Host-microbiota interactions in mammals

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE UNIVERSITY OF OXFORD

BY

EVELIINA HANSKI

HERTFORD COLLEGE

AND

DEPARTMENT OF BIOLOGY

OXFORD

December 2022

DECLARATION

I declare that the work presented in this thesis is my own. Contributions of co-authors are presented for each chapter. This thesis contains less than 50,000 words excluding supplementary material and references.

ACKNOWLEDGEMENTS

Thank you to my funders, the Finnish Government, the Osk. Huttunen Foundation, Otto A. Malm Foundation, Emil Aaltonen Foundation as well as the National Geographic Society for supporting me financially through scholarships and research grants during this PhD.

Thank you, Sarah and Martin, for the irreplaceable supervision and mentorship. Thank you for withstanding my stubbornness and letting me go on a science adventure with a largely (entirely) unknown destination. Thank you to members of the Knowles Lab, Maiden Lab and (the latest addition) Sheppard Lab for scientific and less scientific discussions. Special thanks to Aura, my polar opposite/identical copy, for speaking to me in Finnish (as in directly, not necessarily, and most certainly not entirely, in the Finnish language), introducing me to the Ite saa päättää-philosophy, being my most trusted source of information of all kind and being the person with whom it has been very easy to solemnly swear to be up to no good. Other special thanks of course go to James, my go-to BLASTer and an all-around great person.

Thank you, Eagle and Brown Bear, my island family, for making this PhD possible in the first place, supporting my fieldwork despite my study species clearly being from the wrong class, letting me cuddle the fluffiest of all creatures and giving me your honest opinion on the quality of my puns. Thank you Beggsy, Dobbins, Howard, Billy, and many more island friends for making my time in Skokholm 'mega'. Thank you, my little island and all the stars above you.

Thank you, Alison, Marte, Freddy, Jonny, Becky, Patrik, Thomas, Luke, Kieran, Mark, Jack, Gibbo, Lena, Anya, João, Kamal and many more incredible and curious people for making my time in Oxford special. I will see you again (around the world). Thank you especially Mark, my very first Oxford friend, you have been there for me in so many moments and always joined for a spontaneous pint no questions asked. I hope we will always be friends. Thank you, Alison and Marte, my Teddy Cool Girls, for long nights drinking port, early mornings rowing down the river, adventures in various countries, endless phone calls and a genuinely quality friendship. Thank you, Eeva, Katri, Matti and Sia for supporting me remotely from Funland.

Lastly and most importantly, thank you life and a series of coincidences for bringing me to Oxford and making me walk to the Rusty Bicycle on a rainy November night in 2019. Nothing really was the same after that, was it eh?

TABLE OF CONTENTS

ABSTRACT

The mammalian gut microbiota provides important functions for the host but at the same time varies greatly between species, populations, individuals and even within individuals over time. Drivers of this variation and particularly their relative importance remains poorly understood. Further, a large proportion of gut microbiota research has been conducted using the laboratory mouse as a model organism. However, laboratory mice inhabit an artificial world where individuals are inbred, the environment is stable, and microbial exposures are limited. As a result of domestication, the gut microbiota of laboratory mice differs in notable ways from that of their wild relatives, and studies investigating the gut microbiota of wild house mice have great potential to aid the interpretation of findings from the laboratory. In this thesis, I explore how the gut microbiota varies between laboratory and wild mice, and examine drivers of natural gut microbiota variation among and within wild mouse populations. I show that although wild mice have a compositionally and functionally more diverse gut microbiota with a faster turnover rate that is clearly shaped by their more complex environment, major patterns of gut microbiota assembly during early life are conserved between laboratory and wild house mice. Since the gut microbiota showed strong age-related variation, I developed a non-invasive epigenetic clock-based method for assessing age in wild mice, whose application showed promise for future studies estimating biological (if not chronological) age in this system. Gut microbiota composition was also influenced by spatial factors, with stronger effects on the fungal than bacterial components of the microbiota, and microbial transmission from soil was identified as one possible source of spatial variation. Finally, I investigate gut microbial ecology in a mammalian species that is quite the opposite of a model organism, the critically endangered Saiga antelope which suffers periodic mass mortality events. I characterise healthy gut microbiota in comparison to other similar species, providing important background for future conservation efforts, as well as contributing to our broader understanding of what shapes microbiota among mammals. Together the results from this thesis increase our understanding of the gut microbiota variation but also highlight challenges in studying the gut microbiota, in the laboratory and beyond.

GENERAL INTRODUCTION

The mammalian gut microbiota

All mammals harbour complex communities of microbial organisms on their epithelial barrier surfaces throughout the body (McFall-Ngai et al., 2013). A collection of these microbes in a given site is referred to as the microbiota. A particularly dense and diverse microbiota is found in the mammalian gut. In humans, the gut microbiota contains approximately 2,000 bacterial species that collectively contain more than 100 times more genes than there are in our own genome and at least as many cells as there are human cells (Almeida et al., 2019; Gill et al., 2006; Sender et al., 2016). Species in this gut community range from beneficial to potentially harmful but as a whole the gut microbiota provides important functions for the host, such as immune training and vitamin biosynthesis (Chung et al., 2012; Yatsunenko et al., 2012). At the same time, numerous disease states have been linked to an 'altered' gut microbial community (Turnbaugh et al., 2009; Le Chatelier et al., 2013; Qin et al., 2012; Fujimura et al., 2016; Wang et al., 2011), although findings are often limited to correlation and identification of mechanisms by which the microbiota influences disease pathogenesis has been limited.

Historically, microbes were primarily studied as individual, isolated microbial species. Advances in culture-independent techniques have enabled a rapid expansion of our understanding of the gut microbiota, where entire microbial communities rather than single species can be studied. In particular, 16S rRNA gene sequencing has been widely used to identify members of microbial communities and quantify their relative abundances. While lacking resolution at lower taxonomic levels (Johnson et al., 2019), this well-established and readily available method provides a broad overview of microbiota composition, richness, and evenness. More recently, the method has been coupled with quantitative PCR or similar to quantify bacterial density, allowing measurement of absolute taxon abundances in the gut (Jokela et al., 2022; Vandeputte et al., 2017; Jian et al., 2020). Together these data allow us to

study host–microbiota dynamics from various angles. More advanced sequencing technologies, such as metagenomics, can add further resolution and information to these data, though their high cost still hinders widespread and large-scale application.

Establishment of the gut microbiota

The developing mammalian gut is thought to remain sterile until birth (Walter et al., 2021), when primary colonisers initiate the gradual establishment of a resident microbial community. This initial colonisation by vertically transmitted maternal microbes is followed by – at least in humans – a largely predictable microbial succession with stochastic elements (Rao et al., 2021; Stewart et al., 2018; Bäckhead et al., 2015; Ferretti et al., 2018). New microbes continue to enter the gut through e.g., skin contact and environmental exposure, and the relative abundance of gut microbes is influenced by host selection and stochastic processes.

Important sources of host selection in the developing gut include oxygen levels, the immune system, and diet (Barroso-Batista et al., 2015; Stewart et al., 2018; Sprockett et al., 2018; Albenberg et al., 2014; Friedman et al., 2018). The gut is aerobic at birth but through oxygen depletion it becomes increasingly anoxic, favouring anaerobes over aerotolerant taxa, and through immune maturation a host's ability to suppress pathogenic taxa increases. Further, diet change, particularly a shift from milk to solid food, represents a major selective pressure on the microbiota as species vary in their ability to utilise different substrates. Alongside these deterministic factors shaping the gut microbiota, compositional changes are also induced by stochastic processes, such as drift. Such effects can be strong, particularly for low-abundance species that are more likely to go locally extinct (Sprockett et al., 2018).

As a result of these community processes, the gut microbiota becomes more diverse within individuals but more similar between individuals as they develop (Bäckhead et al., 2015; Ferretti et al., 2018), eventually reaching an 'adult-like' community structure. Still, the gut microbiota continues to fluctuate and adapt throughout life (Murillo et al., 2022; Vandeputte et al., 2021). In particular, the host continues to be exposed to new microbes through environmental and social transmission. Various factors, including baseline microbiota composition, contribute to whether these microbes manage to colonise the gut or not (Sprockett et al., 2018; Bittleston et al., 2020).

Variation in the gut microbiota

Once established, the gut microbiota varies between host species, populations, and across and within individuals (Ochman et al., 2010; Phillips et al., 2012; Yatsunenko et al., 2012; Suzuki & Worobey, 2014; Wang et al., 2014; Smits et al., 2017). Such variation can be detected on many levels, including *taxonomy* (Rojas et al., 2021; Knowles et al., 2019), *diversity* (Rudolph et al., 2022), *absolute* and *compositional abundance* (Jokela et al., 2022; Vandeputte et al., 2017; Risely et al., 2021), *stability* (Lozupone et al., 2012), and *functionality* (Baniel et al., 2022). The mammalian gut microbiota displays a prominent host phylogenetic signal (Song et al., 2020; Kohl et al., 2018; Weinstein et al., 2021; Ochman et al., 2010) and is relatively host species-specific, such that typical features are shared between individuals of a given species and certain microbiota-dependent functions will be altered if the host is colonised with a microbiota from a heterospecific individual (Mallott & Amato, 2021; Chung et al., 2012). At the same time, the gut microbiota varies between individuals to the extent that for instance in humans, the vast majority of gut microbial species in the population are not detected in a majority of individuals (Almeida et al., 2021; Gilbert et al., 2018).

Drivers of gut microbiota variation

Within any given species, various genetic and environmental factors are recognised to influence the composition and diversity of the gut microbiota (Gacesa et al., 2022; Yatsunenko et al., 2012; Suzuki & Worobey, 2014; Grieneisen et al., 2021; Spor et al., 2011; Rudolph et al., 2022; Falony et al., 2016; Kurilshikov et al., 2021; Rothschild et al., 2018). Perhaps the most intuitive driver of gut microbiota variation is diet. Dietary components that are not digested in the upper gastrointestinal tract end up in the large intestine, where the majority of gut microbes reside. These substrates can then be utilised by the resident microbes. As microbes vary in their ability to utilise different substrates, a given diet may promote certain microbial species while depleting others. In line, compositional differences are evident between individuals with distinct diets and the gut microbiota can rapidly respond to dietary changes within individuals (Turnbaugh et al., 2008; Sonnenburg et al., 2010; Arifuzzaman et al., 2022; David et al., 2014; Lancaster et al., 2022). Perhaps the best described diet-induced changes in the gut microbiota are those induced by a Western diet, which is low in fibre but rich in fat and sugar. Consumption of this diet is characterised by an overrepresentation of taxa from phylum Proteobacteria, such as *Escherichia coli,* and reduction in the overall diversity of the gut microbiota (Agus et al., 2016; Martinez-Medina et al., 2014).

The gut microbiota is further influenced by *spatial factors*, such as altitude and latitude (Yatsunenko et al., 2012; Suzuki & Worobey, 2014; Zhang et al., 2018; Suzuki et al., 2019; Olm et al., 2022), *temporal factors*, such as season and time of day (Smits et al., 2017; Maurice et al., 2015; Marsh et al., 2022; Risely et al., 2021), and *host-level factors*, such as age, social interaction, reproductive status, and genetics (Stewart et al., 2018; Koren et al., 2012; DiGiulio et al., 2015; Grieneisen et al., 2021; Rothschild et al., 2018; Raulo et al., 2021; Tung et al., 2015). Despite the insights on environmental and genetic influences on gut microbiota, the

majority of variation often remains unexplained and the relative importance of identified drivers appears to be context-dependent with the explanatory power of a given predictor varying vastly between studies. Variation in gut microbiota is, however, expected to have important effects on host physiology and immunity, and ultimately on fitness, thus an enhanced understanding of gut microbiota variation and its drivers would be valuable.

Effects of gut microbiota variation on the host

Both the gut microbiota and immune system develop in early life. A key role of the mammalian immune system is to defend against pathogenic microbes. Intriguingly, it appears that to do this properly, it first needs to be trained by non-pathogenic microbes. The microbiota's clear impact on host immunity is perhaps most evident in early life, when colonisation by microbes is essential for immune maturation (Gensollen et al., 2016).

The clearest insights on cross-talk between commensals and host immunity come from germ-free (GF) models, where animals lacking microbes present defects in lymphoid tissue architecture and immune functioning, such as a substantial reduction in IgA levels and missing subsets of lymphocytes (Bauer et al., 1963; Ivanov et al., 2009). Microbial colonisation of GF animals restores at least some of these functions, further highlighting the importance of microbes in immune maturation (Hapfelmeier et al., 2010; Chung et al., 2012). Importantly, colonisation with host-specific microbes is required for maximal immune restoration, indicating that exposure to any microbes is not sufficient (Chung et al., 2012).

Interactions between host immunity and the microbiota continue beyond development allowing maintenance of regulatory pathways but also protection against pathogens (Belkaid & Hand, 2014; Chung et al., 2012). By providing suitable conditions for hosting a microbiota, the gut

also provides a site through which pathogens can access the host. Resident microbes can protect the host against pathogen colonisation (known as 'colonisation resistance') through niche competition. Microbiota members can, for instance, alter nutrient availability with consequences for pathogen growth rate and expression of virulence genes (Kamada et al., 2013; Stecher et al., 2011; Britton & Young, 2014). Importantly, resident microbes provide protection by being there: in an established gut community of a healthy individual there are no free niches and local microbes are well-adapted, thus entering pathogens are generally excluded through competition (Kamada et al., 2013; Stecher et al., 2011).

Alongside having a deficient immune system, GF animals present various behavioural abnormalities including exaggerated stress responses and reduced anxiety (Sudo et al., 2004; Diaz Heijtzet al., 2011). These findings on the apparent link between gut microbes and the central nervous system have sparked investigations into the gut–brain axis. Changes in the gut microbiota have now been linked to microstructural (Liu et al., 2019; Ong et al., 2018), cellular (Möhle et al., 2016), and transcriptional changes (Hoban et al., 2016) in the brain and implicated in a variety of developmental disorders (Sharon et al., 2019; Li et al., 2021) and behavioural phenotypes (Schretter et al., 2018; Wu et al., 2021). Nutritional (Donovan, 2017), metabolic (Bouter et al., 2017) and immunological signals (Foster & McVey Neufeld, 2013) have been identified as possible mediators of the bi-directional cross-talk between microbes and the central nervous system, although the exact mechanisms are still largely unknown.

Perhaps the most established link between gut microbes and host health is the gut microbiota's effect on host metabolism (Turnbaugh et al., 2006; Bäckhead et al., 2004; Ley et al., 2005; Turnbaugh et al., 2007; Ley et al., 2006). In particular, it has been demonstrated that gut microbes contribute to energy extraction from food. Colonisation of GF animals with a

'standard' laboratory mouse microbiota drastically increases body fat within a matter of two weeks, even when food consumption is reduced (Bäckhead et al., 2004). Gut microbiota can promote energy harvest through fermentation of indigestible dietary fibre to monosaccharides and short-chain fatty acids (SCFAs), which can then be absorbed and converted to more complex lipids in the liver followed by lipid deposition in adipocytes (Flint et al., 2012). More recently the gut microbiota has been demonstrated to facilitate adaptation to cold environments (Chevalier et al., 2015). Mice exposed to cold undergo compositional changes in their gut microbiota and transplantation of this 'cold' microbiota to GF mice promotes their tolerance to cold and increases insulin sensitivity (Chevalier et al., 2015). Similar indications of gut microbial adaptation to cold have been observed outside the laboratory: wild mice from colder climates produce larger amounts of bacteria-driven energy sources in relation to food consumption, suggesting their gut microbiota to be more efficient in energy extraction from food (Suzuki et al., 2020).

The value of wild microbiota studies

While most gut microbiota studies have been undertaken on humans or laboratory animals, the gut microbiota has been studied in numerous host species (Yatsunenko et al., 2012; Chung et al., 2012; Blyton et al., 2019; Roeselers et al., 2011; Debebe et al., 2017; Baniel et al., 2022; Rojas et al., 2021; Youngblut et al., 2019), each of which perhaps being best suited as a study system for a specific setting.

Human studies, particularly twin studies, have been pivotal in building an understanding of the gut microbiota, particularly in the context of human biology. However, humans are exposed to environmental factors that are difficult to disentangle from each other and from genetic effects, and are too long-lived to effectively investigate fitness effects of the microbiota. The short-

lived laboratory mouse (*Mus musculus*) is perhaps the most extensively studied mammal, providing a vast set of baseline information on which gut microbiota research can build. As such, laboratory mouse models can be extremely powerful for understanding mechanisms of host–microbiota interaction, by which phenotypic effects arise. A key advantage here, however, is also a major limitation: in order to unpick mechanisms of interest, genetic variation and environmental exposures are deliberately limited and tightly controlled for laboratory mice. Further, as a result of domestication, the gut microbiota of the laboratory mouse is substantially different from that of their wild counterparts. As such, results can be difficult to put into wider context, considering their separation from the natural world.

Studying gut microbiota in the wild can have several advantages. In particular, studying gut microbiota of the wild house mouse holds great potential in increasing our understanding of host–microbiota interactions. First, unlike the laboratory mouse, it is outbred and exposed to environmental factors in a fluctuating way, thus providing a natural platform for investigation of both the drivers and consequences of naturally occurring microbiota variation. Understanding links between how variation arises and what influences it has on host will allow us to understand the evolutionary significance of host–microbiota interactions, for instance, whether and how microbes might provide adaptive plasticity. Second, since the house mouse has been used in the laboratory for decades, a vast collection of robust reagents and related technologies have been designed specifically for this species and can be applied to wild house mouse systems. Third, studying gut microbiota of the wild house mouse allows us to test whether observations from laboratory models are an artefact of the constrained laboratory environment or somewhat different microbiota, or whether they reflect natural processes observed in the wild. Finally, studying ‘wild gut microbiota’ beyond model organisms can

increase our understanding of microbiota's role in ecology and provide key information of host biology with possible conservation applications.

However, moving from the laboratory to the wild can come at a cost. Factors that could shape the gut microbiota under natural conditions, such as age, diet, and health status, can be difficult to determine, and genetic relatedness and social contact between wild mice laboursome to measure. Lack of precise age determination is a particularly key limitation, as age is an important driver of gut microbiota variation (Olm et al., 2022; Stewart et al., 2018; Bosco & Noti, 2021). Further, longitudinal study of wild mice can be limited by poor recapture rate, which could be influenced by predation, anthropogenic interference (e.g., pesticides), and abiotic factors (e.g., temperature). The wild mouse population chosen as the focus for this thesis, and the work undertaken herein, attempts to overcome some of these challenges and develop wild house mouse as a powerful natural model system for mammalian gut microbiota research.

Study systems

The research in this thesis uses an island population of the house mouse (*Mus musculus domesticus*) as a model system. Paralleling studies will be conducted using the most widely used laboratory mouse strain, C57BL/6, in order to gain tools for the wild system, as well as to examine findings from the wild in relation to those from the laboratory.

The laboratory mice were housed in two separate facilities at the University of Oxford and King's College London and the wild mice are part of an island population on Skokholm Island, Wales, UK (Figure 1A). This wild population presents a great opportunity for studying the gut microbiota in a wild setting for several reasons. First, house mouse is the only small mammal

species present on this small island (~100 hectares), making trapping efforts straightforward and meaning there are no interspecific interactions with other small mammals that could affect population size or enable cross-species microbial transmission among closely related species. Second, there are no natural predators of house mice breeding on the island, thus predation is limited and presumably does not have a great impact on recapture of marked individuals.

Third, many mice on this island are likely to live a lifestyle that is completely independent of humans. The island is completely free of human inhabitants for the winter months, while at other times up to 25 people stay on the island in a small number of buildings, which are far from the expected home ranges of many mice on the island. Hence, mice on Skokholm live largely independent of humans unlike many of their mainland relatives, which are often subject to anthropogenic influences, such as pesticides. Lastly, the ecology of the Skokholm mouse population was well studied in the 1960–1970s by Richard Berry and others, providing much baseline knowledge about their history, ecology and population dynamics to build upon (Berry & Jakobson, 1971; Berry, 1968; Berry et al., 1973). Skokholm Island is colonised by thousands of nesting seabirds, such as Manx shearwaters, carcasses of which have been identified to possibly form a part of Skokholm mouse diet (based on camera trap footage) alongside seeds and a range of invertebrates, including amphipods (Berry & Jakobson, 1974).

Two geographically separated but closely located sampling sites were set up on the island in 2019. The '*Observatory'* is located in a central, northern part of the island (Fig. 1B) and surrounds a few buildings, including a bird observatory, after which the site was named. In this sampling site, buildings and human-made walls provide protection from harsh weather, and the communal kitchen, compost, and toilet facilities provide possible sources of food for the mice. The second sampling site, the '*Quarry'* (named after part of this sampling site which is

colloquially known as the Quarry of the island), is located in the southwest of the island (Fig. 1B), near and partly surrounding the lighthouse where the island wardens live. Quarry represents a more exposed habitat: the majority of trapping points there are close to cliffs, thus the site is subject to more harsh wind and even waves from the sea. These two sampling sites were decided upon in discussion with the wardens and after some preliminary trappings in April 2019, to target areas with suspected high mouse density and away from highly sensitive seabird nesting areas.

Each sampling site has 150 fixed trapping points, which were scattered around the two sites with the aim of covering as much of the sites as possible without disturbing other wildlife (a symmetric grid of trapping points was not feasible on the island). Distance between trapping points across the two sites ranges from ~0.5km to ~1.3km. Only 5 out of 337 (1.5%) individual mice caught on the fieldtrips mentioned below were captured at both sampling sites, suggesting mouse migration across the two sites is limited despite geographic proximity. Data for this thesis was collected over five fieldtrips: three in 2019 (spring, summer, autumn), one in 2020 (autumn) and one in 2021 (spring). Three seasonal fieldtrips were planned for 2020 as well; however, it was only possible to conduct one fieldtrip in 2020 due to COVID19 restrictions. Fieldtrips were designed to have a capture-mark-recapture (CMR) study design, where we attempted to conduct a minimum 6 nights of live trapping at each site from summer 2019 onwards. RFID technology (PIT-tagging) and ear notching were used to identify individuals.

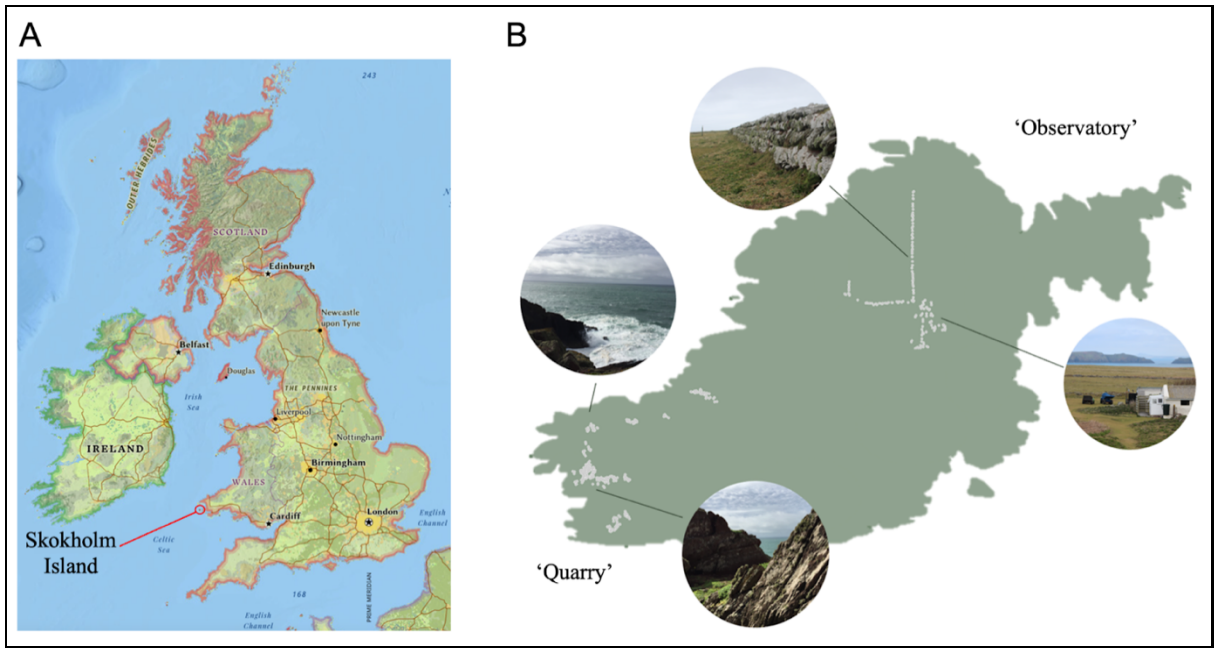


Figure 1. (**A**) Skokholm Island is located 4 km off the coast of Pembrokeshire in south-west Wales, UK. (**B**) Two wild house mouse sampling sites on Skokholm Island. 150 trapping points (white circles) were distributed at each site.

In addition to these two mouse systems – the Skokholm and C57BL/6 mice – samples from additional laboratory mouse strains (n=3) and other wild mouse populations (n=6, including mainland and island populations from Europe and beyond) were acquired from collaborators to put the main study systems into a broader ecological context and test how representative of lab/wild mice they are.

Finally, as a case study of the gut microbiota in the context of a species of conservation concern, I studied the gut microbiota of the critically endangered Saiga antelope, infamous for suffering from recurring mass mortality events (Kock et al., 2018). The Saiga live out of sight in the remote lands of Kazakhstan, the largest land-locked country in the world, making them a difficult species to study. As a result, much of the biology of the Saiga remains unknown. Further, as the mass mortality events have been seemingly caused by overgrowth of *Pasteurella multocida* (Kock et al., 2018), a possible symbiont of the Saiga, the microbiota of the Saiga is of particular interest. Two sub-populations of the Kazakh Saiga were sampled cross-sectionally

in 2019, a year without mass die-offs, in order to characterise the baseline gut microbiota of the Saiga and investigate how the Saiga gut microbiota compares with that of other antelope species.

Research aims and overview of chapters

With the presented study systems, this thesis has four main aims:

I. *Identify drivers of gut microbiota variation in an island house mouse population.*
II. *Optimise and test a lab-based epigenetic clock method for age estimation in wild house mice.*
III. *Investigate gut microbiota assembly patterns in wild and lab mice, and whether they are conserved across these settings.*
IV. *Characterise and assess geographic variation in the gut microbiota of the critically endangered Saiga antelope.*

In Chapter I, I conduct an extensive analysis of gut microbiota variation within wild house mice, focusing on the wild population from Skokholm Island as the main study system. In addition to cross-sectional analyses, I analyse wild house mouse gut microbiota longitudinally, allowing study of within-individual temporal turnover. I show that the gut microbiota of wild mice is shaped by spatial factors, with stronger effects on fungal than bacterial microbiota, and that microbial transmission from soil might explain some of the spatial variation detected. In order to understand gut microbiota variation in house mice across settings, I compare the gut microbiota of Skokholm mice with that of laboratory mice from multiple facilities and strains, as well as other wild house mice from mainland and island populations.

In Chapter II, I attempt to overcome a common limitation of wild small mammal study systems – the lack of accurate age estimation – and estimate chronological age of wild mice using a methylation-based epigenetic clock approach. I show that an epigenetic clock built with faecal samples from laboratory mice can be used to estimate chronological age in these wild mice: the clock successfully predicted adult mice to be older than juvenile mice, and predicted age was generally higher at a later timepoint for mice sampled twice. However, the method did not appear to provide a more accurate estimate of chronological age than body mass, thus I decided not to use predicted ages in subsequent chapters. However, it appeared the method holds promise as a tool for assessing 'biological' age among repeat-captured mice. Further, the framework used in this study (building a clock with samples from captive individuals and using this to estimate age in wild individuals, using faecal samples as a source of host DNA) may be useful when estimating age in other wild species that are hard to capture or even detect.

In Chapter III, I investigate early life gut microbiota assembly dynamics in laboratory and wild mice in parallel. I demonstrate that despite harbouring distinct gut microbiotas on many measures, lab and wild mice share several major patterns of gut microbiota assembly, including similar changes in alpha and beta diversity as well as in the relative abundance of predominant phyla. I argue that these results indicate some degree of intrinsic host programme in gut microbiota assembly that transcends contrasting genetic and environmental backgrounds. As such, these results suggest the broad patterns of gut microbiota assembly are relatively conserved across lab and wild mice, and demonstrate that the widely used lab mouse model – despite their artificial environment and lack of genetic variation – can be used to study age-related microbiota dynamics.

In Chapter IV, I study the gut microbiota in a completely different context and characterise the gut microbiota of the critically endangered Saiga antelope, which has suffered from recurring mass mortality events. It has been suspected that these mass die-offs are caused by outgrowth of a normally commensal symbiont. As such, the Saiga gut microbiota and the possible presence of this symbiont in healthy Saigas is of interest. Here, I describe the general characteristics of the gut microbiota in two geographically distinct populations of the Kazakh Saiga during a 'healthy' year without die-offs, and investigate the extent to which the Saiga gut microbiota resembles that of other antelopes. With this, I demonstrate that the Saiga has a fairly typical gut microbiota for an antelope, that is most similar to a geographically proximate species rather than its closest relative. This characterisation of the Saiga gut microbiota will provide a useful baseline for future work looking into the role of gut microbiota in mammalian mass mortalities.

References

Agus A, Denizot J, Thévenot J, et al. Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. *Sci Rep*. 2016;6:19032. Published 2016 Jan 8. doi:10.1038/srep19032

Albenberg L, Esipova TV, Judge CP, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*. 2014;147(5):1055-63.e8. doi:10.1053/j.gastro.2014.07.020

Almeida A, Mitchell AL, Boland M, et al. A new genomic blueprint of the human gut microbiota. *Nature*. 2019;568(7753):499-504. doi:10.1038/s41586-019-0965-1

Almeida A, Nayfach S, Boland M, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol*. 2021;39(1):105-114. doi:10.1038/s41587-020-0603-3

Arifuzzaman M, Won TH, Li TT, et al. Inulin fibre promotes microbiota-derived bile acids and type 2 inflammation (published online ahead of print, 2022 Nov 2). *Nature*. 2022;10.1038/s41586-022-05380-y. doi:10.1038/s41586-022-05380-y

Bäckhed F, Ding H, Wang T, et al. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A*. 2004;101(44):15718-15723. doi:10.1073/pnas.0407076101

Bäckhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci U S A*. 2007;104(3):979-984. doi:10.1073/pnas.0605374104

Bäckhed F, Roswall J, Peng Y, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life (published correction appears in Cell Host Microbe. 2015 Jun 10;17(6):852. Jun, Wang (corrected to Wang, Jun)) (published correction appears in Cell Host Microbe. 2015 Jun 10;17(6):852). *Cell Host Microbe*. 2015;17(5):690-703. doi:10.1016/j.chom.2015.04.004

Baniel A, Petrullo L, Mercer A, et al. Maternal effects on early-life gut microbiota maturation in a wild nonhuman primate. *Curr Biol*. 2022;32(20):4508-4520.e6. doi:10.1016/j.cub.2022.08.037

Barroso-Batista J, Demengeot J, Gordo I. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat Commun*. 2015;6:8945. Published 2015 Nov 30. doi:10.1038/ncomms9945

Bauer H, Horowitz RE, Levenson SM, Popper H. The response of the lymphatic tissue to the microbial flora. Studies on germfree mice. *Am J Pathol*. 1963;42(4):471-483.

Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell*. 2014;157(1):121-141. doi:10.1016/j.cell.2014.03.011

Berry RJ, Jakobson ME, Triggs GS. Survival in wild-living mice. *Mamm Rev* 1973; 3: 46–57.

Berry RJ, Jakobson ME. Life and death in an island population of the house mouse. *Exp Gerontol*. 1971;6(2):187-197. doi:10.1016/s0531-5565(71)80018-9

Berry RJ, Jakobson ME. Vagility in an island population of the house mouse. *J Zool.* 1974; 173(341–354).

Berry RJ. The ecology of an island population of the house mouse. *J Anim Ecol* 1968; 37: 445–470.

Bittleston LS, Gralka M, Leventhal GE, Mizrahi I, Cordero OX. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nat Commun*. 2020;11(1):1440. Published 2020 Mar 18. doi:10.1038/s41467-020-15169-0

Blyton MDJ, Soo RM, Whisson D, et al. Faecal inoculations alter the gastrointestinal microbiome and allow dietary expansion in a wild specialist herbivore, the koala.

Anim Microbiome. 2019;1(1):6. Published 2019 Aug 21. doi:10.1186/s42523-019-0008-0

Bosco N, Noti M. The aging gut microbiome and its impact on host immunity. *Genes Immun*. 2021;22(5-6):289-303. doi:10.1038/s41435-021-00126-8

Bouter KE, van Raalte DH, Groen AK, Nieuwdorp M. Role of the Gut Microbiome in the Pathogenesis of Obesity and Obesity-Related Metabolic Dysfunction. *Gastroenterology*. 2017;152(7):1671-1678. doi:10.1053/j.gastro.2016.12.048

Britton RA, Young VB. Role of the intestinal microbiota in resistance to colonization by Clostridium difficile. *Gastroenterology*. 2014;146(6):1547-1553. doi:10.1053/j.gastro.2014.01.059

Chevalier C, Stojanović O, Colin DJ, et al. Gut Microbiota Orchestrates Energy Homeostasis during Cold. *Cell*. 2015;163(6):1360-1374. doi:10.1016/j.cell.2015.11.004

Chung H, Pamp SJ, Hill JA, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*. 2012;149(7):1578-1593. doi:10.1016/j.cell.2012.04.037

Crawford PA, Crowley JR, Sambandam N, et al. Regulation of myocardial ketone body metabolism by the gut microbiota during nutrient deprivation. *Proc Natl Acad Sci U S A*. 2009;106(27):11276-11281. doi:10.1073/pnas.0902366106

David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505(7484):559-563. doi:10.1038/nature12820

Debebe T, Biagi E, Soverini M, et al. Unraveling the gut microbiome of the long-lived naked mole-rat. *Sci Rep*. 2017;7(1):9590. Published 2017 Aug 29. doi:10.1038/s41598-017-10287-0

Diaz Heijtz R, Wang S, Anuar F, et al. Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A*. 2011;108(7):3047-3052. doi:10.1073/pnas.1010529108

DiGiulio DB, Callahan BJ, McMurdie PJ, et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A*. 2015;112(35):11060-11065. doi:10.1073/pnas.1502875112

Donovan SM. Introduction to the special focus issue on the impact of diet on gut microbiota composition and function and future opportunities for nutritional modulation of the gut microbiome to improve human health. *Gut Microbes*. 2017;8(2):75-81. doi:10.1080/19490976.2017.1299309

Falony G, Joossens M, Vieira-Silva S, et al. Population-level analysis of gut microbiome variation. *Science*. 2016;352(6285):560-564. doi:10.1126/science.aad3503

Ferretti P, Pasolli E, Tett A, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe*. 2018;24(1):133-145.e5. doi:10.1016/j.chom.2018.06.005

Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol*. 2012;9(10):577-589. Published 2012 Sep 4. doi:10.1038/nrgastro.2012.156

Foster JA, McVey Neufeld KA. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci*. 2013;36(5):305-312. doi:10.1016/j.tins.2013.01.005

Friedman ES, Bittinger K, Esipova TV, et al. Microbes vs. chemistry in the origin of the anaerobic gut lumen (published correction appears in Proc Natl Acad Sci U S A. 2022 Jun 14;119(24):e2207826119). *Proc Natl Acad Sci U S A*. 2018;115(16):4170-4175. doi:10.1073/pnas.1718635115

Fujimura KE, Sitarik AR, Havstad S, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med*. 2016;22(10):1187-1191. doi:10.1038/nm.4176

Gacesa R, Kurilshikov A, Vich Vila A, et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature*. 2022;604(7907):732-739. doi:10.1038/s41586-022-04567-7

Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science*. 2016;352(6285):539-544. doi:10.1126/science.aad9378

Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. *Nat Med*. 2018;24(4):392-400. doi:10.1038/nm.4517

Gill SR, Pop M, Deboy RT, et al. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006;312(5778):1355-1359. doi:10.1126/science.1124234

Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159(4):789-799. doi:10.1016/j.cell.2014.09.053

Grieneisen L, Dasari M, Gould TJ, et al. Gut microbiome heritability is nearly universal but environmentally contingent. *Science*. 2021;373(6551):181-186. doi:10.1126/science.aba5483

Hapfelmeier S, Lawson MA, Slack E, et al. Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science*. 2010;328(5986):1705-1709. doi:10.1126/science.1188454

Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, et al. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*. 2009;137(5):1716-24.e242. doi:10.1053/j.gastro.2009.08.042

Hoban AE, Stilling RM, Ryan FJ, et al. Regulation of prefrontal cortex myelination by the microbiota. *Transl Psychiatry*. 2016;6(4):e774. Published 2016 Apr 5. doi:10.1038/tp.2016.42

Ivanov II, Atarashi K, Manel N, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. 2009;139(3):485-498. doi:10.1016/j.cell.2009.09.033

Jian C, Luukkonen P, Yki-Järvinen H, Salonen A, Korpela K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One*. 2020;15(1):e0227285. Published 2020 Jan 15. doi:10.1371/journal.pone.0227285

Johnson JS, Spakowicz DJ, Hong BY, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10(1):5029. Published 2019 Nov 6. doi:10.1038/s41467-019-13036-1

Jokela R, Korpela K, Jian C, et al. Quantitative insights into effects of intrapartum antibiotics and birth mode on infant gut microbiota in relation to well-being during the first year of life. *Gut Microbes*. 2022;14(1):2095775. doi:10.1080/19490976.2022.2095775

Kamada N, Chen GY, Inohara N, Núñez G. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol*. 2013;14(7):685-690. doi:10.1038/ni.2608

Knowles SCL, Eccles RM, Baltrūnaitė L. Species identity dominates over environment in shaping the microbiota of small mammals. *Ecol Lett*. 2019;22(5):826-837. doi:10.1111/ele.13240

Kock RA, Orynbayev M, Robinson S, et al. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv*. 2018;4(1):eaao2314. Published 2018 Jan 17. doi:10.1126/sciadv.aao2314

Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species /distinguishability and phylosymbiosis along the length of the gastrointestinal tract. *Mol Ecol*. 2018;27(8):1874-1883. doi:10.1111/mec.14460

Koren O, Goodrich JK, Cullender TC, et al. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*. 2012;150(3):470-480. doi:10.1016/j.cell.2012.07.008

Kurilshikov A, Medina-Gomez C, Bacigalupe R, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021;53(2):156-165. doi:10.1038/s41588-020-00763-1

Lancaster SM, Lee-McMullen B, Abbott CW, et al. Global, distinctive, and personal changes in molecular and microbial profiles by specific fibers in humans. *Cell Host Microbe*. 2022;30(6):848-862.e7. doi:10.1016/j.chom.2022.03.036

Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500(7464):541-546. doi:10.1038/nature12506

Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*. 2005;102(31):11070-11075. doi:10.1073/pnas.0504978102

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature*. 2006;444(7122):1022-1023. doi:10.1038/4441022a

Li S, Song J, Ke P, et al. The gut microbiome is associated with brain structure and function in schizophrenia (published correction appears in Sci Rep. 2021 Aug 30;11(1):17643). *Sci Rep*. 2021;11(1):9743. Published 2021 May 7. doi:10.1038/s41598-021-89166-8

Liu Y, Li F, Luo H, et al. Improvement of Deep Brain Stimulation in Dyskinesia in Parkinson's Disease: A Meta-Analysis. *Front Neurol*. 2019;10:151. Published 2019 Feb 25. doi:10.3389/fneur.2019.00151

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489(7415):220-230. doi:10.1038/nature11550

Mallott EK, Amato KR. Host specificity of the gut microbiome. *Nat Rev Microbiol*. 2021;19(10):639-653. doi:10.1038/s41579-021-00562-3

Marsh KJ, Raulo AM, Brouard M, et al. Synchronous Seasonality in the Gut Microbiota of Wild Mouse Populations. *Front Microbiol*. 2022;13:809735. Published 2022 Apr 25. doi:10.3389/fmicb.2022.809735

Martinez-Medina M, Denizot J, Dreux N, et al. Western diet induces dysbiosis with increased E coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. *Gut*. 2014;63(1):116-124. doi:10.1136/gutjnl-2012-304119

Maurice CF, Knowles SC, Ladau J, et al. Marked seasonal variation in the wild mouse gut microbiota. *ISME J*. 2015;9(11):2423-2434. doi:10.1038/ismej.2015.53

McFall-Ngai M, Hadfield MG, Bosch TC, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A*. 2013;110(9):3229-3236. doi:10.1073/pnas.1218525110

Möhle L, Mattei D, Heimesaat MM, et al. Ly6C(hi) Monocytes Provide a Link between Antibiotic-Induced Changes in Gut Microbiota and Adult Hippocampal Neurogenesis. *Cell Rep*. 2016;15(9):1945-1956. doi:10.1016/j.celrep.2016.04.074

Murillo, T., Schneider, D., Fichtel, C. *et al.* Dietary shifts and social interactions drive temporal fluctuations of the gut microbiome from wild redfronted lemurs. *ISME COMMUN.* 2, 3 (2022). https://doi.org/10.1038/s43705-021-00086-0

Ochman H, Worobey M, Kuo CH, et al. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol*. 2010;8(11):e1000546. Published 2010 Nov 16. doi:10.1371/journal.pbio.1000546

Olm MR, Dahan D, Carter MM, et al. Robust variation in infant gut microbiome assembly across a spectrum of lifestyles. *Science*. 2022;376(6598):1220-1223. doi:10.1126/science.abj2972

Ong IM, Gonzalez JG, McIlwain SJ, et al. Gut microbiome populations are associated with structure-specific changes in white matter architecture. *Transl Psychiatry*. 2018;8(1):6. Published 2018 Jan 10. doi:10.1038/s41398-017-0022-5

Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol*. 2007;5(7):e177. doi:10.1371/journal.pbio.0050177

Phillips CD, Phelan G, Dowd SE, et al. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol Ecol*. 2012;21(11):2617-2627. doi:10.1111/j.1365-294X.2012.05568.x

Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55-60. doi:10.1038/nature11450

Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*. 2021;591(7851):633-638. doi:10.1038/s41586-021-03241-8

Raulo A, Allen BE, Troitsky T, et al. Social networks strongly predict the gut microbiota of wild mice. *ISME J*. 2021;15(9):2601-2613. doi:10.1038/s41396-021-00949-3

Risely A, Wilhelm K, Clutton-Brock T, Manser MB, Sommer S. Diurnal oscillations in gut bacterial load and composition eclipse seasonal and lifetime dynamics in wild meerkats. *Nat Commun*. 2021;12(1):6017. Published 2021 Oct 14. doi:10.1038/s41467-021-26298-5

Roeselers G, Mittge EK, Stephens WZ, et al. Evidence for a core gut microbiota in the zebrafish. *ISME J*. 2011;5(10):1595-1608. doi:10.1038/ismej.2011.38

Rojas CA, Ramírez-Barahona S, Holekamp KE, Theis KR. Host phylogeny and host ecology structure the mammalian gut microbiota at different taxonomic scales. *Anim Microbiome*. 2021;3(1):33. Published 2021 Apr 23. doi:10.1186/s42523-021-00094-4

Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018;555(7695):210-215. doi:10.1038/nature25973

Rudolph K, Schneider D, Fichtel C, Daniel R, Heistermann M, Kappeler PM. Drivers of gut microbiome variation within and between groups of a wild Malagasy primate. *Microbiome*. 2022;10(1):28. Published 2022 Feb 9. doi:10.1186/s40168-021-01223-6

Schretter CE, Vielmetter J, Bartos I, et al. A gut microbial factor modulates locomotor behaviour in Drosophila. *Nature*. 2018;563(7731):402-406. doi:10.1038/s41586-018-0634-9

Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol*. 2016;14(8):e1002533. Published 2016 Aug 19. doi:10.1371/journal.pbio.1002533

Sharon G, Cruz NJ, Kang DW, et al. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell*. 2019;177(6):1600-1618.e17. doi:10.1016/j.cell.2019.05.004

Smits SA, Leach J, Sonnenburg ED, et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*. 2017;357(6353):802-806. doi:10.1126/science.aan4834

Song SJ, Sanders JG, Delsuc F, et al. Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *mBio*. 2020;11(1):e02901-19. Published 2020 Jan 7. doi:10.1128/mBio.02901-19

Sonnenburg ED, Zheng H, Joglekar P, et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell*. 2010;141(7):1241-1252. doi:10.1016/j.cell.2010.05.005

Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol*. 2011;9(4):279-290. doi:10.1038/nrmicro2540

Sprockett D, Fukami T, Relman DA. Role of priority effects in the early-life assembly of the gut microbiota. *Nat Rev Gastroenterol Hepatol*. 2018;15(4):197-205. doi:10.1038/nrgastro.2017.173

Stecher B, Hardt WD. Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol*. 2011;14(1):82-91. doi:10.1016/j.mib.2010.10.003

Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-588. doi:10.1038/s41586-018-0617-x

Sudo N, Chida Y, Aiba Y, et al. Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. *J Physiol*. 2004;558(Pt 1):263-275. doi:10.1113/jphysiol.2004.063388

Suzuki T, Suzuki N, Tojo K. Parallel evolution of an alpine type ecomorph in a scorpionfly: Independent adaptation to high-altitude environments in multiple mountain locations. *Mol Ecol*. 2019;28(13):3225-3240. doi:10.1111/mec.15119

Suzuki TA, Martins FM, Phifer-Rixey M, Nachman MW. The gut microbiota and Bergmann's rule in wild house mice. *Mol Ecol*. 2020;29(12):2300-2311. doi:10.1111/mec.15476

Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. *Biol Lett*. 2014;10(2):20131037. Published 2014 Feb 12. doi:10.1098/rsbl.2013.1037

Tung J, Barreiro LB, Burns MB, et al. Social networks predict gut microbiome composition in wild baboons. *Elife*. 2015;4:e05224. Published 2015 Mar 16. doi:10.7554/eLife.05224

Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe*. 2008;3(4):213-223. doi:10.1016/j.chom.2008.02.015

Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;457(7228):480-484. doi:10.1038/nature07540

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449(7164):804-810. doi:10.1038/nature06244

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. 2006;444(7122):1027-1031. doi:10.1038/nature05414

Valles-Colomer M, Falony G, Darzi Y, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol*. 2019;4(4):623-632. doi:10.1038/s41564-018-0337-x

Vandeputte D, De Commer L, Tito RY, et al. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat Commun*. 2021;12(1):6740. Published 2021 Nov 18. doi:10.1038/s41467-021-27098-7

Vandeputte D, Kathagen G, D'hoe K, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*. 2017;551(7681):507-511. doi:10.1038/nature24460

Vrieze A, Van Nood E, Holleman F, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome (published correction appears in Gastroenterology. 2013 Jan;144(1):250). *Gastroenterology*. 2012;143(4):913-6.e7. doi:10.1053/j.gastro.2012.06.031

Walter J, Hornef MW. A philosophical perspective on the prenatal in utero microbiome debate. *Microbiome*. 2021;9(1):5. Published 2021 Jan 12. doi:10.1186/s40168-020-00979-7

Wang J, Linnenbrink M, Künzel S, et al. Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiome of wild

mice. *Proc Natl Acad Sci U S A*. 2014;111(26):E2703-E2710. doi:10.1073/pnas.1402342111

Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472(7341):57-63. doi:10.1038/nature09922

Weinstein SB, Martínez-Mota R, Stapleton TE, et al. Microbiome stability and structure is governed by host phylogeny over diet and geography in woodrats (*Neotoma* spp.). *Proc Natl Acad Sci U S A*. 2021;118(47):e2108787118. doi:10.1073/pnas.2108787118

Wu WL, Adame MD, Liou CW, et al. Microbiota regulate social behaviour via stress response neurons in the brain. *Nature*. 2021;595(7867):409-414. doi:10.1038/s41586-021-03669-y

Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-227. Published 2012 May 9. doi:10.1038/nature11053

Youngblut ND, Reischer GH, Walters W, et al. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun*. 2019;10(1):2200. Published 2019 May 16. doi:10.1038/s41467-019-10191-3

Zhang W, Jiao L, Liu R, et al. The effect of exposure to high altitude and low oxygen on intestinal microbial communities in mice. *PLoS One*. 2018;13(9):e0203701. Published 2018 Sep 12. doi:10.1371/journal.pone.0203701

CHAPTER I

Drivers of gut microbiota variation in an island population of the house mouse, *Mus musculus domesticus*

Authors

Eveliina Hanski[1], Susan Joseph[2], Michael A Curtis[2], Wieteke Holthuijzen[3], Jonathan H Plissner[4], Jens-Kjeld Jensen[5], Aura Raulo[1,6], James Swann[7], John F Baines[8], Maude Quicray[9], Kathryn J Else[10], Janette E Bradley[11], Sarah CL Knowles[1]

Title

Drivers of gut microbiota variation in an island population of the house mouse, *Mus musculus domesticus*

Abstract

The mammalian gut microbiota is a complex community of microbes that has diverse impacts on host biology. House mice constitute the model organism for research on mammals, but laboratory domestication has altered their gut microbiota from that of their wild counterparts. Knowledge about what shapes the gut microbiota of house mice under natural conditions remains limited, but would improve this species' utility as a model organism both inside and outside the lab. Here, we use a large, longitudinal dataset to investigate the factors predicting natural gut microbiota variation within a single wild house mouse population, inhabiting the Skokholm Island, Wales. We reveal prominent temporal and small-scale spatial variation in microbiota composition (with larger spatial effects on fungal than bacterial components of the microbiota), and identify microbial transmission from soil as one possible source of this spatial variation. To put our findings in a wider ecological context, we compare gut microbiota composition and dynamics of Skokholm mice to those of multiple laboratory mouse strains and other wild house mice from mainland and island populations. While the microbiota of Skokholm mice is broadly representative of other wild populations, collectively wild mice possess a microbiota that is not only compositionally distinct, taxonomically and functionally

more diverse than that of lab mice, but also has a faster temporal turnover rate. These findings highlight the much greater spatiotemporal turnover of gut microbes in wild mice compared to laboratory strains, which may affect how they influence host phenotypes in these contrasting settings.

Introduction

The mammalian gut houses a diverse collection of microbial organisms known as the gut microbiota, that provides many important functions for the host. It is involved in several developmental processes, such as growth, immune maturation, and development of the central nervous system (Sharon et al., 2016; Chung et al., 2012; Yan et al., 2016), but also in host processes beyond early life, such as regulation of the immune system, metabolism, and protection against pathogens (Round & Mazmanian, 2009; Fan & Pederson, 2021; Pickard et al., 2017). Most insights into host-microbiota interactions in mammals come from human or laboratory animal studies. In particular, the laboratory mouse (*Mus musculus*) is a widely used model system in gut microbiota research and mammalian research more broadly. It is a powerful system in which virtually everything from genetic background to diet can be tightly controlled, biological processes can be studied across the lifespan as mice are short-lived, and the gut microbiota can be readily modulated, for instance by colonizing germ-free mice with specific combinations of microbes.

However, laboratory mice live in an artificial world where individuals are typically inbred, housed under stable environmental conditions and exposed to a limited number of other individuals, raising concerns about the relevance of gut microbiota findings from laboratory mice, and the influence lab-adapted gut microbiotas may have on other mouse research. Indeed, laboratory mice have a somewhat different gut microbiota from that of their wild house mouse

relatives (Wang et al., 2014; Kreisinger et al., 2014; Wang et al., 2015, Rosshart el al., 2019; Bowerman et al., 2021) and results from recent experiments where laboratory mice were colonised with wild-derived microbiota were shown to have greater applicability for understanding human health (Rosshart, et al., 2019). Considering the model organism status of the house mouse, understanding natural drivers of wild house mouse microbiota is pivotal. An enhanced understanding of house mouse gut microbiota across different genetic and environmental backgrounds is also important for the next generation of wild-reconstituted model organisms (Thomson et al., 2022).

Previous studies on the gut microbiota of wild house mice have indicated geography to be one of the most influential drivers of variation, while individual-level attributes such as age and sex appear to have a lesser effect (Goertz et al., 2019; Weldon et al., 2015; Linnenbrink et al., 2013). Compositional and diversity differences are seen in mice from different countries/regions (Linnenbrink et al., 2013; Weldon et al., 2015) and altitudes (Suzuki et al., 2019) but also on a fine spatial scale within a population (Goertz et al., 2019), with possible downstream effects on the host. For instance, mice from higher altitudes had higher relative abundances of anaerobic bacteria better adapted for environments with limited oxygen, and this 'high altitude microbiota' is enriched in a functional pathway involved in regulation of blood pressure (Suzuki et al., 2019).

Here, we build upon previous work and investigate drivers of gut microbiota variation in a wild house mouse (*Mus musculus domesticus*) population on Skokholm Island, Wales. Using a large dataset containing nearly 1,000 samples from over 300 individual mice, we disentangle the relative importance of different drivers of variation in wild house mouse microbiota. In addition to cross-sectional analysis, we use repeat-sampled individuals to analyse temporal variability

of the wild house mouse gut microbiota for the first time. We further expand on previous studies by including analysis of the fungal microbiota ('mycobiota'; Rosshart et al., 2019), as well as by exploring associations between microbial communities from the natural environment (soil) and the gut that could indicate microbial transmission between these two habitats. To put these findings in context and gain further understanding of house mouse variation and ecology across settings, we also compare microbiota composition and dynamics of wild Skokholm mice with both other wild-captured house mice other islands and mainland settings, as well as those of laboratory mice of various strains, from multiple facilities.

We hypothesised that due to contrasting environmental and genetic backgrounds as well as different microbial exposure patterns, laboratory and wild mice have distinct gut microbiotas in terms of taxonomy, functionality as well as stability. Further, we hypothesised that dispersal of gut microbes might be reduced in island mice living in closed ecosystems compared to mainland mice, and that island and mainland mice might have differences in diet, possibly resulting in different gut microbiota characteristics between mainland and island mice. Finally, we hypothesised that spatial effects on gut microbiota composition are detectable within an island population of the house mouse, with effects on both bacterial and fungal components, and that these spatial effects might be partly driven by variation in environmental microbiotas.

Methods

Sample collection

Six laboratory mouse (*Mus musculus*) colonies from three animal facilities were sampled between November 2020 and May 2021. These included a C57BL/6 colony from King's College, London, C57BL/6 and transgenic SKG colonies from Kennedy Institute, Oxford, and C57BL/6, transgenic CCSP-rtTA and transgenic Pdgfra-creER colonies from BMS, Oxford

(Supplementary Table 1). SKG mice presented intestinal inflammation that had been induced with curdlan injection while all other sampled mice were not subject to interventions. For sample collection, mice were placed on a sterile surface until defaecation. Faecal pellets were immediately preserved in DNA/RNA Shield and stored in -80ºC until DNA extraction.

Wild house mouse (*Mus musculus domesticus*) samples originated from seven populations. The main study population, the Skokholm Island population, was sampled in April–May 2019, July 2019, September–October 2019, August–September 2020, and April–May 2021 on Skokholm Island in Wales, UK, yielding a total of 948 samples from 337 unique mice (1–12 samples per mouse; Suppl. Table 1).

Mice were live-trapped with small Sherman traps with a standardised weight of peanuts (4g) as bait and non-absorbent cotton wool as bedding. Traps were set at dusk and collected at dawn. Traps were set at two geographically separate but closely located sampling sites ('Observatory' and 'Quarry') and trapping was generally conducted in sets of three consecutive nights at one site, after which traps were set for three nights at the other site. Traps that a mouse had been inside were washed thoroughly and sterilised with 20% bleach solution before being re-used. Captured mice were tagged with a subcutaneous passive integrated transponder (PIT) tag for permanent identification, or identified through PIT tag detection upon recapture. Mice were then aged, sexed, measured (from tip of nose to base of tail) and weighed before release within 2 meters of their trapping point. Aging was carried out based on body size and appearance; small (typically <15g of mass and <80mm of length) were ranked as juveniles, full-sized mice as adults (typically >20g and >80mm of length) and anything in between as sub-adults. Sex was determined based on visual inspection and anogenital distance. Body condition score was

determined on a scale from 0–4 by palpating the lower spine and hips to estimate the amount of subcutaneous fat.

Faecal samples were collected from traps in a sterile manner and preserved in DNA/RNA Shield (Zymo Research, Irvine, California, USA). Samples were stored at -20ºC while on the island (up to 6 weeks), after which samples were transported back to the lab and stored at -80ºC until DNA extraction (up to 17 months). In order to investigate the effect of preservative on sample-level gut microbiota profiles, replicate aliquots were collected for a subset of 15 Skokholm Island faecal samples, from 15 separate individuals. Aliquots were stored in either DNA/RNA Shield, RNAlater, absolute ethanol, or without preservative, and then stored and transported as described above. At DNA extraction, all four replicates from a given sample were extracted in the same batch, and processed in a single round of library preparation and amplicon sequencing.

Gut microbiota samples from other wild mouse populations (n=6) were acquired through collaboration. As some of these samples had already been collected, sample type and preservation system varied between populations (Suppl. Table 1). Samples from Midway Atoll and the Faroe Islands were large intestinal contents taken from dissected animals and samples from Cologne, Espelette, Isle of May, and Oxford were faecal pellets. Further, while samples from Skokholm, Wytham, and all laboratory colonies were preserved in DNA/RNA Shield, samples from other wild mice had been preserved differently: Midway Atoll and Faroe Islands samples had been stored in -20ºC or -80ºC in isopropyl alcohol (Midway Atoll) or without preservative (Faroe Islands), however, samples were preserved with DNA/RNA Shield before sample shipping to the UK. Samples from the Isle of May had been stored at -20ºC or -80ºC without preservative. Cologne samples were stored in RNAlater and Espelette samples in PBS.

Others have demonstrated limited differences between large intestinal and faecal microbiota compositions in mice (Kohl et al., 2018; Suzuki & Nachman, 2016; Anders et al., 2021; Knowles et al., 2019). Further, our analysis on the effect of preservative (see above) demonstrated preservative to have limited effect on microbiota composition: samples strongly clustered by mouse ID (Suppl. Figure 1), which explained 75.7% of microbiota variation (PERMANOVA on Aitchison distance, p=0.001; beta dispersion, F=2.4151, p=0.018), while type of preservative only explained 2% of microbiota variation (p=0.020; beta dispersion, F=0.1523, p=0.926). Thus, we expect differences originating from distinct sample types or preservatives to be limited in our study.

To investigate how transmission from the natural environment might affect the gut microbiota of wild house mice, we collected soil samples from both sampling sites ('Quarry' or 'Observatory') on Skokholm Island. The full set of trapping points (n=150) at each sampling site were divided into 6–7 'sub-sites' (6 at one site and 7 at the other) with different types of habitats (1–9 trapping points per sub-site), and 16% of trapping points within each sub-site were then randomly selected for soil sampling for a total of 25 trapping points per sampling site, and 50 in total across the two sampling sites. At each of these points, two soil samples were collected one week apart in September 2020. Soil samples were collected within 50 cm proximity of the trapping point, from approximately 2 cm depth (or shallower where 2 cm was not feasible, such as on rocks) using a sterile metal spatula. Samples were preserved and stored in an identical way to faecal samples from Skokholm (see above).

DNA extraction, library preparation, and sequencing

DNA was extracted from the samples using ZymoBIOMICS DNA MiniPrep kits in 24-tube format according to manufacturer's instructions (sample type-specific instructions were

followed for faecal and soil samples; Zymo Research, Irvine, California, USA). Samples were randomised into 64 extraction batches of up to 23 samples. A negative control (an equivalent volume, 40µL, of DNAse-free H_2O) was included in every extraction batch at varying positions, with the exception of one extraction batch in which a negative control was not included. For samples preserved with DNA/RNA Shield, DNA/RNA Shield was used as lysis solution in the first step of DNA extraction. For other faecal samples, any preservative was removed by centrifuging and pipetting, after which ZymoBIOMICS Lysis Solution was added in the first step of DNA extraction.

Library preparation and amplicon sequencing was completed by the Integrated Microbiome Resource (IMR), Dalhouise University, using the protocol described in Comeau et al. (2017). Briefly, the V4–V5 region of the bacterial 16S rRNA gene was targeted using primers 515(F) and 926(R) (Parada et al., 2016; Walters et al., 2015). For a subset of 60 Skokholm mouse faecal samples, the mycobiota was also characterised, targeting the fungal ITS2 in amplicon sequencing using primers ITS86(F) and ITS4(R) primers (Op De Beeck et al., 2014), also at the IMR. Amplification and sequencing of 16S rRNA gene was conducted in 16 batches of up to 95 samples using the Illumina MiSeq platform (Reagent kit v3, 2x300 bp chemistry). Amplification and sequencing of ITS2 gene conducted in a single batch with the same sequencing platform. All extraction controls (n=63) were sequenced, and a negative PCR control was included for each batch of library preparation (n=16).

Data selection

All laboratory mice sampled were adult (>3 months of age) and reproductively inactive. For comparisons of laboratory and wild mice, to maximise the comparability of wild mice with laboratory mice in terms of age and reproductive state, we excluded wild mice (1) classed as

juvenile or sub-adult and (2) that showed signs of active reproduction (*females*: pregnancy, prominent nipples, lactation, perforate vagina; *males*: large testes). As environmental factors (e.g., temperature, humidity, food type and availability) are stable in the laboratory, we only included wild mice sampled in September, October, or November to minimise seasonal variation (Maurice et al., 2015; Marsh et al., 2022), when sampling date was known (samples from Skokholm Island, Midway Atoll, Isle of May, and Wytham; sampling date was unknown for samples from Faroe Islands, Espelette and Cologne). This sub-sampling is unlikely to have had a major impact on lab-wild comparisons, as the proportion of shared taxa between lab and wild mice (Jaccard similarity) only showed slight variation according to the month wild mice were sampled (Suppl. Fig. 2; permutational Wilcoxon rank sum test for July vs Oct, p=0.452, April vs Oct, p<0.001).

For longitudinal analysis of lab and wild mouse samples, three subsets of data were selected: (1) the above lab-wild dataset was subsetted to include individual mice sampled >3 times. The maximum sampling interval for wild mice in this data subset was 40 days, thus lab mice with >40-day sampling interval were excluded to achieve a comparable time window. (2) The same analysis was repeated with all repeat-sampled lab and wild mice without exclusion based on age, reproductive status, sampling season or sampling interval. (3) To investigate ASV persistence within individuals on comparable timescales in the lab and in the wild, we selected mice from both settings that had been sampled 4–5 times at approximately one-week intervals (regardless of sampling month or year and reproductive status). For the mainland vs island analyses, same data selection was applied, as above for lab vs wild analyses (though herein only wild mice were included), i.e., only samples from September–November months were included, and samples from mice that had recordings of reproductive activity were excluded.

Data processing of bacterial data

Data processing and analyses were carried out in R version 4.1.2 (R Core Team, 2021). Sequences were denoised and amplicon sequence variants (ASVs) inferred using DADA2 version 1.16 (Callahan et al., 2016), with the pipeline run separately for wild mouse, laboratory mouse and soil samples. Taxonomy was assigned using the SILVA rRNA database version 138. R packages DECIPHER and phangorn were used to build a phylogenetic tree, and package iNEXT (Hsieh et al., 2022; Chao et al., 2014) was used to generate sample completeness and rarefaction curves. The read depth threshold was set at 4,000, based on where these curves plateaued. Samples falling below this threshold were excluded from further analysis. Data were not rarefied.

Asymptotic ASV richness and Shannon diversity were equally estimated with iNEXT. Singleton and doubleton ASVs were separately removed from lab and wild mouse datasets, and non-gut microbial ASVs (chloroplasts, mitochondria) were removed from mouse datasets. 464 unique ASVs were detected across 63 DNA extraction controls and 16 library preparation controls, with 0–1,394 reads (mean 75, median 13) for any given control, with the exception of one control for library preparation, which had 255 unique ASVs and 29,300 reads (mean read count of biological samples was 29,235; range 5,662–232,225, median 26,097). All extraction controls from the same 96-well plate (n=5) had <20 reads each, indicating the entire plate was not contaminated during library preparation but rather that the most likely explanation is that a biological sample was mistakenly pipetted into the control well in addition to its designated well. Since all other controls on this plate were negative, we retained this plate of samples in our analyses.

R package decontam was used to test for the presence of potential contaminants. The package was used with the 'prevalence' (presence/absence) method, in which each sequence in biological samples was compared to the corresponding prevalence in negative controls and considered a contaminant if it reached a probability of 0.1 in a Fisher's exact test. With this, we identified 31 contaminants which were removed from the dataset. R package phyloseq (McMurdie et al., 2013) was used to normalise ASV counts to proportional abundance.

Data processing of fungal data

Fungal data was processed as the bacterial data (above), with a few differences: as a processing pipeline, DADA2 ITS workflow version 1.8 was used (Callahan et al., 2016). The fungal ITS2 gene was not sequenced for the negative controls, thus data could not be decontaminated. However, extraction batches in which these samples were processed included negative controls that did not present significant bacterial contamination (above), thus fungal contamination was similarly unlikely. As fungal sequences in faecal samples can include those derived from ingested macrofungi, which can be difficult to separate from symbiotic fungi (Lavrinienko et al., 2021), sequence filtering was conducted to only include putative symbiont sequences. For this, we used BLAST version 2.2.28+ (word size = 20) to scan sequences against Targeted Host-associated Fungi ITS Database version 1.6.1, which includes fungal symbionts previously detected in human or mouse samples. Sequences with an identity percentage ≥97% and a total length of matched segment >180bp (335 out of 1115 (30.0%)) were considered symbiont-derived.

Functional profiles

Functional pathways were predicted from the 16S rRNA data using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2) version 2.5.0 pipeline

(picrust2_pipeline.py) using default options (Douglas et al., 2020) and MetaCyc Metabolic database. Read counts of functional pathways were normalised to relative abundance.

Analysis

Alpha and beta diversity differences across lab and wild, mainland and island, and sub-populations of Skokholm mice were tested with Wilcoxon rank sum tests with 1,000 permutations. Principal coordinates analysis (PCoA) was performed using phylogenetically uninformed (Jaccard, Aitchison) and phylogenetically-informed (unweighted and weighted UniFrac) distances. A centered log-ratio (clr) transformation was performed before ordination on Aitchison distance, using the R package microbiome (Lahti & Shetty, 2017). In clr-transformation, zero relative abundances were replaced with a pseudocount as follows: *min(relative abundance/2)*.

Further measures of beta diversity included a pairwise dissimilarity analysis among samples (one random sample per individual) using R package microbiome, and marginal permutational multivariate analyses of variance (PERMANOVA) on Jaccard and clr-transformed Aitchison distances using the *adonis2* function from R package vegan (Oksanen et al., 2022). In PERMANOVA a range of spatiotemporal and individual-level factors were included as explanatory factors. Month was fitted as a year-specific multilevel factor (e.g., 04/2019), sampling site as a 2-level factor (Quarry or Observatory), minimum distance to a building or sealine as continuous variables, age as a 3-level factor (juvenile/sub-adult/adult), body mass as a continuous variable, body condition as a multilevel factor, and sex and reproductive activity as a 2-level factors (female/male, active/inactive). Interaction terms between e.g., sex and reproductive activity were fitted with separate models. Beta dispersion were tested using the betadisper function in package vegan.

Random Forest regressions were used to identify key taxa driving differences between laboratory and wild mice and between mainland and island mice, with mean decrease in Gini used as a measure of importance. A Mantel test was used to measure correlation between geographic proximity and gut microbial dissimilarity between (1) mouse-mouse, (2) mouse-soil, and (3) soil-soil sample pairs, using the R package ade4. A brm model from the R package brms (Bürkner, 2017) was used to investigate whether (1) sampling site or (2) geographical proximity predicted fungal dissimilarity more strongly than bacterial dissimilarity. For this, pairwise Jaccard dissimilarity (measured separately for bacteria and fungi) was used as response variable, with either sampling site or geographic proximity, together with their interaction with microbiota component (bacteria/fungi) as predictors. In both models, a multi-membership random effect was included to account for the non-independence of dyadic data.

Ethical statement

Samples from transgenic mice were collected under Home Office licence PPL PP0574716 held at the University of Oxford. Skokholm wild mouse samples were collected under Home Office license PPL PB0178858 held at the University of Oxford, and with a research permit from the Islands Conservation Advisory Committee (ICAC), and Natural Resources Wales. Protocol for soil sampling was discussed with Skokholm Island wardens prior to sample collection and conducted with minimal disturbance to vegetation.

Results

Wild mice have compositionally distinct and taxonomically and functionally more diverse gut microbiota than laboratory mice

Across six laboratory mouse colonies and seven wild mouse populations, wild mice had significantly higher alpha diversity than lab mice overall (both estimated ASV richness and

Shannon diversity; permutational Wilcoxon rank sum test, $p<0.001$; wild, $n=436$; lab, $n=146$). However, alpha diversity varied greatly between colonies and populations such that it was not consistently higher in every wild-laboratory population comparison (Fig. 1A, Suppl. Fig. 3). Among laboratory mice, alpha diversity was highest in transgenic SKG mice ($n=3$), which had curdlan-induced intestinal inflammation. Due to lack of steady state SKG mice, this effect could not be separated from the effect of strain (steady-state SKG mice have also been reported to present microbiota alterations; Rehaume et al., 2014). Alpha diversity was similarly high in steady state transgenic CCSP-rtTA mice ($n=7$), a strain in which activation of specific genes in the respiratory epithelium can be induced (Perl et al., 2009).

The gut microbiota of wild mice was also more compositionally variable among individuals within each population than for lab mice (Jaccard dissimilarity, Wilcoxon rank sum test with permutations, $p<0.001$; wild, $n=436$; lab, $n=146$; Fig. 1B), and wild mice had a significantly higher number of unique predicted functional pathways per individual on average than lab mice (permutational Wilcoxon rank sum test, $p<0.001$; wild, $n=436$; lab, $n=146$; Fig. 1C). Median within-population number of pathways was lower in all laboratory colonies than in wild populations, with the exception of SKG mice ($n=3$) that had intestinal inflammation and also the highest alpha diversity among lab colonies (Fig. 1A, Suppl. Fig. 3). However, there was no universal relationship between alpha diversity and number of functional pathways (Fig. 1A, Fig. 1C, Suppl. Fig. 3).

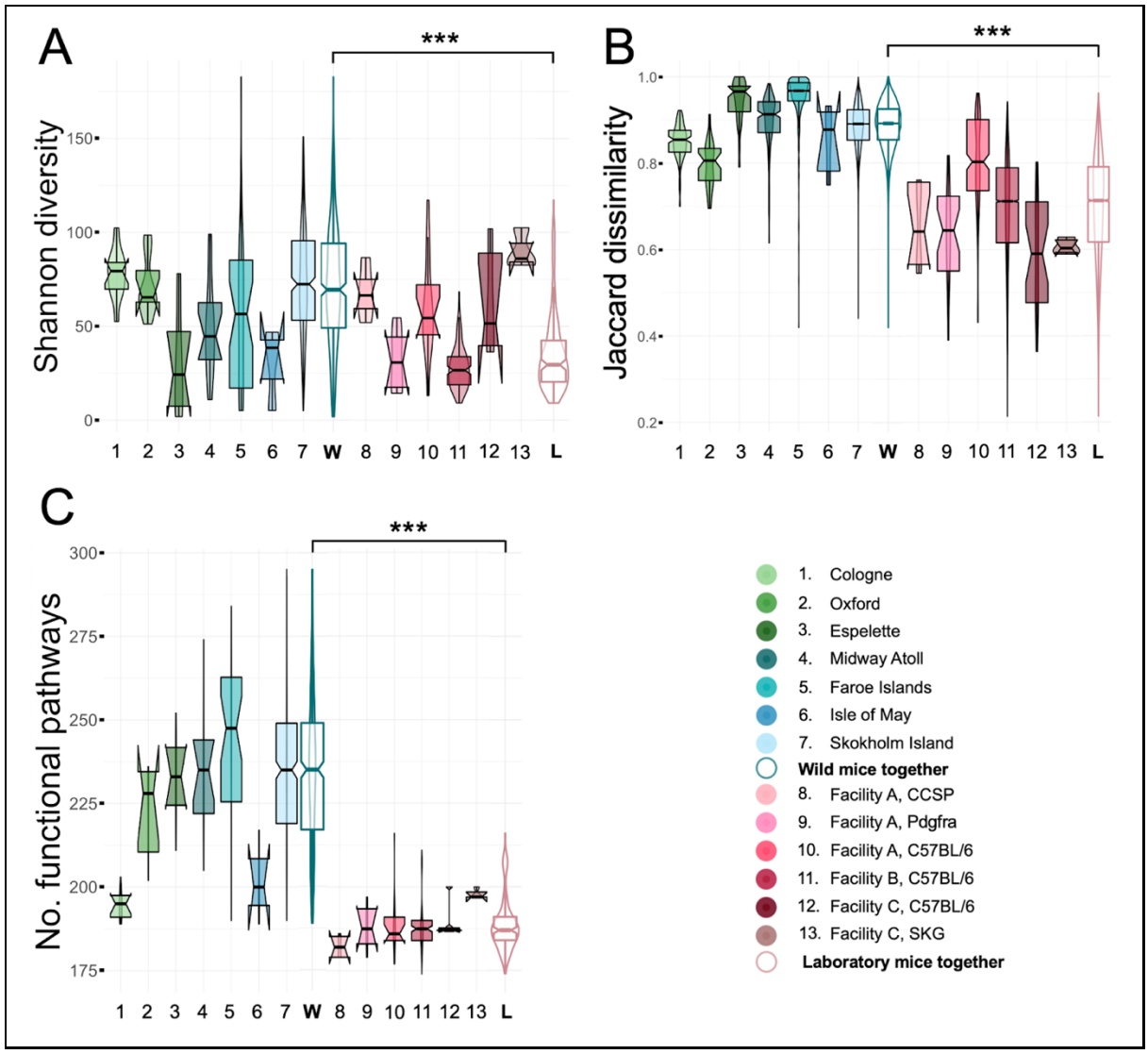


Figure 1. (**A**) Asymptotic Shannon diversity, (**B**) pairwise, within-population/colony Jaccard dissimilarity, and (**C**) number of unique functional pathways per mouse in wild (n=436) and laboratory (n=146) mice from seven populations and six colonies, respectively. Samples are from >3-month-old laboratory mice and wild mice estimated to be adults based on body size. Boxplots are individual wild mouse populations (*green* = mainland populations, *blue* = island populations) or laboratory mouse colonies (*pink*). Empty boxplots are all wild (*teal*) or laboratory (*pink*) mouse samples together. Statistical differences between lab and wild mice were tested with permutational Wilcoxon rank sum tests (***; p<0.001).

The gut microbiota of lab and wild mice was also compositionally distinct, with samples clustering by source (lab/wild) and largely separately from each other on both non-phylogenetic and phylogenetic distance metrics (Fig. 2A, Suppl. Fig. 4). Samples from Skokholm Island population clustered somewhat separately from other wild mouse samples on non-phylogenetic distances (Jaccard, Aitchison), but not on phylogenetic distances

(unweighted and weighted Unifrac; Fig. 2A, Suppl. Fig. 4). Despite the clustering patterns, source (lab/wild) explained only 1.5% of gut microbial variation (PERMANOVA on Jaccard distance, p=0.001; source explained just 1.1% of variation on Aitchison distance, p=0.001). Since wild mouse samples showed more variability in composition than lab mouse samples (dispersion test, F=59.337, p=0.001), this effect of source may be influenced by differences in both the mean and variability in composition.

Overall, wild mice had a higher ratio of Firmicutes to Bacteroidota (Fig. 2B). At the bacterial family level, the lab mouse microbiota was dominated by Muribaculaceae (mean relative abundance 46.7%, standard deviation 14.1%), while wild mice did not present such consistent dominance by a single family (Fig. 2C). In Random Forest regression models, the ten ASVs most important for distinguishing wild from lab mice all had either very low or zero relative abundance in wild mice and belonged to the bacterial families Muribaculaceae, Desulfovibrionaceae, and Erysipelothricaeae (Suppl. Fig. 5).

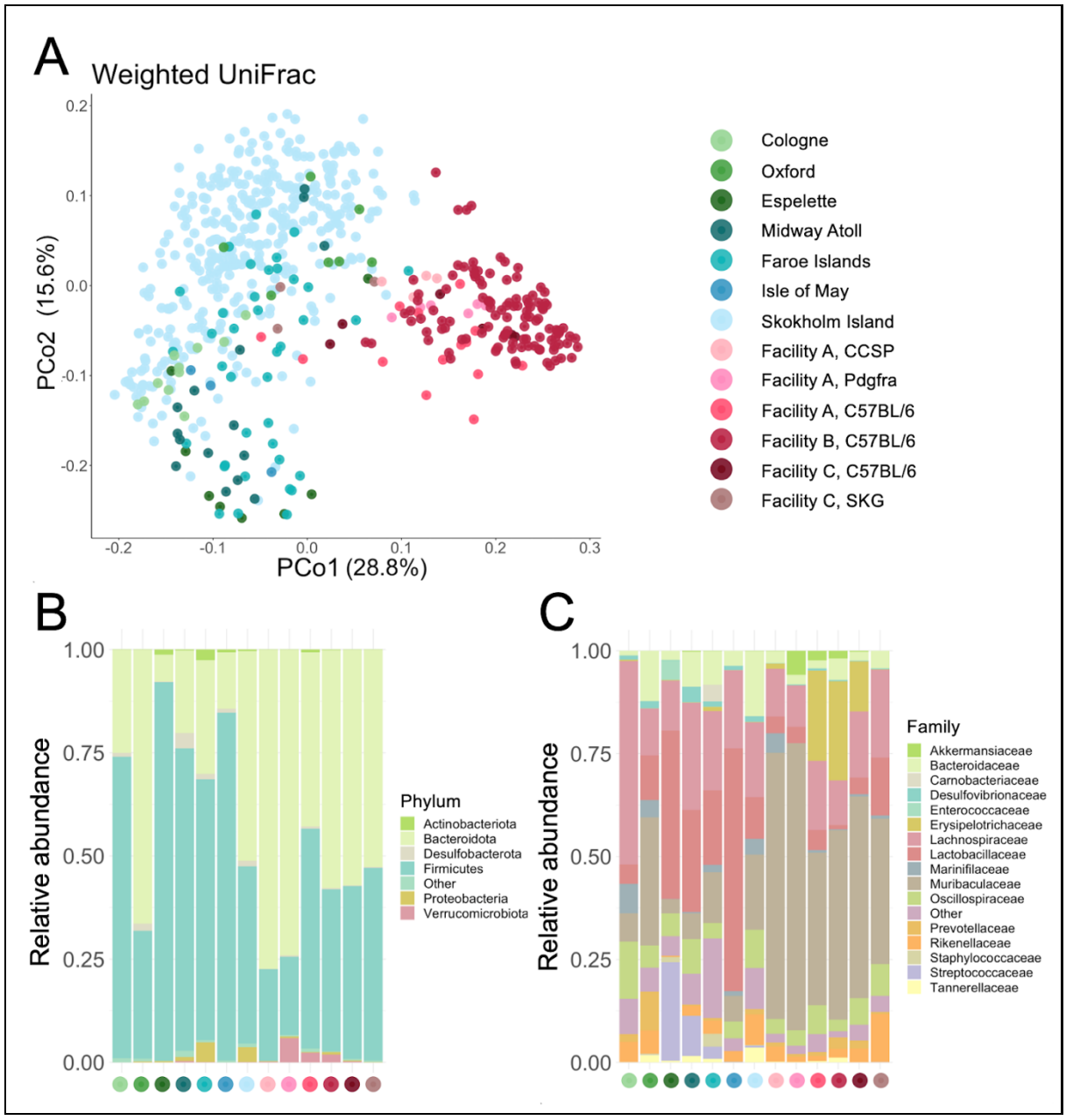


Figure 2. (**A**) Principal coordinate analysis of wild (*n*=436) and laboratory (*n*=146) mouse samples from seven populations and six colonies, respectively, on weighted UniFrac distance. Circles are individual samples coloured by population/colony (*green* = mainland wild mice, *blue* = island wild mice, *pink* = laboratory mice). (**B**–**C**) Mean relative abundance of bacterial (**B**) phyla and (**C**) families across seven wild populations and six laboratory colonies. Bars are individual populations/colonies, as indicated with coloured circles on x-axis.

Gut microbiota turnover rate is faster within wild than laboratory mice

We next hypothesised that wild mice are exposed to a higher and more variable pool of microbes than lab mice and consequently should have higher within-host turnover in gut microbial taxa. Consistent with this hypothesis, among mice repeat-sampled over short time-

spans (mice with >3 samples each with ≤40 days between sampling dates), gut microbiota dissimilarity increased with sampling interval among repeat-sampled wild mice but was fairly constant irrespective of sampling interval in lab mice (Fig. 3A–B; log-linear model on Aitchison distance; wild mice R^2=0.29, p<0.001; lab mice R^2=0.0094, p=0.268). Over longer sampling intervals, gut microbial turnover also increased with sampling interval in lab mice (log-linear model; R^2=0.34, p<0.001; all samples from repeat-sampled laboratory mice from C57BL/6 colony from Animal Facility B, n=99; Suppl. Fig. 6A). In wild mice, within-individual dissimilarity increased with sampling interval among sample pairs less than 20 days apart, but then reached a plateau (quadratic plateau model; R^2=0.44, critical point of inflexion=20.1 days; all samples from repeat-sampled Skokholm Island mice without exclusion based on season, age, or reproductive status, n=762; Suppl. Fig. 6B).

This observed faster short-term microbial turnover in wild mice can be illustrated by considering ASV changes during a fixed one-month period in both systems. Over one month, 25% of ASVs detected at the start were still detected at the end in lab mice, whereas this was true of only 12.5% of ASVs in wild mice (Fig. 3C–D). Similarly, around 85% of total relative abundance in the lab mouse microbiota at the end comprised ASVs detectable at the start, whereas for wild mice this was only 50% (Fig. 3E–F).

Lab mice harboured 35 ‘persistent’ ASVs (defined as ASVs detected at all timepoints in all included lab mice (n=4), regardless of abundance) that originated from nine bacterial families (Muribaculaceae, Rikenellaceae, Marinifilaceae, Oscillospiraceae, Bacteroidaceae, Akkermansiaceae, Lachnospiraceae, Sutterellaceae, and Erysipelotrichaceae). The combined relative abundance of these ASVs was 52.3–83.0% (mean 72.1%, median 75.7%) in these longitudinally studied mice, and similar abundances were seen in other samples from the

colony in which all 35 ASVs were detected (32 out of 123 samples; combined relative abundance 29.0–83.0%, mean 62.1%, median 63.9%). These persistent ASVs included 27 ASVs that were detected in >90% and six ASVs (all Muribaculaceae) that were detected in all samples from the colony.

In contrast, only 7 persistent ASVs were identified in wild mice (*n*=3). These ASVs were from families Lactobacillaceae (*n*=1), Oscillospiraceae (*n*=5) and Deferribacteraceae (*n*=1) and had a relative abundance of 2.3–23.6% (mean 8.9%, median 5.6%) in these three mice. Similar proportions were detected again when studying all samples from the population in which the 7 persistent ASVs were detected (*n*=575; 60.7% of all 948 samples); combined relative abundance ranged from 0.05 to 69.3% (mean 8.2%, median 6.0%). Across 15,733 ASVs detected across all of the Skokholm samples (*n*=948), only 7 ASVs were detected in >90% samples (at any abundance). Only two of these ASVs were not in the identified persistent ASVs (these two additional ASVs were from Oscillospiraceae and Lachnospiraceae families). These results suggest gut microbiota changes faster in the wild than in the lab, and that taxa that are found across individuals are also those that persist within individuals over time.

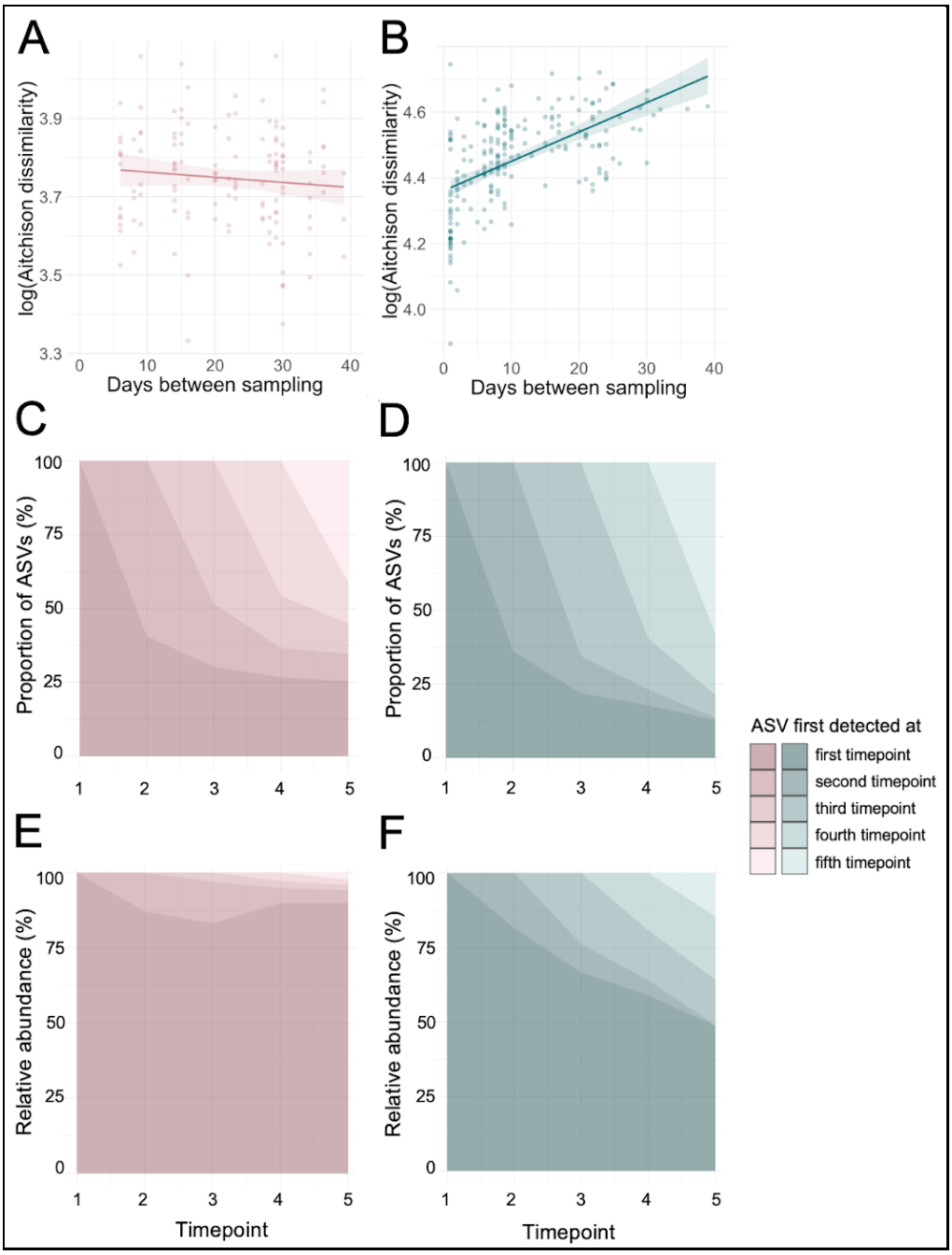


Figure 3. Gut microbiota turnover in adult laboratory (*left*) and wild (*right*) mice. (**A**–**B**) Within-individual microbiota dissimilarity on Aitchison distance (log-transformed) between sample pairs from the same (**A**) laboratory or (**B**) wild mouse (98 samples from 16 laboratory mice, 5–7 samples per mouse; 72 samples from 14 wild mice, 4–10 samples per mouse) against time between samples. Laboratory mice were from a single C57BL/6 colony (Animal facility B, Supplementary Table 1). The relationship is fitted with a log-linear model. Lab mice: F_{131}=1.239, adjusted R^2=0.001806, p=0.2677; wild mice: F_{164}=*65.12,* adjusted R^2=0.2799, p<0.001. (**C**–**D**) Mean proportion and (**E–F**) mean relative abundance of amplicon sequence variants (ASVs) at five timepoints in (**C**, **E**) laboratory (n=4) and (**D**, **F**) wild (n=4 for timepoints 1–4, n=3 for timepoint 5) mice based on timepoint the ASV was first detected. Laboratory and wild mice for which similar sampling intervals (4–5 samples ~1 week apart) were available were selected for the analysis. Timepoints are days 0, 9, 15, 23, and 29 for laboratory mice and days 0, 6–8, 12–16, 22–23, and 28–32 for wild mice.

Mainland and island mice harbour similar gut microbiotas

We next investigated whether among wild mice there is a difference in the gut microbiota of mainland versus island mice, and if so, whether Skokholm mice are similar to other island mice in terms of their gut microbiota. House mice are one of the most successful invasive species and have colonised many islands (Cuthbert & Hilton, 2004; Berry & Tricker, 1969; Kappes et al., 2022; Berry, 1968) where they are often of larger body size (Phifer-Rixey and Nachman, 2015). This provides an opportunity to examine whether there have been consistent shifts in gut microbiota composition after island colonisation. We hypothesised that mainland and island mice might present different gut microbiota characteristics that may arise from (1) reduced dispersal of gut microbes to island mice living in ecosystems that are more isolated and closed compared to mainland mice, and (2) differences in diet, due to island mice possibly having access to seabird carcasses and remains of marine vertebrates, which should be rarer for mainland mouse populations.

Setting (mainland/island) showed some compositional distinction, but this was subtle with setting explaining only 0.7% of gut microbial variation on Aitchison distance (PERMANOVA, p=0.002; setting explained 0.5% of variation on Jaccard distance; PERMANOVA, p=0.010). Compositional distinction by setting was not visible on principal coordinates analyses (PCoAs) irrespective of the distance metric used (Jaccard, Aitchison, unweighted and weighted UniFrac; Fig. 4A–B, Suppl. Fig. 7). Some of the results on Jaccard distance may have been influenced by beta dispersion which varied significantly between island and mainland mice (F=23.533, p=0.001; beta dispersion did not vary significantly on Aitchison distance, F=0.5186, p=0.504).

Random Forest regression models were used to identify ASVs driving the distinction between mainland and island mice. The top ten ASVs all had a higher relative abundance in mainland

than island mice and belonged to either the order Bacteroidales or Lachnospirales (Suppl. Fig. 8). The third most important taxon was *Muribaculum intestinale*, which was also an important driver of the lab/wild distinction, with higher relative abundance in lab than wild mice (Suppl. Fig. 5). Overall, the gut microbiota of island mice was more variable than that of mainland mice (Jaccard distance; permutational Wilcoxon rank sum test, $p<0.001$), although compositional variation in the microbiota among individuals varied among populations of both mainland and island mice (Fig. 4C). Particularly, mainland population from Espelette (~15 km from coastline) had significantly higher mean variation than the other two mainland populations (Cologne, Oxford; ~200 and ~70 km from coastline, respectively). Overall, we detected gut microbiota differences between mainland and island mice, but these differences appeared subtle.

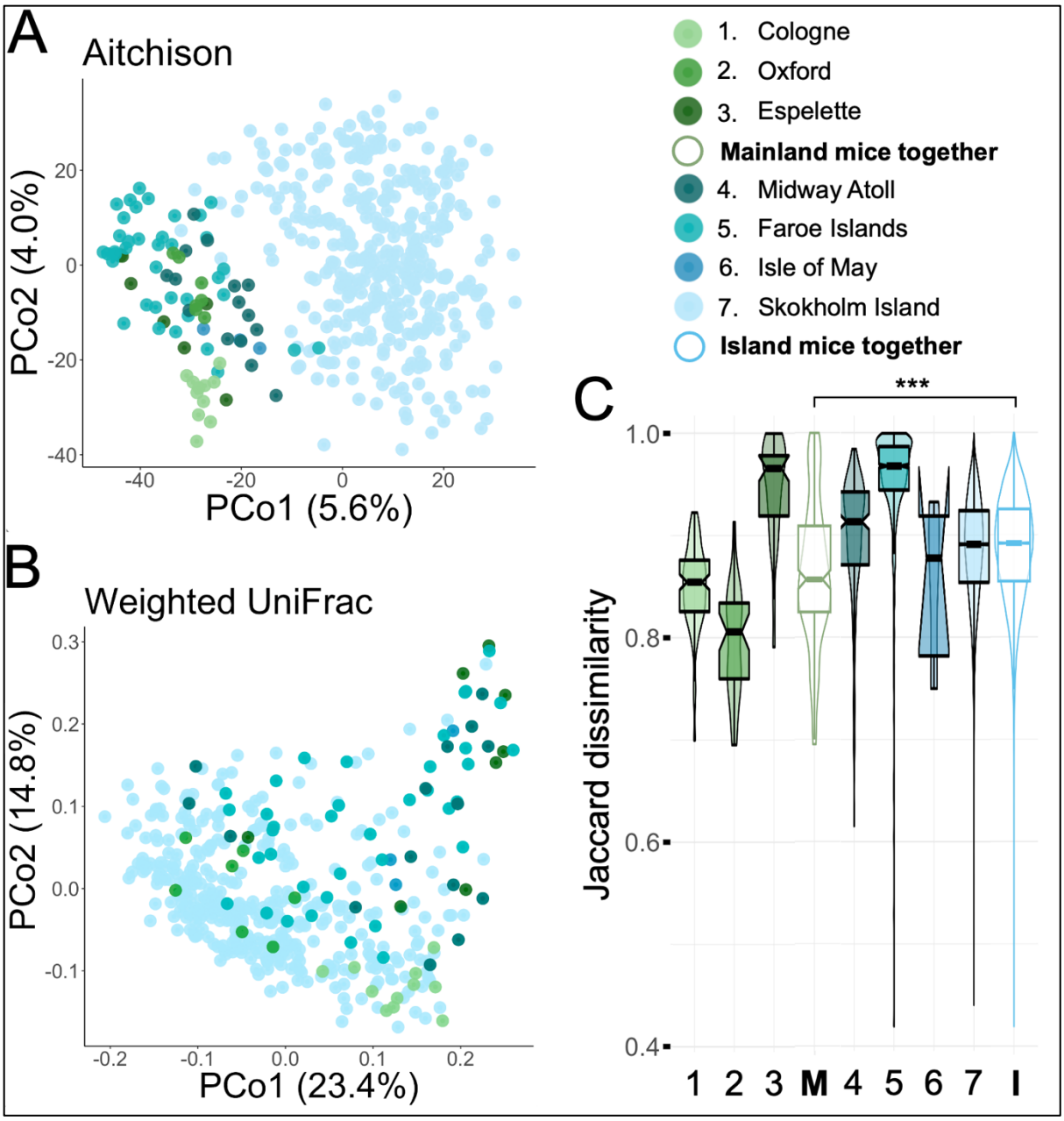


Figure 4. (**A–B**) Principal coordinate analysis of mainland (n=26) and island (n=410) mouse samples from three mainland and four island populations on (**A**) Aitchison and (**B**) weighted UniFrac distances. (**C**) Pairwise Jaccard dissimilarity within mainland and island populations. Empty boxplots are mainland (*green*) or island (*blue*) mouse samples together. Statistical differences between mainland and island mice were tested with permutational Wilcoxon rank sum tests (***; p<0.001).

Factors predicting gut microbiota composition within an island mouse population

Finally, we investigated gut microbiota variation within a population by focusing on the Skokholm Island population. When limiting analysis to spatial and temporal predictors, sampling time (at the resolution of month) was the strongest and only significant biological predictor of gut microbial composition, explaining 1.2% of variation on Aitchison distance (Table 1A). When including individual-level variables, additional significant predictors were sampling site (Fig. 5H) and body mass, although both explained less than 1% of variation. Although age category did not predict microbiota composition when body mass was in the model, a significant age effect was detected when body mass was removed from the model (Age category, R^2=0.013, F=0.8968, p=0.002), as these two terms are somewhat collinear. All significant predictors had significant beta dispersion which may have influenced the detected effects. Among adult mice only, the same significant predictors of gut microbiota composition were detected, with a weak but significant effect of body mass still present (Table 1B). Here, sampling site was identified as a significant predictor both when excluding and including individual-level variables.

A

Variable	Spatial and temporal variables only (*n*=340)				Spatial, temporal, and host variables (*n*=133)			
	degrees of freedom	partial R^2	F statistic	p-value	degrees of freedom	partial R^2	F statistic	p-value
Read count	1	0.00088	0.3095	**0.043**	1	0.00044	0.0624	0.370
MiSeq run	13	0.03951	1.0712	**0.001**	12	0.08337	0.9854	**0.030**
Month, year	4	0.01243	1.0948	**0.001**	2	0.01781	1.2627	**0.001**
Sampling site	1	0.00066	0.2327	0.065	1	0.00223	0.3156	**0.038**
Distance to buildings	1	0.00000	0.0000	1.000	1	0.00000	0.00000	1.000
Distance to sealine	1	0.00063	0.2238	0.080	1	0.00107	0.1518	0.196
Age					2	0.00819	0.5808	0.195
Body mass					1	0.00273	0.3875	**0.024**
Body condition					1	-0.00047	-0.0672	0.664
Age:Body mass					2	0.00752	0.5350	0.339
Age:Body condition					2	0.00648	0.4608	0.640
Sex					1	0.00075	0.1071	0.273
Reproductive activity					1	-0.00139	-0.1977	0.876
Sex:Reproductive activity					1	0.00066	0.0934	0.310

B

Variable	Spatial and temporal variables only (*n*=244)				Spatial, temporal, and host variables (*n*=101)			
	degrees of freedom	partial R^2	F statistic	p-value	degrees of freedom	partial R^2	F statistic	p-value
Read count	1	0.00170	0.4328	**0.006**	1	0.00023	0.0239	0.495
MiSeq run	13	0.05291	1.0343	**0.001**	12	0.10381	0.9174	0.383
Month, year	6	0.03631	1.5381	**0.001**	2	0.02096	1.1114	**0.001**
Sampling site	1	0.00112	0.2849	**0.046**	1	0.00418	0.4429	**0.008**
Distance to buildings	1	0.00000	0.0000	1.000	1	0.00000	0.00000	1.000
Distance to sealine	1	0.00072	0.1842	0.130	1	0.00122	0.1293	0.215
Body mass					1	0.00340	0.3607	**0.034**
Body condition					1	0.00081	0.0862	0.324
Sex					1	0.00252	0.2667	0.071
Reproductive activity					1	-0.00060	-0.0637	0.693
Sex:Reproductive activity					1	0.00105	0.1117	0.271

Table 1. Spatial, temporal and host-level predictors of gut microbial composition in an island mouse population (**A**) across mice of all ages and (**B**) only in adult mice. The power of spatial and temporal variables in predicting gut microbial composition were tested without (*left*) and with (*right*) host variables using marginal permutational multivariate analyses (PERMANOVAs) on Aitchison distance. For each model, one sample was selected randomly per individual (**A**: without host variables; *n*=340, with host variables; *n*=133; **B**: without host variables, *n*=244; with host variables, *n*=101). Significant *p*-values (<0.05) are in bold. Beta dispersion was tested for significant terms (factors only) and was significant for all (*p*=0.001 for 'month, year' in all models; *p*=0.012 for sampling site in A with host variables model; *p*=0.004 and *p*=0.001 for sampling site in B without and with host variables models, respectively). Interaction terms were fitted in separate models, all other results are from models without interactions terms.

Geographic distance predicts gut microbiota dissimilarity with stronger effects on fungal rather than bacterial component

Mice from two geographically distinct but closely located sampling sites ('Observatory' and 'Quarry'; Fig. 5H) had differences in both bacterial and fungal microbiota composition (only previously identified symbiotic fungi were included, see Methods; Fig. 5A, 5D). Mice from Observatory had a higher relative abundance of Lachnospiraceae and Muribaculaceae bacteria as well as Didymellaceae and Sclerotiniaceae fungi. Meanwhile, the relative abundances of Enterobacteriaceae bacteria, Bulleribasidiaceae and Pleosporaceae fungi were higher in mice from Quarry.

Samples clustered by sampling site more strongly for fungal than bacterial components of the microbiota (Fig. 5B, Fig. 5E, Suppl. Fig. 9), with sampling site explaining 8.3% of fungal but only 4.1% of bacterial community variation (PERMANOVA on Jaccard distance, $p<0.001$ for both, $n=38$), indicating mice from the two sites varied more in their fungal rather than bacterial microbiota. However, both bacterial and fungal dissimilarities were significantly higher between mice from different sampling sites than between mice from the same sampling site (Jaccard dissimilarity, permutational Wilcoxon rank sum test, $p<0.001$ for both types of symbionts; Fig. 5C, 5F), although fungal dissimilarity was higher than bacterial dissimilarity (Fig. 5C, 5F).

At a finer spatial resolution, geographic distance between exact trap locations of mice had a stronger correlation with fungal than with bacterial dissimilarity (Mantel test on geographical distance and Jaccard dissimilarity; fungi, $r=0.32$, $p<0.001$; bacteria, $r=0.14$, $p=0.005$). Consistent with this, the microbiota component (bacteria/fungi) had a strong effect on the predictive power of space on gut microbial dissimilarity, regardless of spatial resolution (brm

model; *sampling site*: bacteria, posterior mean 0.15, 95% credible intervals (CI) 0.05–0.26; fungi, posterior mean 0.98, 95% CI 0.79–1.16; *geographic proximity*: bacteria, posterior mean 0.28, 95% CI 0.12–0.44, fungi, posterior mean: 0.98, CI 0.79–1.16, Fig. 5G).

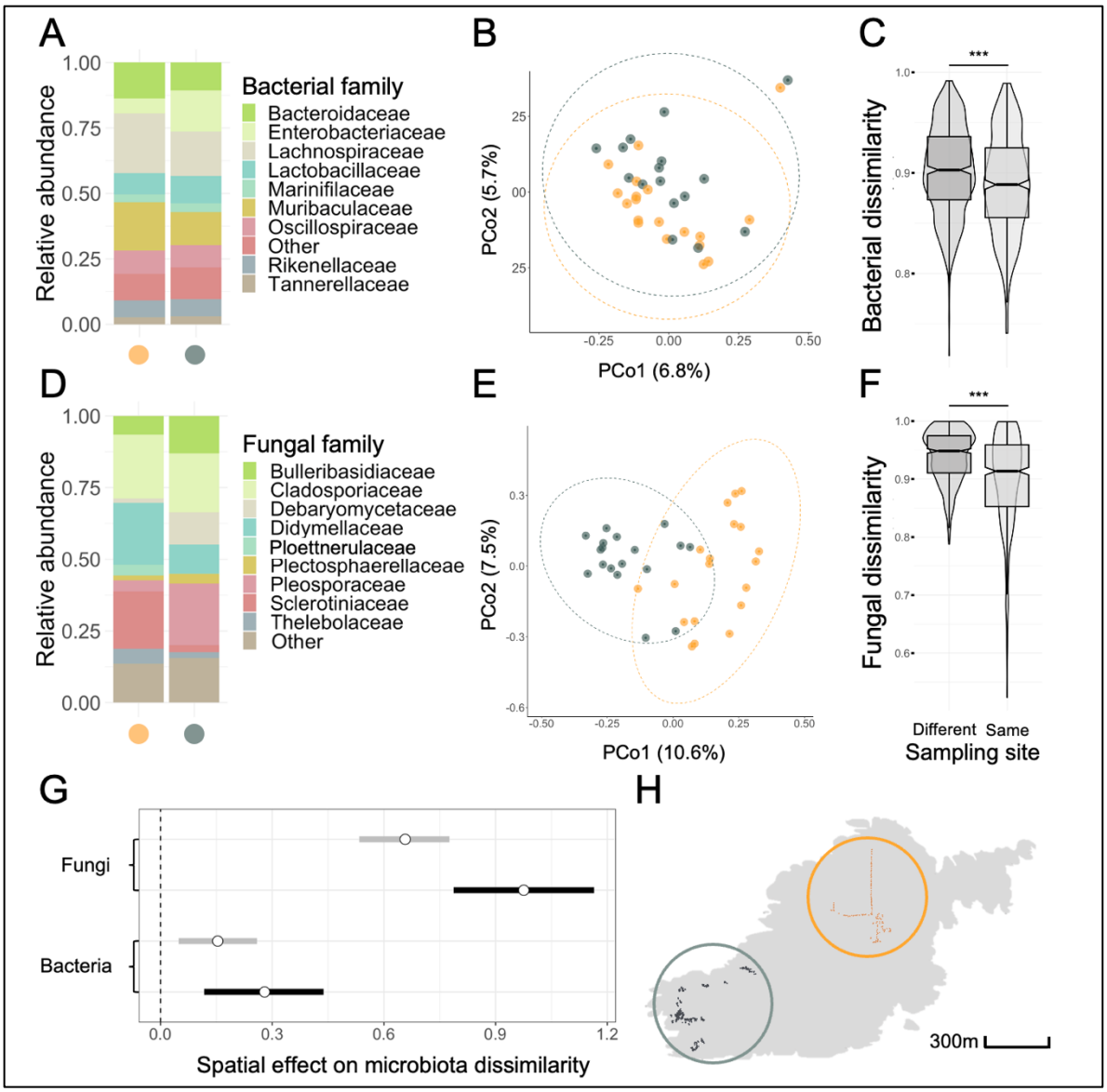


Figure 5. (**A–C**) Bacterial and (**D–F**) fungal variation in gut microbiota of mice (n=37) from two geographically distinct sampling sites (*yellow* = Observatory, *blue* = Quarry). (**A**, **D**) Relative abundance of (**A**) bacterial and (**B**) fungal families. (**B**, **E**) Principal coordinates analysis (PCoA) on Jaccard distance based on (**B**) bacteria and (**E**) fungi. (**C**, **F**) Pairwise Jaccard dissimilarity of (**C**) bacterial and (**F**) fungal microbiota in sample pairs from same or different sampling sites. Differences in dissimilarity between groups (same vs different sampling site) were tested with permutational Wilcoxon rank sum tests (***; p<0.001). (**G**) Estimated spatial effects on Jaccard dissimilarity (slope of the relationship between sampling site similarity (same vs different) or geographic proximity and Jaccard index) and 95% credible intervals from a Bayesian regression (brm) model that included an interaction term between microbiota component (fungi/bacteria) and sampling site similarity (same vs different) *or* geographic distance. Colour indicates the model (measure of spatial proximity used): *grey* = sampling site, *black* = geographic distance. Both measures of spatial proximity have a significant association with microbial dissimilarity such that mice from different sites or with a larger geographic proximity have more dissimilar microbiotas than mice captured closer to each other (same site or shorter distance). (**H**) Sampling sites on Skokholm Island, Wales (*yellow* = Observatory, *blue* = Quarry). Dots are individual sampling locations within sites (n=150 at each site).

Microbial similarity between mice and soil increases with geographic proximity

Considering the compositional differences between mice from the two sampling sites, we investigated whether soil microbiota may influence mouse gut microbiota composition. For this, we analysed 86 soil and 41 mouse (bacterial) microbiota samples collected within a single month (September 2020) and limited ASVs used in analyses to those detected in both soil and mouse samples (at any prevalence and abundance, n=185). These ASVs found in both soil and mouse faecal material made up 1.9–79.7% (mean 13.6%) of reads in mouse gut microbiota profiles, and 4.1–32.5% (mean 16.5%) of soil microbiota relative abundance. However, prevalence of these ASVs was low across samples, such that no ASV was found in >50% of mice and >10% of soil samples (regardless of abundance).

Still, microbiota dissimilarity was significantly higher in mouse-soil sample pairs that originated from different than same sampling sites using both Jaccard and Aitchison distances (permutational Wilcoxon rank sum tests; p<0.001, Fig. 6B). A similar pattern was detected in mouse-mouse and soil-soil sample pairs, when using the same sample and taxa filtering criteria as for mouse-soil sample pairs, i.e., only including taxa detected in both mouse and soil samples (permutational Wilcoxon rank sum tests; p<0.001 for both; Fig. 6A, 6C). Further, the proportion of shared ASVs between mouse-soil sample pairs increased with the geographic proximity of their sampling locations across the island (Mantel test on Jaccard distance, r=0.081, p=0.005). These findings suggest gut microbiota may be influenced by environmental transmission of microbes from soil, but that overall the gut microbial communities in wild mouse faeces are qualitatively and compositionally very distinct from those found in soil, as expected.

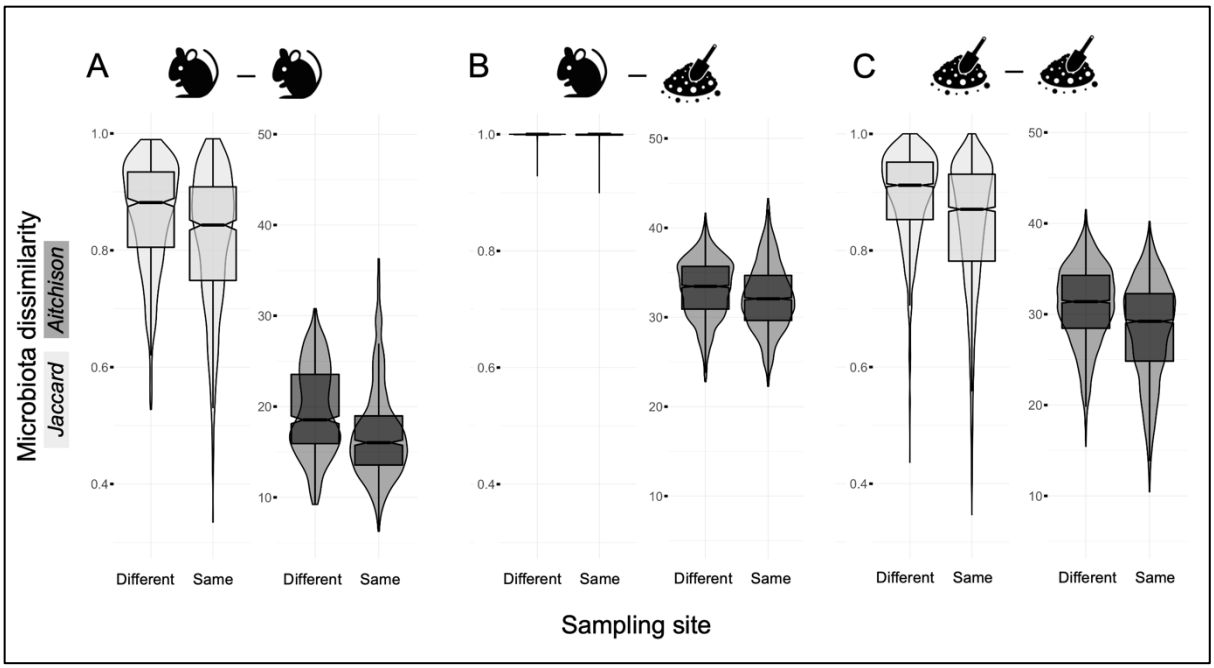


Figure 6. Jaccard (*grey*) and Aitchison (*black*) dissimilarity in bacterial community (**A**) mouse-mouse, (**B**) mouse-soil, and (**C**) soil-soil sample pairs from the same or different sampling sites (Quarry or Observatory, Fig. 5C). Samples (soil; n=86, mouse; n=41) were collected within a single month from 25 locations across each site. Dissimilarity was measured using amplicon sequence variants (ASVs) that were detected in both mice and soil (at any prevalence and abundance, excluding singletons and doubletons; n=185). Microbiota dissimilarity was significantly higher in sample pairs from different than from the same sampling site within each sample pair type and on both measures (Jaccard, Aitchison; permutational Wilcoxon rank sum tests, p<0.001 for all different vs same sampling site comparisons).

Discussion

We investigated drivers of gut microbiota variation in a wild house mouse (*Mus musculus domesticus*) population on the Skokholm Island, Wales, using a large dataset of nearly 1,000 samples from over 300 individual mice. We studied gut microbiota variation both cross-sectionally as well as longitudinally, and explored both the bacterial and fungal components of the microbiota. We further investigated similarities between soil and gut microbiota, which may reflect transmission of microbes between these two habitats. To put our study population into a wider ecological context, we first compared the gut microbiota of Skokholm mice to that of laboratory mice from multiple facilities and strains, as well as other wild mice from both mainland and island populations.

In line with previous studies (Rosshart et al., 2019; Wang et al., 2014; Kreisinger et al., 2014; Wang et al., 2015), wild mice had a taxonomically and functionally more diverse gut microbiota than lab mice, although alpha diversity in particular varied greatly between laboratory colonies and wild mouse populations. This included variation between laboratory strains from the same animal facility as well as variation within a single laboratory strain (C57BL/6) from different animal facilities, similar to previous findings (Bowerman et al. 2021), further highlighting the importance of controlling for such covariates to ensure experimental reproducibility (Ericsson et al., 2021). The number of unique functional pathways was higher in wild compared to lab mice, independent of alpha diversity of taxa. We hypothesise this may arise from wild mice likely having a more diverse diet, which may favour microbial taxa with a wider range of functional properties that can effectively utilise different substrates (Frese et al., 2015). Further, exposure of wild mice to a wider range of environmental conditions, such as temperature, water availability or salinity at varying levels may also have contributed to higher number of functional pathways (Chevalier et al., 2015; Sepulveda & Moeller, 2020).

As in other mammals, the predominant phyla in both wild and laboratory mice were Firmicutes and Bacteroidota, a ratio of which is often used as a key feature of the gut microbiota (Ley et al., 2005). Wild mice had a higher ratio of Firmicutes to Bacteroidota, contradicting findings from Rosshart et al. (2017) where the relative abundance of Firmicutes was higher in lab compared to wild house mice. One possible contributing factor to this difference is that our wild mouse samples come from higher latitudes that those used in Rosshart et al. (2017), and previous work in humans has shown a positive relationship between the Firmicutes:Bacteroidota ratio and latitude in the human gut microbiota (Suzuki & Worobey, 2014). However, the relative abundance of Firmicutes in the Midway Atoll population was

comparable to our other populations from higher latitudes, proving other factors besides latitude must also play a role in shaping the relative abundance of these two dominant phyla. Wild mice had a higher abundance of Proteobacteria than lab mice (in line with Rosshart et al., 2017), a phylum containing various common pathogens, such as *Salmonella* and *Escherichia,* often excluded in the commonly used specific-pathogen-free (SPF) laboratory mouse facilities, which may partly explain the lower relative abundance of Proteobacteria in lab mice.

Contrary to observations in Thomson et al. (2022), we detected Akkermansiaceae, Streptococcaceae, and Enterobacteriaceae in wild mice as well as lab mice. Streptococcaceae and Enterobacteriaceae occurred at high prevalence across wild mouse samples (~76% and ~59%, respectively). A closer investigation of population differences revealed that these families had low relative abundance in most wild populations (0.004–3.0%), but were more common in others (16.7% and 10.7% in Espelette and Midway Atoll populations, respectively). These findings illustrate the vast cross-populational variation seen in wild house mouse gut microbiota and highlight the advantage of including multiple populations from a range of locations for more comprehensive understanding of the wild mouse gut microbiota (Weldon et al., 2015; Suzuki et al., 2019).

At the family level, the lab mouse gut microbiota was dominated by Muribaculaceae (mean relative abundance 46.7%), which is considered an important family in the mouse gut (Lagkouvardos et al., 2016; Lagkouvardos et al., 2019; Ormerod et al., 2016) but was present at 17% relative abundance on average in our wild mouse samples. Consistent with this, muribaculaceaen taxa were key in driving microbiota differences between lab and wild mice, in line findings from Bowerman et al. (2021). One such taxon was *Muribaculum intestinale*, which was omnipresent in lab mice but very rare in wild mice (detected in 6% of all wild mouse

samples with <1% relative abundance). Identification of taxa conserved across house mice across laboratories and wild populations could provide indication of taxa that may provide important, conserved functions for this host species, or functions that may have been lost with domestication.

As could be expected from exposure to a larger and more variable pool of microbes and environmental conditions, the gut microbiota showed more inter-individual variation among wild mouse populations compared to lab mouse colonies, and wild mice showed a much faster within-host turnover of gut microbial taxa, with far fewer ASVs stably persisting over an approximate 40-day period. We found that taxa that persist within individuals over time are also those that are prevalent across individuals, and that this pool of 'core' microbes was larger in lab than in wild mice, further reflecting limited and more stable microbial exposure patterns in the lab. Temporal dynamics were investigated within a single wild mouse population and a single lab mouse colony, and further work is required to investigate whether the observed dynamics are representative of wild and lab mice in general.

We next investigated whether wild mice from mainland and island populations harbour unique gut microbiota characteristics that could arise from diet differences on islands, or the closed nature of island ecosystems where influx of new microbes may be more limited than in mainland ecosystems. Overall, wild mice across mainland and island populations harboured similar gut microbiotas both in terms of composition and within-population dissimilarity. Remarkably, samples from Midway Atoll (North Pacific Ocean) where house mice were introduced over 75 years ago (Duhr et al., 2019), did not cluster separately from other populations despite their extreme geographic isolation (>10,000 km from all other sampled populations). However, we did detect subtle differences between mainland and island

microbiotas, with setting (mainland/island) explaining a small proportion (~0.5–1%) of gut microbiota variation, and certain ASVs distinguishing mainland mice from island mice.

In particular, *Muribaculum intestinale*, a typical member of the laboratory gut microbiota (Lagkouvardos et al., 2016; Lagkouvardos et al., 2019), was a key taxon driving this distinction, with a higher relative abundance in mainland mice (although mean relative abundance was very low (<1%) in both mainland and island mice). The same bacterium distinguished lab mice from wild mice, with lab mice having a higher relative abundance. As such, the differences in the relative abundance of *M. intestinale* could reflect more limited anthropogenic interference of island rather than mainland mice included in this study. Overall, gut microbiota differences between mainland and island mice were subtle; however, even small taxonomic differences may be functionally important. Further work would be required to test whether the island syndrome, where species (including house mice) on islands are predictably different from their mainland counterparts (Gray et al., 2015), applies to or is affected by the gut microbiota.

Finally, we investigated gut microbiota variation within an extensively sampled population from Skokholm Island, Wales. In our previous analyses, we demonstrated that this population was similar to other wild populations, thus confirming it to be relatively representative of wild mice. In line with a previous wild house mouse study (Goertz et al., 2019), we detected fine-scale spatial effects on gut microbiota composition, although these were relatively weak and intriguingly, spatial effects were stronger for the fungal rather than the bacterial microbiota. As fungal DNA in faeces may originate from ingested macrofungi (Lavrinienko et al., 2021), we limited the analysis to fungi that had been previously identified as symbionts in human or laboratory mouse gut (see Methods). As such, it is more likely that with this filtering approach

we excessively removed fungal symbionts (as wild house mice may have fungal symbionts that have not been previously detected in humans or laboratory mice) rather than retained non-symbiotic macrofungi, and that the results reflect spatial differences in symbiotic fungi rather than diet. Future work is needed to assess if the increased spatial variation in fungal communities is specific to the Skokholm population or general pattern seen in other wild mouse populations as well.

To investigate whether exposure to environmental microbes might explain spatial variation in gut microbiota, we contrasted mouse gut and soil (bacterial) microbiotas. Overall mouse and soil microbiotas were drastically different and very few ASVs were detected in both habitats, similar to previous comparisons of host-associated and soil microbiotas (Li et al., 2016; Zhou et al., 2016; Brown et al., 2022). Still, spatial proximity increased microbial similarity between mouse and soil samples, suggesting environmental transmission of microbes from soil to mouse gut may contribute to within-population spatial variation of the gut microbiota. More investigation is required to determine the direction(s) of possible transmission, as well as to exclude the possibility that soil-associated microbes were detected in faecal samples through environmental contamination, as faecal samples were collected from live traps where mice had been overnight thus any soil on e.g., paws of mice may have mixed with faecal samples.

Alongside spatial patterns, we did detect temporal changes with month predicting approximately 2% of gut microbiota variation. However, this dataset did not include sufficient resolution to test for repeatable seasonal shifts in the gut microbiota, as reported for other wild mammal species (Smits et al., 2017; Maurice et al., 2015; Marsh et al., 2022; Risely et al., 2021). Similar to previous studies, we detected limited influences of individual-level variables such as sex and reproductive status (Weldon et al., 2015; Linnenbrink et al., 2013; Goertz et

al., 2019), though we did detect weak but significant effects of body mass and age on microbiota composition, that could benefit from more detailed investigation.

Overall, we find clear differences between the gut microbiota of wild house mice and laboratory mice with differences seen in taxonomic composition, diversity, functionality as well as stability. An analysis of a large dataset from a representative wild mouse population demonstrates spatial effects on gut microbiota that are stronger on the fungal rather than bacterial component. Microbial transmission from soil was identified as a possible source of spatial influence on the gut microbiota. Still, the vast majority of variation continued to remain unexplained, highlighting the need for further research on gut microbiota variation in wild house mice.

Funding

This work was financially supported by The Osk. Huttunen Foundation studentship and the National Geographic Society (Early Career grant reference No. EC-58520R-19) to EH, the European Research Council under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851550) to SK, and the UKRI UKRI Medical Research Council (award MR/P012175/2) to SJ and MAC.

Acknowledgements

We thank Giselle Eagle and Richard Brown (wardens of Skokholm Island), the Friends of Skokholm and Skomer, the Wildlife Trust of South and West Wales and field assistants for their help in enabling the Skokholm wild mouse data collection.

Author information

Affiliations

[1]Department of Biology, University of Oxford, UK

[2]Centre for Host-Microbiome Interactions, King's College London, UK

[3]University of Tennessee, USA

[4]U.S. Fish and Wildlife Service, Midway Atoll NWR, USA

[5]Í Geilini, Nólsoy, Faroe Islands

[6]Department of Computing, University of Turku, Finland

[7]Kennedy Institute, University of Oxford, UK

[8]Max Planck Institute for Evolutionary Biology, Plön, Germany

[9]Department of Functional and Evolutionary Entomology, University of Liège, Belgium

[10]Lydia Becker Institute of Immunology and Inflammation, University of Manchester, UK

[11]School of Life Sciences, University of Nottingham, Nottingham, UK

Contributions

EH and SCLK set up the Skokholm mouse study system. EH, AR and SCLK collected the Skokholm mouse data, and WH, JHP, JJ, JFB, MQ, KJE and JEB collected wild mouse samples from other populations. EH, SJ, MAC, and JS contributed to the lab mouse samples. EH collected the soil samples. EH conducted the laboratory work, analysed the data, and wrote the manuscript. All authors contributed to the final manuscript.

References

Amato KR, Leigh SR, Kent A, et al. The gut microbiota appears to compensate for seasonal diet variation in the wild black howler monkey (Alouatta pigra). *Microb Ecol*. 2015;69(2):434-443. doi:10.1007/s00248-014-0554-7

Anders JL, Moustafa MAM, Mohamed WMA, Hayakawa T, Nakao R, Koizumi I. Comparing the gut microbiome along the gastrointestinal tract of three sympatric species of wild rodents. *Sci Rep*. 2021;11(1):19929. Published 2021 Oct 7. doi:10.1038/s41598-021-99379-6

Berry RJ, Tricker BJK. Competition and extinction: the mice of Foula, with notes on those of Fair Isle and St Kilda. *J Zool.* 1969. doi:10.1111/j.1469-7998.1969.tb02145.x

Berry RJ. The ecology of an island population of the house mouse. *J Anim Ecol* 1968; 37: 445–470.

Bowerman, K.L., Knowles, S.C.L., Bradley, J.E. *et al.* Effects of laboratory domestication on the rodent gut microbiome. *ISME COMMUN.* 1, 49 (2021). https://doi.org/10.1038/s43705-021-00053-9

Brown MD, Shinn LM, Reeser G, et al. Fecal and soil microbiota composition of gardening and non-gardening families. *Sci Rep*. 2022;12(1):1595. Published 2022 Jan 31. doi:10.1038/s41598-022-05387-5

Bürkner P (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software*, 80(1), 1–28. doi:10.18637/jss.v080.i01.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature Methods*, 13, 581-583. doi: 10.1038/nmeth.3869.

Chang CS, Liao YC, Huang CT, et al. Identification of a gut microbiota member that ameliorates DSS-induced colitis in intestinal barrier enhanced Dusp6-deficient mice. *Cell Rep*. 2021;37(8):110016. doi:10.1016/j.celrep.2021.110016

Chao A, Gotelli NJ, Hsieh TC, Sande EL, Ma KH, Colwell RK, Ellison AM (2014). "Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies." *Ecological Monographs*, 84, 45–67.

Chevalier C, Stojanović O, Colin DJ, et al. Gut Microbiota Orchestrates Energy Homeostasis during Cold. *Cell*. 2015;163(6):1360-1374

Chung H, Pamp SJ, Hill JA, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*. 2012;149(7):1578-1593. doi:10.1016/j.cell.2012.04.037

Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems*. 2017;2(1):e00127-16. Published 2017 Jan 3. doi:10.1128/mSystems.00127-16

Cuthbert R, Hilton G. Introduced house mice Mus musculus: a significant predator of threatened and endemic birds on Gough Island, South Atlantic Ocean? *Biol Conservation.* 2004;117:483-489. doi:10.1016/j.biocon.2003.08.007

Depner M, Taft DH, Kirjavainen PV, et al. Maturation of the gut microbiota during the first year of life contributes to the protective farm effect on childhood asthma. *Nat Med*. 2020;26(11):1766-1775. doi:10.1038/s41591-020-1095-x

Douglas GM, Maffei VJ, Zaneveld JR, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685-688. doi:10.1038/s41587-020-0548-6

Douglas GM, Maffei VJ, Zaneveld JR, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685-688. doi:10.1038/s41587-020-0548-6

Duhr M, Flint EN, Hunter SA, et al. Control of house mice preying on adult albatrosses at Midway Atoll National Wildlife Refuge. 2019. Available here:

https://www.sprep.org/attachments/Publications/articles/control-house-mice-preying-adult-albatrosses-midway-atoll-national-wildlife-refuge.pdf Accessed on 4.12.2022.

Ericsson AC, Franklin CL. The gut microbiome of laboratory mice: considerations and best practices for translational research. *Mamm Genome*. 2021;32(4):239-250. doi:10.1007/s00335-021-09863-7

Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol*. 2021;19(1):55-71. doi:10.1038/s41579-020-0433-9

Frese SA, Parker K, Calvert CC, Mills DA. Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome*. 2015;3:28. Published 2015 Jul 1. doi:10.1186/s40168-015-0091-8

Goertz S, de Menezes AB, Birtles RJ, et al. Geographical location influences the composition of the gut microbiota in wild house mice (Mus musculus domesticus) at a fine spatial scale. *PLoS One*. 2019;14(9):e0222501. Published 2019 Sep 26. doi:10.1371/journal.pone.0222501

Gray MM, Parmenter MD, Hogan CA, et al. Genetics of Rapid and Extreme Size Evolution in Island Mice. *Genetics*. 2015;201(1):213-228. doi:10.1534/genetics.115.177790

Hsieh TC, Ma KH, Chao A (2022). *iNEXT: Interpolation and Extrapolation for Species Diversity*. R package version 3.0.0, http://chao.stat.nthu.edu.tw/wordpress/software_download/.

Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet, et al (2022). vegan: Community Ecology Package. R package version 2.6-2. https://CRAN.R-project.org/package=vegan

Kappes PJ, Siers SR, Leinbach I, et al. Relative palatability and efficacy of brodifacoum-25D conservation rodenticide pellets for mouse eradication on Midway Atoll. *Biol Invasions* 24, 1375–1392 (2022). https://doi.org/10.1007/s10530-021-02714-1

Knowles SCL, Eccles RM, Baltrūnaitė L. Species identity dominates over environment in shaping the microbiota of small mammals. *Ecol Lett*. 2019;22(5):826-837. doi:10.1111/ele.13240

Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species distinguishability and phylosymbiosis along the length of the gastrointestinal tract. *Mol Ecol*. 2018;27(8):1874-1883. doi:10.1111/mec.14460

Kreisinger J, Cížková D, Vohánka J, Piálek J. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Mol Ecol*. 2014;23(20):5048-5060. doi:10.1111/mec.12909

Lagkouvardos I, Pukall R, Abt B, et al. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota [published correction appears in Nat Microbiol. 2016 Oct 17;1(11):16219]. *Nat Microbiol*. 2016;1(10):16131. Published 2016 Aug 8. doi:10.1038/nmicrobiol.2016.131

Lagkouvardos, I., Lesker, T.R., Hitch, T.C.A. *et al.* Sequence and cultivation study of *Muribaculaceae* reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* 7, 28 (2019). https://doi.org/10.1186/s40168-019-0637-2

Lavrinienko, A., Scholier, T., Bates, S.T. *et al.* Defining gut mycobiota for wild animals: a need for caution in assigning authentic resident fungal taxa. *anim microbiome* 3, 75 (2021). https://doi.org/10.1186/s42523-021-00134-z

Leo Lahti, Sudarshan Shetty et al. (2017). Tools for microbiota analysis in R. URL: http://microbiota.github.com/microbiota.

Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*. 2005;102(31):11070-11075. doi:10.1073/pnas.0504978102

Li H, Li T, Yao M, et al. Pika Gut May Select for Rare but Diverse Environmental Bacteria. *Front Microbiol*. 2016;7:1269. Published 2016 Aug 17. doi:10.3389/fmicb.2016.01269

Linnenbrink M, Wang J, Hardouin EA, Künzel S, Metzler D, Baines JF. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol Ecol*. 2013;22(7):1904-1916. doi:10.1111/mec.12206

Marsh KJ, Raulo AM, Brouard M, et al. Synchronous Seasonality in the Gut Microbiota of Wild Mouse Populations. *Front Microbiol*. 2022;13:809735. Published 2022 Apr 25. doi:10.3389/fmicb.2022.809735

Maurice CF, Knowles SC, Ladau J, et al. Marked seasonal variation in the wild mouse gut microbiota. *ISME J*. 2015;9(11):2423-2434. doi:10.1038/ismej.2015.53

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiota census data. *PLoS One*. 2013;8(4):e61217. Published 2013 Apr 22. doi:10.1371/journal.pone.0061217

Merrill BD, Carter MM, Olm MR, et al. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing microbes. bioRxiv. Accessed 1.12.2022. doi:10.1101/2022.03.30.486478

Ochman H, Worobey M, Kuo CH, et al. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol*. 2010;8(11):e1000546. Published 2010 Nov 16. doi:10.1371/journal.pbio.1000546

Op De Beeck M, Lievens B, Busschaert P, Declerck S, Vangronsveld J, Colpaert JV. Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *PLoS One*. 2014;9(6):e97629. Published 2014 Jun 16. doi:10.1371/journal.pone.0097629

Orkin JD, Campos FA, Myers MS, Cheves Hernandez SE, Guadamuz A, Melin AD. Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. *ISME J*. 2019;13(1):183-196. doi:10.1038/s41396-018-0256-0

Ormerod, K.L., Wood, D.L.A., Lachner, N. *et al.* Genomic characterization of the uncultured *Bacteroidales* family *S24-7* inhabiting the guts of homeothermic animals.*Microbiome* 4, 36 (2016). https://doi.org/10.1186/s40168-016-0181-2

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol*. 2016;18(5):1403-1414. doi:10.1111/1462-2920.13023

Perl AK, Zhang L, Whitsett JA. Conditional expression of genes in the respiratory epithelium in transgenic mice: cautionary notes and toward building a better mouse trap. *Am J Respir Cell Mol Biol*. 2009;40(1):1-3. doi:10.1165/rcmb.2008-0011ED

Phifer-Rixey M, Nachman MW. Insights into mammalian biology from the wild house mouse Mus musculus. *Elife* 2015; 2015: 1–13.

Phillips CD, Phelan G, Dowd SE, et al. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol Ecol*. 2012;21(11):2617-2627. doi:10.1111/j.1365-294X.2012.05568.x

Pickard JM, Zeng MY, Caruso R, Núñez G. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev*. 2017;279(1):70-89. doi:10.1111/imr.12567

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reese AT, Chadaideh KS, Diggins CE, et al. Effects of domestication on the gut microbiota parallel those of human industrialization. *Elife*. 2021;10:e60197. Published 2021 Mar 23. doi:10.7554/eLife.60197

Rehaume LM, Mondot S, Aguirre de Cárcer D, et al. ZAP-70 genotype disrupts the relationship between microbiota and host, leading to spondyloarthritis and ileitis in SKG mice. *Arthritis Rheumatol*. 2014;66(10):2780-2792. doi:10.1002/art.38773

Ren T, Boutin S, Humphries MM, et al. Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. *Microbiome*. 2017;5(1):163. Published 2017 Dec 21. doi:10.1186/s40168-017-0382-3

Risely A, Wilhelm K, Clutton-Brock T, Manser MB, Sommer S. Diurnal oscillations in gut bacterial load and composition eclipse seasonal and lifetime dynamics in wild meerkats. *Nat Commun*. 2021;12(1):6017. Published 2021 Oct 14. doi:10.1038/s41467-021-26298-5

Rosshart SP, Herz J, Vassallo BG, et al. Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science*. 2019;365(6452):eaaw4361. doi:10.1126/science.aaw4361

Rosshart SP, Vassallo BG, Angeletti D, et al. Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance. *Cell*. 2017;171(5):1015-1028.e13. doi:10.1016/j.cell.2017.09.016

Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease [published correction appears in Nat Rev Immunol. 2009 Aug;9(8):600]. *Nat Rev Immunol*. 2009;9(5):313-323. doi:10.1038/nri2515

Schnorr SL, Candela M, Rampelli S, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun*. 2014;5:3654. Published 2014 Apr 15. doi:10.1038/ncomms4654

Sepulveda J, Moeller AH. The Effects of Temperature on Animal Gut Microbiomes. *Front Microbiol*. 2020;11:384. Published 2020 Mar 10. doi:10.3389/fmicb.2020.00384

Sharon G, Sampson TR, Geschwind DH, Mazmanian SK. The Central Nervous System and the Gut Microbiome. *Cell*. 2016;167(4):915-932. doi:10.1016/j.cell.2016.10.027

Smits SA, Leach J, Sonnenburg ED, et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*. 2017;357(6353):802-806. doi:10.1126/science.aan4834

Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiota in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-588. doi:10.1038/s41586-018-0617-x

Suzuki TA, Martins FM, Nachman MW. Altitudinal variation of the gut microbiota in wild house mice. *Mol Ecol*. 2019;28(9):2378-2390. doi:10.1111/mec.14905

Suzuki TA, Nachman MW. Spatial Heterogeneity of Gut Microbial Composition along the Gastrointestinal Tract in Natural Populations of House Mice. *PLoS One*. 2016;11(9):e0163720. Published 2016 Sep 26. doi:10.1371/journal.pone.0163720

Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. *Biol Lett*. 2014;10(2):20131037. Published 2014 Feb 12. doi:10.1098/rsbl.2013.1037

Thomson CA, Morgan SC, Ohland C, McCoy KD. From germ-free to wild: modulating microbiome complexity to understand mucosal immunology. *Mucosal Immunol*. 2022;15(6):1085-1094. doi:10.1038/s41385-022-00562-3

Walters W, Hyde ER, Berg-Lyons D, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2015;1(1):e00009-15. Published 2015 Dec 22. doi:10.1128/mSystems.00009-15

Wang J, Kalyan S, Steck N, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun*. 2015;6:6440. Published 2015 Mar 4. doi:10.1038/ncomms7440

Wang J, Linnenbrink M, Künzel S, et al. Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiota of wild mice. *Proc Natl Acad Sci U S A*. 2014;111(26):E2703-E2710. doi:10.1073/pnas.1402342111

Weldon L, Abolins S, Lenzi L, Bourne C, Riley EM, Viney M. The Gut Microbiota of Wild Mice. *PLoS One*. 2015;10(8):e0134643. Published 2015 Aug 10. doi:10.1371/journal.pone.0134643

Yan J, Herzog JW, Tsang K, et al. Gut microbiota induce IGF-1 and promote bone formation and growth. *Proc Natl Acad Sci U S A*. 2016;113(47):E7554-E7563. doi:10.1073/pnas.1607235113

Yassour M, Vatanen T, Siljander H, et al. Natural history of the infant gut microbiota and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016;8(343):343ra81. doi:10.1126/scitranslmed.aad0917

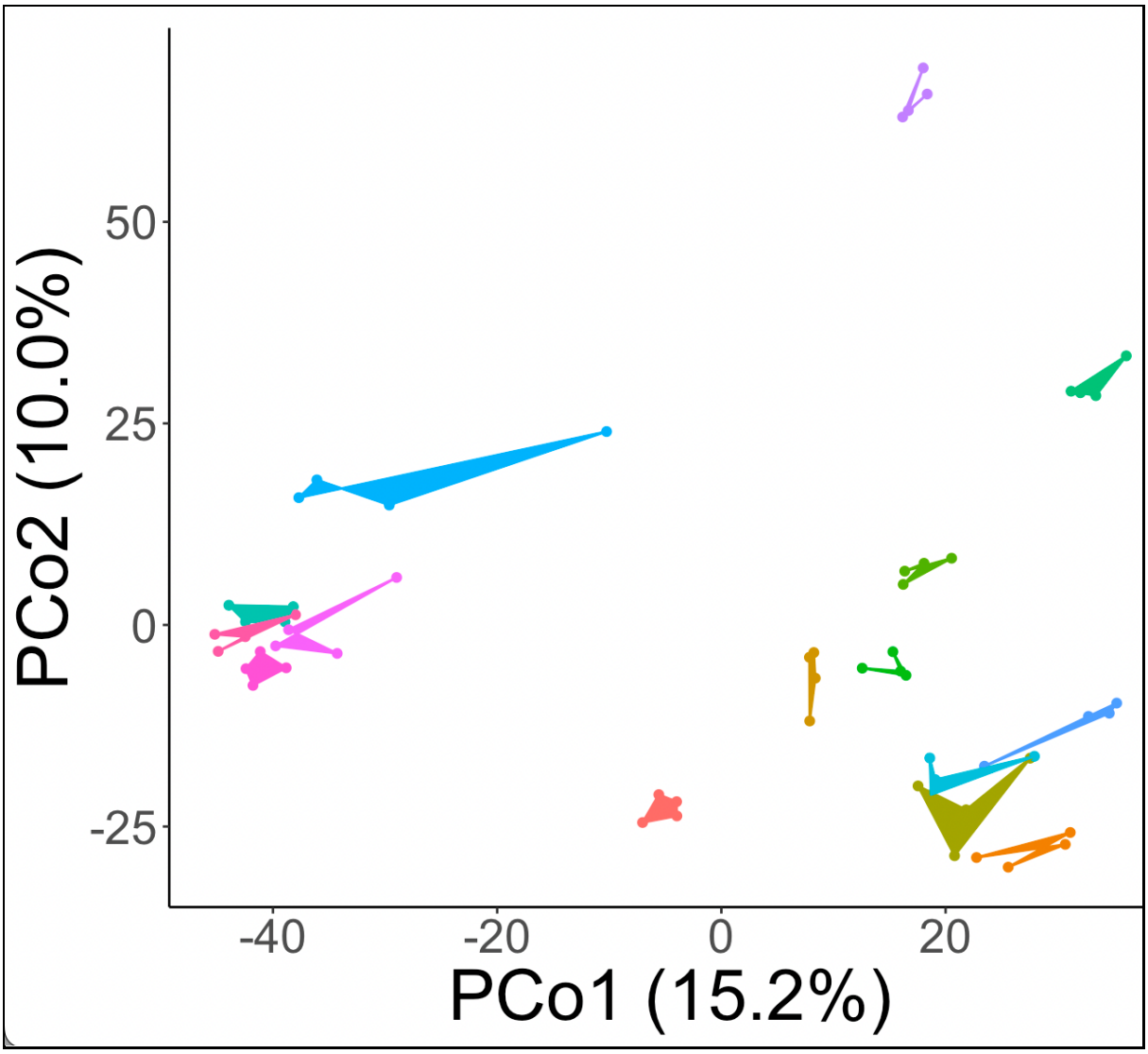
Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222-227. Published 2012 May 9. doi:10.1038/nature11053

Zhou D, Zhang H, Bai Z, et al. Exposure to soil, house dust and decaying plants increases gut microbial diversity and decreases serum immunoglobulin E levels in BALB/c mice. *Environ Microbiol*. 2016;18(5):1326-1337. doi:10.1111/1462-2920.12895

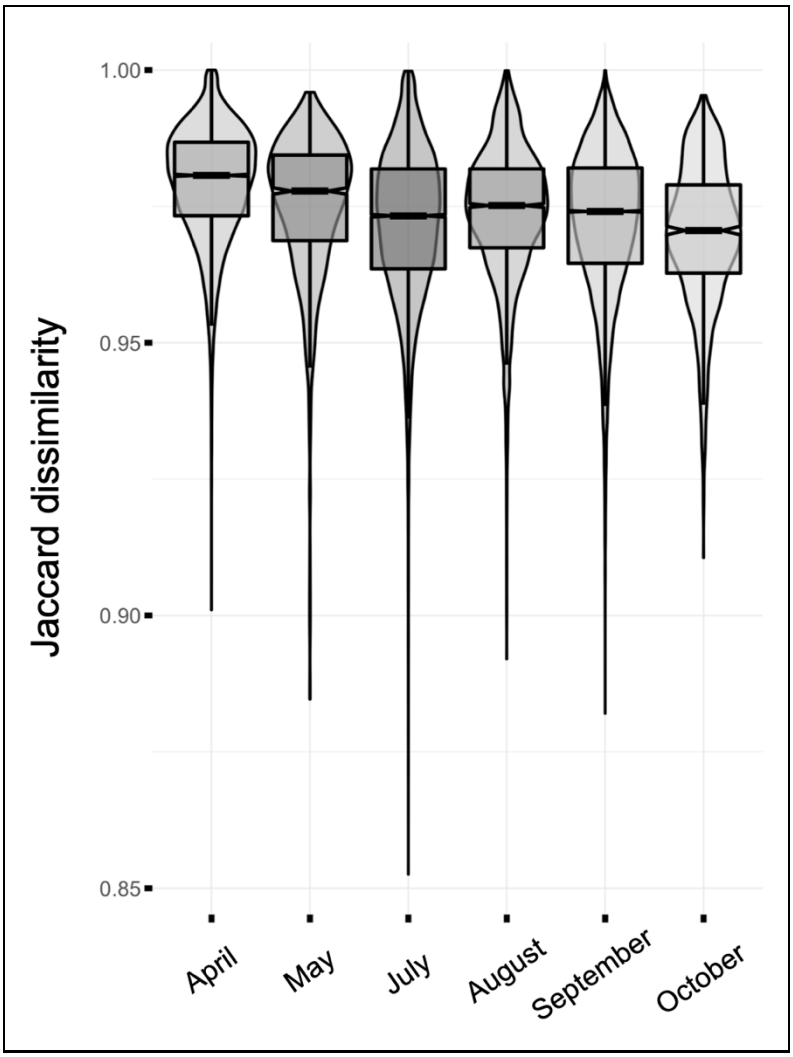
Supplementary data

Supplementary Table 1. Origin, sample type and preservation of samples used in the study. Samples were shipped on dry ice unless preservative allowed sample shipping at room temperature (DNA/RNA Shield).

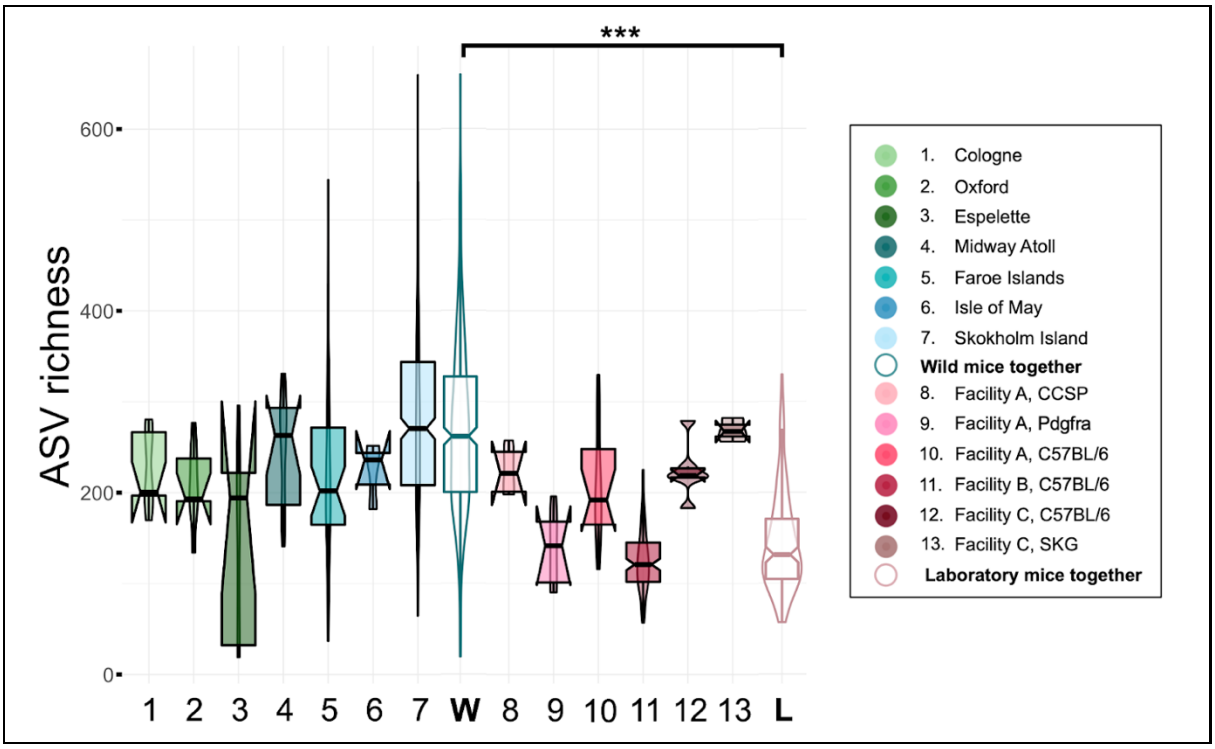
Source	**Strain (*abbreviation*)**	**Facility/population**	**Sample type**	**Preservative, storage temperature**	**Sample size**
Laboratory	C57BL/6	Animal facility A (BMS, Oxford, UK)	Faeces	DNA/RNA Shield, -80◦C	39
Laboratory	C57BL/6	Animal facility B (Kennedy Institute, Oxford, UK)	Faeces	DNA/RNA Shield, -80◦C	6
Laboratory	C57BL/6	Animal facility C (King's College, London, UK)	Faeces	DNA/RNA Shield, -80◦C	123
Laboratory	SKG	Animal facility B (Kennedy Institute, Oxford, UK)	Faeces	DNA/RNA Shield, -80◦C	3
Laboratory	Pdgfra-CreER (*Pdgfra*)	Animal facility A (BMS, Oxford, UK)	Faeces	DNA/RNA Shield, -80◦C	6
Laboratory	CCSP-rtTA (*CCSP)*	Animal facility A (BMS, Oxford, UK)	Faeces	DNA/RNA Shield, -80◦C	7
Wild	N/A	Wytham Field Station, Oxford, UK	Faeces	DNA/RNA Shield, -80◦C	12
Wild	N/A	Skokholm Island, Wales, UK	Faeces	DNA/RNA Shield, -20◦C/-80◦C	948
Wild	N/A	Isle of May, Scotland, UK	Faeces	None, -80◦C	10
Wild	N/A	Cologne, Germany	Faeces	RNAlater, -80◦C	11
Wild	N/A	Espelette, France	Faeces	PBS, -80◦C	8
Wild	N/A	Midway Atoll, North Pacific Ocean	Intestinal contents (colon)	70% isopropyl alcohol until moved to DNA/RNA Shield for shipping from US to UK, -20◦C/-80◦C	20
Wild	N/A	Faroe Islands, North Atlantic Ocean	Intestinal contents (colon)	DNA/RNA Shield, -20◦C	38



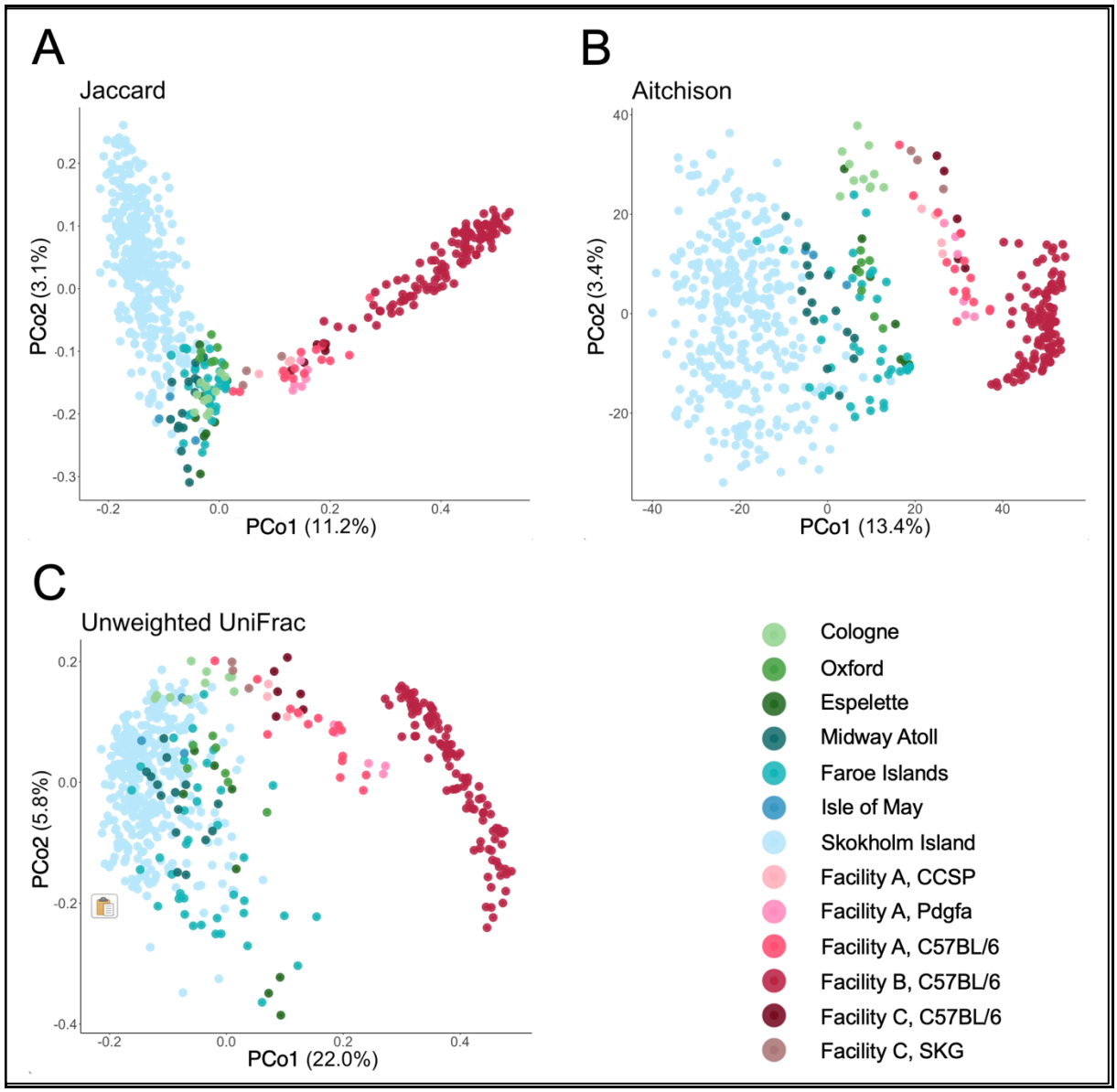
Supplementary Figure 1. Principal coordinate analysis of four aliquots from 15 Skokholm mice on Aitchison distances. Aliquots were stored in DNA/RNA Shield, RNAlater, absolute ethanol or without preservation buffer. Colours represent individual mice, points are different aliquots.



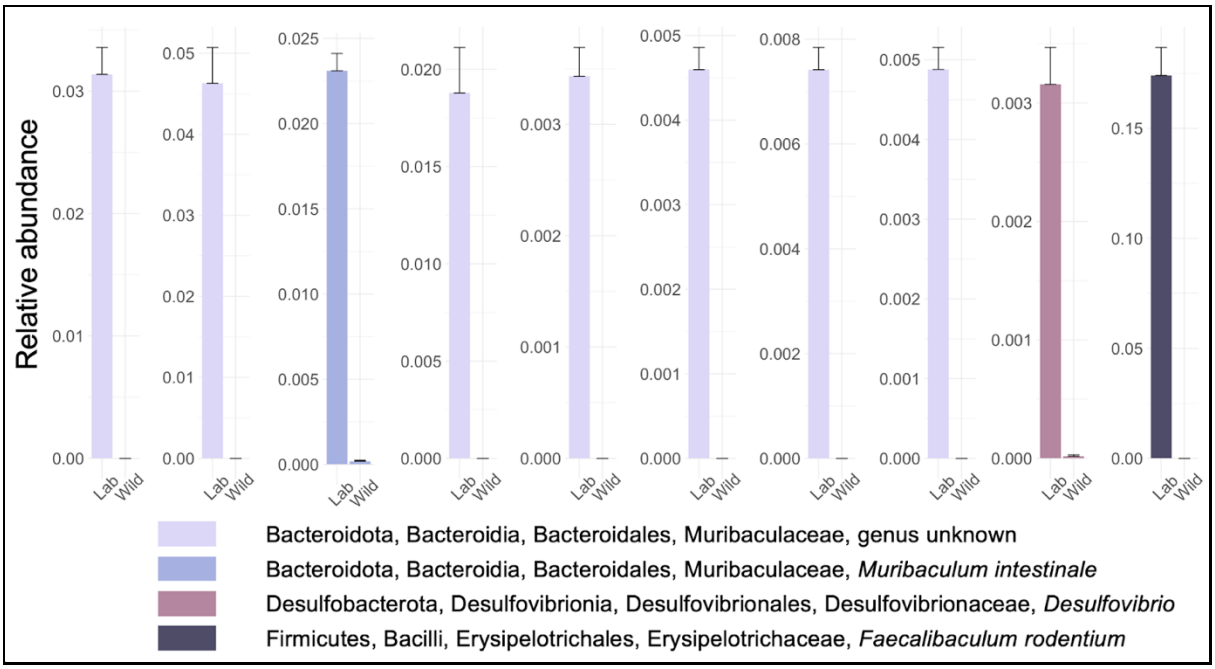
Supplementary Figure 2. Jaccard dissimilarity between laboratory-wild mouse sample pairs across six different sampling months. Laboratory samples include samples from six colonies and wild mouse samples are from Skokholm Island population (Supplementary Table 1). Only samples from adult mice (>3 months of age in laboratory, >17g of body mass in wild mice).



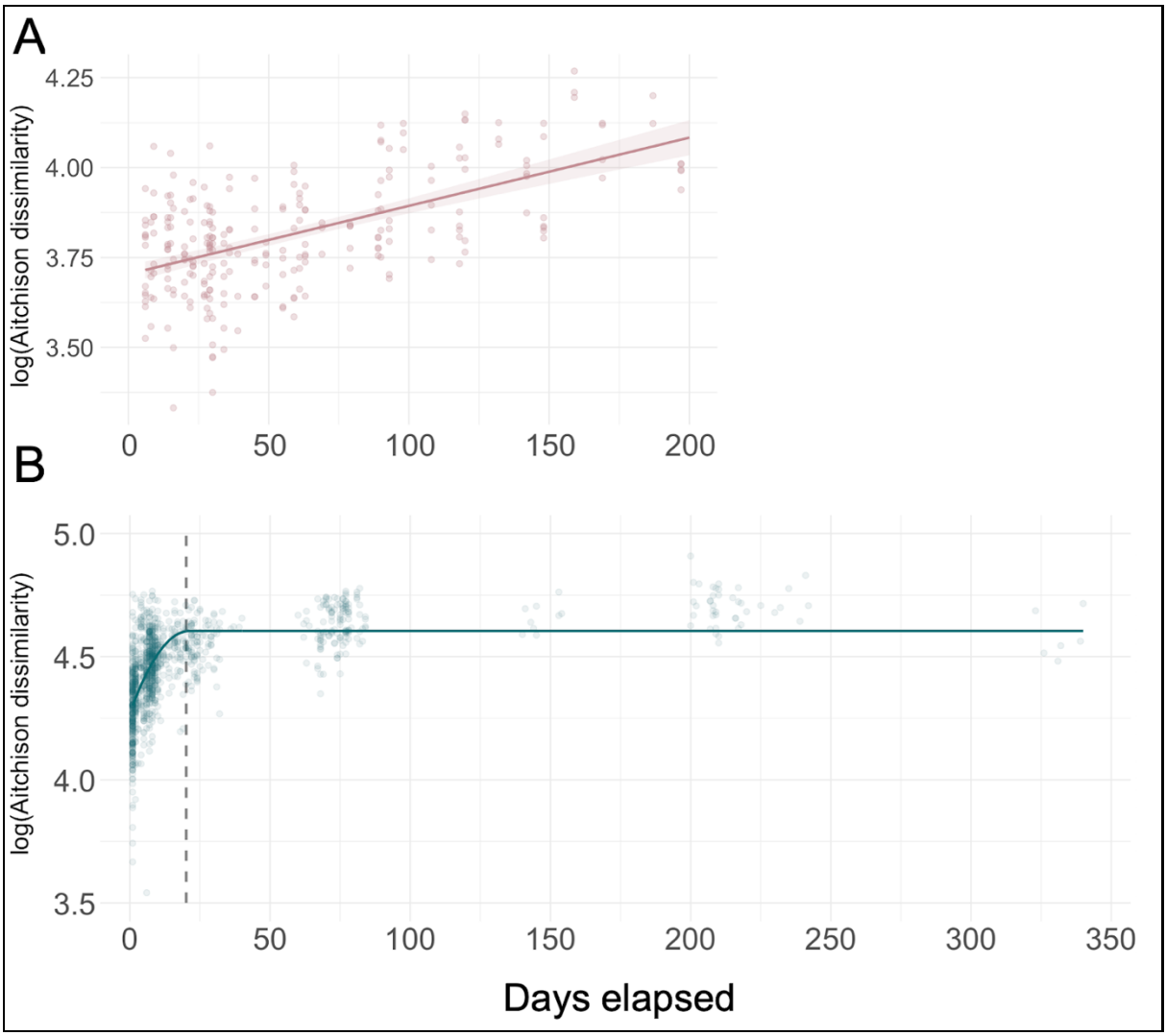
Supplementary Figure 3. Asymptotic ASV richness in wild (*n*=436) and laboratory (*n*=146) mice. Boxplots are individual wild mouse populations (mainland populations, *green*; island populations, *blue*) or laboratory mouse colonies (*pink*). Empty boxplots are all wild (*blue*) or laboratory (*pink*) mouse samples together. Statistical differences between lab and wild mice were tested with permutational Wilcoxon rank sum tests (***; $p<0.001$).



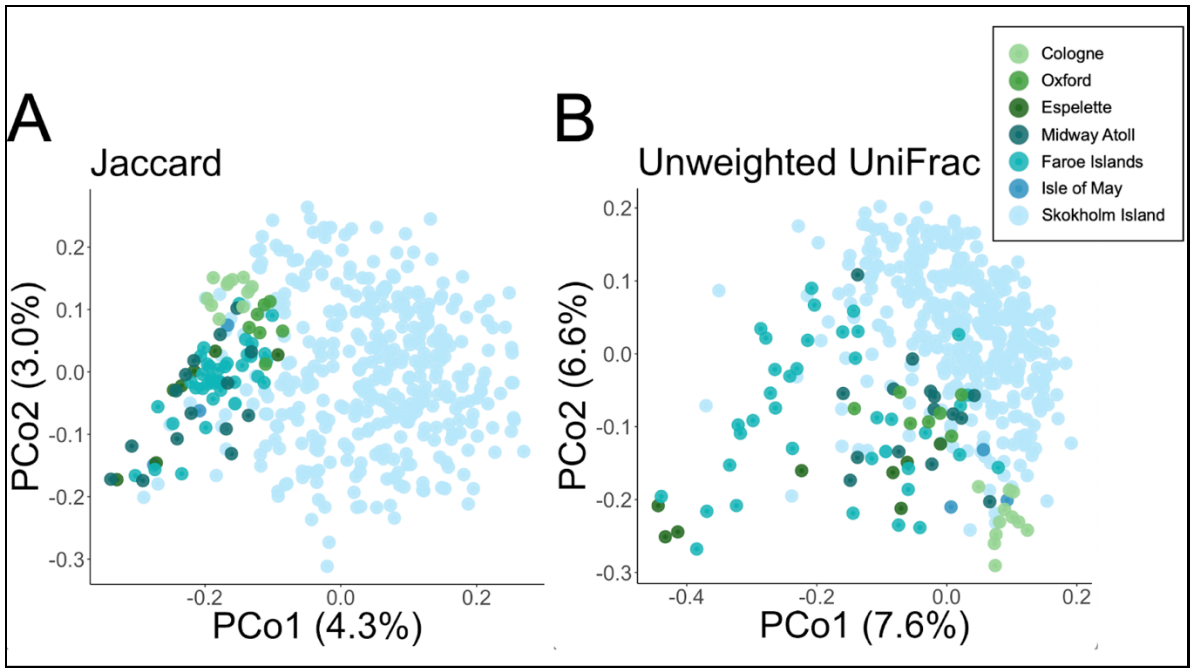
Supplementary Figure 4. (**A–C**) Principal coordinate analysis (PCoA) of wild and laboratory samples on (**A**) Jaccard, (**B**) Aitchison, and (**C**) unweighted UniFrac distances from seven wild populations and six laboratory colonies.



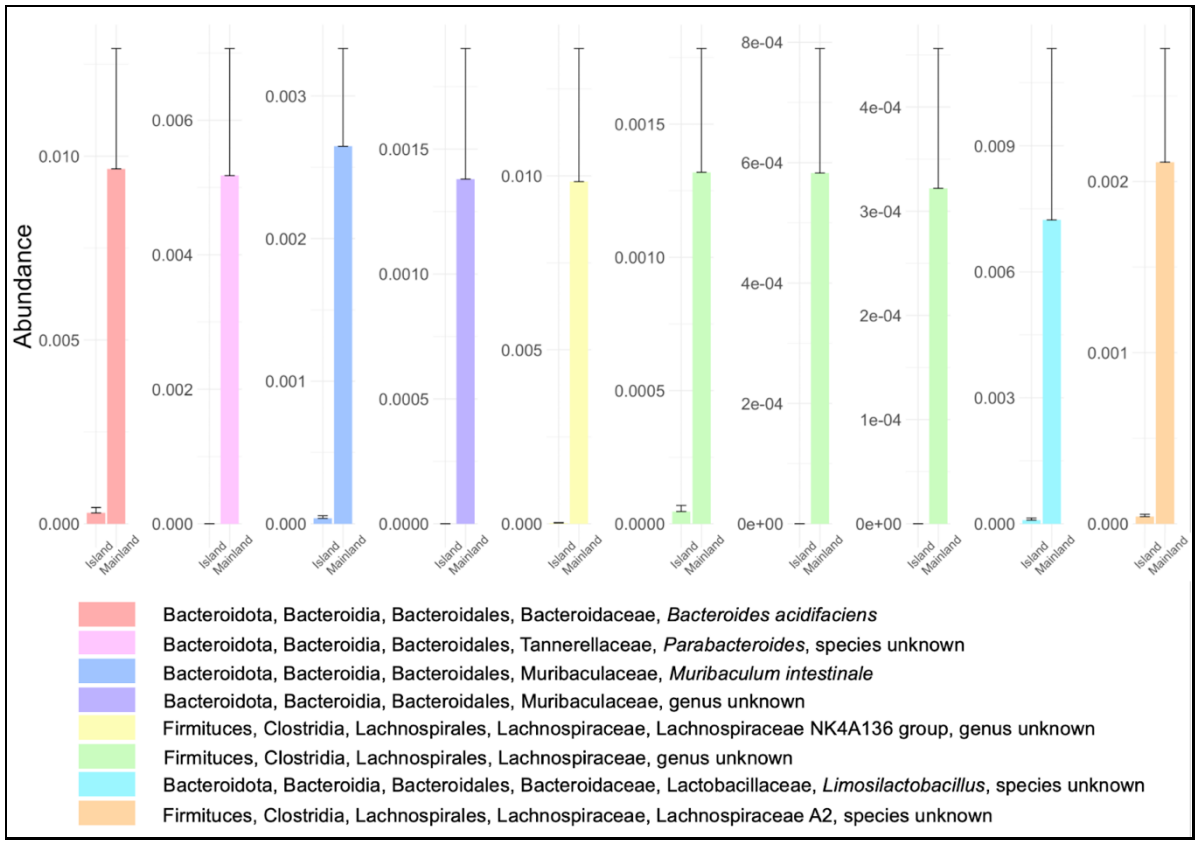
Supplementary Figure 5. Relative abundance of top ten amplicon sequence variants (ASVs) predicting source (lab/wild), ordered in decreasing order of importance from left to right. ASVs were identified with Random Forest regressions, with mean decrease in Gini used as a measure of importance, such that the higher the score, the higher the importance (OOB estimate of error = 0%).



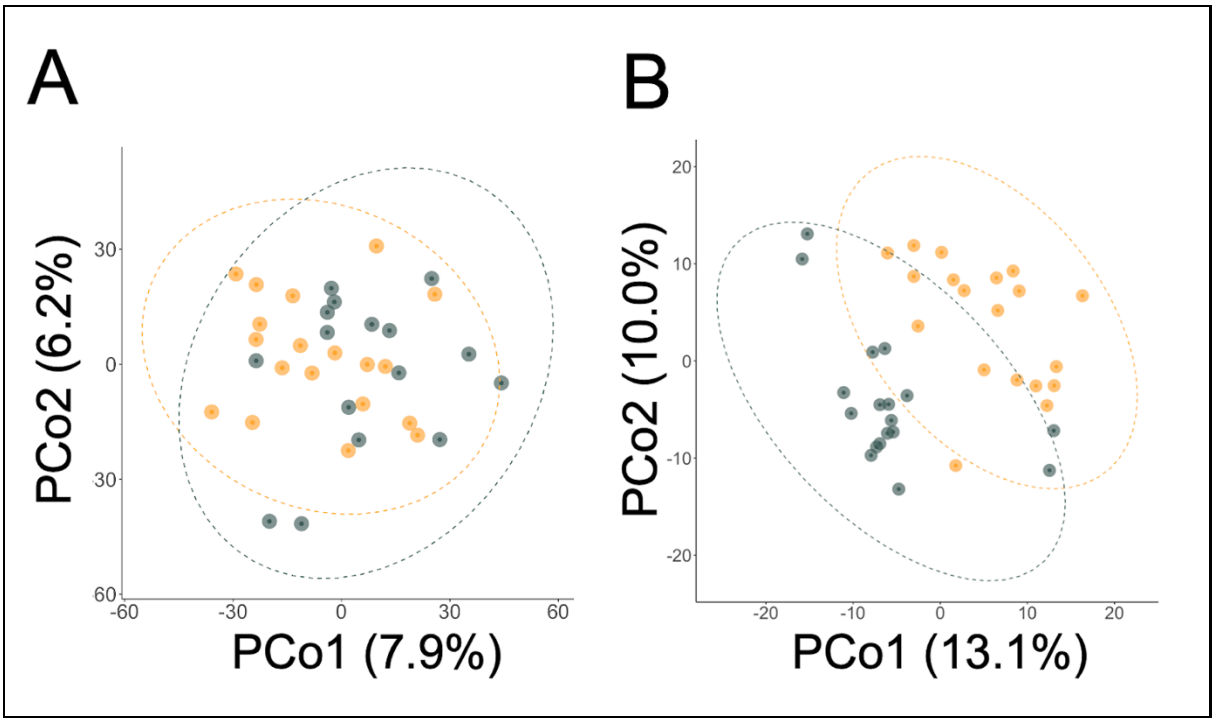
Supplementary Figure 6. Within-individual microbiota dissimilarity on Aitchison distance (log-transformed) between sample pairs from the same (**A**) laboratory or (**B**) wild mouse (99 samples from 16 laboratory mice, 5–7 samples per mouse; 762 samples from 217 wild mice, 2–12 samples per mouse) against time between samples. Laboratory mice were from a single C57BL/6 colony (King's College). The relationship is fitted with (**A**) a log-linear model (lab mice; F_{254}=130.0, adjusted R^2=0.3359, p<0.001) and (**B**) a quadratic plateau model (wild mice; adjusted R^2=0.4394, critical point of inflexion=20.1 days). Note the distinct y-axes scales to aid visibility of dissimilarity patterns across time.



Supplementary Figure 7. Principal coordinate analysis of mainland (*green*) and island (*blue*) mouse samples on (**A**) Jaccard and (**B**) unweighted UniFrac distances.



Supplementary Figure 8. Relative abundance of top ten amplicon sequence variants (ASVs) predicting setting (island/mainland), ordered in decreasing order of importance from left to right. ASVs were identified with Random Forest regressions, with mean decrease in Gini used as a measure of importance (the higher the score, the higher the importance; OOB estimate of error = 3.9%).



Supplementary Figure 9. Principal coordinates analysis (PCoA) on Aitchison distance based on (**A**) bacteria and (**B**) fungi in wild mice from two geographically distinct sampling sites on Skokhlm Island (*yellow* = Observatory, *blue* = Quarry).

CHAPTER II

A targeted methylation-based sequencing approach for aging wild mice

Authors

Eveliina Hanski[1], Susan Joseph[2], Aura Raulo[1], Aine O'Toole[3], Sarah C L Knowles[1], Tom J Little[3]

Title

A targeted methylation-based sequencing approach for aging wild mice

Abstract

Myriad biological processes from immune maturation to chronic diseases are tied to age. Most insights into age-related processes come from humans or laboratory animals. Wild animals have been the target of fewer studies, and one barrier is that the precise age of wild animals is often unknown. One approach for estimating age in wild individuals takes advantage of age-related changes in DNA methylation. Here, we built an epigenetic clock using faecal samples, which have the benefit of being non-invasive, from standard laboratory house mice (C57BL/6, *Mus musculus)* and then used it to gain age estimates for wild mice of unknown age. We show that this clock accurately predicts adult wild mice to be older than juveniles, and that wild mice typically increase in epigenetic age over time, although the rate of epigenetic aging varies widely between individuals. Further, we demonstrate that from early life onwards wild mice are epigenetically older than lab mice, and that this is not explained by accelerated aging post-capture. This suggests these different epigenetic age profiles in mice from different genetic and environmental backgrounds arise very early in life and may be driven by peri- and postnatal effects on offspring DNA methylation.

Introduction

Age is a key characteristic for any organism, with various biological processes from immune maturation to reproduction being age-related. Most insights on age-related processes come from human or laboratory animal studies, both of which have limitations. Humans are long-lived and exposed to environmental factors that are difficult to disentangle. On the other hand, typically short-lived laboratory animals are often inbred and live in an artificial world that lacks natural environmental variation and stress. Studying outbred wild animals has the potential to overcome some of these limitations and complement previous findings from humans and laboratory animals. Wild populations may prove particularly powerful study systems for understanding age-related processes such as immune maturation or senescence at the level of both individual and population.

Yet, measuring age in wild individuals can be challenging as date of birth is often unknown and alternative measures of age are often imprecise (e.g., body size) or destructive (e.g., post mortem measurement of eye lens weight), raising ethical concerns and preventing longitudinal studies. An alternative approach for measuring age in the wild lies in the concept of epigenetic age (Jarman et al., 2015; Anderson et al., 2021; Pinho et al., 2022; De Paoli-Iseppi et al., 2017; Polanowski et al., 2014; Thompson et al., 2017; Wright et al., 2018; Mayne et al., 2022; Larison et al., 2021; Bors et al., 2021; Sullivan et al., 2022). Epigenetic age can be estimated by measuring DNA methylation rates across or at specific sites in the genome. DNA methylation refers to the addition of methyl groups to DNA molecules, usually at CpG sites (those where a cytosine is followed by a guanine; Moore et al., 2013). At some CpG sites, the proportion of methylated cytosines appears to change linearly with age, and together these sites can be used to derive an ‘epigenetic clock’. By training a clock with samples from individuals of known age, the clock can then be used to predict age in individuals of unknown age, possibly providing

a more accurate estimate of chronological age than is otherwise obtainable from visible characteristics in the wild (Mayne et al., 2022; Larison et al., 2021).

Alongside tracking chronological age, epigenetic clocks also appear to capture signals of biological age. Accelerated epigenetic aging has been linked to various communicable and non-communicable diseases in both humans and laboratory mice (Joyce et al., 2021; Morales Berstein et al., 2022; Ambatipudi et al., 2017; Harvanek et al., 2021; Cao et al., 2022; Peng et al., 2019). Insights also come from the wild: high social rank is associated with accelerated epigenetic aging in wild baboons (Anderson et al., 2021), and hibernation slows down aging in marmots and bats (Pinho et al., 2022; Sullivan et al., 2022). Thus, the use of epigenetic clocks may provide a means of estimating chronological age in wild animals while simultaneously providing insight into biological aging in a natural setting.

Previous studies using epigenetic clocks have focused on humans (Joyce et al., 2021; Morales Berstein et al., 2022; Ambatipudi et al., 2017; Harvanek et al., 2021; Cao et al., 2022; Peng et al., 2019) or either laboratory (Han et al., 2018; Kerepesi et al., 2022) *or* wild animals (Prado et al., 2021; Polanowski et al., 2014; Anderson et al., 2021; De Paoli-Iseppi et al., 2018; Lemaître et al., 2022), rather than integrating the two. Studying epigenetic age in both lab and wild animals of the same species in a comparative manner could help us understand drivers of biological aging and its variability in individuals from contrasting genetic and environmental backgrounds. Further, using faecal samples as a source of DNA would make the method non-invasive, allowing longitudinal sampling without ethical limitations to sampling frequency, and enable epigenetic age estimation even in species that are hard to capture.

Here, we hypothesised that both laboratory and wild house mice undergo age-related DNA methylation, possibly allowing application of a lab-based epigenetic clock in age prediction of wild mice. Thus, we set out to test whether we could develop a method of accurately aging wild house mice of unknown age, and compare epigenetic age and aging patterns across laboratory and wild settings. We used faecal samples to do this, in order to minimise ethical and logistical barriers of collecting data for these purposes. We built an epigenetic clock using samples from laboratory mice and used this lab-based clock to predict age in house mice from a wild population. Our results show the potential of such an approach but also indicate contrasting genetic and environmental backgrounds between lab and wild individuals limit accuracy in predicted age.

Methods

Sample collection

A total of 137 faecal samples were collected from 65 individual *Mus musculus* C57BL/6 laboratory mice (30 females, 35 males) from two animal facilities. The samples were collected in May–November 2021 at the Biomedical Services Building, Oxford, UK (Animal facility A), and King's College, London, UK (Animal facility B). The chronological age of the mice varied from 7 to 339 days. The mice were kept in standard housing and were not subject to any interventions before or during sampling. Body mass was recorded for mice from Animal facility B during sample collection. Body mass was estimated for 25 lab mice (from Animal facility A) under seven weeks of chronological age for which body mass was not recorded during sample collection. Body mass estimation was done based on Spangenberg et al. (2014) for 7–20-day old pups and The Jackson Laboratory C57BL/6 body mass references for 3–7-week-old pups (Jax, 2022a; Jax, 2022b). For the latter age group, estimation was done separately for females and males. To collect faecal samples, mice were briefly placed on a

sterile surface until defecation. Faecal pellets were collected in a sterile manner, immediately preserved in DNA/RNA Shield, and stored frozen at -80°C until further processing.

Wild house mouse (*Mus musculus domesticus*) sampling was conducted in April–May 2019, July 2019, September–October 2019, August–September 2020, and April–May 2021 on Skokholm Island, Wales, UK. Mice were trapped overnight using small Sherman live traps baited with peanuts and non-absorbent cotton wool for bedding, and with a spray of sesame oil outside the trap as a lure. Across each of two broad sampling sites (one near the coast and one in the island interior), on each trapping night 150 traps were set at dusk and checked at dawn. To prevent cross-contamination, any traps showing signs of mouse presence were washed and sterilised before being reset using bleach solution (including a ≥60 min soak in 20% bleach) to destroy bacterial cells and DNA.

All newly captured mice were permanently identified by subcutaneous injection of a passive integrated transponder (PIT) tag. Upon each capture, each mouse was either tagged or identified (if a recapture), aged, sexed, and measured before being released at its trapping point. Age category was determined based on size and pelage characteristics: small (typically <15g of mass and <80mm of length) mice were classified as juveniles and full-sized mice (typically >20g and >80mm of length) were classified as adults. Mice falling between the two categories were ranked as sub-adults. Sex was determined using anogenital distance and reproductive state. At each capture, body mass and body condition were recorded. Body condition was scored from 1 to 4 by palpating the lower back area to estimate the amount of subcutaneous fat.

Faecal samples were collected from traps in a sterile manner, preserved in DNA/RNA Shield and stored in a -20°C freezer until returned to the laboratory frozen (maximum 6 weeks after sample collection). Collected faecal samples were then stored in -80°C until DNA extraction. A total of 215 samples were selected from all collected samples (>900) for further processing by first selecting a longitudinal dataset (mice sampled ≥2 over time) with as much variation as possible in morphometric variables (age, body mass, sex, and reproductive status), as well as environmental variables (sampling season and sampling site) and then supplementing this with additional (equally variable) cross-sectional samples to increase number of individuals for cross-sectional analyses. Variation in variables was achieved by randomly selecting approximately equal numbers of samples e.g., across juvenile, sub-adult and adult mice.

DNA extraction, bisulfite conversion and PCR amplification

DNA was extracted from faecal samples using the ZymoBIOMICS DNA MiniPrep Kit according to the manufacturer's protocol (Zymo Research, Irvine, California, USA). DNA was then bisulfite-converted using the Zymo EZ DNA Methylation-Gold Kit to convert unmethylated cytosines to urasil and then thymine (Zymo Research, Irvine, California, USA). PCR amplification was conducted for five genes previously reported to correlate with chronological age in *Mus musculus*; *Prima1, Hsf4, Kcns1, Gm9312*, and *Gm7325* (Han et al., 2018). Amplification was conducted using the PyroMark PCR Kit according to manufacturer's instructions and primers for the five genes (QIAGEN, Hilden, Germany; Supplementary Table 2; Han et al., 2018). Amplification success was confirmed using gel electrophoresis. The five amplicons (PCR products) were pooled for each sample into 5-gene libraries. DNA was quantified with Qubit Fluorometer High Sensitivity dsDNA kit and normalised to 6.25 ng/µl (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Sequencing, basecalling, and demultiplexing

The Oxford Nanopore Technology (ONT) platform was used for sequencing of the libraries (Oxford Nanopore Technologies, Oxford, UK). We first used the ONT Ligation Sequencing Kit (SQK-LSK109), to repair and dA-tail the DNA ends, followed by ligation of sequencing adaptors to the prepared ends. We then barcoded libraries using the ONT Native Barcoding Expansion kit (EXP-NBD104 or EXP-NBD196). Approximately 15 ng of the prepared library was loaded onto a prepared ONT MinION Mk1B R9.4.1 flow cell and sequenced using the ONT MinKNOW software v21.10.4.

Libraries were sequenced across a total of six runs resulting in a mean of 49,969 reads per sample. A negative control (where DNAse free H_2O was used instead of pooled amplicons at the start of Nanopore pipeline) was included in three sequencing runs, and these generated a mean of 200 (range 17–519) reads. One flow cell was used twice, and washed between runs with the ONT Flow Cell Wash kit (EXP-WSH003). Different barcodes were used for negative controls across the two sequencing runs where the same flow cell was used to enable testing for carry-over of reads (17 potential carry-over reads). All ONT procedures were conducted according to manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK).

Raw sequencing data was basecalled and demultiplexed using High Accuracy basecalling on the ONT Guppy software v5.0.11. The basecalled fastq files were then run through the Apollo software v0.1 (https://github.com/WildANimalClocks/apollo) to acquire methylation rates for each CpG site within the five genes. Using the alignment with the reference genes, target sites (cytosines within CpGs) were identified in each read and determined as either methylated (cytosine) or unmethylated (urasil). The process was continued for each read, resulting in a proportion of methylated cytosines at each CpG site.

Analyses

The data was analysed and visualised in R (v4.1.2). The epigenetic clock was built using a cross-sectional dataset from C57BL/6 mice from two facilities (n=50). We performed generalized linear regression with the package glmnet v4.1–3 (Friedman et al., 2010) to identify CpG sites whose methylation is predicted by chronological age, using a LASSO model (mixing parameter alpha=1) and a leave-one-out cross validation (nfolds=nrow). We failed to acquire sufficient read counts and thus methylation rates for CpG sites from the gene *Gm7325* in 20 wild mouse samples (9% of all wild mouse samples). Inclusion of CpG sites from this gene appeared to have little effect when building the epigenetic clock (Pearson's r=0.996 when included and r=0.994 when excluded), hence we did not include CpG sites from *Gm7325* (n=11) in the final clock. The CpG-specific weights from the clock were then used to predict epigenetic age in 30 additional samples from C57BL/6 mice (n=15, two samples per animal) as well as in wild house mice for which methylation rates at the CpG sites included in the clock were successfully measured (n=201; 93% of all wild mouse samples) using linear modeling. Relationship between epigenetic age and chronological age was tested with a Pearson's correlation test.

To test for the effect of covariates on predicted epigenetic age in the validation dataset, we fit a linear model with epigenetic age as the dependent variable and chronological age, sex, cage, and sequencing run ID as predictor variables. Wilcoxon rank sum test with 1,000 permutations was used to test whether the predicted epigenetic age of wild mice ranked as adults was older than that of wild mice ranked as juveniles. The ability of the clock to predict increases in epigenetic age in mice sampled over time was tested with a binomial test, with probability of predicted age at second timepoint being older than at first timepoint being 0.5. Linear modeling was used to test whether increase in time predicted increase in epigenetic age. Age category

and sex at first timepoint were included as predictor variables. Mice with less than 25 days between timepoints were excluded from the longitudinal analysis based on the mean absolute error (MAE) between predicted epigenetic and chronological age in validation dataset (MAE=25 days, Fig. 1B).

To test whether source (lab/wild) predicted epigenetic age when controlling for the effects of body mass, we fit a linear model with epigenetic age as the dependent variable and source and an interaction of body mass and sex as predictor variables. For this analysis, wild mice with recordings of pregnancy or prominent nipples (indication of ongoing/recent pregnancy) were excluded as these significantly predicted body mass. Lastly, to study the rate of epigenetic aging across the two mouse systems we used (1) linear modeling to test whether the rate of epigenetic aging varied between lab and wild mice with increase in epigenetic age as response variable and interaction of increase in time and source as explanatory variable, and (2) Levene's test to investigate whether rate of epigenetic aging is more variable within wild mice. Here, we used the ratio between change in epigenetic age and change in time as response variable and source (lab/wild) as the explanatory variable.

Ethical statement

Wild mouse work was conducted under Home Office license PPL PB0178858 held at the University of Oxford, and with a research permit from the Islands Conservation Advisory Committee (ICAC), and Natural Resources Wales. No mice were subject to intervention for the purposes of this study.

Results

Construction of a non-invasive epigenetic clock

We used faecal samples from C57BL/6 laboratory mice (*Mus musculus*, n=50, one sample per animal) from two animal facilities to generate a DNA methylation-based epigenetic clock. The chronological age of the mice ranged from 7 to 339 days (mean 96, median 40), covering life stages from the pre-weaning postnatal phase to middle-age (Flurkey et al., 2007). We focused on five genes containing CpG sites whose degree of methylation has previously been shown to correlate with chronological age in laboratory mice (Suppl. Fig. 1; Han et al., 2018; Petkovich et al., 2017; Stubbs et al., 2017). We measured methylation at all 73 CpG sites in these genes by first bisulphite treating DNA, PCR amplifying the focal genes, and then sequencing using the Nanopore MinION platform.

An epigenetic clock was ultimately built using 62 of these CpG sites from four genes (*Hsf4*, *Gm9312*, *Kcns1*, and *Prima1*) in which methylation rate was successfully measured across the majority of laboratory and wild mouse samples (see Methods). Elastic net regression identified 18 CpG sites from three of the targeted genes, two from *Hsf4*, five from *Gm9312*, and eleven from *Kcns1* (Suppl. Table 2), that showed a positive linear relationship with chronological age with a mean absolute error (MAE) of 7 days (Pearson's r=0.996, p<0.001; Fig. 1A). We validated the clock by applying it to a set of C57BL/6 mice not used in training the clock (n=15, one sample per animal). Predicted age was strongly correlated with chronological age (Pearson's r=0.933, p<0.001, MAE=25 days; Fig. 1B). Neither sex, cage nor sequencing run had a significant effect on predicted age (linear model; F_8=12.3, chronological age, p<0.001; other variables, p>0.1). This demonstrates that non-invasive faecal samples can be used to generate an epigenetic clock in laboratory mice with equivalent or higher accuracy in

estimating chronological age as a clock previously derived using blood samples (Han et al., 2018; MAE=35–41 days in two validation datasets).

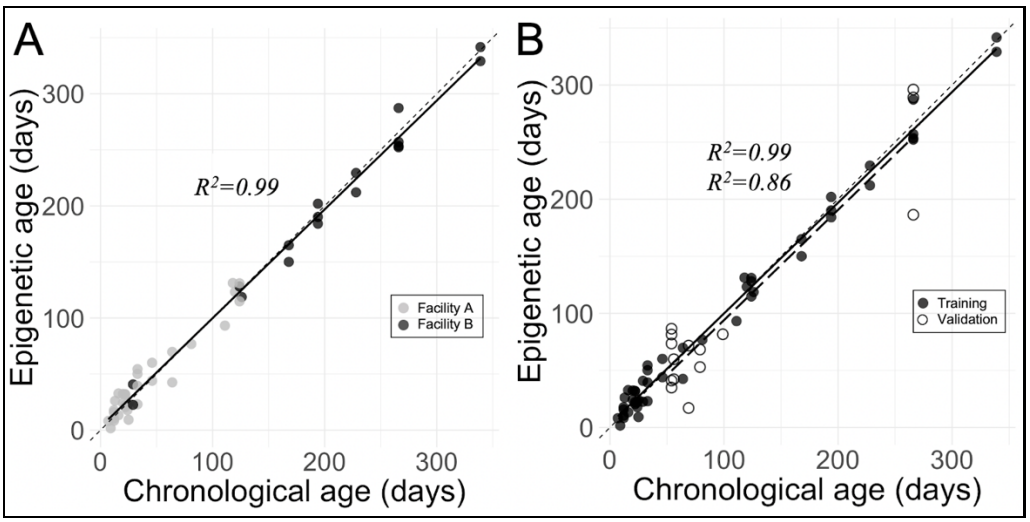


Figure 1. (**A**) DNA methylation-based epigenetic age in relation to chronological age in C57BL/6 mice from two animal facilities (slope estimate = 0.97 ± 0.02 standard error, p<0.001, mean absolute error (MAE) = 7 days, n=50). Circles are individual mice with colour indicating animal facility (*grey* = Animal facility A, *black* = Animal facility B). Solid line is a linear regression line (R^2=0.99). Dashed line is a reference line (y=x). (**B**) The epigenetic clock was used to predict age in 15 validation samples from C57BL/6 mice from Animal facility B (slope estimate = 0.97 ± 0.10 standard error, p<0.001, MAE=25 days). Filled circles are mice from clock training set (n=50, Fig. 1A), empty circles are mice from validation set (n=15). Lines are linear regression lines (*solid* = training set, R^2=0.99; *long dash* = validation set, R^2=0.86; *short dash* = reference line (y=x)).

Chronological age prediction in wild mice

We next applied this lab-mouse derived epigenetic clock to 199 faecal samples from 116 wild house mice to test if we can use it to estimate chronological age in wild individuals of unknown age. Mice of all available body sizes were included with the aim of capturing as much age variation as possible (body mass range 5.9–43.0g, mean 19.1, median 20.1). The epigenetic age of wild mice varied from -20 to 659 days (mean 273, median 260; 2 out of 199 samples (1%) had a negative epigenetic age). To assess the performance of the clock on these wild mice, we first asked if the clock predicted whether a mouse had been categorised as 'adult' or 'juvenile' when captured (see Methods). 'Adult' mice were estimated to be significantly older

than 'juvenile' mice using the epigenetic clock (permutational Wilcoxon rank sum test, $p<0.001$, $n=163$; Fig. 2A).

We then used data from wild mice that were repeat-captured to test whether they increased in epigenetic age over time. Among 34 wild mice sampled twice between 30 and 340 days apart (mean 129, median 81), 30 (88%) were epigenetically older at the latter timepoint (binomial test for H0: p=0.5; $p<0.001$); however, the slope of this relationship varied among wild mice (Fig. 2B). Still, the number of days between sampling points positively predicted change in epigenetic age (linear model, $F_{1,28}=2.2$, $p=0.0162$; Pearson's $r=0.451$, $p=0.007$, $n=34$), and this relationship was stronger when excluding the four mice whose epigenetic age estimates decreased over time (linear model, $F_{1,24}=5.6$, $p<0.001$; Pearson's $r=0.609$, $p<0.001$, $n=30$). We did not detect significant effects from investigated covariates (sex, reproductive status, body mass, or season at first sampling point; linear model, $F_{18}=2.044$, $p>0.05$ for all). These results indicate an epigenetic clock trained with samples from inbred lab mice can be used to provide an estimate of chronological age in outbred wild mice, though not one that is highly precise.

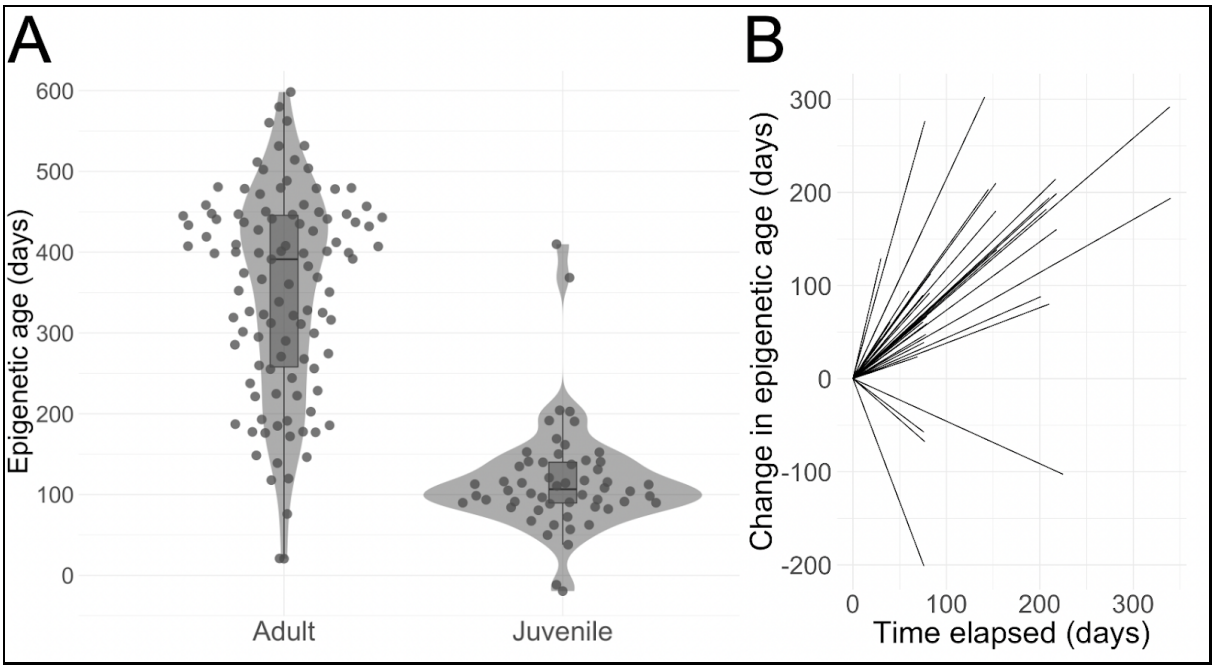


Figure 2. (**A**) Epigenetic age of wild mice phenotypically characterised as adult (*n*=107) and juvenile (*n*=56) predicted with a lab-based epigenetic clock. Age category was based on body size and appearance (see Methods). Median epigenetic age was 391 days in adults and 107 days in juveniles. Circles are individual samples. (**B**) Change in epigenetic age between two timepoints in wild mice sampled twice 30–340 days apart (*n*=34). Lines represent individual mice. Epigenetic age increased with time for 30 out of 34 (88%) mice.

Wild mice are multiple times older epigenetically than laboratory mice

We next assessed whether epigenetic age for a given chronological age differed between wild and laboratory mice. In the absence of known chronological age for wild mice, we used body mass to provide an upper limit age estimate for individuals classed as juveniles. In mice, body mass in very early life largely reflects developmental stage and therefore age (Gerber et al., 2021; Ferrari et al., 2015), and is less influenced by factors, such as reproductive state, which can drive mass variation among adults. Others have reported that 12–13-day-old wild house mice from mainland Europe weigh around 7g (range 3.6–10.5g, mean 6.8; Gerber et al., 2021) and another study has shown that 14-day-old wild-derived house mice from Gough Island (home to the largest wild house mice recorded; data is from pups born in laboratory) weigh around 8.5g (range ~7–10.5g, raw data not available; Gray et al., 2015; Suppl. Fig. 1). We examined epigenetic age from a set of juvenile wild Skokholm Island mice that fall within this

body mass range and thus which we expect to be no more than 25 days old (n=20, body mass 5.9–10.3g with a mean of 8.5g). Among these individuals, the mean epigenetic age estimate was 120 days (range 50–368, median 96), which is more than four times the expected chronological age (Suppl. Fig. 1).

To explore whether wild mice were epigenetically older than lab mice beyond early life, we explored the relationship between epigenetic age and body mass across mice of all sizes. While the reliability of body mass as an indicator of age declines after initial growth during first few weeks of life, it continues to increase with chronological age in both C57BL/6 lab and wild mice beyond this time and thus can be used as a rough estimate of age in adults as well (Jax, 2022a; Jax, 2022b; Gerber et al., 2021; Gray et al., 2015). In a cross-sectional subset of 43 lab and 112 wild mice (wild mice with visual indication of ongoing/recent pregnancy were excluded, see Methods), animal source (lab/wild) predicted epigenetic age (linear model, $F_{1,140}$=14.69, p<0.001; model included an interaction term between sex and body mass), with wild mice being several times epigenetically older on average than lab mice (Fig. 3A). In line, wild mice had higher rates of DNA methylation across mice of all sizes, particularly at CpG sites within the *Gm9312* gene (Suppl. Fig. 2), explaining the higher epigenetic age profile.

To further examine whether older epigenetic age profiles among wild mice might be due to accelerated aging through exposure to environmental stressors, such as food shortage and climate variation, we studied the rate of epigenetic aging ($\frac{change\ in\ epigenetic\ age}{days\ elapsed}$) across lab and wild mice for which two timepoints were available. In lab mice, the rate of epigenetic aging was again close to 1 (slope estimate = 1.08 ± 0.19 standard error) as observed when originally validating the clock using cross-sectional data (slope estimate = 0.97 ± 0.10 standard error; Fig. 1B). This rate appeared lower in wild mice (slope estimate = 0.60 ± 0.21 standard error);

however, it was not statistically different from that seen in lab mice (linear model, $F_{1,45}$=7.49, p=0.195, an interaction term between source and change in time was used to predict change in epigenetic age; Fig. 3B).

Further, although the rate of epigenetic aging appeared to be more variable among wild than lab mice, the variance in epigenetic aging rate did not differ statistically across settings (Levene's test, $F_{1,47}$=0.1333, p=0.7167; correlation between change in epigenetic age and days elapsed: lab mice, Pearson's r=0.85, p<0.001; wild mice, Pearson's r=0.57, p<0.001; Fig. 3B). Overall, these results suggest that wild mice undergo more intensive DNA methylation and thus epigenetic aging in very early life compared to lab mice.

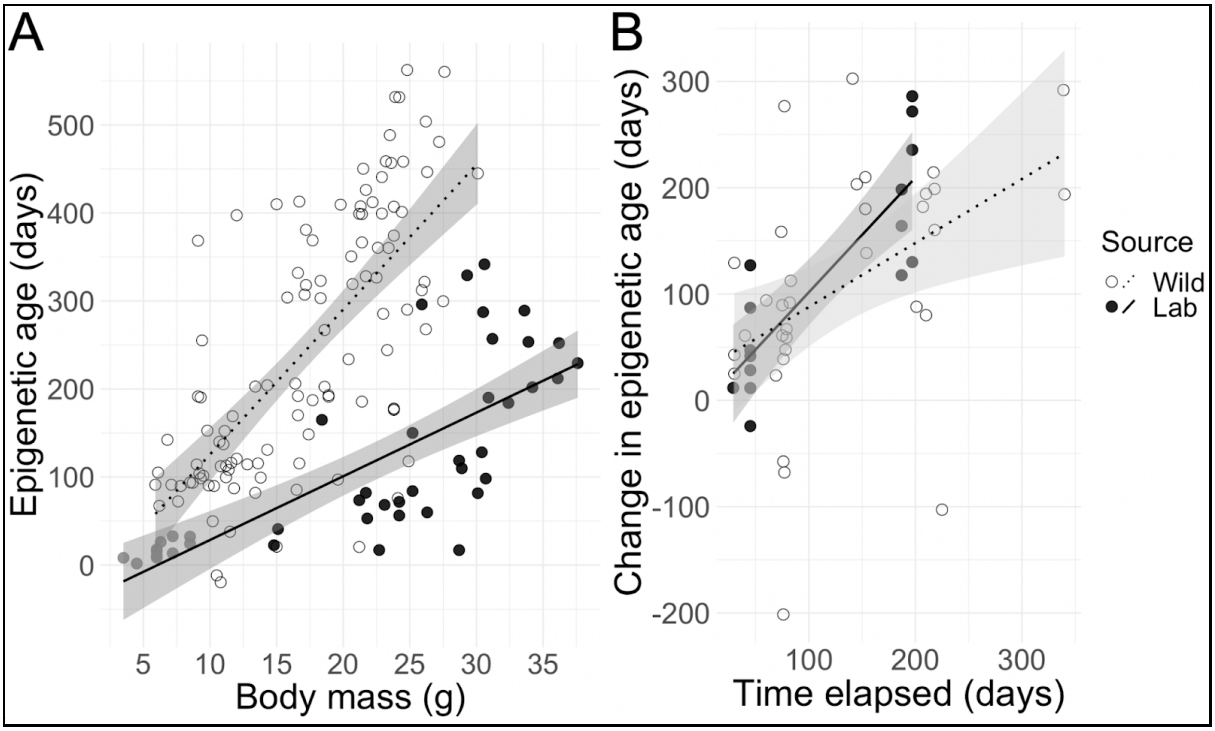


Figure 3. (**A**) Epigenetic age of wild (n=112) and laboratory mice (n=43) in relation to body mass. Empty circles are wild mice and filled circles are lab mice; black circles are lab mice for which body mass was recorded during sample collection, and grey circles are lab mice for which body mass was estimated post hoc (see Methods). Lines are linear regression lines (*dashed* = wild mice, *solid* = lab mice) with 95% confidence interval bands. (**B**) Change in epigenetic age in relation to time elapsed in laboratory (n=15) and wild mice (n=28) sampled twice minimum 25 days apart. Empty circles are wild mice and filled circles are lab mice. Lines are linear regression lines (*dashed* = wild mice, *solid* = lab mice) with 95% confidence interval bands.

Discussion

Here, we present an approach for estimating age in wild mice, by building an epigenetic clock using samples from inbred C57BL/6 laboratory mice and using it to estimate age in outbred wild mice of unknown chronological age. The clock could effectively distinguish wild juveniles from adult mice, and typically predicted increases in age in individuals sampled over time (predicted age was older in 88.0%) wild mice, showing a similar success to the wild baboon study of Anderson et al (2021)).

However, while the clock accurately predicted age in an independent set of laboratory mice, it struggled to provide an accurate estimate of chronological age when applied to mice from an entirely different ecological context. Others have had better success in applying clocks built with captive individuals to wild individuals (Mayne et al., 2022; Robeck et al., 2021; correlation between chronological and epigenetic age in these studies 0.67–0.98 *vs* 0.45–0.61 correlation between change in time and change in epigenetic age in our study). However, these studies have built epigenetic clocks using samples from captive individuals where there is genetic and environmental variation, such as zoos or outdoor enclosures. Our clock was built with samples from inbred lab mice housed under very stable environmental conditions, but applied to wild mice that are outbred and exposed to a highly variable temperate climate. Studies of different lab strains have confirmed that epigenetic clocks may behave differently in different genetic backgrounds. For instance, Han et al demonstrated DBA/2 mice to be up to twice as old epigenetically as C57BL/6 mice (Han et al., 2018).

Moreover, DNA methylation may be influenced by environmental factors (Zocher et al., 2021; Parrott et al., 2014; Viitaniemi et al., 2019). As such, the contrasting genetic and environmental backgrounds in our mouse systems may partly explain why age estimates in wild mice had low

accuracy. Finally, while all samples from lab mice were preserved immediately after defecation, the time between defecation and sample preservation varied in wild mice for which samples were collected from traps where the mice had been up to a few hours. It is possible some DNA degradation by nuclease and gut microbes occurred before the samples were preserved (in DNA/RNA Shield, which inactivates microbes and preserves DNA integrity), affecting the methylation profiles.

The predicted (epigenetic) age of wild mice from Skokholm Island varied from 21 to 659 days (mean 275, median 267; excluding 2 (1% of all 199) samples which had negative epigenetic ages, -12 and -20 days). Despite our epigenetic clock having a 1:1 relationship with chronological age in laboratory mice, several lines of evidence suggest these epigenetic age estimates for wild mice are overestimates. First, when investigating juveniles, in which body mass is an accurate predictor of age (Jax, 2022a; Jax, 2022b; Gerber et al., 2021; Gray et al., 2015), we found that wild juveniles were predicted to be several times older than their expected chronological age from body mass (Gerber et al., 2021; Gray et al., 2015). Second, by comparing the epigenetic age estimates to body mass in mice of all sizes, we found that wild mice were several times epigenetically older at any given body mass than lab mice, with similar patterns seen in CpG site methylation rates. While accelerated weight gain in *ad libitum*-fed lab mice may contribute to a lower epigenetic age among adult lab mice compared to wild mice, these findings suggested that the more challenging environment experienced by wild mice may accelerate epigenetic clocks.

To test whether the older epigenetic age profile of wild mice could be explained by accelerated aging post-weaning (i.e., from when they are trappable) we investigated the rate of epigenetic aging using individuals captured and sampled twice over time. If anything, the rate of

epigenetic age appeared slightly slower and more variable in wild mice, though this observation relied on a relatively small sample size and was not statistically significant. Moreover, as we observed heightened epigenetic age in wild compared to lab mice even during the first ~2 weeks of life, we hypothesize that peri- and early postnatal effects on offspring DNA methylation may vary between laboratory and wild mice and contribute to their different epigenetic age profiles. Various human, mouse, and other animal studies have demonstrated the association between prenatal maternal experience (such as food shortage, diet, infection, substance exposure, and stress) and offspring DNA methylation patterns, with differences detectable from the prenatal (foetal) phase into later life (Tobi et al., 2009; Heijmans et al., 2008; Lan et al., 2013; Richetto et al., 2017; Camerota et al., 2021; Joubert et al., 2016; Kertes et al., 2016; Vangeel et al., 2017).

While our approach of training an epigenetic clock with lab individuals and using it to estimate age in wild individuals does not allow precise estimation of chronological age, our results demonstrate such an approach can still be effective in distinguishing between juvenile and adult individuals at the very least. Such information may be useful in contexts where a faecal deposit is found but the individual is not observed, such as in field-based projects of animals that are hard or impossible to capture. At the same time, this method can provide interesting insights into biological aging when applied to wild animals of known chronological age or to individuals sampled longitudinally (De Paoli-Iseppi et al., 2017; Powell & Proulx, 2003; Brivio et al., 2015). Considering the much greater variability in epigenetic aging rates we observed in wild compared to laboratory animals, our results suggest wild systems may provide a better environment in which to study drivers of epigenetic age acceleration. In the present study we were not able to identify drivers of epigenetic aging rates in the wild. However, given the wide

variation in epigenetic aging rates observed, our sample size may not have provided sufficient power to do this, and future studies with this goal would need to be larger.

We used faeces as a source of host DNA and demonstrated that an epigenetic clock can be built with faecal samples at a similar or even improved accuracy to a previously published blood-based mouse epigenetic clock (Han et al., 2018). As faecal samples can be collected non-invasively, a given individual can be sampled over time without limitations to sampling frequency. Further, it may be possible to collect faecal samples without capturing the animal, thus the method may be particularly useful when estimating age in wild species hard to capture or even detect.

In summary, our data indicate the potential to use a non-invasive, DNA methylation-based epigenetic clock built with samples from laboratory mice to estimate the age of entirely wild mice. This approach does not provide highly precise estimates of chronological age, but may capture signals of biological aging in longitudinal studies, making it a promising tool for studies of ontogeny and senescence in wild settings.

Funding

This work was funded by The Osk. Huttunen Foundation studentship and the National Geographic Society (Early Career grant reference No. EC-58520R-19) to EH, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851550), a NERC fellowship (NE/L011867/1) to SCLK, the British Ecological Society (BES) to TJL.

Acknowledgements

We thank Giselle Eagle and Richard Brown, the wardens of Skokholm Island, the Friends of Skokholm and Skomer, the Wildlife Trust of South and West Wales and field assistants for their help in enabling the wild mouse data collection.

Author information

Affiliations

[1]University of Oxford

[2]King's College London

[3]University of Edinburgh

Contributions

EH and SCLK set up the wild mouse study system. EH, SJ, and AR collected the samples. AO and TJL developed the software Apollo. EH conducted the laboratory work and analysed the data with support from TJL, SCLK and AR. EH wrote the manuscript with contributions from all authors.

Corresponding authors

Tom Little (tom.little@ed.ac.uk), Sarah Knowles (sarah.knowles@biology.ox.ac.uk)

References

Ambatipudi S, Horvath S, Perrier F, et al. DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur J Cancer*. 2017;75:299-307. doi:10.1016/j.ejca.2017.01.014

Anderson JA, Johnston RA, Lea AJ, et al. High social status males experience accelerated epigenetic aging in wild baboons. *Elife*. 2021;10:e66128. Published 2021 Apr 6. doi:10.7554/eLife.66128

Berry RJ, Jakobson ME. Life and death in an island population of the house mouse. *Exp Gerontol.* 1971;6: 187–197.

Bors EK, Baker CS, Wade PR, et al. An epigenetic clock to estimate the age of living beluga whales. *Evol Appl*. 2021;14(5):1263-1273. Published 2021 Feb 3. doi:10.1111/eva.13195

Brivio F, Grignolio S, Sica N, Cerise S, Bassano B. Assessing the Impact of Capture on Wild Animals: The Case Study of Chemical Immobilisation on Alpine Ibex. *PLoS One*. 2015;10(6):e0130957. Published 2015 Jun 25. doi:10.1371/journal.pone.0130957

Camerota M, Graw S, Everson TM, et al. Prenatal risk factors and neonatal DNA methylation in very preterm infants. *Clin Epigenetics*. 2021;13(1):171. Published 2021 Sep 10. doi:10.1186/s13148-021-01164-9

Cao X, Li W, Wang T, et al. Accelerated biological aging in COVID-19 patients. *Nat Commun*. 2022;13(1):2135. Published 2022 Apr 19. doi:10.1038/s41467-022-29801-8

De Paoli-Iseppi R, Deagle BE, McMahon CR, Hindell MA, Dickinson JL, Jarman SN. Measuring Animal Age with DNA Methylation: From Humans to Wild Animals. *Front Genet*. 2017;8:106. Published 2017 Aug 17. doi:10.3389/fgene.2017.00106

Ferrari, Manuela; Lindholm, Anna K; König, Barbara (2015). *The risk of exploitation during communal nursing in house mice, Mus musculus domesticus.* Animal Behaviour, 110:133-143.

Flurkey, Currer, and Harrison, 2007. 'The mouse in biomedical research.' in James G. Fox (ed.), American College of Laboratory Animal Medicine series (Elsevier, AP: Amsterdam; Boston).

Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22.

Gerber N, Auclair Y, König B, Lindholm AK. Population density and temperature influence the return on maternal investment in wild house mice. *Front Ecol Evol.* 2021;8:602359. https://doi.org/10.3389/fevo.2020.602359

Gray MM, Parmenter MD, Hogan CA, et al. Genetics of Rapid and Extreme Size Evolution in Island Mice. *Genetics*. 2015;201(1):213-228. doi:10.1534/genetics.115.177790

Han Y, Eipel M, Franzen J, et al. Epigenetic age-predictor for mice based on three CpG sites. *Elife*. 2018;7:e37462. Published 2018 Aug 24. doi:10.7554/eLife.37462

Harvanek ZM, Fogelman N, Xu K, Sinha R. Psychological and biological resilience modulates the effects of stress on epigenetic aging. *Transl Psychiatry*. 2021;11(1):601. Published 2021 Nov 27. doi:10.1038/s41398-021-01735-7

Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008;105(44):17046-17049. doi:10.1073/pnas.0806560105

Jarman SN, Polanowski AM, Faux CE, et al. Molecular biomarkers for chronological age in animal ecology. *Mol Ecol*. 2015;24(19):4826-4847. doi:10.1111/mec.13357

JAX®. Body weight information for aged C57BL/6J. Available at: https://www.jax.org/jax-mice-and-services/strain-data-sheet-pages/body-weight-chart-aged-b6 (Accessed: 30.11.2022)

JAX®. Body weight information for C57BL/6J. Available at: https://www.jax.org/jax-mice-and-services/strain-data-sheet-pages/body-weight-chart-000664 (Accessed: 30.11.2022)

Joubert BR, Felix JF, Yousefi P, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet*. 2016;98(4):680-696. doi:10.1016/j.ajhg.2016.02.019

Joyce BT, Gao T, Zheng Y, et al. Epigenetic Age Acceleration Reflects Long-Term Cardiovascular Health. *Circ Res*. 2021;129(8):770-781. doi:10.1161/CIRCRESAHA.121.318965

Kerepesi C, Meer MV, Ablaeva J, et al. Epigenetic aging of the demographically non-aging naked mole-rat. *Nat Commun*. 2022;13(1):355. Published 2022 Jan 17. doi:10.1038/s41467-022-27959-9

Kertes DA, Kamin HS, Hughes DA, Rodney NC, Bhatt S, Mulligan CJ. Prenatal Maternal Stress Predicts Methylation of Genes Regulating the Hypothalamic-Pituitary-Adrenocortical System in Mothers and Newborns in the Democratic Republic of Congo. *Child Dev*. 2016;87(1):61-72. doi:10.1111/cdev.12487

Lan X, Cretney EC, Kropp J, et al. Maternal Diet during Pregnancy Induces Gene Expression and DNA Methylation Changes in Fetal Tissues in Sheep. *Front Genet*. 2013;4:49. Published 2013 Apr 5. doi:10.3389/fgene.2013.00049

Larison B, Pinho GM, Haghani A, et al. Epigenetic models developed for plains zebras predict age in domestic horses and endangered equids. *Commun Biol*. 2021;4(1):1412. Published 2021 Dec 17. doi:10.1038/s42003-021-02935-z

Lemaître JF, Rey B, Gaillard JM, et al. DNA methylation as a tool to explore ageing in wild roe deer populations. *Mol Ecol Resour*. 2022;22(3):1002-1015. doi:10.1111/1755-0998.13533

Mayne B, Mustin W, Baboolal V, et al. Age prediction of green turtles with an epigenetic clock. *Mol Ecol Resour*. 2022;22(6):2275-2284. doi:10.1111/1755-0998.13621

Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013;38(1):23-38. doi:10.1038/npp.2012.112

Morales Berstein F, McCartney DL, Lu AT, et al. Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. *Elife*. 2022;11:e75374. Published 2022 Mar 29. doi:10.7554/eLife.75374

Parrott BB, Bowden JA, Kohno S, et al. Influence of tissue, age, and environmental quality on DNA methylation in Alligator mississippiensis. *Reproduction*. 2014;147(4):503-513. Published 2014 Mar 2. doi:10.1530/REP-13-0498

Peng C, Cardenas A, Rifas-Shiman SL, et al. Epigenetic age acceleration is associated with allergy and asthma in children in Project Viva. *J Allergy Clin Immunol*. 2019;143(6):2263-2270.e14. doi:10.1016/j.jaci.2019.01.034

Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab*. 2017;25(4):954-960.e6. doi:10.1016/j.cmet.2017.03.016

Pinho GM, Martin JGA, Farrell C, et al. Hibernation slows epigenetic ageing in yellow-bellied marmots. *Nat Ecol Evol*. 2022;6(4):418-426. doi:10.1038/s41559-022-01679-1

Polanowski AM, Robbins J, Chandler D, Jarman SN. Epigenetic estimation of age in humpback whales. *Mol Ecol Resour*. 2014;14(5):976-987. doi:10.1111/1755-0998.12247

Powell RA, Proulx G. Trapping and marking terrestrial mammals for research: integrating ethics, performance criteria, techniques, and common sense. *ILAR J*. 2003;44(4):259-276. doi:10.1093/ilar.44.4.259

Prado NA, Brown JL, Zoller JA, et al. Epigenetic clock and methylation studies in elephants. *Aging Cell*. 2021;20(7):e13414. doi:10.1111/acel.13414

Richetto J, Massart R, Weber-Stadlbauer U, Szyf M, Riva MA, Meyer U. Genome-wide DNA Methylation Changes in a Mouse Model of Infection-Mediated Neurodevelopmental Disorders. *Biol Psychiatry*. 2017;81(3):265-276. doi:10.1016/j.biopsych.2016.08.010

Robeck TR, Fei Z, Lu AT, et al. Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun Biol*. 2021;4(1):642. Published 2021 May 31. doi:10.1038/s42003-021-02179-x

Spangenberg E, Wallenbeck A, Eklöf AC, Carlstedt-Duke J, Tjäder S. Housing breeding mice in three different IVC systems: maternal performance and pup development. *Lab Anim*. 2014;48(3):193-206. doi:10.1177/0023677214531569

Stubbs TM, Bonder MJ, Stark AK, et al. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol*. 2017;18(1):68. Published 2017 Apr 11. doi:10.1186/s13059-017-1203-5

Sullivan IR, Adams DM, Greville LJS, Faure PA, Wilkinson GS. Big brown bats experience slower epigenetic ageing during hibernation. *Proc Biol Sci*. 2022;289(1980):20220635. doi:10.1098/rspb.2022.0635

Thompson MJ, vonHoldt B, Horvath S, Pellegrini M. An epigenetic aging clock for dogs and wolves. *Aging (Albany NY)*. 2017;9(3):1055-1068. doi:10.18632/aging.101211

Tobi EW, Lumey LH, Talens RP, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet*. 2009;18(21):4046-4053. doi:10.1093/hmg/ddp353

Vangeel EB, Pishva E, Hompes T, et al. Newborn genome-wide DNA methylation in association with pregnancy anxiety reveals a potential role for *GABBR1*. *Clin Epigenetics*. 2017;9:107. Published 2017 Oct 3. doi:10.1186/s13148-017-0408-5

Viitaniemi HM, Verhagen I, Visser ME, Honkela A, van Oers K, Husby A. Seasonal Variation in Genome-Wide DNA Methylation Patterns and the Onset of Seasonal Timing of Reproduction in Great Tits. *Genome Biol Evol*. 2019;11(3):970-983. doi:10.1093/gbe/evz044

Wright PGR, Mathews F, Schofield H, et al. Application of a novel molecular method to age free-living wild Bechstein's bats. *Mol Ecol Resour*. 2018;18(6):1374-1380. doi:10.1111/1755-0998.12925

Zocher S, Overall RW, Lesche M, Dahl A, Kempermann G. Environmental enrichment preserves a young DNA methylation landscape in the aged mouse hippocampus. *Nat Commun*. 2021;12(1):3892. Published 2021 Jun 23. doi:10.1038/s41467-021-23993-1

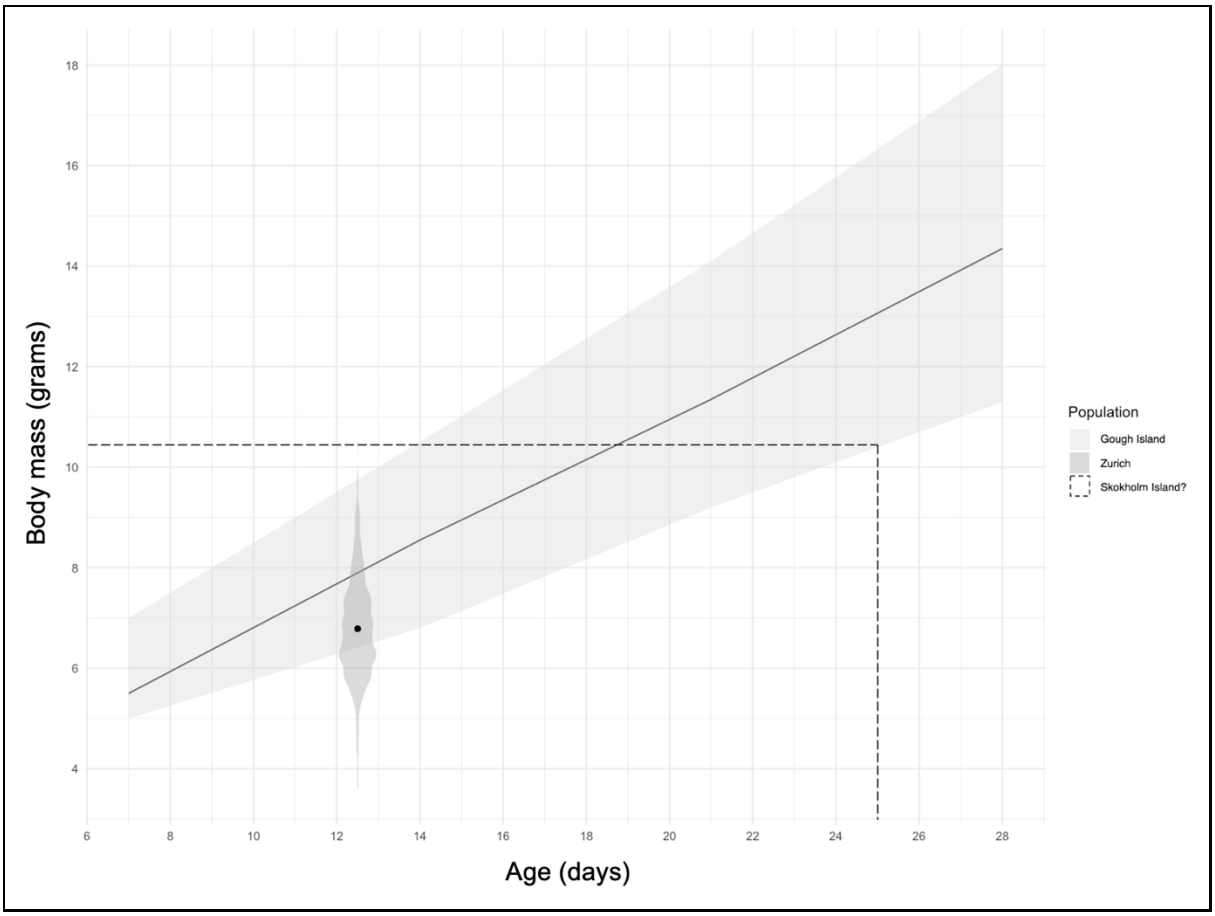
Supplementary material

Supplementary Table 1. Primers used to target five *Mus musculus* genes (*Hsf4, Gm9312, Fm7325, Kcns1, Prima1*), generated based on Han et al (2018).

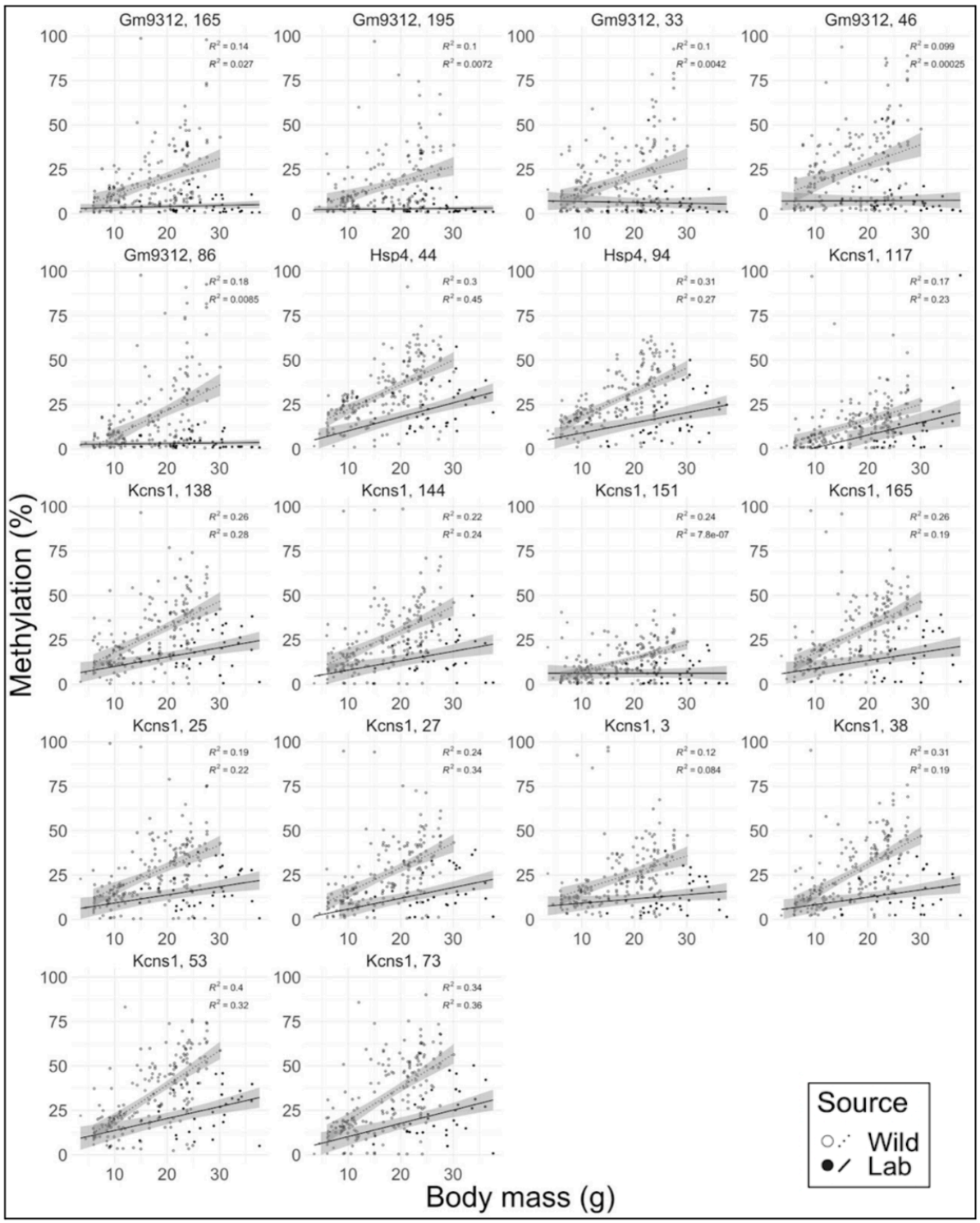
Gene	Primer sequence
Hsf4	GTGAGTAGTAAGGTGGGATAAATTGTAGAAAAAATG (Forward), TCCCTACTCTCCTACACTCCTCTCAAAACTTA (Reverse)
Gm9312	TTGTTTTGGGGTATTAGAAATTTTTTT (Forward), CCTAACCATACTAAACCAAATCTCTATATCTAAAT (Reverse)
Gm7325	TGTTGGTTGAGGATAAAGAGTAGATAGTTTAGTAGAGT (Forward), TTCCCTTTACAAATACAAATCCTACCATA (Reverse)
Kcns1	GGTTGAGAGGGTGGTAGAAGAAGTTG (Forward), ACTCCCCTCCATCCCTACCATATACATCCA (Reverse)
Prima1	TTGTGTTTAATTAGGAGAGGTAAATTATGAATTAGGTTTATA (Forward), CAAAATTAATTACACCAACTTATAACCTACTATTC (Reverse)

Supplementary Table 2. Elastic net regression used to build an epigenetic clock identified 18 CpG sites (in red) from three of five targeted genes (*Hsf4, Gm9312, Fm7325, Kcns1, Prima1*): two from *Hsf4*, five from *Gm9312*, and eleven from *Kcns1*.

Gene	Sequence
Hsf4	GGAAGGTATTAATGTTGGTATTTTTGGTTTTGTTTATGTGTTTCGGATGGTGTTTTTTGTTTGTAGGTATTTGCGTTGCGAGGCGATGATAGTCGATGGCGTTCGGAAGATTTGAGTCGATTGTTGGGAGAGGTG
Gm9312	AGGTGTGGGCGTAGTCGGAGGGTATTGGGTATCGGGTATTAAGCGCGGAAGTTTATTAGGTGTTTAGGGCGTAGCGCGATTTTCGCGATTTTAGTTTTTCGTTCGTTGCGGGTTACGTTATCGTTTATTTTTTCGATGGTCGTCGCGGTGTTTCGGTATTGGGTCGAGCGCGTGGTGAAATTAGAGGTCGTGGGCGTTTTGTAGTTTTTAAGGGTTTCGGTTACGA
Kcns1	CGCGTGTTGGGAGTTAGTAGTAGGCGCGACGATATTTCGAAGTTGAATTAAGCGATGTAGAAGTATTTTAGGCGGCGTAGTATCGGGTCGTCGCGTATTTTTTTTGCGTTGCGATTCGCGGTTATTGTAGTTATCGTCGTCGTCGTTTTTCGAGTTTGGTATTTCGGTAGGTTG



Supplementary Figure 1. Relationship between body mass and chronological age during early life in two wild house mouse populations. The black solid line shows the mean body mass for wild-derived (Gough Island) mice born in laboratory of a given age with shading indicating lower and upper limits (data reproduced from Figure 3 in Gray et al., 2015). The violin plot indicates body mass distribution in 12–13-day old mice (mean 12.8 days, median 13.0 days, *n*=438) from Zurich, Switzerland (Gerber et al., 2021). The point indicates average body mass in Zurich mice (6.8g; ranges 3.6–10.5g, median 6.7g). Dashed lines indicate the upper limit of body mass (10.5g) for a set of juvenile Skokholm Island mice for which epigenetic age was estimated. According to age-mass relationships in Zurich and Gough, these Skokholm Island mice are estimated to be under 25 days old.



Supplementary Figure 2. Methylation rate at 18 CpG sites included in the epigenetic clock across body mass in laboratory and wild mice. Gene and CpG start position are indicated as panel titles. R^2-values are presented for each CpG site for wild (*top*) and lab (*bottom*) mice.

CHAPTER III

Conserved gut microbiota assembly patterns across laboratory and wild mice

Authors

Eveliina Hanski[1], Aura Raulo[1,2], Sarah C L Knowles[1]

Title

Conserved gut microbiota assembly patterns across laboratory and wild mice

Abstract

The gut microbiota is a dynamic ecosystem that varies between individuals and populations but also within an individual over time. Changes in the gut microbiota have been linked to various age-related processes such as growth, immune maturation, and reproduction, making age an intriguing variable in the investigation of ecological dynamics of the microbiota. Studies investigating how the gut microbiota develops in early life and changes with age have been limited to either laboratory or wild animals. However, we have little understanding of how age-related dynamics compare across these two contexts, where many key factors will vary, including environmental exposures, social contacts, diet, and even the microbial taxa that colonise the gut. Importantly, it remains unclear how representative laboratory mouse models can be of ecological processes in the gut microbiota, considering their microbiota is markedly different from wild mice. Here, we used a standardised pipeline to directly compare early life gut microbiota assembly patterns in laboratory-reared and wild house mice (*Mus musculus domesticus*). We found that despite having contrasting genetic and environmental backgrounds and compositionally distinct gut microbiotas, lab and wild mice share several common features of gut microbiota assembly. Specifically, both alpha and beta diversity as well as changes in the relative abundance of predominant phyla – Firmicutes, Bacteroidota, and Proteobacteria – followed remarkably similar, and similarly timed, temporal patterns in lab and wild mice, independent of the different lower-level taxa present in these systems. These results support

the notion that early life gut microbiota dynamics observed in laboratory settings are representative of the ecological processes at play in the wild. Our results indicate some degree of intrinsic host programme in gut microbiota assembly that transcends very different taxonomic profiles as well as genetic and environmental backgrounds of the host, and suggest age-related microbiota changes consider the emergent community properties of the microbiota. Thus, despite their artificial environment, lab models can provide meaningful insights into these dynamics.

Introduction

The mammalian gut is densely inhabited by a diverse collection of microbial organisms, collectively referred to as the gut microbiota. The gut microbiota is an important factor for host biology with conserved roles in host processes ranging from metabolism to behaviour (Visconti et al., 2019; Turnbaugh et al., 2006; Liberti et al., 2022). Composition and diversity of this miniature ecosystem varies not only between individuals, populations, and species but also within an individual over time (Rudolph et al., 2022; Nishida & Ochman, 2018; Vandeputte et al., 2017). In particular, at the beginning of a host's life, the gut microbiota undergoes a period of maturation with changes in composition and diversity (Stewart et al., 2018; Yassour et al., 2016; Depner et al., 2020), guided by both intrinsic ecological processes (Gonze et al., 2018) as well as selective processes imposed by the physiology of the host, as it undergoes various developmental processes, such as immune maturation and the development of the central nervous system (Chung et al., 2012; Sharon et al., 2016). Perturbation of the gut microbiota during this critical time of development has been linked to developmental alterations (Olin et al., 2018; Darabi et al., 2019), indicating an active role for the gut microbiota in host development. Thus, a detailed understanding of the gut microbiota maturation process is essential for comprehensive understanding of host development.

Many human studies have provided valuable insights into maturation of the human gut microbiota in early life (Stewart et al., 2018; Niu et al., 2020; Bäckhead et al., 2015; Ferretti et al., 2018; Shao et al., 2019). During this maturation process, the gut is first colonised by microbes from the mother (through birth and breastfeeding) and the environment. After the initial colonisation, the gut microbiota undergoes compositional changes through which it becomes more diverse within an individual but more homogenous across individuals (Bäckhead et al., 2015; Ferretti et al., 2018).

Alteration of the general maturation trajectory can happen at various stages and may have long-lasting consequences (Olin et al., 2018; Darabi et al., 2019). Firstly, transmission of maternal microbes can be disrupted at birth leading to primary colonisation by generalist pathogens from the environment instead of specialist symbionts from the mother. For example, relative abundance of pathogens associated with hospital environment, e.g., *Klebsiella*, has been shown to be elevated in infants delivered via caesarean section (CS) (Bäckhead et al., 2015; Shao et al., 2019). Further, CS infants have lower levels of bacteria from the genus Bacteroides, a trait associated with lower gut microbiota diversity and slower maturation rate (Stewart et al., 2018). After the initial colonisation, the stereotypic gut microbiota maturation trajectory can be further altered by care practises, such as types of feeding (breastfeeding vs formula feeding) and timing of breastfeeding cessation (Bäckhead et al., 2015; Stewart et al., 2018; Shao et al., 2019). Particularly, breastfeeding cessation can accelerate gut microbiota maturation, marked by an increase in Firmicutes (Stewart et al., 2018).

Similar patterns of gut microbiota maturation have been observed in some, but perhaps not all, wild primates (Baniel et al., 2022; Petrullo et al., 2022; Reese et al., 2021). However, previous animal studies have focused on a single population, wild or laboratory, with little direct

comparison across populations. For laboratory model species, a key gap in understanding is whether the patterns and processes of microbiota maturation observed in laboratory settings are representative of those occurring in natural populations. In the wild, several factors such as genetic background, diet, environmental, and social exposures will show much greater variation, with potentially important implications for gut microbiota composition, diversity, and function. While the house mouse (*Mus musculus*) is the mammalian model species of choice in microbiota research, little is known about gut microbiota dynamics in this species outside of the lab. In particular, it remains unknown whether fundamental age-related dynamics in the mouse gut observed in the lab are paralleled in natural populations, despite markedly different gut microbiota compositions (Rosshart et al., 2019; Wang et al., 2014; Kreisinger et al., 2014; Wang et al., 2015).

Here, we directly compare early life gut microbiota assembly dynamics in laboratory and wild house mice, using a standardised 16S-pipeline. Given the proposed key role of gut microbiota in developmental processes that are expected to be conserved across contexts, we hypothesise that despite differentiation in community composition and diversity, some broad trends of gut microbiota assembly will be conserved between lab and wild settings. In particular, based on findings from previous studies (Bäckhead et al., 2015; Wang et al., 2020; Stewart et al., 2018; Baniel et al., 2022; Derrien et al., 2019), we hypothesise that gut microbiota diversity increases in early life within-individuals but decreases between individuals, and that the ratio between aerotolerant and anaerobic taxa change in favour of anaerobes.

Methods

Sample collection

Faecal samples were collected from 7–81-day old C57BL/6 laboratory mice (*Mus musculus, n*=40) in November 2020 and October 2021 at the Biomedical Services Building, Oxford, UK (the colony is originally from Jackson Laboratories but bred at the facility for the last 17 years). Mice were housed in a specific-pathogen-free (SPF) facility, where the following pathogenic taxa were excluded: *Helicobacter* species, *Pasteurella pneumotropica,* beta-haemolytic streptococci*, Streptococcus pneumoniae, Citrobacter rodentium, Corynebacterium kutscheri, Salmonella* species, and *Streptobacillus moniliformis*. Mice were not subject to any interventions before or during sample collection. Mice were placed on a sterile surface for faecal collection. Faecal pellets were preserved in Zymo DNA/RNA Shield and stored at -80°C until DNA extraction (up to 12 months).

Wild house mice (*Mus musculus domesticus*) were live-trapped on Skokholm Island (Wales, UK), in April–May 2019, July 2019, September–October 2019, August–September 2020, and April–May 2021. Trapping was carried out using small Sherman traps provisioned with peanuts, non-absorbent cotton wool for bedding, and with a spray of sesame oil outside the trap as a lure. Trapping was conducted in two broad sampling sites on the island ('Observatory' and 'Quarry'). On each trapping night, 150 traps were set at dusk and checked at dawn at one of the sampling sites. Used traps (where signs of a visiting animal were detected, whether captured or not) were washed and sterilised with 20% bleach solution before re-use. Upon first capture, mice were either tagged with a subcutaneous passive integrated transponder (PIT) tag for permanent identification. All captures were therefore individually identified, sexed, and weighed to the nearest 0.1g before release within 2m of their trapping point. Sex was determined based on visual inspection, anogenital distance, and reproductive state. Faecal

pellets were collected from traps, and preserved in DNA/RNA Shield. Samples were stored at -20°C until return to the laboratory (within 6 weeks), after which they were stored at -80°C until DNA extraction (up to 17 months).

DNA extraction, library preparation, and sequencing

DNA extraction was performed using ZymoBIOMICS DNA MiniPrep Kit (spin-column format) according to manufacturer's instructions (Zymo Research, Irvine, California, USA). Samples were randomised into 53 extraction batches of up to 23 samples. A negative extraction control (DNAse-free H_2O) was included in each extraction batch at a varying position with the exception of one extraction batch, in which a negative control was not included. The microbiota of each faecal sample was characterised using amplicon sequencing, targeting the V4–V5 region of the 16S rRNA gene using the 515F–926R primers (Parada et al., 2016; Walters et al., 2015). Samples were amplified and sequenced in 16 sequencing runs of up to 95 samples using the Illumina MiSeq platform (Reagent kit v3, 2x300 bp chemistry). Library preparation and sequencing was conducted at the Integrated Microbiome Resource, Dalhousie University, using the protocol described in Comeau et al. (2017). Each sequencing run included a negative control for PCR reaction and a separate negative control for sequencing. Additionally, all extraction controls (n=52) were sequenced amongst the true samples (extraction controls were randomised into batches of true samples).

Data selection

The lab mouse dataset was restricted to mice under 91 days of age in order to focus on the early life phase when individuals are still developing (lab mice are considered 'adult' from 3 months onwards; Flurkey et al., 2007). The wild mouse samples analysed here were processed and sequenced as part of a larger sample set. While it is impossible to precisely age wild mice, in

order to focus on the early life stage we excluded from analyses any captures we could confidently predict to be over 90 days old (those recaptured a minimum 76 days after first being captured, as trappable mice are likely to be at least 14 days old). Further, only samples from September and October (from the years 2019 and 2020) were included to minimise seasonal effects on the gut microbiota (Maurice et al., 2015; Marsh et al., 2022).

Further, female wild mice with signs of ongoing or recent pregnancy (bulging central body, enlarged/prominent nipples) as well as male wild mice with large or extra-large testes (testes were classified as abdominal, small, large or extra-large based on appearance and size), as these had significant effects on body mass, which was used as a proxy of age. We further excluded captures with a missing body mass measurement. Only mice under 25 grams were included with the aim of excluding pregnant females that may not have been recognised as such. Applying these criteria resulted in a set of 315 wild mouse samples from 199 individual mice. Body mass ranged from 5.4 to 24.9 grams (mean 16.8, median 17.2) among these captures.

Data processing

Data processing and subsequent analysis was done using R version 4.1.2 (R Core Team, 2021). Demultiplexed sequencing reads were processed through the DADA2 pipeline (Callahan et al., 2016) to infer amplicon sequence variants (ASVs) and assign taxonomy using the SILVA rRNA database version 138. A phylogenetic tree was built using R packages DECIPHER and phangorn. R package iNEXT (Hsieh et al., 2016) was used to generate sample completeness and rarefaction curves and the read depth threshold was set where these curves plateaued, at 4,000. Samples with a lower read depth were excluded from further analysis. Data were not rarefied. iNEXT was further used to measure asymptotic ASV richness and Shannon diversity. Singleton, doubleton, and non-gut microbial ASVs (mitochondria, chloroplasts) were removed

separately for lab and wild mouse datasets (sample selection for these datasets is described below). ASV counts were normalised to proportional abundance using R package phyloseq (McMurdie et al., 2013). Aerotolerance and sporulation capacity of bacterial genera were extracted from Bergey's Manual of Systematics of Archaea and Bacteria and additional references where information was not available in the manual (Supplementary Table 1). Bacteria were categorised into either aerotolerants or obligate anaerobes; the latter only when explicitly listed *obligate* anaerobes. Aerotolerance and spore-forming ability were determined based on genus-level taxonomy, unless information was not available, in which case family-level information was inspected and used if all genera were stated to have same aerotolerance and/or sporulation ability.

A total of 41 unique ASVs were detected in negative controls for DNA extraction (n=52) and library preparation (n=14) with a maximum read count of 363 for any given control, with the exception of one library preparation control. This control had 255 unique ASVs and 29,300 reads, while mean read count of biological samples was 29,235 (range 5,662–232,225, median 26,097). All extraction controls (n=5) from the same 96-well plate had <20 reads each, indicating the entire plate was not contaminated during library preparation (extraction and library preparation negative controls were included on the plate in a randomised order, thus several rows and columns had negative controls). Instead, the most likely explanation is that a biological sample was mistakenly pipetted into the control well in addition to its designated well. As all extraction controls on this plate were considered negative, we retained this plate of samples in our analyses.

We tested for the presence of any potential contaminants using the R package decontam using the 'prevalence' (presence/absence) method, in which each sequence in biological samples is

compared to the prevalence in negative controls (excluding the library preparation control that was contaminated with a biological sample, see above). A sequence was considered a contaminant if it reached a probability of 0.1 in the Fisher's exact test used in decontam. 31 contaminants were identified with this method, and filtered from the data before subsequent analysis. These included 10 ASVs that belonged to Muribaculaceae (unknown genus for all; other contaminants belonged to 14 families with 1–3 ASVs per family).

Functional profiles

MetaCyc pathways were predicted from the 16S rRNA data using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2) version 2.5.0 pipeline (picrust2_pipeline.py) using default options (Douglas et al., 2020). Read counts of functional pathways were normalised to proportional abundance. Analysis of functional profiles was restricted to pathway-level due to discrepancies between functional profiles at category level depending on which database was used to assign functional categories.

Analysis

Differences in alpha diversity across lab and wild mice were tested with Wilcoxon rank sum tests with 1,000 permutations. Principal coordinate analysis was used to investigate (1) compositional difference of lab and wild mouse microbiotas and (2) relationship between age and gut microbiota composition in lab and wild mice separately using four distances: Jaccard, Aitchison, unweighted UniFrac and weighted UniFrac. For Aitchison distance, a centered log-ratio (clr) transformation was performed before ordination using the R package microbiome (Lahti & Shetty, 2017), where zero relative abundances were replaced with a pseudocount (min(relative abundance/2)).

Beta diversity was measured with (1) marginal permutational multivariate analysis of variance (PERMANOVA) from the R package vegan (Oksanen et al., 2017) and (2) pairwise dissimilarity among samples, using R package microbiome (Lahti & Shetty, 2017). Quadratic plateau models from the R package easynls (Arnhold, 2017) were used to assess the relationship between age/body mass and (1) alpha diversity, (2) first or second axis from principal coordinate analyses, or (3) distance to reference early life microbiota. Reference early life microbiota was estimated separately for laboratory and wild mice by taking the mean abundance in mice under 14 days of age (lab mice) or under 10 grams of body mass (wild mice).

Statistical differences in pairwise dissimilarity of different age groups were tested with Wilcoxon rank sum tests with 1,000 permutations. Taxonomic and phenotypic abundance of bacteria were modelled and plotted using locally weighted scatterplot smoothing (LOESS) regression from the R package ggplot2 (Wickham, 2016). NCBI Nucleotide Basic Local Alignment Search Tool (Nucleotide BLAST) was used to BLAST a single ASV that had been assigned taxonomy at family (Enterobacteriaceae) but not at genus level using SILVA database (Suppl. Table 2, Suppl. Fig. 6B).

Ethical statement

Wild mouse data collection was done under Home Office license PPL PB0178858 held at the University of Oxford, and with a research permit from the Islands Conservation Advisory Committee (ICAC), and Natural Resources Wales. Mice were not subject to intervention as part of this study.

Results and discussion

To investigate whether trends of gut microbiota assembly are conserved between lab and wild settings despite differentiation in community composition and diversity, we compared early life gut microbiota dynamics in the widely used laboratory mouse strain C57BL/6 to that of wild individuals of the same species (*Mus musculus*), from a single well-studied population, living a feral existence on the Skokholm Island in Wales, UK. These wild mice are outbred and exposed to highly variable environmental conditions in contrast to inbred lab mice living under highly stable environmental conditions. We used 16S rRNA amplicon sequencing to characterise gut microbiota composition and predicted function from 355 faecal samples from 40 laboratory mice (one sample per individual) and 199 wild mice (315 samples; 1–8 samples per individual). After removal of singletons and doubletons, a total of 7,649 amplicon sequence variants (ASVs) were recovered. Taxonomic assignment using SILVA rRNA gene database was successful for 6,481 ASVs (84.7%) at family level and 3,935 ASVs (51.4%) at genus level. These rates were comparable across the two systems, indicating lab and wild mice have similar numbers of known bacterial genera (90.0% and 84.2% ASVs assigned to family, 50.3% and 51.7% assigned to genus, for lab and wild mice respectively).

Laboratory and wild mice have distinct gut microbiotas

First, using a randomly selected cross-sectional subset of the data, we demonstrate that laboratory and wild mice have taxonomically distinct gut microbiotas, as previously reported (Rosshart et al., 2019; Wang et al., 2014; Kreisinger et al., 2014; Wang et al., 2015). Wild mice had higher alpha diversity in terms of both estimated ASV richness and Shannon diversity (permutational Wilcoxon rank sum tests, $p<0.001$ for both, Fig. 1A–B). Only 7.8% of ASVs (484 out of 6,235 in the dataset) were detected in both lab and wild mice; however, these comprised on average 42.7% of wild mouse microbiota and 48.2% of lab mouse microbiota

(standard deviation 10.5% and 22.5%, respectively), indicating microbes shared between lab and wild mice form a large proportion of gut microbiota in both systems.

Still, as observed in previous studies (Rosshart et al., 2019; Kreisinger et al., 2014), gut microbiota composition significantly differed between lab and wild mice, with samples clustering based on source using both non-phylogenetic (Jaccard, Aitchison) and phylogenetic (weighted and unweighted UniFrac) community distance metrics and on various levels of taxonomic resolution (Fig. 1C, Suppl. Fig. 1). Despite this, source (lab/wild) explained a very small amount of compositional variation in the dataset overall (0.9% in Aitchison dissimilarity, PERMANOVA, p=0.001; beta dispersion, F=0.0329, p=0.855), lending to the high variability of especially the wild mouse samples (Fig. 1C–D).

Lab and wild mouse gut microbiotas resembled each other at phylum level with Firmicutes and Bacteroidota forming 92.9–94.0% of relative abundance in both settings at approximately 1:1 ratio (Fig. 1E). However, compositional similarity decreased at lower taxonomic levels, such as family and genus (Fig. 1F–G), in line with previous work (Rosshart et al., 2019). It is plausible that some of these differences were driven by differential age distribution across the two systems. In particular, the lab mouse dataset included samples from pre-weaned individuals with strikingly different gut microbiota from weaned individuals (Suppl. Fig. 7A), while the wild mice were presumably weaned at the time of capture as they were independent and entering live traps (precise age and weaning status unknown). However, samples continued to cluster by source when excluding pre-weaned lab mice (mice <22 days of age, Suppl. Fig. 2). Together, these data show that while a large proportion of lab and wild mouse gut microbiotas is formed by taxa found in both systems, compositionally these mice harbour

distinct gut microbiotas. Further, wild mice have higher gut microbial diversity, as expected from their greater environmental exposure and genetic variance.

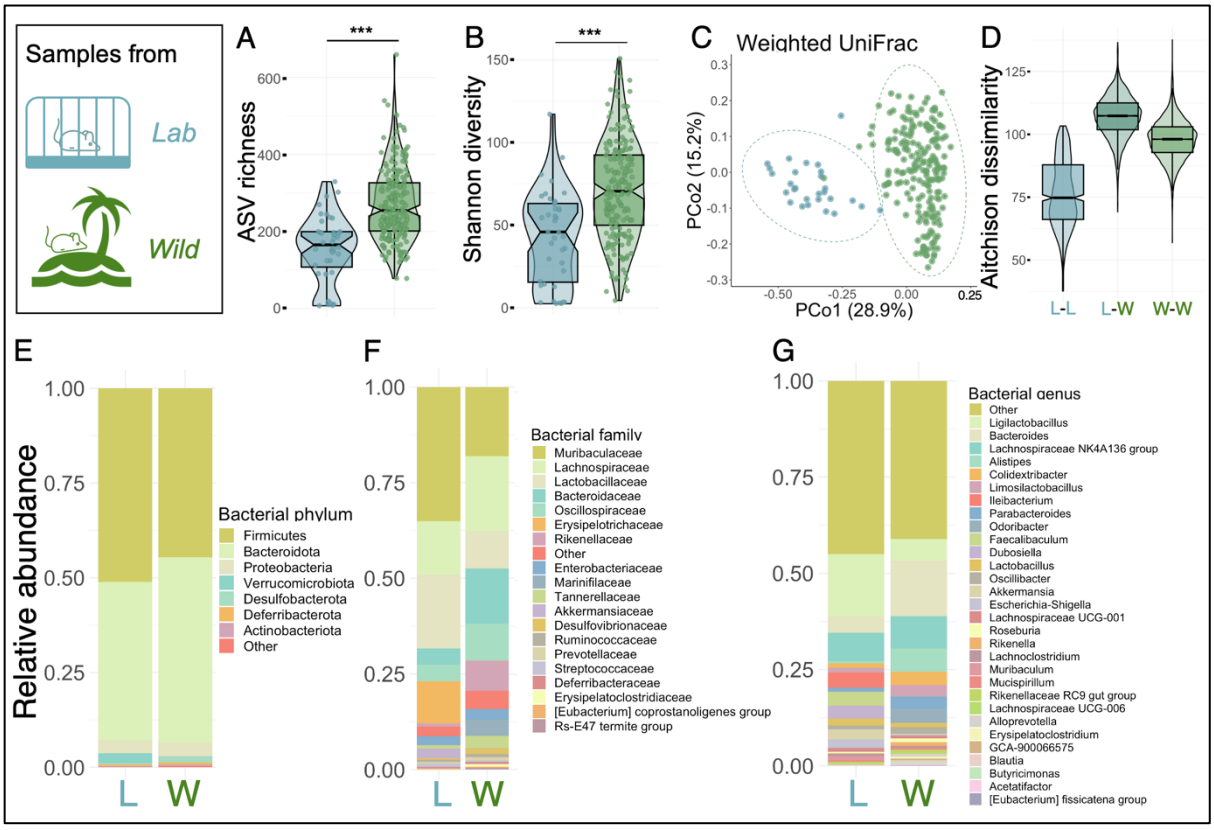


Figure 1. Comparison of laboratory and wild mouse gut microbiotas using 239 faecal samples from 40 laboratory and 199 wild mice (1 sample per mouse). (**A–B**) Alpha diversity; (**A**) asymptotic ASV richness and (**B**) asymptotic Shannon diversity. Statistical differences between lab and wild mice were tested with permutational Wilcoxon rank sum tests (***; p<0.001). (**C**) Principal coordinates analysis (PCoA) on weighted UniFrac distance. (**D**) Pairwise comparison on gut microbial dissimilarity on Aitchison distance in lab-lab (L-L), lab-wild (L-W) and wild-wild (W-W) sample pairs. (**E–G**) Mean relative abundance of bacterial (**E**) phyla, (**F**) families and (**G**) genera in laboratory (L) and wild (W) mice. Colour in figures A–D indicates sample source: *blue* = laboratory, *green* = wild, *teal* = lab/wild.

Conserved taxon-independent gut microbiota assembly patterns

Considering an increase in gut microbial diversity in early life has been detected in various host species (Bäckhead et al., 2015; Baniel et al., 2022; Frese et al., 2015; Janiak et al., 2021), we predicted an increase in alpha diversity during early life for both laboratory and wild mice. Since wild-caught mice cannot be precisely aged, we used body mass as a proxy as it has

previously been shown to predict age relatively accurately in wild mice, especially among very young individuals (Gerber et al., 2021; Gray et al., 2015; Ferrari et al., 2015).

In the wild mice studied here, body mass ranged from 5.4g to 24.9g. Based on previously published age-mass relationships in wild or wild-derived mice (Gray et al., 2015; Ferrari et al., 2015), the youngest wild mice in this data are estimated to be ≤13 days old. The age of laboratory mice in the present study ranged from 7 to 81 days. Body mass is associated with age also in laboratory mice (Jax, 2022a; Jax, 2022b; Spangenberg et al., 2014); however, lab mice are fed *ad libitum* and housed in an environment that is free of several common pathogens while wild mice are likely to undergo food shortages and infections with possible effects on body mass, thus age-body mass relationship varies between lab and wild mice (Jax, 2022a; Jax, 2022b; Spangenberg et al., 2014; Ferrari et al., 2015; Gray et al., 2015). As such, use of body mass as a proxy of age in lab mice would not inherently make the lab and wild mouse data more comparable. Hence, age-related dynamics in lab mice were investigated using actual age rather than body mass.

Age/body mass was strongly associated with alpha diversity in both lab and wild mice, with ASV richness increasing to a plateau by adulthood in both systems, similar to findings in Baniel et al. (2022). In lab mice, a noticeable plateau was detected at approximately 30 days of age (Fig. 2A). The plateau in ASV richness was more subtle in wild mice, occurring at around 20 grams of body mass (Fig. 2B), corresponding to an estimated ≥23 days of age (Ferrari et al., 2015). We then asked whether age is associated with gut microbial composition in addition to alpha diversity. Microbiota variability among individuals (the average Jaccard dissimilarity within a given age or body mass class) decreased with age in both systems; however, the

decrease between first and last age class was only significant in wild mice (Wilcoxon rank sum test, lab, p=0.2035; wild, p<0.001, Fig. 2C).

An Aitchison dissimilarity ordination performed separately for lab and wild mice suggested age (body mass in wild mice) to be associated with the first or second axis of variation, explaining 22.1% and 3.4% of all gut microbial variation in lab and wild mice, respectively (PCoA; PCo1 for laboratory mice, PCo2 for wild mice, Fig. 2D–F; distance measure was selected based on which one (Jaccard/Aitchison) was most strongly associated with age in both lab and wild mice, Suppl. Fig. 3). Age was strongly associated with gut microbiota composition in lab mice up to 26 days of age (Fig. 2D), similar to that seen in alpha diversity (Fig. 2A). In wild mice, on the other hand, gut microbiota composition was associated with body mass up to around 23 grams (Fig. 2E), again closely in line with alpha diversity dynamics (Fig. 2B). The lower explanatory power of 'age' (body mass) in gut microbial variation of wild mice might be partly driven by the inaccuracy of body mass as an age proxy (for instance, heavier mice may not be older but could also be pregnant, and likewise lighter mice may have suffered from food shortage rather than be young), but it could also be that age signal on gut microbiota is weaker in the wild, where individuals are genetically diverse and exposed to variable environment.

We further investigated gut microbiota maturation by measuring compositional dissimilarity to a reference microbiota of juvenile mice (measured by taking the mean of all taxa abundances in lab mice <14 days of age (n=5), or wild mice ≤9g of body mass (n=16), which is the estimated equivalent of 14 days of age; Ferrari et al., 2015; Gerber et al., 2021; Gray et al., 2015). Dissimilarity to this token of early life increased with age, with a plateau reached at around 28 days of age in lab mice and around 14 grams in wild mice (corresponding to ≥23

days of age, Fig. 1F–G). These trends in diversity and composition suggest that gut microbiotas become richer within individuals but more homogenous across individuals with age in both lab and wild mice, consistent with findings in humans (Derrien et al., 2019) and gelada baboons (Baniel et al., 2022) but potentially not all mammals (Reese et al., 2021).

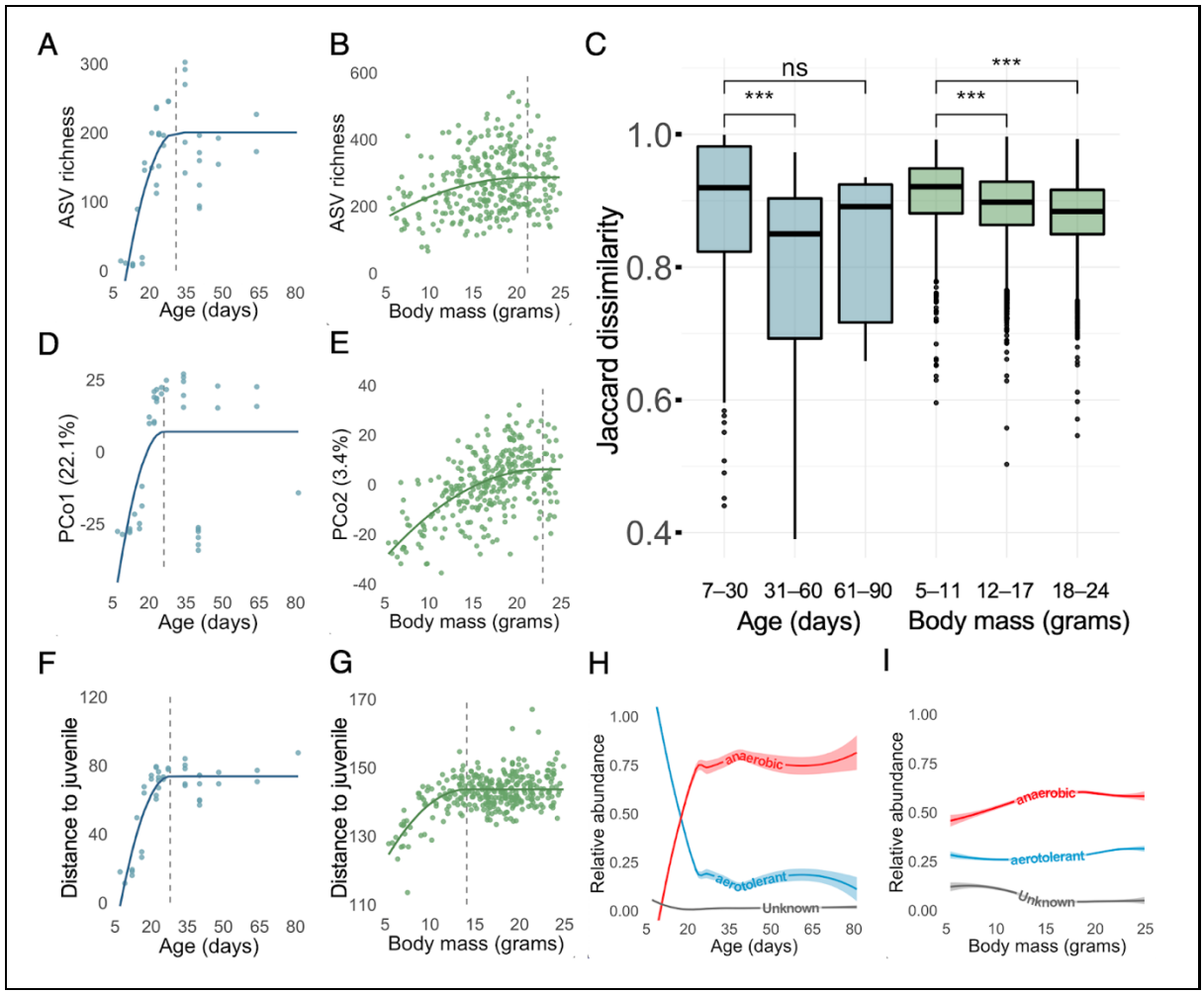


Figure 2. Taxon-independent trends in laboratory and wild mouse gut microbiota. (**A–B**) Quadratic plateau models on asymptotic ASV richness. Vertical lines indicate critical points of plateau. (**A**) Lab mice: R^2=0.515, critical point of inflexion=30.2d. (**B**) Wild mice: R^2=0.069, critical point of inflexion=21.2g. (**C**) Beta diversity (based on Jaccard dissimilarity) in lab and wild mice. Samples were assigned into three age/body mass classes at equal intervals. Significance was tested with permutational Wilcoxon rank sum test (***; p<0.001; ns, p=0.216). (**D–E**) Quadratic plateau models on the relationship between first (lab mice, **D**) or second (wild mice, **E**) principal coordinates (PCo) and age (body mass for wild mice). Ordination was performed on Aitchison dissimilarity. Y-axes are not comparable in D and E as datasets were ordinated separately. Vertical lines indicate critical points of inflexion. Lab mice: R^2=0.303, critical point of inflexion=26.0d. Wild mice: R^2=0.329, critical point of inflexion=22.9g. **(F–G)** Quadratic plateau models on the relationship between Aitchison dissimilarity to reference juvenile microbiota and age (body mass for wild mice). Reference juvenile microbiota was measured separately for laboratory and wild mice by taking the mean of all taxa abundances in mice <14 days of age (laboratory mice, n=5), or ≤9g of body mass (wild mice, n=16). Vertical lines indicate critical points of inflexion. Lab mice: R^2=0.763, critical point of inflexion=27.8d. Wild mice: R^2=0.301, critical point of inflexion=14.1g. (**H–I**) Relative abundance of aerotolerant and anaerobic bacteria across age/body mass in (**H**) lab and (**I**) wild mice respectively. Lines are locally estimated scatterplot smoothing (LOESS) lines with 95% confidence interval bands. *Anaerobic* = obligate anaerobes, *aerotolerant* = everything else with known aerotolerance, *unknown* = bacteria with unknown/mixed aerotolerance.

As the mammalian gut becomes increasingly anoxic in early life (Albenberg et al., 2014), we hypothesised that the ratio between aerotolerant and anaerobic bacteria would shift in early life, favouring anaerobes over aerotolerant bacteria, as shown in humans (Guittar et al., 2019; Bäckhead et al., 2015). In lab mice, we detected a clear pattern between age and bacterial aerotolerance consistent with this expectation. Over the first 30 days of life, aerotolerant bacteria initially dominated the gut microbiota but declined in relative abundance while anaerobic taxa increased, until they reached approximately 1:3 ratio with anaerobic taxa dominating (Fig. 2H). In wild mice, we similarly detected an increase in anaerobic taxa while the relative abundance of aerotolerant taxa did not change with age but retained a higher relative abundance than in lab mice (mean relative abundance of aerotolerant taxa 29.3% in wild mice >15g vs 17.4% in lab mice >20d; Fig. 2I). Further, taxa with unknown/mixed aerotolerance decreased, plateauing at around the same time with anaerobic taxa, at around 15 grams of body mass (Fig. 2G).

The less pronounced changes in the ratio of aerotolerant and anaerobic taxa might have been due to the likely absence of pre-weaned individuals in the wild mouse data (lab dataset included pre-weaned individuals). The higher relative abundance of genera for which aerotolerance could not be determined may also have contributed to these somewhat different patterns (mean 7.9% in wild mice vs 1.7% in lab mice). The higher prevalence of aerotolerant bacteria in wild mice could be driven by distinct microbial exposure patterns between lab and wild mice. Laboratory mice were housed in individually ventilated cages with limited and invariable exposure to environmental microbes, whereas wild mice are exposed to a large, changing pool of microbes in their environment (Raulo et al., 2021). As environmental microbes can be expected to tolerate oxygen on some level or form aerotolerant spores, wild mice are likely to be exposed to a higher number of aerotolerant and/or sporulating microbes.

There were no clear age-related patterns in the proportion of spore-forming bacteria over time in either system, although data for making such inferences was more limited, with a high proportion of bacterial genera having unknown sporulation capacity (Suppl. Fig. 4, Suppl. Table 1). Overall, these results suggest that while the ratio between aerotolerant and anaerobic bacteria differs between lab and wild populations, the relative abundance of bacteria adapted for anoxic conditions increases with age in both lab and wild mouse gut microbiotas. This is in line with the development of the mammalian gut from an aerobic to increasingly anaerobic environment, potentially due to depletion of oxygen by aerobic microbes alongside chemical reactions of the host (e.g., lipid oxidation) (Albenberg et al., 2014; Friedman et al., 2018).

Conserved taxon-dependent gut microbiota assembly patterns

Despite harbouring gut microbiotas that clearly differed in diversity and composition at lower taxonomic levels (Fig. 1), at the phylum level gut microbial composition of laboratory and wild mice was very similar (Fig. 1E), in line with previous studies (Rosshart et al., 2019; Kreisinger et al., 2014). We therefore compared early life dynamics of the predominant phyla (Firmicutes, Bacteroidota, and Proteobacteria; phyla with >2% mean relative abundance in both systems) across both systems. Remarkably, these dominant phyla displayed very similar early life trends in both systems, with the relative abundances of Firmicutes and Proteobacteria decreasing and that of Bacteroidota increasing (Fig. 3A–B) in early life. The ratio between Firmicutes and Bacteroidota relative abundance became more equal at around 25–30 days for lab mice and at around 15–20 grams of body mass for wild mice (corresponding to ≥17 days of age; Ferrari et al., 2015).

This phylum-level change coincided with the shift in abundance ratio between aerobic and anaerobic bacteria (Fig. 2F–G), suggesting it may be related to depletion of oxygen in the gut.

In line, reduction in the relative abundance of Firmicutes was driven by aerotolerant *Ligilactobacillus* particularly in lab mice (Suppl. Fig. 5A–B). This taxon has been found in milk (Quilodran-Vega et al., 2020), thus it may represent those vertically transmitted from female mice to pups during lactation and the reduction in its relative abundance may reflect cessation of milk consumption (Qi et al., 2022). Indeed, relative abundance of *Ligilactobacillus* decreased in lab mice from an average 73.7% (46.9–90.8%) at 7–14 days of age (n=5) to an average 1.2% (0.1–2.4%) at 19–21 days of age (n=2), when pups were weaned (Suppl. Fig. 5A). In lab mice that had been weaned for a minimum of 7 days (n=17), relative abundance of *Ligilactobacillus* was on average 0.3% (0.0–0.8%). Relative abundance of *Ligilactobacillus* was substantially lower in wild mice of all ages (Suppl. Fig. 5B), which would be expected if the bacterium was primarily transmitted during lactation, as the wild mice were weaned at time of trapping. Anaerobic Muribaculaceae (previously known as S24–7; unknown genus-level taxonomy) was a key driver of Bacteroidota relative abundance increase in both lab and wild mice (Suppl. Fig. 5C–D).

In line with patterns seen in humans and other primates (Jokela et al., 2022; Bäckhead et al., 2015; Baniel et al., 2022), the relative abundance of Proteobacteria decreased in both systems and stabilised at around 25 days of age for lab mice and at 15 grams of body mass for wild mice (Fig. 3A–B). Shared patterns were also observed in ASV richness. In both lab and wild mice, the richness of ASVs in Firmicutes and Bacteroidota increased in early life while Proteobacteria ASV richness remained stable (Fig. 3C–D). The increase and subsequent decrease in relative abundance but consistently low diversity of Proteobacteria in early life may reflect immune maturation, since the phylum includes several common pathogens, such as *Salmonella* and *Escherichia* from the Enterobacteriaceae family, and early-life immune maturation increases the host's ability to suppress such pathogenic taxa.

Indeed, changes in the relative abundance of Proteobacteria were driven by taxa belonging to the Enterobacteriaceae family in both laboratory and wild mice (Suppl. Fig. 6). In laboratory mice, these ASVs (*n*=2) were from *Escherichia/Shigella* and *Muribacter* genera. In wild mice, reduction in Proteobacteria relative abundance was driven by a single ASV; however, genus-level taxonomy could not be assigned with the SILVA database for this ASV (see Methods). A Nucleotide BLAST search of the sequence produced 100% species identity match (with 100% query cover and 0.0 E-value) with 99 species, 83 of which belonged to *Enterobacter*, a genus with several pathogenic strains (Davin-Regli et al., 2019; other genera included *Raoultella*, *Kluyvera*, *Klebsiella*, *Erwinia*, *Pantoea*, and *Cedecea*). Across all lab mouse samples, there were only three ASVs assigned to Enterobacteriaceae, while in wild mouse samples there were 35 such ASVs. The presence of Enterobacteriaceae and other proteobacterial taxa was limited in lab mice by selective exclusion of known pathogens as per common practice (see Methods; Dobson et al., 2019; Scavizzi et al., 2021), such as *Citrobacter* and *Salmonella* (from Enterobacteriaceae), both of which were detected in wild mouse samples.

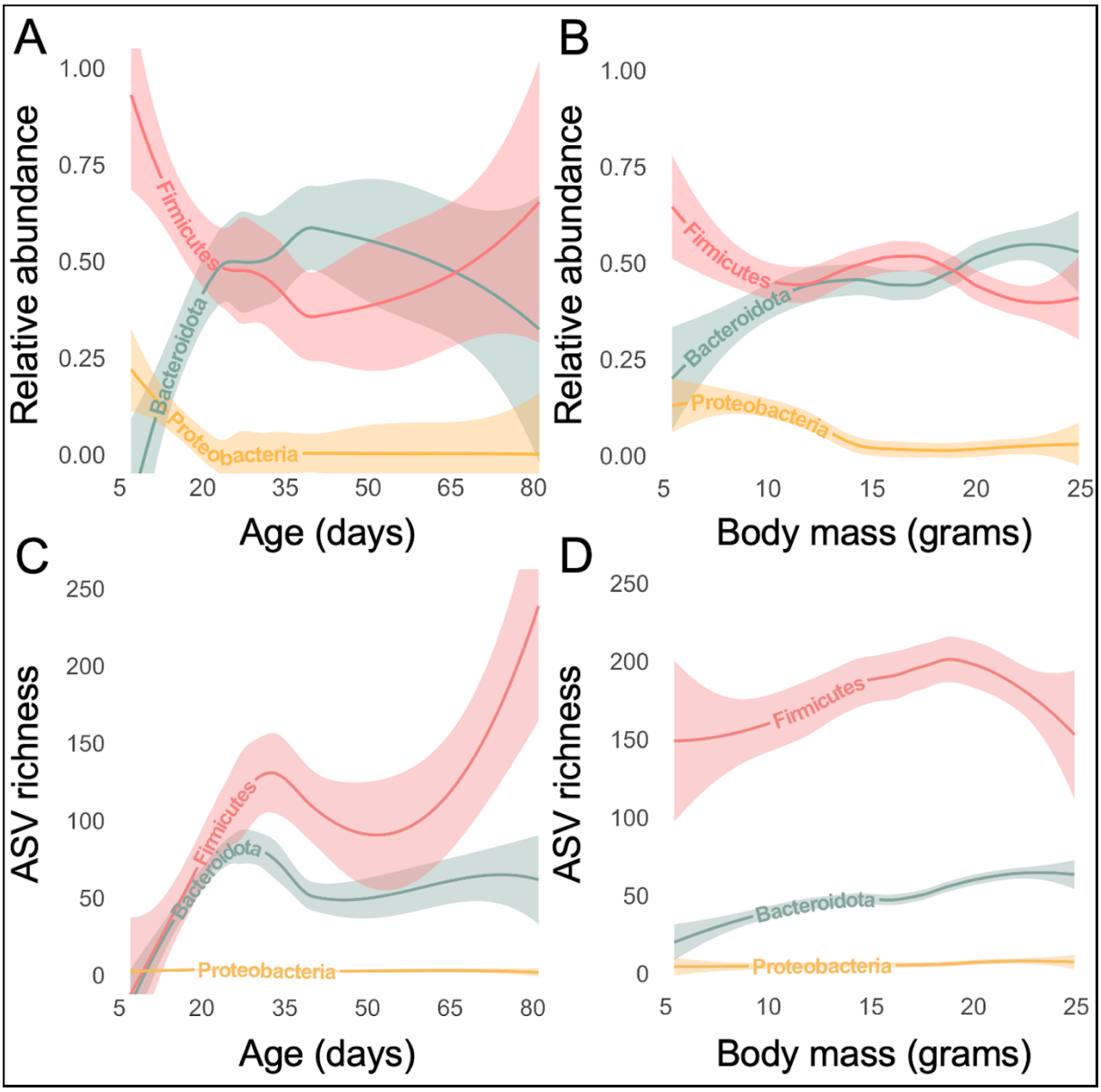


Figure 3. Age-related dynamics in bacterial phyla in (**A**, **C**) lab and (**B**, **D**) wild mice. (**A–B**) Relative abundance of predominant phyla (only phyla with >2% mean relative abundance in both systems are presented; abundances were measured from all taxa). (**C–D**) ASV richness (total count of unique ASVs) in predominant phyla. Lines are locally estimated scatterplot smoothing (LOESS) lines with 95% confidence interval bands.

Finally, we investigated predicted functional profiles of the gut microbiota. The relative abundance of functional pathways had a drastic shift at 17 days of age in lab mice (Suppl. Fig. 7C). Similarly, a compositional restructuring of bacterial genera occurred at 17 days of age (Suppl. Fig. 7A). Given the timing, we hypothesize these changes in lab mice to be driven by the initiation of solid food consumption. As laboratory mice were weaned in this study at 19–

21 days of age but these microbiota changes precede this, the cessation of milk consumption is unlikely the primary driver of this compositional change. Instead, laboratory mice begin eating solid food at 12 days of age, when their teeth begin to erupt and develop (Jax, 2022c). No clear shift in functional profile nor bacterial genus composition was detected in wild mice (Suppl. Fig. 7B, 7D), which were all weaned. Thus, any possible dynamics preceding weaning or initiation of solid food consumption would not necessarily be detectable in this wild mouse dataset. While initiation of solid food consumption appeared to drive some compositional and functional changes in lab mouse gut microbiota, most age-related microbiota changes were detected at 1–2 weeks post weaning (corresponding to 2–3 weeks since initiation of solid food consumption; Fig. 2A, 2D, 2F), indicating cessation of milk consumption – rather than initiation of solid food consumption alone – may be an important driver of gut microbiota maturation. This is in line with what has been demonstrated in humans and other primates (Bäckhead et al., 2015; Wang et al., 2020; Stewart et al., 2018; Baniel et al., 2022).

Considering laboratory mice are widely used in research on gut microbiota and its links to host physiology, we set out to investigate whether dynamics of gut microbiota maturation in laboratory mice are relatable to that in a natural population. A comparison of age-related gut microbial dynamics across standard laboratory mice and entirely wild mice indicated that several gut microbial assembly patterns are conserved across mice from these two systems despite them having remarkably different gut microbiotas. Gut microbiota matured during the first month of life (by 15–20g of body mass in wild mice, which is estimated to correspond to ≥17 days of age; Ferrari et al., 2015). Paralleling results from humans, during this time of maturation, gut microbiota became more diverse within individuals but at the same time more homogenous across individuals and bacteria adapted to the increasingly anaerobic conditions of the developing gut increased in relative abundance.

Despite having taxonomically distinct gut microbiotas, similar trends were detected in compositional changes of predominant phyla. In particular, changes in relative abundance and richness of proteobacterial taxa presented remarkably similar trends across lab and wild mice, potentially reflecting maturation of the immune system. Together these results indicate a level of an intrinsic host programme in early life succession of the gut microbiota that transcends contrasting genetic and environmental backgrounds. As such, our results demonstrate that in the context of early life meaningful insights can be drawn from lab models.

Funding

This work was funded by The Osk. Huttunen Foundation studentship and the National Geographic Society (Early Career grant reference No. EC-58520R-19) to EH, and the European Research Council under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851550) to SK.

Author information

Affiliations

[1]Department of Biology, University of Oxford

[2]Department of Computing, University of Turku

Contributions

EH and SCLK set up the wild mouse study system. EH and AR collected the data. EH conducted laboratory work and analysis. EH wrote the manuscript. All authors contributed to the final manuscript.

References

Albenberg L, Esipova TV, Judge CP, et al. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology*. 2014;147(5):1055-63.e8. doi:10.1053/j.gastro.2014.07.020

Bäckhed F, Roswall J, Peng Y, et al. Dynamics and Stabilization of the Human Gut microbiota during the First Year of Life [published correction appears in Cell Host Microbe. 2015 Jun 10;17(6):852. Jun, Wang [corrected to Wang, Jun]] [published correction appears in Cell Host Microbe. 2015 Jun 10;17(6):852]. *Cell Host Microbe*. 2015;17(5):690-703. doi:10.1016/j.chom.2015.04.004

Baniel A, Petrullo L, Mercer A, et al. Maternal effects on early-life gut microbiota maturation in a wild nonhuman primate. *Curr Biol*. 2022;32(20):4508-4520.e6. doi:10.1016/j.cub.2022.08.037

Chung H, Pamp SJ, Hill JA, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*. 2012;149(7):1578-1593. doi:10.1016/j.cell.2012.04.037

Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems*. 2017;2(1):e00127-16. Published 2017 Jan 3. doi:10.1128/mSystems.00127-16

Darabi B, Rahmati S, HafeziAhmadi MR, Badfar G, Azami M. The association between caesarean section and childhood asthma: an updated systematic review and meta-analysis. *Allergy Asthma Clin Immunol*. 2019;15:62. Published 2019 Oct 29. doi:10.1186/s13223-019-0367-9

Davin-Regli A, Lavigne JP, Pagès JM. *Enterobacter* spp.: Update on Taxonomy, Clinical Aspects, and Emerging Antimicrobial Resistance. *Clin Microbiol Rev*. 2019;32(4):e00002-19. Published 2019 Jul 17. doi:10.1128/CMR.00002-19

Depner M, Taft DH, Kirjavainen PV, et al. Maturation of the gut microbiota during the first year of life contributes to the protective farm effect on childhood asthma. *Nat Med*. 2020;26(11):1766-1775. doi:10.1038/s41591-020-1095-x

Derrien M, Alvarez AS, de Vos WM. The Gut Microbiota in the First Decade of Life. *Trends Microbiol*. 2019;27(12):997-1010. doi:10.1016/j.tim.2019.08.001

Dobson GP, Letson HL, Biros E, Morris J. Specific pathogen-free (SPF) animal status as a variable in biomedical research: Have we come full circle? EBioMedicine. 2019 Mar;41:42-43. doi: 10.1016/j.ebiom.2019.02.038. Epub 2019 Feb 23. PMID: 30803932; PMCID: PMC6443024.

Douglas GM, Maffei VJ, Zaneveld JR, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685-688. doi:10.1038/s41587-020-0548-6

Emmanuel Arnhold (2017). easynls: Easy Nonlinear Model. R package version 5.0. https://CRAN.R-project.org/package=easynls

Ferrari, Manuela; Lindholm, Anna K; König, Barbara (2015). *The risk of exploitation during communal nursing in house mice, Mus musculus domesticus.* Animal Behaviour, 110:133-143.

Ferretti P, Pasolli E, Tett A, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut microbiota. *Cell Host Microbe*. 2018;24(1):133-145.e5. doi:10.1016/j.chom.2018.06.005

Flurkey, Currer, and Harrison, 2007. 'The mouse in biomedical research.' in James G. Fox (ed.), American College of Laboratory Animal Medicine series (Elsevier, AP: Amsterdam; Boston).

Frese SA, Parker K, Calvert CC, Mills DA. Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome*. 2015;3:28. doi:10.1186/s40168-015-0091-8

Friedman ES, Bittinger K, Esipova TV, et al. Microbes vs. chemistry in the origin of the anaerobic gut lumen [published correction appears in Proc Natl Acad Sci U S A. 2022 Jun 14;119(24):e2207826119]. *Proc Natl Acad Sci U S A*. 2018;115(16):4170-4175. doi:10.1073/pnas.1718635115

Gerber N, Auclair Y, König B, Lindholm AK. Population density and temperature influence the return on maternal investment in wild house mice. *Front Ecol Evol.* 2021;8:602359. https://doi.org/10.3389/fevo.2020.602359

Gonze D, Coyte KZ, Lahti L, Faust K. Microbial communities as dynamical systems. *Curr Opin Microbiol*. 2018;44:41-49. doi:10.1016/j.mib.2018.07.004

Gray MM, Parmenter MD, Hogan CA, et al. Genetics of Rapid and Extreme Size Evolution in Island Mice. *Genetics*. 2015;201(1):213-228. doi:10.1534/genetics.115.177790

Guittar J, Shade A, Litchman E. Trait-based community assembly and succession of the infant gut microbiota. *Nat Commun*. 2019;10(1):512. Published 2019 Feb 1. doi:10.1038/s41467-019-08377-w

Hsieh TC, Ma KH, Chao A (2022). *iNEXT: Interpolation and Extrapolation for Species Diversity*. R package version 3.0.0, http://chao.stat.nthu.edu.tw/wordpress/software_download/.

Janiak MC, Montague MJ, Villamil CI, et al. Age and sex-associated variation in the multi-site microbiome of an entire social group of free-ranging rhesus macaques. *Microbiome*. 2021;9(1):68. doi:10.1186/s40168-021-01009-w

Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet, et al (2022). vegan: Community Ecology Package. R package version 2.6-2. https://CRAN.R-project.org/package=vegan

JAX®. Body weight information for aged C57BL/6J. Available at: https://www.jax.org/jax-mice-and-services/strain-data-sheet-pages/body-weight-chart-aged-b6 (Accessed: 30.11.2022)

JAX®. Body weight information for C57BL/6J. Available at: https://www.jax.org/jax-mice-and-services/strain-data-sheet-pages/body-weight-chart-000664 (Accessed: 30.11.2022)

JAX®. Mice pup appearance by age. Available at: https://oacu.oir.nih.gov/system/files/media/file/2021-02/jaxpupsposter.pdf (Accessed: 30.11.2022)

Jokela R, Korpela K, Jian C, et al. Quantitative insights into effects of intrapartum antibiotics and birth mode on infant gut microbiota in relation to well-being during the first year of life. *Gut Microbes*. 2022;14(1):2095775. doi:10.1080/19490976.2022.2095775

Kashyap PC, Marcobal A, Ursell LK, et al. Complex interactions among diet, gastrointestinal transit, and gut microbiota in humanized mice. *Gastroenterology*. 2013;144(5):967-977. doi:10.1053/j.gastro.2013.01.047

Kreisinger J, Cížková D, Vohánka J, Piálek J. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Mol Ecol*. 2014;23(20):5048-5060. doi:10.1111/mec.12909

Leo Lahti, Sudarshan Shetty et al. (2017). Tools for microbiota analysis in R. URL: http://microbiota.github.com/microbiota.

Liberti J, Kay T, Quinn A, et al. The gut microbiota affects the social network of honeybees. *Nat Ecol Evol*. 2022;6(10):1471-1479. doi:10.1038/s41559-022-01840-w

Marsh KJ, Raulo AM, Brouard M, et al. Synchronous Seasonality in the Gut Microbiota of Wild Mouse Populations. *Front Microbiol*. 2022;13:809735. Published 2022 Apr 25. doi:10.3389/fmicb.2022.809735

Maurice CF, Knowles SC, Ladau J, et al. Marked seasonal variation in the wild mouse gut microbiota. *ISME J*. 2015;9(11):2423-2434. doi:10.1038/ismej.2015.53

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiota census data. *PLoS One*. 2013;8(4):e61217. Published 2013 Apr 22. doi:10.1371/journal.pone.0061217

Nishida AH, Ochman H. Rates of gut microbiota divergence in mammals. *Mol Ecol*. 2018;27(8):1884-1897. doi:10.1111/mec.14473

Niu J, Xu L, Qian Y, et al. Evolution of the Gut microbiota in Early Childhood: A Cross-Sectional Study of Chinese Children. *Front Microbiol*. 2020;11:439. Published 2020 Apr 3. doi:10.3389/fmicb.2020.00439

Olin A, Henckel E, Chen Y, et al. Stereotypic Immune System Development in Newborn Children. *Cell*. 2018;174(5):1277-1292.e14. doi:10.1016/j.cell.2018.06.045

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol*. 2016;18(5):1403-1414. doi:10.1111/1462-2920.13023

Petrullo L, Baniel A, Jorgensen MJ, Sams S, Snyder-Mackler N, Lu A. The early life microbiota mediates maternal effects on offspring growth in a nonhuman primate. *iScience*. 2022;25(3):103948. Published 2022 Feb 18. doi:10.1016/j.isci.2022.103948

Qi C, Zhou J, Tu H, et al. Lactation-dependent vertical transmission of natural probiotics from the mother to the infant gut through breast milk. *Food Funct*. 2022;13(1):304-315. Published 2022 Jan 4. doi:10.1039/d1fo03131g

Quilodrán-Vega S, Albarracin L, Mansilla F, et al. Functional and Genomic Characterization of *Ligilactobacillus salivarius* TUCO-L2 Isolated From *Lama glama*Milk: A Promising Immunobiotic Strain to Combat Infections. *Front Microbiol.* 2020;11:608752. doi: *10.3389/fmicb.2020.608752*

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raulo A, Allen BE, Troitsky T, et al. Social networks strongly predict the gut microbiota of wild mice. *ISME J*. 2021;15(9):2601-2613. doi:10.1038/s41396-021-00949-3

Reese AT, Phillips SR, Owens LA, et al. Age Patterning in Wild Chimpanzee Gut Microbiota Diversity Reveals Differences from Humans in Early Life. *Curr Biol*. 2021;31(3):613-620.e3. doi:10.1016/j.cub.2020.10.075

Rosshart SP, Herz J, Vassallo BG, et al. Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science*. 2019;365(6452):eaaw4361. doi:10.1126/science.aaw4361

Rudolph K, Schneider D, Fichtel C, Daniel R, Heistermann M, Kappeler PM. Drivers of gut microbiota variation within and between groups of a wild Malagasy primate. *microbiota*. 2022;10(1):28. Published 2022 Feb 9. doi:10.1186/s40168-021-01223-6

Scavizzi F, Bassi C, Lupini L, Guerriero P, Raspa M, Sabbioni S. A comprehensive approach for microbiota and health monitoring in mouse colonies using metagenomic shotgun sequencing. *Anim Microbiome*. 2021;3(1):53. Published 2021 Jul 29. doi:10.1186/s42523-021-00113-4

Shao Y, Forster SC, Tsaliki E, et al. Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature*. 2019;574(7776):117-121. doi:10.1038/s41586-019-1560-1

Sharon G, Sampson TR, Geschwind DH, Mazmanian SK. The Central Nervous System and the Gut microbiota. *Cell*. 2016;167(4):915-932. doi:10.1016/j.cell.2016.10.027

Spangenberg E, Wallenbeck A, Eklöf AC, Carlstedt-Duke J, Tjäder S. Housing breeding mice in three different IVC systems: maternal performance and pup development. *Lab Anim*. 2014;48(3):193-206. doi:10.1177/0023677214531569

Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiota in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-588. doi:10.1038/s41586-018-0617-x

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiota with increased capacity for energy harvest. *Nature*. 2006;444(7122):1027-1031. doi:10.1038/nature05414

Vandeputte D, Kathagen G, D'hoe K, et al. Quantitative microbiota profiling links gut community variation to microbial load. *Nature*. 2017;551(7681):507-511. doi:10.1038/nature24460

Visconti A, Le Roy CI, Rosa F, et al. Interplay between the human gut microbiota and host metabolism. *Nat Commun*. 2019;10(1):4505. Published 2019 Oct 3. doi:10.1038/s41467-019-12476-z

Walters W, Hyde ER, Berg-Lyons D, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2015;1(1):e00009-15. Published 2015 Dec 22. doi:10.1128/mSystems.00009-15

Wang J, Kalyan S, Steck N, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun*. 2015;6:6440. Published 2015 Mar 4. doi:10.1038/ncomms7440

Wang J, Linnenbrink M, Künzel S, et al. Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiota of wild mice. *Proc Natl Acad Sci U S A*. 2014;111(26):E2703-E2710. doi:10.1073/pnas.1402342111

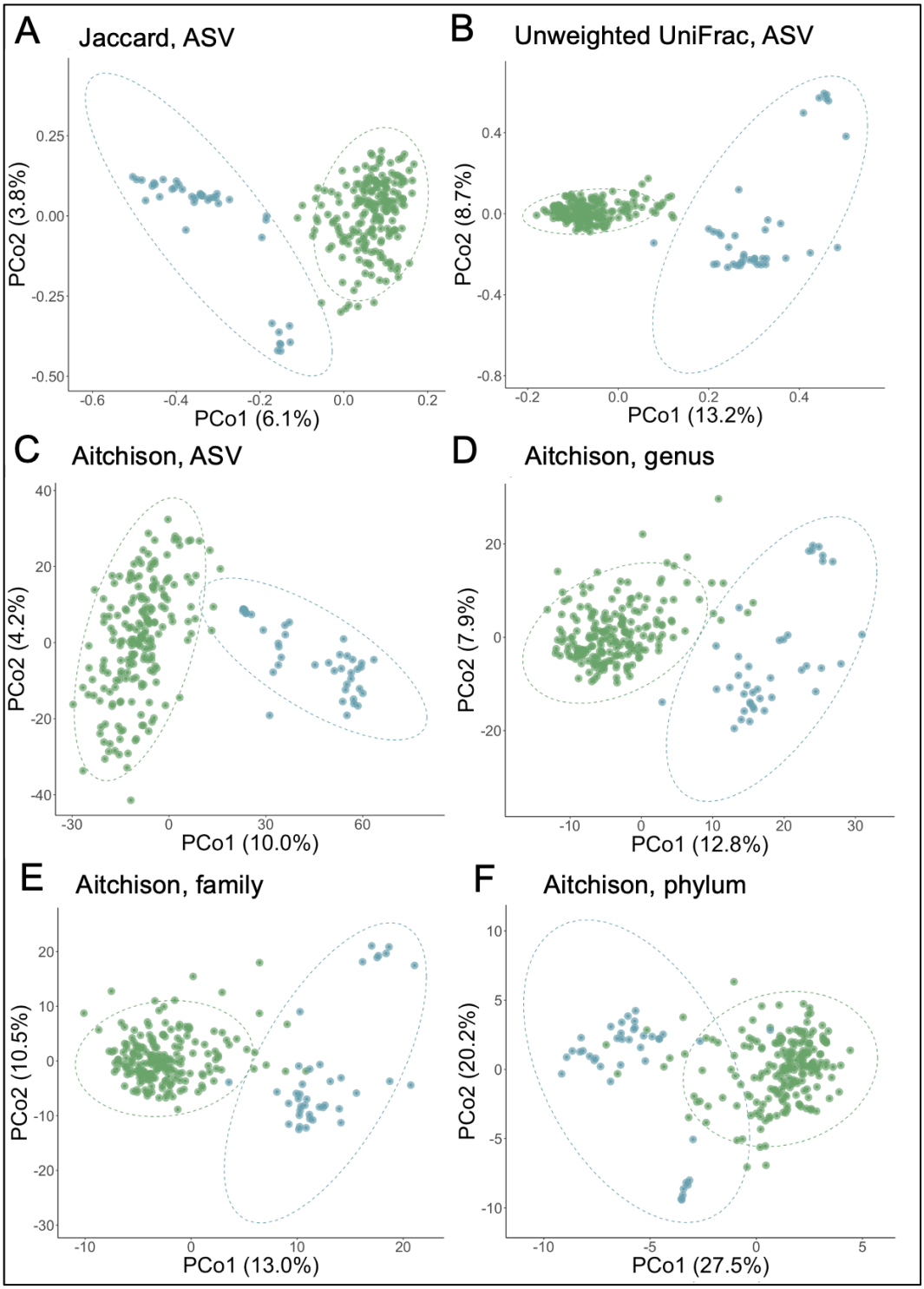
Wang S, Ryan CA, Boyaval P, Dempsey EM, Ross RP, Stanton C. Maternal Vertical Transmission Affecting Early-life Microbiota Development. *Trends Microbiol*. 2020;28(1):28-45. doi:10.1016/j.tim.2019.07.010

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

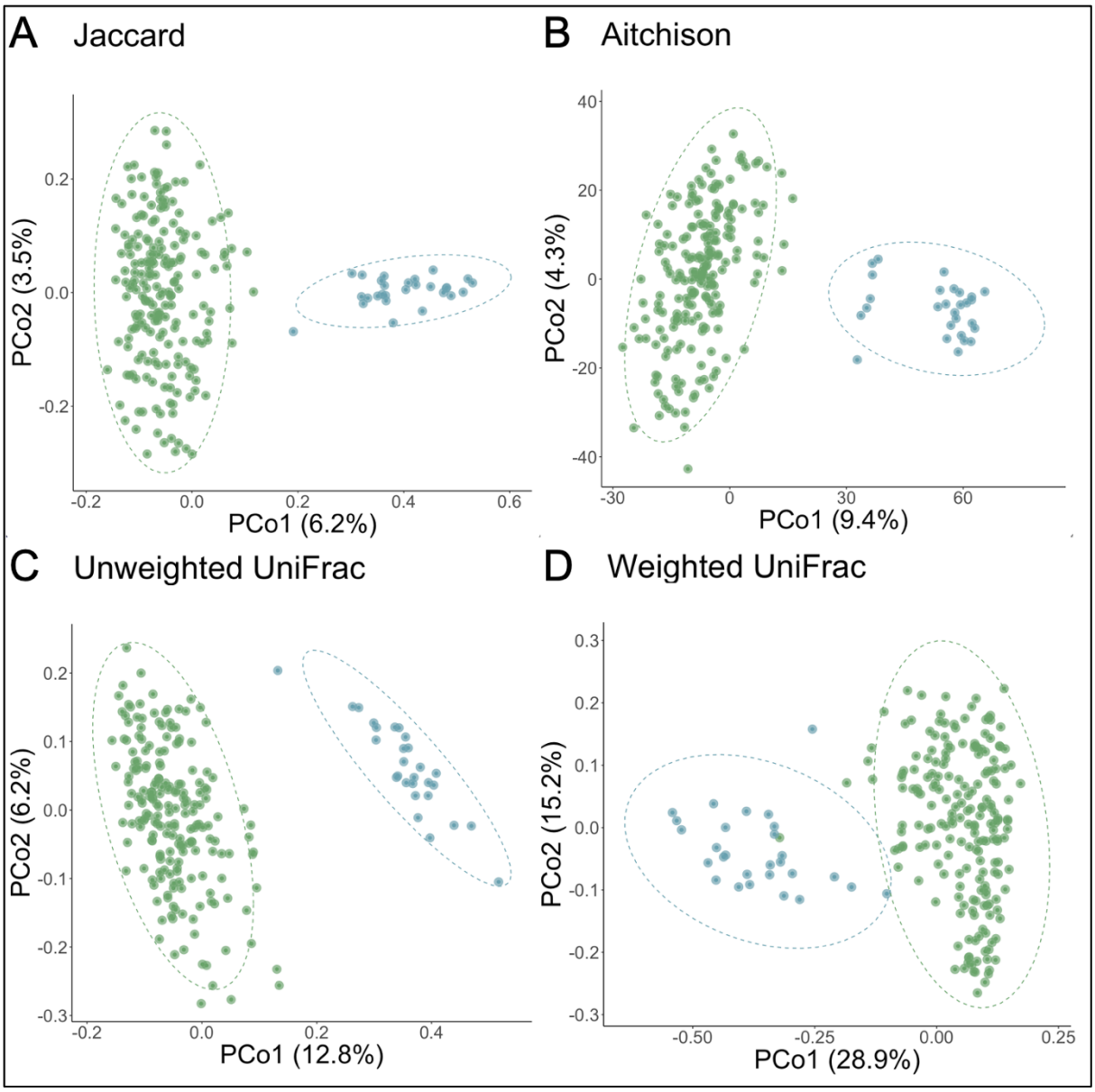
Yassour M, Vatanen T, Siljander H, et al. Natural history of the infant gut microbiota and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med*. 2016;8(343):343ra81. doi:10.1126/scitranslmed.aad0917

Zhou J, Xu G, Li X. *Limosilactobacillus reuteri* FN041 prevents atopic dermatitis in pup mice by remodeling the ileal microbiota and regulating gene expression in Peyer's patches after vertical transmission. *Front Nutr.* 2022;9:987400. doi: 10.3389/fnut.2022.987400

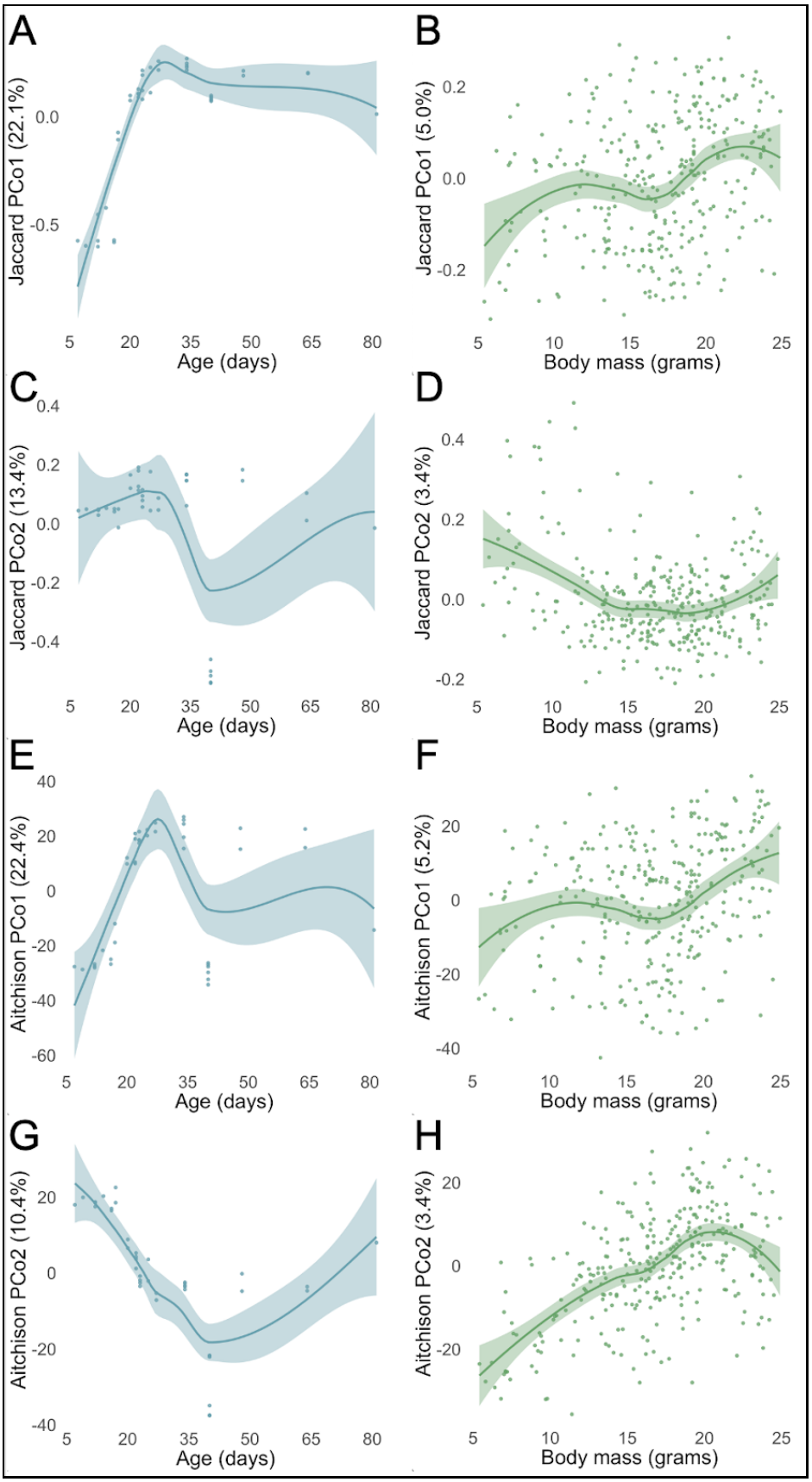
Supplementary data



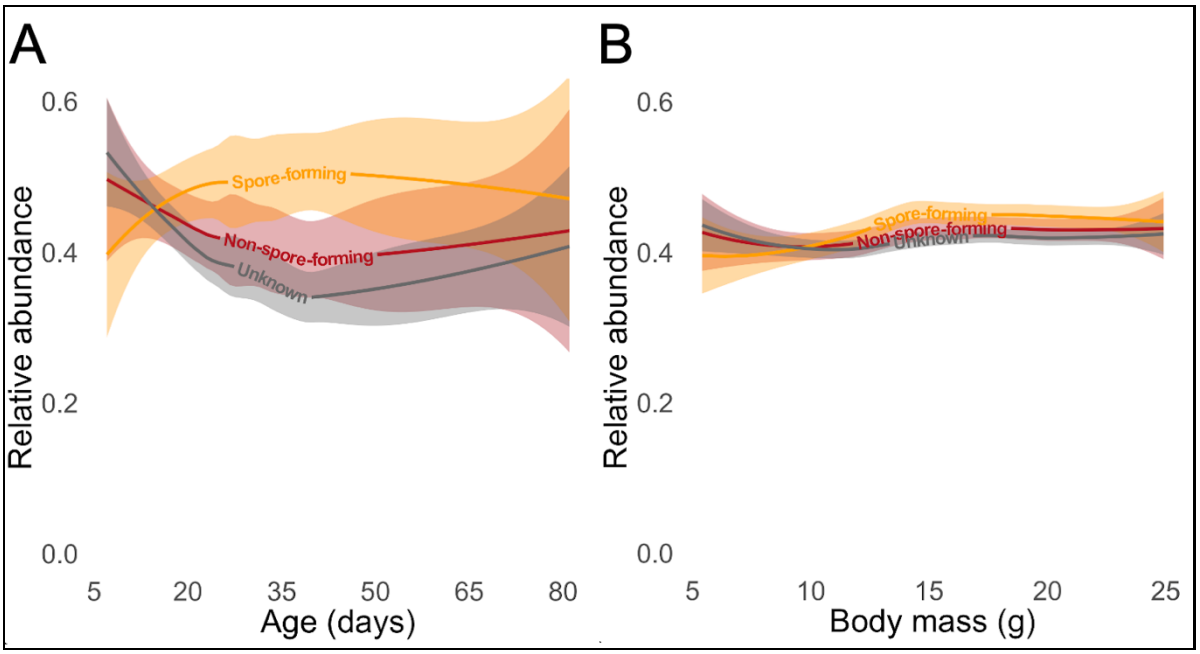
Supplementary Figure 1. Principal coordinates analysis (PCoA) of 239 faecal samples from 40 laboratory and 199 wild mice using (**A**) Jaccard, (**B**) unweighted UniFrac, and (**C–F**) Aitchison distances at (**A–C**) amplicon sequence variant (ASV), (**D**) genus, (**E**) family, or (**F**) phylum level. Colour indicates sample source (*blue* = laboratory, *green* = wild).



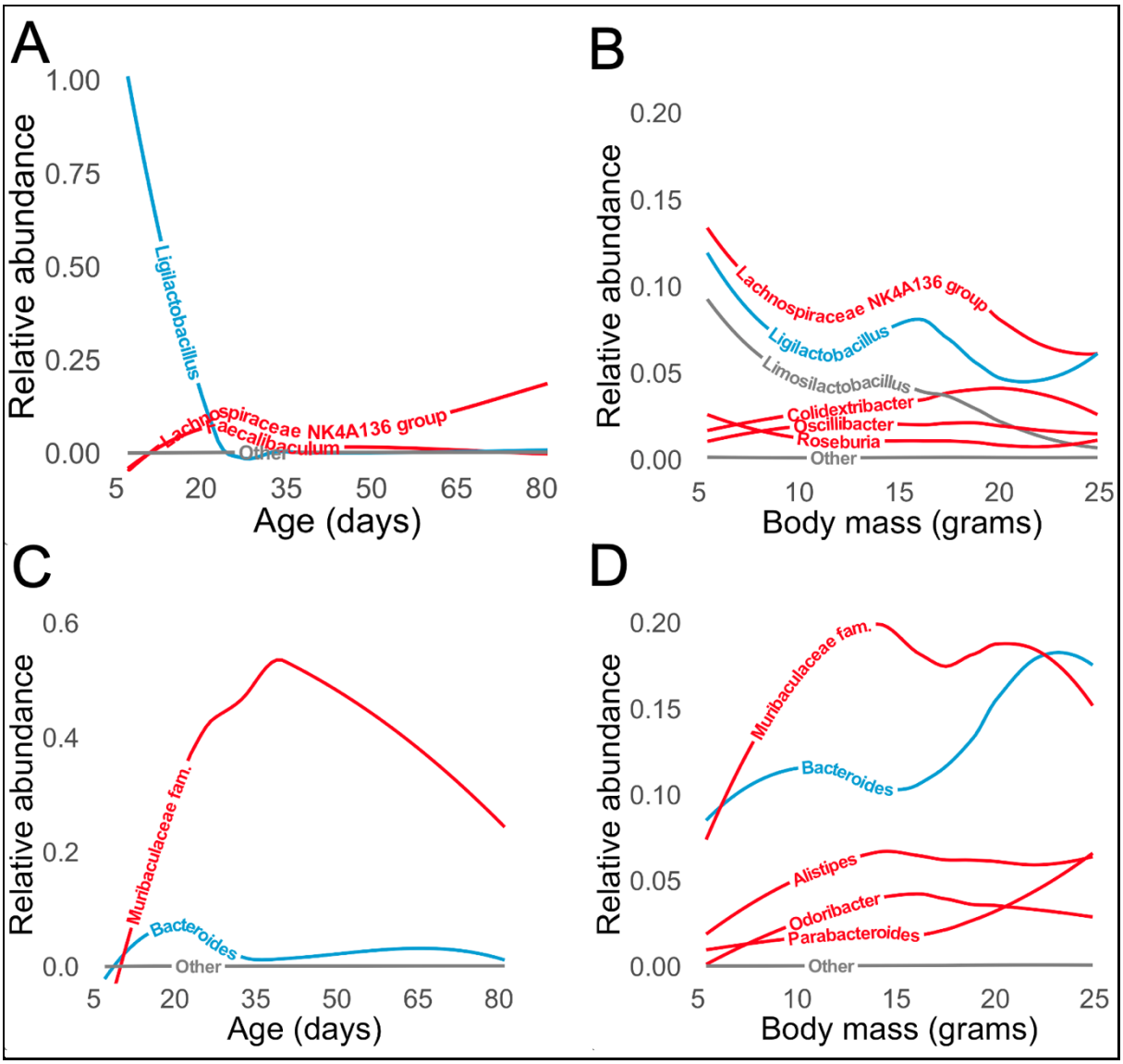
Supplementary Figure 2. Principal coordinates analysis (PCoA) of 227 faecal samples from 28 laboratory and 199 wild mice using (**A**) Jaccard, (**B**) Aitchison, (**C**) unweighted UniFrac, and (**D**) weighted UniFrac distances at amplicon sequence variant (ASV) level. Pre-weaned lab mice have been excluded. Colour indicates sample source (*blue* = laboratory, *green* = wild).



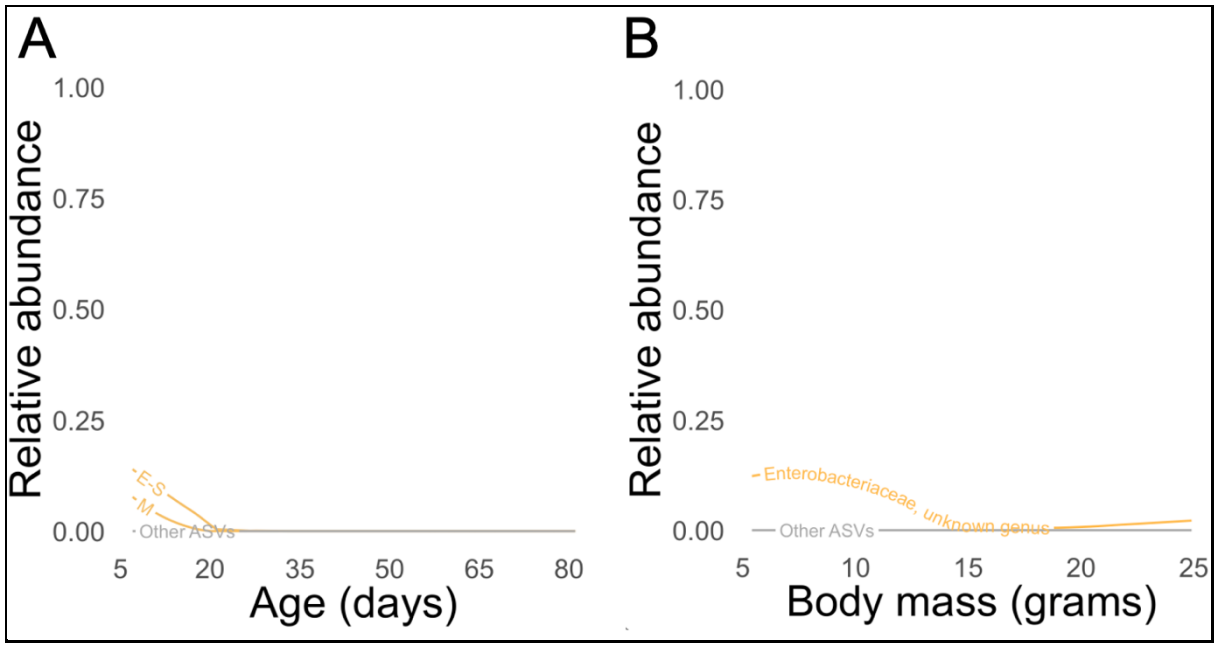
Supplementary Figure 3. First and second principal coordinates on (**A**–**D**) Jaccard and (**E**–**H**) Aitchison distance plotted against age in laboratory mice (*blue*) and body mass in wild mice (*green*). Lines are locally estimated scatterplot smoothing (LOESS) lines with 95% confidence interval bands.



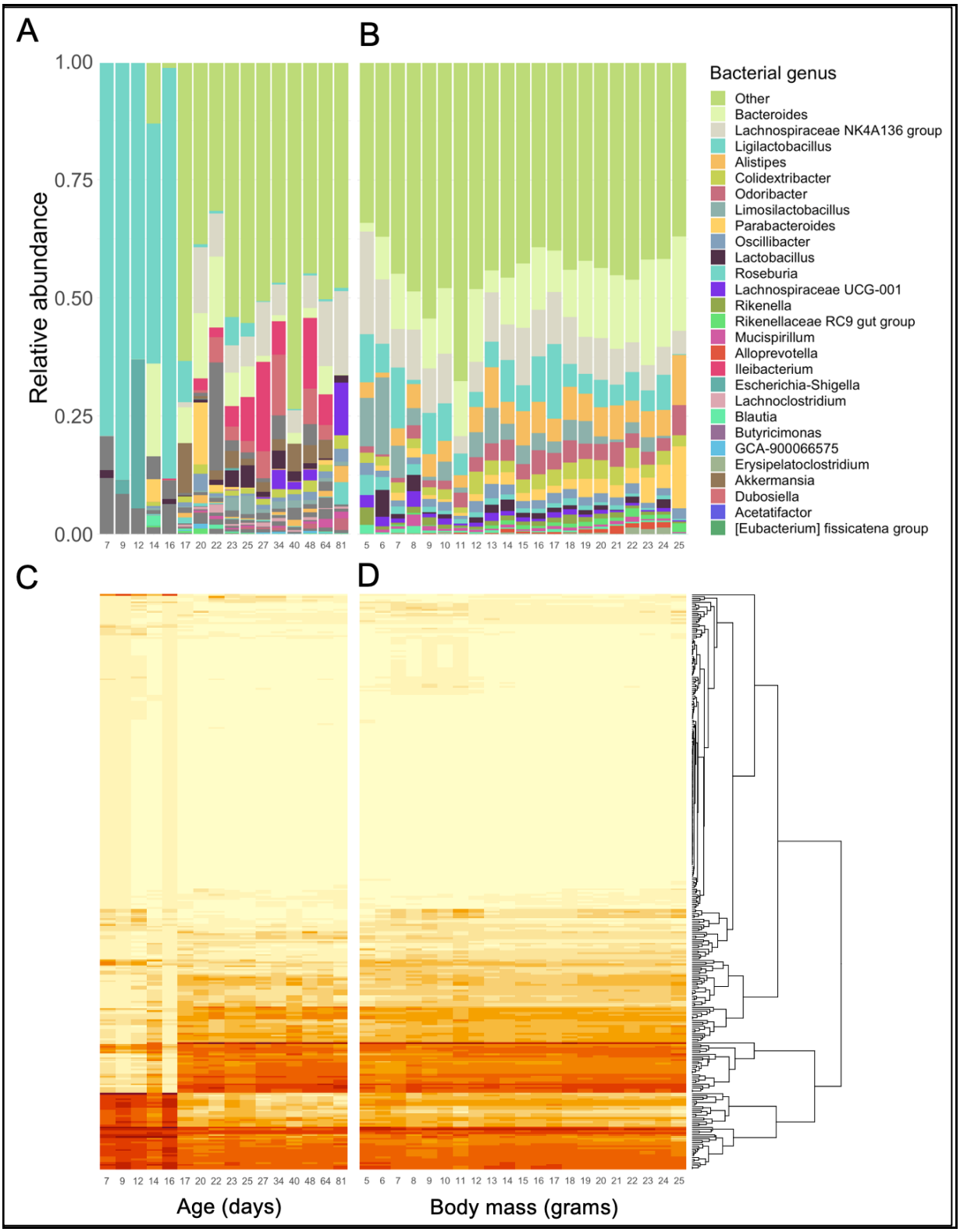
Supplementary Figure 4. Relative abundance of sporulating (*yellow*) and non-sporulating (*red*) bacteria. Bacterial genera with unknown sporulation ability (*grey*) made up 33.3% of microbial relative abundance in both systems (range 0.0–99.9% in individual laboratory mice and 0.2–96.1% in wild mice). Lines are locally estimated scatterplot smoothing (LOESS) lines with 95% confidence interval bands.



Supplementary Figure 5. Relative abundance of genera from predominant phyla, (**A–B**) Firmicutes and (**C–D**) Bacteroidota, in (A, C) laboratory and (B, D) wild mice. Relative abundances are from the whole microbiota, rather than within phylum. Lines are locally estimated scatterplot smoothing (LOESS) lines. Confidence intervals are not plotted for easier interpretation. Lines are coloured by genus aerotolerance (*red* = obligate anaerobes, *blue* = aerotolerant, *grey* = unknown/mixed aerotolerance). Note the variable y-axes scales to aid visibility of taxa with low abundance.



Supplementary Figure 6. Relative abundance of ASVs assigned to the phylum Proteobacteria in (**A**) lab and (**B**) wild mouse gut microbiota. E-S=*Escherichia-Shigella*, M=*Muribacter*.



Supplementary Figure 7. Composition of bacterial genera and functional pathways in laboratory and wild mice. **(A–B)** Mean relative abundance of bacterial genera across each available unit of age/body mass in (**A**) lab and (**B**) wild mice. (**C–D**) Mean relative abundance of predicted functional pathways across each available unit of age/body mass in (**C**) lab and (**D**) wild mice. Rows are unique pathways ordered by hierarchical clustering as indicated with dendrogram. Colour darkness increases with relative abundance (range 0.0–2.7%).

Supplementary Table 1. Aerotolerance and spore-forming ability of bacterial genera detected across laboratory and wild mice. Bergey's Manual of Systematics of Archae and Bacteria alongside additional references (listed below) were used to determine aerotolerance (A = aerotolerant, OA = obligate anaerobe) and spore-forming ability (SF= spore forming, NSF= nonsporeforming). Bacteria were classified as obligate anaerobes **only** when explicitly listed as *obligate* anaerobes. Aerotolerance and spore-forming ability were determined based on genus listed in *Genus*. If multiple genera were assigned for a given ASV (e.g., "Methylobacterium-Methylorubrum", genus based on which aerotolerance and spore-forming ability were determined is listed in *Comments*. If genus-level information was not available, family-level information was inspected and used if all genera from a given family were stated to have same aerotolerance and sporulation ability (indicated in *Comments* as 'Based on -ceae').

Family	*Genus*	*Aerotolerance*	*Sporulation*	*Reference*	*Comments*
[Clostridium] methylpentosum group	Other	Unknown	Unknown		
[Eubacterium] coprostanoligenes group	Other	Unknown	Unknown		
67-14	Other	Unknown	Unknown		
A4b	Other	Unknown	Unknown		
Abditibacteriaceae	Abditibacterium	Unknown	Unknown		
Acetobacteraceae	Acetobacter	A	NSF	Bergey's Manual	
Acetobacteraceae	Acidicaldus	A	Unknown	Johnson et al., 2006	
Acetobacteraceae	Acidiphilium	A	NSF	Bergey's Manual	
Acetobacteraceae	Acidisoma	A	NSF	Bergey's Manual	
Acetobacteraceae	Endobacter	A	Unknown	Bergey's Manual	Based on Acetobacteraceae
Acetobacteraceae	Gluconobacter	A	NSF	Bergey's Manual	
Acetobacteraceae	Other	A	Unknown	Bergey's Manual	Based on Acetobacteraceae
Acetobacteraceae	Rhodovastum	A	Unknown	Bergey's Manual	Based on Acetobacteraceae
Acetobacteraceae	Roseomonas	A	Unknown	Bergey's Manual	
Acholeplasmataceae	Anaeroplasma	OA	Unknown	Bergey's Manual	
Acholeplasmataceae	Other	Unknown	Unknown		
Acidaminococcaceae	Phascolarctobacterium	OA	NSF	Bergey's Manual	
Acidobacteriaceae (Subgroup 1)	Edaphobacter	A	NSF	Bergey's Manual	
Acidothermaceae	Acidothermus	A	NSF	Bergey's Manual	
Actinomycetaceae	Actinomyces	A	NSF	Bergey's Manual	
Aerococcaceae	Facklamia	A	NSF	Bergey's Manual	

Aerococcaceae	Other	A	NSF	Bergey's Manual	Based on Aerococcaceae
Aeromonadaceae	Oceanisphaera	A	NSF	Bergey's Manual	Based on Aeromonadaceae
Akkermansiaceae	Akkermansia	OA	Unknown	Bergey's Manual	
AKYG1722	Other	Unknown	Unknown		
Alcaligenaceae	Achromobacter	A	NSF	Bergey's Manual	
Alcaligenaceae	Alcaligenes	A	Unknown	Bergey's Manual	Based on Alcaligenaceae
Alcaligenaceae	Candidimonas	A	NSF	Bergey's Manual	
Alcaligenaceae	Eoetvoesia	A	NSF	Bergey's Manual	Based on Alcaligenaceae
Alcaligenaceae	Other	A	NSF	Bergey's Manual	Based on Alcaligenaceae
Alcaligenaceae	Paenalcaligenes	A	Unknown	Bergey's Manual	
Alcaligenaceae	Verticiella	A	Unknown	Vandamme et al., 2015	Based on Verticia
Alicyclobacillaceae	Tumebacillus	A	SF	Bergey's Manual	
Amoebophilaceae	Candidatus Cardinium	Unknown	Unknown		
Anaerofustaceae	Anaerofustis	OA	NSF	Bergey's Manual	
Anaerovoracaceae	[Eubacterium] brachy group	OA	NSF	Bergey's Manual	Based on Eubacterium
Anaerovoracaceae	[Eubacterium] nodatum group	OA	NSF	Bergey's Manual	Based on Eubacterium
Anaerovoracaceae	Anaerovorax	OA	NSF	Bergey's Manual	
Anaerovoracaceae	Family XIII AD3011 group	A	NSF	Bergey's Manual	Based on Anaerovorax
Anaerovoracaceae	Family XIII UCG-001	A	NSF	Bergey's Manual	Based on Anaerovorax
Anaerovoracaceae	Other	Unknown	Unknown		
Anaplasmataceae	Wolbachia	Unknown	Unknown		
Atopobiaceae	Coriobacteriaceae UCG-002	A	NSF	Bergey's Manual	Based on Atopobiaceae
Atopobiaceae	Coriobacteriaceae UCG-003	A	NSF	Bergey's Manual	Based on Atopobiaceae
Azospirillaceae	Skermanella	A	NSF	Bergey's Manual	
Bacillaceae	Allobacillus	Unknown	Unknown		
Bacillaceae	Bacillus	A	SF	Bergey's Manual	
Bacillaceae	Falsibacillus	Unknown	Unknown		
Bacillaceae	Microaerobacter	Unknown	Unknown		
Bacillaceae	Natronobacillus	Unknown	Unknown		

Bacillaceae	Other	Unknown	Unknown		
Bacteroidaceae	Bacteroides	A	NSF	Bergey's Manual	
Bacteroidales RF16 group	Other	A	NSF	Bergey's Manual	Based on Bacteroidales RF16 group
Barnesiellaceae	Barnesiella	OA	NSF	Bergey's Manual	
Barnesiellaceae	Coprobacter	Unknown	Unknown		
Barnesiellaceae	Other	Unknown	Unknown		
Bdellovibrionaceae	Bdellovibrio	A	Unknown	Bergey's Manual	
Beijerinckiaceae	1174-901-12	Unknown	Unknown		
Beijerinckiaceae	Bosea	A	Unknown	Bergey's Manual	
Beijerinckiaceae	Methylobacterium-Methylorubrum	A	NSF	Bergey's Manual	Based on Methylobacterium
Beijerinckiaceae	Microvirga	A	NSF	Kanso & Patel, 2003	
Beijerinckiaceae	Other	A	NSF	Bergey's Manual	Based on Beijerinckiaceae
Beijerinckiaceae	Roseiarcus	Unknown	Unknown		
Bifidobacteriaceae	Bifidobacterium	A	NSF	Bergey's Manual	
Bradymonadaceae	Bradymonas	Unknown	Unknown		
Brevibacteriaceae	Brevibacterium	A	Unknown	Bergey's Manual	
Brevibacteriaceae	Spelaeicoccus	Unknown	Unknown		
Bryobacteraceae	Bryobacter	A	NSF	Bergey's Manual	
Budviciaceae	Budvicia	A	NSF	Bergey's Manual	
Butyricicoccaceae	Butyricicoccus	OA	Unknown	Trachsel et al., 2018	
Butyricicoccaceae	Other	Unknown	Unknown		
Butyricicoccaceae	UCG-008	Unknown	Unknown		
Butyricicoccaceae	UCG-009	OA	SF	Bergey's Manual	Based on Clostridia
Caldicoprobacteraceae	Caldicoprobacter	OA	SF	Bergey's Manual	
Candidatus Hepatincola	Other	Unknown	Unknown		
Carnobacteriaceae	Atopostipes	A	NSF	Bergey's Manual	
Carnobacteriaceae	Carnobacterium	A	NSF	Bergey's Manual	
Carnobacteriaceae	Granulicatella	A	NSF	Bergey's Manual	
Carnobacteriaceae	Marinilactibacillus	A	NSF	Bergey's Manual	
Catellicoccaceae	Catellicoccus	Unknown	Unknown		
Caulobacteraceae	Brevundimonas	A	NSF	Bergey's Manual	
Caulobacteraceae	Caulobacter	A	Unknown	Bergey's Manual	

Caulobacteraceae	Phenylobacterium	A	NSF	Bergey's Manual	
Cellulomonadaceae	Cellulomonas	A	NSF	Bergey's Manual	
Cellulomonadaceae	Oerskovia	A	NSF	Bergey's Manual	
Cellulomonadaceae	Paraoerskovia	A	NSF	Bergey's Manual	
Cellulomonadaceae	Pseudactinotalea	Unknown	Unknown		
Cellvibrionaceae	Cellvibrio	A	Unknown	Bergey's Manual	
Chitinophagaceae	Chitinophaga	A	Unknown	Bergey's Manual	Based on Chitinophagaceae
Chitinophagaceae	Haoranjiania	A	Unknown	Bergey's Manual	
Chitinophagaceae	Other	A	Unknown	Bergey's Manual	
Chitinophagaceae	Segetibacter	A	Unknown	Bergey's Manual	
Christensenellaceae	Christensenella	OA	Unknown	Morotomi et al., 2012	
Christensenellaceae	Christensenellaceae R-7 group	OA	Unknown	Morotomi et al., 2013	
Christensenellaceae	Other	OA	Unknown	Morotomi et al., 2014	
Chroococcidiopsaceae	Aliterella	OA	Unknown		
Chthoniobacteraceae	Candidatus Udaeobacter	A	NSF	Bergey's Manual	
Clostridiaceae	Candidatus Arthromitus	OA	SF	Schnupf et al., 2015	
Clostridiaceae	Clostridium sensu stricto 1	OA	SF	Bergey's Manual	Based on Clostridiaceae
Clostridiaceae	Clostridium sensu stricto 5	OA	SF	Bergey's Manual	Based on Clostridiaceae
Clostridiaceae	Clostridium sensu stricto 6	OA	SF	Bergey's Manual	Based on Clostridiaceae
Clostridiaceae	Clostridium sensu stricto 7	OA	SF	Bergey's Manual	Based on Clostridiaceae
Clostridiaceae	Other	OA	SF	Bergey's Manual	Based on Clostridiaceae
Comamonadaceae	Comamonas	A	Unknown	Bergey's Manual	
Comamonadaceae	Lampropedia	A	Unknown	Bergey's Manual	
Comamonadaceae	Limnohabitans	A	Unknown	Bergey's Manual	
Comamonadaceae	Other	Unknown	NSF	Bergey's Manual	Based on Comamonadaceae
Comamonadaceae	Polaromonas	A	Unknown	Bergey's Manual	
Comamonadaceae	Simplicispira	Unknown	NSF	Bergey's Manual	Based on Comamonadaceae
Comamonadaceae	Variovorax	A	NSF	Bergey's Manual	

Comamonadaceae	Xenophilus	Unknown	NSF	Bergey's Manual	Based on Comamonadaceae
Coralloluteibacterium	Other	Unknown	Unknown		
Coriobacteriaceae	Collinsella	OA	NSF	Bergey's Manual	
Coriobacteriales Incertae Sedis	Other	Unknown	Unknown		
Corynebacteriaceae	Corynebacterium	A	NSF	Bergey's Manual	
Corynebacteriales Incertae Sedis	Tomitella	Unknown	Unknown		
Crocinitomicaceae	Fluviicola	A	Unknown	Bergey's Manual	
Crocinitomicaceae	Other	Unknown	Unknown		
Cryomorphaceae	Other	A	NSF	Bergey's Manual	Based on Cryomorphaceae
Cyanobiaceae	Synechococcus CC9902	Unknown	Unknown		
Cyclobacteriaceae	Algoriphagus	A	Unknown	Bergey's Manual	
Cyclobacteriaceae	Other	A	NSF	Bergey's Manual	Based on Cyclobacteriaceae
Deferribacteraceae	Mucispirillum	A	NSF	Bergey's Manual	Based on Deferribacteraceae
Defluviitaleaceae	Defluviitaleaceae UCG-011	A	SF	Bergey's Manual	Based on Defluviitaleaceae
Demequinaceae	Demequina	Unknown	Unknown		
Dermabacteraceae	Brachybacterium	A	NSF	Bergey's Manual	
Dermabacteraceae	Helcobacillus	Unknown	NSF	Bergey's Manual	Based on Dermabacteraceae
Dermacoccaceae	Flexivirga	A	NSF	Bergey's Manual	Based on Dermacoccaceae
Dermacoccaceae	Other	A	NSF	Bergey's Manual	Based on Dermacoccaceae
Desulfovibrionaceae	Bilophila	OA	NSF	Bergey's Manual	
Desulfovibrionaceae	Desulfovibrio	OA	NSF	Bergey's Manual	
Desulfovibrionaceae	Lawsonia	OA	NSF	Bergey's Manual	
Desulfovibrionaceae	Other	OA	NSF	Bergey's Manual	Based on Desulfovibrionaceae
DEV007	Other	Unknown	Unknown		
Devosiaceae	Arsenicitalea	Unknown	Unknown		
Devosiaceae	Devosia	A	NSF	Bergey's Manual	
Devosiaceae	Other	Unknown	Unknown		
Devosiaceae	Pelagibacterium	Unknown	Unknown		
Dietziaceae	Dietzia	A	NSF	Bergey's Manual	

Diplorickettsiaceae	Diplorickettsia	A	NSF	Mediannikov et al., 2010	
Diplorickettsiaceae	Other	Unknown	Unknown		
Diplorickettsiaceae	Rickettsiella	A	NSF	Bergey's Manual	
Dysgonomonadaceae	Dysgonomonas	A	Unknown	Bergey's Manual	
Eggerthellaceae	Adlercreutzia	OA	NSF	Bergey's Manual	
Eggerthellaceae	Asaccharobacter	A	NSF	Bergey's Manual	Based on Eggerthellaceae
Eggerthellaceae	DNF00809	Unknown	Unknown	Bergey's Manual	Based on Eggerthellaceae
Eggerthellaceae	Eggerthella	OA	NSF	Bergey's Manual	
Eggerthellaceae	Enterorhabdus	A	NSF	Clavel et al., 2009	
Eggerthellaceae	Gordonibacter	OA	NSF	Wurdemann et al., 2009, Bergey's Manual	
Eggerthellaceae	Other	A	NSF	Bergey's Manual	Based on Eggerthellaceae
Eggerthellaceae	Parvibacter	A	NSF	Clavel et al., 2013	
Enterobacteriaceae	Aquamonas	Unknown	NSF	Bergey's manual, Dorlands medical dictionary	Based on Enterobacteriaceae
Enterobacteriaceae	Atlantibacter	Unknown	NSF	Bergey's manual, Dorlands medical dictionary	Based on Enterobacteriaceae
Enterobacteriaceae	Buttiauxella	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Cedecea	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Citrobacter	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Escherichia-Shigella	A	NSF	Bergey's Manual	Based on Escherichia
Enterobacteriaceae	Klebsiella	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Kluyvera	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Kosakonia	Unknown	NSF	Bergey's manual, Dorlands medical dictionary	Based on Enterobacteriaceae
Enterobacteriaceae	Other	Unknown	NSF	Bergey's manual, Dorlands medical dictionary	Based on Enterobacteriaceae

Enterobacteriaceae	Raoultella	Unknown	NSF	Bergey's manual, Dorlands medical dictionary	Based on Enterobacteriaceae
Enterobacteriaceae	Salmonella	A	Unknown	Bergey's Manual	
Enterobacteriaceae	Yokenella	A	NSF	Bergey's Manual	
Enterococcaceae	Enterococcus	A	NSF	Bergey's Manual	
Enterococcaceae	Other	Unknown	Unknown		
Entomoplasmatales Incertae Sedis	Candidatus Hepatoplasma	Unknown	Unknown		
Erwiniaceae	Erwinia	A	NSF	Bergey's Manual	
Erwiniaceae	Other	Unknown	Unknown		
Erwiniaceae	Pantoea	A	NSF	Bergey's Manual	
Erwiniaceae	Siccibacter	Unknown	Unknown		
Erwiniaceae	Tatumella	A	NSF	Bergey's Manual	
Erysipelatoclostridiaceae	Candidatus Stoquefichus	A	NSF	Bergey's Manual	Based on Erysipelatoclostridiaceae
Erysipelatoclostridiaceae	Catenibacterium	OA	NSF	Bergey's Manual	
Erysipelatoclostridiaceae	Coprobacillus	OA	NSF	Bergey's Manual	
Erysipelatoclostridiaceae	Erysipelatoclostridium	OA	NSF	Yutin & Galperin, 2013	
Erysipelatoclostridiaceae	Erysipelotrichaceae UCG-003	A	NSF	Bergey's Manual	Based on Erysipelatoclostridiaceae
Erysipelatoclostridiaceae	Other	A	NSF	Bergey's Manual	Based on Erysipelatoclostridiaceae
Erysipelotrichaceae	[Clostridium] innocuum group	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Erysipelotrichaceae	Dubosiella	OA	NSF	Cox et al., 2017	
Erysipelotrichaceae	Erysipelotrichaceae UCG-008	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Erysipelotrichaceae	Faecalibaculum	OA	NSF	Chang et al., 2015	
Erysipelotrichaceae	Faecalitalea	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Erysipelotrichaceae	Holdemanella	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Erysipelotrichaceae	Ileibacterium	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Erysipelotrichaceae	Other	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae

Erysipelotrichaceae	Turicibacter	A	NSF	Bergey's Manual	
Erysipelotrichaceae	ZOR0006	A	NSF	Bergey's Manual	Based on Erysipelotrichaceae
Eubacteriaceae	Other	OA	NSF	Bergey's Manual	Based on Eubacteriaceae
Euzebyaceae	Other	Unknown	Unknown		
Exiguobacteraceae	Exiguobacterium	A	NSF	Bergey's Manual	
Family XI	Other	Unknown	Unknown		
Family XI	Tissierella	OA	NSF	Bergey's Manual	
Flavobacteriaceae	Aequorivita	A	NSF	Bergey's Manual	
Flavobacteriaceae	Aquibacter	A	NSF	Bergey's Manual	Based on Flavobacteriaceae
Flavobacteriaceae	Arenibacter	A	NSF	Bergey's Manual	
Flavobacteriaceae	Aurantiacicella	A	NSF	Bergey's Manual	Based on Flavobacteriaceae
Flavobacteriaceae	Flavobacterium	A	NSF	Bergey's Manual	
Flavobacteriaceae	Gelidibacter	A	NSF	Bergey's Manual	
Flavobacteriaceae	Gillisia	A	NSF	Bergey's Manual	
Flavobacteriaceae	Imtechella	A	NSF	Bergey's Manual	Based on Flavobacteriaceae
Flavobacteriaceae	Leeuwenhoekiella	A	Unknown	Bergey's Manual	
Flavobacteriaceae	Muricauda	A	NSF	Bergey's Manual	
Flavobacteriaceae	Myroides	A	NSF	Bergey's Manual	
Flavobacteriaceae	Other	A	NSF	Bergey's Manual	Based on Flavobacteriaceae
Flavobacteriaceae	Subsaxibacter	A	NSF	Bergey's Manual	
Frankiaceae	Frankia	A	SF	Bergey's Manual	
Frankiaceae	Jatrophihabitans	A	NSF	Bergey's Manual	
Fusobacteriaceae	Cetobacterium	A	NSF	Bergey's Manual	
Fusobacteriaceae	Fusobacterium	OA	NSF	Bergey's Manual	
Gaiellaceae	Gaiella	A	NSF	Bergey's Manual	
Garciellaceae	Rhabdanaerobium	OA	SF	Bergey's Manual	
Gemellaceae	Gemella	A	NSF	Bergey's Manual	
Geminicoccaceae	Candidatus Alysiosphaera	Unknown	Unknown		
Geminicoccaceae	Geminicoccus	Unknown	Unknown		
Geminicoccaceae	Other	Unknown	Unknown		
Gemmataceae	Fimbriiglobus	A	Unknown	Bergey's Manual	
Gemmataceae	Gemmata	A	Unknown	Bergey's Manual	

Gemmataceae	Other	A	Unknown	Bergey's Manual	Based on Gemmataceae
Gemmatimonadaceae	Gemmatimonas	A	NSF	Bergey's Manual	
Gemmatimonadaceae	Other	Unknown	Unknown		
Geodermatophilaceae	Antricoccus	A	SF	Bergey's Manual	Based on Geodermatophilaceae
Geodermatophilaceae	Blastococcus	A	NSF	Bergey's Manual	
Geodermatophilaceae	Klenkia	A	NSF	Bergey's Manual	
Geodermatophilaceae	Modestobacter	A	NSF	Bergey's Manual	
Gottschalkia	Other	Unknown	Unknown		
Granulosicoccaceae	Granulosicoccus	Unknown	Unknown		
Hafniaceae	Edwardsiella	A	Unknown	Bergey's Manual	
Hafniaceae	Hafnia-Obesumbacterium	A	Unknown	Bergey's Manual	Based on Hafniaceae
Halomonadaceae	Chromohalobacter	A	NSF	Bergey's Manual	
Halomonadaceae	Halomonas	A	NSF	Bergey's Manual	
Halomonadaceae	Salinicola	A	NSF	Bergey's Manual	
Helicobacteraceae	Helicobacter	A	NSF	Bergey's Manual	
Hydrogenoanaerobacterium	Other	Unknown	Unknown		
Hyphomicrobiaceae	Hyphomicrobium	A	NSF	Bergey's Manual	
Hyphomicrobiaceae	Pedomicrobium	A	Unknown	Bergey's Manual	
Iamiaceae	Iamia	A	Unknown	Bergey's Manual	
Ilumatobacteraceae	CL500-29 marine group	Unknown	Unknown		
Ilumatobacteraceae	Ilumatobacter	Unknown	Unknown		
Ilumatobacteraceae	Other	Unknown	Unknown		
Intrasporangiaceae	Humibacillus	A	NSF	Bergey's Manual	
Intrasporangiaceae	Intrasporangium	A	NSF	Bergey's Manual	
Intrasporangiaceae	Janibacter	A	NSF	Bergey's Manual	
Intrasporangiaceae	Knoellia	A	NSF	Bergey's Manual	
Intrasporangiaceae	Oryzihumus	A	NSF	Bergey's Manual	
Intrasporangiaceae	Other	A	NSF	Bergey's Manual	Based on Intrasporangiaceae
Intrasporangiaceae	Pedococcus-Phycicoccus	A	NSF	Bergey's Manual	Based on Intrasporangiaceae
Isosphaeraceae	Aquisphaera	A	Unknown	Bergey's Manual	

Isosphaeraceae	Candidatus Nostocoida	A	Unknown	Bergey's Manual	Based on Isosphaeraceae
Isosphaeraceae	Isosphaera	A	Unknown	Bergey's Manual	
Isosphaeraceae	Other	A	Unknown	Bergey's Manual	Based on Isosphaeraceae
Isosphaeraceae	Paludisphaera	A	Unknown	Bergey's Manual	
Isosphaeraceae	Singulisphaera	A	Unknown	Bergey's Manual	
Isosphaeraceae	Tundrisphaera	A	Unknown	Bergey's Manual	Based on Isosphaeraceae
JG30-KF-CM45	Other	Unknown	Unknown		
Kaistiaceae	Kaistia	Unknown	Unknown		
Kineosporiaceae	Angustibacter	A	NSF	Bergey's Manual	
Kineosporiaceae	Other	Unknown	Unknown		
Kineosporiaceae	Quadrisphaera	Unknown	NSF	Bergey's Manual	
Ktedonobacteraceae	G12-WMSP1	A	SF	Bergey's Manual	Based on Ktedonobacteraceae
Ktedonobacteraceae	HSB OF53-F07	A	SF	Bergey's Manual	Based on Ktedonobacteraceae
Labraceae	Labrys	A	Unknown	Bergey's Manual	
Lachnospiraceae	28-Apr	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Acetivibrio] ethanolgignens group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] eligens group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] fissicatena group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] hallii group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] oxidoreducens group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] ventriosum group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Eubacterium] xylanophilum group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Ruminococcus] gauvreauii group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	[Ruminococcus] gnavus group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae

Lachnospiraceae	[Ruminococcus] torques group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	A2	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Acetatifactor	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Agathobacter	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Anaerosporobacter	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Anaerostipes	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	ASF356	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Blautia	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Cellulosilyticum	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Coprococcus	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Cuneatibacter	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Dorea	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Eisenbergiella	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Epulopiscium	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Fusicatenibacter	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	GCA-900066575	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Hungatella	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnoclostridium	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae FCS020 group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae NK4A136 group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae NK4B4 group	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae UCG-001	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae UCG-002	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae

Lachnospiraceae	Lachnospiraceae UCG-004	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae UCG-006	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae UCG-008	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Lachnospiraceae UCG-010	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Marvinbryantia	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Murimonas	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Other	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	possible genus Sk018	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Robinsoniella	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Roseburia	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Sellimonas	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Tuzzerella	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lachnospiraceae	Tyzzerella	OA	Unknown	Bergey's Manual	Based on Lachnospiraceae
Lactobacillaceae	Agrilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Bombilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Companilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Dellaglioa	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	HT002	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Lacticaseibacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Lactiplantibacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Lactobacillus	A	NSF	Bergey's Manual	
Lactobacillaceae	Latilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Leuconostoc	Unknown	NSF	Bergey's Manual	
Lactobacillaceae	Levilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae

Lactobacillaceae	Ligilactobacillus	A	Mixed	Marta et al., 2021; Zheng et al., 2020; Mozota et al., 2021	
Lactobacillaceae	Limosilactobacillus	Unknown	Unknown	Zheng et a., 2020	
Lactobacillaceae	Other	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Paucilactobacillus	Unknown	NSF	Bergey's Manual	Based on Lactobacillaceae
Lactobacillaceae	Weissella	A	NSF	Bergey's Manual	
Legionellaceae	Legionella	A	SF	Bergey's Manual	
Leptolyngbyaceae	Leptolyngbya PCC-6306	Unknown	Unknown		
Listeriaceae	Listeria	A	NSF	Bergey's Manual	Based on Listeriaceae
Marinifilaceae	Butyricimonas	Unknown	Unknown		
Marinifilaceae	Odoribacter	OA	unknown	Hardham et al., 2008	
Marinifilaceae	Other	Unknown	Unknown		
Marinifilaceae	Sanguibacteroides	Unknown	Unknown		
Marinilabiliaceae	Natronoflexus	Unknown	Unknown	Bergey's Manual	Based on Marinilabiliaceae
Marinilabiliaceae	Other	Unknown	Unknown	Bergey's Manual	Based on Marinilabiliaceae
Marinobacteraceae	Marinobacter	A	NSF	Bergey's Manual	
Methyloligellaceae	Methyloligella	Unknown	Unknown		
Methyloligellaceae	Other	Unknown	Unknown		
Microbacteriaceae	Agrococcus	A	NSF	Bergey's Manual	
Microbacteriaceae	Amnibacterium	A	NSF	Bergey's Manual	
Microbacteriaceae	Curtobacterium	A	SF	Bergey's Manual	
Microbacteriaceae	Frigoribacterium	A	NSF	Bergey's Manual	
Microbacteriaceae	Homoserinibacter	A	NSF	Bergey's Manual	
Microbacteriaceae	Leucobacter	Unknown	NSF	Bergey's Manual	
Microbacteriaceae	Lysinimonas	A	NSF	Bergey's Manual	
Microbacteriaceae	Microbacterium	A	NSF	Bergey's Manual	
Microbacteriaceae	Microterricola	A	Unknown	Bergey's Manual	
Microbacteriaceae	Mycetocola	A	NSF	Bergey's Manual	
Microbacteriaceae	Other	A	NSF	Bergey's Manual	Based on Microbacteriaceae

Microbacteriaceae	Parafrigoribacterium	A	NSF	Bergey's Manual	Based on Microbacteriaceae
Microbacteriaceae	Plantibacter	A	NSF	Bergey's Manual	
Microbacteriaceae	Pseudoclavibacter	A	NSF	Bergey's Manual	
Micrococcaceae	Arthrobacter	A	NSF	Bergey's Manual	
Micrococcaceae	Glutamicibacter	A	NSF	Bergey's Manual	
Micrococcaceae	Kocuria	A	NSF	Bergey's Manual	
Micrococcaceae	Nesterenkonia	A	NSF	Bergey's Manual	
Micrococcaceae	Other	Unknown	Unknown		
Micrococcaceae	Paeniglutamicibacter	A	NSF	Bergey's Manual	
Micrococcaceae	Pseudarthrobacter	A	NSF	Bergey's Manual	
Micromonosporaceae	Actinoplanes	A	SF	Bergey's Manual	
Micromonosporaceae	Micromonospora	A	SF	Bergey's Manual	
Micromonosporaceae	Other	A	SF	Bergey's Manual	Based on Micromonosporaceae
Micromonosporaceae	Xiangella	A	Unknown	Bergey's Manual	
Monoglobaceae	Monoglobus	OA	NSF	Kim et al., 2017	
Moraxellaceae	Acinetobacter	A	NSF	Bergey's Manual	
Moraxellaceae	Alkanindiges	A	NSF	Bergey's Manual	Based on Moraxellaceae
Moraxellaceae	Enhydrobacter	A	NSF	Bergey's Manual	
Moraxellaceae	Psychrobacter	A	Unknown	Bergey's Manual	
Morganellaceae	Cosenzaea	Unknown	Unknown		
Morganellaceae	Moellerella	A	Unknown	Bergey's Manual	
Morganellaceae	Morganella	A	NSF	Bergey's Manual	
Morganellaceae	Other	Unknown	Unknown		
Morganellaceae	Proteus	A	Unknown	Bergey's Manual	
Morganellaceae	Providencia	A	Unknown	Bergey's Manual	
Muribaculaceae	Muribaculum	OA	NSF	Bergey's Manual	
Muribaculaceae	Other	OA	Unknown	Lagkouvardos et al., 2019	
MWH-CFBk5	Other	Unknown	Unknown		
Mycobacteriaceae	Mycobacterium	A	NSF	Bergey's Manual	
Mycobacteriaceae	Other	Unknown	Unknown		
Mycoplasmataceae	Candidatus Bacilloplasma	A	Unknown	Bergey's Manual	Based on Mycoplasmataceae
Mycoplasmataceae	Mycoplasma	Unknown	Unknown	Bergey's Manual	

Mycoplasmataceae	Other	A	Unknown	Bergey's Manual	Based on Mycoplasmataceae
Myxococcaceae	P3OB-42	Unknown	Unknown		
Nakamurellaceae	Nakamurella	A	NSF	Bergey's Manual	
Nannocystaceae	Enhygromyxa	Unknown	Unknown		
Neisseriaceae	Other	A	Unknown	Bergey's Manual	Based on Neisseriaceae
Neisseriaceae	Vitreoscilla	A	NSF	Bergey's Manual	
Nitrospiraceae	Nitrospira	A	Unknown	Bergey's Manual	
Nocardiaceae	Gordonia	A	NSF	Bergey's Manual	
Nocardiaceae	Nocardia	A	SF	Bergey's Manual	
Nocardiaceae	Other	A	Unknown	Bergey's Manual	Based on Nocardiaceae
Nocardiaceae	Rhodococcus	A	Unknown	Bergey's Manual	
Nocardiaceae	Williamsia	A	NSF	Bergey's Manual	
Nocardioidaceae	Aeromicrobium	A	NSF	Bergey's Manual	
Nocardioidaceae	Marmoricola	A	NSF	Bergey's Manual	
Nocardioidaceae	Mumia	A	SF	Bergey's Manual	Based on Nocardioidaceae
Nocardioidaceae	Nocardioides	A	NSF	Bergey's Manual	
Nocardioidaceae	Other	A	Unknown	Bergey's Manual	Based on Nocardioidaceae
Nocardiopsaceae	Nocardiopsis	A	SF	Bergey's Manual	
Nostocaceae	Calothrix PCC-6303	Unknown	Unknown		
Nostocaceae	Other	Unknown	Unknown		
Nostocaceae	Rivularia PCC-7116	Unknown	Unknown	Bergey's Manual	Based on Rivularia
Oligoflexaceae	Oligoflexus	Unknown	Unknown		
Oscillospiraceae	Colidextribacter	OA	NSF	Ricaboni et al., 2017	
Oscillospiraceae	Flavonifractor	OA	SF	Bergey's Manual	Based on Eubacterium
Oscillospiraceae	Intestinimonas	OA	SF	Kläring et al., 2013, Bergey's manual	
Oscillospiraceae	NK4A214 group	OA	Unknown	Tindal et al., 2019	
Oscillospiraceae	Oscillibacter	OA	NSF	Iino et al., 2007	
Oscillospiraceae	Oscillospira	OA	SF	Bergey's Manual	
Oscillospiraceae	Other	A	Unknown	Bergey's Manual	Based on Oscillospiraceae

Oscillospiraceae	Papillibacter	OA	NSF	Bergey's Manual	
Oscillospiraceae	Pseudoflavonifractor	Unknown	Unknown		
Oscillospiraceae	UCG-002	Unknown	Unknown		
Oscillospiraceae	UCG-003	Unknown	Unknown		
Oscillospiraceae	UCG-005	OA	Unknown	Bergey's Manual	Based on Oscillospiraceae
Oscillospiraceae	UCG-007	Unknown	Unknown		
Oxalobacteraceae	Duganella	A	NSF	Bergey's Manual	
Oxalobacteraceae	Massilia	A	NSF	Bergey's Manual	
Oxalobacteraceae	Noviherbaspirillum	A	Unknown	Bergey's Manual	Based on Oxalobacteraceae
Oxalobacteraceae	Oxalicibacterium	A	Unknown	Bergey's Manual	Based on Oxalobacteraceae
Oxalobacteraceae	Oxalobacter	OA	NSF	Bergey's Manual	
Oxalobacteraceae	Paraherbaspirillum	A	Unknown	Bergey's Manual	Based on Oxalobacteraceae
Paenibacillaceae	Ammoniphilus	A	NSF	Bergey's Manual	
Paenibacillaceae	Paenibacillus	A	SF	Bergey's Manual	
Paludibacteraceae	H1	Unknown	Unknown		
Pasteurellaceae	Conservatibacter	Unknown	Unknown	Bergey's Manual	
Pasteurellaceae	Mesocricetibacter	A	Unknown	Bergey's Manual	
Pasteurellaceae	Muribacter	Unknown	Unknown	Bergey's Manual	
Pasteurellaceae	Rodentibacter	A	unknown	Bergey's Manual, Benga, Sager & Christensen 2018	
Pectobacteriaceae	Nissabacter	Unknown	Unknown		
Peptococcaceae	Other	OA	Unknown	Bergey's Manual	Based on Peptococcaceae
Peptococcaceae	Peptococcus	OA	NSF	Bergey's Manual	
Peptostreptococcaceae	[Eubacterium] tenue group	Unknown	Unknown		
Peptostreptococcaceae	Clostridioides	A	Unknown	Bergey's Manual	Based on Peptostreptococcaceae
Peptostreptococcaceae	Intestinibacter	A	Unknown	Bergey's Manual	Based on Peptostreptococcaceae
Peptostreptococcaceae	Paeniclostridium	A	Unknown	Bergey's Manual	Based on Peptostreptococcaceae

Peptostreptococcaceae	Paraclostridium	A	SF	Bergey's Manual	
Peptostreptococcaceae	Romboutsia	OA	Unknown	Gerritsen et al., BioRxiv 2019	
Peptostreptococcaceae	Sporacetigenium	OA	NSF	Bergey's Manual	
Peptostreptococcaceae	Terrisporobacter	A	Unknown	Bergey's Manual	Based on Peptostreptococcaceae
Phormidesmiaceae	Phormidesmis ANT.LACV5.1	Unknown	Unknown		
Phormidiaceae	Other	Unknown	Unknown		
Phormidiaceae	Tychonema CCAP 1459-11B	Unknown	Unknown		
Pirellulaceae	Blastopirellula	A	NSF	Bergey's Manual	
Pirellulaceae	Bythopirellula	Unknown	Unknown		
Pirellulaceae	Other	Unknown	Unknown		
Pirellulaceae	Pir4 lineage	Unknown	Unknown	Bergey's Manual	Based on Pirellulaceae
Pirellulaceae	Pirellula	A	Unknown	Bergey's Manual	
Pirellulaceae	Rhodopirellula	A	NSF	Bergey's Manual	
Pirellulaceae	Rubripirellula	A	Unknown	Bergey's Manual	
Planococcaceae	Kurthia	A	NSF	Bergey's Manual	
Planococcaceae	Lysinibacillus	A	SF	Bergey's Manual	
Planococcaceae	Other	Unknown	Unknown		
Planococcaceae	Paenisporosarcina	A	SF	Krish murthi et al., 2009	
Planococcaceae	Planomicrobium	A	NSF	Bergey's Manual	
Planococcaceae	Psychrobacillus	A	Unknown	Bergey's Manual	Based on Planococcaceae
Planococcaceae	Solibacillus	A	SF	Krish murthi et al., 2009	
Planococcaceae	Sporosarcina	A	SF	Bergey's Manual	
Polyangiaceae	Aetherobacter	A	Unknown	Garcia et al. 2016	
Porphyromonadaceae	Falsiporphyromonas	OA	Unknown	Bergey's Manual	Based on Porphyromonadaceae
Porphyromonadaceae	Other	OA	Unknown	Bergey's Manual	Based on Porphyromonadaceae
Prevotellaceae	Alloprevotella	OA	NSF	Bergey's Manual	Based on Prevotellaceae
Prevotellaceae	Other	A	NSF	Bergey's Manual	Based on Prevotellaceae

Prevotellaceae	Paraprevotella	OA	NSF	Bergey's Manual	Based on Prevotellaceae
Prevotellaceae	Prevotella_7	A	NSF	Bergey's Manual	Based on Prevotella
Prevotellaceae	Prevotella_9	A	NSF		
Prevotellaceae	Prevotellaceae UCG-001	A	NSF	Bergey's Manual	Based on Prevotellaceae
Prevotellaceae	Prevotellaceae UCG-004	A	NSF	Bergey's Manual	Based on Prevotellaceae
Promicromonosporaceae	Isoptericola	A	NSF	Bergey's Manual	
Propionibacteriaceae	Friedmanniella	A	NSF	Bergey's Manual	
Propionibacteriaceae	Marinilutecoccus	Unknown	Unknown		
Propionibacteriaceae	Microlunatus	A	NSF	Bergey's Manual	
Propionibacteriaceae	Other	Unknown	Unknown		
Propionibacteriaceae	Tessaracoccus	A	NSF	Bergey's Manual	
Pseudomonadaceae	Pseudomonas	A	Unknown	Bergey's Manual	
Pseudonocardiaceae	Actinomycetospora	A	SF	Bergey's Manual	
Pseudonocardiaceae	Actinophytocola	A	Unknown	Bergey's Manual	Based on Pseudonocardiaceae
Pseudonocardiaceae	Crossiella	A	Unknown	Bergey's Manual	Based on Pseudonocardiaceae
Pseudonocardiaceae	Kibdelosporangium	A	Unknown	Bergey's Manual	
Pseudonocardiaceae	Other	A	Unknown	Bergey's Manual	Based on Pseudonocardiaceae
Pseudonocardiaceae	Pseudonocardia	A	SF	Bergey's Manual	
Puniceicoccaceae	Cerasicoccus	A	Unknown	Bergey's Manual	
Reyranellaceae	Reyranella	Unknown	Unknown		
Rhizobiaceae	Ahrensia	A	NSF	Bergey's Manual	
Rhizobiaceae	Aliihoeflea	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium	A	NSF	Bergey's Manual	Based on Rhizobium and Allorhizobium
Rhizobiaceae	Aminobacter	A	NSF	Bergey's Manual	
Rhizobiaceae	Aquamicrobium	Unknown	NSF	Bergey's Manual	
Rhizobiaceae	Aurantimonas	A	Unknown	Bergey's Manual	Based on Rhizobiaceae

Rhizobiaceae	Aureimonas	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Brucella	A	NSF	Bergey's Manual	
Rhizobiaceae	Corticibacterium	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Falsochrobactrum	Unknown	NSF	Bergey's Manual	
Rhizobiaceae	Hoeflea	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Jiella	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Mesorhizobium	A	NSF	Bergey's Manual	
Rhizobiaceae	Neorhizobium	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Ochrobactrum	A	Unknown	Bergey's Manual	
Rhizobiaceae	Other	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Paenochrobactrum	A	NSF	Bergey's Manual	
Rhizobiaceae	Phyllobacterium	A	Unknown	Bergey's Manual	
Rhizobiaceae	Pseudaminobacter	A	Unknown	Bergey's Manual	
Rhizobiaceae	Shinella	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhizobiaceae	Tianweitania	A	Unknown	Bergey's Manual	Based on Rhizobiaceae
Rhodanobacteraceae	Chujaibacter	Unknown	Unknown		
Rhodanobacteraceae	Dokdonella	Unknown	Unknown		
Rhodanobacteraceae	Luteibacter	Unknown	Unknown		
Rhodanobacteraceae	Mizugakiibacter	Unknown	Unknown		
Rhodanobacteraceae	Oleiagrimonas	Unknown	Unknown		
Rhodanobacteraceae	Rhodanobacter	A	NSF	Bergey's Manual	
Rhodobacteraceae	Actibacterium	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Albirhodobacter	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Amaricoccus	A	NSF	Bergey's Manual	
Rhodobacteraceae	Boseongicola	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Defluviimonas	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Falsirhodobacter	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae

Rhodobacteraceae	Gemmobacter	A	NSF	Bergey's Manual	
Rhodobacteraceae	Jannaschia	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Limibaculum	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Maribius	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Oceaniovalibus	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Octadecabacter	A	Unknown	Bergey's Manual	
Rhodobacteraceae	Other	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Paenirhodobacter	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Paracoccus	A	NSF	Bergey's Manual	
Rhodobacteraceae	Plastorhodobacter	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Pseudorhodobacter	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Pseudoruegeria	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Rhodobacter	Unknown	Unknown	Bergey's Manual	
Rhodobacteraceae	Rhodobaculum	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Roseivivax	A	Unknown	Bergey's Manual	
Rhodobacteraceae	Roseovarius	A	Unknown	Bergey's Manual	
Rhodobacteraceae	Rubellimicrobium	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Sulfitobacter	A	NSF	Bergey's Manual	
Rhodobacteraceae	Thioclava	A	Unknown	Bergey's Manual	Based on Rhodobacteraceae
Rhodobacteraceae	Yoonia-Loktanella	A	NSF	Bergey's Manual	Based on Yoonia
Rhodocyclaceae	Azovibrio	A	Unknown	Bergey's Manual	
Rhodocyclaceae	Thauera	A	NSF	Bergey's Manual	
Rhodomicrobiaceae	Rhodomicrobium	A	Unknown	Bergey's Manual	
Rhodothermaceae	Rubrivirga	A	Unknown	Bergey's Manual	Based on Rhodothermaceae
Rickettsiaceae	Rickettsia	Unknown	Unknown	Bergey's Manual	
Rikenellaceae	Alistipes	OA	NSF	Bergey's Manual	
Rikenellaceae	Other	Unknown	Unknown		
Rikenellaceae	Rikenella	OA	NSF	Bergey's Manual	

Rikenellaceae	Rikenellaceae RC9 gut group	A	unknown	Bergey's Manual	Based on Rikenellaceae
Rs-E47 termite group	Other	Unknown	Unknown		
Rubinisphaeraceae	Planctomicrobium	Unknown	Unknown		
Rubinisphaeraceae	SH-PL14	Unknown	Unknown		
Rubritaleaceae	Luteolibacter	A	Unknown	Bergey's Manual	Based on Rubritaleaceae
Ruminococcaceae	[Eubacterium] siraeum group	OA	NSF	Bergey's Manual	Based on Eubacterium
Ruminococcaceae	Anaerotruncus	OA	SF	Bergey's Manual	
Ruminococcaceae	Angelakisella	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Candidatus Soleaferrea	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Caproiciproduce ns	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	DTU089	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Faecalibacterium	OA	NSF	Bergey's Manual	
Ruminococcaceae	Fournierella	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Harryflintia	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Incertae Sedis	OA	SF	Bergey's Manual, Browne et al., 2016	Based on Ruminococcaceae
Ruminococcaceae	Negativibacillus	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Other	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Paludicola	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Pygmaiobacter	OA	Unknown	Bergey's Manual	Based on Ruminococcaceae
Ruminococcaceae	Ruminococcus	OA	NSF	Bergey's Manual, Browne et al., 2016	
Ruminococcaceae	Subdoligranulum	OA	NSF	Bergey's Manual	
Ruminococcaceae	UBA1819	OA	NSF	Bergey's Manual	Based on Faecalibacterium
Salinisphaeraceae	Salinisphaera	A	NSF	Bergey's Manual	
Sandaracinaceae	Other	Unknown	Unknown		
Sanguibacteraceae	Sanguibacter-Flavimobilis	Unknown	NSF	Bergey's Manual	Based on Sanguibacter

Saprospiraceae	Lewinella	A	NSF	Bergey's Manual	Based on Saprospiraceae
Saprospiraceae	Other	A	NSF	Bergey's Manual	Based on Saprospiraceae
SC-I-84	Other	Unknown	Unknown		
Schlesneriaceae	Planctopirus	Unknown	Unknown		
Schlesneriaceae	Schlesneria	A	Unknown	Bergey's Manual	
Shewanellaceae	Shewanella	A	NSF	Bergey's Manual	
Solirubrobacteraceae	Conexibacter	A	NSF	Bergey's Manual	
Solirubrobacteraceae	Other	Unknown	Unknown	Bergey's Manual	Based on Solirubrobacteraceae
Solirubrobacteraceae	Parviterribacter	Unknown	Unknown		
Solirubrobacteraceae	Patulibacter	A	NSF	Bergey's Manual	
Solirubrobacteraceae	Solirubrobacter	A	NSF	Bergey's Manual	
Sphingobacteriaceae	Pedobacter	A	NSF	Bergey's Manual	
Sphingobacteriaceae	Sphingobacterium	A	NSF	Bergey's Manual	
Sphingomonadaceae	Altererythrobacter	Unknown	Unknown		
Sphingomonadaceae	Erythrobacter	A	Unknown	Bergey's Manual	
Sphingomonadaceae	Novosphingobium	A	NSF	Takeuchi et al., 2001, Bergey's manual	
Sphingomonadaceae	Qipengyuania	Unknown	Unknown		
Sphingomonadaceae	Sphingomonas	A	NSF	Bergey's Manual	
Sphingomonadaceae	Sphingopyxis	A	NSF	Bergey's Manual	
Spirosomaceae	Persicitalea	Unknown	Unknown		
Spirosomaceae	Rhabdobacter	Unknown	Unknown		
Sporichthyaceae	Longivirga	A	Unknown	Bergey's Manual	Based on Sporichthyaceae
Sporichthyaceae	Other	A	Unknown	Bergey's Manual	Based on Sporichthyaceae
Sporomusaceae	Dendrosporobacter	OA	SF	Bergey's Manual	
Sporomusaceae	Other	Unknown	Unknown		
Sporomusaceae	Sporomusa	A	SF	Bergey's Manual	
Staphylococcaceae	Corticicoccus	Unknown	Unknown		
Staphylococcaceae	Jeotgalicoccus	A	NSF	Bergey's Manual	
Staphylococcaceae	Macrococcus	Unknown	Unknown		
Staphylococcaceae	Other	Unknown	Unknown		

Staphylococcaceae	Staphylococcus	A	NSF	Bergey's Manual	
Stappiaceae	Labrenzia	Unknown	Unknown		
Stappiaceae	Other	Unknown	Unknown		
Streptococcaceae	Lactococcus	A	NSF	Bergey's Manual	
Streptococcaceae	Other	A	NSF	Bergey's Manual	Based on Streptococcaceae
Streptococcaceae	Streptococcus	A	NSF	Bergey's Manual	
Streptomycetaceae	Kitasatospora	A	SF	Bergey's Manual	Based on Streptomycetaceae
Streptomycetaceae	Other	A	SF	Bergey's Manual	Based on Streptomycetaceae
Streptomycetaceae	Streptomyces	A	SF	Bergey's Manual	
Streptosporangiaceae	Streptosporangium	A	SF	Bergey's Manual	
Sulfobacillaceae	Other	Unknown	Unknown		
Sumerlaeaceae	Sumerlaea	Unknown	Unknown		
Sutterellaceae	Other	Unknown	Unknown		
Sutterellaceae	Parasutterella	OA	NSF	Nagai et al., 2009	
Sutterellaceae	Sutterella	A	Unknown	Bergey's Manual	
Tannerellaceae	Candidatus Vestibaculum	Unknown	Unknown		
Tannerellaceae	Macellibacteroides	Unknown	Unknown		
Tannerellaceae	Other	Unknown	Unknown		
Tannerellaceae	Parabacteroides	OA	NSF	Sakamoto & Benno, 2006	
Thermoactinomycetaceae	Other	A	SF	Bergey's Manual	Based on Thermoactinomycetaceae
Thermoactinomycetaceae	Risungbinella	A	SF	Bergey's Manual	Based on Thermoactinomycetaceae
Trueperaceae	Truepera	A	NSF	Bergey's Manual	
Tsukamurellaceae	Tsukamurella	A	NSF	Bergey's Manual	
UCG-010	Other	Unknown	Unknown		
Vagococcaceae	Vagococcus	A	NSF	Bergey's Manual	
Veillonellaceae	Dialister	OA	NSF	Bergey's Manual	
Veillonellaceae	Veillonella	A	NSF	Bergey's Manual	
Vibrionaceae	Vibrio	A	NSF	Bergey's Manual	
WD2101 soil group	Other	Unknown	Unknown		

Weeksellaceae	Candidatus Hemobacterium	Unknown	Unknown		
Weeksellaceae	Chishuiella	Unknown	Unknown		
Weeksellaceae	Chryseobacterium	A	NSF	Bergey's Manual	
Weeksellaceae	Empedobacter	A	NSF	Bergey's Manual	
Wohlfahrtiimonadaceae	Ignatzschineria	Unknown	Unknown		
Wohlfahrtiimonadaceae	Wohlfahrtiimonas	Unknown	Unknown		
Xanthobacteraceae	Afipia	A	Unknown		
Xanthobacteraceae	Bradyrhizobium	A	NSF	Bergey's Manual	
Xanthobacteraceae	Other	Unknown	Unknown		
Xanthobacteraceae	Pseudolabrys	Unknown	Unknown		
Xanthobacteraceae	Pseudorhodoplanes	Unknown	Unknown		
Xanthobacteraceae	Rhodoplanes	A	NSF	Bergey's Manual	
Xanthobacteraceae	Rhodopseudomonas	A	Unknown	Bergey's Manual	
Xanthomonadaceae	Luteimonas	A	Unknown	Bergey's Manual	
Xanthomonadaceae	Lysobacter	A	Unknown	Bergey's Manual	
Xanthomonadaceae	Pseudoxanthomonas	A	NSF	Bergey's Manual	
Xanthomonadaceae	SN8	Unknown	Unknown		
Xanthomonadaceae	Stenotrophomonas	A	NSF	Bergey's Manual	
Xanthomonadaceae	Thermomonas	A	NSF	Bergey's Manual	
Xenococcaceae	Pleurocapsa PCC-7319	A	Unknown	Bergey's Manual	
Yersiniaceae	Other	Unknown	Unknown		
Yersiniaceae	Rahnella	A	Unknown	Bergey's Manual	
Yersiniaceae	Serratia	A	Unknown	Bergey's Manual	
Yersiniaceae	Yersinia	A	NSF	Bergey's Manual	

Supplementary Table 2. DNA sequence of an amplicon sequence variant (ASV) driving reduction in relative abundance of Proteobacteria in wild mice (Supplementary Figure 6B).

ACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGT
CGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTA
GAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCG
AAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAG
ATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCC
GGAGCTAACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTA

Supplementary references

Benga, L., Sager, M., & Christensen, H. (2018). From the [Pasteurella] pneumotropica complex to Rodentibacter spp.: an update on [Pasteurella] pneumotropica. In *Veterinary Microbiology* (Vol. 217, pp. 121–134). Elsevier B.V. https://doi.org/10.1016/j.vetmic.2018.03.011

Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D., & Lawley, T. D. (2016). Culturing of "unculturable" human microbiota reveals novel taxa and extensive sporulation. *Nature*, *533*(7604), 543–546. https://doi.org/10.1038/nature17645

Cai, S., & Dong, X. (2010). Cellulosilyticum ruminicola gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of Clostridium lentocellum as Cellulosilyticum lentocellum comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *60*(4), 845–849. https://doi.org/10.1099/ijs.0.014712-0

Chang DH, Rhee MS, Ahn S, et al. Faecalibaculum rodentium gen. nov., sp. nov., isolated from the faeces of a laboratory mouse [published correction appears in Antonie Van Leeuwenhoek. 2016 Mar;109(3):481]. *Antonie Van Leeuwenhoek*. 2015;108(6):1309-1318. doi:10.1007/s10482-015-0583-3

Clavel, T., Charrier, C., Wenning, M., & Haller, D. (2013). Parvibacter caecicola gen. nov., sp. nov., a bacterium of the family Coriobacteriaceae isolated from the caecum of a mouse. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART7), 2642–2648. https://doi.org/10.1099/ijs.0.045344-0

Clavel, T., Duck, W., Charrier, C., Wenning, M., Elson, C., & Haller, D. (2010). Enterorhabdus caecimuris sp. nov., a member of the family Coriobacteriaceae isolated from a mouse model of spontaneous colitis, and emended description of the genus Enterorhabdus Clavel et al. 2009. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 60, Issue 7, pp. 1527– 1531). Microbiology Society. https://doi.org/10.1099/ijs.0.015016-0

Cox LM, Sohn J, Tyrrell KL, et al. Description of two novel members of the family Erysipelotrichaceae: Ileibacterium valens gen. nov., sp. nov. and Dubosiella newyorkensis, gen. nov., sp. nov., from the murine intestine, and emendation to the description of Faecalibaculum rodentium [published correction appears in Int J Syst Evol Microbiol. 2017 Oct;67(10):4289]. *Int J Syst Evol Microbiol*. 2017;67(5):1247-1254. doi:10.1099/ijsem.0.001793

Garcia R, Stadler M, Gemperlein K, Müller R. *Aetherobacter fasciculatus* gen. nov., sp. nov. and *Aetherobacter rufus* sp. nov., novel myxobacteria with promising biotechnological applications. *Int J Syst Evol Microbiol*. 2016;66(2):928-938. doi:10.1099/ijsem.0.000813

Gerritsen, J., Hornung, B., Ritari, J., Paulin, L., Rijkers, G., Schaap, P., de Vos, W., & Smidt, H. (2019). A comparative and functional genomics analysis of the genus Romboutsia provides insight into adaptation to an intestinal lifestyle. *BioRxiv*, 845511. https://doi.org/10.1101/845511

Hardham JM, King KW, Dreier K, et al. Transfer of Bacteroides splanchnicus to Odoribacter gen. nov. as Odoribacter splanchnicus comb. nov., and description of Odoribacter denticanis sp. nov., isolated from the crevicular spaces of canine periodontitis patients. *Int J Syst Evol Microbiol*. 2008;58(Pt 1):103-109. doi:10.1099/ijs.0.63458-0

Iino, T., Mori, K., Tanaka, K., Suzuki, K. I., & Harayama, S. (2007). Oscillibacter valericigenes gen. nov., sp. nov., a valerate-producing anaerobic bacterium isolated from the alimentary canal of a Japanese corbicula clam. *International Journal of*

Systematic and Evolutionary Microbiology, *57*(8), 1840–1845. https://doi.org/10.1099/ijs.0.64717-0

Jeong, H., Lim, Y. W., Yi, H., Sekiguchi, Y., Kamagata, Y., & Chun, J. (2007). Anaerosporobacter mobilis gen. nov., sp. nov., isolated from forest soil. *International Journal of Systematic and Evolutionary Microbiology*, *57*(8), 1784–1787. https://doi.org/10.1099/ijs.0.63283-0

Johnson DB, Stallwood B, Kimura S, Hallberg KB. Isolation and characterization of Acidicaldus organivorus, gen. nov., sp. nov.: a novel sulfur-oxidizing, ferric iron-reducing thermo-acidophilic heterotrophic Proteobacterium. *Arch Microbiol*. 2006;185(3):212-221. doi:10.1007/s00203-006-0087-7

Kanso, S., & Patel, B. K. C. (2003). Microvirga subterranea gen. nov., sp. nov., a moderate thermophile from a deep subsurface Australian thermal aquifer. *International Journal of Systematic and Evolutionary Microbiology*, *53*(2), 401–406. https://doi.org/10.1099/ijs.0.02348- 0

Kim, B. C., Jeon, B. S., Kim, S., Kim, H., Um, Y., & Sang, B. I. (2015). Caproiciproducens galactitolivorans gen. Nov., sp. nov., a bacterium capable of producing caproic acid from galactitol, isolated from a wastewater treatment plant. *International Journal of Systematic and Evolutionary Microbiology*, *65*(12), 4902–4908. https://doi.org/10.1099/ijsem.0.000665

Kläring, K., Hanske, L., Bui, N., Charrier, C., Blaut, M., Haller, D., Plugge, C. M., & Clavel, T. (2013). Intestinimonas butyriciproducens gen. nov., sp. nov., a butyrate-producing bacterium from the mouse intestine. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART 12), 4606–4612. https://doi.org/10.1099/ijs.0.051441-0

Krishnamurthi, S., Bhattacharya, A., Mayilraj, S., Saha, P., Schumann, P., & Chakrabarti, T. (2009). Description of Paenisporosarcina quisquiliarum gen. nov., sp. nov., and reclassification of Sporosarcina macmurdoensis Reddy et al. 2003 as Paenisporosarcina macmurdoensis comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *59*(6), 1364–1370. https://doi.org/10.1099/ijs.0.65130-0

Lagkouvardos, I., Lesker, T. R., Hitch, T. C. A., Gálvez, E. J. C., Smit, N., Neuhaus, K., Wang, J., Baines, J. F., Abt, B., Stecher, B., Overmann, J., Strowig, T., & Clavel, T. (2019). Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome*, *7*(1), 28. https://doi.org/10.1186/s40168- 019-0637-2

Liu X, Mao B, Gu J, et al. *Blautia*-a new functional genus with potential probiotic properties?. *Gut Microbes*. 2021;13(1):1-21. doi:10.1080/19490976.2021.1875796

Mediannikov, O., Sekeyová, Z., Birg, M.-L., & Raoult, D. (2010). A Novel Obligate Intracellular Gamma-Proteobacterium Associated with Ixodid Ticks, Diplorickettsia massiliensis, Gen. Nov., Sp. Nov. *PLoS ONE*, *5*(7), e11478. https://doi.org/10.1371/journal.pone.0011478

Morotomi M, Nagai F, Watanabe Y. Description of Christensenella minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. *Int J Syst Evol Microbiol*. 2012;62(Pt 1):144-149. doi:10.1099/ijs.0.026989-0

Mozota M, Castro I, Gómez-Torres N, et al. Administration of *Ligilactobacillus salivarius* MP101 in an Elderly Nursing Home during the COVID-19 Pandemic: Immunological and Nutritional Impact. *Foods*. 2021;10(9):2149. Published 2021 Sep 11. doi:10.3390/foods10092149

Nagai, F., Morotomi, M., Sakon, H., & Tanaka, R. (2009). Parasutterella excrementihominis gen. nov., sp. nov., a member of the family Alcaligenaceae isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, *59*(7), 1793–1797. https://doi.org/10.1099/ijs.0.002519-0

Pfeiffer, N., Desmarchelier, C., Blaut, M., Daniel, H., Haller, D., & Clavel, T. (2012). Acetatifactor muris gen. nov., sp. nov., a novel bacterium isolated from the intestine of an obese mouse. *Archives of Microbiology*, *194*(11), 901–907. https://doi.org/10.1007/s00203-012-0822-1

Ricaboni D, Mailhe M, Cadoret F, Vitton V, Fournier PE, Raoult D. *'Colidextribacter massiliensis'* gen. nov., sp. nov., isolated from human right colon. *New Microbes New Infect*. 2016;17:27-29. Published 2016 Nov 28. doi:10.1016/j.nmni.2016.11.023

Rosero JA, Killer J, Sechovcová H, et al. Reclassification of *Eubacterium rectale* (Hauduroy *et al.* 1937) Prévot 1938 in a new genus *Agathobacter* gen. nov. as *Agathobacter rectalis* comb. nov., and description of *Agathobacter ruminis* sp. nov., isolated from the rumen contents of sheep and cows. *Int J Syst Evol Microbiol*. 2016;66(2):768-773. doi:10.1099/ijsem.0.000788

Sakamoto M, Benno Y. Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. *Int J Syst Evol Microbiol*. 2006;56(Pt 7):1599-1605. doi:10.1099/ijs.0.64192-0

Schnupf, P., Gaboriau-Routhiau, V., Gros, M., Friedman, R., Moya-Nilges, M., Nigro, G., Cerf- Bensussan, N., & Sansonetti, P. J. (2015). Growth and host interaction of mouse segmented filamentous bacteria in vitro. *Nature*, *520*(7545), 99–103. https://doi.org/10.1038/nature14027

Takeuchi, M., Hamana, K., & Hiraishi, A. (2001). Proposal of the genus Sphingomonas sensu stricto and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. *International Journal of Systematic and Evolutionary Microbiology*, *51*(4), 1405–1417. https://doi.org/10.1099/00207713-51-4-1405

Tindall BJ. The names *Hungateiclostridium*Zhang *et al.* 2018, *Hungateiclostridium thermocellum* (Viljoen *et al.* 1926) Zhang *et al.*2018, *Hungateiclostridium cellulolyticum*(Patel *et al.* 1980) Zhang *et al.* 2018, *Hungateiclostridium aldrichii* (Yang *et al.* 1990) Zhang *et al.* 2018, *Hungateiclostridium alkalicellulosi* (Zhilina *et al.* 2006) Zhang *et al.*2018, *Hungateiclostridium clariflavum*(Shiratori *et al.* 2009) Zhang *et al.* 2018, *Hungateiclostridium straminisolvens* (Kato *et al.* 2004) Zhang *et al.* 2018 and *Hungateiclostridium saccincola* (Koeck *et al.*2016) Zhang *et al.* 2018 contravene Rule 51b of the International Code of Nomenclature of Prokaryotes and require replacement names in the genus *Acetivibrio* Patel *et al.* 1980. *Int J Syst Evol Microbiol*. 2019;69(12):3927-3932. doi:10.1099/ijsem.0.003685

Trachsel, J., Humphrey, S., & Allen, H. K. (2018). Butyricicoccus porcorum sp. nov., a butyrate- producing bacterium from swine intestinal tract. *International Journal of Systematic and Evolutionary Microbiology*, *68*(5), 1737–1742. https://doi.org/10.1099/ijsem.0.002738

Vandamme PA, Peeters C, Cnockaert M, et al. Bordetella bronchialis sp. nov., Bordetella flabilis sp. nov. and Bordetella sputigena sp. nov., isolated from human respiratory specimens, and reclassification of Achromobacter sediminum Zhang et al. 2014 as Verticia sediminum gen. nov., comb. nov. *Int J Syst Evol Microbiol*. 2015;65(10):3674-3682. doi:10.1099/ijsem.0.000473

Würdemann D, Tindall BJ, Pukall R, et al. Gordonibacter pamelaeae gen. nov., sp. nov., a new member of the Coriobacteriaceae isolated from a patient with Crohn's disease,

and reclassification of Eggerthella hongkongensis Lau et al. 2006 as Paraeggerthella hongkongensis gen. nov., comb. nov. *Int J Syst Evol Microbiol*. 2009;59(Pt 6):1405-1415. doi:10.1099/ijs.0.005900-0

Yutin, N., & Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology*, *15*(10), 2631–2641. https://doi.org/10.1111/1462-2920.12173

Zheng J, Wittouck S, Salvetti E, et al. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus*Beijerinck 1901, and union of *Lactobacillaceae*and *Leuconostocaceae*. *Int J Syst Evol Microbiol*. 2020;70(4):2782-2858. doi:10.1099/ijsem.0.004107

CHAPTER IV

Gut microbiota of the critically endangered Saiga antelope across two wild populations in a year without mass mortality

Authors

Eveliina Hanski[1], Munib Khanyari[2,3], Jingdi Li[1], Kieran A Bates[1], Steffen Zuther[4,5], Martin C J Maiden[1], Richard Kock[6], Sarah C L Knowles[1]

Title

Gut microbiota of the critically endangered Saiga antelope across two wild populations in a year without mass mortality

Abstract

The Saiga are migratory antelopes inhabiting the grasslands of Eurasia. Over the last century, Saiga have been pushed to the brink of extinction by mass mortality events and intense poaching. Yet, despite the high profile of the Saiga as an animal of conservation concern, little is known of its biology. The gut microbiota of Saiga has not been studied, despite its potential importance in health. Here, we characterise the gut microbiota of Saiga from two geographically distinct populations in Kazakhstan and compare it with that of other antelope species. We identified a consistent gut microbial diversity and composition among individuals and across two populations during a year without die-offs, with over 85% of bacterial genera being common to both populations. We further show that the Saiga gut microbiota resembled that of five other antelopes. The putative causative agent of mass die-offs, *Pasteurella multocida*, was not detected in the Saiga microbiota. Our findings provide the first description of the Saiga gut microbiota, generating a baseline for future work investigating the microbiota's role in health and mass die-offs, and supporting the conservation of this critically endangered species.

Introduction

The Saiga antelope (*Saiga tatarica* ssp.) is a long-distance migratory ungulate, famous for its distinctive pendulant nose (Figure 1A). Once ranging across nearly the whole of Eurasia, the majority of Saiga now inhabit Kazakhstan, the largest landlocked country in the world. The Kazakh Saiga (*Saiga tatarica tatarica*) is of particular interest due to its vulnerability to mass mortality events, seemingly caused by *Pasteurella multocida* (Kock et al., 2018). These mass mortality events, together with anthropogenic impacts such as poaching, have reduced the population of Kazakh Saiga by 95% in recent decades, from an estimated 1,200,000 to circa 50,000 between mid-1970s and 2003 (Bekenov et al., 1998; IUCN, 2018). Since 2002, the Saiga has been listed as 'critically endangered' by the International Union for Conservation of Nature (IUCN, 2018); however, much about the biology of the Saiga remains unknown. They live out of sight in the remote grasslands and deserts of Kazakhstan and have proved challenging to keep in captivity (Rduch et al., 2016), making this species particularly difficult to study. Since the early 21st century, the Kazakh Saiga have been living as two geographically distinct populations in the Betpak-Dala and Ural regions, both of which have suffered from mass die-offs.

A better understanding of the biology of the Saiga antelope could provide insights into their susceptibility to mass mortality events with potential conservation application. One key factor that can have important effects on mammal biology is the gut microbiota, the diverse community of microorganisms residing in the intestinal tract. While this community contains microorganisms that range from beneficial to pathogenic, as a whole it provides important functions for the host by regulating key physiological processes such as immune maturation and nutrient extraction (Semova et al., 2012; Chung et al., 2012). Correspondingly, disruption

of the gut microbiota can have adverse effects on the host and increase susceptibility to both infectious and non-infectious diseases (Clemente et al., 2012).

Many natural factors can influence gut microbiota composition in mammals, including diet, social interactions, infection, and aging (David et al., 2014; Raulo et al., 2021; Knutie, 2018; Langille et al., 2014). A growing literature indicates how the microbiota of wild animals can be altered by changes to their environment such as exposure to chemicals, habitat destruction, infectious disease, urbanization, and housing in captivity (Kohl et al., 2015; Kakumanu et al., 2016; Amato et al., 2013; Knutie et al., 2018; Dillard et al., 2022; Teyssier et al., 2018; Gibson et al., 2019; McKenzie et al., 2017; Alberdi et al., 2021; Malukiewicz et al., 2022), yet knowledge about the implications for animal health remains limited. Given the lability and health impacts of the gut microbiota, an improved understanding of host-microbiota interactions also holds conservation potential (Trevelline et al., 2019; Diaz & Reese, 2021). For example, characterisation of the Saiga gut microbiota during a year without mass mortality will allow future investigation potential role of the microbiota in disease, and might inform the design of microbiota manipulation in captivity, e.g., through diet or faecal microbiota transplants, to facilitate captive breeding programmes for species recovery (Trevelline et al., 2019).

Here, to our knowledge, we provide the first characterisation of the Saiga gut microbiota during a year without die-offs, explore the extent of microbiota variation across two geographically distinct populations, and examine how the Saiga microbiota compares with that of other antelope species. We hypothesised that if gut microbiota is involved in making the Saiga susceptible to mass mortalities caused by a pathobiont, this may manifest in an ‘altered’ gut microbiota, such as one that has a lower diversity in comparison to other related antelopes.

These data provide a baseline understanding of the gut microbiota in this critically endangered species, which future work examining the potential significance of the microbiota for mass mortality events can build upon.

Methods

Sample collection

Faecal samples were used to provide a non-invasive characterisation of the gut microbiota. Faecal samples of Saiga antelope were collected opportunistically from two populations in north-west (within 25 kilometres from 49°59′ N, 47°40′ E; 'Ural population') and central (within 25 kilometres from 49°30′ N, 61°51′ E; 'Betpak-Dala population') Kazakhstan in May 2019, which is during and immediately after the peak calving period.

Faecal samples were collected in a non-intrusive manner. Saiga individuals were located and observed (either directly or through binoculars) until defecation and once they moved on from that location, a fresh faecal sample was collected. The Saiga is the only wild ungulate in the region, with distinct pellet-shaped faeces, and tends to avoid livestock (Singh & Milner-Gulland, 2011), minimising the risk of faeces mis-identification. During sampling, faecal matter was picked up using clean disposable gloves and dissected with sterile forceps in order to get two separate 100 mg aliquots from inside faecal pellets, to avoid environmental contamination. The time between defecation and collection of samples was within an hour in most cases. While it was not possible to ensure that every faecal sample was from a different individual as Saiga were not marked, our collection methodology made it unlikely to sample the same individual twice: faecal sampling was conducted at various locations over several days within the two sites (Ural, Betpak-Dala) where tens of thousands of Saigas had gathered to calve.

Faecal samples were immediately preserved in DNA/RNA Shield, a preservation solution that protects DNA against degradation and allows sample storage at ambient temperature (Zymo Research, Irvine, California, USA). After a maximum of 3 weeks at ambient temperature in the field, samples were kept at -20°C in Kazakhstan until shipping at ambient temperature to the UK in March 2020 (under an import permit from the UK Plant and Animal Health Agency). Upon receipt samples were kept at -80°C until DNA extraction.

DNA extraction, PCR amplification, and amplicon sequencing

Genomic DNA was extracted using the ZymoBIOMICS DNA Miniprep Kit, following the manufacturer's protocol (Zymo Research, Irvine, California, USA). A total of 80 samples (32 samples from the Ural population and 48 samples from the Betpak-Dala population) were randomised into four extraction batches. The 80 samples included 10 duplicate aliquots which were used to check the robustness of our sample and data processing pipeline. DNA was eluted in 50 μL DNAse-free H_2O, and one negative extraction control (DNAse-free H_2O) was included in each extraction batch. Library preparation and sequencing was conducted at the Integrated Microbiome Resource, Dalhousie University, as described in Comeau et al. (2017). The V4–V5 region of the 16S rRNA gene was amplified by PCR using the 515F–926R primers (Parada et al., 2016; Walters et al., 2015). All samples were amplified and sequenced in one batch using the Illumina MiSeq platform (Reagent kit v3, 2x300 bp chemistry). The sequencing run included a negative control for the PCR reaction and a negative control for the sequencing. All four extraction controls were sequenced either on the sequencing run in question or on subsequent sequencing runs.

Online data acquisition

A Web of Knowledge search was conducted in November 2021 to identify datasets which would allow comparison of the Saiga gut microbiota to that of other wild antelope species. Search keywords included *gut microbiome, gut microbiota, antelope*, and *ungulate.* Studies of captive animals were excluded. Only publicly available datasets for which V3–V4, V4, or V4–V5 16S rRNA primers and the Illumina sequencing platform had been used were considered. The retrieved and included datasets include the following five antelope species: Tibetan antelope (*Patholops hodgsonii*), Przewalski's gazelle (*Procapra przewalskii*), Impala (*Aepyceros melampus*), Springbok (*Antidorcas masupialis*), and Sable antelope (*Hippotragus niger*) (Supplementary Table 1).

16S data processing

To make all datasets fully comparable, publicly available antelope datasets were downloaded and the raw sequencing reads were processed together with the Saiga sequencing reads using a standardized pipeline, as follows. Downloaded files were converted to match a format that was compatible with Quantitative Insights into Microbial Ecology (QIIME2, 2020.11 distribution) (zipped fastq files with a unique name). First, FastQC v0.11.9 (Andrews, 2010) and MultiQC v1.12 (Philip et al., 2016) were used to visualize read quality, before cutadapt v3.4 (Martin, 2011) was used to remove adapters and/or primers where still present. Due to differences in targeted 16S rRNA region(s) (Suppl. Table 1), for comparative analyses across antelope species, sequencing reads were trimmed to include the V4 region (515F-806R) only to make amplicon sequence variants comparable using cutadapt (Martin, 2011). In analyses of Saiga microbiota samples alone, such trimming was not performed, in order to provide as much resolution as possible.

Trimmed or original reads containing V4 region were then processed as follows. Low-quality reads were filtered using qiime2 quality-filter q-score (default settings, QIIME2 2020.11) before using the Deblur workflow (Amir et al., 2017) to denoise sequences into ASVs, as suggested by QIIME2 (Bolyen et al., 2019). Within Deblur, trimming length was determined by manually viewing the quality plot for each study. ASV taxonomy was assigned using a classifier trained on the full-length 16S rRNA gene SILVA v138.1 database (Quast et al., 2013), and ASVs taxonomically assigned as 'mitochondria' or 'chloroplast' as well as those not identifiable at the kingdom level were removed. Phylogenetic trees of the remaining ASVs were built using the SEPP qiime2 plugin (https://github.com/bioshared/q2-fragment-insertion) with a reference phylogeny (sepp-refs-gg-13-8).

Negative controls for DNA extraction (n=4) and library preparation (n=1) of Saiga antelope samples collectively contained 13 ASVs with a maximum read count for any given ASV per control of 9. The R package decontam was used to test for potential contaminants in the Saiga dataset, for which negative controls were available. The decontam test was conducted using the 'prevalence' (presence/absence) method, which compares each sequence in biological samples to the prevalence in negative extraction and PCR controls. A sequence was considered a contaminant if it reached a probability of 0.1 in the Fisher's exact test used in decontam. No potential contaminants were identified in the Saiga dataset with this method.

Sample completeness and rarefaction curves were generated with R package iNEXT (Hsieh et al., 2022; Chao et al., 2014) for all included datasets combined and the read depth threshold (below which samples were excluded from further analysis) was set at 4000, based on where these curves plateaued. Data were not rarefied. Singletons and doubletons (ASVs with a total of either one or two sequences across the dataset) were removed prior to beta (not alpha)

diversity analyses to guard against the possible influence of remaining contaminants and sequencing errors. Microbiome profiles from duplicate Saiga samples were inspected using a principal coordinate analysis (PCoA) based on Aitchison distance in package vegan, and the effect of sample ID for microbiota composition was tested using permutational multivariate analysis of variance (PERMANOVA) on Aitchison distance. Prior to ordination on Aitchison distance, a centered log-ratio (clr) transformation was performed using the package microbiome (Lahti & Shetty, 2017), with relative abundance values of zero replaced with a pseudocount (min(relative abundance/2)). Duplicates were removed from the dataset before further analysis.

Analyses

Data was analysed and visualised in R (v4.1.2) using packages phyloseq (McMurdie & Holmes, 2013), vegan (Oksanen et al., 2022), microbiome (Lahti & Shetty, 2017), ALDEx2 (Fernandes et al., 2013), UpSetR (Conway et al., 2017), and ggplot2 (Wickham, 2016). For the inspection and comparison of taxonomic compositions, ASV counts were transformed into relative abundance per sample. We searched for the presence of *Pasteurella multocida* in the Saiga gut microbiota both by investigating taxonomic assignments from the SILVA database as well as by using the NCBI Nucleotide Basic Local Alignment Search Tool (Nucleotide BLAST). For the latter, we BLAST searched all ASVs either assigned to Gammaproteobacteria (the bacterial class containing *P. multocida*) or not assigned a class by SILVA, against the *P. multocida* type sequence (taxid: 747).

Differential abundance testing of bacterial taxa across the two Saiga populations was conducted using package ALDEx2 (Fernandes et al., 2013). Monte-Carlo sampling (mc.samples=128) from a Dirichlet distribution was used to generate a distribution of clr transformed values for

all taxa. Welch's and on rank sum tests were then performed on the clr transformed values. Taxa with a Benjamini-Hochberg corrected p-value (q-value) of less than 0.05 and an effect size greater than 1 were considered to have a significantly different abundance between the two populations. Asymptotic alpha diversity (ASV richness and Shannon diversity) was estimated using package iNEXT (Hsieh et al., 2022; Chao et al., 2014). Differences in alpha diversity and Jaccard dissimilarity among samples from each population were tested using Wilcoxon rank sum tests with 1,000 permutations.

Beta diversity metrics were calculated with the package vegan, and PERMANOVA was performed using Aitchison and Jaccard distances. Ordination was conducted with the package vegan using PCoA. A clr transformation was conducted before ordination for Aitchison distance (zero relative abundances replaced with pseudocounts as described in *Data processing*). Ordination plots were produced using the phyloseq package (McMurdie & Holmes, 2013). A cladogram showing the phylogenetic relationships among the six antelope species was retrieved from the TimeTree database.

Results

Sequencing outcome of the Saiga antelope samples

A total of 80 faecal DNA samples from 70 Saiga antelope (Kazakhstan) (Fig. 1A–B) were sequenced using the Illumina MiSeq platform. Eight Saiga samples were lost during read depth filtering. The remaining 72 samples included eight duplicate samples (duplicate samples from a single faecal deposition, see Methods) that were sequenced to assess the robustness of our pipeline. Duplicates generally clustered together with their corresponding samples (Suppl. Fig. 1) and sample ID strongly predicted microbiota composition across these repeat samples, explaining 67.1% of all gut microbial variation (PERMANOVA based on Aitchison distance,

$p<0.001$), indicating the sampling method produced a repeatable representation of the faecal microbiota. These duplicates were removed from the dataset before further analysis. After filtering out singleton and doubleton amplicon sequence variants (ASVs) in the dataset, the two populations of Saiga collectively harboured a total of 4,036 unique ASVs, with a mean of 743 ASVs per sample (range 470 to 1,057). Using the SILVA database (v138.1), 92.9% ASVs could be assigned to family level, 74.7% to genus level, and 0.7% to species level. Due to the low level of assignment to species level, the highest taxonomic level we considered in analyses was genus.

Composition and diversity of the Saiga antelope gut microbiota

Of the 4,036 ASVs found in Saiga, 2,689 (66.6%) were detected in both populations, while 570 (14.1%) were unique to the Ural population (n=25) and 777 (19.3%) were unique to the Betpak-Dala population (n=39). Across both populations, 16 bacterial phyla, 24 classes, 47 orders, 82 families, and 163 genera were identified (Fig. 1C). The two populations shared 15 out of 16 phyla (93.8%) and 91.5–91.7% of classes and orders. When inspecting gut microbiota profiles at the family and genus level, 85.4% and 85.9% of taxa were shared between the two populations, respectively. The shared bacterial families and genera formed 93.4–96.1% and 80.3–84.4%, respectively, of the total relative abundance in the Betpak-Dala and Ural populations, such that most bacteria in the Saiga microbiota belong to genera that are common to both populations. Unique genera in each population were not of any particular lower resolution taxa (such as family or class); instead, the Betpak-Dala population had 8 unique genera from 5 different classes. Similarly, the Ural population had 15 unique genera from 8 different classes.

At the phylum level, the Saiga gut microbiota was heavily dominated by Firmicutes and Bacteroidota, which together comprised 94.9% of all reads on average per sample (Fig. 1D). The Ural population harboured one unique phylum, Campilobacterota, that was rare and formed just 0.00002% of total abundance in this population. The shared microbiota (comprised of ASVs found in both populations, irrespective of relative abundance and prevalence) was proportionately dominated by the bacterial families Oscillospiraceae (20.8% and 16.6% of total relative abundance of taxa in Betpak-Dala and Ural, respectively) and Rikenellaceae (15.6% and 9.9%%), followed by 68 additional shared families (Suppl. Fig. 2). The shared microbiota contained 140 genera, with the predominant genus being Oscillospiraceae UCG-005 (16.9% and 11.8%). A small number of individuals across the two populations had an increased relative abundance of Proteobacteria (Fig. 1D), which includes various common potentially pathogenic taxa, such as *Salmonella, Escherichia,* and *Helicobacter.* A genus-level inspection of the Saiga gut microbiota revealed *Escherichia*/*Shigella* to account for Proteobacteria overrepresentation in these samples (Suppl. Fig. 2).

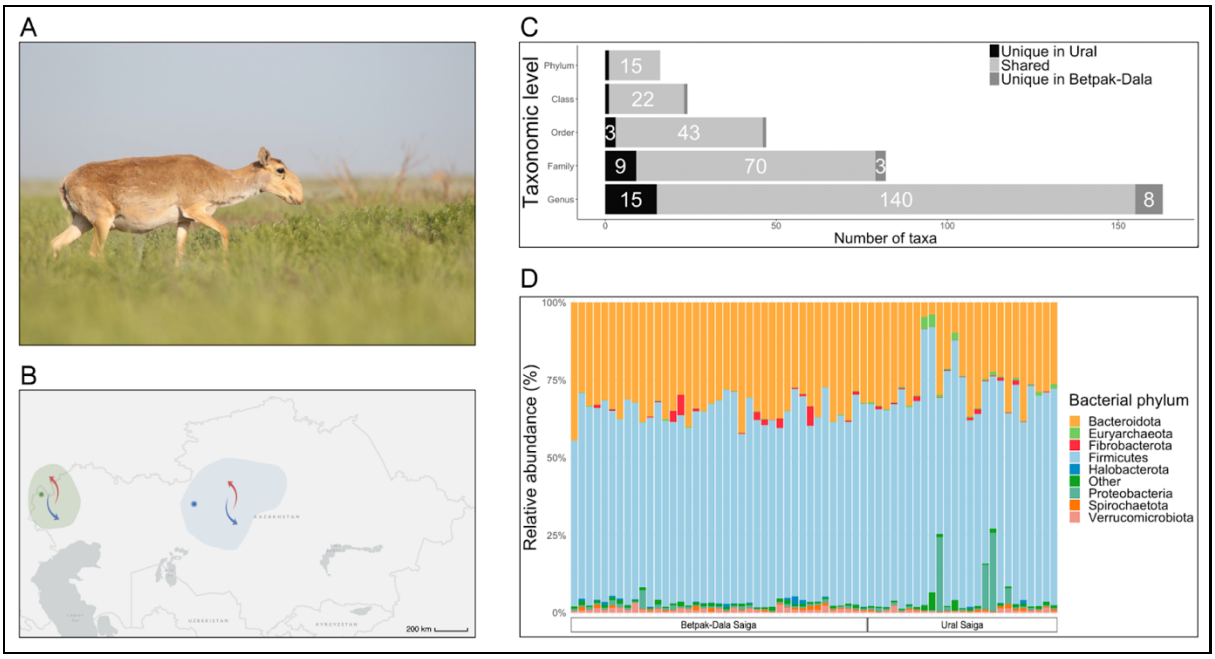


Figure 1. (**A**) A female Saiga antelope (*Saiga tatarica*). Photo credit: Albert Salemgareyev. (**B**) Geographic ranges of the two sampled Saiga antelope populations (*blue* = Betpak-Dala population, *green* = Ural population). Overall migration directions are indicated with arrows (*red* = migration direction in spring, *blue* = migration direction in autumn). Approximate centroids of faecal sample collection sites are indicated with dark dots within the habitats. (**C**) Number of shared and unique taxa between the two populations at five taxonomic levels. Numbers on bars indicate how many taxa were unique to Ural (black), Betpak-Dala population (dark grey), or that were shared between the two populations (light grey). Number of unique/shared taxa is written when ≥3. (**D**) Gut microbiota composition of the Saiga antelope at phylum level. Rare taxa (mean relative abundance <0.03% and prevalence <0.10% across samples) are under 'Other'. Stacked bars are individual samples. Horizontal bars indicate Saiga population.

Gammaproteobacteria, the class containing *P. multocida*, the bacterium implicated in Saiga mass mortalities, was detected in both populations and formed 0.3% (Betpak-Dala) and 2.9% (Ural) of total abundance at class level; however, no ASVs were assigned to the family Pasteurellaceae nor the genus *Pasteurella.* Furthermore, a Nucleotide BLAST search was conducted for ASVs for which SILVA had either assigned the class Gammaproteobacteria or had failed to assign a class. This search did not produce a ≥98% species identity match for *P. multocida*, further confirming this bacterium was not detected in this Saiga gut microbiota dataset.

Despite the high proportion of shared ASVs across the two populations (shared ASVs formed 77.6% and 82.5% of all ASVs detected in the Betpak-Dala and Ural populations, respectively), the gut microbiota compositions of the two populations were detectably different, with 10.4% of variation in Aitchison dissimilarity being explained by population identity (PERMANOVA, $p<0.0001$; population identity explained 8.6% of Jaccard dissimilarity, $p<0.001$). Samples clustered by population when ordinated with the exception of 3 samples, and these patterns were consistent across categorical (Aitchison, Jaccard) and phylogenetic distances (weighted and unweighted UniFrac) (Fig. 2A, Suppl. Fig. 3), indicating consistently that the two populations have distinct gut microbial communities. The three samples that did not cluster by populations were not the samples with increased relative abundance of Proteobacteria (Fig. 1D).

To investigate whether the relative abundance of any specific taxa significantly differed between the two populations, we performed a differential abundance analysis across taxonomic levels from phylum to genus. A total of 14 taxa including three orders, five families, and six genera significantly differed in relative abundance between populations (Fig. 2B). In addition to the shared taxa with significantly different abundances, the Ural population had 9 unique bacterial families including Campylobacteraceae, Moraxellaceae, and Acetobacteraceae. The Betpak-Dala population had 3 unique families; Methanosarcinaceae, Oxalobacteraceae, and Microbacteriaceae. At genus level, the Ural population had 3 unique taxa while the Betpak-Dala population had 9 unique taxa.

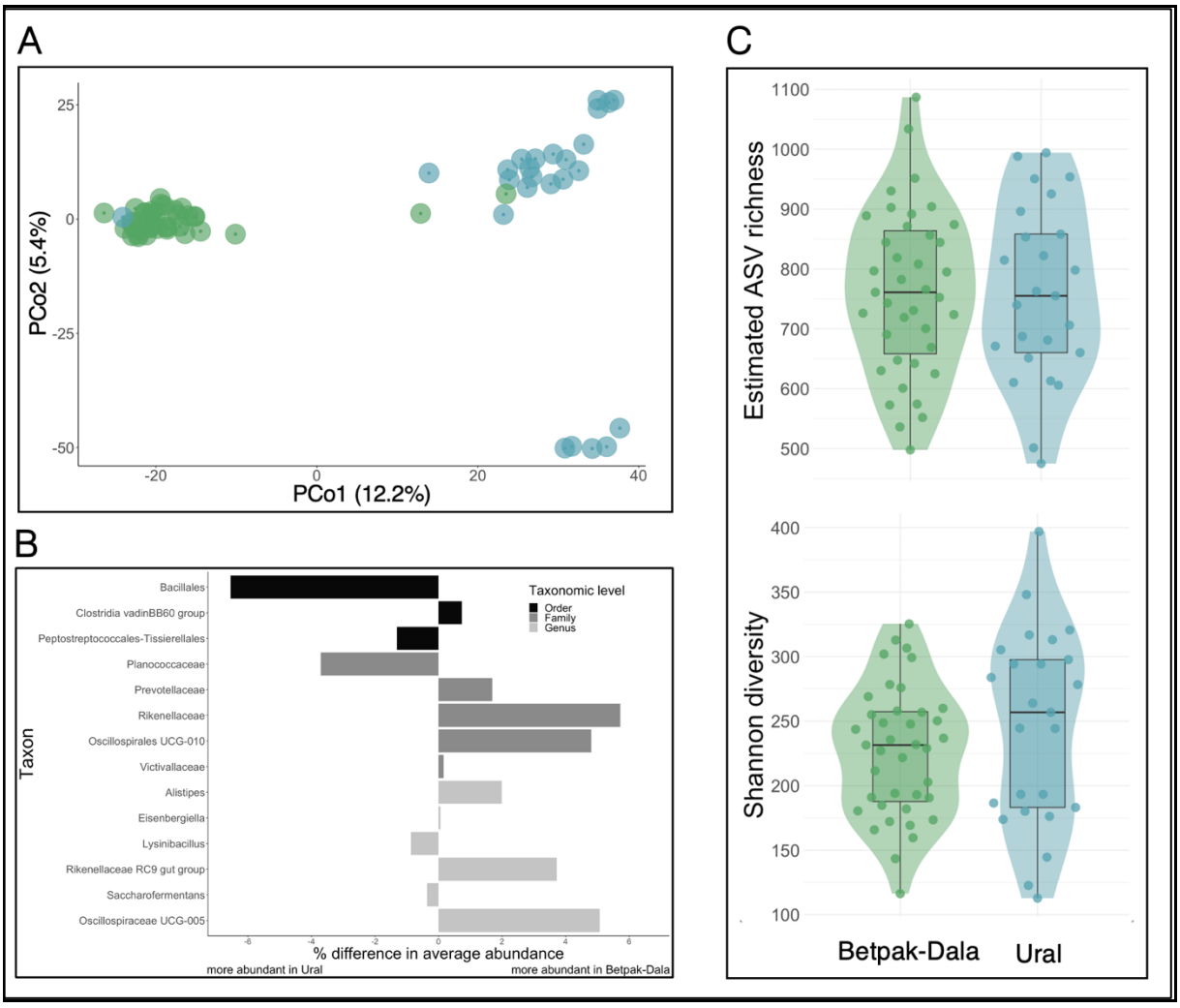


Figure 2. (**A**) Principal coordinate analysis of Betpak-Dala (green) and Ural (blue) Saiga gut microbiota (dis)similarity based on Aitchison distance. Circles are individual samples. (**B**) Bacterial taxa with significantly differing relative abundance between the Betpak-Dala and the Ural Saiga populations. Bars indicate the mean relative abundance in the Betpak-Dala population minus the mean relative abundance in the Ural population. Only taxa with a Benjamini-Hochberg corrected p-value of less than 0.05 and an effect size (standardized mean difference) greater than 1 are shown. (**C**) Sample-level asymptotic ASV richness (top) and Shannon diversity (bottom) of Betpak-Dala (*green*) and Ural (*blue*) Saiga. Circles are individual samples. Horizontal bar indicates median alpha diversity. Statistical differences between Betpak-Dala and Ural Saiga were tested with permutational Wilcoxon rank sum tests (p>0.05 for both ASV richness and Shannon diversity).

The two Saiga populations had very similar ASV richness (permutational Wilcoxon rank sum test, p=0.978; estimated ASV richness range in Ural Saiga 475–994, mean 759, median 755; range in Betpak-Dala Saiga 498–1087, mean 763, median 761). Similarly, while the Ural Saiga had a slightly higher Shannon diversity index, this difference was not statistically significant

(permutational Wilcoxon rank sum test, p=0.262; range in Ural Saiga 113–397, mean 245, median 257; range in Betpak-Dala Saiga 116–325, mean 227, median 232; Fig. 2C).

Comparison between Saiga antelope and five other antelope species

The Saiga gut microbiota was compared with that of five other antelope species for which publicly available 16S rRNA microbiota data could be retrieved: the Tibetan antelope (*Patholops hodgsonii*), Przewalski's gazelle (*Procapra przewalskii*), Impala (*Aepyceros melampus*), Springbok (*Antidorcas masupialis*), and Sable antelope (*Hippotragus niger*). The samples originated from China, South Africa, and Namibia (Suppl. Table 1). The investigated antelope species typically live in similar habitats namely deserts, grasslands, and shrublands (IUCN, 2018). The target region of 16S rRNA gene varied across the datasets, thus all 16S amplicon sequences were trimmed to the V4 region for this comparative analysis to make them comparable (see Methods).

The gut microbiota composition of the Saiga was, at a high taxonomic level, broadly similar to that of the five other antelope species. The predominant phyla in all species were Firmicutes (54.1%–70.9% of total abundance) and Bacteroidota (23.8%–39.9%). At family level, Oscillospiraceae and Rikenellaceae formed the two predominant bacterial families (25.6%–34.2%) in all antelope species except the Tibetan antelope, which had Oscillospiraceae and Lachnospiraceae as the most abundant families, followed by Rikenellaceae (Fig. 3A). At the ASV level, Saiga shared 13.6% and 14.2% of their ASVs with Tibetan antelope and Przewalski's gazelle, respectively (Suppl. Fig. 4A), whereas no common ASVs were detected between Saiga (or the other two Asian antelopes, Przewalski's gazelle and Tibetan antelope) and the African antelopes (Impala, Springbok, and Sable antelope).

Host species identity, which could not be distinguished from dataset identity, had a significant effect on gut microbiota composition explaining 24.1% of variation in bacterial community structure across the dataset (PERMANOVA based on Aitchison distance, $p<0.001$; host species identity explained 19.5% of Jaccard dissimilarity, $p<0.001$), although beta dispersion varied significantly among the datasets and could have affected these results ($F=458.74$, $p=0.001$). In ordination analyses, Saiga samples clustered separately from samples from all other antelopes on Jaccard, unweighted UniFrac and Aitchison distances but together with other samples on weighted UniFrac distance (Fig. 3B, Suppl. Fig. 4B–D).

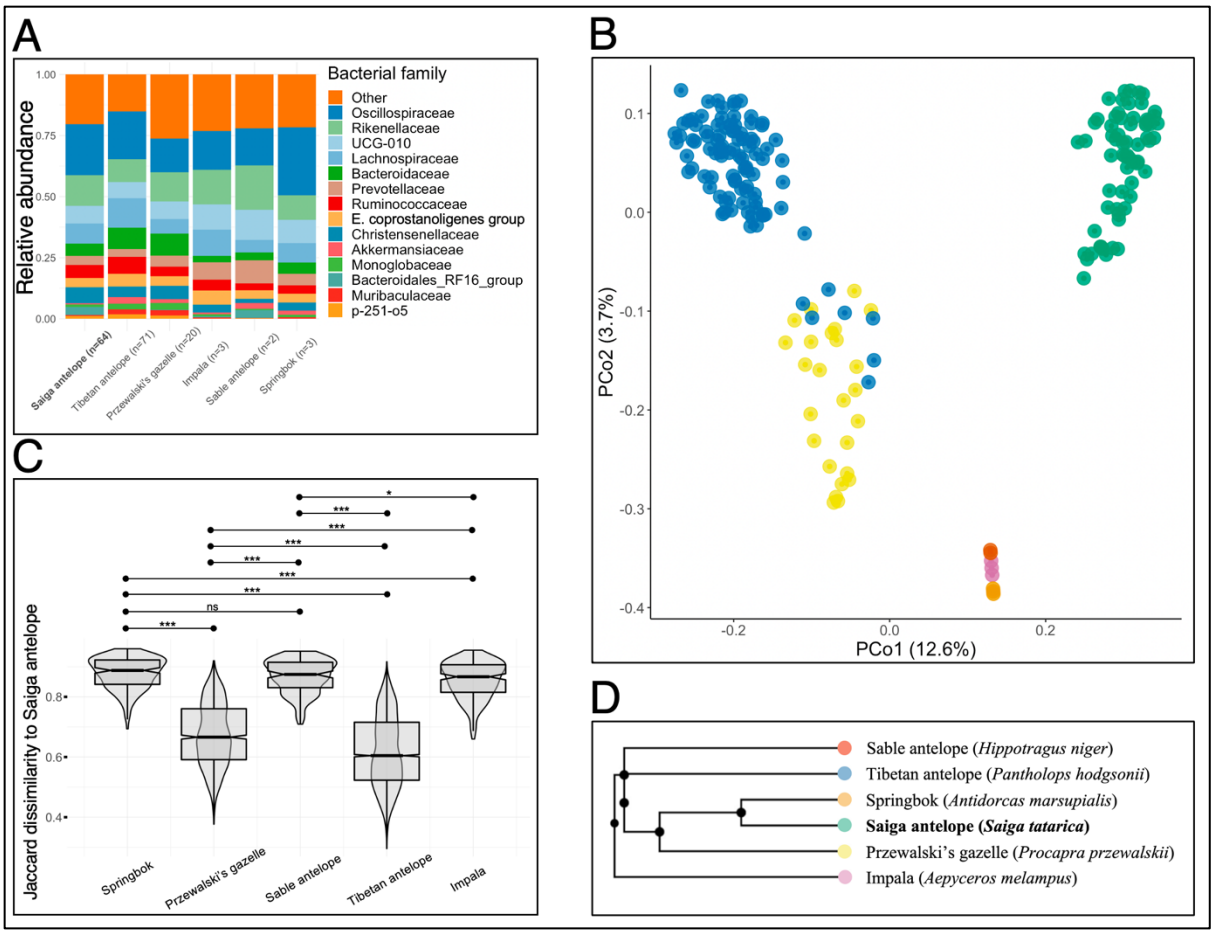


Figure 3. (**A**) The gut microbiota composition of six antelope species at family level. Rare taxa (mean relative abundance <0.03% and prevalence <0.10% across samples) and taxa for which bacterial family could not be assigned are under 'Other'. (**B**) Principal coordinate analysis of gut microbiota (dis)similarity of six antelope species based on Jaccard distance. Taxa have been agglomerated to the family level before ordination. Circles are individual samples. Colour indicates host species (*green* = Saiga antelope, *yellow* = Przewalski's gazelle, *blue* = Tibetan antelope, *orange* = Springbok, *pink* = Impala, *red* = Sable antelope). (**C**) Pairwise dissimilarity of bacterial genera (Jaccard distance) between Saiga antelope and (left to right) Springbok, Przewalski's gazelle, Sable antelope, Tibetan antelope, or Impala (antelopes ordered by phylogenetic relatedness to Saiga antelope starting with closest relative). Differences were tested with permutational Wilcoxon rank sum tests (***, $p<0.001$; ns, $p>0.05$). (**D**) A cladogram of six antelopes. The nodes indicate a common ancestor and the lines are relative to evolutionary timescale. The cladogram of the six antelopes was retrieved from the TimeTree database.

The Saiga shared the majority of its gut microbial taxa with at least one antelope at all investigated taxonomic levels (phylum → family). It shared all of its 14 phyla with at least one other antelope species and 12 phyla (85.7%) with at least three other antelope species (Suppl. Fig. 5A). The other antelopes collectively harboured an additional eight phyla not found in the

Saiga (Patescibacteria, Campilobacterota, Myxococcota, Nitrospirota, Acidobacteriota, Chloroflexi, Fusobacteriota, and Synergistota). At class and order levels, the Saiga shared 95.5% and 81.8% of its taxa, respectively, with at least three other antelope species. At family level, the Saiga shared 75.7% of taxa with at least three other antelopes (Suppl. Fig. 5B). These 56 bacterial families formed 94.2% of the total relative abundance of the Saiga gut microbiota. The Saiga harboured two unique bacterial families not detected in any of the other antelope species. These families detected only in Saiga were rare, representing less than 1% of total relative abundance in both Saiga populations.

In terms of shared genera (Jaccard dissimilarity measured at genus-level, due to lack of shared ASVs between Asian and African antelopes), the Saiga microbiota was more similar in composition to that of other Asian antelope species (Tibetan antelope, Przewalski's gazelle) than African species (Sable antelope, Springbok, Impala), despite the Saiga being phylogenetically more closely related to the African species (Fig. 3C–D). However, when considering the phylogenetic distance between genera (unweighted UniFrac distance on genera), the Saiga microbiota was more similar to that of African, rather than Asian, antelope species (Suppl. Fig. 6).

Compared to the other antelope species examined, Saiga had the highest alpha diversity (ASV richness and Shannon diversity; Suppl. Fig. 7). The difference was significant between Saiga and all other antelopes (permutational Wilcoxon rank sum tests, $p<0.001$ for all). However, the only significant difference in ASV richness was between Saiga and Tibetan antelope (permutational Wilcoxon rank sum test, $p<0.001$; $p>0.06$ for all other comparisons).

Discussion

The critically endangered Saiga antelope is a species of interest because of its vulnerability to mass mortality events. Together with anthropogenic factors, such as poaching and agricultural, road infrastructure expansion, mass mortality events in Saiga can lead to the extinction of regional populations with observed losses of 75% and 88% of the affected population in 1988 and 2015, respectively (Robinson et al., 2019). A better understanding of the baseline biology of the Saiga could provide insights into their susceptibility to mass mortality events and tools for more successful husbandry in captivity (Trevelline et al., 2019; Diaz & Reese, 2021). As a step towards this, we present a gut microbiota profile of the Saiga, the first to our knowledge, from samples taken in a year without mass mortality and compare it to the gut microbiota of five other antelope species.

Similar to other mammals (Rojas et al., 2021), the gut microbiota of the Saiga was dominated by Firmicutes and Bacteroidota, which together formed 95% of total relative abundance. Two subpopulations of the Kazakh Saiga were sampled in the Betpak-Dala and Ural regions. These sampling locations were over 1,000 km apart and the estimated geographic ranges of these migrating populations are at least 500 km apart (ACBK, 2011). Despite the geographic separation, the two populations shared approximately 85–94% of taxa at all inspected taxonomic levels from phylum to genus, and the shared genera formed over 80% of total relative abundance in both populations.

The Betpak-Dala and Ural Saiga populations displayed similar gut microbial alpha diversity, but differed in microbiota composition. Others investigating gut microbiota diversity across two wild populations of the Przewalski's gazelle, a close relative of the Saiga, did not detect significant population differences in either alpha or beta diversity (Liu et al., 2021). These

populations are, however, geographically closer to each other (sampling locations <200 km apart) than the Betpak-Dala and Ural Saiga populations (sampling locations >1,000 km apart), which could contribute to the greater cross-population similarity in the gut microbiota of Przewalski's gazelle compared to Saiga. Geographical proximity could affect gut microbiota similarity through, for instance, similar vegetation type and thus diet as well as population mixing and gut microbe transmission (Knowles et al., 2019; Raulo et al., 2021).

Due to the opportunistic nature of sampling and the tendency of the Saiga to avoid people, we were unable to collect more information concerning the individual animals. Hence, it remains unknown how much of the variation within and across the populations was driven by factors such as age and sex. Similarly, while various steps were taken to minimise the potential effect of exposure to oxygen and environmental microbes (see Methods), it is possible that the microbial composition of faecal matter was altered by these factors. Considering the two populations inhabit different areas at least 500 km apart, it is likely some of the gut microbial variation was driven by differences in habitat and diet, as has been found in other wild mammalian species (Kartzinel et al., 2019). The faecal samples used in the study were collected during or shortly after the Saiga calving period in May. Hence, the gut microbiota profiles reflect those in the calving period. While this might vary from the Saiga gut microbiota outside calving period, for example due to differences in hormone levels (Mallott et al., 2020), this is a relevant time point for sampling the Saiga since the previous mass mortality events have occurred during calving (Kock et al., 2018), when the birthing females are likely to be stressed and potentially immunocompromised.

As a commensal, *Pasteurella multocida* (the putative causative agent of the Saiga mass mortalities) is most often found in the oral, nasopharyngeal, and respiratory microbiota (Wilson

& Ho, 2013) and previous work has demonstrated the presence of the bacterium in the respiratory tract of healthy Saiga (Sanchez-Monge, 2016). Analysis of tissue samples from the 2015 Saiga mass mortality event provided indication of a possible *P. multocida* invasion from the gut (the bacterium was detected in intestinal mucosa, alongside other body sites; Kock et al., 2018), thus we searched for the presence of this bacterium in our dataset. We did not detect *P. multocida* in the Saiga gut microbiota and the closest taxonomic rank was the class Gammaproteobacteria.

The absence of *P. multocida* could be an artefact of the method since we characterised the gut microbiota by targeting a region of the 16S rRNA gene, a commonly used method that provides limited fine-scale taxonomic resolution, particularly at species level (Johnson et al., 2019). The method should provide a representative profiling at lower levels, such as at class and family levels (Heidrich et al., 2022). Considering the class Gammaproteobacteria was the closest taxonomic rank detected to *P. multocida,* our 16S data does suggest the bacterium was not present in the Saiga gut microbiota. Further, the primers used to amplify the V4–V5 region of the 16S rRNA gene (515F–926R) in the Saiga gut microbiota have detected Pasteurellaceae, the family under which *P. multocida* is classified, in another gut microbiota dataset (originating from wild *Mus musculus* faecal samples; mean relative abundance of Pasteurellaceae 0.01%; Hanski, unpublished). Considering this, the presence of *P. multocida* should have been captured at family level at least, further suggesting *P. multocida* was not present in the Saiga gut microbiota at the time of sampling, or alternatively its abundance was below the detection threshold of the methods used.

To put the Saiga gut microbiota into a wider phylogenetic perspective, we compared it to that of five other antelope species (Tibetan antelope, Przewalski's antelope, Sable antelope,

Springbok, and Impala) for which 16S rRNA V4 gut microbiota data from wild individuals was publicly available. Saiga shared the majority of its taxa with other antelope species across taxonomic levels from phylum to genus. Still, microbiota compositions varied significantly between Saiga and the other antelopes. This could have been affected by environmental factors, such as geographical location and diet, as well as experimental factors, such as DNA extraction kit and sequencing batch, which varied between the datasets and could not be controlled for due to small sample size (Knowles et al., 2019; Zhu 2022; Shaffer et al., 2022). The finding that the hosts shared a high number of taxa at the various taxonomic levels despite these methodological limitations suggests the Saiga does not present an outlier within antelopes from the gut microbiota perspective.

In terms of shared taxa, the Saiga gut microbiota was more similar in composition to antelopes from the same rather than different continent when similarity was measured with non-phylogenetically informed Jaccard distance, despite Saiga being phylogenetically more closely related to the African rather than Asian antelopes. However, this conclusion depended on the distance metric considered; when a phylogenetic distance metric was used (unweighted UniFrac), the Saiga microbiota was most similar to that of an African antelope species and its closest relative (Springbok), rather than geographically more proximate Asian antelope species. These results suggest that phylosymbiosis, under which microbiota similarity is expected to correlate with phylogenetic relatedness and which is observed in several (Song et al., 2020; Kohl et al., 2018; Weinstein et al., 2021; Ochman et al., 2010) but perhaps not all animals (Moeller et al., 2017), may be detected in the studied antelope species only when phylogenetic distance of gut microbes is considered. When phylogenetic relatedness of microbes is not considered, geography rather than host phylogeny appears to have a stronger influence on Saiga microbiota in relation to other antelopes.

Overall, our results indicate the gut microbiota of two geographically disparate Saiga antelope populations is taxonomically rather consistent, but varies in relative abundance of bacterial taxa. We did not detect *Pasteurella multocida* – the bacterium thought to cause Saiga mass mortalities – in the Saiga gut microbiota during this year without die-offs. Finally, we showed that the Saiga gut microbiota resembles that of other antelopes. With this, we provide a baseline description of the gut microbiota in this critically endangered species, on which future work examining the potential role of the microbiota in mass mortality events can build.

Acknowledgements

EH was supported by a scholarship from the Osk. Huttunen Foundation (Finland). MK was supported by the National Geographic Society's grant (EC-54710C-19). The research was funded by grants from the Emil Aaltonen Foundation (Finland), and fieldwork was supported by the Altyn Dala Conservation Initiative. SCLK was supported by a NERC fellowship (NE.L011867/1) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 851550).

Author information

Affiliations

[1]Department of Biology, University of Oxford

[2]School of Biological Sciences, University of Bristol

[3]Nature Conservation Foundation, Mysore, India

[4]Association for the Conservation of Biodiversity of Kazakhstan, Nur-Sultan, Kazakhstan

[5]Frankfurt Zoological Society, Frankfurt, Germany

[6]Centre for Emerging, Endemic and Exotic Diseases, The Royal Veterinary College, University of London.

Contributions

EH designed the Saiga antelope sample collection protocol. RK and MK collected the Saiga antelope samples. SZ assisted with fieldwork logistics. EH conducted DNA extraction. JL and EH processed the raw data. EH analysed the data with input from KAB and SCLK. EH, RK, SCLK, and MCJM contributed to study conceptualisation. EH wrote the manuscript with contributions from all authors.

Corresponding author

Sarah Knowles (sarah.knowles@biology.ox.ac.uk)

References

Alberdi A, Martin Bideguren G, Aizpurua O. Diversity and compositional changes in the gut microbiota of wild and captive vertebrates: a meta-analysis. *Sci Rep*. 2021;11(1):22660. Published 2021 Nov 22. doi:10.1038/s41598-021-02015-6

Amato KR, Yeoman CJ, Kent A, et al. Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiotas. *ISME J*. 2013;7(7):1344-1353. doi:10.1038/ismej.2013.16

Amir A, McDonald D, Navas-Molina JA, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. 2017;2(2):e00191-16. Published 2017 Mar 7. doi:10.1128/mSystems.00191-16

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Association for the Conservation of Biodiversity of Kazakhstan (ACBK). Annual Report, 2011.

Bekenov AB, Grachev IA, Milner-Gulland EJ. The ecology and management of the Saiga antelope in Kazakhstan. *Mammal Rev.* 1998;28(1) doi:10.1046/j.1365-2907.1998.281024

Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiota data science using QIIME 2 [published correction appears in Nat Biotechnol. 2019 Sep;37(9):1091]. *Nat Biotechnol*. 2019;37(8):852-857. doi:10.1038/s41587-019-0209-9

Chao A, Gotelli NJ, Hsieh TC, Sande EL, Ma KH, Colwell RK, Ellison AM (2014). "Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies." *Ecological Monographs*, **84**, 45–67.

Chung H, Pamp SJ, Hill JA, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell*. 2012;149(7):1578-1593. doi:10.1016/j.cell.2012.04.037

Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012;148(6):1258-1270. doi:10.1016/j.cell.2012.01.035

Comeau AM, Douglas GM, Langille MG. Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *mSystems*. 2017;2(1):e00127-16. Published 2017 Jan 3. doi:10.1128/mSystems.00127-16

Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938-2940. doi:10.1093/bioinformatics/btx364

David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiota. *Nature*. 2014;505(7484):559-563. doi:10.1038/nature12820

Diaz J, Reese AT. Possibilities and limits for using the gut microbiota to improve captive animal health. *Anim Microbiome*. 2021;3(1):89. Published 2021 Dec 29. doi:10.1186/s42523-021-00155-8

Dillard BA, Chung AK, Gunderson AR, Campbell-Staton SC, Moeller AH. Humanization of wildlife gut microbiota in urban environments. *Elife*. 2022;11:e76381. Published 2022 May 31. doi:10.7554/eLife.76381

Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354

F. Sanchez-Monge, “Calving status and commensalism of Pasteurella multocida in the surviving Betpak-dala saiga population in May 2016, after a mass mortality event in May 2015,” thesis, Royal Veterinary College, University of London (2016).
Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One*. 2013;8(7):e67019. Published 2013 Jul 2. doi:10.1371/journal.pone.0067019
Gibson KM, Nguyen BN, Neumann LM, et al. Gut microbiota differences between wild and captive black rhinoceros - implications for rhino health. *Sci Rep*. 2019;9(1):7570. Published 2019 May 28. doi:10.1038/s41598-019-43875-3
Heidrich V, Inoue LT, Asprino PF, et al. Choice of 16S Ribosomal RNA Primers Impacts Male Urinary Microbiota Profiling. *Front Cell Infect Microbiol*. 2022;12:862338. Published 2022 Apr 21. doi:10.3389/fcimb.2022.862338
Hsieh TC, Ma KH, Chao A (2022). *iNEXT: Interpolation and Extrapolation for Species Diversity*. R package version 3.0.0, http://chao.stat.nthu.edu.tw/wordpress/software_download/.
IUCN. 2018. The IUCN Red List of Threatened Species. Version 2018-2. Available at: www.iucnredlist.org. (Accessed: 29 November 2022).
Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet, et al (2022). vegan: Community Ecology Package. R package version 2.6-2. https://CRAN.R-project.org/package=vegan
Johnson JS, Spakowicz DJ, Hong BY, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiota analysis. *Nat Commun*. 2019;10(1):5029. Published 2019 Nov 6. doi:10.1038/s41467-019-13036-1
Kakumanu ML, Reeves AM, Anderson TD, Rodrigues RR, Williams MA. Honey Bee Gut Microbiome Is Altered by In-Hive Pesticide Exposures. *Front Microbiol*. 2016;7:1255. Published 2016 Aug 16. doi:10.3389/fmicb.2016.01255
Kartzinel TR, Hsing JC, Musili PM, Brown BRP, Pringle RM. Covariation of diet and gut microbiota in African megafauna. *Proc Natl Acad Sci U S A*. 2019;116(47):23588-23593. doi:10.1073/pnas.1905666116
Knowles SCL, Eccles RM, Baltrūnaitė L. Species identity dominates over environment in shaping the microbiota of small mammals. *Ecol Lett*. 2019;22(5):826-837. doi:10.1111/ele.13240
Knutie SA. Relationships among introduced parasites, host defenses, and gut microbiota of Galapagos birds. *Ecosphere.* 2018;9(5)e02286.
Kock RA, Orynbayev M, Robinson S, et al. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv*. 2018;4(1):eaao2314. Published 2018 Jan 17. doi:10.1126/sciadv.aao2314.
Kohl KD, Cary TL, Karasov WH, Dearing MD. Larval exposure to polychlorinated biphenyl 126 (PCB-126) causes persistent alteration of the amphibian gut microbiota. *Environ Toxicol Chem*. 2015;34(5):1113-1118. doi:10.1002/etc.2905
Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species distinguishability and phylosymbiosis along the length of the gastrointestinal tract. *Mol Ecol*. 2018;27(8):1874-1883. doi:10.1111/mec.14460
Langille MG, Meehan CJ, Koenig JE, et al. Microbial shifts in the aging mouse gut. *Microbiome*. 2014;2(1):50. Published 2014 Dec 5. doi:10.1186/s40168-014-0050-9
Leo Lahti, Sudarshan Shetty et al. (2017). Tools for microbiota analysis in R. Version . URL: http://microbiota.github.com/microbiota.
Liu R, Shi J, Shultz S, Guo D, Liu D. Fecal Bacterial Community of Allopatric Przewalski's Gazelles and Their Sympatric Relatives. *Front Microbiol*. 2021;12:737042. Published 2021 Sep 24. doi:10.3389/fmicb.2021.737042

Mallott EK, Borries C, Koenig A, Amato KR, Lu A. Reproductive hormones mediate changes in the gut microbiota during pregnancy and lactation in Phayre's leaf monkeys. *Sci Rep*. 2020;10(1):9961. Published 2020 Jun 19. doi:10.1038/s41598-020-66865-2

Malukiewicz J, Cartwright RA, Dergam JA, et al. The gut microbiota of exudivorous marmosets in the wild and captivity. *Sci Rep*. 2022;12(1):5049. Published 2022 Mar 23. doi:10.1038/s41598-022-08797-7

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

McKenzie VJ, Song SJ, Delsuc F, et al. The Effects of Captivity on the Mammalian Gut Microbiome. *Integr Comp Biol*. 2017;57(4):690-704. doi:10.1093/icb/icx090

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiota census data. *PLoS One*. 2013;8(4):e61217. Published 2013 Apr 22. doi:10.1371/journal.pone.0061217

Moeller AH, Suzuki TA, Lin D, Lacey EA, Wasser SK, Nachman MW. Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proc Natl Acad Sci U S A*. 2017;114(52):13768-13773. doi:10.1073/pnas.1700122114

Ochman H, Worobey M, Kuo CH, et al. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol*. 2010;8(11):e1000546. Published 2010 Nov 16. doi:10.1371/journal.pbio.1000546

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiotas with mock communities, time series and global field samples. *Environ Microbiol*. 2016;18(5):1403-1414. doi:10.1111/1462-2920.13023

Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-D596. doi:10.1093/nar/gks1219

Raulo A, Allen BE, Troitsky T, et al. Social networks strongly predict the gut microbiota of wild mice. *ISME J*. 2021;15(9):2601-2613. doi:10.1038/s41396-021-00949-3

Raulo A, Allen BE, Troitsky T, et al. Social networks strongly predict the gut microbiota of wild mice. *ISME J*. 2021;15(9):2601-2613. doi:10.1038/s41396-021-00949-3

Rduch V, Zimmermann W, Vogel K, et al. Saiga antelopes (*Saiga tatarica*) at Cologne Zoo: husbandry experiences and observations. *Zeitschrift des Kölner Zoos*. 2016;59(3)135-149.

Robinson S, Milner-Gulland EJ, Grachev Y, et al. Opportunistic bacteria and mass mortality in ungulates: lessons from an extreme event. *Ecosphere.* 2019;10(6):e02671. doi: 10.1002/esv2.2671.

Rojas CA, Ramírez-Barahona S, Holekamp KE, Theis KR. Host phylogeny and host ecology structure the mammalian gut microbiota at different taxonomic scales. *Anim Microbiome*. 2021;3(1):33. Published 2021 Apr 23. doi:10.1186/s42523-021-00094-4

Semova I, Carten JD, Stombaugh J, et al. Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish. *Cell Host Microbe*. 2012;12(3):277-288. doi:10.1016/j.chom.2012.08.003

Shaffer JP, Carpenter CS, Martino C, Salido RA, Minich JJ, Bryant M, Sanders K, Schwartz T, Humphrey G, Swafford AD, Knight R. A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. Biotechniques. 2022 Jun;73(1):34-46. doi: 10.2144/btn-2022-0032. Epub 2022 Jun 17. PMID: 35713407; PMCID: PMC9361692.

Singh NJ, Milner-Gulland EJ. Conserving a moving target: planning protection for a migratory species as its distribution changes. *Journal of Applied Ecology*. 2011(48)35–46.

Song SJ, Sanders JG, Delsuc F, et al. Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *mBio*. 2020;11(1):e02901-19. Published 2020 Jan 7. doi:10.1128/mBio.02901-19

Teyssier A, Rouffaer LO, Saleh Hudin N, et al. Inside the guts of the city: Urban-induced alterations of the gut microbiota in a wild passerine. *Sci Total Environ*. 2018;612:1276-1286. doi:10.1016/j.scitotenv.2017.09.035

Trevelline BK, Fontaine SS, Hartup BK, Kohl KD. Conservation biology needs a microbial renaissance: a call for the consideration of host-associated microbiota in wildlife management practices. *Proc Biol Sci*. 2019;286(1895):20182448. doi:10.1098/rspb.2018.2448

Walters W, Hyde ER, Berg-Lyons D, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2015;1(1):e00009-15. Published 2015 Dec 22. doi:10.1128/mSystems.00009-15

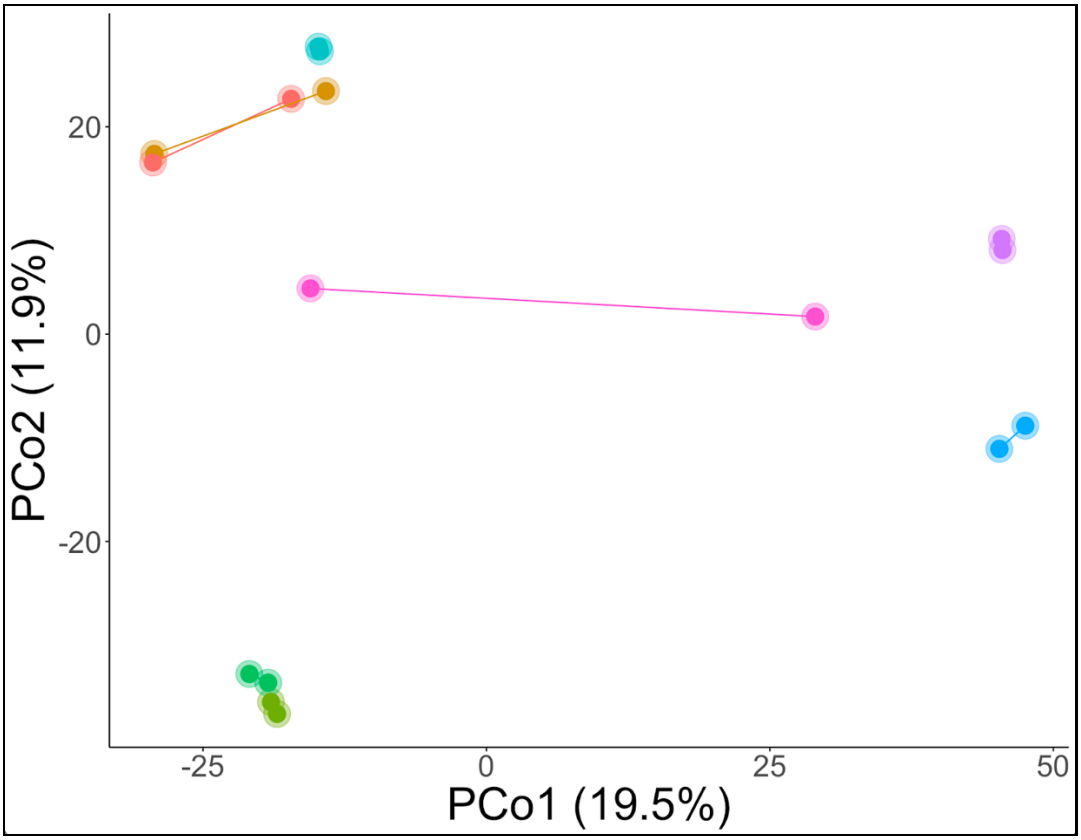
Weinstein SB, Martínez-Mota R, Stapleton TE, et al. Microbiome stability and structure is governed by host phylogeny over diet and geography in woodrats (*Neotoma* spp.). *Proc Natl Acad Sci U S A*. 2021;118(47):e2108787118. doi:10.1073/pnas.2108787118

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

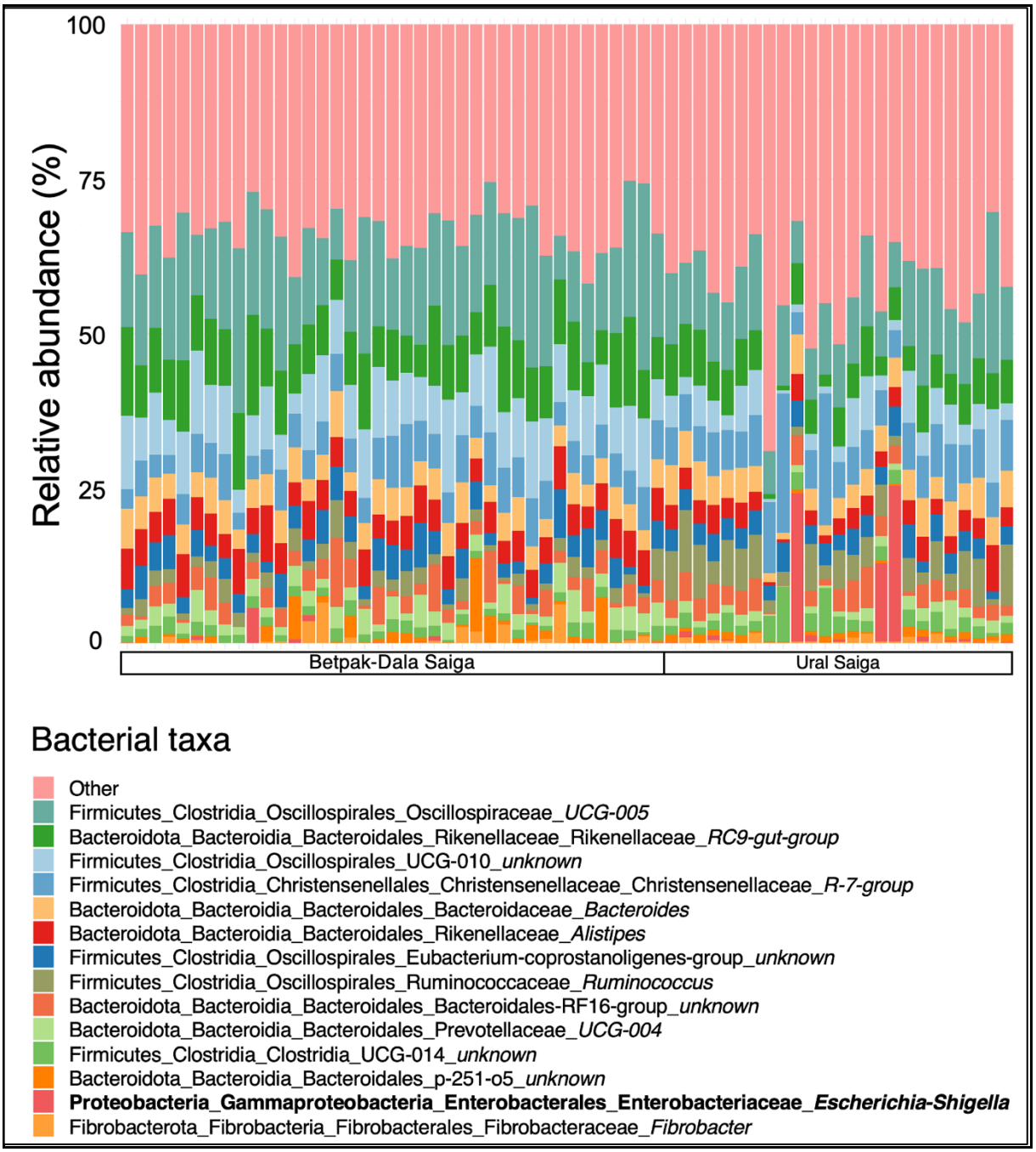
Wilson BA, Ho M. Pasteurella multocida: from zoonosis to cellular microbiology. *Clin Microbiol Rev*. 2013;26(3):631-655. doi:10.1128/CMR.00024-13

Zhu L. New challenges in the study of the evolution of wild animals and their gut microbiome. Ecol Evol. 2022 May 7;12(5):e8904. doi: 10.1002/ece3.8904. PMID: 35571763; PMCID: PMC9077628.

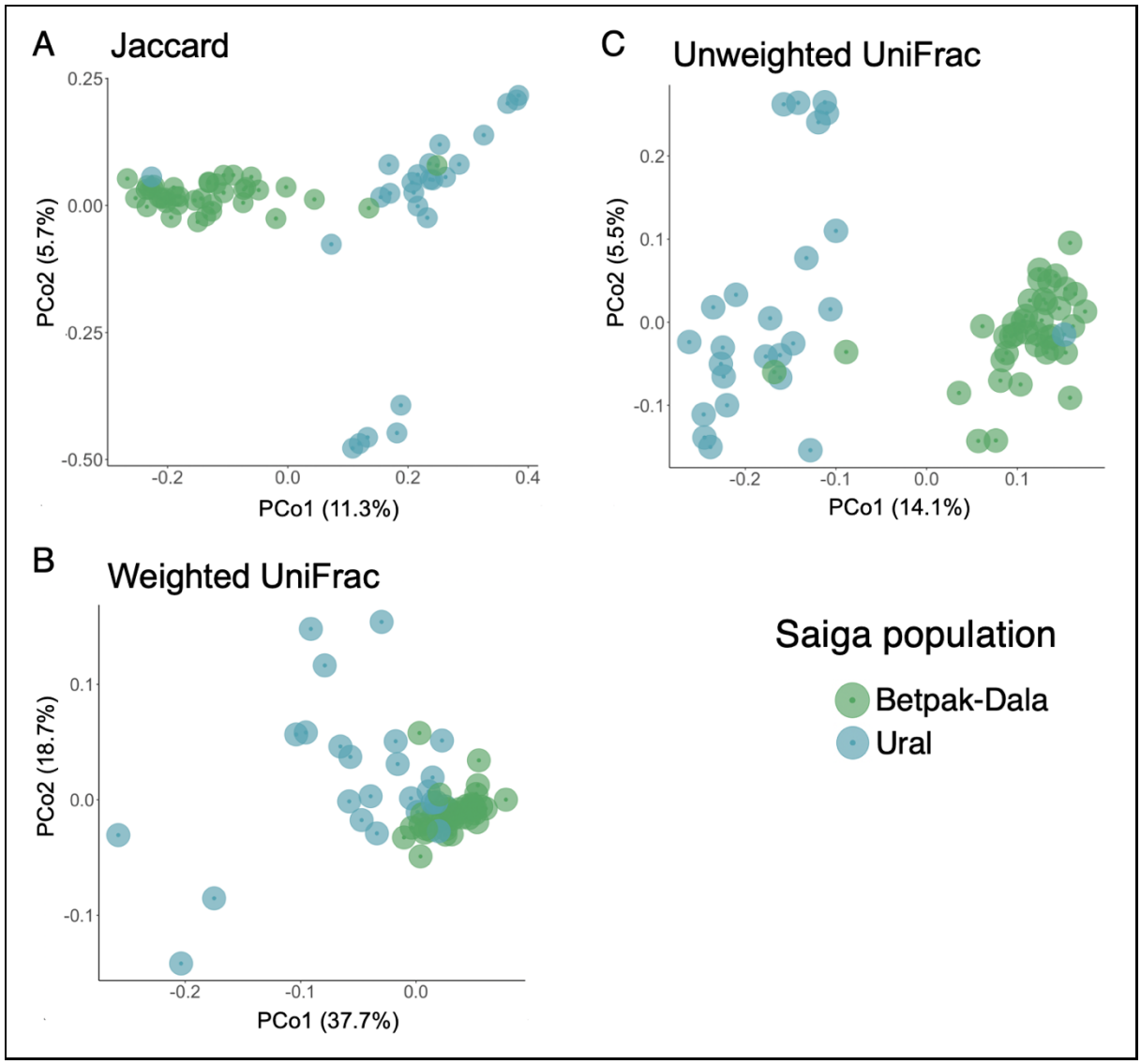
Supplementary material



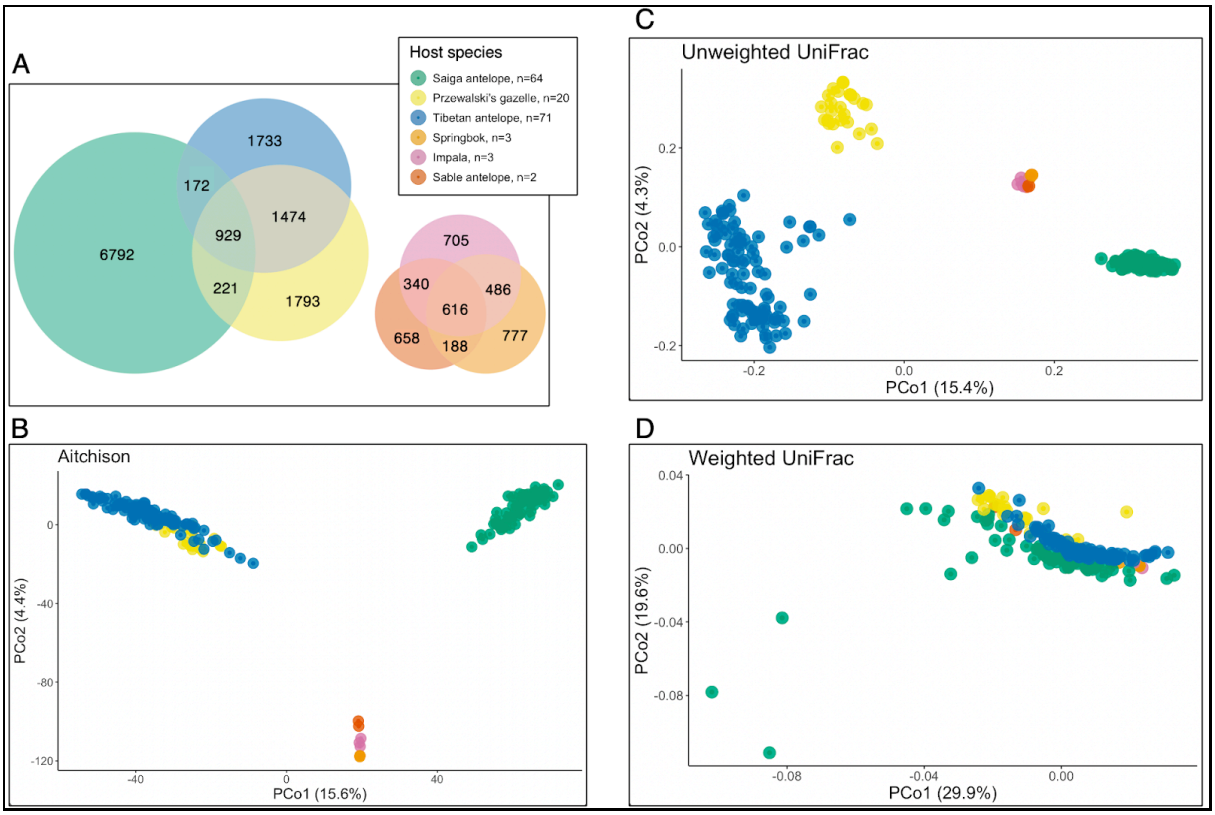
Supplementary Figure 1. Principal coordinate analysis (PCoA) on Aitchison distance for eight Saiga samples for which duplicates were included. Colour indicates sample ID (duplicates for each, connected with lines).



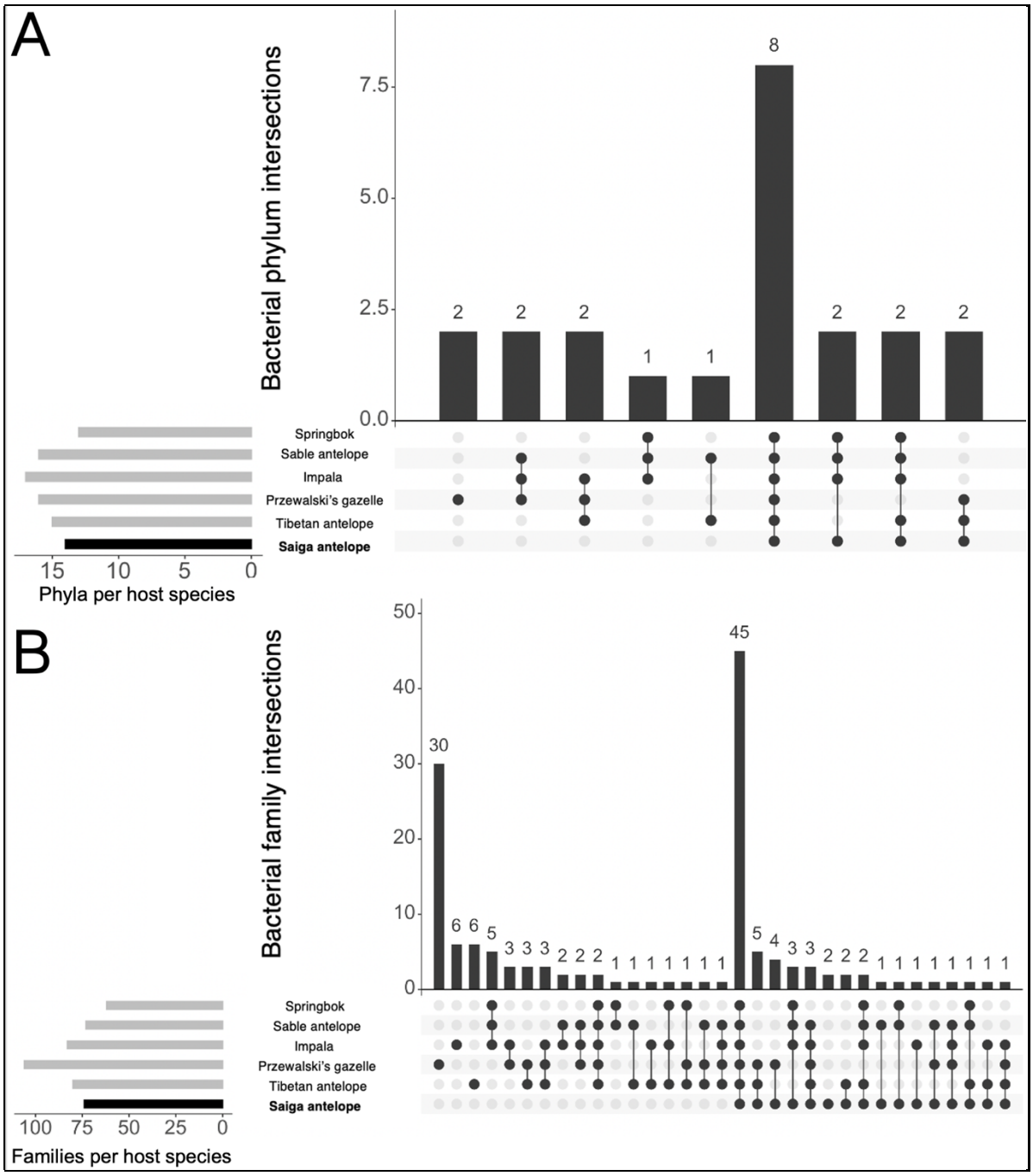
Supplementary Figure 2. Gut microbiota composition of the Saiga at family level. Stacked bars represent individual samples with horizontal bars indicating Saiga population. Rare taxa (mean relative abundance <0.03% and prevalence <0.10% across samples) and taxa for which bacterial family could not be assigned are under 'Other'.



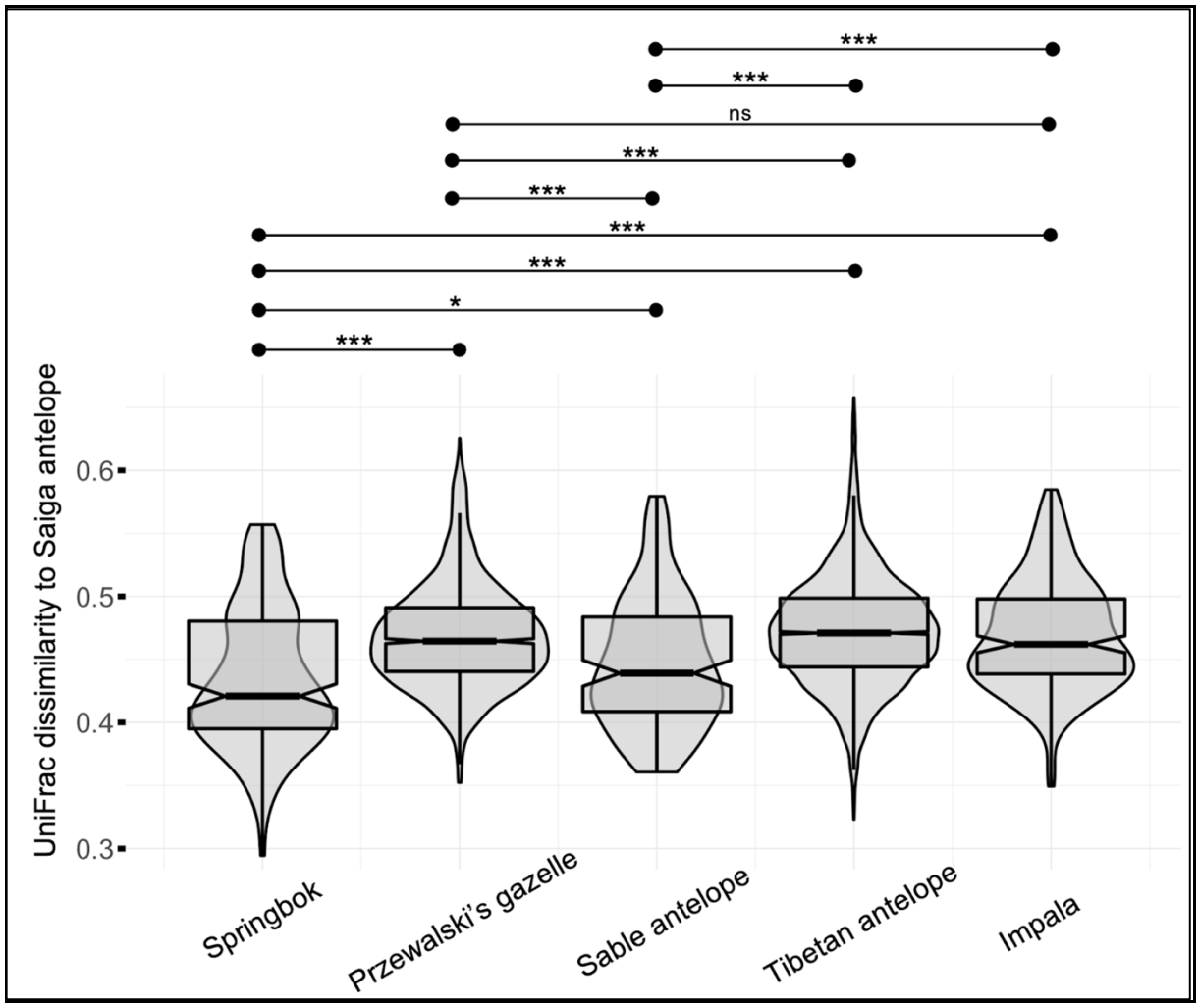
Supplementary Figure 3. Principal coordinate analysis of Betpak-Dala (*green*) and Ural (*blue*) Saiga gut microbiota on (**A**) Jaccard, (**B**) weighted UniFrac, and (**C**) unweighted UniFrac distances. Circles are individual samples.



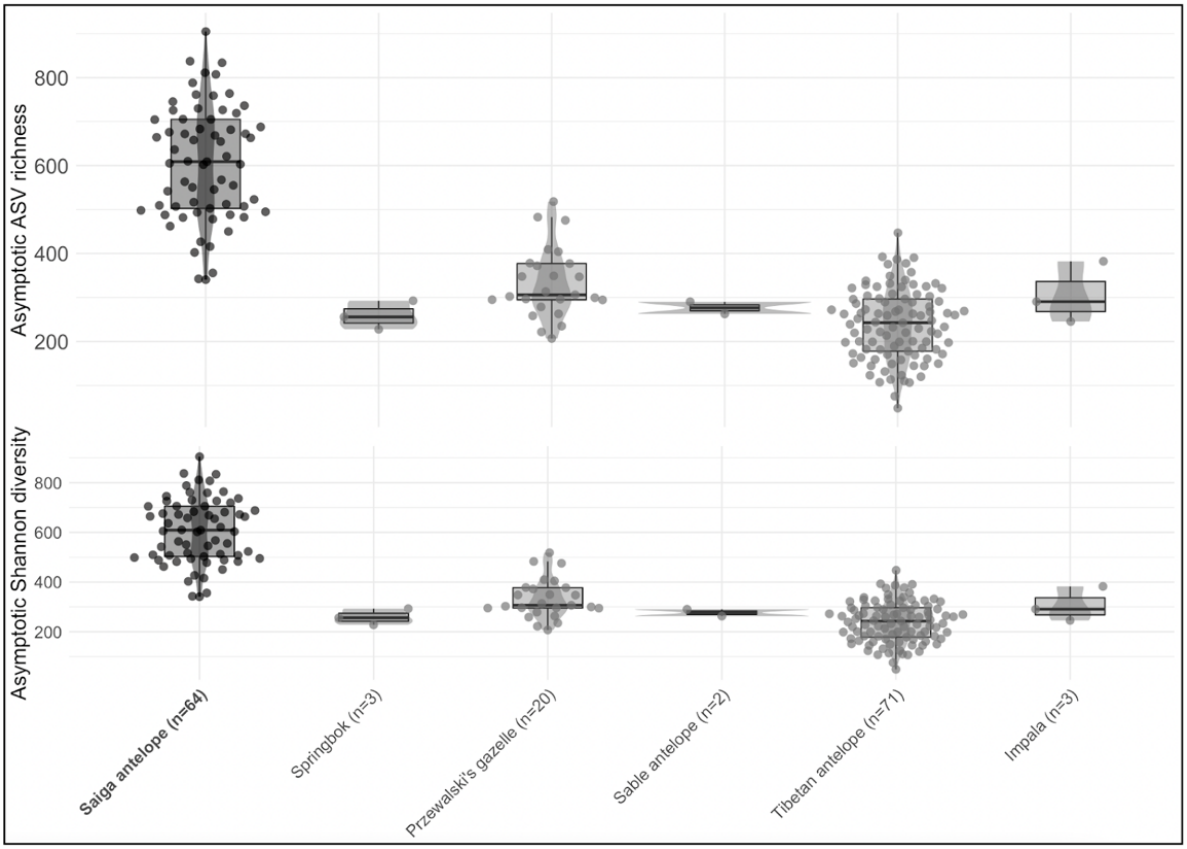
Supplementary Figure 4. (**A**) Euler diagram on shared and unique amplicon sequence variants in gut microbiota of six antelope species. (**B-D**) Principal coordinate analysis of gut microbiota (dis)similarity of six antelopes at amplicon sequence variants (ASV) level based on (**B**) Aitchison, (**C**) unweighted UniFrac, and (**D**) weighted UniFrac distances. Circles are individual samples. Colour indicates host species (*green* = Saiga antelope, *yellow* = Przewalski's gazelle, *blue* = Tibetan antelope, *orange* = Springbok, *pink* = Impala, *red* = Sable antelope).



Supplementary Figure 5. UpSet plots showing quantitative intersections of bacterial (**A**) phyla and (**B**) families between six antelopes. The numbers above the vertical bars indicate the number of common taxa between host species.



Supplementary Figure 6. Pairwise dissimilarity of bacterial genera (unweighted UniFrac distance) between Saiga antelope and (from left to right:) Springbok, Przewalski's gazelle, Sable antelope, Tibetan antelope, or Impala (antelopes ordered by phylogenetic relatedness to Saiga antelope starting with closest relative). Antelope species are ordered by phylogenetic relatedness to Saiga with alphabetical ordering where relatedness to Saiga cannot be distinguished (Sable antelope, Tibetan antelope). Differences were tested with permutational Wilcoxon rank sum tests (***, $p<0.001$; *, $p=0.034$; ns, $p=0.998$).



Supplementary Figure 7. Sample-level asymptotic estimates of ASV richness (*top*) and Shannon diversity (*bottom*) for six antelope species. Circles are individual samples. Antelope species are ordered by phylogenetic relatedness to Saiga with alphabetical ordering where relatedness to Saiga cannot be distinguished (Sable antelope, Tibetan antelope).

Supplementary Table 1. Datasets included in cross-host species analyses.

Host species	**Country of sample collection**	**Study reference**	**Sample size**	**16s rRNA region (primers)**	**Data access code**
Saiga antelope (*Saiga tatarica tatarica*)	Kazakhstan	Hanski et al.	64	V4–V5 (515F–926R)	
Tibetan antelope (*Pantholops hodgsonii*)	China	Shi et al., 2021	71	V3–V4 (338F-806R)	PRJNA673151
Przewalski's gazelle (*Procapra przewalskii*)	China	Liu et al., 2021	20	V3–V4 (338F-806R)	PRJNA684634
Impala (*Aepyceros melampus*)	South Africa	McKenzie et al., 2017	3	V4 (515F-806R)	Qiita: 10936
Springbok (*Antidorcas marsupialis*)	Namibia, South Africa	McKenzie et al., 2017	3	V4 (515F-806R)	Qiita: 10936
Sable antelope (*Hippotragus niger*)	South Africa	McKenzie et al., 2017	2	V4 (515F-806R)	Qiita: 10936

GENERAL DISCUSSION

In this thesis, I have explored how the mammalian gut microbiota varies across species, populations, and individuals as well as within individuals over time. Extensive within-species differences in the gut microbiota were detected between laboratory and wild mice, with wild mice having taxonomically and functionally more diverse gut microbiota with a faster turnover rate. Yet despite these differences, remarkably similar gut microbiota assembly patterns were detected in mice from these contrasting genetic and environmental backgrounds. As these age-related dynamics suggest age is an important variable for gut microbiota research, I developed an epigenetic method for age estimation from faeces. This method was capable of providing an approximate age estimate, but may be more useful for investigating biological rather than chronological aging in wild populations.

Gut microbiota differences between laboratory and wild house mice

The vast majority of gut microbiota research on the house mouse (*Mus musculus*) model system comes from the laboratory, where mice live in an artificial world that does not exist outside the laboratory, raising concerns about the relevance of such studies. As such, studies investigating differences between laboratory and wild mice, as well as drivers of gut microbiota in natural settings can aid interpretation of findings from laboratory-based gut microbiota studies. In Chapters I and III, I conducted extensive comparative studies of the gut microbiota across laboratory and wild mice. In line with previous findings (Rosshart et al., 2019; Wang et al., 2014; Kreisinger et al., 2014; Wang et al., 2015), the gut microbiota differed between lab and wild mice at nearly all levels of investigation, both taxonomically and in terms of diversity and functionality.

I complemented previous studies investigating gut microbiota differences between laboratory and wild mice with two novel angles: within-individual temporal dynamics as well as

aerotolerance phenotypes of the gut microbiota. The gut microbiota of wild mice had a faster within-host turnover rate and lower number of taxa that persisted over time than that of lab mice (Chapter I, Fig. 3), and the relative abundance of bacteria that can tolerate oxygen was almost twice as high in wild than in lab mice (Chapter III, Fig. 2). These results suggest that wild mice are exposed to a larger and more variable pool of microbes than are lab mice, which are housed in small, individually ventilated cages with only up to a few cage mates and minimal environmental fluctuation. Considering the differences in relative abundance of aerotolerant taxa, it may be that environmentally rather than socially transmitted bacteria form a larger part of gut microbiota in the wild compared to lab mice, assuming environmentally transmitted bacteria tolerate oxygen. In Chapter I, I identified soil as a possible source of environmentally transmitted microbes in wild mice as microbial similarity between soil and faecal samples increased with geographic proximity.

The different gut microbiota of the laboratory mouse from that of its wild relatives is the result of domestication; however, the move from the wild to the laboratory may have disproportionately affected certain features of the gut microbiota. Above I have discussed how the gut microbiota varies between lab and wild mice. I hypothesised that since the gut microbiota plays an important role in various developmental processes, patterns of gut microbiota assembly may be conserved across mice from these contrasting systems. In line with this idea, I found that laboratory and wild mice presented remarkably similar general patterns of gut microbiota assembly (Chapter III). For instance, as found in other mammalian species studies (Derrien et al., 2019; Baniel et al., 2022), as young mice developed the gut microbiota became more diverse within individuals but more homogenous across individuals in both lab and wild settings, and the relative abundance and richness of Proteobacteria showed highly similar changes across the two systems, despite lab mice having substantially fewer

proteobacterial taxa. These conserved assembly patterns illustrate the effect of host selection (such as depleting levels of oxygen in the gut, the shift from milk to solid food, and activation of the immune system) on gut microbial communities, which appear to be stronger than the effects of host genetics and environmental background.

Ultimately, what the microbes living in the gut are capable of doing is likely to be more relevant for host biology than what specific taxa we find, as the microbiota is known to display a high degree of functional redundancy and plasticity (Tian et al., 2020). As such, future investigation of functional differences in the gut microbiota of wild and lab mice would be a useful next step to provide key information about the consequences of this gut microbial variation. For instance, despite sharing similar assembly patterns in terms of taxonomy, the gut microbiota of lab and wild mice may nonetheless undergo differential functional changes in early life.

In this thesis, functionality of the gut microbiota was inspected using functional prediction from 16S data. In comparison to lab mice, wild mice had a higher number of unique functional pathways (Chapter I, Fig. 1). This is perhaps what one might expect given their more variable diet and abiotic environment. Diversity in diet can lead to diversity in gut microbial functionality since different functional pathways are required for utilisation of different dietary substrates. Variable abiotic conditions might also promote a higher number of functional pathways, as for instance cold temperature can increase a host's need for thermoregulatory pathways, some of which may be provided by gut microbes (Chevalier et al., 2015). However, I also detected similarities in functional profiles of adult lab and wild mice, particularly in the relative abundance of functional pathways (limited to visual inspection; Chapter III, Suppl. Fig. 7). These functional pathways may represent those that are important for the host

regardless of setting (lab/wild); for instance, those that are important in developmental processes or reproduction.

However, I limited analysis of the functional profiles to the level of pathways, as initial analysis suggested discrepancies between functional profiles at category level depending on which database was used to assign functional categories, highlighting limitations in the prediction based functional profiling used. Higher resolution functional profiling using e.g., metagenomic approaches would be needed for more thorough investigation of functional differences between laboratory and wild mice. This could allow us to identify functional implications of domestication of the house mouse (such as potential loss of functional pathways not needed in laboratory settings) and test whether functions that may have been lost during this process can be reacquired, e.g., when lab mice are exposed to dietary or abiotic changes. Future work could also investigate the degree of redundancy in the gut microbiota, and therefore the functional implications of lower diversity and altered composition in laboratory mice compared to their wild counterparts.

Thus, despite their 'altered' gut microbiota, it is possible to draw conclusions from laboratory mouse microbiota studies that are also relevant in natural populations. However, this is likely to be true only for certain phenomena associated with the gut microbiota, such as those strongly influenced by host selection. Further, this may also be true only for the taxonomic rather than functional aspects of the gut microbiota. Considering findings from Chapters I and III, my results highlight the importance of wild house mouse studies for understanding the relevance of gut microbiota findings obtained in laboratory settings.

Gut microbiota variation in the wild

Studying the gut microbiota of wild rather than laboratory mice has the advantage of representing the real world, where populations are genetically diverse, the environment changes, and space is shared with other animals; thus, findings can be easier to put into a wider ecological context. In line with previous work (Suzuki et al., 2020; Goertz et al., 2019; Weldon et al., 2015, Linnenbrink et al., 2013), I detected spatial influences on gut microbiota variation across and within wild house populations. In particular, broad and fine-scale spatial differences in the microbiota were detected within an extensively studied island mouse population (Chapter I, Fig. 5).

Interestingly, spatial variation was stronger for the fungal than the bacterial microbiota, possibly explained by abiotic differences (such as water availability) with effects on fungal communities in the host and environment. As such, the fungal microbiota, or 'mycobiota', specifically may provide a sensitive system for capturing spatial variation and transmission of host-associated symbiotic communities. Various factors from food availability to social interaction and the presence of other host species could explain why the gut microbiota shows spatial variation. Microbial communities in soil may present one such factor. Indeed, I found that microbial similarity between mice and soil increases with geographical proximity, which suggests some transmission from soil to gut may occur (Chapter I, Fig. 6). As microbial similarity increased with proximity also within soil samples, spatially varying soil microbiota could indeed explain some of the spatial variation seen in the mouse gut microbiota.

To separate microbial transmission from soil from other spatial factors shaping the gut microbiota, one could experimentally expose individuals from different environments to either standardised or different batches of soil, and characterise gut microbiota before, during, and

after the experiment. Alongside spatial influences on microbiota variation, weak effects of temporal and individual-level factors were detected. Particularly, influences from age-related factors were detected, in line with findings in Chapter III, although the effect size was smaller when contrasted against other possible drivers of gut microbiota. Despite the identified drivers of gut microbial composition across and within individuals, the vast majority of variation (>90%) remained unexplained. This may be due to the inability to identify other deterministic drivers (due to e.g., lack of data, such as temperature and diet, or bioinformatic approaches used), but it may also be that a substantial part of microbial variation arises from stochasticity.

Studying 'wild gut microbiota' beyond model systems has potential to increase our understanding of microbiota's role in ecology and provide important information of host biology with possible conservation applications. In the final chapter of my thesis, I characterised the gut microbiota of the critically endangered Saiga antelope, (in)famous for its distinctive pendulant nose as well as for suffering from recurring mass mortality events (Kock et al., 2018). It has been suspected that these mass mortalities are caused by outgrowth of *P. multocida*, a normally commensal symbiont of the Saiga. As such, the Saiga gut microbiota and the possible presence of this symbiont in healthy Saigas has been of interest.

While I predicted that the Saiga microbiota might be atypical in some way (e.g., particularly low in diversity), if it has a role in allowing pathobiont outgrowth during die-offs, the Saiga gut microbiota was similar to that of other antelopes from Asia and Africa. This suggested no obvious sign that the baseline gut microbiota of the Saiga differed in ways that could predispose them to mass die-offs, although more work is needed to investigate the possible link between host-associated microbiota and susceptibility to mass mortalities more thoroughly, such as characterisation of respiratory microbiota, functional profiling and/or deeper sequencing of gut

microbiota, and importantly investigation of possible changes in gut microbiota during a mass mortality event. The gut microbiota profile of the Saiga during a year without die-offs provides a baseline for future work investigating the microbiota's role in mass mortalities, and also more widely contributes to our understanding of the Saiga and its ecology.

Estimating age in wild populations

As seen in Chapter III, age is a key variable in gut microbiota dynamics. Thus, a downside of using wild mice as a model system is that their precise age is often unknown. In Chapter II, I attempted to overcome this limitation by predicting age in wild mice with an epigenetic clock built using calibrated age data from laboratory mice. This method worked to some extent, but it did not seem to provide more accurate estimates of age than e.g., body mass, thus the method was not used in other chapters. As others who have had better success predicting age in wild individuals of unknown age have used captive animals from zoos to build an epigenetic clock (Mayne et al., 2022; Robeck et al., 2021), I hypothesize the inaccuracy of my clock may arise from laboratory and wild mice being genetically too different and living in contrasting environments with laboratory mice lacking stressors like food shortage and disease, that may be important drivers of DNA methylation in the wild. To refine this method further and better predict chronological age in wild mice, it may be more fruitful to train an epigenetic clock using individuals from semi-natural settings rather than lab animals. For example, the barn-dwelling mice that are intensively studied at the University of Zurich (Ferrari et al., 2022) provide a system where close monitoring of breeding attempts and individuals means age can often be known precisely, but individuals are outbred and the environment is variable, which may increase the clock's accuracy for application in wild settings.

While the approach used did not provide an accurate estimate of *chronological* age, it may be that it captures useful signals of *biological age* (the 'wear and tear' of the body), which can be influenced by various genetic and environmental factors (Han et al., 2018; Joubert et al., 2016). As wild mice had substantially higher levels of DNA methylation in general (Chapter II, Suppl. Fig. 2), and experience far more varying environmental conditions that can affect their rate of biological aging, studying the drivers of biological aging could be powerfully addressed in the future with large datasets from wild mouse populations involving repeat captures to estimate change in epigenetic markers over time. Such studies may be more informative for understanding drivers of biological aging, or at least complement, those conducted in highly controlled lab environments. Perhaps future research will even identify links between the gut microbiota and biological aging.

Final conclusions

The mammalian gut microbiota is evidently important for the host, but many aspects of it remain poorly understood. In particular, what drives variation across and within individuals over time remains largely unknown. The laboratory mouse is widely used as a model system in gut microbiota research, despite its gut microbiota being distinct from that of wild house mice. In this thesis, I have illustrated how the setting (laboratory vs wild) can affect some but perhaps not all aspects of the gut microbiota and beyond. Laboratory and wild mice have gut microbiotas that are compositionally substantially different but these microbial communities undergo similar assembly processes in early life. Further, laboratory and wild mice age epigenetically at fairly similar rates, but on entirely different scales, with wild mice having remarkably higher levels of DNA methylation throughout life. I have further demonstrated that spatial influences on gut microbiota variation are stronger on the fungal than bacterial part of the microbiota, and that microbial transmission from soil may be one source of spatial

variation. Together the results from this thesis advance gut microbiota research by aiding interpretation of lab-based studies as well as by highlighting the potential in using novel approaches informed by wild populations.

References

Baniel A, Petrullo L, Mercer A, et al. Maternal effects on early-life gut microbiota maturation in a wild nonhuman primate. *Curr Biol*. 2022;32(20):4508-4520.e6. doi:10.1016/j.cub.2022.08.037

Chevalier C, Stojanović O, Colin DJ, et al. Gut Microbiota Orchestrates Energy Homeostasis during Cold. *Cell*. 2015;163(6):1360-1374

Derrien M, Alvarez AS, de Vos WM. The Gut Microbiota in the First Decade of Life. *Trends Microbiol*. 2019;27(12):997-1010. doi:10.1016/j.tim.2019.08.001

Ferrari M, Lindholm AK, Ozgul A, Oli MK, König B. Cooperation by necessity: condition- and density-dependent reproductive tactics of female house mice. *Commun Biol*. 2022;5(1):348. Published 2022 Apr 12. doi:10.1038/s42003-022-03267-2

Goertz S, de Menezes AB, Birtles RJ, et al. Geographical location influences the composition of the gut microbiota in wild house mice (Mus musculus domesticus) at a fine spatial scale. *PLoS One*. 2019;14(9):e0222501. Published 2019 Sep 26. doi:10.1371/journal.pone.0222501

Han Y, Eipel M, Franzen J, et al. Epigenetic age-predictor for mice based on three CpG sites. *Elife*. 2018;7:e37462. Published 2018 Aug 24. doi:10.7554/eLife.37462

Joubert BR, Felix JF, Yousefi P, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet*. 2016;98(4):680-696. doi:10.1016/j.ajhg.2016.02.019

Kock RA, Orynbayev M, Robinson S, et al. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv*. 2018;4(1):eaao2314. Published 2018 Jan 17. doi:10.1126/sciadv.aao2314

Kreisinger J, Cížková D, Vohánka J, Piálek J. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. *Mol Ecol*. 2014;23(20):5048-5060. doi:10.1111/mec.12909

Linnenbrink M, Wang J, Hardouin EA, Künzel S, Metzler D, Baines JF. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol Ecol*. 2013;22(7):1904-1916. doi:10.1111/mec.12206

Mayne B, Mustin W, Baboolal V, et al. Age prediction of green turtles with an epigenetic clock. *Mol Ecol Resour*. 2022;22(6):2275-2284. doi:10.1111/1755-0998.13621

Robeck TR, Fei Z, Lu AT, et al. Multi-species and multi-tissue methylation clocks for age estimation in toothed whales and dolphins. *Commun Biol*. 2021;4(1):642. Published 2021 May 31. doi:10.1038/s42003-021-02179-x

Rosshart SP, Herz J, Vassallo BG, et al. Laboratory mice born to wild mice have natural microbiota and model human immune responses. *Science*. 2019;365(6452):eaaw4361. doi:10.1126/science.aaw4361

Suzuki TA, Martins FM, Phifer-Rixey M, Nachman MW. The gut microbiota and Bergmann's rule in wild house mice. *Mol Ecol*. 2020;29(12):2300-2311. doi:10.1111/mec.15476

Tian L, Wang XW, Wu AK, et al. Deciphering functional redundancy in the human microbiome. *Nat Commun*. 2020;11(1):6217. Published 2020 Dec 4. doi:10.1038/s41467-020-19940-1

Wang J, Kalyan S, Steck N, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun*. 2015;6:6440. Published 2015 Mar 4. doi:10.1038/ncomms7440

Wang J, Linnenbrink M, Künzel S, et al. Dietary history contributes to enterotype-like clustering and functional metagenomic content in the intestinal microbiota of wild

mice. *Proc Natl Acad Sci U S A*. 2014;111(26):E2703-E2710. doi:10.1073/pnas.1402342111

Weldon L, Abolins S, Lenzi L, Bourne C, Riley EM, Viney M. The Gut Microbiota of Wild Mice. *PLoS One*. 2015;10(8):e0134643. Published 2015 Aug 10. doi:10.1371/journal.pone.0134643