorchid", "darkolivegreen1",
"lightskyblue", "darkgreen", "deeppink", "khaki2", "firebrick", "brown1", "darkorange1", "cyan1", "royalblue4",
"darksalmon", "darkblue"))+
  theme(legend.position="right")+
  theme(legend.title = element_blank(),text = element_text(size=15,face="bold"),axis.text.x =
element_text(angle=90, hjust=1,size=15,face="bold"),legend.text=element_text(size=20, face="italic"))+
  labs(y = "Relative Abundance (%)", x = "")+
  guides(fill=guide_legend(ncol=4))
p1
```

Figure 3B

```
#after contamination removal --> use pseq.clean ----
#phylum-level - aggregate to phylum
pseq.phylum.uncontaminated <- aggregate_taxa(pseq.clean,"Phylum")
pseq.phylum.rel.uncontaminated <- microbiome::transform(pseq.phylum.uncontaminated,"compositional")
#remove unknown phyla
pseq.phylum.rel.uncontaminated = subset_taxa(pseq.phylum.rel.uncontaminated, Phylum!="Unknown")
dat.uncontaminated <- psmelt(pseq.phylum.rel.uncontaminated)
dat.uncontaminated$Phylum <- gsub("p__","",dat.uncontaminated$Phylum)

p2 <- ggplot(data=dat.uncontaminated, aes(x=Sample, y=Abundance, fill=Phylum))
p2<-p2 + geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = c("darkblue", "darkgoldenrod1", "darkseagreen", "darkorchid", "darkolivegreen1",
"lightskyblue", "darkgreen", "deeppink", "khaki2", "firebrick", "brown1", "darkorange1", "cyan1", "royalblue4",
"darksalmon", "darkblue"))+
  theme(legend.position="none") +
  theme(text = element_text(size=15,face="bold"),axis.text.x = element_text(angle=90,
hjust=1,size=15,face="bold"),legend.text=element_text(size=15, face="italic"))+
  labs(y = "Relative Abundance (%)", x = "")+
```

```
  guides(fill=guide_legend(nrow=4,byrow=TRUE))
p2

leg <- get_legend(p1)
p1 <- p1 + theme(legend.position = "none")
ggarrange(p1,p2, common.legend = T, labels = c("A","B"),font.label = list(size=20))
```

16S Core Microbiome

```
#clean
pseq.unfiltered.clean <-prune_samples(pseq.unfiltered.clean@sam_data$Sampleid!="KIT-CTRL",pseq.unfiltered.clean)
pseq.unfiltered.rel <- microbiome::transform(pseq.unfiltered.clean, "compositional")
pseq.unfiltered.rel@tax_table <-gsub("g__","",pseq.unfiltered.rel@tax_table)

#combine unidentified and NA taxa
df <- data.frame(pseq.unfiltered.rel@tax_table)
df$Genus <- gsub("unidentified",NA, df$Genus)
tax_table(pseq.unfiltered.rel) <- as.matrix(df)
#determine detection and prevalence for core microbiome ----
det <- c(0, 0.1, 0.5, 2, 5, 20)/100
#prevalences <- seq(from = 10, to = 100, by = 10)
prevalences <- c(10,20,30,40,50,60,70,80,90,100)
#prev 0%, 50%,
```

SI Figure 1A

```
p.16<-plot_core(pseq.unfiltered.rel, prevalences = prevalences,
        detections = det, plot.type = "lineplot") +
  xlab("Detection threshold (%)") +
  theme_bw()

#the lines are parallel between 0.75-0.5 prevalence
#chosen prevalence: 60%, chosen det. threshold: 0.008%
#plot core genera ----
pseq.core.taxa <- aggregate_taxa(pseq.unfiltered.rel,"Genus")
pseq.core.taxa <- core(pseq.core.taxa, detection = 0.008/100, prevalence = .6)
#clean
tax <- as.matrix(pseq.core.taxa@tax_table)
tax<- tax[-c(32,46),]
tax_table(pseq.core.taxa) <- tax
```

Figure 4

```
p <- plot_core(pseq.core.taxa, plot.type = "heatmap",
               prevalences = prevalences,
               detections = det,
               colours = rev(brewer.pal(5, "Spectral")),
               min.prevalence = .5, horizontal = F)
p + theme(axis.text = element_text(size = 12, face="bold.italic"),
          axis.title = element_text(size=15, face="bold"),
           legend.title = element_text(face="bold",size=15))+
  coord_fixed(ratio=.2)
```

ITS Core Microbiome

```
#convert to relative abundance, clean, and remove controls ----
pseq.core.rel <- microbiome::transform(pseq.clean, "compositional")
pseq.core.rel<-prune_samples(pseq.core.rel@sam_data$Sampleid!="CON-CTRL",pseq.core.rel)
pseq.core.rel@tax_table <-gsub("g__","",pseq.core.rel@tax_table)
```

```
#combine unidentified and NA taxa
df <- data.frame(pseq.core.rel@tax_table)
df$Genus <- gsub("unidentified",NA, df$Genus)
tax_table(pseq.core.rel) <- as.matrix(df)
#determine detection and prevalence for core microbiome ----
det <- c(0, 0.1, 0.5, 2, 5, 20)/100
prevalences <- seq(from = 10, to = 100, by = 5)
```

SI Figure 1B

```
p.ITS <-plot_core(pseq.core.rel, prevalences = prevalences,
        detections = det, plot.type = "lineplot") +
  xlab("Detection threshold (%)") +
  theme_bw()
#a core microbiome of under 50 taxa
#the lines are parallel between 0.25-0.60 prevalence
#chosen prevalence: 50%, chosen det. treshold: 0.008%

#plot core taxa ----
pseq.core.taxa <- aggregate_taxa(pseq.core.rel,"Genus")
pseq.core.taxa <- core(pseq.core.taxa, detection = 0.008/100, prevalence = .5)
```

Figure 5

```
p <- plot_core(pseq.core.taxa, plot.type = "heatmap",
               prevalences = prevalences,
               detections = det,
               colours = rev(brewer.pal(5, "Spectral")),
               min.prevalence = .5, horizontal = F)
p+ theme(axis.text = element_text(size = 12, face="bold.italic"),
         axis.title = element_text(size=15, face="bold"),
         legend.title = element_text(face="bold",size=15))+
  coord_fixed(ratio=1)

#combine 16S and ITS cores into one figure
p.final <-ggarrange(p.16, p.ITS, nrow = 1, common.legend = T, labels = c("A","B"),
                    font.label = list(size=20,face="bold"))
p.final
```

16S PERMANOVA (SI Table 7.1-7.2)

```
#use compositional abundance
#clean
pseq.rel<- microbiome::transform(pseq.clean, "compositional")
pseq.rel<-prune_samples(pseq.rel@sam_data$Sampleid!="KIT-CTRL",pseq.rel)
pseq.rel <- phyloseq_rm_na_tax(pseq.rel)
#make matrix of wunifrac and jensen-shannon distances for groups in which no subject answered "Uncertain"
perm.wunifrac <- phyloseq ::distance (pseq.rel,"wunifrac")
perm.jsd <- phyloseq::distance(pseq.rel, "jsd")
#permanova and disp tests

#age
vegan::adonis(perm.wunifrac ~ phyloseq::sample_data(pseq.rel)$Age)
age.wunifrac <- vegan::betadisper(perm.wunifrac, phyloseq::sample_data(pseq.rel)$Age)
anova(age.wunifrac)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Age)
age.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Age)
anova(age.jsd)
```

```
#antibiotics
pseq.ant = subset_samples(pseq.rel, Antibiotics != "Uncertain")
perm.wunifrac.ant <- phyloseq ::distance (pseq.ant,"wunifrac")
perm.jsd.ant <- phyloseq::distance(pseq.ant, "jsd")
vegan::adonis(perm.wunifrac.ant ~ phyloseq::sample_data(pseq.ant)$Antibiotics)
ant.wunifrac <- vegan::betadisper(perm.wunifrac.ant, phyloseq::sample_data(pseq.ant)$Antibiotics)
anova(ant.wunifrac)
vegan::adonis(perm.jsd.ant~phyloseq::sample_data(pseq.ant)$Antibiotics)
ant.jsd<-vegan::betadisper(perm.jsd.ant , phyloseq::sample_data(pseq.ant)$Antibiotics)
anova(ant.jsd)

#biological sex
vegan::adonis(perm.wunifrac ~ phyloseq::sample_data(pseq.rel)$Sex)
sex.wunifrac <- vegan::betadisper(perm.wunifrac, phyloseq::sample_data(pseq.rel)$Sex)
anova(sex.wunifrac)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Sex)
sex.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Sex)
anova(sex.jsd)

#Diarrhoea
vegan::adonis(perm.wunifrac ~ phyloseq::sample_data(pseq.rel)$Diarrhoea)
diar.wunifrac <- vegan::betadisper(perm.wunifrac, phyloseq::sample_data(pseq.rel)$Diarrhoea)
anova(diar.wunifrac)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Diarrhoea)
diar.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Diarrhoea)
anova(diar.jsd)

#Intestinal infections
pseq.int = subset_samples(pseq.rel, Intestinal.Infections != "Uncertain")
perm.wunifrac.int <- phyloseq ::distance (pseq.int,"wunifrac")
perm.jsd.int <- phyloseq::distance(pseq.int, "jsd")
vegan::adonis(perm.wunifrac.int ~ phyloseq::sample_data(pseq.int)$Intestinal.Infections)
int.wunifrac <- vegan::betadisper(perm.wunifrac.int, phyloseq::sample_data(pseq.int)$Intestinal.Infections)
anova(int.wunifrac)

vegan::adonis(perm.jsd.int~phyloseq::sample_data(pseq.int)$Intestinal.Infections)
int.jsd<-vegan::betadisper(perm.jsd.int , phyloseq::sample_data(pseq.int)$Intestinal.Infections)
anova(int.jsd)

#Malaria medication
pseq.mal = subset_samples(pseq.rel, Malaria.Medication != "Uncertain")
perm.wunifrac.mal <- phyloseq ::distance (pseq.mal,"wunifrac")
perm.jsd.mal <- phyloseq::distance(pseq.mal, "jsd")
vegan::adonis(perm.wunifrac.mal ~ phyloseq::sample_data(pseq.mal)$Malaria.Medication)
mal.wunifrac <- vegan::betadisper(perm.wunifrac.mal, phyloseq::sample_data(pseq.mal)$Malaria.Medication)
anova(mal.wunifrac)

vegan::adonis(perm.jsd.mal~phyloseq::sample_data(pseq.mal)$Malaria.Medication)
mal.jsd<-vegan::betadisper(perm.jsd.mal , phyloseq::sample_data(pseq.mal)$Malaria.Medication)
anova(mal.jsd)

#Travel
vegan::adonis(perm.wunifrac ~ phyloseq::sample_data(pseq.rel)$Travel)
trav.wunifrac <- vegan::betadisper(perm.wunifrac, phyloseq::sample_data(pseq.rel)$Travel)
anova(trav.wunifrac)
```

```
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Travel)
trav.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Travel)
anova(trav.jsd)

#Village
vegan::adonis(perm.wunifrac ~ phyloseq::sample_data(pseq.rel)$Village)
vil.wunifrac <- vegan::betadisper(perm.wunifrac, phyloseq::sample_data(pseq.rel)$Village)
anova(vil.wunifrac)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Village)
vil.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Village)
anova(vil.jsd)
```

SI Table 6.1-6.3

```
alphaObserved = estimate_richness(pseq.clean, measures="Observed")
alphaSimpson = estimate_richness(pseq.clean, measures="Shannon")
alphaChao = estimate_richness(pseq.clean, measures="Chao1")

alpha.stats <- cbind(alphaObserved, sample_data(pseq.clean))
alpha.stats2 <- cbind(alpha.stats, alphaSimpson)
alpha.stats3 <- cbind(alpha.stats2, alphaChao)

#for two factors
kruskal.test(Shannon~Antibiotics, data = alpha.stats3)

#for more than 2 factors
dunn.test(alpha.stats3$Shannon, alpha.stats3$Malaria, method="hochberg")
```

ITS PERMANOVA (SI Table 8.1-8.2)

```
#use compositional abundance
#clean
pseq.rel<- microbiome::transform(pseq.clean, "compositional")
pseq.rel<-prune_samples(pseq.rel@sam_data$Sampleid!="CON-CTRL",pseq.rel)
pseq.rel <- phyloseq_rm_na_tax(pseq.rel)
#make matrix of bray-curtis and jensen-shannon distances for groups in which no subject answered "Uncertain"
perm.bray <- phyloseq ::distance (pseq.rel,"bray")
perm.jsd <- phyloseq::distance(pseq.rel, "jsd")

#permanova and disp tests
#age
vegan::adonis(perm.bray ~ phyloseq::sample_data(pseq.rel)$Age)
age.bray <- vegan::betadisper(perm.bray, phyloseq::sample_data(pseq.rel)$Age)
anova(age.bray)

vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Age)
age.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Age)
anova(age.jsd)
#antibiotics
pseq.ant = subset_samples(pseq.rel, Antibiotics != "Uncertain")
perm.bray.ant <- phyloseq ::distance (pseq.ant,"bray")
perm.jsd.ant <- phyloseq::distance(pseq.ant, "jsd")
vegan::adonis(perm.bray.ant ~ phyloseq::sample_data(pseq.ant)$Antibiotics)
ant.bray <- vegan::betadisper(perm.bray.ant, phyloseq::sample_data(pseq.ant)$Antibiotics)
anova(ant.bray)
vegan::adonis(perm.jsd.ant~phyloseq::sample_data(pseq.ant)$Antibiotics)
ant.jsd<-vegan::betadisper(perm.jsd.ant , phyloseq::sample_data(pseq.ant)$Antibiotics)
```

```
anova(ant.jsd)

#biological sex
vegan::adonis(perm.bray ~ phyloseq::sample_data(pseq.rel)$Sex)
sex.bray <- vegan::betadisper(perm.bray, phyloseq::sample_data(pseq.rel)$Sex)
anova(sex.bray)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Sex)
sex.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Sex)
anova(sex.jsd)

#Diarrhoea
vegan::adonis(perm.bray ~ phyloseq::sample_data(pseq.rel)$Diarrhoea)
diar.bray <- vegan::betadisper(perm.bray, phyloseq::sample_data(pseq.rel)$Diarrhoea)
anova(diar.bray)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Diarrhoea)
diar.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Diarrhoea)
anova(diar.jsd)

#Intestinal infections
pseq.int = subset_samples(pseq.rel, Intestinal.Infections != "Uncertain")
perm.bray.int <- phyloseq ::distance (pseq.int,"bray")
perm.jsd.int <- phyloseq::distance(pseq.int, "jsd")
vegan::adonis(perm.bray.int ~ phyloseq::sample_data(pseq.int)$Intestinal.Infections)
int.bray <- vegan::betadisper(perm.bray.int, phyloseq::sample_data(pseq.int)$Intestinal.Infections)
anova(int.bray)
vegan::adonis(perm.jsd.int~phyloseq::sample_data(pseq.int)$Intestinal.Infections)
int.jsd<-vegan::betadisper(perm.jsd.int , phyloseq::sample_data(pseq.int)$Intestinal.Infections)
anova(int.jsd)

#Malaria medication
pseq.mal = subset_samples(pseq.rel, Malaria.Medication != "Uncertain")
perm.bray.mal <- phyloseq ::distance (pseq.mal,"bray")
perm.jsd.mal <- phyloseq::distance(pseq.mal, "jsd")
vegan::adonis(perm.bray.mal ~ phyloseq::sample_data(pseq.mal)$Malaria.Medication)
mal.bray <- vegan::betadisper(perm.bray.mal, phyloseq::sample_data(pseq.mal)$Malaria.Medication)
anova(mal.bray)
vegan::adonis(perm.jsd.mal~phyloseq::sample_data(pseq.mal)$Malaria.Medication)
mal.jsd<-vegan::betadisper(perm.jsd.mal , phyloseq::sample_data(pseq.mal)$Malaria.Medication)
anova(mal.jsd)

#Travel
vegan::adonis(perm.bray ~ phyloseq::sample_data(pseq.rel)$Travel)
trav.bray <- vegan::betadisper(perm.bray, phyloseq::sample_data(pseq.rel)$Travel)
anova(trav.bray)
vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Travel)
trav.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Travel)
anova(trav.jsd)

#Village
vegan::adonis(perm.bray ~ phyloseq::sample_data(pseq.rel)$Village)
vil.bray <- vegan::betadisper(perm.bray, phyloseq::sample_data(pseq.rel)$Village)
anova(vil.bray)

vegan::adonis(perm.jsd~phyloseq::sample_data(pseq.rel)$Village)
vil.jsd<-vegan::betadisper(perm.jsd , phyloseq::sample_data(pseq.rel)$Village)
```

```
anova(vil.jsd)
```

Function: Annote_Figure

```
#annotate figures
#age
age <-annotate_figure(
  age,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="1) Age",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 35,
  fig.lab.face = "bold"
)
age

#antibiotics
ant <-annotate_figure(
  ant,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="2) Antibotics",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 30,
  fig.lab.face = "bold"
)
ant

#biological sex
sex <-annotate_figure(
  sex,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="3) Biological Sex",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 30,
  fig.lab.face = "bold"
sex

#diarrhoea
diar <-annotate_figure(
  diar,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="4) Diarrhoea",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 35,
  fig.lab.face = "bold"
```

```
)
diar

#intestinal infection
int <-annotate_figure(
  int,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="5) Intestinal \nInfection",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 30,
  fig.lab.face = "bold"
)
int

#malaria
mal <-annotate_figure(
  mal,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="6) Malaria \nMedication",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 30,
  fig.lab.face = "bold"
)
Mal

#travel
trav <-annotate_figure(
  trav,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="7) Travel",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 35,
  fig.lab.face = "bold"
)
trav

#village
vil <-annotate_figure(
  vil,
  top = NULL,
  bottom = NULL,
  left = NULL,
  right = NULL,
  fig.lab ="8) Village",
  fig.lab.pos = c("top.left"),
  fig.lab.size = 30,
```

```
  fig.lab.face = "bold"
)
vil
```

16S Community Composition

Figure 6A

```
#alpha div ----
p.alp=plot_richness(pseq.unfiltered.clean, color="Sampleid", measures=c("Simpson", "Shannon"),nrow = 3)
p.alp + geom_point(size=10, alpha=0.7)+
  theme(strip.text = element_text(size=30, face= "bold"),
        legend.position = "none",
        axis.text = element_text(size = 20, face="bold"),
        plot.title = element_text(size=35, face="bold"))+
  labs (x="",y="")+
  ggtitle("A")

#beta div must be run on filtered phyloseq object ----
#run 1-create-phyloseq-16S
#remove controls for beta div
#relative abundance
pseq.rel <- microbiome::transform(pseq.clean, 'compositional')
pseq.rel<-prune_samples(pseq@sam_data$Sampleid!="KIT-CTRL",pseq.rel)

#weighted unifrac and jensen-shannon matrices
wunifrac_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "wunifrac")
jsd_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "jsd")
```

Figure 6 B, C

```
p.nmds.wu<-plot_ordination(pseq.rel, wunifrac_matrix, color = "Village") + geom_point(size = 10)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.text = element_text(size = 35, face="bold"),
        legend.title = element_text(size=35, face="bold"),
        plot.title = element_text(size=35, face="bold"))+
  ggtitle("B")
  p.nmds.wu
p.nmds.jsd<-plot_ordination(pseq.rel, jsd_matrix, color = "Village")+geom_point(size = 10)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.text = element_text(size = 35, face="bold"),
        legend.title = element_text(size=35, face="bold"),
        plot.title = element_text(size=35, face="bold"))+
  ggtitle("C")
p.nmds.jsd
#run annotate-figure
```

SI Figure 2

```
#beta div must be run on filtered phyloseq object ----
#run 1-create-phyloseq-16S
#remove controls for beta div
#relative abundance
pseq.rel <- microbiome::transform(pseq.clean, 'compositional')
pseq.rel<-prune_samples(pseq@sam_data$Sampleid!="KIT-CTRL",pseq.rel)

#weighted unifrac and jensen-shannon matrices
wunifrac_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "wunifrac")
```

```
jsd_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "jsd")

#Age
pseq.rel@sam_data$Age <- factor(pseq.rel@sam_data$Age, levels = c("Late-adolescent","Early-adult","Middle-adult","Late-adult"))
age1 <- phyloseq::plot_ordination(pseq.rel, wunifrac_matrix, color = "Age") + geom_point(size = 10)+
theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
age2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Age") + geom_point(size = 10)
leg <- get_legend(age1)
age <-ggarrange(age1, age2, leg,nrow = 1,common.legend=T,labels = c("Weighted UniFrac"
                                                    ,"Jensen-Shannon"),
          font.label = list(size=35,face="bold"),
          legend = "right",
          widths = c(2,2,.15),
          label.x = .1)
#run annotate-figure

#Antibiotics
#remove uncertain subjects
pseq.rel.ant = subset_samples(pseq.rel, Antibiotics != "Uncertain")
pseq.rel.ant<-prune_samples(pseq.rel.ant@sam_data$Sampleid!="KIT-CTRL",pseq.rel.ant)
pseq.rel.ant@sam_data$Antibiotics <- factor(pseq.rel.ant@sam_data$Antibiotics, levels = c("Yes","No"))
wunifrac_matrix.ant <- ordinate(pseq.rel.ant, method = "NMDS", distance = "wunifrac")
jsd_matrix.ant <- ordinate(pseq.rel.ant, method = "NMDS", distance = "jsd")
ant1 <- phyloseq::plot_ordination(pseq.rel.ant, wunifrac_matrix.ant, color = "Antibiotics") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
ant2 <- phyloseq::plot_ordination(pseq.rel.ant, jsd_matrix.ant, color = "Antibiotics") + geom_point(size = 10)
ant.leg <- get_legend(ant1)
ant <-ggarrange(ant1, ant2, ant.leg,nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
          font.label = list(size=35,face="bold"),
          widths = c(2,2,.15),
          label.x = .15,
          legend = "right")
#run annotate-figure

#Biological sex
sex1 <- phyloseq::plot_ordination(pseq.rel, wunifrac_matrix, color = "Sex") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
sex2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Sex") + geom_point(size = 10)
sex.leg <- get_legend(sex1)
sex <-ggarrange(sex1, sex2, sex.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
          font.label = list(size=30,face="bold"),
          widths = c(2,2,.15),
          label.x = .3,
          legend = "right")
#run annotate-figure

#Diarrhoea
diar1 <- phyloseq::plot_ordination(pseq.rel, wunifrac_matrix, color = "Diarrhoea") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
```

```
      legend.title = element_text(size=30,face="bold"))
diar2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Diarrhoea") + geom_point(size = 10)
diar.leg <- get_legend(diar1)
diar <-ggarrange(diar1, diar2, diar.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
            font.label = list(size=30,face="bold"),
            widths = c(2,2,.15),
            label.x = .3,
            legend = "right")
#run annotate-figure

#16S Intestinal Infections
#remove uncertain subjects
pseq.rel.int = subset_samples(pseq.rel, Intestinal.Infections != "Uncertain")
pseq.rel.int<-prune_samples(pseq.rel.int@sam_data$Sampleid!="KIT-CTRL",pseq.rel.int)
pseq.rel.int@sam_data$Intestinal.Infections <- factor(pseq.rel.int@sam_data$Intestinal.Infections, levels = c("Yes","No"))
wunifrac_matrix.int <- ordinate(pseq.rel.int, method = "NMDS", distance = "wunifrac")
jsd_matrix.int <- ordinate(pseq.rel.int, method = "NMDS", distance = "jsd")
int1<- phyloseq::plot_ordination(pseq.rel.int, wunifrac_matrix.int, color = "Intestinal.Infections") + geom_point(size = 10)+
  scale_color_discrete("Intestinal Infections")+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
int2<- phyloseq::plot_ordination(pseq.rel.int, jsd_matrix.int, color = "Intestinal.Infections") + geom_point(size = 10)+
  scale_colour_discrete("Intestinal Infections")
int.leg <- get_legend(int1)
int <-ggarrange(int1, int2, int.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
            font.label = list(size=30,face="bold"),
            widths = c(2,2,.15),
            label.x = .15,
            legend = "right")
#run annotate-figure

#16S Malaria Medication
#remove uncertain subjects
pseq.rel.mal = subset_samples(pseq.rel, Malaria.Medication != "Uncertain")
pseq.rel.mal<-prune_samples(pseq.rel.mal@sam_data$Sampleid!="KIT-CTRL",pseq.rel.mal)
pseq.rel.mal@sam_data$Malaria.Medication <- factor(pseq.rel.mal@sam_data$Malaria.Medication, levels = c("Yes","No"))
wunifrac_matrix.mal <- ordinate(pseq.rel.mal, method = "NMDS", distance = "wunifrac")
jsd_matrix.mal <- ordinate(pseq.rel.mal, method = "NMDS", distance = "jsd")
mal1<- phyloseq::plot_ordination(pseq.rel.mal, wunifrac_matrix.mal, color = "Malaria.Medication") + geom_point(size = 10)+
  scale_color_discrete("Malaria Medication")+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
mal2<- phyloseq::plot_ordination(pseq.rel.mal, jsd_matrix.mal, color = "Malaria.Medication") + geom_point(size = 10)+
  scale_colour_discrete("Malaria Medication")
mal.leg <- get_legend(mal1)
mal <-ggarrange(mal1, mal2, mal.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
            font.label = list(size=30,face="bold"),
            widths = c(2,2,.15),
```

```
            label.x = .15,
            legend = "right")
#run annotate-figure

#16S Travel
trav1<- phyloseq::plot_ordination(pseq.rel, wunifrac_matrix, color = "Travel") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
trav2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Travel") + geom_point(size = 10)
trav.leg <- get_legend(trav1)
trav <-ggarrange(trav1, trav2, trav.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
            font.label = list(size=30,face="bold"),
            widths = c(2,2,.15),
            label.x = .15,
            legend = "right")
#run annotate-figure

#16S Village
#for Supplementary
vil1<- phyloseq::plot_ordination(pseq.rel, wunifrac_matrix, color = "Village") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
vil2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Village") + geom_point(size = 10)
vil.leg <- get_legend(vil1)
vil <-ggarrange(vil1, vil2, vil.leg, nrow = 1,common.legend=T,labels = c("Weighted UniFrac","Jensen-Shannon"),
             font.label = list(size=30,face="bold"),
             widths = c(2,2,.15),
             label.x = .15,
             legend = "right")
#run annotate-figure
```

Figure 6D and E

```
#barplots of village average distances ----
#first get p-values using pairwise adonis
i.wunifrac="wunifrac"
pseq.rel %>%
  phyloseq::distance(method = i.wunifrac) ->  dist.wunifrac
wunifrac.p<-mctoolsr::calc_pairwise_permanovas(dist.wunifrac, as(sample_data(pseq.rel), "data.frame"), "Village")

#then construct graphs
wu = phyloseq::distance(pseq.rel, "wunifrac")
wu.m = melt(as.matrix(wu))
wu.m = wu.m %>%
  filter(as.character(Var1) != as.character(Var2)) %>%
  mutate_if(is.factor, as.character)

#village
sd.village <- data.frame(pseq.rel@sam_data$Sampleid, pseq.rel@sam_data$Village)
sd.village <- sd.village %>%
  rename(Village = pseq.rel.sam_data.Village)
sd.village <- sd.village %>%
  rename(Sampleid = pseq.rel.sam_data.Sampleid)
colnames(sd.village) = c("Var1", "Type1")
```

```
wu.sd = left_join(wu.m, sd.village, by = "Var1")
colnames(sd.village) = c("Var2", "Type2")
wu.sd = left_join(wu.sd, sd.village, by = "Var2")

#remove self from wu.sd
wu.sd = wu.sd %>%
  filter(as.character(Type1) != as.character(Type2)) %>%
  mutate_if(is.factor, as.character)

#duinpos vs denui
duinpos.denui <- subset(wu.sd, wu.sd$Type1 == "Duinpos")
duinpos.denui <- subset(duinpos.denui, duinpos.denui$Type2 == "Den/ui")
duinpos.denui.avg <- mean(duinpos.denui$value)

#duinpos vs MP
duinpos.MP <- subset(wu.sd, wu.sd$Type1 == "Duinpos")
duinpos.MP <- subset(duinpos.MP, duinpos.MP$Type2 == "Mountain Pos")
duinpos.MP.avg <- mean (duinpos.MP$value)

#duinpos vs OM
duinpos.OM <- subset(wu.sd, wu.sd$Type1 == "Duinpos")
duinpos.OM <- subset(duinpos.OM, duinpos.OM$Type2 == "!Om!o")
duinpos.OM.avg <- mean(duinpos.OM$value)

#denui vs MP
denui.MP <- subset(wu.sd, wu.sd$Type1 == "Den/ui")
denui.MP <- subset(duinpos.MP, denui.MP$Type2 == "Mountain Pos")
denui.MP.avg <- mean(duinpos.MP$value)

#denui vs OM
denui.OM <- subset(wu.sd, wu.sd$Type1 == "Den/ui")
denui.OM <- subset(denui.OM, denui.OM$Type2 == "!Om!o")
denui.OM.avg <- mean(denui.OM$value)

#MP vs OM
MP.OM <- subset(wu.sd, wu.sd$Type1 == "Mountain Pos")
MP.OM <- subset(MP.OM, MP.OM$Type2 == "!Om!o")
MP.OM.avg <- mean(MP.OM$value)

vals <- c(duinpos.denui.avg,duinpos.MP.avg,duinpos.OM.avg,denui.MP.avg,denui.OM.avg,MP.OM.avg)
col <- c("Duinpos vs Den/ui", "Duinpos vs Mountain Pos", "Duinpos vs !Om!o",
       "Den/ui vs Mountain Pos", "Den/ui vs !Om!o","Mountain Pos vs !Om!o")

village.df.wu <- data.frame(col,vals)
village.df.wu$FDRp <- wunifrac.p$pvalFDR
village.df.wu$FDRp <- sub("^","p =", village.df.wu$FDRp)

p.bar.wu<-ggplot(village.df.wu, aes(x=reorder(col,vals), y=vals,fill= col)) +
  geom_bar(stat = "identity") +
  theme_bw()+
  theme(axis.title = element_text(size = 30,face="bold"),
        axis.text.x = element_blank(),
        axis.text.y = element_text(size=25, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(face="bold",size=30),
```

```
      legend.position = "right",
      aspect.ratio = 5/5)+
  geom_text(aes(label=FDRp),vjust=0, fontface = "bold",size=8)+
  expand_limits(y=0.5)+
  labs (y="Weighted UniFrac", x="")+
  scale_fill_discrete()
p.bar.wu

#jsd
#get p-values
i.jsd="jsd"
pseq.rel %>%
  phyloseq::distance(method = i.jsd) ->  dist.jsd

jsd.p<-mctoolsr::calc_pairwise_permanovas(dist.jsd, as(sample_data(pseq.rel), "data.frame"), "Village")

jsd = phyloseq::distance(pseq.rel, "jsd")
jsd.m = melt(as.matrix(jsd))
jsd.m = jsd.m %>%
  filter(as.character(Var1) != as.character(Var2)) %>%
  mutate_if(is.factor, as.character)

#village
sd.village <- data.frame(pseq.rel@sam_data$Sampleid, pseq.rel@sam_data$Village)
sd.village <- sd.village %>%
  rename(Village = pseq.rel.sam_data.Village)
sd.village <- sd.village %>%
  rename(Sampleid = pseq.rel.sam_data.Sampleid)
colnames(sd.village) = c("Var1", "Type1")
jsd.sd = left_join(jsd.m, sd.village, by = "Var1")
colnames(sd.village) = c("Var2", "Type2")
jsd.sd = left_join(jsd.sd, sd.village, by = "Var2")

#remove self from jsd.sd
jsd.sd = jsd.sd %>%
  filter(as.character(Type1) != as.character(Type2)) %>%
  mutate_if(is.factor, as.character)

#duinpos vs denui
duinpos.denui <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
duinpos.denui <- subset(duinpos.denui, duinpos.denui$Type2 == "Den/ui")
duinpos.denui.avg <- mean(duinpos.denui$value)

#duinpos vs MP
duinpos.MP <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
duinpos.MP <- subset(duinpos.MP, duinpos.MP$Type2 == "Mountain Pos")
duinpos.MP.avg <- mean (duinpos.MP$value)

#duinpos vs OM
duinpos.OM <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
duinpos.OM <- subset(duinpos.OM, duinpos.OM$Type2 == "!Om!o")
duinpos.OM.avg <- mean(duinpos.OM$value)

#denui vs MP
denui.MP <- subset(jsd.sd, jsd.sd$Type1 == "Den/ui")
```

```
denui.MP <- subset(duinpos.MP, denui.MP$Type2 == "Mountain Pos")
denui.MP.avg <- mean(duinpos.MP$value)

#denui vs OM
denui.OM <- subset(jsd.sd, jsd.sd$Type1 == "Den/ui")
denui.OM <- subset(denui.OM, denui.OM$Type2 == "!Om!o")
denui.OM.avg <- mean(denui.OM$value)

#MP vs OM
MP.OM <- subset(jsd.sd, jsd.sd$Type1 == "Mountain Pos")
MP.OM <- subset(MP.OM, MP.OM$Type2 == "!Om!o")
MP.OM.avg <- mean(MP.OM$value)

vals <- c(duinpos.denui.avg,duinpos.MP.avg,duinpos.OM.avg,denui.MP.avg,denui.OM.avg,MP.OM.avg)
col <- c("Duinpos vs Den/ui", "Duinpos vs Mountain Pos", "Duinpos vs !Om!o",
        "Den/ui vs Mountain Pos", "Den/ui vs !Om!o","Mountain Pos vs !Om!o")

village.df.jsd <- data.frame(col,vals)
village.df.jsd$FDRp <- jsd.p$pvalFDR
village.df.jsd$FDRp <- sub("^","p =", village.df.jsd$FDRp)

p.bar.jsd <-ggplot(village.df.jsd, aes(x=reorder(col,vals), y=vals,fill= col)) +
  geom_bar(stat = "identity") +
  theme_bw()+
  theme(axis.title = element_text(size = 30,face="bold"),
        axis.text.x = element_blank(),
        axis.text.y = element_text(size=25, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(face="bold",size=30),
        legend.position = "right",
        aspect.ratio = 3/3)+
  geom_text(aes(label=FDRp),vjust=0, fontface = "bold", size=8)+
  labs (y="Jensen-Shannon", x="")+
  scale_fill_discrete()
p.bar.jsd
p.final <-ggarrange(p.bar.wu, p.bar.jsd, nrow = 1, common.legend = T, labels = c("D","E"),
                font.label = list(size=35,face="bold"))
p.final
```

Figure 6F

```
#Village ----
outVillage = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
                 p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
                 group = "Village", struc_zero = F, neg_lb = F, tol = 1e-5,
                 max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_village = outVillage$res_global
#adj p values < 0.05
tab_q_vil = res_global_village[, "q_val", drop = FALSE]
tab_q_vil<-filter(tab_q_vil, q_val <= 0.05)
tab_q_vil <- rownames_to_column(tab_q_vil,var="Genus")
tab_q_vil$Genus <- gsub("g__","",tab_q_vil$Genus)
#run round_df function first
tab_q_vil <- round_df(tab_q_vil,5)
tab_q_vil$q_val <- sub("^","q = ",tab_q_vil$q_val)
```

```
write.csv(tab_q_vil,file="<file/path>ancombc_village.csv")

#remove controls, convert to relative abundance, aggregate to genus level
pseq.clean<-prune_samples(pseq.clean@sam_data$Sampleid!="KIT-CTRL",pseq.clean)
pseq.ra <- microbiome :: transform (pseq.clean,'compositional')
pseq.genus <- tax_glom(pseq.ra, "Genus", NArm=T)
#DA_genera will change depending on what was identified by ANCOM-BC
DA_genera <-
c("g__Bacteroides","g__Barnesiella","g__Odoribacter","g__Muribaculaceae","g__Paraprevotella","g__Prevotel
laceae_NK3B31_group","g__Alistipes","g__Gastranaerophilales","g__Desulfovibrio","g__Catenibacterium","g
__Solobacterium","g__RF39","g__Clostridia_UCG-
014","g__Clostridia_vadinBB60_group","g__[Eubacterium]_eligens_group","g__[Eubacterium]_xylanophilum
_group","g__[Ruminococcus]_gauvreauii_group","g__[Ruminococcus]_torques_group","g__Agathobacter","g_
_Blautia","g__Butyrivibrio","g__Dorea","g__Lachnospiraceae_AC2044_group","g__Lachnospiraceae_FCS020
_group","g__Lachnospiraceae_ND3007_group","g__Lachnospiraceae_NK4A136_group","g__Lachnospiraceae
_NK4B4_group","g__Lachnospiraceae_UCG-007","g__Lachnospiraceae_UCG-
010","g__Marvinbryantia","g__Roseburia","g__[Eubacterium]_coprostanoligenes_group","g__Butyricicoccus",
"g__Colidextribacter","g__NK4A214_group","g__UCG-
002","g__[Eubacterium]_siraeum_group","g__Candidatus_Soleaferrea","g__Faecalibacterium","g__Fournierell
a","g__Ruminococcaceae","g__Ruminococcus","g__UCG-
010","g__Intestinibacter","g__Romboutsia","g__Phascolarctobacterium","g__Sutterella")

#turn pseq into df
melt.pseq.genus <-psmelt(pseq.genus)
#subset to only include selected genera
melt.pseq.genus <- subset(melt.pseq.genus, Genus %in% DA_genera)
#clean and incorporate q-values from ancom-bc
melt.pseq.genus$Genus <- gsub("g__","",melt.pseq.genus$Genus)
melt.pseq.genus.final <- merge(melt.pseq.genus,tab_q_vil,by="Genus")
melt.pseq.genus.final$plot_text = paste(melt.pseq.genus.final$Genus," ", melt.pseq.genus.final$q_val)

#village
ggplot(data = melt.pseq.genus.final, aes(x = Village, y = Abundance)) +
  geom_boxplot(outlier.shape  = NA) +
  geom_jitter(aes(color = Genus), height = 0, width = .2, size=5) +
  labs(x = "", y = "Relative Abundance (%)\n") +
  facet_wrap(~ plot_text, scales = "free",ncol = 2)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.title = element_blank(),
        legend.position = "none",
        strip.text = element_text(size=20, face= "bold.italic"),
        plot.title = element_text(size=35, face="bold"))+
  scale_color_discrete(name="Differentially abundant genera")+
  ggtitle("F")
```

Differentially abundant genera

```
#Age ----
outAge = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
              p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
              group = "Age", struc_zero = F, neg_lb = F, tol = 1e-5,
              max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_age = outAge$res_global
#adj p values < 0.05
tab_q_age = res_global_age[, "q_val", drop = FALSE]
tab_q_age<-filter(tab_q_age, q_val <= 0.05)
```

```
tab_q_age <- rownames_to_column(tab_q_age,var="Genus")
tab_q_age$Genus <- gsub("g__","",tab_q_age$Genus)
#run round_df function first
tab_q_age <- round_df(tab_q_age,5)
tab_q_age$q_val <- sub("^","q = ",tab_q_age$q_val)

pseq.genus.age.2 = subset_samples(pseq.genus, Age != "Early-adult")
outAge2 = ancombc(phyloseq = pseq.genus.age.test,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
             group = "Age", struc_zero = F, neg_lb = F, tol = 1e-5,
             max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_age2 = outAge2$res_global
#adj p values < 0.05
tab_q_age2 = res_global_age2[, "q_val", drop = FALSE]
tab_q_age2<-filter(tab_q_age2, q_val <= 0.05)
write.csv(tab_q_age2,file="<file/path>ancombc_age_no_early-adults.csv")

#Antibiotics ----
#remove TDP06 and TDP04
pseq.genus.Ant = subset_samples(pseq.genus, Antibiotics != "Uncertain")
outAnt = ancombc(phyloseq = pseq.genus.Ant,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
            group = "Antibiotics", struc_zero = F, neg_lb = F, tol = 1e-5,
            max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resAnt = outAnt$res

#adjusted p
tab_q_ant = resAnt$q_val
tab_q_ant<- tab_q_ant["AntibioticsYes"]
tab_q_ant<-filter(tab_q_ant, AntibioticsYes <= 0.05)
tab_q_ant <- rownames_to_column(tab_q_ant,var="Genus")
tab_q_ant$Genus <- gsub("g__","",tab_q_ant$Genus)
#run round_df function first
tab_q_ant <- round_df(tab_q_ant,5)
tab_q_ant$q_val <- sub("^","q = ",tab_q_ant$AntibioticsYes)

write.csv(tab_q_ant,file="<file/path>ancombc_antibiotics.csv")

#Intestinal Infection ----
#remove TDE03, TDE08, TOM01
pseq.genus.Int = subset_samples(pseq.genus, Intestinal.Infections != "Uncertain")
outInt = ancombc(phyloseq = pseq.genus.Int,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
            group = "Intestinal.Infections", struc_zero = F, neg_lb = F, tol = 1e-5,
            max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resInt = outInt$res
#adjusted p
tab_q_int = resInt$q_val
tab_q_int<- tab_q_int["Intestinal.InfectionsYes"]
tab_q_int<-filter(tab_q_int, Intestinal.InfectionsYes <= 0.05)
tab_q_int <- rownames_to_column(tab_q_int,var="Genus")
tab_q_int$Genus <- gsub("g__","",tab_q_int$Genus)
```

```
#run round_df function first
tab_q_int <- round_df(tab_q_int,5)
tab_q_int$q_val <- sub("^","q = ",tab_q_int$Intestinal.InfectionsYes)
write.csv(tab_q_int,file="<file/path>ancombc_intestinal_infections.csv")

#Malaria.Medication ----
#remove Uncertain
pseq.genus.Mal = subset_samples(pseq.genus, Malaria.Medication != "Uncertain")
outMal = ancombc(phyloseq = pseq.genus.Mal,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
            group = "Malaria.Medication", struc_zero = F, neg_lb = F, tol = 1e-5,
            max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resMal = outMal$res

#adjusted p
tab_q_mal = resMal$q_val
tab_q_mal<- tab_q_mal["Malaria.MedicationYes"]
tab_q_mal<-filter(tab_q_mal, Malaria.MedicationYes <= 0.05)
tab_q_mal <- rownames_to_column(tab_q_mal,var="Genus")
tab_q_mal$Genus <- gsub("g__","",tab_q_mal$Genus)
#run round_df function first
tab_q_mal <- round_df(tab_q_mal,5)
tab_q_mal$q_val <- sub("^","q = ",tab_q_mal$Malaria.MedicationYes)
write.csv(tab_q_mal,file="<file/path>ancombc_malaria.csv")

#Travel ----
outTrav = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
             p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
             group = "Travel", struc_zero = F, neg_lb = F, tol = 1e-5,
             max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_trav = outTrav$res_global
#adj p values < 0.05
tab_q_trav = res_global_trav[, "q_val", drop = FALSE]
tab_q_trav<-filter(tab_q_trav, q_val <= 0.05)
tab_q_trav <- rownames_to_column(tab_q_trav,var="Genus")
tab_q_trav$Genus <- gsub("g__","",tab_q_trav$Genus)
#run round_df function first
tab_q_trav <- round_df(tab_q_trav,5)
tab_q_trav$q_val <- sub("^","q = ",tab_q_trav$q_val)
write.csv(tab_q_trav,file="<file/path>ancombc_travel.csv")

#Village ----
outVillage = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
               p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
               group = "Village", struc_zero = F, neg_lb = F, tol = 1e-5,
               max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_village = outVillage$res_global
#adj p values < 0.05
tab_q_vil = res_global_village[, "q_val", drop = FALSE]
tab_q_vil<-filter(tab_q_vil, q_val <= 0.05)
tab_q_vil <- rownames_to_column(tab_q_vil,var="Genus")
tab_q_vil$Genus <- gsub("g__","",tab_q_vil$Genus)
```

```
#run round_df function first
tab_q_vil <- round_df(tab_q_vil,5)
tab_q_vil$q_val <- sub("^","q = ",tab_q_vil$q_val)
write.csv(tab_q_vil,file="<file/path>ancombc_village.csv")
#Diarrhoea ----
outDiar = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
               p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
               group = "Diarrhoea", struc_zero = F, neg_lb = F, tol = 1e-5,
               max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resDiar = outDiar$res

#adjusted p
tab_q_diar = resDiar$q_val
tab_q_diar<- tab_q_diar["DiarrhoeaYes"]
tab_q_diar<-filter(tab_q_diar, "DiarrhoeaYes" <= 0.05)
tab_q_diar <- rownames_to_column(tab_q_diar,var="Genus")
tab_q_diar$Genus <- gsub("g__","",tab_q_diar$Genus)
#run round_df function first
tab_q_diar <- round_df(tab_q_diar,5)
tab_q_diar$q_val <- sub("^","q = ",tab_q_diar$q_val)
write.csv(tab_q_diar,file="<file/path>ancombc_diar.csv")

#Sex ----
outGen = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
               p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
               group = "Sex", struc_zero = F, neg_lb = F, tol = 1e-5,
               max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resGen = outGen$res
#adjusted p
tab_q_gen = resGen$q_val
tab_q_gen<- tab_q_gen["SexMale"]
tab_q_gen<-filter(tab_q_gen, SexMale <= 0.05)
write.csv(tab_q_gen,file="<file/path>ancombc_sex.csv")
```

ITS Community Composition

Figure 7A

```
#alpha div
#leave controls in analysis
p.alp=plot_richness(pseq.clean, color="Sampleid", measures=c("Simpson", "Shannon"),nrow = 3)
p.alp<-p.alp+ geom_point(size=10, alpha=0.7) +
  theme(strip.text = element_text(size=30, face= "bold"),
        legend.position = "none",
        axis.text = element_text(size = 30, face="bold"),
        plot.title = element_text(size=35,face="bold"))+
  ggtitle("A")+
  labs (x="",y="")

#relative abundance
pseq.rel <- microbiome::transform(pseq.clean, 'compositional')
pseq.rel<-prune_samples(pseq@sam_data$Sampleid!="CON-CTRL",pseq.rel)

#Bray-Curtis and jensen-shannon matrices
bc_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "bray")
```

```
jsd_matrix <- ordinate(pseq.rel, method = "NMDS", distance = "jsd")
```

Figure 7B and C

```
#Village
#for article
p.nmds.bc <- plot_ordination(pseq.rel, bc_matrix, color = "Village") + geom_point(size = 10)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.text = element_text(size = 35, face="bold"),
        legend.title = element_text(size=35, face="bold"),
        plot.title = element_text(size=35, face="bold"))+
  ggtitle("B")
p.nmds.bc
p.nmds.jsd <- plot_ordination(pseq.rel, jsd_matrix, color = "Village")+geom_point(size = 10)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.text = element_text(size = 35, face="bold"),
        legend.title = element_text(size=35, face="bold"),
        plot.title = element_text(size=35, face="bold"))+
  ggtitle("C")
p.nmds.jsd
```

SI Figure 3

```
#for Supplementary
vil1<- phyloseq::plot_ordination(pseq.rel, bc_matrix, color = "Village") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
vil2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Village") + geom_point(size = 10)
vil.leg <- get_legend(vil1)
vil <-ggarrange(vil1, vil2, vil.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis",”Jensen-Shannon"),
                font.label = list(size=30,face="bold"),
                widths = c(2,2,.15),
                label.x = .15,
                legend = "right")
#run annotate-figure

#Age
pseq.rel@sam_data$Age <- factor(pseq.rel@sam_data$Age, levels = c("Late-adolescent","Early-adult","Middle-
adult","Late-adult"))
age1 <- phyloseq::plot_ordination(pseq.rel, bc_matrix, color = "Age") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
age2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Age") + geom_point(size = 10)
leg <- get_legend(age1)
age <-ggarrange(age1, age2, leg,nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
                font.label = list(size=35,face="bold"),
                legend = "right",
                widths = c(2,2,.15),
                label.x = .1)
#run annotate-figure

#Antibiotics
#remove uncertain subjects
pseq.rel.ant = subset_samples(pseq.rel, Antibiotics != "Uncertain")
pseq.rel.ant<-prune_samples(pseq.rel.ant@sam_data$Sampleid!="CON-CTRL",pseq.rel.ant)
```

```
pseq.rel.ant@sam_data$Antibiotics <- factor(pseq.rel.ant@sam_data$Antibiotics, levels = c("Yes","No"))
bc_matrix.ant <- ordinate(pseq.rel.ant, method = "NMDS", distance = "bray")
jsd_matrix.ant <- ordinate(pseq.rel.ant, method = "NMDS", distance = "jsd")
ant1 <- phyloseq::plot_ordination(pseq.rel.ant, bc_matrix.ant, color = "Antibiotics") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
ant2 <- phyloseq::plot_ordination(pseq.rel.ant, jsd_matrix.ant, color = "Antibiotics") + geom_point(size = 10)
ant.leg <- get_legend(ant1)
ant <-ggarrange(ant1, ant2, ant.leg,nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
                font.label = list(size=35,face="bold"),
                widths = c(2,2,.15),
                label.x = .25,
                legend = "right")
#run annotate-figure

#Biological sex
sex1 <- phyloseq::plot_ordination(pseq.rel, bc_matrix, color = "Sex") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
sex2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Sex") + geom_point(size = 10)
sex.leg <- get_legend(sex1)
sex <-ggarrange(sex1, sex2, sex.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
                font.label = list(size=30,face="bold"),
                widths = c(2,2,.15),
                label.x = .4,
                legend = "right")
#run annotate-figure

#Diarrhoea
diar1 <- phyloseq::plot_ordination(pseq.rel, bc_matrix, color = "Diarrhoea") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
diar2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Diarrhoea") + geom_point(size = 10)
diar.leg <- get_legend(diar1)
diar <-ggarrange(diar1, diar2, diar.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
                 font.label = list(size=30,face="bold"),
                 widths = c(2,2,.15),
                 label.x = .3,
                 legend = "right")
#run annotate-figure

#Intestinal Infections
#remove uncertain subjects
pseq.rel.int = subset_samples(pseq.rel, Intestinal.Infections != "Uncertain")
pseq.rel.int<-prune_samples(pseq.rel.int@sam_data$Sampleid!="CON-CTRL",pseq.rel.int)
pseq.rel.int@sam_data$Intestinal.Infections <- factor(pseq.rel.int@sam_data$Intestinal.Infections, levels =
c("Yes","No"))
bc_matrix.int <- ordinate(pseq.rel.int, method = "NMDS", distance = "bray")
jsd_matrix.int <- ordinate(pseq.rel.int, method = "NMDS", distance = "jsd")
int1<- phyloseq::plot_ordination(pseq.rel.int, bc_matrix.int, color = "Intestinal.Infections") + geom_point(size =
10)+
  scale_color_discrete("Intestinal Infections")+
  theme(legend.text=element_text(size=30,face="bold"),
        legend.title = element_text(size=30,face="bold"))
int2<- phyloseq::plot_ordination(pseq.rel.int, jsd_matrix.int, color = "Intestinal.Infections") + geom_point(size =
10)+
```

```
  scale_colour_discrete("Intestinal Infections")
int.leg <- get_legend(int1)
int <-ggarrange(int1, int2, int.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
          font.label = list(size=30,face="bold"),
          widths = c(2,2,.15),
          label.x = .3,
          legend = "right")
#run annotate-figure

#Malaria Medication
#remove uncertain subjects
pseq.rel.mal = subset_samples(pseq.rel, Malaria.Medication != "Uncertain")
pseq.rel.mal<-prune_samples(pseq.rel.mal@sam_data$Sampleid!="CON-CTRL",pseq.rel.mal)
pseq.rel.mal@sam_data$Malaria.Medication <- factor(pseq.rel.mal@sam_data$Malaria.Medication, levels = c("Yes","No"))
bc_matrix.mal <- ordinate(pseq.rel.mal, method = "NMDS", distance = "bray")
jsd_matrix.mal <- ordinate(pseq.rel.mal, method = "NMDS", distance = "jsd")

mal1<- phyloseq::plot_ordination(pseq.rel.mal, bc_matrix.mal, color = "Malaria.Medication") + geom_point(size = 10)+
  scale_color_discrete("Malaria Medication")+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
mal2<- phyloseq::plot_ordination(pseq.rel.mal, jsd_matrix.mal, color = "Malaria.Medication") + geom_point(size = 10)+
  scale_colour_discrete("Malaria Medication")
mal.leg <- get_legend(mal1)
mal <-ggarrange(mal1, mal2, mal.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis"Jensen-Shannon"),
          font.label = list(size=30,face="bold"),
          widths = c(2,2,.15),
          label.x = .15,
          legend = "right")
#run annotate-figure

#Travel
trav1<- phyloseq::plot_ordination(pseq.rel, bc_matrix, color = "Travel") + geom_point(size = 10)+
  theme(legend.text=element_text(size=30,face="bold"),
      legend.title = element_text(size=30,face="bold"))
trav2<-phyloseq::plot_ordination(pseq.rel, jsd_matrix, color = "Travel") + geom_point(size = 10)
trav.leg <- get_legend(trav1)
trav <-ggarrange(trav1, trav2, trav.leg, nrow = 1,common.legend=T,labels = c("Bray-Curtis","Jensen-Shannon"),
           font.label = list(size=30,face="bold"),
           widths = c(2,2,.15),
           label.x = .15,
           legend = "right")
#run annotate-figure
```

Figure 7 D and E

```
#barplots of village average distances ----
#first get p-values using pairwise adonis
i.bray="bray"
pseq.rel %>%
  phyloseq::distance(method = i.bray) ->  dist.bray
bray.p<-mctoolsr::calc_pairwise_permanovas(dist.bray, as(sample_data(pseq.rel), "data.frame"), "Village")
bc = phyloseq::distance(pseq.rel, "bray")
bc.m = melt(as.matrix(bc))
```

```
bc.m = bc.m %>%
  filter(as.character(Var1) != as.character(Var2)) %>%
  mutate_if(is.factor, as.character)

#village
sd.village <- data.frame(pseq.rel@sam_data$Sampleid, pseq.rel@sam_data$Village)
sd.village <- sd.village %>%
  rename(Village = pseq.rel.sam_data.Village)
sd.village <- sd.village %>%
  rename(Sampleid = pseq.rel.sam_data.Sampleid)
colnames(sd.village) = c("Var1", "Type1")
bc.sd = left_join(bc.m, sd.village, by = "Var1")
colnames(sd.village) = c("Var2", "Type2")
bc.sd = left_join(bc.sd, sd.village, by = "Var2")

#remove self from bc.sd
bc.sd = bc.sd %>%
  filter(as.character(Type1) != as.character(Type2)) %>%
  mutate_if(is.factor, as.character)

#duinpos vs denui
duinpos.denui <- subset(bc.sd, bc.sd$Type1 == "Duinpos")
duinpos.denui <- subset(duinpos.denui, duinpos.denui$Type2 == "Den/ui")
duinpos.denui.avg <- mean(duinpos.denui$value)

#duinpos vs MP
duinpos.MP <- subset(bc.sd, bc.sd$Type1 == "Duinpos")
duinpos.MP <- subset(duinpos.MP, duinpos.MP$Type2 == "Mountain Pos")
duinpos.MP.avg <- mean (duinpos.MP$value)

#duinpos vs OM
duinpos.OM <- subset(bc.sd, bc.sd$Type1 == "Duinpos")
duinpos.OM <- subset(duinpos.OM, duinpos.OM$Type2 == "!Om!o")
duinpos.OM.avg <- mean(duinpos.OM$value)

#denui vs MP
denui.MP <- subset(bc.sd, bc.sd$Type1 == "Den/ui")
denui.MP <- subset(duinpos.MP, denui.MP$Type2 == "Mountain Pos")
denui.MP.avg <- mean(duinpos.MP$value)

#denui vs OM
denui.OM <- subset(bc.sd, bc.sd$Type1 == "Den/ui")
denui.OM <- subset(denui.OM, denui.OM$Type2 == "!Om!o")
denui.OM.avg <- mean(denui.OM$value)

#MP vs OM
MP.OM <- subset(bc.sd, bc.sd$Type1 == "Mountain Pos")
MP.OM <- subset(MP.OM, MP.OM$Type2 == "!Om!o")
MP.OM.avg <- mean(MP.OM$value)

vals <- c(duinpos.denui.avg,duinpos.MP.avg,duinpos.OM.avg,denui.MP.avg,denui.OM.avg,MP.OM.avg)
col <- c("Duinpos vs Den/ui", "Duinpos vs Mountain Pos", "Duinpos vs !Om!o",
       "Den/ui vs Mountain Pos", "Den/ui vs !Om!o","Mountain Pos vs !Om!o")
village.df.bc <- data.frame(col,vals)
village.df.bc$FDRp <- bray.p$pvalFDR
```

```
village.df.bc$FDRp <- sub("^","p =", village.df.bc$FDRp)
p.bar.bc<-ggplot(village.df.bc, aes(x=reorder(col,vals), y=vals,fill= col)) +
  geom_bar(stat = "identity") +
  theme_bw()+
  theme(axis.title = element_text(size = 30,face="bold"),
        axis.text.x = element_blank(),
        axis.text.y = element_text(size=25, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(face="bold",size=30),
        legend.position = "right",
        aspect.ratio = 5/5)+
  geom_text(aes(label=FDRp),vjust=0, fontface = "bold",size=8)+
  expand_limits(y=0.5)+
  labs (y="Bray-Curtis", x="")+
  scale_fill_discrete()
p.bar.bc

#jsd
i.jsd="jsd"
pseq.rel %>%
  phyloseq::distance(method = i.jsd) ->  dist.jsd
jsd.p<-mctoolsr::calc_pairwise_permanovas(dist.jsd, as(sample_data(pseq.rel), "data.frame"), "Village")
jsd = phyloseq::distance(pseq.rel, "jsd")
jsd.m = melt(as.matrix(jsd))
jsd.m = jsd.m %>%
  filter(as.character(Var1) != as.character(Var2)) %>%
  mutate_if(is.factor, as.character)
#village
sd.village <- data.frame(pseq.rel@sam_data$Sampleid, pseq.rel@sam_data$Village)
sd.village <- sd.village %>%
  rename(Village = pseq.rel.sam_data.Village)
sd.village <- sd.village %>%
  rename(Sampleid = pseq.rel.sam_data.Sampleid)
colnames(sd.village) = c("Var1", "Type1")
jsd.sd = left_join(jsd.m, sd.village, by = "Var1")
colnames(sd.village) = c("Var2", "Type2")
jsd.sd = left_join(jsd.sd, sd.village, by = "Var2")

#remove self from jsd.sd
jsd.sd = jsd.sd %>%
  filter(as.character(Type1) != as.character(Type2)) %>%
  mutate_if(is.factor, as.character)

#duinpos vs denui
duinpos.denui <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
duinpos.denui <- subset(duinpos.denui, duinpos.denui$Type2 == "Den/ui")
duinpos.denui.avg <- mean(duinpos.denui$value)

#duinpos vs MP
duinpos.MP <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
duinpos.MP <- subset(duinpos.MP, duinpos.MP$Type2 == "Mountain Pos")
duinpos.MP.avg <- mean (duinpos.MP$value)

#duinpos vs OM
duinpos.OM <- subset(jsd.sd, jsd.sd$Type1 == "Duinpos")
```

```
duinpos.OM <- subset(duinpos.OM, duinpos.OM$Type2 == "!Om!o")
duinpos.OM.avg <- mean(duinpos.OM$value)

#denui vs MP
denui.MP <- subset(jsd.sd, jsd.sd$Type1 == "Den/ui")
denui.MP <- subset(duinpos.MP, denui.MP$Type2 == "Mountain Pos")
denui.MP.avg <- mean(duinpos.MP$value)

#denui vs OM
denui.OM <- subset(jsd.sd, jsd.sd$Type1 == "Den/ui")
denui.OM <- subset(denui.OM, denui.OM$Type2 == "!Om!o")
denui.OM.avg <- mean(denui.OM$value)

#MP vs OM
MP.OM <- subset(jsd.sd, jsd.sd$Type1 == "Mountain Pos")
MP.OM <- subset(MP.OM, MP.OM$Type2 == "!Om!o")
MP.OM.avg <- mean(MP.OM$value)

vals <- c(duinpos.denui.avg,duinpos.MP.avg,duinpos.OM.avg,denui.MP.avg,denui.OM.avg,MP.OM.avg)
col <- c("Duinpos vs Den/ui", "Duinpos vs Mountain Pos", "Duinpos vs !Om!o",
       "Den/ui vs Mountain Pos", "Den/ui vs !Om!o","Mountain Pos vs !Om!o")
village.df.jsd <- data.frame(col,vals)
village.df.jsd$FDRp <- jsd.p$pvalFDR
village.df.jsd$FDRp <- sub("^","p =", village.df.jsd$FDRp)

p.bar.jsd <-ggplot(village.df.jsd, aes(x=reorder(col,vals), y=vals,fill= col)) +
  geom_bar(stat = "identity") +
  theme_bw()+
  theme(axis.title = element_text(size = 30,face="bold"),
        axis.text.x = element_blank(),
        axis.text.y = element_text(size=25, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(face="bold",size=30),
        legend.position = "right",
        aspect.ratio = 3/3)+
  geom_text(aes(label=FDRp),vjust=0, fontface = "bold", size=8)+
  labs (y="Jensen-Shannon", x="")+
  scale_fill_discrete()
p.bar.jsd
p.final <-ggarrange(p.bar.bc, p.bar.jsd, nrow = 1, common.legend = T, labels = c("D","E"),
                font.label = list(size=35,face="bold"))
p.final
```

Figure 7F

```
#Remove KIT-CTRL
pseq.clean <- prune_samples(pseq.clean@sam_data$Sampleid!="CON-CTRL",pseq.clean)

#2. Aggregate to genus level and remove uncultured and unknown genera
pseq.genus <- aggregate_taxa(pseq.clean,"Genus")
pseq.genus <- subset_taxa(pseq.genus, Genus != "g__uncultured")
pseq.genus <- subset_taxa(pseq.genus, Genus != "Unknown")
#Village ----
outVillage = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
                 p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
```

```
                group = "Village", struc_zero = F, neg_lb = F, tol = 1e-5,
                max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_village = outVillage$res_global
#adj p values < 0.05
tab_q_vil = res_global_village[, "q_val", drop = FALSE]
tab_q_vil<-filter(tab_q_vil, q_val <= 0.05)
tab_q_vil <- rownames_to_column(tab_q_vil,var="Genus")
tab_q_vil$Genus <- gsub("g__","",tab_q_vil$Genus)
#run round_df function first
tab_q_vil <- round_df(tab_q_vil,5)
tab_q_vil$q_val <- sub("^","q = ",tab_q_vil$q_val)
write.csv(tab_q_vil,file="<file/path>/ancombc_village.csv")

#run 1-create-phyloseq-decontam-ITS.R, then remove controls, convert to relative abundance and aggregate to genus
pseq.clean<-prune_samples(pseq.clean@sam_data$Sampleid!="CON-CTRL",pseq.clean)
pseq.ra <- microbiome :: transform (pseq.clean,'compositional')
pseq.genus <- tax_glom(pseq.ra, "Genus", NArm=T)
#DA_genera identified by ANCOM-BC
DA_genera<-
c("g__Stagonospora","g__Aspergillus","g__Saccharomyces","g__Candida","g__Porodisculus","g__Schizophyllum","g__Panellus","g__Mycoacia")
#turn pseq into df
melt.pseq.genus <-psmelt(pseq.genus)
#subset to only include selected genera
melt.pseq.genus <- subset(melt.pseq.genus, Genus %in% DA_genera)
#clean
melt.pseq.genus$Genus <- gsub("g__","",melt.pseq.genus$Genus)
melt.pseq.genus.final <- merge(melt.pseq.genus,tab_q_vil,by="Genus")
melt.pseq.genus.final$plot_text = paste(melt.pseq.genus.final$Genus," ", melt.pseq.genus.final$q_val)
#village
vil.da <- ggplot(data = melt.pseq.genus.final, aes(x = Village, y = Abundance)) +
  geom_boxplot(outlier.shape  = NA) +
  geom_jitter(aes(color = Genus), height = 0, width = .2, size=5) +
  labs(x = "", y = "Relative Abundance (%)\n") +
  facet_wrap(~ plot_text, scales = "free",ncol = 2)+
  theme(axis.title = element_text(size = 20, face = "bold"),
        axis.text = element_text(size = 20, face="bold"),
        legend.title = element_blank(),
        legend.position = "none",
        strip.text = element_text(size=20, face= "bold.italic"),
        plot.title = element_text(size=35, face="bold"))+
  scale_color_discrete(name="Differentially abundant genera")+
  ggtitle("F")

#Age ----
outAge = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
            Intestinal.Infections+Diarrhoea+Malaria.Medication",
                p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
                group = "Age", struc_zero = F, neg_lb = F, tol = 1e-5,
                max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_age = outAge$res_global
#adj p values < 0.05
tab_q_age = res_global_age[, "q_val", drop = FALSE]
tab_q_age<-filter(tab_q_age, q_val <= 0.05)
tab_q_age <- rownames_to_column(tab_q_age,var="Genus")
```

```
tab_q_age$Genus <- gsub("g__","",tab_q_age$Genus)
#run round_df function first
tab_q_age <- round_df(tab_q_age,5)
tab_q_age$q_val <- sub("^","q = ",tab_q_age$q_val)
write.csv(tab_q_age,file="<file/path>/ancombc_age.csv")

pseq.genus.age.2 = subset_samples(pseq.genus, Age != "Early-adult")
outAge2 = ancombc(phyloseq = pseq.genus.age.2,  formula = "Antibiotics+Sex+Travel+Age+Village+
        Intestinal.Infections+Diarrhoea+Malaria.Medication",
           p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
           group = "Age", struc_zero = F, neg_lb = F, tol = 1e-5,
           max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_age2 = outAge2$res_global
#adj p values < 0.05
tab_q_age2 = res_global_age2[, "q_val", drop = FALSE]
tab_q_age2<-filter(tab_q_age2, q_val <= 0.05)
tab_q_age2 <- rownames_to_column(tab_q_age2,var="Genus")
tab_q_age2$Genus <- gsub("g__","",tab_q_age2$Genus)
#run round_df function first
tab_q_age2 <- round_df(tab_q_age2,5)
tab_q_age2$q_val <- sub("^","q = ",tab_q_age2$q_val)
write.csv(tab_q_age2,file="<file/path>/ancombc_age_no_early-adults.csv")

#Sex ----
outGen = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
        Intestinal.Infections+Diarrhoea+Malaria.Medication",
           p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
           group = "Sex", struc_zero = F, neg_lb = F, tol = 1e-5,
           max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resGen = outGen$res

#adjusted p
tab_q_gen = resGen$q_val
tab_q_gen<- tab_q_gen["SexMale"]
tab_q_gen<-filter(tab_q_gen, SexMale <= 0.05)
tab_q_gen <- rownames_to_column(tab_q_gen,var="Genus")
tab_q_gen$Genus <- gsub("g__","",tab_q_gen$Genus)
#run round_df function first
tab_q_gen <- round_df(tab_q_gen,5)
tab_q_gen$q_val <- sub("^","q = ",tab_q_gen$SexMale)
write.csv(tab_q_gen,file="<file/path>/ancombc_sex.csv")

#Intestinal Infection ----
#remove TDE03, TDE08, TOM01
pseq.genus.Int = subset_samples(pseq.genus, Intestinal.Infections != "Uncertain")
outInt = ancombc(phyloseq = pseq.genus.Int,  formula = "Antibiotics+Sex+Travel+Age+Village+
        Intestinal.Infections+Diarrhoea+Malaria.Medication",
           p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
           group = "Intestinal.Infections", struc_zero = F, neg_lb = F, tol = 1e-5,
           max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resInt = outInt$res

#adjusted p
tab_q_int = resInt$q_val
tab_q_int<- tab_q_int["Intestinal.InfectionsYes"]
```

```
tab_q_int<-filter(tab_q_int, Intestinal.InfectionsYes <= 0.05)
tab_q_int <- rownames_to_column(tab_q_int,var="Genus")
tab_q_int$Genus <- gsub("g__","",tab_q_int$Genus)
#run round_df function first
tab_q_int <- round_df(tab_q_int,5)
tab_q_int$q_val <- sub("^","q = ",tab_q_int$Intestinal.InfectionsYes)
write.csv(tab_q_int,file="<file/path>/ancombc_intestinal_infections.csv")

#Malaria.Medication ----
#remove Uncertain
pseq.genus.Mal = subset_samples(pseq.genus, Malaria.Medication != "Uncertain")
outMal = ancombc(phyloseq = pseq.genus.Mal,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
            group = "Malaria.Medication", struc_zero = F, neg_lb = F, tol = 1e-5,
            max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resMal = outMal$res

#adjusted p
tab_q_mal = resMal$q_val
tab_q_mal<- tab_q_mal["Malaria.MedicationYes"]
tab_q_mal<-filter(tab_q_mal, Malaria.MedicationYes <= 0.05)
tab_q_mal <- rownames_to_column(tab_q_mal,var="Genus")
tab_q_mal$Genus <- gsub("g__","",tab_q_mal$Genus)
#run round_df function first
tab_q_mal <- round_df(tab_q_mal,5)
tab_q_mal$q_val <- sub("^","q = ",tab_q_mal$Malaria.MedicationYes)
write.csv(tab_q_mal,file="<file/path>/ancombc_malaria.csv")

#Travel ----
outTrav = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
            p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
            group = "Travel", struc_zero = F, neg_lb = F, tol = 1e-5,
            max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_trav = outTrav$res_global

#adj p values < 0.05
tab_q_trav = res_global_trav[, "q_val", drop = FALSE]
tab_q_trav<-filter(tab_q_trav, q_val <= 0.05)
tab_q_trav <- rownames_to_column(tab_q_trav,var="Genus")
tab_q_trav$Genus <- gsub("g__","",tab_q_trav$Genus)
#run round_df function first
tab_q_trav <- round_df(tab_q_trav,5)
tab_q_trav$q_val <- sub("^","q = ",tab_q_trav$q_val)
write.csv(tab_q_trav,file="<file/path>/ancombc_travel.csv")


#Village ----
outVillage = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
              p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
              group = "Village", struc_zero = F, neg_lb = F, tol = 1e-5,
              max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
res_global_village = outVillage$res_global
```

```
#adj p values < 0.05
tab_q_vil = res_global_village[, "q_val", drop = FALSE]
tab_q_vil<-filter(tab_q_vil, q_val <= 0.05)
tab_q_vil <- rownames_to_column(tab_q_vil,var="Genus")
tab_q_vil$Genus <- gsub("g__","",tab_q_vil$Genus)
#run round_df function first
tab_q_vil <- round_df(tab_q_vil,5)
tab_q_vil$q_val <- sub("^","q = ",tab_q_vil$q_val)
write.csv(tab_q_vil,file="<file/path>/ancombc_village.csv")

#Diarrhoea ----
outDiar = ancombc(phyloseq = pseq.genus,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
             p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
             group = "Diarrhoea", struc_zero = F, neg_lb = F, tol = 1e-5,
             max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resDiar = outDiar$res

#adjusted p
tab_q_diar = resDiar$q_val
tab_q_diar<- tab_q_diar["DiarrhoeaYes"]
tab_q_diar<-filter(tab_q_diar, "DiarrhoeaYes" <= 0.05)
tab_q_diar <- rownames_to_column(tab_q_diar,var="Genus")
tab_q_diar$Genus <- gsub("g__","",tab_q_diar$Genus)
#run round_df function first
tab_q_diar <- round_df(tab_q_diar,5)
tab_q_diar$q_val <- sub("^","q = ",tab_q_diar$q_val)
write.csv(tab_q_diar,file="<file/path>/ancombc_diar.csv")

#Antibiotics ----
#remove TDP06 and TDP04
pseq.genus.Ant = subset_samples(pseq.genus, Antibiotics != "Uncertain")
outAnt = ancombc(phyloseq = pseq.genus.Ant,  formula = "Antibiotics+Sex+Travel+Age+Village+
          Intestinal.Infections+Diarrhoea+Malaria.Medication",
             p_adj_method = "BH", zero_cut = 1, lib_cut = 0,
             group = "Antibiotics", struc_zero = F, neg_lb = F, tol = 1e-5,
             max_iter = 100, conserve = TRUE, alpha = 0.05, global = TRUE)
resAnt = outAnt$res

#adjusted p
tab_q_ant = resAnt$q_val
tab_q_ant<- tab_q_ant["AntibioticsYes"]
tab_q_ant<-filter(tab_q_ant, AntibioticsYes <= 0.05)
tab_q_ant <- rownames_to_column(tab_q_ant,var="Genus")
tab_q_ant$Genus <- gsub("g__","",tab_q_ant$Genus)
#run round_df function first
tab_q_ant <- round_df(tab_q_ant,5)
tab_q_ant$q_val <- sub("^","q = ",tab_q_ant$AntibioticsYes)
write.csv(tab_q_ant,file="<file/path>/ancombc_antibiotics.csv")
```

References

Adler, C.J. *et al.* (2013) "Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions," *Nature Genetics 2013 45:4*, 45(4), pp. 450–455. doi:10.1038/ng.2536.

Aitchison, J. (1982) "The Statistical Analysis of Compositional Data," *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2), pp. 139–160. doi:10.1111/J.2517-6161.1982.TB01195.X.

Almeida, A. *et al.* (2020) "A unified catalog of 204,938 reference genomes from the human gut microbiome," *Nature Biotechnology 2020 39:1*, 39(1), pp. 105–114. doi:10.1038/s41587-020-0603-3.

Amato, K.R. *et al.* (2019) "Shifting Climates, Foods, and Diseases: The Human Microbiome through Evolution," *BioEssays*, 41(10), pp. 1–9. doi:10.1002/bies.201900034.

Amato, K.R. *et al.* (2021) "The human gut microbiome and health inequities," *Proceedings of the National Academy of Sciences*, 118(25), p. 2017947118. doi:10.1073/PNAS.2017947118.

Amir, A. *et al.* (2017) "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns," *mSystems*, 2(2). doi:10.1128/MSYSTEMS.00191-16.

Anderson, I. *et al.* (2016) "Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): a population study," *The Lancet*, 388(10040), pp. 131–157. doi:10.1016/S0140-6736(16)00345-7.

Angebault, C. *et al.* (2018) "Combined bacterial and fungal intestinal microbiota analyses: Impact of storage conditions and DNA extraction protocols," *PLOS ONE*, 13(8), p. e0201174. doi:10.1371/JOURNAL.PONE.0201174.

Angelakis, E. *et al.* (2019) "Treponema species enrich the gut microbiota of traditional rural populations but are absent from urban individuals," *New Microbes and New Infections* [Preprint]. doi:10.1016/j.nmni.2018.10.009.

Ayeni, F.A. *et al.* (2018) "Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria," *Cell Reports*, 23(10), pp. 3056–3067. doi:10.1016/j.celrep.2018.05.018.

Bager, P., Wohlfahrt, J. and Westergaard, T. (2008) "Caesarean delivery and risk of atopy and allergic disesase: Meta-analyses," *Clinical and Experimental Allergy*, 38(4), pp. 634–642. doi:10.1111/j.1365-2222.2008.02939.x.

Bailén, M. *et al.* (2020) "Bioinformatic strategies to address limitations of 16rRNA short-read amplicons from different sequencing platforms," *Journal of Microbiological Methods*, 169, p. 105811. doi:10.1016/J.MIMET.2019.105811.

Behjati, S. and Tarpey, P.S. (2013) "What is next generation sequencing?," *Archives of Disease in Childhood. Education and Practice Edition*, 98(6), p. 236. doi:10.1136/ARCHDISCHILD-2013-304340.

Bentley, G.R., Goldberg, T. and Jasielqska, G. (2010) "The Fertility of Agricultural and Non-Agricultural Traditional Societies," *http://dx.doi.org/10.1080/00324720310001470 06*, 47(2), pp. 269–281. doi:10.1080/0032472031000147006.

Bharti, R. and Grimm, D.G. (2021) "Current challenges and best-practice protocols for microbiome analysis," *Briefings in Bioinformatics*, 22(1), pp. 178–193. doi:10.1093/BIB/BBZ155.

Bik, E.M. (2016) "The hoops, hopes, and hypes of human microbiome research," *Yale Journal of Biology and Medicine*, 89(3), pp. 363–373.

Blaser, M.J. and Falkow, S. (2009) *What are the consequences of the disappearing human microbiota?*, *Nature Reviews Microbiology 2009 7:12*. Nature Publishing Group. doi:10.1038/nrmicro2245.

Blekhman, R. *et al.* (2015) “Host genetic variation impacts microbiome composition across human body sites,” *Genome Biology 2015 16:1*, 16(1), pp. 1–12. doi:10.1186/S13059-015-0759-1.

Bolyen, E. *et al.* (2019) “Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2,” *Nature Biotechnology*. Nature Publishing Group, pp. 852–857. doi:10.1038/s41587-019-0209-9.

Bordenstein, S.R. and Theis, K.R. (2015) “Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes,” *PLOS Biology*, 13(8), p. e1002226. doi:10.1371/JOURNAL.PBIO.1002226.

Bowyer, R.C.E. *et al.* (2019) “Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study,” *Microorganisms 2019, Vol. 7, Page 17*, 7(1), p. 17. doi:10.3390/MICROORGANISMS7010017.

Brand-Miller, J.C., Mann, N.J. and Cordain, L. (2018) “Paleolithic nutrition:what did our ancestors eat? .”

Buford, T.W. (2017) “(Dis)Trust your gut: the gut microbiome in age-related inflammation, health, and disease,” *Microbiome 2017 5:1*, 5(1), pp. 1–11. doi:10.1186/S40168-017-0296-0.

Cabrera-Rubio, R. *et al.* (2012) “The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery,” *The American Journal of Clinical Nutrition*, 96(3), pp. 544–551. doi:10.3945/AJCN.112.037382.

Callahan, B.J. *et al.* (2016) “DADA2: High-resolution sample inference from Illumina amplicon data,” *Nature Methods*, 13(7), pp. 581–583. doi:10.1038/nmeth.3869.

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

Cando-Dumancela, C. *et al.* (2021) “A guide to minimize contamination issues in microbiome restoration studies,” *Restoration Ecology*, 29(4), p. e13358. doi:10.1111/REC.13358.

Carlet, J. (2012) “The gut is the epicentre of antibiotic resistance,” *Antimicrobial Resistance and Infection Control*. BioMed Central Ltd., pp. 1–7. doi:10.1186/2047-2994-1-39.

Carrera-Bastos, P. *et al.* (2011) “The western diet and lifestyle and diseases of civilization,” *Article in Research Reports in Clinical Cardiology* [Preprint]. doi:10.2147/RRCC.S16919.

Carson, S.L. *et al.* (2019) “Health Challenges and Assets of Forest-Dependent Populations in Cameroon,” *EcoHealth 2019 16:2*, 16(2), pp. 287–297. doi:10.1007/S10393-019-01411-9.

Casals-Pascual, C., Vergara, A. and Vila, J. (2018) “Intestinal microbiota and antibiotic resistance: Perspectives and solutions,” *Human Microbiome Journal*, 9, pp. 11–15. doi:10.1016/J.HUMIC.2018.05.002.

Chen, E. and Miller, G.E. (2013) “Socioeconomic Status and Health: Mediating and Moderating Factors.” doi:10.1146/annurev-clinpsy-050212-185634.

Chen, L. *et al.* (2018) “GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data,” *PeerJ*, 6(4), p. e4600. doi:10.7717/PEERJ.4600.

Chen, Y. *et al.* (2011) “Correlation between gastrointestinal fungi and varying degrees of chronic hepatitis B virus infection,” *Diagnostic Microbiology and Infectious Disease*, 70(4), pp. 492–498. doi:10.1016/J.DIAGMICROBIO.2010.04.005.

Chin, V.K. *et al.* (2020) “Mycobiome in the Gut: A Multiperspective Review,” *Mediators of Inflammation*, 2020. doi:10.1155/2020/9560684.

Clark, L.P., Millet, D.B. and Marshall, J.D. (2017) “Changes in Transportation-Related Air Pollution Exposures by Race-Ethnicity and Socioeconomic Status: Outdoor Nitrogen Dioxide in the United States in 2000 and 2010,” *Environmental Health Perspectives*, 125(9). doi:10.1289/EHP959.

Clemente, J.C. *et al.* (2015) "The microbiome of uncontacted Amerindians," *Science Advances*, 1(3), p. e1500183. doi:10.1126/sciadv.1500183.

Cong, X. *et al.* (2016) "Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender," *PLOS ONE*, 11(4), p. e0152751. doi:10.1371/JOURNAL.PONE.0152751.

Cordain, L. *et al.* (2005) "Origins and evolution of the Western diet: health implications for the 21st century," *The American Journal of Clinical Nutrition*, 81(2), pp. 341–354. doi:10.1093/AJCN.81.2.341.

Cordain, L. (2006) *Evolution of the Human Diet: The Known, the Unknown, and the Unknowable* . Edited by P.S. Ungar.

Cottier, F. and Pavelka, N. (2012) "Complexity and dynamics of host–fungal interactions," *Immunologic Research 2012 53:1*, 53(1), pp. 127–135. doi:10.1007/S12026-012-8265-Y.

Crittenden, A.N. and Schnorr, S.L. (2017) "Current views on hunter-gatherer nutrition and the evolution of the human diet," *American Journal of Physical Anthropology*, 162(S63), pp. 84–109. doi:10.1002/ajpa.23148.

Cross, T.W.L., Kasahara, K. and Rey, F.E. (2018) "Sexual dimorphism of cardiometabolic dysfunction: Gut microbiome in the play?," *Molecular Metabolism*, 15, pp. 70–81. doi:10.1016/J.MOLMET.2018.05.016.

Dambuza, I.M. and Brown, G.D. (2015) "C-type lectins in immunity: recent developments," *Current Opinion in Immunology*, 32, pp. 21–27. doi:10.1016/J.COI.2014.12.002.

Dao, M.C. *et al.* (2016) "Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology," *Gut*, 65(3), pp. 426–436. doi:10.1136/GUTJNL-2014-308778.

Davis, C.D. (2016) "The gut microbiome and its role in obesity," *Nutrition Today*, 51(4), pp. 167–174. doi:10.1097/NT.0000000000000167.

Davis, N.M. *et al.* (2017) "Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data," *bioRxiv*, p. 221499. doi:10.1101/221499.

Denker, H., Thompson, S. and Jarvis, A. (2012) *Living with wildlife-Living with wildlife Nyae Nyae Conservancy. The story of Nyae Nyae-Ju/'hoansi for "the place without mountains, but rocky."*

Dethlefsen, L., McFall-Ngai, M. and Relman, D.A. (2007) "An ecological and evolutionary perspective on humang-microbe mutualism and disease," *Nature*. Nature Publishing Group, pp. 811–818. doi:10.1038/nature06245.

Domingues, R.G. and Hepworth, M.R. (2020) "Immunoregulatory Sensory Circuits in Group 3 Innate Lymphoid Cell (ILC3) Function and Tissue Homeostasis," *Frontiers in Immunology*, 0, p. 116. doi:10.3389/FIMMU.2020.00116.

Donaldson, G.P., Lee, S.M. and Mazmanian, S.K. (2016) "Gut biogeography of the bacterial microbiota," *Nature reviews. Microbiology*, 14(1), p. 20. doi:10.1038/NRMICRO3552.

Du, S.F. *et al.* (2014) "China in the period of transition from scarcity and extensive undernutrition to emerging nutrition-related non-communicable diseases, 1949-1992," *Obesity Reviews*, 15(S1), pp. 8–15. doi:10.1111/OBR.12122/SUPPINFO.

Dubois, G. *et al.* (2017) "The Inuit gut microbiome is dynamic over time and shaped by traditional foods," *Microbiome*, 5(1), p. 151. doi:10.1186/s40168-017-0370-7.

Dunne, J.L. *et al.* (2014) "The intestinal microbiome in type 1 diabetes," *Clinical & Experimental Immunology*, 177(1), pp. 30–37. doi:10.1111/cei.12321.

Edgar, R.C. (2010) “Search and clustering orders of magnitude faster than BLAST,” *Bioinformatics*, 26(19), pp. 2460–2461. doi:10.1093/BIOINFORMATICS/BTQ461.

Edgar, R.C. (2016) “UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing,” *bioRxiv*, p. 081257. doi:10.1101/081257.

Eisenhofer, R. *et al.* (2019) “Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations,” *Trends in Microbiology*, 27(2), pp. 105–117. doi:10.1016/J.TIM.2018.11.003.

Eisenstein, M. (2020) “The hunt for a healthy microbiome,” *Nature*. NLM (Medline). doi:10.1038/d41586-020-00193-3.

Eribo, O.A. *et al.* (2019) “The gut microbiome in tuberculosis susceptibility and treatment response: guilty or not guilty?,” *Cellular and Molecular Life Sciences 2019 77:8*, 77(8), pp. 1497–1509. doi:10.1007/S00018-019-03370-4.

Eshed, V. *et al.* (2004) “Musculoskeletal stress markers in Natufian hunter-gatherers and Neolithic farmers in the Levant: The upper limb,” *American Journal of Physical Anthropology*, 123(4), pp. 303–315. doi:10.1002/AJPA.10312.

Eshed, V., Gopher, A. and Hershkovitz, I. (2006) “Tooth wear and dental pathology at the advent of agriculture: New evidence from the Levant,” *American Journal of Physical Anthropology*, 130(2), pp. 145–159. doi:10.1002/AJPA.20362.

Evans, A. (2019) “How Cities Erode Gender Inequality: A New Theory and Evidence from Cambodia:,” *https://doi.org/10.1177/0891243219865510*, 33(6), pp. 961–984. doi:10.1177/0891243219865510.

Ezzy, A.C. *et al.* (2019) “Storage and handling of human faecal samples affect the gut microbiome composition: A feasibility study,” *Journal of Microbiological Methods*, 164, p. 105668. doi:10.1016/J.MIMET.2019.105668.

Fiedorová, K. *et al.* (2019) “The impact of DNA extraction methods on stool bacterial and fungal microbiota community recovery,” *Frontiers in Physiology*, 10, p. 821. doi:10.3389/FMICB.2019.00821/BIBTEX.

De Filippo, C. *et al.* (2010) “Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa,” *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), pp. 14691–14696. doi:10.1073/pnas.1005963107.

De Filippo, C. *et al.* (2017) “Diet, Environments, and Gut Microbiota. A Preliminary Investigation in Children Living in Rural and Urban Burkina Faso and Italy,” *Frontiers in Microbiology*, 8(OCT), p. 1979. doi:10.3389/fmicb.2017.01979.

Francino, M.P. (2016) “Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances,” *Frontiers in Microbiology*. Frontiers Media S.A., p. 1543. doi:10.3389/fmicb.2015.01543.

Franzosa, E.A. *et al.* (2015) “Identifying personal microbiomes using metagenomic codes,” *Proceedings of the National Academy of Sciences of the United States of America*, 112(22), pp. E2930–E2938. doi:10.1073/pnas.1423854112.

Fredricks, D.N., Smith, C. and Meier, A. (2005) “Comparison of six DNA extraction methods for recovery of fungal DNA as assessed by quantitative PCR,” *Journal of Clinical Microbiology*, 43(10), pp. 5122–5128. doi:10.1128/JCM.43.10.5122-5128.2005/ASSET/8903969E-BEF4-4244-A023-247890634536/ASSETS/GRAPHIC/ZJM0100559160003.JPEG.

Galloway-Peña, J. and Hanson, B. (2020) “Tools for Analysis of the Microbiome,” *Digestive Diseases and Sciences 2020 65:3*, 65(3), pp. 674–685. doi:10.1007/S10620-020-06091-Y.

Gargallo, E. (2020) “Community Conservation and Land Use in Namibia: Visions, Expectations and Realities,” *https://doi.org/10.1080/03057070.2020.1705617*, 46(1), pp. 129–147. doi:10.1080/03057070.2020.1705617.

Gaulke, C.A. *et al.* (2018) “Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota,” *mBio*, 9(5). doi:10.1128/MBIO.01348-18.

Geneva: World Health Organization (2020) *Global tuberculosis report 2020*.

Gloor, G.B. *et al.* (2017) “Microbiome Datasets Are Compositional: And This Is Not Optional,” *Frontiers in Microbiology*, 0(NOV), p. 2224. doi:10.3389/FMICB.2017.02224.

Gomez, A. *et al.* (2016a) “Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns,” *Cell Reports*, 14(9), pp. 2142–2153. doi:10.1016/j.celrep.2016.02.013.

Gomez, A. *et al.* (2016b) “Gut Microbiome of Coexisting BaAka Pygmies and Bantu Reflects Gradients of Traditional Subsistence Patterns,” *Cell Reports*, 14(9), pp. 2142–2153. doi:10.1016/j.celrep.2016.02.013.

Gomez, A., Luckey, D. and Taneja, V. (2015) “The gut microbiome in autoimmunity: Sex matters,” *Clinical Immunology*, 159(2), pp. 154–162. doi:10.1016/J.CLIM.2015.04.016.

Grieneisen, L.E. *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*, 286(1901). doi:10.1098/RSPB.2019.0431.

Gupta, V.K., Paul, S. and Dutta, C. (2017) “Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity,” *Frontiers in Microbiology*, 8(JUN). doi:10.3389/fmicb.2017.01162.

Hallen-Adams, H.E. *et al.* (2015) “Fungi inhabiting the healthy human gastrointestinal tract: a diverse and dynamic community,” *Fungal Ecology*, 15, pp. 9–17. doi:10.1016/J.FUNECO.2015.01.006.

Hallen-Adams, H.E. and Suhr, M.J. (2016) “Fungi in the healthy human gastrointestinal tract,” *http://dx.doi.org/10.1080/21505594.2016.1247140*, 8(3), pp. 352–358. doi:10.1080/21505594.2016.1247140.

Hansen, R., Gerasimidis, K. and Turner, S. (2019) “Asthma causation and the gastrointestinal microbiome and metabolome: Might there be a signal, or is it just noise?,” *Journal of Allergy and Clinical Immunology*. Mosby Inc., pp. 401–403. doi:10.1016/j.jaci.2019.04.026.

Haro, C. *et al.* (2016) “Intestinal microbiota is influenced by gender and body mass index,” *PLoS ONE*, 11(5). doi:10.1371/journal.pone.0154090.

Harrison, X.A. *et al.* (2021) “Fungal microbiomes are determined by host phylogeny and exhibit widespread associations with the bacterial microbiome,” *Proceedings of the Royal Society B*, 288(1957), p. 20210552. doi:10.1098/RSPB.2021.0552.

Hassanzadeh, S. and Gholamnezhad, M. (2020) “Investigating the importance and causes of rheumatoid arthritis and its effective treatments: A review study,” *Revista Latinoamericana de Hipertension*, 15(1), pp. 59–63. doi:10.5281/zenodo.4074226.

Hernández-Santos, N. and Klein, B.S. (2017) “Through the Scope Darkly: The Gut Mycobiome Comes into Focus,” *Cell Host & Microbe*, 22(6), pp. 728–729. doi:10.1016/J.CHOM.2017.11.013.

Hildebrand, H. *et al.* (2008) “Early-life exposures associated with antibiotic use and risk of subsequent Crohn’s disease,” *Scandinavian Journal of Gastroenterology*, 43(8), pp. 961–966. doi:10.1080/00365520801971736.

Hitchcock, R. (2020) “Dietary transitions of foragers in southern Africa,” *Nature Food 2020 1:3*, 1(3), pp. 183–184. doi:10.1038/s43016-020-0048-3.

Hoffmann, C. *et al.* (2013) “Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents,” *PLOS ONE*, 8(6), p. e66019. doi:10.1371/JOURNAL.PONE.0066019.

Hornung, B.V.H., Zwittink, R.D. and Kuijper, E.J. (2019) “Issues and current standards of controls in microbiome research,” *FEMS Microbiology Ecology*, 95(5), p. 45. doi:10.1093/FEMSEC/FIZ045.

Hu, Y. *et al.* (2019) “Gut microbiota associated with pulmonary tuberculosis and dysbiosis caused by anti-tuberculosis drugs,” *Journal of Infection*, 78(4), pp. 317–322. doi:10.1016/J.JINF.2018.08.006.

Hughes, J.B. and Hellmann, J.J. (2005) “The Application of Rarefaction Techniques to Molecular Inventories of Microbial Diversity,” *Methods in Enzymology*, 397, pp. 292–308. doi:10.1016/S0076-6879(05)97017-1.

Huttenhower, C. *et al.* (2012) “Structure, function and diversity of the healthy human microbiome,” *Nature*, 486(7402), pp. 207–214. doi:10.1038/nature11234.

Iliev, I.D. *et al.* (2012) “Interactions Between Commensal Fungi and the C-Type Lectin Receptor Dectin-1 Influence Colitis,” *Science*, 336(6086), pp. 1314–1317. doi:10.1126/SCIENCE.1221789.

IMAMURA-HAYAKI, K. (1996) “Gathering Activity among the Central Kalahari San,” *African study monographs. Supplementary issue.*, 22, pp. 47–66. doi:10.14989/68377.

Intriago, M. *et al.* (2019) “Clinical Characteristics in Patients with Rheumatoid Arthritis: Differences between Genders,” *Scientific World Journal*, 2019. doi:10.1155/2019/8103812.

Isaac, G. (2010) “The diet of early man: Aspects of archaeological evidence from lower and middle Pleistocene sites in Africa,” *https://doi.org/10.1080/00438243.1971.9979481*, 2(3), pp. 278–299. doi:10.1080/00438243.1971.9979481.

Jackson, G. (2021) “Critical junctures, agrarian change, and the (re)production of vulnerability in a marginalised Indigenous society,” *World Development*, 145, p. 105538. doi:10.1016/J.WORLDDEV.2021.105538.

Jiang, T.T. *et al.* (2017) “Commensal Fungi Recapitulate the Protective Benefits of Intestinal Bacteria,” *Cell Host & Microbe*, 22(6), pp. 809-816.e4. doi:10.1016/J.CHOM.2017.10.013.

Jiao, K., Xu, M. and Liu, M. (2018) “Health status and air pollution related socioeconomic concerns in urban China,” *International Journal for Equity in Health 2018 17:1*, 17(1), pp. 1–11. doi:10.1186/S12939-018-0719-Y.

Johanesen, P. *et al.* (2015) “Disruption of the Gut Microbiome: Clostridium difficile Infection and the Threat of Antibiotic Resistance,” *Genes*, 6(4), pp. 1347–1360. doi:10.3390/genes6041347.

John, G.K. and Mullin, G.E. (2016) “The Gut Microbiome and Obesity,” *Current Oncology Reports 2016 18:7*, 18(7), pp. 1–7. doi:10.1007/S11912-016-0528-7.

Kabwe, M.H. *et al.* (2020) “The gut mycobiota of rural and urban individuals is shaped by geography,” *BMC Microbiology 2020 20:1*, 20(1), pp. 1–12. doi:10.1186/S12866-020-01907-3.

Kaul, A. *et al.* (2017) “Analysis of Microbiome Data in the Presence of Excess Zeros,” *Frontiers in Microbiology*, 0(NOV), p. 2114. doi:10.3389/FMICB.2017.02114.

KE, F. *et al.* (2016) “Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation,” *Nature medicine*, 22(10), pp. 1187–1191. doi:10.1038/NM.4176.

Keohane, D.M. *et al.* (2020) “Microbiome and health implications for ethnic minorities after enforced lifestyle changes,” *Nature Medicine 2020 26:7*, 26(7), pp. 1089–1095. doi:10.1038/s41591-020-0963-8.

Kho, Z.Y. and Lal, S.K. (2018) “The Human Gut Microbiome – A Potential Controller of Wellness and Disease,” *Frontiers in Microbiology*, 0(AUG), p. 1835. doi:10.3389/FMICB.2018.01835.

Klem, F. *et al.* (2017) “Prevalence, Risk Factors, and Outcomes of Irritable Bowel Syndrome After Infectious Enteritis: A Systematic Review and Meta-analysis,” *Gastroenterology*, 152(5), pp. 1042-1054.e1. doi:10.1053/J.GASTRO.2016.12.039.

Kong, H.H. and Morris, A. (2017) "The emerging importance and challenges of the human mycobiome," *http://dx.doi.org/10.1080/21505594.2017.1279780*, 8(3), pp. 310–312. doi:10.1080/21505594.2017.1279780.

Kovatcheva-Datchary, P. *et al.* (2015) "Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella," *Cell Metabolism*, 22(6), pp. 971–982. doi:10.1016/J.CMET.2015.10.001.

Kumbhare, S. V *et al.* (2019) "Factors influencing the gut microbiome in children: from infancy to childhood," *Journal of Biosciences 2019 44:2*, 44(2), pp. 1–19. doi:10.1007/S12038-019-9860-Z.

Kurilshikov, A. *et al.* (2021) "Large-scale association analyses identify host factors influencing human gut microbiome composition," *Nature Genetics 2021 53:2*, 53(2), pp. 156–165. doi:10.1038/s41588-020-00763-1.

Kushak, R.I. and Winter, H.S. (2018) "Intestinal microbiota, metabolome and gender dimorphism in autism spectrum disorders," *Research in Autism Spectrum Disorders*, 49, pp. 65–74. doi:10.1016/j.rasd.2018.01.009.

Lang, S. *et al.* (2020) "Intestinal Fungal Dysbiosis and Systemic Immune Response to Fungi in Patients With Alcoholic Hepatitis," *Hepatology*, 71(2), pp. 522–538. doi:10.1002/HEP.30832.

Larsen, C.S. *et al.* (2019) "Bioarchaeology of Neolithic Çatalhöyük reveals fundamental transitions in health, mobility, and lifestyle in early farmers," *Proceedings of the National Academy of Sciences*, 116(26), pp. 12615–12623. doi:10.1073/PNAS.1904345116.

Lathi, L. and Shetty, S. (2017) *Tools for microbiome analysis in R, version 1.1.10013*. Available at: https://github.com/microbiome/microbiome/.

LEDERBERG, J. and MCCRAY, A.T. (2001) "`Ome Sweet `Omics--A Genealogical Treasury of Words," *The Scientist*, 15(7), pp. 8–8.

Lee, R.B. (2010) "Ecology of Food and Nutrition Mongongo: The ethnography of a major wild food resource." doi:10.1080/03670244.1973.9990351.

Lee, R.B. (2017) "What Hunters Do for a Living, or, How to Make Out on Scarce Resources," *Man the Hunter*, pp. 30–48. doi:10.4324/9780203786567-6.

Li, Q. *et al.* (2014) "Dysbiosis of Gut Fungal Microbiota is Associated With Mucosal Inflammation in Crohn's Disease," *Journal of Clinical Gastroenterology*, 48(6), p. 513. doi:10.1097/MCG.0000000000000035.

Li, Y. (2018) "Epigenetic Mechanisms Link Maternal Diets and Gut Microbiome to Obesity in the Offspring," *Frontiers in Genetics*, 0(AUG), p. 342. doi:10.3389/FGENE.2018.00342.

Lim, S.J. and Bordenstein, S.R. (2020) "An introduction to phylosymbiosis," *Proceedings of the Royal Society B*, 287(1922). doi:10.1098/RSPB.2019.2900.

Lind, A.L. and Pollard, K.S. (2021) "Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing," *Microbiome 2021 9:1*, 9(1), pp. 1–18. doi:10.1186/S40168-021-01015-Y.

Love, M.I., Huber, W. and Anders, S. (2014) "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology 2014 15:12*, 15(12), pp. 1–21. doi:10.1186/S13059-014-0550-8.

Lupton, J.R. (2002) *Dietary Reference Intakes For Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*. Available at: https://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=s10490 (Accessed: January 15, 2022).

Ma, Y. *et al.* (2018) "Help, hope and hype: ethical considerations of human microbiome research and applications," *Protein & Cell 2018 9:5*, 9(5), pp. 404–415. doi:10.1007/S13238-018-0537-4.

Malla, M.A. *et al.* (2019) "Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment," *Frontiers in Immunology*, 0(JAN), p. 2868. doi:10.3389/FIMMU.2018.02868.

Mancabelli, L. *et al.* (2017) "Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations," *Environmental Microbiology*, 19(4), pp. 1379–1390. doi:10.1111/1462-2920.13692.

Mar Rodríguez, M. *et al.* (2015) "Obesity changes the human gut mycobiome," *Scientific Reports 2015 5:1*, 5(1), pp. 1–15. doi:10.1038/srep14600.

Marlowe, F. (2002) "Why the Hadza are Still Hunter-Gatherers," pp. 247–275.

Martin, C.R. *et al.* (2018) "The Brain-Gut-Microbiome Axis," *CMGH*. Elsevier Inc, pp. 133–148. doi:10.1016/j.jcmgh.2018.04.003.

Martinez Arbizu, P. (2017) "pairwiseAdonis: Pairwise Multilevel Comparison using Adonis."

Martínez, I. *et al.* (2015) "The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes," *Cell Reports*, 11(4), pp. 527–538. doi:10.1016/j.celrep.2015.03.049.

Masood, F. (2020) "View of The Effect of Agriculture on Health in Neolithic Populations in the Levant | Pathways," *Pathways*, 1, pp. 84–95.

McInnes, R.S. *et al.* (2020) "Horizontal transfer of antibiotic resistance genes in the human gut microbiome," *Current Opinion in Microbiology*. Elsevier Ltd, pp. 35–43. doi:10.1016/j.mib.2020.02.002.

McKnight, D.T. *et al.* (2019) "Methods for normalizing microbiome data: An ecological perspective," *Methods in Ecology and Evolution*, 10(3), pp. 389–400. doi:10.1111/2041-210X.13115.

McMurdie, P.J. and Holmes, S. (2014) "Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible," *PLOS Computational Biology*, 10(4), p. e1003531. doi:10.1371/JOURNAL.PCBI.1003531.

Miller, G.E. *et al.* (2016) "Lower Neighborhood Socioeconomic Status Associated with Reduced Diversity of the Colonic Microbiota in Healthy Adults," *PLOS ONE*, 11(2), p. e0148952. doi:10.1371/JOURNAL.PONE.0148952.

Milton, K. (1999) "A hypothesis to explain the role of meat eating in human evolution," *Evolutionary Anthropology* [Preprint].

Milton, K. (2000) "Hunter-gatherer diets—a different perspective," *The American Journal of Clinical Nutrition*, 71(3), pp. 665–667. doi:10.1093/AJCN/71.3.665.

Minich, J.J. *et al.* (2019) "Quantifying and Understanding Well-to-Well Contamination in Microbiome Research," *mSystems*, 4(4). doi:10.1128/MSYSTEMS.00186-19.

Mizrahi-Man, O., Davenport, E.R. and Gilad, Y. (2013) "Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs," *PLOS ONE*, 8(1), p. e53608. doi:10.1371/JOURNAL.PONE.0053608.

Moeller, A.H. *et al.* (2014) "Rapid changes in the gut microbiome during human evolution," *PNAS*, 111(46), pp. 16431–16435. doi:10.1073/pnas.1419136111.

Moeller, A.H. (2017) "The shrinking human gut microbiome," *Current Opinion in Microbiology*, 38, pp. 30–35. doi:10.1016/J.MIB.2017.04.002.

Moran, N.A. and Sloan, D.B. (2015) "The Hologenome Concept: Helpful or Hollow?," *PLOS Biology*, 13(12), p. e1002311. doi:10.1371/JOURNAL.PBIO.1002311.

Nagpal, R. *et al.* (2018) "Gut microbiome and aging: Physiological and mechanistic insights," *Nutrition and Healthy Aging*. IOS Press, pp. 267–285. doi:10.3233/NHA-170030.

Nash, A.K. *et al.* (2017) "The gut mycobiome of the Human Microbiome Project healthy cohort," *Microbiome 2017 5:1*, 5(1), pp. 1–13. doi:10.1186/S40168-017-0373-4.

Neuwirth, E. (2014) "RColorBrewer: ColorBrewer Palettes."

Nuria, S. *et al.* (2013) "Microbial Targets for the Development of Functional Foods Accordingly with Nutritional and Immune Parameters Altered in the Elderly," *http://dx.doi.org/10.1080/07315724.2013.827047*, 32(6), pp. 399–406. doi:10.1080/07315724.2013.827047.

Odintsova, V., Tyakht, A. and Alexeev, D. (2017) "Guidelines to Statistical Analysis of Microbial Composition Data Inferred from Metagenomic Sequencing," *Current Issues in Molecular Biology 2017, Vol. 24, Pages 17-36*, 24(1), pp. 17–36. doi:10.21775/CIMB.024.017.

O'Keefe, S.J.D. *et al.* (2015) "Fat, fibre and cancer risk in African Americans and rural Africans," *Nature Communications 2015 6:1*, 6(1), pp. 1–14. doi:10.1038/ncomms7342.

Oliphant, K. and Allen-Vercoe, E. (2019) "Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health," *Microbiome 2019 7:1*, 7(1), pp. 1–15. doi:10.1186/S40168-019-0704-8.

Ott, S.J. *et al.* (2009) "Fungi and inflammatory bowel diseases: Alterations of composition and diversity," *http://dx.doi.org/10.1080/00365520801935434*, 43(7), pp. 831–841. doi:10.1080/00365520801935434.

Owers, K.A. *et al.* (2017) "Adaptation to infectious disease exposure in indigenous Southern African populations," *Proceedings of the Royal Society B: Biological Sciences*, 284(1852). doi:10.1098/RSPB.2017.0226.

Paulson, J.N. *et al.* (2013) "Differential abundance analysis for microbial marker-gene surveys," *Nature Methods 2013 10:12*, 10(12), pp. 1200–1202. doi:10.1038/nmeth.2658.

Pérez-Cobas, A.E. *et al.* (2013) "Differential Effects of Antibiotic Therapy on the Structure and Function of Human Gut Microbiota," *PLOS ONE*, 8(11), p. e80201. doi:10.1371/JOURNAL.PONE.0080201.

Perry, G.H. *et al.* (2007) "Diet and the evolution of human amylase gene copy number variation," *Nature Genetics*, 39(10), pp. 1256–1260. doi:10.1038/ng2123.

Polvi, E.J. *et al.* (2015) "Opportunistic yeast pathogens: Reservoirs, virulence mechanisms, and therapeutic strategies," *Cellular and Molecular Life Sciences*, 72(12), pp. 2261–2287. doi:10.1007/S00018-015-1860-Z/FIGURES/1.

Prodan, A. *et al.* (2020) "Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing," *PLOS ONE*. Edited by J.-S. Seo, 15(1), p. e0227434. doi:10.1371/journal.pone.0227434.

Qin, J. *et al.* (2010) "A human gut microbial gene catalogue established by metagenomic sequencing," *Nature*, 464(7285), pp. 59–65. doi:10.1038/nature08821.

Raimondi, S. *et al.* (2019) "Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping, and colonization," *Frontiers in Microbiology*, 10, p. 1575. doi:10.3389/FMICB.2019.01575/BIBTEX.

Rampelli, S. *et al.* (2015) "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota," *Current Biology*, 25(13), pp. 1682–1693. doi:10.1016/j.cub.2015.04.055.

Rangel, I. *et al.* (2015) "The relationship between faecal-associated and mucosal-associated microbiota in irritable bowel syndrome patients and healthy subjects," *Alimentary Pharmacology & Therapeutics*, 42(10), pp. 1211–1221. doi:10.1111/APT.13399.

Ranjan, R. *et al.* (2016) "Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing," *Biochemical and Biophysical Research Communications*, 469(4), pp. 967–977. doi:10.1016/J.BBRC.2015.12.083.

Razavi, A.C. *et al.* (2019) "Sex, gut microbiome, and cardiovascular disease risk," *Biology of Sex Differences 2019 10:1*, 10(1), pp. 1–14. doi:10.1186/S13293-019-0240-Z.

Rifkin, R.F. *et al.* (2020) "Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome," *Microbiome*, 8(1), pp. 1–23. doi:10.1186/s40168-020-00832-x.

Roberfroid, M. *et al.* (2010) "Prebiotic effects: Metabolic and health benefits," *British Journal of Nutrition*. Cambridge University Press, pp. S1–S63. doi:10.1017/S0007114510003363.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, 26(1), pp. 139–140. doi:10.1093/BIOINFORMATICS/BTP616.

Robson, A.J. and Kaplan, H.S. (2003) "The Evolution of Human Life Expectancy and Intelligence in Hunter-Gatherer Economies," *American Economic Review*, 93(1), pp. 150–169. doi:10.1257/000282803321455205.

Rook, G.A.W., Raison, C.L. and Lowry, C.A. (2014) "Microbial 'old friends', immunoregulation and socioeconomic status," *Clinical & Experimental Immunology*, 177(1), pp. 1–12. doi:10.1111/CEI.12269.

Rupnik, M., Wilcox, M.H. and Gerding, D.N. (2009) "Clostridium difficile infection: New developments in epidemiology and pathogenesis," *Nature Reviews Microbiology*. Nature Publishing Group, pp. 526–536. doi:10.1038/nrmicro2164.

Salazar, N. *et al.* (2014) "The human intestinal microbiome at extreme ages of life. Dietary intervention as a way to counteract alterations," *Frontiers in Genetics*. Frontiers Media S.A., p. 406. doi:10.3389/fgene.2014.00406.

Salonen, A. *et al.* (2012) "The adult intestinal core microbiota is determined by analysis depth and health status," *Clinical Microbiology and Infection*, 18(SUPPL. 4), pp. 16–20. doi:10.1111/J.1469-0691.2012.03855.X.

Salter, S.J. *et al.* (2014) "Reagent and laboratory contamination can critically impact sequence-based microbiome analyses," *BMC Biology 2014 12:1*, 12(1), pp. 1–12. doi:10.1186/S12915-014-0087-Z.

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 2017, Vol. 18, Page 330*, 18(2), p. 330. doi:10.3390/IJMS18020330.

Sanidad, K.Z. and Zeng, M.Y. (2020) "Neonatal gut microbiome and immunity," *Current Opinion in Microbiology*, 56, pp. 30–37. doi:10.1016/J.MIB.2020.05.011.

Schaik, W. van (2015) "The human gut resistome," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1670), p. 20140087. doi:10.1098/RSTB.2014.0087.

Schnorr, S.L. *et al.* (2014) "Gut microbiome of the Hadza hunter-gatherers," *Nature Communications 2014 5:1*, 5(1), pp. 1–12. doi:10.1038/ncomms4654.

Schoch, C.L. *et al.* (2012) "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi," *Proceedings of the National Academy of Sciences*, 109(16), pp. 6241–6246. doi:10.1073/PNAS.1117018109.

Sharpton, T.J. (2014) “An introduction to the analysis of shotgun metagenomic data,” *Frontiers in Plant Science*, 0(JUN), p. 209. doi:10.3389/FPLS.2014.00209.

Shuwen, H. *et al.* (2019) “Protective effect of the ‘food-microorganism-SCFAs’ axis on colorectal cancer: from basic research to practical application,” *Journal of Cancer Research and Clinical Oncology 2019 145:9*, 145(9), pp. 2169–2197. doi:10.1007/S00432-019-02997-X.

Sinha, R. *et al.* (2015) “The microbiome quality control project: baseline study design and future directions,” *Genome Biology 2015 16:1*, 16(1), pp. 1–6. doi:10.1186/S13059-015-0841-8.

Sivaprakasam, S., Prasad, P.D. and Singh, N. (2016) “Benefits of short-chain fatty acids and their receptors in inflammation and carcinogenesis,” *Pharmacology and Therapeutics*. Elsevier Inc., pp. 144–151. doi:10.1016/j.pharmthera.2016.04.007.

Slashinski, M.J. *et al.* (2012) “‘Snake-oil,’ ‘quack medicine,’ and ‘industrially cultured organisms:’ biovalue and the commercialization of human microbiome research,” *BMC Medical Ethics 2012 13:1*, 13(1), pp. 1–8. doi:10.1186/1472-6939-13-28.

Sokol, H. *et al.* (2017) “Fungal microbiota dysbiosis in IBD,” *Gut*, 66(6), pp. 1039–1048. doi:10.1136/GUTJNL-2015-310746.

Sommer, F. and Bäckhed, F. (2013) “The gut microbiota-masters of host development and physiology,” *Nature Reviews Microbiology*, pp. 227–238. doi:10.1038/nrmicro2974.

Sommer, M.O.A. (2015) “Advancing gut microbiome research using cultivation,” *Current Opinion in Microbiology*, 27, pp. 127–132. doi:10.1016/J.MIB.2015.08.004.

Song, S.J. *et al.* (2016) “Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies,” *mSystems*, 1(3), pp. 21–37. doi:10.1128/MSYSTEMS.00021-16/SUPPL_FILE/SYS001162019ST5.PDF.

Sonnenburg, E.D. and Sonnenburg, J.L. (2014) “Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates,” *Cell Metabolism*, 20(5), pp. 779–786. doi:10.1016/J.CMET.2014.07.003.

Suzuki, T.A. *et al.* (2021) “Codiversification of gut microbiota with humans,” *bioRxiv*, p. 2021.10.12.462973. doi:10.1101/2021.10.12.462973.

Tang, Q. *et al.* (2020) “Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices,” *Frontiers in Cellular and Infection Microbiology*, 0, p. 151. doi:10.3389/FCIMB.2020.00151.

Thorburn, A.N., Macia, L. and Mackay, C.R. (2014) “Diet, Metabolites, and ‘Western-Lifestyle’ Inflammatory Diseases,” *Immunity*, 40(6), pp. 833–842. doi:10.1016/J.IMMUNI.2014.05.014.

Thursby, E. and Juge, N. (2017) “Introduction to the human gut microbiota,” *Biochemical Journal*, 474(11), pp. 1823–1836. doi:10.1042/BCJ20160510.

Tishkoff, S.A. *et al.* (2007) “Convergent adaptation of human lactase persistence in Africa and Europe,” *Nature genetics*, 39(1), p. 31. doi:10.1038/NG1946.

Tollefson, D. *et al.* (2013) “Burden of tuberculosis in indigenous peoples globally: A systematic review,” *International Journal of Tuberculosis and Lung Disease*, 17(9), pp. 1139–1150. doi:10.5588/IJTLD.12.0385.

Tringe, S.G. and Hugenholtz, P. (2008) “A renaissance for the pioneering 16S rRNA gene,” *Current Opinion in Microbiology*, 11(5), pp. 442–446. doi:10.1016/J.MIB.2008.09.011.

Tropini, C. *et al.* (2017) “The Gut Microbiome: Connecting Spatial Organization to Function,” *Cell Host & Microbe*, 21(4), pp. 433–442. doi:10.1016/J.CHOM.2017.03.010.

Turnbaugh, P.J. *et al.* (2006) "An obesity-associated gut microbiome with increased capacity for energy harvest," *Nature 2006 444:7122*, 444(7122), pp. 1027–1031. doi:10.1038/nature05414.

Turnbaugh, P.J. *et al.* (2007) "The Human Microbiome Project," *Nature 2007 449:7164*, 449(7164), pp. 804–810. doi:10.1038/nature06244.

Underhill, D.M. and Iliev, I.D. (2014) "The mycobiota: interactions between commensal fungi and the host immune system," *Nature Reviews Immunology 2014 14:6*, 14(6), pp. 405–416. doi:10.1038/nri3684.

Vangay, P. *et al.* (2018) "US Immigration Westernizes the Human Gut Microbiome," *Cell*, 175(4), pp. 962-972.e10. doi:10.1016/j.cell.2018.10.029.

Vich Vila, A. *et al.* (2018) "Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome," *Science Translational Medicine*, 10(472), p. eaap8914. doi:10.1126/SCITRANSLMED.AAP8914.

de Vrieze, J. (2014) "Gut Instinct," *Science*, 14 January, pp. 241–243.

Walter, J. and Ley, R. (2011) "The Human Gut Microbiome: Ecology and Recent Evolutionary Changes," *Annual Review of Microbiology*, 65(1), pp. 411–429. doi:10.1146/annurev-micro-090110-102830.

Weisdorf, J.L. (2005) "From Foraging To Farming: Explaining The Neolithic Revolution," *Journal of Economic Surveys*, 19(4), pp. 561–586. doi:10.1111/J.0950-0804.2005.00259.X.

Weiss, S. *et al.* (2014) "Tracking down the sources of experimental contamination in microbiome studies," *Genome Biology 2014 15:12*, 15(12), pp. 1–3. doi:10.1186/S13059-014-0564-2.

Weiss, S. *et al.* (2017) "Normalization and microbial differential abundance strategies depend upon data characteristics," *Microbiome 2017 5:1*, 5(1), pp. 1–18. doi:10.1186/S40168-017-0237-Y.

West, C.E. *et al.* (2015) "The gut microbiota and inflammatory noncommunicable diseases: Associations and potentials for gut microbiota therapies," *Journal of Allergy and Clinical Immunology*. Mosby Inc., p. 14. doi:10.1016/j.jaci.2014.11.014.

Wilson, A.S. *et al.* (2020) "Diet and the Human Gut Microbiome: An International Review," *Digestive Diseases and Sciences*, 65(3), pp. 723–740. doi:10.1007/S10620-020-06112-W/FIGURES/2.

Wlodarska, M. *et al.* (2011) "Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated Citrobacter rodentium-induced colitis," *Infection and Immunity*, 79(4), pp. 1536–1545. doi:10.1128/IAI.01104-10.

Wong, E. *et al.* (2021) "Manipulating the neonatal gut microbiome: current understanding and future perspectives," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, p. fetalneonatal-2021-321922. doi:10.1136/ARCHDISCHILD-2021-321922.

Wood, L.G. (2017) "Diet, Obesity, and Asthma," *https://doi.org/10.1513/AnnalsATS.201702-124AW*, 14, pp. S332–S338. doi:10.1513/ANNALSATS.201702-124AW.

Xia, Y. (2020) "Correlation and association analyses in microbiome study integrating multiomics in health and disease," *Progress in Molecular Biology and Translational Science*, 171, pp. 309–491. doi:10.1016/BS.PMBTS.2020.04.003.

Yatsunenko, T. *et al.* (2012) "Human gut microbiome viewed across age and geography," *Nature*, pp. 222–227. doi:10.1038/nature11053.

Zagorsky, J.L. and Smith, P.K. (2017) "The association between socioeconomic status and adult fast-food consumption in the U.S.," *Economics & Human Biology*, 27, pp. 12–25. doi:10.1016/J.EHB.2017.04.004.

Zeder, M.A. (2011) “The origins of agriculture in the Near East,” *Current Anthropology*, 52(SUPPL. 4), p. 221. doi:10.1086/659307/ASSET/IMAGES/LARGE/FG2.JPEG.

Zhao, L. *et al.* (2018) “Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes,” *Science*, 359(6380), pp. 1151–1156. doi:10.1126/SCIENCE.AAO5774.

Zoetendal, E.G. *et al.* (2002) “Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces,” *Applied and Environmental Microbiology*, 68(7), pp. 3401–3407. doi:10.1128/AEM.68.7.3401-3407.2002.

Zuo, T. *et al.* (2018) “Gut fungal dysbiosis correlates with reduced efficacy of fecal microbiota transplantation in Clostridium difficile infection,” *Nature Communications 2018 9:1*, 9(1), pp. 1–11. doi:10.1038/s41467-018-06103-6.