The Intestinal Microbiota of two Arid-adapted rodents

Peter Christopher William Osborne

A thesis presented for the degree of
Doctor of Philosophy

Earlham Institute
University of East Anglia
UK
December 2022

Abstract

The potential impact of aridity on host-associated microbiomes has been little investigated previously. This study describes the results of bioinformatic and culturing based analysis of a number of faecal samples from two arid-adapted rodent species, *Acomys cahirinus* and *Acomys russatus*; from individuals living sympatrically in the Judean Desert, Israel. 81 faecal samples were collected from two sampling points in June and November of 2016, some animals providing a sample on each occasion. These were sequenced and subjected to bioinformatic analysis to determine the taxonomic composition of the faecal microbiota. Metagenomic bins were generated from the reads obtained from the faecal samples and these were used with the faecal reads to determine if there were statistically significant differences between each host species microbiome and within each species between the two sampling months. These bins were also taxonomically identified and functionally annotated to assist with the development of genomic databases by addition of material from a novel host environment. Guided by bioinformatic analysis, targeted isolation of lactic acid bacteria from the *Acomys* faecal samples was carried out using selective media. From this, 28 distinct lactic acid bacteria isolates were obtained, sequenced and assembled. Putative taxonomic identities for these isolates were obtained, suggesting some are novel species of lactic acid bacteria. 8 of the isolates were also used for halotolerance testing to assess whether an observed host phenotype might also be detected in members of the microbiota. Growth of some isolates on media with 3.5% salinity (comparable to sea water) was observed. Statistically significant differences between the two host species were observed, with limited differences within hosts at each time point. Potentially beneficial functions for the hosts were identified in the faecal microbiome and detected in isolates cultured from faecal samples.

Contents

List of Figures

List of Tables

List of Abbreviations

DNA Deoxyribonucleic Acid

16S Referring to the 16S rRNA gene

ASV Amplicon Sequence Variant

iMGMC Integrated Mouse Gut Metagenome Catalogue

ITS Referring to the fungal Internal Transcribed Spacer region of nuclear DNA

LAB Lactic Acid Bacteria

MAGs In context, Metagenome Assembled Genomes

NGS Next Generation Sequencing

NCBI National Center for Biotechnology Information

OTU Operational Taxonomic Unit

PCR Polymerase Chain Reaction

QC In context, quality control

RPM In context, Reads per-Million

WGS Whole Genome Shotgun Sequencing

Acknowledgements

There are a large number of people I need to thank for their help and support, both direct and indirect, in getting this thesis over the finish line. My studies were funded by the BBSRC and could not have been carried out without this financial backing.

Everyone in the Hall group in both Norwich and Munich who helped me out, especially when access to the labs was restricted. Nancy, Iliana and Antia in particular need credit here as well for their part in getting the wet lab work done. Without them this project could not have been completed. Lindsay herself was a sterling secondary supervisor, supportive and hands-off when appropriate. The good folk at the Earlham Institute, especially in the Haerty and Di Palma groups. The general support, banter and atmosphere in the combined group was extremely helpful in pushing through during the years and maintaining a baseline level of competence. Though stress inducing the lab meetings certainly helped improve the project and the advice on different tools and resources was extremely handy. Will and Graham were both especially useful in their own ways, and Angela was an appreciated presence with metagenomics knowledge in the group. Becky in particular was a fantastic addition to the group. I'm sad we only got two years at EI together but she'll accomplish great things in the future and hopefully remember the cranky old student who bugged her daily. Outside the combined groups the crew in G209 and G208 were great fun and reminded me that science is not supposed to be a dour and serious profession at all times. I was immensely fortunate to always end up in offices full of great people.

My family provided constant support and encouragement, I wouldn't have gotten to this point without their help and backing.

Dave and Laura. I was lucky beyond measure to have had the finest postdoc in the institute in my group when I started and throughout the 4 years. After the first year I was on the way out the door but Dave kept me going and showed what a true scientist can be like. The warmth and kindness they both offered by inviting me into their lives was unexpected and I cannot thank them both enough.

Wilfried, for taking me on in my hour of need and being a fantastic supervisor. Always an encouraging word and always making time in an inhumanely busy schedule if I wanted it. Introducing neat ideas to me and supporting my progress as well as always knowing who the right person to go to for technical help saved me countless headaches. I appreciate that he lost out by accepting me but am immensely grateful he did.

Dr's Anita, Matt and Sam. Thrown together by chance and friends by choice, I owe all three a debt I cannot repay. Helping me keep sane, commiserating about the challenges and being great craic. Truly I would not have finished my studies without them. All three contributed far more to EI than they received and will excel at whatever they do in the future. I was incredibly fortunate to have been in the same intake as them.

Aims and objectives of the project

The aim of this research project is to determine whether and how the faecal microbiome differs between two arid adapted rodent species, and whether and how it differs between two points within the rodent species. Ultimately it aims to assess if there are any links between the faecal microbiome and survival in arid conditions of two rodent species, *Acomys cahirinus* and *Acomys russatus* living sympatrically in the Judean Desert in Israel.

Research objectives

The research objectives are as follows:

1. To assess the ability of existing metagenomic tools and databases to analyse faecal microbiome data from both Acomys species.
2. To establish the taxonomic composition of the faecal microbiome of both *Acomys* species.
3. To functionally characterise the faecal microbiome of both *Acomys* species.
4. To investigate whether bacteria isolated from *Acomys* faecal samples have phenotypic traits which might be associated with observed phenotypic traits of their host.

Research questions
The study proposes the following research questions:

1. What tools are frequently used to analyse metagenomic data?
2. How effective are these tools when used with samples likely to contain novel diversity?
3. What taxa make up the faecal microbiome of *A. cahirinus* and *A. russatus*?
4. Are there links between the host species and/or the time of sampling and the faecal microbiome of *Acomys* species?
5. Can bacterial isolates from *Acomys* be obtained using selective media designed for use with *Apodemus* mice?
6. Are bacteria isolated from *Acomys* faecal samples capable of growing in hypersaline conditions?

Hypotheses

The following hypotheses are proposed:

1. The faecal microbiota differs significantly between two differentially arid-adapted *Acomys* species.
2. The faecal microbiota differs significantly within two differentially arid-adapted *Acomys* species depending on the season, Summer or Winter, sampled.
3. The faecal microbiota of two differentially arid-adapted *Acomys* species provide functional traits assisting in the aridity tolerance of the host rodent.

Description of chapters

Chapter 1 - Introduction

Chapter 1. gives a general introduction to the study of the microbiota, the different approaches often employed in investigations of the microbiota and defines some key terms. It also highlights the strengths and weaknesses of different bioinformatic approaches to microbiota investigations, recounts some examples of notable findings from different approaches and examines existing findings concerning arid associated microbial communities.

Chapter 2 - Methods

Chapter 2. describes the different methods used in this project. They include sampling information, bioinformatics tools used along with their parameters and the experimental setup of isolation of bacteria from *Acomys* faecal samples. It concludes with a list of all software used and their version.

Chapter 3 - Results

Chapter 3. presents the results of the different analyses conducted in this project, presenting in turn the findings of the different methodologies used.

Chapter 4 - Discussion

Chapter 4. expands on the results and provides some more context for them, relationships to other studies are discussed and the results are interpreted in light of the initial hypotheses.

Chapter 5 - Conclusions

Chapter 5. discusses the inherent limitations of the project, the profound impact of the global COVID-19 pandemic on the project and what further investigation was planned and could be conducted in the future.

Declaration

This thesis is the result of my own work and includes nothing which is the product of work done in collaboration **except where indicated in the text and listed below**.

This thesis involved contributions from a number of individuals. In the relevant section of the Methods chapter the names and contributions are described, they are also listed here below for ease of consultation.

1. The *Acomys* trapping, tagging and sampling was conducted in Israel by Ella Pasternak.
2. The DNA extraction of the *Acomys* faecal samples was carried out by Iliana Serghiou
3. Assembly and binning of reads from the *Acomys* faecal samples after sequencing was carried out by Professor Chris Quince and Dr. Sébastien Raguideau
4. The culturing and isolation of bacteria on lactic acid selective media from Acomys faecal samples was conducted by Nancy Teng
5. Sequencing of the bacterial isolates obtained from *Acomys* faecal samples was carried out by Dave Baker
6. Lactic acid bacteria isolates obtained from *Apodemus* faecal samples were provided by Dr Magdalena Kujawska
7. Experimental halotolerance culturing of *Acomys* and *Apodemus* isolates was carried out by Antia Acuna-Gonzalez
8. *Acomys* reference genomes were generated by Dr Shane McCarthy, Dr David Thybert and Dr Kerstin Howe

Peter Osborne

December 2022

Chapter 1

Introduction

- The concept of the Microbiota and its study is outlined in broad terms.
- The use of computational tools to study the Microbiota is discussed.
- The use of culturing and traditional microbiological approaches in combination with bioinformatics to study the Microbiota is discussed.
- Broad patterns and major findings from investigation of environmental Microbiota are discussed.
- Broad patterns and major findings from investigation of host-associated Microbiota are discussed.
- The relationship between the Microbiota and host tolerance of environmental stresses is discussed, with a focus on aridity tolerance and the relative lack of published research examining the Microbiota of arid-adapted species.

1.1 A general outline

The existence of microorganisms in the wider environment was not a recent discovery, they had been known of, to some degree, for centuries [1] and studied in depth for decades [2, 3, 4, 5]. With the discovery of the structure of the Deoxyribonucleic Acid (DNA) molecule [6] and subsequent advances in accessing [7, 8, 9], sequencing [10, 11, 12, 13] and analysing genomic information [14, 15] it became clear that there were communities of microorganisms almost everywhere on earth [16, 17, 18]. With time and study the concept of the 'Microbiota' began to develop, defined as 'The assemblage of microorganisms present in a defined environment' [19]. Researchers discovered that the Microbiota can provide a wide range of functions for organisms which host them, ranging from access to recalcitrant energy sources [20], protection from pathogens [21], ensuring normal development [22] and tolerance of environmental stresses [23]. The microbiota became a target of investigation in particular for medical research due to the potential it appeared to have for addressing illnesses [24] and promoting or maintaining good health [25]. The possibility that the microbiota might also play a key role in the ability of host organisms to remain healthy or even survive in their standard conditions led to investigation from a conservation perspective [26, 27]. The compounds produced by the microorganisms of microbiota from different environments are of themselves a source of interest, in particular as novel components for medicinal [28] or industrial processes [29]. Investigation of the Microbiota can be carried out through computational analysis of sequenced reads, traditional microbiological methods or both in combination.

1.1.1 A note on terminology

There are two other terms commonly used in discussion related to this field, the 'Microbiome' and the 'Metagenome'. The Microbiome is defined as 'The entire habitat, including the microorganisms (bacteria, archaea, lower and higher eurkaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions' [19]. The Metagenome is defined as 'The collection of genomes and genes from the members of a microbiota' [19]. In this study the term 'Microbiota' is used most frequently as the project focuses on the taxonomy of the microbial communities being studied along with their relationship to the host organisms. When quoting or referring to the work of others one of the other two terms may instead be used when following the use of the cited authors.

1.2 Bioinformatic investigation of the Microbiota

The two principal means of bioinformatic investigation of microbial communities from DNA sequencing of the environment they reside in are targetted sequencing of particular genes, typically the 16S rRNA gene when dealing with Bacteria, and untargetted sequencing of any genetic material that can be obtained from the environment; this latter approach is typically referred to as 'Shotgun sequencing'.

1.2.1 16S and other marker gene based approaches

16S rRNA sequencing (Referring to the 16S rRNA gene (16S)) takes advantage of the ubiquity of the gene and the very high level of conservation it retains across time [30]. Since the development of the approach [31] it has become an extremely common and widely used approach for study of microbial communities, especially as the cost of sequencing in general decreased [32, 33] along with the rapid increase in affordable computation. The conserved regions of the gene can be targetted with specific primers and PCR employed to greatly increase the amount of material available for analysis from a sample, which can make the approach quite useful in situations with a limited sample supply. The variable regions allow for computational analysis to identify [34] and delineate taxa [35, 36] on the basis of their similarity to each other in these gene regions. The use of 16S sequencing for these purposes has been a mainstay of research for decades and there has been much published on the advantages [37] and limitations [38, 39, 40, 41, 42] of the approach. Clarridge [43] provides a good outline of the actual mechanics of generating 16S sequencing data for analysis.

The use of the 16S rRNA gene for establishing phylogenetic relationships between different microbes, or equivalent marker genes in Fungi, has been a major part of metagenomics for years. It has been presumed that differences in the sequence of the gene were due to speciation and reflected real phylogenetic differences between the genomes the genes originated in. However in recent years there has been evidence emerging that the gene is subject to the same processes of horizontal transfer [44, 45] as other genes. Hassler *et. al.* [46] conducted an investigation into the degree of similarity between phylogenies produced using the 16S rRNA gene and those obtained using concatenated core genes; they found the 16S rRNA gene to be a poor tool for phylogenetics due to poor species and strain delineation.

16S sequencing has some direct, practical advantages over Whole Genome Shotgun Sequencing (Whole Genome Shotgun Sequencing (WGS)). It is usually cheaper to conduct the extraction, amplification and sequencing of a specific region of DNA as opposed to attempting to capture as much of the sum total of genetic material present in a sample. A number of primers exist which allow for targetting of specific bacterial taxa of relevance to investigators [47, 48] and if the identity of a microorganism needs to be established with a high degree of confidence relatively rapidly - such as for diagnostic [49, 50, 51] or medical research purposes [52] - then it is likely the superior method of the two discussed here. The relationship between different primers and certain domains has been reviewed multiple times, including by Baker *et. al.* [53]. The 16S gene was used in a number of different investigative approaches, as discussed by Zoetendal *et. al.* [54] but over time the most commonly employed method was, and remains so at the time of writing, 16S rRNA gene sequencing. The development of Next Generation Sequencing (Next Generation Sequencing (NGS)) and Long Read Sequencing have made 16S sequencing-based investigations faster [55], larger in scope [56] and with a greater degree of resolving power - through sequencing of the entire 16S gene [57] - while simultaneously introducing questions about the underlying principles of the approach [58]. These being the inability to provide a spe-

cific taxonomic identification at a level below the genus, the total absence of directly observable functional information and reliance on inferred function along with the errors created when amplifying the reads using PCR (Polymerase Chain Reaction (PCR)). Steps were, and still are, taken to try and improve the approach through creation of new methods [59] and the refinement of existing ones [60].

16S sequencing has been used in countless investigations into the Microbiota of different environments, including host-associated ones. Ley *et. al.* [61] employed it in their investigation into the intestinal microbiota its relationship to obesity, detecting phylum level differences associated with obesity. Dupont *et. al.* [62] used the approach to study the microbiota of a carnivorous sponge (*Asbestopluma hypogea (Cladorhizidae)*) from the deep ocean. Panke-Buisse *et. al.* [63] used 16S sequencing to link differences in soil microbiota to the flowering times of *Arabidopsis thaliana*. Marteinsson *et. al.* [64] utilised 16S sequencing alongside other approaches when studying the microbial community found in the water beneath an Icelandic ice cap which helped them conclude there may be a connection between two subglacial lakes. Newton *et. al.* [65] uncovered faecal pollution of lake waters from a nearby urban area in Michigan through the use of 16S sequencing. More recently, Mohd-Yusof *et. al.* [66] used 16S sequencing when they investigated the bacterial pathogens which could be found in the microbiota of a fruit eating bats on two Malaysian islands. Leclerc *et. al.* [67] used 16S rRNA gene sequencing to study the diversity of archaea from a range of anaerobic digesters treating a range of material while Pereira da Fonseca *et. al.* [68] used 16S rRNA gene sequencing to assess the bacterial community found on Brazilian bank notes.

The utility of 16S sequencing for examining Archaea in addition to Bacteria can also be seen in the many studies which have employed it to investigate the proportion of different microbial communities made up of Archaea. Cunha *et. al.* [69] used the approach to assess both the bacterial and archaeal component of the rumen microbiota of Brazilian goats (*Capra hircus*). Probst *et. al.* [70] used 16S sequencing to inspect the archaeal component of the human skin microbiota and reported that up to 4.2% was composed of Archaea; though more recent work by Umbach *et. al.* [71] also using 16S sequencing proposes a value between 1 and 2% instead.

Fungal Internal Transcribed Spacer (Referring to the fungal Internal Transcribed Spacer region of nuclear DNA (ITS)) regions have been employed in the same manner as the 16S rRNA gene for investigation of the fungal component of various microbial communities; as the approach shares many of the underlying rationale concerning species specific regions, measureable evolutionary rates and relative ease of sequencing [72]. In their study of the microbiota in crystalline rock 2 kilometers below the surface Miettinen *et. al.* [73] used both 16S sequencing to assess the bacterial and archaeal members of the community but also employed ITS region sequencing to analyse the fungi which were present. ITS sequencing let Abdelfattah *et. al.* examine the fungal members of the microbiota of the Olive (*Olea europaea*), identifying 195 different Operational Taxonomic Units. Suhr *et. al.* [74] used ITS sequencing in their study of the fungal members of the human gut microbiota, detecting both longer term resident fungi and transient members. Rittenour *et. al.* [75] employed ITS sequencing to investigate the fungal community inside the homes of asthmatic children. More broadly, the potential to use marker based approaches for the investigation of microbial eukaryotes within a microbiota has great potential - especially in light of the relatively greater size of eukaryotic genomes as compared to prokaryotic ones; the development of tools like taxaTarget by Commichaux *et. al.* [76] demonstrate the desire in the microbial ecology community to develop and implement such approaches.

16S sequencing computational tools and approaches

The 16S region sequence reads can be analysed using a variety of tools but with two broad, overarching approaches. These are whether the researcher is interested in the taxonomy of the microbiota as determined by Operational Taxonomic Unit (OTU)s or by Phylotype (similarity to a reference sequence). These can be interpreted respectively as whether the researcher is interested in the diversity of taxa within a sample irrespective of any external references, or the composition of the community in terms of previously identified taxa. In some circumstances, medical diganostics most obviously, the phylotyping approach is the most useful as the clinician or investigator is primarily concerned with the presence or absence of some particular taxa which have already been identified and are present in an external reference database [77, 78]. The OTU based approach may be of more use for researchers who are investigating a microbiota which has been little studied previously and where they might be comparing taxonomic diversity to some other variable within their study; for instance the work of Siddiqui *et. al.* [79] looking at the diversity of the female urine microbiota. In recent years there has been a move from OTUs to Amplicon Sequence Variant (Amplicon Sequence Variant (ASV)) due to the increasingly clear strain-level specificity of some genes and the restricted size of the core genome within an OTU, as discussed by Fernádez *et. al.* [80].

Multiple pipelines exist for the processing of 16S sequencing data, both at the OTU and ASV level for researchers taking that approach. QIIME [81] (Quantitative Insights Into Microbial Ecology) and QIIME2 [82] are quite commonly used, with many studies having used QIIME to produce high impact publications. MOTHUR [83] is another frequently used pipeline in 16S sequencing studies, such as Desai *et. al.*'s [84] study into the relationship between dietary fibre, the gut microbiota and pathogen susceptibility. DADA2 [85] is an ASV level pipeline which is both standalone and has been included within QIIME2. USEARCH [86] as a tool includes the pipelines UPARSE [87] and UNOISE3 [88]. Prodan *et. al.* [89] provide a review of these pipelines which concludes that the best balance between resolution and specificity was offered by USEARCH-UNOISE3. Typical measures reported in 16S sequencing studies are the alpha and beta diversities, usually presenting multiple metrics for each including - but not limited to - species richness, Chao1, Shannon index, Bray-Curtis and Jaccard distance. Dedicated R packages exist for 16S analysis after the initial data generation and processing of the sequencing output including Vegan [90] and Phyloseq [91]. As 16S sequencing does not capture genetic content outside the amplified region tools have been developed to infer function from the taxonomic identities provided by the 16S analysis, these include PICRUSt [92], PICRUSt2 [93], Tax4Fun [94] and Piphillin [95].

The choice of reference database used in a 16S sequencing based investigation can have a major influence on detected taxa and abundances as measured by the different tools which can be utilised. The most frequently employed databases include Silva [96, 97], RDP [98] (Ribosomal Database Project) and Greengenes [99], some researchers may use multiple simultaneously and take a consensus result from two or more of these databases - though this can lead to difficulties due to the reported errors when providing the same data for comparison to the different databases [100]. More tailored databases have also been created which focus on microorganisms associated with particular environments of interest to many investigators, such as CORE [101] for the oral microbiota, DAIRYdb [102] for milk and related dairy product microbiota and MiDAS [103] for the microbiota of wastewater. Researchers looking to create a custom 16S reference database for their own project can avail themselves of published guides and methods on how to achieve this, such as the work of Dueholm *et. al.* [104] with synthetic long read sequencing.

Refinements of 16S sequencing such as the development of ASV approaches with their asso-

ciated benefits [105] or sequencing the 16S-23SrRNA encoding region rather than the 16S alone [106] have gone some way to addressing the limitations [107, 108, 109, 110] of the approach but it remains the case that outside of descriptive studies or use as a rapid diagnostic the approach cannot provide as much insight into the actual genomes - and thus total genetic content - of the microbes of assessed microbiota.

1.2.2 Shotgun sequencing

Shotgun sequencing was the approach used in this project and has been used in a wide variety of studies investigating different microbial communities, becoming increasingly common as the cost of sequencing and - crucially - the cost of computation decreased. In contrast to 16S or other marker gene based approaches, shotgun sequencing of the genetic material from an environment can allow the direct detection of particular genes. Functions can therefore be observed rather than inferred based on taxonomic similarity, through capturing entire genes [111] or allowing the assembly of entire genomes [112]. It can allow for analysis and interpretation of sequencing data of a microbial community even if there is nothing known about the taxa it contains. Shotgun sequencing of a microbiota can allow for strain level differentiation [113], which might provide essential context for understanding the community. Assembly of microbial genomes from sequencing reads requires the entirety of the genome to be sequenced, even if in fragments, so assemblies cannot be produced using 16S sequencing [114]. Kashaf *et. al.* [115] provide a protocol for obtaining prokaryotic genomes from short-read shotgun sequencing of microbiota samples.

That said, shotgun sequencing can often detect significantly fewer taxa than 16S sequencing, as found by Tessler *et. al.* [116] in their investigation of water samples. It can also lead to false diversity, species counts and diversity metrics being artificially inflated by misaligned reads - a problem which can become more acute with increased sequencing depth [117]. Taxonomic classification from shotgun sequencing also relies on comparison to reference databases, this might be using marker sequences [118] or short nucleotide sections called 'kmers' [119]. As with 16S sequencing, shotgun sequencing can be of limited utility if there is a large proportion of unknown or unclassified taxa in the samples being investigated. A common solution, creating tailored databases containing the taxa likely to be detected in the samples [120], may not be practical with some samples at this stage due to the limited knowledge of what might be considered typical composition for them. Studies which employ shotgun sequencing are unlikely to capture an entire microbial genome from stochastic environmental sampling. Even with the decreasing cost of sequencing, a much greater depth of coverage is required to exhaustively sequence all organisms in the samples and obtain genomes of comparable quality to isolation and cultivation. This is especially true for less abundant members of the microbiota or when trying to ensure coverage of the functional potential [121] of a microbial community as opposed to sequencing purely for taxonomic quantification. The growing number of Metagenome Assembled Genomes which have been published [122, 123, 124] demonstrate though the utility of shotgun sequencing and that it is not an absolute requirement to have cultured microbes in order to carry out in-depth research [125, 126]. Cost of sequencing has decreased with time [32, 33] but there is still a notable and sometimes prohibitive difference in price between carrying out shogtun sequencing of a microbita sample versus targeted sequencing of specific regions of DNA such as 16S or ITS regions. The use of long read technology for shotgun sequencing may provide a great deal of information as compared to short read shotgun sequencing [127] alone. Combining long and short read data can be especially informative [128].

There are difficulties which can be encountered when employing different types of taxonomic

classification software on microbiota samples. In order to identify the members of a microbial community from genetic sequence data it is necessary to carry out computational analysis of the sequences. Taxonomic identification might be performed through different approaches. The first focuses on matching a part of the sequence data to a specific reference marker/markers (e.g. Metaphlan versions). Another approach relies on databases of predefined kmers (relatively short sections of bases) with the sample reads being compared to the databases for matches; then using statistics to identify the most likely taxon to contain them (e.g. Kraken versions). The other approach involves translation of the genetic sequence data into amino acid sequence and comparison to a protein database (e.g. Kaiju versions [129]). There are inherent limitations with each of these methods, including database incompleteness for marker and kmer based approaches. Marker based tools such as Metaphlan inherently correlate the likelihood of detection to the presence of the specific marker(s) in the sequenced reads [130] leading low abundance taxa to be missed at lower sequencing depths. Kmer-based methods suffer from being more computationally demanding -especially for RAM- even after multiple versions which improve their efficiency [131]. More directly related to the biology being investigated, kmer-based approaches can provide a number of false positive detections [132], caused by highly conserved regions shared across taxa combined with shorter kmers. This is especially relevant for low abundance taxa in which the conserved regions may be the only sequenced data. Many of the discussed methods depend on reference databases, as a consequence of the tendency towards sampling a limited number of host organisms and environments; these reference databases can cause issues when trying to analyse samples from previously unexplored sources. A good outline for the design of a shotgun-sequencing based investigation of microbial communities is provided by Quince *et. al.* [133].

The marker-based approach to taxnomic classification, when using shotgun sequencing rather than 16S/ITS, depends in a large part on the set of markers used. For Metaphlan, one of the most commonly used marker-based taxonomic classifier with shotgun sequencing data, the authors of the original paper [134] highlight that by using clade-specific marker genes they can conclude that a clade is present in a sample. The advantages of this approach is that it is computationally efficient, since mapping of sample reads is to this smaller subset of all collected genetic information; 4% at the time of the original paper. It also allows for easier storage through use of this smaller reference database. The same basic principle is in effect with the updated versions of Metaphlan, including Metaphlan 4 [135], with the number of marker genes being expanded and being looked at in the context of identifying 'species-level genome bins' rather than just traditional clades. Metaphlan 3 [136] can provide strain level resolution through the use of strain-level specific markers. By restricting itself to the clade or 'species-level genome bin' level the Metaphlan markers negate concerns around the impact of horizontal gene transfer and gene duplication which might arise through other approaches. The authors of mOTUs [137] initially used 40 universal single copy marker genes, these markers had been used prior to this for delineating prokaryotic species [138]. The authors then reduced this down to the 10 most suitable genes. The updated mOTUs, mOTUs2 [139], includes an expanded set of marker genes after taking advantage of the increased number of metagenomic studies undertaken to boost the number of mOTUs to grow the database. The older tool, MetaPhyler [140], employed 31 phylogenetic marker genes and the authors highlighted the importance they placed on using the length of the gene fragment for tuning the taxonomic classification. A major philosophical distinction, which has become less prevalent over time and with the iterating and development of the tools, is whether marker genes should only be sought in entire genomes. That is to say whether a marker should only be considered as a reliable and true marker for the presence of the organism if the genome of said organism has been fully sequenced and published. This restriction inherently limits the diversity and applicability of the tool associated with it but eliminates totally any possibility that detection of

a marker might instead be associated with the presence of some other organism in the database. It does not, however, eliminate the possiblity that what was previously considered a exclusive marker might be shared with novel taxa in a sample which have not been previously sequenced. There is also always the potential for contamination, though processing of data with tools such as Decontam [141] can alleivate this risk the use of marker-based approaches is as vulnerable to false positives from contamination as any other approach. Though the authors were interested in refining 16S sequencing analysis, the tool MATAM [142] which assembles full (or as close to it as possible) length marker genes from short read shotgun sequencing data after identifying reads originating from that marker, suggests a possible refinement of marker-based approaches without increasing storage and computational demands excessively; something the authors themselves suggest in their discussion.

Assembling shotgun sequencing reads into assemblies, bins and even draft genomes can enable the use of alignment based approaches for phylogenetic analysis. It can also allow phylogenetics to be carried out through methods relying on the concatenation of multiple core or univeral genes, as performed by Fitzpatrick *et. al.* [143] with a number of fungal genomes. Tools like PhyloPhlAn 3.0 [144] which rely on the use of multiple genes are dependent therefore on the sequencing of entire genomes or shotgun sequencing rather than sequencing restricted to a single gene in order to work. Alignment free based tools for phylogenetics have also been developed which use assemblies or genomes. An example of this is the tool JolyTree [145], which has been used in a number of publications [146, 147, 148, 149] across a wide variety of microbiological investigations to produce phylogenetic trees from large sequence files without the need for computationally demanding alignments or reliance on marker genes. There continue to be further advances in alignment-free methods [150] for phylogenetics using assemblies, bins and genomes. If entire genome alignment is desired then aligners like Mugsy [151] and Mauve [152] have been used in the investigation of microbial phylogenetics [153, 154] and can be applied to metagenomics investigations limited only by the availability of computational resources and time. These can be used either to produce a tree directly or to give areas or blocks of alignment which can be used for phylogenetic analysis [155]. Shotgun sequencing can also allow the sequencing of alternative marker genes, such as *rpoB* [156], for phylogenetic analysis; though amplicon sequencing targeted at the marker gene of interest may be more practical.

Shotgun sequencing tools, pipelines and annotation

A considerable number of bioinformatic tools have been developed for processing and analysing sequencing data from shotgun sequencing of microbial communities, including those mentioned previously for taxonomic classification of microbes within the samples. These include more theoretical developments such as the unified probalistic framework developed by Xia *et. al.* [157] which is called GRAMMy and was designed for use with many read assignment tools or the machine learning method proposed by Bai *et. al.* [158]. MetaFast [159] is a graph-based tool for reference-free comparison of shotgun sequencing data from microbiota samples, developed by Ulyantsev *et. al.*. The entire workflow from the initial read processing to the production of metagenomic bins can be carried out by the MetaWRAP [160] pipeline, developed by Uritskiy, DiRuggiero and Taylor. HOME-BIO [161] is a recently published and thorough pipeline for the processing and analysis of shotgun sequencing data from microbiota samples, published by Ferravante *et. al.*. Processed shotgun reads can be assembled into contigs using tools developed for this purpose like MEGAHIT [162] or adaptations of existing assembly software such as metaSPAdes [163]. Binning of assembled contigs from shotgun sequencing reads of microbiota samples can be conducted using a number of tools, including MyCC [164], CONCOCT [165], GraphBin [166] and Canopy [167]. In the same manner that ITS sequencing is a maker based approach to access

the fungal community within a microbiota, Donovan *et. al.* [168] developed FindFungi which is a pipeline for the detection of fungi within shotgun sequencing data. FunOMIC [169] is another fungal identification pipeline which was developed by Xie and Manichanh. HumanMycobiomeScan [170] is a similar tool for detection of fungi in shotgun sequencing samples developed by Soverini *et. al.* and uses a fungal genome reference database - which has been presented as leading to inherently unreliable results which the tool was assessed with databases containing other domains of life [171]. Strain level analysis of microbial communities can be accomplished using shotgun sequencing data using tools like PStrain [172], PanPhlAn [113], mixtureS [173] and Strain Finder [174].

For the assignation of functions to either reads or MAGs there have been a variety of computational tools devised. These include SmashCommunity by Arumugam *et. al.* [175], SUPERFOCUS by Gueiros Z. Silva *et. al.* [176], the MOCAT2 [177] pipeline from Kultima *et. al.* includes annotation of the sequencing reads and builds on a more limited annotation of the produced assemblies in the initial MOCAT [178] pipeline. Randle-Boggis *et. al.* [179] evaluated different annotation tools which were then available in 2016. The progress in shotgun sequencing based investigations of microbial communities has been rapid and new approaches to processing and analysing the data are published frequently, reviews of the different bioinformatic tools available [180, 181, 182, 183, 184, 185] provide insight into both the state of the field at the time they were written as well as demonstrating the longevity of some tools and the fleeting relevance of others. Viral sequences can be obtained, either intentionally or as a byproduct, during shotgun sequencing and these can be used for analysis of the viral component of the microbiota being sampled [186].

1.3 Traditional microbiological investigation of the Microbiota

The culturing of microorganisms has played a crucial role in the investigation and understanding of the microscopic world for over a century [187]. For much of the history of microbiology, culturing was one of two primary means of microbial investigation alongside visual examination [188]. A number of important discoveries have been made from the culturing of microorganisms including the accidental discovery of Penicillin by Fleming [189], the unintended isolation of *Helicobacter pylori* by B. Marshall [190] and the original work by Koch [191]. Moreover, clinical diagnosis can sometimes depend on culturing of pathogens isolated from the patient [192, 193] typically when trying to diagnose sepsis [194] or determine the best course of treatment for an identified pathogen [195]. Culturing though should not be thought of as exclusive to biomedical research. Prior to the advent of sequencing of environmental DNA, it was the principal way of investigating the microbial community which may be present in an environment. Cuadros-Orellana *et. al.* [196] isolated halophilic archaea which could grow in aromatic compounds from a number of hypersaline sites with Ozcan *et. al.* [197] carrying out similar work to obtain extremely halophilic archaea from different saline environments around Turkey. Duck *et. al.* [198] isolated flagellated bacteria from mice serving as models of colitis and Gatson *et. al.* [199] isolated a Bacillus species from a millenia old Mexican tomb. Vicente *et. al.* [200] isolated a black yeast like fungus from the tissue of infected patients and D'Elia *et. al.* [201] isolated fungi from the subglacial Lake Vostok in Antarctica. These examples demonstrate the wide range of uses culturing had to determine both if there were microorganisms present in an environment and something of the nature of any microbes found. Scientific advances have seen a new era in the application of culturing technology. 'Culturomics' as a term describes high throughput culturing of microorganisms [202] ; the approach combines automation, targeted media design and careful control of selection [203]. The limitations of environmental sequencing in isolation such as database dependency, read depth biases and potential inability to distinguish between strains [204] all helped contribute to the resurgence of culturing which culturomics represents. Pfleiderer *et. al.* [205] used culturomics to identify 11 new species from 12,700 colonies isolated from a single human faecal sample. Diop *et. al.* [206] identified a new species during their culturomics investigation of the microbiota of a commercial table salt and Angelakis *et. al.* [207] isolated 2,500 colonies from air samples collected in Saudi Arabia. There have also been other advances in media-based microbial investigation, such as 'reverse genomics' [208], the use of microcapsules [209] and development of microfluidics approaches [210]. Lewis *et. al.* [211] offer a thorough review of recent developments in culturing focussed on culturing previously uncultured microorganisms.

Initially it was thought that if a taxon was detected by sequencing but was not readily culturable then it was 'uncultivable'. An increasing number of studies combine metagenomics with culturing approaches to capture microbes which might otherwise be missed. To assist with this they might involve using metatranscriptomics to identify the necessary substrates to provide in media to cultivate specific microorganisms, as seen with work by Bomar *et. al.* [212] to obtain a *Rikenella*-like bacterium resident in the gut of the medicinal leech. They may also involve culturing approaches meant to isolate specific types of microorganisms whose presence is suggested by metagenomic investigation, such as the isolation of a nitrogen fixing and acidophilic bacterium by Tyson *et. al.* [213] ; which was directed by their analysis of assembled nucleotide fragments obtained from environment sequencing.

1.4 Host-associated microbiota

The advent of sequencing of environmental samples, whehther through 16S or shotgun-based approaches, alongside the decreasing cost of computational resources and the renaissance in culture-based methods have granted researchers access to the microbial communities which can be found living on and in macroscopic organisms. These include both plants and animals, with much research focussed on the microbiota of different human body sites; the intestinal tract in particular being much studied [214, 215]. Other regions of the human body have been somewhat less studied with regards to their microbiota; though the skin [216], vaginal [217] and - in more recent years - lung microbiota [218] have all been investigated. Links between health conditions such as obesity [219], cancer [220], HIV [221] and diabetes [222] with the microbiota are of great interest to clinicians and researchers alike.

Outside the animal kingdom, plant associated microbiota have also been studied [223, 224, 225], especially in relation to pathogen or pest resistance [226, 227]. Selim *et. al.* [228] uncovered an Archaeal strain which reduced the uptake of cobalt and so reduced its toxicity in Maize (*Zea maize*) plants when pre-inoculated into the soil. Johnston-Monje, Gutiĕrrez and Lopez-Lavalle [229] investigated the microbiota of juvenile plants of a variety of species and found the majority of the microbial communities present had been vertically transmitted via the seeds. The different fungal components of the soil microbiota associted with Oil palm (*Elaeis* members) were investigated by Kirkman *et. al.* [230] who found consierable variation across the different plantations studied and in the different compartments within the soil examined. Xu *et. al.* provide a good overview of the major findings in the study of plant-associated microbial communities in their recent review [231]. Busby *et. al.* [232] review studies into different plant microbiota from the viewpoint of enhancing reforestation efforts.

1.4.1 Animals

Much of the work conducted on animal microbiota has focused on animals of scientific, economic or cultural importance. Dirksen *et. al.* [233] released a comprehensive and manipulable core microbiome of *C. elegans* built on a previous meta-analysis of earlier microbiota investigations [234]. Horses (*Equus ferus caballus*) are of considerable economic and cultural importance and have had their associated microbiota investigated numerous times. O'Donnell *et.al.* [235] reported the presence of 35 species in a core microbiome from 6 thoroughbred racehorses, a more recent work by Gilroy *et.al.* [236] used shotgun sequencing along with assembly and binning to produce 55 high quality MAGs from young thoroughbred horses. Different microbiota studies have been carried out on dogs (*Canis lupus familiaris*) over the years, McDonald *et. al.* [237] in 2016 found the canine oral microbiome contained mostly Proteobacteria and Bacteroidetes members. More recently Craddock *et. al.* [238] used metagenomics to suggest potential links between certain phenotypic traits of interest; such as increased abundance of *Faecalibacterium prausnitzii* being negatively associated with motivation in working dogs. Li *et. al.* [239] reported the various factors influencing the composition of the faecal microbiome of Asian elephants (*Elephas maximus*), Mittal *et. al.* [240] examined the gut microbiome of three big cat species and Michel *et. al.* [241] found that three genetically differentiated populations of Grauer's gorillas (*Gorilla beringei graueri*) had a shared core gut microbiota with some variation between the populations. Links have even been found between the gut microbiota and the circadian rhythm of host animals, reported by Leone *et. al.* [242] in mice (*Mus musculus*).

The microbiota of an endangered bird, the Crested Ibis (*Nipponia nippon*) was investigated by Ran, Wan and Fang [243] who found that there was an association between the gut microbiota

and reproductive capacity; something of considerable interest to conservationists. The influence of environmental factors on the potential composition of animal-associated microbial communities was explored by Berg *et. al.* [244] in their work using *C. elegans* and different soil environments. Transmission of the gut microbiota from mother to chick along with the different impacts of the environment and host genetics for establishing the community were investigated by Ding *et. al.* [245] in Chickens (*Gallus gallus domesticus*). The role of the diet in shaping the taxonomic makeup of the caecal microbiota in goats provided with different diets was examined by Jin *et. al.* [246]. Delgado *et. al.* in their somewhat related examination of the rumen microbiota and feed efficiency in Holstein Cattle (*Bos taurus*) found the most efficient cattle had a greater relative abundance of *Bacteroidetes* and *Prevotella* in their microbiota. The potential for study of animal-associated microbiota to assist in aquaculture was shown by the work of Zheng *et. al.* [247] which used 16S sequencing to examine the impact of feed supplementation in farmed Tilapia *Oreochromis niloticus*. The stability of the core faecal microbiota of Merino sheep (*Ovis aries*) was reported by Mamun *et. al.* [248], one of a number of studies which have examined different animal-associated microbial communities over time to assess whether their consistency.

Rodents

Rodents have had their associated microbial communities studied for some years, the role of both rats and mice as very commonly used experimental animals meaning they were some of the first and still amongst the most frequently investigated species. Work by Campbell *et. al.* [249] uncovered some of the factors which can impact the composition of the mouse gut microbiota. Maurice *et. al.* [250] found that there was a strong seasonal shift, which they believed was due to dietary changes, in the wild wood mouse (*Apodemus sylvaticus*) gut microbiota. A review of the published literature around microbiota focussed investigations of rats has been written by Čoklo, Rešetar and Kraljević Pavelić [251].

In addition to the above described bioinformatic investigations there have been a variety of investigations of the microbiota of different rodents using culturing-based approaches. Many of these previous studies have focussed on either laboratory mice or have applied a range of experimental conditions. Killer *et. al.* [252] isolated three strains of gram positive bacteria from wild mice (*Mus musculus*) which were identified at the time as novel species within the -as it was configured then- *Lactobacillus* genus. Rats (*Rattus* species) and mice (*Mus musculus*) living in and around pig farms in north-east Spain were found to harbour *Clostridium difficile* in their gut microbiota by Andrés-Lasheras *et. al.* [253] ; which they may play a role in spreading. Hilton *et. al.* [254] isolated *Salmonella enterica* from faecal samples obtained in the West Midlands of the UK from wild, urban dwelling *Rattus norvegicus*. An example of a laboratory mouse derived isolate from animals not being subjected to any experimental treatment was the *Faecalibaculum rodentium* isolated from mouse faeces by Chang *et. al.* [255]. Rodents of unidentified species in Singapore provided droppings found near food waste disposal areas, these were found by Ong *et. al.* [256] to contain antimicrobial resistant *E. coli* through isolation and culturing. The Himalayan Marmot (*Marmota himalayana*) has provided a number of novel species through cultivation of different samples, including from work by Hu *et. al.* [257], Niu *et. al.* [258, 259, 260] - including isolates obtained from respiratory tract samples - and Meng *et. al.* [261]. Soh *et. al.* [262] isolated a novel species from a lab mouse strain while studying the utilisation of different carbon sources by members of the mouse microbiota. Twenty different bacterial and seven fungal species were isolated from the faeces of Balkan Snow Voles (*Dinaromys bogdanovi*) in captivity by Lukac *et. al.* [263] in the first microbiological investigation of that species. Ben-Tzvi [264] isolated *E. coli* from faecal samples obtained in the Negev Desert from multiple species of *Gerbillinae* as part of an investigation into colicin production. Neumann *et. al.* [265] isolated a

distinct lineage of *Fibrobacter* from Capybara (*Hydrochoerus hydrochaeris*) faeces using selective media.

A variety of studies have looked at links between the microbiota of rodents and their health, typically through the prism of using them as models for human medical research. This is not without some risk of misinterpretation when linking changes or trends in the microbiota in the rodents back to humans, as discussed by Walter *et. al.* [266], or from insufficiently accurate and thorough reporting of experimental variables [267]. Hu *et. al.* [268] found adolescent rats when exposed to a range of environmental chemicals saw changes in the composition of their gut microbiome and an associated decrease in bodyweight. The ability to exercise freely was linked in rats to increased relative abundance of taxa including *Lactobacillus* and *Eubacterium rectale* in the gut microbiota [269], demonstrating a relationship between the microbial community and exercise. A number of rodent studies have linked depressive disorders to the intestinal microbiota [270], including the inducing of behvaioural and physiological changes associated with depression through faecal transplant of samples from depressed human patients into rats [271]. Rajpal *et. al.* [272] observed that improvements in metabolic markers in obese mice were associated with a change in the proportion of Firmicutes in the gut microbiota following administration of the antibiotic Ceftazidime; continuing research finding links between the intestinal microbiota and metabolic disorders [273, 274, 275]. Increased inflammation both within the GIT [276] and systemically [277] has been linked to changes in the intestinal microbiota in obese rodents, demonstrating the importance of the microbiota for avoiding longer term systemic issues for the host [278, 279]. Potential links between the gut microbiota, endocrine system and bone metabolism are reviewed by Tu *et. al.* [280] - highlighting the variety of health-crucial systems which are influenced by the microbiota. Xu *et. al.* [281] found that chronic stress in adolescent rats led to a temporary change in the taxonomic composition of the gut microbiota but lasting metabolic profile changes and Jameson *et. al.* [282] found that two wild rodent species had changes in their behaviour following the administration of antibiotics which altered the composition of their intestinal microbiota. The latter study is interesting as there has been a very limited amount of research into the relationships between the microbiota, health and behaviour in wild animals. The long-lived Blind mole-rat (*Spalax leucodon*) was investigated by Sibai *et. al.* [283] who found that the most commonly detected family of bacteria in faecal samples from the animals was Muribaculaceae; which they speculate may contribute to the longevity of the host.

Different rodent species have had microbial communities associated with them examined. Linnenbink *et. al.* [284] examined wild *Mus musculus domesticus* across an area spanning sections of France and Germany, collecting samples from 121 mice to conduct 16S analysis on the caecal mucosa-associated microbiota. They found a significant difference in the abundance of *Helicobacter* in the mucosa as compared to the caecum contents, along with a significant difference for *Mucispirillum* and *Oscillibacter*. Their major finding was that geography, the location the mouse was sampled from, was the most significant factor impacting the microbial diversity between the sampled mice; as opposed to the genetic distance between the mice. This suggested that the various factors which are ultimately subject to the environment, such as diet, had the larger influence over the microbiota as compared to the individual host genetics. Weldon *et. al.* [285] found supporting results, the greater influence of sampling site over genetic differences in their investigation of wild *Mus musculus domesticus*. Weldon *et. al.* also found the most common bacterial families in their samples to be Lachnospiraceae and Ruminococcaceae, at 47% and 15% each. Rodents in different environments having different microbiota compositions can also be found in the results from Williams *et. al*'s [286] investigation of *Mus musculus* in New York residential buildings. Though they also report differences based on geographic location more interestingly was the relatively high proportion of antibiotic resistance genes they uncovered, which

they speculate has a direct link to human use of antibiotics - though also pointing out that antibiotic resistance has been found in the microbiota of wild and presumably unexposed rodents previously [287]. A trend which has been reported in a number of rodent species is that location within the intestinal tract can have a sizeable impact on the microbiota composition, observed in both wild *Mus musculus* [288] and Woodrats (*Neotoma*) [289]. Kreisinger *et. al.* [290] are amongst those who have reported significant differences in rodents between the intestinal microbiota of laboratory and wild animals of the same species, though their levels of alpha and beta diversity were the same the actual taxa varied considerably. An experiment by Leung *et. al.* [291] in which laboratory mice (*Mus musculus*) were released into a close-to-wild environment as part of an experiment investigating gut nematode susceptibility found that the microbiota changed rapidly away from the lab composition. Rosshart *et. al.* [292] used wild and laboratory *Mus musculus domesticus* in their research and found both that the two had significantly different microbiota compositions and that the wild microbiota was more beneficial for the host in terms of promoting fitness and reducing inflammation than the laboratory microbiota. They believe that in the wild the microbiota needs to confer these advantages to the host to provide baseline protection for it against infectious disease and 'naturally occurring mutagens', which would go some way to explaining similar results found by Hild *et. al.* [293].

Afrizal *et. al.* [294] published an expanded mouse intestinal bacteria collection (miBC) which contains 212 publicly available and taxonomically classified bacterial strains; resources of this kind are possible due to the considerable work which has been carried out exploring and describing the microbiota of mice. Other rodent species or animals in general subject to the same level of investigation would likely yield sufficient data to construct similar collections and tailored databases.

1.5 Environmental tolerance and the Microbiota

Despite the increasing number of metagenomic studies published, as sequencing gets cheaper and more accurate alongside decreasing computational costs, an area which has seen less focus is the potential of the microbiota to allow animals to tolerate harsh conditions [295]. Coryell *et. al.* [296] observed that arsenic toxicity was modulated by the presence of a stable and healthy gut microbiota in mice - in the process linking protection against arsenic poisoning to the microbiome in humans as well. Schmidt *et. al.* [297] assessed the changes in microbiota composition in a euryhaline fish as salinity of the surrounding water was altered - they reported that the wider water microbiota remained distinct from that of the fish at all salinities and the fish-associated microbiota changed significantly with salinity. Cheaib *et. al.* [298] found that the toxic metal Cadmium led to decreased colonisation resistance in Yellow Perch (*Perca flavescens*), with the metal having different impacts on the microbiota of different body sites. An uncommon but no less severe environmental stress shown to have an impact on the microbiota of animals is radioactive contamination. Lavrinienko *et. al.* [299] found that the gut microbiota of bank voles (*Myodes glareolus*) was influenced by soil radioactivity, though the skin microbiota was not. UV irradiation from prolonged periods soaring in the high atmosphere was shown by Graves *et. al.* [300] to influence the microbiota of vulture feathers - favouring microbes capable of resisting multiple stresses. Barelli *et. al.* [301] showed that human disruption of the natural habitat of two primate species led to alterations in the bacterial composition of their gut microbiota.

Arango *et. al.* [302] reported a connection between an environmental stressor and the microbiota of an animal. Their investigation of the Eastern subterranean termite (*Reticulitermes flavipes*) found that elevated temperatures lead to a decrease in the richness and diversity of the gut microbiota along with increased mortality and susceptibility to subsequent low temperatures. Chevalier *et. al.* [303] identified that colder temperatures resulted in both changes to the composition of the gut microbiota of their treated mice (*Mus musculus*) and to the insulin sensitivity of individuals into which cold exposed mice-derived microorganisms were introduced. This observation established a potential link between the microbiota and host tolerance of the thermal stress. Chi *et. al.* [304] built on prior research to investigate the functional impact on the microbiota in mice exposed to arsenic - finding that genes associated with pyruvate metabolism were reduced. Along with this they recorded altered alpha diversity and demonstrated the challenges for the host due to microbiota alterations; in addition to the impact of the stress directly on the host. Wang *et. al.* [305] also investigated the microbiota and arsenic stress, the authors demonstrated the potential of the microbiota to confer a survival advantage for the host when exposed to arsenic through increased excretion of the poison in stool. Outside of the mammals, Jaramillo and Castañeda [306] found that a transient heat stress altered the gut microbiota composition of *Drosophila subobscura* and that the gut microbiota itself conferred increased temperature tolerance (below 35℃) as compared to axenic individuals in their study. Sepulveda and Moeller [307] highlight a number of examples of environmental temperatures influencing animal gut microbiota in their review. Casero *et. al.* [308] demonstrated that a stress somewhat unlikely to ever be encountered in nature can still alter the gut microbiota composition with their experiments with low levels of linear energy transfer radiation. The authors identified a difference between lower and higher levels of radiation exposure potentially linked to the manner and type of response by the microorganisms. The potential influence on the gut microbiota of radiation had been investigated previously by Kim *et. al.* [309] amongst others and is a stressor which can, depending on the type of radiation, act directly on internal microbial communities within an animal host.

A comparatively more likely environmental stress an animal may encounter is heavy metal toxicity. Breton *et. al.* [310] used lead and cadmium to observe the impacts on the murine gut

microbiota of non-absorbed pollutants. The authors reported changes in composition at lower taxonomic levels (family and genus). Šrut *et. al.* [311] in a more recent investigation, looking specifically at cadmium in soil and the earthworm gut microbiota, found a similar result along with an increase in alpha diversity. Salinity as an environmental stress can be encountered in a range of environments and by numerous animals [312, 313], as such it is also encountered by microbial communities associated with these animals. Dulski *et. al.* [314] examined the impact of salinity on the gut microbiota of pike (*Esox lucius*) fry, in the context of potentially using sodium chloride in aquaculture, and found no statistically significant differences between their control groups and those reared at 3 and 7% salinity. By contrast, Castillo *et.al.* [315] found that salinity significantly impacted the β diversity of the microbiota from the water strider *Telmatometra withei*. These different results highlight how the impact of a stress on a microbiota can be dependent on the body site and selection pressure on community stability. It is perhaps not unexpected that a stress acting on a host organism will also impact the microbiota of the animal, especially in cases of external body site communities directly exposed to the stressor [316], with investigations suggesting different mechanisms for the host-environment and microbiota-host interactions to play out. Fontaine *et. al.* [317] found that the microbiota of tadpoles (*Lithobates clamitans*) conferred some protection against both elevated and decreased temperatures at the acute scale and to longer term exposure to heat stress - showing that an animal's microbiota may not have a simple unidirectional response to an environmental stressor. Kokou *et. al.* [318] observed that the host, while undergoing selection due to an environmental stressor (in their case thermal stress) appeared to be influencing the microbial community and its tolerance of the same stress through selection as well. Zhang *et. al.* [319] found that increasing temperatures led to a breakdown in the symbiont provided detoxification of an insecticide in the brown planthopper (*Nilaparvata lugens*). This highlights the potential for an environmental stress to impact the tolerance of a different stressor conferred by the microorganisms resident inside a host animal. Wang *et. al.* [320] observed a similar trend with zebrafish (*Danio rerio*) though with different temperature regimes altering the beta diversity of the microbiota, which then led to differing degrees of susceptibility to subsequent radiation exposure. The potential for the microbiota of an animal to impact its tolerance of multiple environmental threats simultaneously requires further investigation, looking at animals from an environment where a combination of stressors are in play.

A great proportion of the land surface area of the Earth can be classified as arid, more than 46 million square kilometres is covered by deserts and along with drylands the arid areas of the planet are home to almost 2.1 billion people. Quite apart from the great human diversity, arid areas are home to an array of plant and animal species; some found nowhere else. Arid areas are defined by their low annual precipitation, receiving between $< 10 - 500$ mm annually [321]. Given the absolute requirement all known forms of life have for water [322] these environments are challenging to survive in. Nonetheless, micro- to macroscopic life can be found in arid environments around the globe. Yu and Steinberger [323] demonstrate the presence of bacteria and fungi in the exposed soil of the Negev Desert, Polar Bears (*Ursus maritimus*) and Arctic Foxes (*Vulpes lagopus*) are just some of the most well known animals [324] found in the arid north and Cacti [325] can be found growing in some of the driest places on Earth. Macroscopic life has been shown to have adapted to the need to source and retain water, including in arid environments; both plants [326] and in animals [327]. Microbial communities in arid environments have been less studied until recent developments in technology allowed for higher throughput sequencing and culture independent approaches. This is despite a number of researchers investigating the microbial communities of a range of extreme environments. Rincón-Molina *et. al.* [328] found bacteria thriving in the hot and acidic crater-lake of an active volcano, Liu *et. al.* [329] uncovered a microbial population rich in hydrocarbon-degrading bacteria near the deepest point in the ocean and Bowers *et. al.* [330] demonstrated the presence of bacterial (and fungal) life 3,204 metres

above sea level. The demonstrated presence of microbial life in naturally extreme environments has also been matched with the discovery of microorganisms living in environments made harsh through the actions of man.

Baron *et. al.* [331] observed the continued presence of a bacterial community in the monochloramine treated hot-water system of a hospital, Fomina *et. al.* [332] saw significant changes in - but the continued presence of - a fungal community in soil contaminated by depleted uranium and Almatroudi *et. al.* [333] observed the survival (after prolonged culture) of *S. aureus* in dry biofilms following autoclaving at 121℃ for 30 minutes. There is a limited amount of published work which assesses the microbiota as a means of tolerating environmental stresses more broadly, rather than focusing on a specific stress. The microbiota of animals adapted to harsh conditions, particularly those facing multiple stressors simultaneously, has not been the subject of many dedicated investigations. It is difficult therefore to definitively state that the microbiota is contributing to any extent in an organism's ability to live in harsh conditions when there is more than one environmental stress acting on the organism. Arid environments present multiple challenges for the animals which live in them, cover approximately 30% of the Earth's land surface area [334] and have already been demonstrated as leading to adaptive behavioural traits in resident animals [327].

Arid-adapted animals can be found across both hot and cold deserts, providing an example of how life can adapt to harsh conditions which present a common challenge despite being geographically and meteorologically distinct. The typically remote nature of arid environments combined with the difficulties of carrying out work in the field, especially with animals, has led to a focus on animals which are of commercial or social interest [335, 336]. Some examples of these studies are work by Couch *et. al.* [337] with Bighorn sheep in the Mojave desert, Gharaechahi *et. al.* [338] with Dromedaries, Bird *et. al.* [339] with Muskoxen (*Ovibos moschatus*) and He *et. al.* [340] with Bactrian camels. In their examination of the Bighorn Sheep (*Ovis canadensis*) Couch and collaborators found proximity based correlations in the microbiota. Though this is not exclusive to arid-adapted animals [341] it is the case that in environments with few water sources and specialised flora, such as arid environments, the microbiota of animals is going to be influenced by the limited number of environmental sources. Given the restricted nutritional supplies to be found within arid environments it is essential for arid-adapted animals to be able to extract the maximum amount of resources from their diets. He *et. al.* in their study of Bactrian camel (*Camelus bactrianus*) microbiota along the digestive system found a major fraction was composed of unknown Ruminococcaceae. They suggest that this might assist the camels in their consumption of the salty and hard to digest plants found in their arid habitat. Gharechahi *et. al.* [342] found that the Dromedary Camel (*Camelus dromedarius*) had a rumen microbiota which was functionally distinct from those of many other ruminants, sharing some traits with the Moose (*Alces alces*). **Table 1.1** summarises some of the details of the limited number of studies undertaken into the microbiota of arid adapted animals.

There have been some individual investigations into other species, like that of the Mongolian gerbil (*Meriones unguiculatus*) by Nouri *et. al.* [343]. Animals which can inhabit arid and non-arid environments have been studied in comparative terms, between populations living in the two or more environments, with some of these studies either focusing on or including microbiota investigation [344, 345]. Eisenhofer *et. al.* [346] in their comparison of the intestinal microbiota of captive and wild southern hairy-nosed wombats (*Lasiorhinus latifrons*) found that the semi-arid conditions in which the wild individuals lived contributed to a gut microbiota with greater potential to access and use recalcitrant plant energy sources as compared to the captive populations. This suggested that an arid (or semi-arid in this instance) adapted microbiota might allow access to a wider range of otherwise unusable energy sources. Graves *et. al.* [300] looked at the

plumage of New World vultures (Cathartiformes) which spend considerable time exposed to UV and ionising radiation as well as in an extremely dry environment while soaring and found that extremophiles were well represented in their feather samples; demonstrating the diversity and severity of stresses which any external microbiota would need to tolerate. Ribeiro *et. al.* [347] examined the faecal microbiota of the Karoo scrub-robin (*Cercotrichas coryphaeus*) as part of a wider investigation of the impact of aridity on physiology. The authors detected enrichment of bacteria from *Sphingomonas* and *Leucobacter* amongst other genera in birds from more arid, inland locations. There have been metagenomic investigations of sympatric animals, including those which are relatively closely related. Li *et. al.* [348] used 16S sequencing with two bird species overwintering and found that sympatric individuals within each species had distinct faecal microbiota communities associated with dietary differences; they also identified a host effect on microbiota composition. Perofsky *et. al.* [349] examined six sympatric mammal species in Madagascar, three lemurs and three non-primates, within an area of 1 square kilometre finding that microbiome taxonomic composition was largely influenced by diet, evolutionary relationships and shared terrestriality. Shanmuganandam *et. al.* [350] compared the faecal microbiota of sympatric European brown hares (*Lepus europaeus*) and European rabbits (*Oryctolagus cuniculus*) in Australia. The authors found similarity between the two species at the higher taxonomic levels but differences in the most abundant bacterial genera between them. Menke *et. al.* [351] looked at more distantly related African animals living sympatrically in Namibia, cheetahs (*Acinonyx jubatus*) and black-backed jackals (*Canis mesomelas*). The authors found similar results, in terms of patterns and trends, as Shanmuganandam *et. al.* ; similar taxa at higher levels like phyla but increasing diversity with lower taxonomic levels.

Sequencing method	No. hosts sampled	Host organism	Prokaryotic taxa of interest	Suggested prokaryotic role	Publication
16S rRNA	3	Dromedary Camel	*Prevotella ruminicola, Ruminococcus flavefaciens, Fibrobacter succinogenes*	Produce glycoside hydrolase enzymes, Synergise with fibrolytic bacteria to improve fibre digestion, High efficiency in degrading crystalline and amorphous cellulose, Prolific cellulose degrader	Gharechahi J. *et. al.* 2015
16S rRNA	18	Bactrian Camel	*Blautia* species, Christensenellaceae members	May provide anti-inflammatory effects in young camels, May help regulate intestinal environment, Linked to immunomodulation and healthy homeostasis	He J. *et. al.* 2019
16S rRNA	11	Bactrian Camel	Unclassified Ruminococcaceae, Unclassified Clostridiales, *Akkermansia* species	Feed fermentation to cope with low quality forage, Help prevent diabetes even with high blood glucose levels, Help prevent hypertension even with high salt diet	He J. *et. al.* 2018
WGS	3	Dromedary Camel	*Fibrobacter succinogenes*, *Ruminococcus* species, Fibrobacteres members, Spirochaetes members, Bacteroidetes members	Potential lignocelluose degrader, Degrading crystalline cellulose, Help with lignocellulose rich diet, Use PUL enzymes to assimilate complex carbohydrates	Gharaechahi J. *et. al.* 2018

WGS	6	Dromedary camel	*Bacteroides thetaiotaomicron*	Production of starch -degrading enzymes	Bhatt V.D. *et. al.* 2013
16S rRNA	3	Muskox	Ruminococcaceae members	Help digest highly lignified winter forage diet	Salgado-Flores A. *et. al.* 2016

Table 1.1: Table summarising the details and key findings of a number of microbiome studies into arid-adapted animals. Modified from Osborne *et. al.* [295]

Those studies which have been conducted into the microbiota of arid-adapted animals have produced some exciting results, including for selection against motility proteins by microorganisms in arid environments [352] and selection for increased pigmentation by exposed microorganisms in extremely dry sites [353] along with those described above. As discussed earlier, prior research into the microbiota of arid-adapted animals has been largely conducted on ungulates. It would be beneficial for researchers to consider examining the microbiota of different clades of arid-adapted animals. A number of arid-adapted rodent species exist [354], some extremely well adapted to arid environments [355] and some which can be found in both arid and non-arid environments [356]. Rodents adapted to life in arid conditions have been the subject of considerable scientific investigation around behavioural [357], metabolic [358] and renal adaptations [359] to their environments. The potential of the rodent microbiota in assisting life in arid conditions has not yet been investigated.

Chapter 2

Methods

- The underlying experimental setup is described
- The sources of external data used in the project are provided
- Information on the experimental animals used is provided
- The different bioinformatic and microbiological methods are provided
- The versions of bioinformatic tools used are provided

2.1 Sample and data collection

2.1.1 Ethics

In the course of this project no animals were sacrificed nor reared in laboratory conditions. The animals used in this experiment were wild and aside from the trapping, marking with ear punches and collection of faecal samples were not subject to any other procedure or interacted with. All experiments were carried out in accordance with the Israeli Ministry of Health guide for the care and use of laboratory animals and all experimental protocols were approved by the Tel-Aviv University Institutional Animal Care and Use Committee (IACUC protocol approval number 04-18-056, 04-19-031; permit number L15055). Faecal samples shipped to the UK were not subject to the Nagoya Protocol as Israel is not a signatory. The Animals (Scientific Procedures) Act 1986 (ASPA) was not applicable to this project as no animals were sacrificed or reared in laboratory conditions; nor did any of the *Acomys* undergo any procedure which caused pain, suffering, distress or lasting harm. Samples were collected as part of a broader project being undertaken by collaborators. Samples were imported under a licence to import animal/poultry products under the provisions of the Importation of Animal Products and Poultry Products Order 1980 (as amended), permit authorisation number: ITIMP21.0586.

2.1.2 *Acomys cahirinus* and *Acomys russatus*, experimental organisms and background

Acomys cahirinus ('Cairo Spiny Mouse') and *Acomys russatus* ('Golden Spiny Mouse') are two rodent species from the *Acomys* genus. They are both adapted to aridity albeit to different extents, with *A. russatus* being able to tolerate harsher conditions than *A. cahirinus*. Both can be found living sympatrically where their ranges overlap, including around the site of the enclosure in which the individuals used in this study lived. The enclosure was property of the University of Tel Aviv, Israel, and was located near the Ein Gedi Reserve in the Judean Desert; approximate coordinates were within 5km of `31°28'00.0"N, 35°23'00.0"E`. A photograph of an example *Acomys cahirinus* individual can be seen in **2.1A** and an example *Acomys russatus* individual in **2.1B** - note images are not to scale; *A. russatus* individuals can achieve greater mass than *A. cahirinus* though both have similar mass and size ranges.

A. russatus has a habitat range covering the arid regions of the Middle East, including Egypt, Israel, Jordan, Saudi Arabia and Yemen [360]. It inhabits rocky areas on slopes and dried river beds where it can use boulders for cover. Shkolnik and Borut [361] found that the *A. russatus* basal metabolic rate is 34.5% lower than predicted for their body weight, they could survive ambient temperatures up to 42.5℃, could concentrate the urea in their urine up to 4700 - 4800 mM and could drink sea water. In Egypt and Israel it is not recorded as reproducing during winter, gestation in *A. russatus* is 40 days and litters typically contain 1 to 4 young depending on the age of the mother [362]. They are born well developed with open eyes and a fur coat, maturing fully within 3 months [363]. *A. russatus* are omnivorous and obtain the water they require to survive from their diet; plants, invertebrates and with a particular preference for snails [364]. Shafrir and Adler [365] fed *A. russatus* sucrose and fat diets, observing that the experimental animals gained weight rapidly and significantly on both the regular and fat-rich seed diet, they also had increased fat deposition from the fat-rich diet. By preference *A. russatus* is nocturnal, and is so wherever in its range it is not sympatric with *A. cahirinus*, exposure to *A. cahirinus* causes a shift in circadian rhythms in *A. russatus* to a diurnal lifestyle [366]. It is listed as Least Concern by the IUCN Red List [367].

The other Acomys species used in this work is *A. cahirinus*. This species has a much wider habitat range, from Ethiopia north through the Eastern Mediterranean, Middle East and into Iran, the Arabian Peninsula and with populations across Mauritania and Mali [360]. It inhabits rocky areas across its range, typically canyons or cliffs; it has been observed amongst smaller rocks than those preferred by *A. russatus* [368]. *A. cahirinus* are observed to spend less time in deeper crevices and boulder cover than *A. russatus* [369]. *A. cahirinus* has a similar diet to *A. russatus*, being omnivorous and consuming seeds, green plant matter, arthropods and snails (when available) [370]. Testing by Shkolnik and Borut found that *A. cahirinus* could not survive ambient temperatures of 40℃, that the mean minimal metabolism of *A. cahirinus* was 13.5% lower than predicted by body weight and it could concentrate chloride in its urine to around 700-800 mM. Scantlebury *et. al.* [371] found that *A. cahirinus* individuals from a xeric (arid) environment assimilated energy more efficiently than those from a nearby but mesic (non-arid) environment. *A. cahirinus* produces 2-3 precocial young in a litter [372], after a gestation of 39 days and the species was found to undergo menstruation [373]. They are social animals and are found in groups in the wild in which the strength of social relationships is determined by familiarity [374] and recognition of siblings [375] or offspring [376]. *A. cahirinus* is exclusively nocturnal across its entire range, when sympatric with *A. russatus* the latter switches to a diurnal lifestyle. This switch is not caused by aggressive behaviour on the part of *A. cahirinus* [377]. Baudoin *et. al.* [378] demonstrated that *A. cahirinus* urine could induce reduced foraging behaviour in *A. russatus*. *A. cahirinus* is listed as Least Concern by the IUCN Red List [379].

Where the two species are sympatric *A. russatus* adopts a diurnal cycle from its usual nocturnal cycle when not sympatric, *A. cahirinus* remains nocturnal. Diets between the two species overlap more in winter as they switch to primarily consuming plants. In summer the arthropods they predate have their own temporal patterns leading to different availability to the two *Acomys* species [370].

These two species were used here as they provided a existing and natural setup whereby two closely related and wild species could be investigated for possible links between aridity and the intestinal microbiota. They are both arid-adapted though to differing degrees, they are closely related and can be found in sympatry in some areas - including the experimental location. Their diets are largely similar - to the extent that they have developed mechanisms to avoid competition through *A. russatus* switching to a diurnal lifestyle as described above. That they are wild animals is also important for the purposes of advancing metagenomics, it is necessary to expand existing collections of reference material to include an increasing diversity of wild animals [380] as well as to gain greater understanding of how different they may be from more commonly sequenced laboratory, captive or domesticated animals. Their use in this project furthers the growing research focus on wild animals as subjects of metagenomic investigation [381, 382, 383, 384].

Halotolerance

As part of an investigation into thermogenesis in *Acomys cahirinus*, Scantlebury *et. al.* [385] provided water with salinity up to 2.5%, which the experimental subjects could tolerate. Wube *et. al.* [386] investigated the impact of dietary salinity on both reproductive status and energy intake of both *A. russatus* and *A. cahirinus*; a salinity of 3.5% was used for *A. cahirinus* and 5% for *A. russatus* with both able to survive the treatment. Shanas *et. al.* [387] also used a salinity of up to 3.5% for their work studying the impact of dietary salt and season on the daily rhythm of body temperature. Shanas and Haim [388] also investigated the impact of dietary salinity and vasopressin on reproduction in *A. russatus* and also used salinity up to 5% in the course of their work. Ron and Haim [389] found *A. russatus* responded to increasing salinity (ranging from 1 to

(a) *Acomys cahirinus*. Credit Vera Kuttelvaserova (213793495), reproduced under Extended Adobe Stock License.

(b) *Acomys russatus*. Credit Eric Isselee (213793495), reproduced under Extended Adobe Stock License.

Figure 2.1: Photographs of individuals of both *Acomys* species used in this project.

7%) by decreasing their urine volume and increasing the osmolarity of urine. Shkolnik and Borut [361] carried out some of the earliest work with *A. russatus* and *A. cahirinus* and identified both the greater salt tolerance of *A. russatus* compared to its congener and the ability of *A. russatus* to drink sea water. Harriman [390] identified that *A. cahirinus* did not prefer saline solutions when fresh water was available, highlighting that though both *Acomys* species can tolerate saline conditions they do not have a preference for them. Seawater has an average salinity of 3.5%, slightly saline water a salinity of 0.05 to 0.15% and highly saline water a salinity of 0.7 to 1.5% [391]. Drinking water typically has a salinity below 0.05%.

2.1.3 *Acomys* sample collection and experimental setup

A pictorial representation of the sample collection and experimental setup can be seen in **Figure 2.2**. As part of an ongoing project an enclosure was erected in the Judean Desert at Ein Gedi (approximate coordinates $31°28'N, 35°23'E$). The enclosure contained wild animals including both species of *Acomys*; these individuals being already present in the location living wild and were not experimental or lab-reared or bred animals. The enclosure prevented the *Acomys* from leaving the site but in no other way impacted on their health or behaviour, the enclosure was exposed to the air and no other interventions were made.

In June of 2016 baited traps were set in the enclosure and a number of *Acomys* individuals of each species caught; they had their ears marked with punched tags as the considerable regenerative abilities of the two species, discussed in **Subsection 2.1.2**, meant this did not cause any harm or ill effect for the animals. The animals were released after tagging, if in the course of the trapping the animals defecated then the faecal pellets were collected. These were stored in test tubes without any protective media or compound - the sample collection was a supplemental component of a larger and ongoing project and so the samples were not being explicitly collected with the intetion of metagenomic analysis. As a consequence of this the species of the animal was recorded but not the age or sex, though there is no established method for determining the age of wild *Acomys cahirinus* or *Acomys russatus*. The tubes containing the samples were labelled with the tag details of the animal they originated from. After fieldwork had been completed the stored faecal samples were transported to the University of Tel Aviv where they were stored at -20°C, this was the coldest freezer available for storage of faecal samples in the facility. In November of 2016 the same trapping and collection procedure was carried out, as the traps were set in the enclosure containing wild animals only some of the animals caught in June were recaptured in November. Once again if the animal yielded a faecal sample then the pellets were collected, stored at ambient temperature in test tubes and then transported at the end of field work to the University of Tel Aviv where they were then stored at -20°C. All fieldwork took

place over a single day on each occasion with samples collected by the same individual, **Ella Pasternak**, pellets were not handled by the collector prior to being placed in the test tube. All samples were then shipped by air to the UK in November 2016, transported frozen at -80℃.

As the animals were wild and were being collected from traps as part of a larger ongoing research project being conducted by collaborators there was no sample size set initially, as it would be conditional on animals being caught in the traps anyway. The author had planned to conduct further sampling in which case it would have been beneficial to determine what a desired minimum sample size would have been, however this was prevented by the global COVID-19 pandemic. This was also why there was no variation in the time of day in which faecal samples were collected, traps were examined and any captured animals handled as and when by the individual in the field conducting the larger investigation of collaborators. Had further sampling been possible it would have been potentially interesting to try and collect samples at different times of day, though as these are wild animals living in an enclosure there would be no guarantee that samples could be obtained at varying times - especially given the different activity patterns followed by the two *Acomys* species when sympatric. An approach such as that outlined by Ferdous *et. al.* [392] would be useful in calculating the ideal sample size to try and obtain if further sampling was possible.

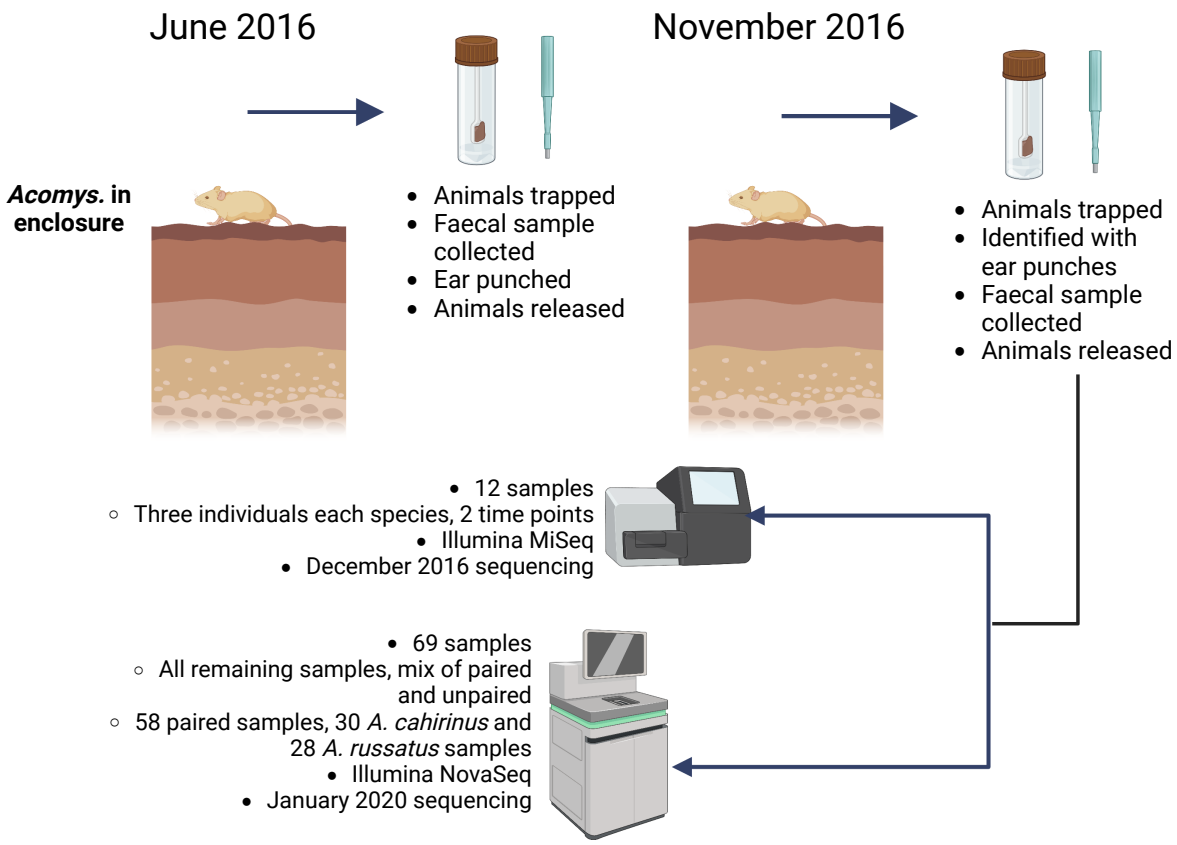

Figure 2.2: Figure outlining the experimental setup for the collection of *Acomys* faecal samples. Created with BioRender

2.1.4 Extraction and sequencing

There were two different times when extraction and sequencing from the samples was undertaken. Initially in December of 2016 12 samples were selected to use as a 'pilot' project to determine the feasibility of investigating the faecal microbiota of *Acomys* DNA was extracted by **Iliana Serghiou** using a MP Biomedicals FastDNA™ SPIN Kit for Soil (MP Biomedicals) following the standard protocol with libraries being prepared at the Sanger Centre. These samples were sequenced for 150bp PE on an Illumina MiSeq. As shown in **Figure 2.2** these samples were a June and November sample from the same individual, for three individuals of each of the two species for a total of 12 samples undergoing extraction and sequencing.

The second extraction and sequencing occured in January 2020, this was also carried out by **Iliana Serghiou**. Extraction was carried out using a MP Biomedicals FastDNA™ SPIN Kit for Soil (MP Biomedicals). The protocol differed slightly from the standard version. Firstly 980 µL of Sodium Phosphate Buffer was used in the sample preparation step instead of 978 µL, also during this stage 120 µL of MT Buffer was used instead of 122 µL. During the following homogenisation step the FastPrep instrument was used for 180 seconds instead of 40 seconds; centrifugation was carried out after this for 15 minutes instead of 5-10 minutes. In the binding step, 700 µL of the mixture of the supernatant and resuspended binding matrix was filtered and centrifuged instead of a maximum of 600 µL. During the drying stage, the dry spin on the centrifuge was carried out for 1 minute instead of 2 minutes. Following this the SPIN filter was air dried at room temperature for 10 minutes and then for 5 minutes at 37℃ instead of just air dried at room temperature for 5 minutes. In the elution step, 65 µL of DES was added to the Binding Matrix instead of a range from 50-100 µL. The protocol otherwise does not differ from that available at: `https://www.mpbio.com/media/document/file/manual/dest/f/a/s/t/d/FastDNA_SPIN_Kit_for_Soil_UM_2021_WEB.pdf` with the DNA obtained being stored at -20℃ prior to being shipped frozen on dry ice to Novogene (Beijing, China) for sequencing. Shotgun DNA libraries were prepared using a NEB kit and then were sequenced on the NovaSeq 6000 platform (Illumina) for paired end sequenced reads of 150bp.

Shotgun sequencing was chosen as the approach for this project to enable as great a capture as possible of the entire genetic material associated with the *Acomys* microbiota - as discussed previously the advantages of 16S sequencing would not be applicable to this project due to a considerable lack of similar prior studies and so shotgun sequencing was chosen as the most appropriate approach. It also enabled the assembly and binning of the reads, which would not be possible with 16S sequencing and the direct annotation of the metagenomic bins so predicted functions came from the output of the annotation tool rather than being inferred from an amplicon-based taxonomy.

2.1.5 Statistics

The various statistical tests employed are described in the relevant methods section where they are used. All statistical analysis was carried out using R [393] (versions listed in context) through RStudio [394] (v. 2022.12.0+353) unless otherwise stated.

2.1.6 External data for tools testing

The sequencing reads generated from a mock human microbial community were kindly provided by **Dr. Richard Leggett** (Earlham Institute, UK); specifically HM277-D (BEI Resources, Man-

assas, VA, USA) and used in the tools testing stage as one of a number of baseline metagenomic communities aside from those obtained during the project from the *Acomys* faecal samples. Two sets of shotgun read files were downloaded from the ENA for study PRJEB32890 [395] at `https://www.ebi.ac.uk/ena/browser/view/PRJEB32890`, specifically the read files were for sequencing of biosamples SAMEA5690310 and SAMEA5690311, both wild mice (*Mus musculus*) examined during the study. Two sets of shotgun read files were also downloaded from the ENA for study PRJEB7759 [396] at `https://www.ebi.ac.uk/ena/browser/view/PRJEB7759`, specifically the read files for sequencing of biosamples SAMEA3134373 and SAMEA3134374, both lab mice (*Mus musculus*) examined in the course of the study.

These data sources were used in the tools testing phase of the project as the author believed they would provide a more reliable measure of the precision and accuracy of the taxonomic classification tools. This was as they were from species which had been subject to considerable metagenomic investigation previously and all the tested tools had been used in published literature on metagenomic samples from these host species. This was in contrast to both *Acomys* species which had never been subject to metagenomic investigation prior to this project. A mix of wild and laboratory mouse samples was chosen to provide a combination of both well characterised microbial communities (lab mouse samples) and ones relatively more similar in origin to the *Acomys* species (wild mouse samples).

The human mock community contained 20 different bacterial species from 17 genera. There are two species from the *Staphylococcus*, *Staphylococcus aureus* and *Staphylococcus epidermis*. There are three species from the *Streptococcus*, *Streptococcus agalactiae*, *Streptococcus mutans* and *Streptococcus pneumoniae*. Some names differ from those in the reference databases used by the different taxonomic classification tools, the community was chosen at the time as it contained multiple species associated with animal-associated bacteria (albeit human) and by including species from within the same genera could potentially be used to test species level resolution of the different taxonomic classifiers. Mock metagenomic datasets have been used to assess and validate a variety of bioinformatic tools [397], techniques [398, 399] and approaches [400, 401, 402] before and the author believed the approach had merit here, given the inclusion of some rodent data from real microbiota samples, as it removes any vagaries around unknown or unclassifiable material. As it was the result of real sequencing of reads from the defined, mock community, rather than a purely simulated sample it also served as a test for the tools ability to function with a sample which could contain the same biases and errors as a real sample.

2.1.7 Bacterial reference genomes and assemblies for phylogenetic comparisons

In addition to the external data described in subsection 2.1.6 use was made in this project of additional sources of external data. These included a number of bacterial genomes and Metagenome assembled genomes (In context, Metagenome Assembled Genomes (MAGs)). These external sources of data were used when creating phylogenetic trees for both binned read files discussed in subsection 2.4.2 and for bacterial isolate genome assemblies as discussed in sub-subsection 2.7.2. These external data sources in broad terms included the following:

- 25 high quality MAGs chosen at random from those created by the Integrated Mouse Gut Metagenome Catalogue (Integrated Mouse Gut Metagenome Catalogue (iMGMC))

- 9 Lactic Acid Bacteria (LAB) genomes from the National Center for Biotechnology Information (NCBI) RefSeq database chosen at random, this type of bacteria was chosen based on the results from three of the taxonomic classification tools on the sequencing reads from the *Acomys* faecal samples
- 14 LAB genomes from the NCBI RefSeq database. LABs chosen based on the results of the taxonomic classification of metagenomic bins produced during this project
- 80 bacterial genomes or assemblies from NCBI databases chosen specifically based on the results of one of the taxonomic classification tools run on the sequencing reads from the *Acomys* faecal samples
- 45 bacterial genomes or assemblies from NCBI databases chosen specifically based on the results obtained in a publication which investigated five desert rodent species' gut microbial ecology [403]

The three classification tools which informed the selection of 9 random LAB genomes were Kraken 2, Kaiju and Metaphlan 3 and the methods employed with them and their relevant citations are given in their sub-subsections within section 2.3. The single classification tool which informed the 80 specific genomes or assemblies downloaded was mOTUs, the methods employed with it and the relevant citation are given in the appropriate sub-subsection within 2.3. The 14 LABs were chosen based on the results from GTDB-Tk analysis of the metagenomic bins produced during this project, the methods employed with it and the relevant citation are given in the appropriate sub-subsection within subsection 2.4.2. The names of all the downloaded genomes, assemblies and MAGs can be found in **Table 5.2** in the **Supplemental Material** along with the rationale for their use. Genomes and assemblies from the NCBI were downloaded from the relevant directory under `https://ftp.ncbi.nlm.nih.gov/genomes/all`, iMGMC MAGs were extracted from the large file downloaded from `https://zenodo.org/record/3631711/files/iMGMC-hqMAGs-dereplicated_genomes.tar.gz`.

2.2 Processing read files

2.2.1 Quality control and removal of host or human contamination

All metagenomic read files were processed using the same procedure to remove any reads which were of low quality, resulted from human contamination or were from the original host organism. FastP [404] (v. 0.19.7) was used with default settings to trim adaptors and for general quality control of the read files. Reads which passed In context, quality control (QC) were then subject to BBDuk [405] (v. 38.06) analysis to remove any reads which were from the host organism or human contamination. For the two *Acomys* species reference genomes which had been generated by **Dr Shane McCarthy, Dr David Thybert** and **Dr Kerstin Howe** at the Wellcome Sanger Institute (available on request) were used. For the *Mus musculus* reference genome, the reference from NCBI found at `https://ftp.ncbi.nlm.nih.gov/genomes/refseq/vertebrate_mammalian/Mus_musculus/reference/GCF_000001635.27_GRCm39/GCF_000001635.27_GRCm39_genomic.fna.gz` and downloaded on 2022-02-07 was used. For the human reference genome the *Homo sapiens* reference genome file found at `https://ftp.ncbi.nlm.nih.gov/refseq/H_sapiens/annotation/GRCh38_latest/refseq_identifiers/GRCh38_latest_genomic.fna.gz` and downloaded on 2021-12-31 from NCBI was used. Default settings were used with BBDuk.

2.2.2 Processing for tool testing

Reads originating from the human mock microbial community which passed the FastP QC step did not undergo any further processing prior to use. The 12 *Acomys* sample read files and the 4 downloaded *Mus* read files were subsampled to a uniform depth using Seqtk (v. 1.0-r77-dirty) [406]. A depth of 4,450,000 reads was chosen as the smallest read count after QC from all rodent files was 4,464,759 from an *Acomys* sample. Reads were subsampled to a uniform depth in order to allow for comparisons across the three rodent species. The same randomly generated seed was used with Seqtk for all samples when subsampling.

2.2.3 Processing for analysis

All *Acomys* read files were processed following the same method as the 12 used for the tools testing stage. Quality control was carried out with FastP with default parameters, contaminant removal was carried out using BBDuk with the two *Acomys* reference genomes and the same human reference genome listed in the prior subsection. Subsampling was also carried out using Seqtk however in this instance two uniform depths were chosen as there was a large difference between the sequencing depths of the reads obtained on the two occasions. The read files from those 12 samples which were sequenced in 2016 were subsampled to a depth of 4,460,000 reads. The read files from those 69 samples sequenced in 2020 were subsampled to a depth of 7,600,000 reads. Two different uniform depths were chosen as there was a minimum 3 million read difference between the smallest post-processing read file from the 2020 sequencing and the largest post-processing read file from the 2016 sequencing. The particular details for read count changes in each processing step can be found in **Table 5.1** in the **Supplemental Material**.

2.3 Taxonomic classification of metagenomic reads

Four different taxonomic classification tools were employed. These were Metaphlan 2 [136] (v. 3.0.10), Kraken 2 [407] (v. 2.0.8), Kaiju [129] (v. 1.7.3) and MOTUs [139, 408] (v. 3.0.3.1). In brief, Metaphlan and mOTUs rely on alignment of the reads to databases of marker genes, with the number of marker genes and the specifics of the matching described in the respective publications. Kraken 2 matches the reads against a reference database by searching for and then extending matching sections of nucleotides called 'kmers', the specific details are described in the publication. Kaiju translates the read files into amino acid sequences and compares these against a protein reference database; the specifics are described in the publication.

These four tools were chosen as they are all intended for use analysing shotgun sequencing reads, which was the form of sequencing used in the project and to provide a range of classification methodologies. The author believed that using multiple tools, and using tools which employed different underlying methods, would be important as otherwise it would be difficult to know whether the results reflected the real state of the *Acomys* microbiota or a failing of a single tool. This is especially important when dealing with samples from a species which has never undergone metagenomic investigation and from a subset of animals, arid-adapted rodents, which have undergone extremely limited metagenomic investigation. All of the tools have been used in metagenomic investigations previously, have been used with rodent microbiota studies specifically [409, 410, 411, 412] and most have been included in publications which compared the performance - both in terms of computational resources and accuracy - of different taxonomic classification tools such as that by Breitwieser *et. al.* [132].

Alpha and Beta diversity metrics were not used in this project as the large proportion of reads could not be classified by three of the four classifiers (discussed in **Chapter 3**), meaning that any taxonomic classification based diversity metrics would only be applicable to a small fraction of the entire data set. While this may have had some merit, being functionally equivalent to analysis of shallow shotgun sequencing like that discussed by Hillmann *et. al.* [37], the author believed that it would be more useful to discuss the small proportion of the reads which could be classified in terms of relative abundance and composition instead. In addition the mapping based approach covered a greater proportion of the reads and did not require diversity-metric based approaches.

2.3.1 Taxonomic classification for tool testing

For each of the 4 taxonomic classification tools used in the tools testing stage of the project the following methods were employed with each tool.

Metaphlan 3

Metaphlan 3 was run using the reference database version `mpa_v30_CHOCOPhlAn_201901` which was downloaded on February 01, 2022 from `https://drive.google.com/drive/folders/1_HaY16mT7mZ_Z8JtesH8zCfG9ikWcLXG`. Default commands were used with Metaphlan 3 with the exception of using the `t rel_ab_w_read_stats` flag to obtain both the relative abundance and the estimated read counts for each clade. Metaphlan was run on the processed read files generated from the 12 *Acomys* samples sequenced in 2016 along with the 4 downloaded and processed *Mus* read files and the processed sequence reads obtained from sequencing of the human mock community.

Kraken 2

Kraken 2 was run using the reference database version `k2_pluspf_20210517` which was downloaded on February 01, 2022 from `https://genome-idx.s3.amazonaws.com/kraken/k2_pluspf_20210517.tar.gz`. When running Kraken 2 default flags were used save for changing the minimum confidence score required using the `confidence` flag. As part of the testing process this was varied from the values of 0.1, 0.25, 0.5, 0.66, 0.75 and 0.95. Kraken 2 was run on the processed read files generated from the 12 *Acomys* samples sequenced in 2016 along with the 4 downloaded and processed *Mus* read files and the processed sequence reads obtained from sequencing of the human mock community.

Kaiju

Kaiju was run using the reference database version `kaiju_db_nr_euk_20210224` which was downloaded on February 01, 2022 from `https://kaiju.binf.ku.dk/database/kaiju_db_nr_euk_2021-02-24.tgz`. When running Kaiju default flags were used save for changing the number of errors allowed using the `e` flag. As part of the testing process this was varied from 0, 1, 2, 3, 4, 5, 10 and 20. Kaiju was run on the processed read files generated from the 12 *Acomys* samples sequenced in 2016 along with the 4 downloaded and processed *Mus* read files and the processed sequence reads obtained from sequencing of the human mock community.

mOTUs

mOTUs was run using the reference database version `db_mOTU_v3.0.1` which was downloaded on March 19, 2023 from `https://zenodo.org/record/5140350/files/db_mOTU_v3.0.1.tar.gz?download=1`. When running mOTUs default flags were used save for running the program twice on each file, once without the addition of any flags and once including the flags `-u -p -q` which respectively instructed the tool to print the full name of any detected species, print the NCBI identifiers for any detections and print the full rank taxonomy for any detections. mOTUs was run on the processed read files generated from the 12 *Acomys* samples sequenced in 2016 along with the 4 downloaded and processed *Mus* read files and the processed sequence reads obtained from sequencing of the human mock community.

2.3.2 Taxonomic classification for analysis

For each of the 4 taxonomic classification tools used in the main sample analysis stage of the project the following methods were employed with each tool.

Metaphlan 3

Metaphlan 3 was run with default settings save for the use of the `t rel_ab_w_read_stats` flag to obtain both the relative abundance and the estimated read counts for each clade. The same reference database was used as described in the previous subsection. Metaphlan 3 was run on all 81 *Acomys* processed read files from both rounds of sequencing.

Kraken 2

Informed by the results of the tools testing stage, Kraken 2 was run using default settings save for setting the `confidence` flag to 0.5 to require a 50% minimum confidence score for classifications. The same reference database was used as described in the previous subsection. Kraken 2 was run on all 81 *Acomys* processed read files from both rounds of sequencing.

Kaiju

Informed by the results of the tools testing stage, Kaiju was run using default settings save for setting the `e` flag to 0 to require no errors in the classification. The same reference database was used as described in the previous subsection. Kaiju was run on all 81 *Acomys* processed read files from both rounds of sequencing.

mOTUs

mOTUs was run using the default settings and run twice on each file as described in the previous subsection to produce two output files providing different levels of information. The same reference database was used as described in the previous subsection. mOTUs was run on all 81 *Acomys* processed read files from both rounds of sequencing, in light of the results from the testing stage concerning the influence of pre-processing, mOTUs was run on input read files which had been subject to quality control and host/contaminant removal but had **not** been subsampled to a uniform depth. The percentages of classified reads from this stage is reported as a percentage of the total number of reads classified in each file processed up to but not including the subsampling to a uniform depth stage.

2.4 Assembly and binning of metagenomic reads

The assembly and binning of reads was carried out by **Dr Sebastien Raguideau** and **Professor Chris Quince** using their Metahood [413] pipeline. To briefly summarise the approach, all reads were processed with FastQC [414] and TrimGalore [415]. Processed reads were then co-assembled with Megahit [162]; default parameters were used with these tools. The contigs generated were then binned with Concoct [165] and Metabat2 [416], both run with the default settings and contigs were assigned into bins based on consensus of their placement by both tools. Bins which met a minimum completeness score of 70% from CheckM [417] analysis were considered 'quality' bins. CheckM calculated completeness by making a taxonomic classification of the query to identify genes which would be expected to be present based on the phylogeny, it then searched the query for these and measured completeness as the fraction of expected genes which were detected.

2.4.1 Processing of bin files

The bins generated by the Metahood pipeline from the 81 *Acomys* samples which met or exceeded the 70% quality threshold were processed using the following methods. CheckM (v. 1.1.2) was run again on these bin files, the `checkm lineage_wf` command was run first and then the generated lineage file was used with the `checkm qa` command while using the `-o 2` flag to generate an extended summary file of the measured statistics. New cutoff values for completeness and contamination of $\geq 80\%$ and $\leq 5\%$ respectively were used to filter the bin files for further analysis. The bins which passed the filtering step were then analysed with FastANI [418] (v. 1.3) to detect any bins which might be identical to each other generated during the Metahood analysis. Any bins with a 99.9% or greater ANI similarity score to each other were considered to be potentially identical and one was randomly selected to use for further analysis, bins which did not meet this threshold were also retained for further analysis. dRep [419] (v. 2.5.0) was used in addition to determine if any bins were extremely similar to each other and pick a best representative of any determined to be very similar. Default settings were used save for the flag `-pa` being used with a value of 0.95.

2.4.2 Identities and phylogeny of metagenomic bins

Taxonomic classification of filtered bins was conducted using GTDB-Tk [420] (v. 1.7.0) using the database `release R202`. The classify workflow was run first using the command `gtdbtk classify_wf` with the flags `-min_perc_aa 10` and `-min_af 0.65`. The *de novo* workflow was then run using the `-min_perc_aa 10` flag and setting the outgroup taxon as the Proteobacteria phylum using the flag `-outgroup_taxon p__Proteobacteria` and using the `-bacteria` flag to indicate the bins files should be treated as if they were bacterial genome files. Proteobacteria was chosen as the outgroup based on the results of the taxonomic classification of the reads from which the bins were produced. The `-min_perc_aa 10` flag instructed the tool to exclude bins which did not have at least 10% of amino acids in the multiple sequence alignment conducted by GTDB-Tk. The `-min_af 0.65` flag required a bin to have a minimum alignment fraction of 0.65 to a species cluster to be assigned to it by the tool.

Barrnap [421] (v. 0.9) was used to identify any Ribosomal RNA genes in the processed and filtered bins. Barrnap was run with the flags `-kingdom bac` to use the included Bacterial database and `-reject 0.1` to reject any detections of ribosomal genes which had a length of less than 10% of the expected length.

2.4.3 Annotation of metagenomic bins

Annotation of metagenomic bin files was conducted using Prokka [422] (v. 1.14.6) using default settings save for setting the flag `-kingdom Bacteria` to indicate that the bin files should be treated as if they were bacterial genomes.

2.5 Mapping of sample reads to metagenomic bins

2.5.1 Mapping

Prior to mapping the read files from the 81 *Acomys* samples all masked bin files were converted into single contig versions and then concatenated together into a single file using the command line. The subsampled read files were then mapped to this single bin reference file using Minimap2 [423] (v. 2.24r1122) with the flags `-ax sr` to produce output in a SAM file format and to indicate that the preset for genomic short read mapping should be used. After mapping files were generated Samtools [424] (v. 1.10) to produce BAM files for further analysis. Samtools was then used to filter the BAM to remove reads which either did not map themselves or had their mate fail to map; using the flag `-F 12` with the `samtools view` command. The BAM files were then sorted using the command `samtools sort` and indexed using the command `samtools index`.

2.5.2 Processing mapping data

Salmon [425] (v. 0.13.1) was used to generate values for 'reads per million' (In context, Reads per-Million (RPM)) for the mapping of the processed read files to the concatenated bins reference file. The bin file was indexed using Salmon and then Salmon was run with the following command: `salmon quant -t index file -l A -a bam file -threads 2 -seqBias -gcBias -writeUnmappedNames -numBootstraps 100 -no-version-check -o output file`. This command instructed Salmon to run and correct for hexamer bias, correct for GC bias, compute estimates of abundances using 100 bootstraps, infer the library type for the BAM file automatically and use two threads while running. The output `.sf` files contain the RPM value though labelled as TPM in the file.

Bin RPMs were evaluated to determine if any were differentially abundant between the two *Acomys* species, between the two sampling times or between the two sampling times within each host species. Only the RPMs from samples sequenced in 2020 and subsampled to a depth of 7,600,000 reads were used for this step to remove the potential confounding factor of different extraction and sequencing methods and dates. The geometric mean RPM for each bin was calculated for each of the host species, sampling month and both combined factors. For difference between the sampling times within each host species a paired Wilcoxon signed rank test [426] was used, the obtained P-values being corrected with the Benjamini-Hochberg method [427] for multiple testing. The `log2()+1` values were calculated for each RPM and used when considering the magnitude of any differences between the sampling times within each species.

For difference between each host species a Kruskal Wallis test [428] was carried out on the geometric mean RPM values for each bin comparing the results for *Acomys cahirinus* and *Acomys russatus*. The obtained P-values were corrected with the Benjamini-Hochberg method for multiple testing. The `log2()+1` values were calculated for each RPM and used when considering the magnitude of any differences between the host species.

For difference between each of the two sampling points a Kruskal Wallis test was carried out on the geometric mean RPM values for each bin comparing the results for all June and all November samples. The obtained P-values were corrected with the Benjamini-Hochberg method for multiple testing. The `log2()+1` values were calculated for each RPM and used when considering the magnitude of any differences between the sampling times.

Bins which showed a statistically significant (Q value ≤ 0.05) and substantial (`log2()+1` fold

change +/- 1 or greater) difference for one or more of the factors being examined (host species, sampling month and both combined) were categorised as either 'Host significant', 'Season significant' or 'Host within significant'. These categories were then subcategorised by which of the 2 host species, 2 sampling months or 4 combinations of the two factors the bin(s) showed a significant difference by. The Prokka annotations of these bins were then examined, looking at the count of predicted products and conducting a hypergeometric test to determine if any had a statistically significant difference (Q value ≤ 0.05) in their counts within these bins versus all bins.

All statistical analysis was carried out using R [393] (v. 4.2.0).

2.6 Lactic acid bacteria isolation from faecal samples

2.6.1 Isolation of lactic acid bacteria

Selective isolation of Lactic Acid Bacteria Lactic Acid Bacteria (LAB) was carried out on those faecal samples which had material remaining after DNA extraction. The selection for LAB was informed by the results of the different taxonomic classification tools on the read files obtained from the samples, relative ease of isolation and cultivation and the well reported presence and numerous proposed roles for such microorganisms in the intestinal microbiota of multiple animals; all discussed in the Introduction. Isolation and culturing of LABs from the faecal samples was conducted by **Nancy Teng** of the Quadram Institute using the following protocol:

1. De Man Rogosa and Sharpe media supplemented with Cysteine (0.5 mg μL^{-1}) was prepared as the selective media to isolate LAB from the *Acomys* faecal samples. Both broth and agar were produced. For 500 mL of MRS agar, 27.5 mg of MRS dry powder was weighed out and dissolved in 500 mL of deionised water along with 8 g of agar.

2. The media was autoclaved and left to temper at 50 ℃ for 40 minutes before 0.001% (v/v) of Cysteine (0.5 mg μL^{-1}) was added.

3. Once well mixed, plates were poured in a MSC II safety cabinet and left to dry for at least 30 minutes before storage at 4 ℃.

4. Production of broth followed the same process but without the addition of agar.

5. 100 mg of each faecal sample was aliquoted and homogenised in 1 mL of reduced sterile PBS (Sigma-Aldrich, UK). A dilution series down to 10^{-5} was prepared.

6. 200 µg L^{-1} of the faecal slurry dilution 10^{-4} was spread onto an MRS agar plate using a disposable hockey stick spreader. The same was repeated for dilution 10^{-5}.

7. The plates were left to dry in a safety cabinet for 15 minutes before being moved to the anaerobic cabinet to incubate anaerobically for 48 hours.

8. After 48 hours, non-confluent and morphologically distinct colonies were picked using a 5 µm sterile loop, from both dilution series, to a total of up to five colonies per sample. Each colony was quadrant streaked onto its own agar plate to purify the isolate.

9. The plate was left to incubate anaerobically for a further 48 hours.

10. After this incubation a single colony off each plate was selected and quadrant streaked onto a fresh plate. If a new colony appeared this was selected as well and streaked onto a new plate.

11. This process was continued at least twice more, bringing the total number of purification streakings to three.

12. Once the isolates were deemed to look pure, a colony was picked off and used to inoculate 10 mL of MRS+Cys broth.

13. The inoculated broth was incubated for 48 hours anaerobically.

14. After 48 hours, the inoculated broth was spun down at 2,600 rpm for 15 minutes to obtain a bacterial pellet. The supernatant was poured off, being careful not to disturb the pellet. With the remaining supernatant the pellet was resuspended.

15. 200 µL of the mixture was set aside for DNA extraction (described in **Subsection 2.6.2**).
16. The remaining mixture was resuspended with 1 mL of MRS with 20% glycerol and stored at -80 ℃.

The non-*Acomys* samples used were originally sequenced as part of an investigation into *Bifidobacterium* [429] and the read files obtained from this were provided by **Magdalena Kujawska** of the Quadram Institute. Isolation was carried out according to the method described in the cited publication.

2.6.2 Extraction of DNA from isolated lactic acid bacteria

The isolates obtained from the *Acomys* faecal samples were sequenced by **Dave Baker** at the Quadram Institute (Norwich, UK). DNA from the selected isolates was extracted from 200 µL of the mixture obtained after resuspension of the bacterial pellet. This was aliquoted into the lysing matrix tube E from the MP Biomedicals FastDNA SPIN Kit for Soil (California, USA). DNA was quantified using Qubit dsDNA BR assay kit (Invitrogen) with the adjustment for using 5 µL per tube as opposed to 10 µL per tube. After this, samples were diluted to 5 ng µL^{-1} and sent for sequencing to 60X depth using the Illumina MiSeq platform; a negative control was sequenced simultaneously to this.

The DNA extraction and sequencing of the *Apodemus* samples was as described in the cited publication.

2.6.3 Assembly, classification and annotation of bacterial assemblies

All sequenced reads were assembled the same way, both those from the *Acomys* faecal samples and those obtained from the *Apodemus* samples.

Assembly

All read files were analysed with FastP (v. 0.23.1) to remove low quality reads, trim adaptors and carry out quality control; default settings were used. Paired end read files containing reads which passed quality control (QC) were assembled using SPADES [430] (v. 3.13.1) using default settings save for the use of the flag `-careful`.

All assemblies were processed with Reformat from BBTools (part of the BBMap suite, v. 38.06) to remove contigs which were less than 500 base pairs in length. Default settings were used save for setting the `minlength=` flag to a value of 500.

CheckM (v. 1.2.0) was used to assess the completeness and contamination percentages of the assemblies. Cutoff values for completeness and contamination of $\geq 80\%$ and $\leq 5\%$ respectively were used to filter the assembly files for further analysis.

The assemblies were analysed with FastANI (v. 1.3) to detect any assemblies which might be identical to each other generated during the Metahood analysis. Any assemblies with a 99.9% or greater ANI similarity score to each other were considered to be potentially identical and one was randomly selected to use for further analysis, assemblies which did not meet this threshold were also retained for further analysis. dRep (v. 2.5.0) was used in addition to determine if any bins were extremely similar to each other and pick a best representative of any determined to be very similar. Default settings were used save for the flag `-pa` being used with a value of 0.999.

Any isolates which did meet or exceed the ANI threshold had their CheckM results compared to determine if either had higher completeness or lower contamination values, if so then that assembly was retained for further analysis. In the event that both values were identical for each assembly one was chosen at random to retain for further analysis.

Taxonomic classification of assemblies

GTDB-Tk (v. 1.7.0) was used to identify the assemblies taxonomically, using the command `gtdbtk classify_wf` to run the classification workflow. The flag `-min_perc_aa` was set to 10, filtering out genomes with less than 10% amino acid similarity to the generated alignments. The `min_af` flag was set to 0.65 to give a minimum threshold for alignment to a species cluster.

Detection of ribosomal genes in assemblies

Barrnap (v. 0.9) was used to identify any Ribosomal RNA genes in the assemblies. Barrnap was run with the flags `-kingdom bac` to use the included Bacterial database and `-reject 0.1` to reject any detections of ribosomal genes which had a length of less than 10% of the expected length.

Annotation of assemblies

Prokka (v. 1.14.6) was used to annotate the isolate assemblies. Default settings were used with the exception of setting the `-kingdom` flag to `Bacteria`

2.7 Phylogenies

2.7.1 Phylogeny of bins and assemblies

Marker-based phylogenetic trees could not be produced using ribosomal genes as there were no complete or partial ribosomal genes present in all bins, an alignment based tree was created instead. To begin all processed bin and assembly files were softmasked using BBMask, a tool from the BBMap [405] (v. 38.79) suite. Default parameters were used with the following exceptions: changing the kmer size range to sweep for exact repeats to 5-15 with the flags `minkr=5` and `maxkr=15`, changing the entropy value used for low complexity masking from 0.7 to 0.85 using the flag `entropy=0.85` and changing the minimum length of the repeat area to mask from 40 to 30 with the flag `minlen=30`. Following this a guide tree was created using JolyTree [431] which was used solely and only as a guide tree for the Cactus aligner [432] and had to have any numbers removed from the tree to be compatible with Cactus. Cactus aligns entire nucleotide sequencing files against each other in a multiple genome alignment. Cactus was run with the command `cactus -binariesMode local -cleanWorkDir never -noLinkImports -noMoveExports -maxCores 64 -defaultMemory 2G -maxMemory 1024G -defaultDisk 64G -maxDisk 512G`. Cactus was run on a directory containing the processed bin fasta files and processed isolate assembly files to produce a tree of the bins and assemblies. Cactus in this instance did not provide estimated branch lengths for the phylogenetic tree.

2.7.2 Phylogeny of assemblies

An alignment based phylogenetic tree was created to examine the relationships between the assemblies exclusively. The hardmasked files produced using BBMask were used for this process. Each assembly had all contigs concatenated into a single contig for each file and then these were all combined into a single multifasta file. This was used as the input for SibeliaZ [433] (v. 1.0.0), which was run with the following command: `sibeliaz -k 15 -b 200 -m 50 -a 1800 -t 1 -f 256`. Following this the MAF file produced was converted into a fasta file with a single line of data for each of the assemblies in the annotation using the Galaxy tool 'MAF to FASTA', which can be accessed at `https://usegalaxy.org/?tool_id=MAF_To_Fasta1&version=1.0.1`. The output options for this were set to the '`One Sequence per Species`'. After this FastTree [434] (v. 2.1.11) was used to produce a phylogentic tree in the Newick format from the fasta file, the command run was: `FastTree -gtr -nt input_fasta_file.fa > output_newick_file.nwk`. This newick file was uploaded to ITOL [435] which was used to produce the visualisations of the phylogenetic tree of the assessed files.

2.7.3 Combined phylogeny with external references

A broader phylogenetic tree was created including the isolate assemblies, the previously generated metagenomic bin and 173 reference genomes or high quality MAGs from the iMGMC. The external data used in the latter tree is detailed in subsection **2.1.7**. In this instance the tree was generated using Cactus in the way as that described in **Subsection 2.7.1**. As with the earlier tree generated by Cactus the tool does not incorporate branch length information in the tree. The tree produced was uploaded to ITOL to produce a visualisation.

2.8 Mapping of isolate assemblies to metagenomic reads

Isolate assemblies were mapped to the single concatenated reference file made of all processed *Acomys* reads described in **Subsection 2.5.1** using Minimap2 (v. 0.13.1) in the same manner as described in that subsection. Processing of the mapped read files was carried out in the same way, using Salmon (v. 0.13.1) and calculating differentially abundant isolates by host species, sampling point and the two factors combined, as described in subsection 2.5.2. As with the mapping to the metagenomic bin files, the analysis of the mapping to the isolate assemblies focused on those samples which had been subsampled to a greater read depth and for which there were paired June and November samples.

2.9 Halotolerance assessment through culturing

All culturing work described in this section was carried out by **Antia Acuna-Gonzalez** of the Quadram Institute, Norwich. In order to test the hypothesis that a demonstrable host phenotypic trait might also be observed in the microbiota a number of isolates obtained from the faecal samples were subjected to halotolerance culturing. As a positive control *Pediococcus pentosaceus* F 166 was obtained from the National Collection of Type Cultures (UKHSA, Salisbury, UK), NCTC number: 8066. This was shipped to the Earlham Institute in a freeze-dried form in August 2022 and stored refrigerated at 4 ℃ until September 2022 when it was brought to the Quadram Institute for use in the culturing experiment. The control was resuspended by putting it into 4 ml of the modified MRS-CYS media, pouring 2 ml of this into 10 ml of broth modified MRS-CYS media and culturing it for 48 hours anaerobically. The media was then centrifuged for 10 minutes at 4,000 rpm and resuspended in 2 ml of MRS with 20% Glycerol. This species was chosen as the positive control as it had been previously reported growing in conditions of elevated salinity [436] (6%).

10 isolates were assessed, 2 from the *Apodemus* and 8 from the *Acomys*. One of the two *Apodemus* isolates came from each of the two sampling locations described in that publication. Of the eight *Acomys* isolates, 3 were from samples obtained from *Acomys cahirinus* and 5 from samples obtained from *Acomys russatus*. The same media was employed as that used for the earlier described selection and isolation of LABs, save that sodium chloride was added to achieve a set range of salinities. The salinity percentages used were: 0, 1, 1.75, 2.5, 3.5, 5, 7.5 and 10%. The following protocol was used for the halotolerance testing:

1. Cultures obtained during the LAB isolation were grown anaerobically for 48 hours on the selective media described in 2.6.1.
2. Single colonies were chosen from the media and then inoculated in 5 mL of MRS+Cys broth.
3. These were incubated for a further 48 hours growth under anaerobic conditions.
4. The broth was then homogenised by manual pipetting and 5 µL of this mixture was inoculated in the wells of a 96 well plate with different salinities using a multichannel pipette. For the blank, negative control, 5 µL of 0% salinity media was inoculated in the blank wells.
5. The plate was placed into a Tecan F50 plate reader to allow for shaking of the media during growth; plates were placed without lids and with a gas permeable membrane.
6. Growth was measured over 48 hours in the plate reader.
7. Shaking occured every 15 minutes and lasted for 30 seconds.
8. The plate reader measured absorbance, as determined by optical density (OD600), measuring every 15 minutes.
9. This protocol was repeated 3 times for a total of 3 replicates.

The salinity ranges chosen were picked based on information regarding the halotolerance of LABs [437], the *Acomys* capacity to drink seawater [386] or reported halotolerances for laboratory mice (*Mus* strains) [438].

To prepare the media salt was added to the modified MRS-CYS broth to reach a total volume of 250 mL at the required salinity. Media was prepared in each bottle to the appropriate salinity and volume, on the day of the experiment it was stirred over 1 minute while pipetting the other salinities for homogenisation. It was then pipetted with a multichannel pipette into the wells.

2.9.1 Halotolerance informed analysis

Assemblies were categorised as either 'Not halotolerant', 'Slightly' or 'Halotolerant'. Assemblies which did not show any growth in any replicates in media with ≤ 2.5% were those categorised as 'Not halotolerant'. Assemblies which showed growth in at least one replicate in 2.5% salinity media were categorised as 'Slightly halotolerant' and assemblies which showed growth in media with ≤ 3.5% in at least one replicate were classified as 'Halotolerant'. The Prokka annotations of the assemblies were then examined, looking at the count of predicted products and conducting a hypergeometric test to determine if any had a statistically significant difference (Q value ≤ 0.05) in their counts in assemblies within one of the categories versus all assemblies.

Assemblies were also categorised by the host *Acomys* species the faecal sample had originated from and a hypergeometric test was conducted to determine if there was a statistically significant difference (Q value ≤ 0.05) in the counts of predicted products in assemblies from each species versus the other.

All statistical analysis was carried out using R (v. 4.2.0).

2.10 List of all bioinformatic software used by author and version

Where multiple versions of a software tool were used the second appearance in the following list is distinguished with bold text.

- FastP v. 0.19.7
- BBDuk v. 38.06
- Seqtk v. 1.0-r77-dirty (SIC)
- MetaPhlAn2 v. 3.0.10
- mOTUs2 v. 3.0.3.1
- Kaiju v. 1.7.3
- Kraken2 v. 2.0.8
- CheckM v. 1.1.2
- FastANI v. 1.3
- dRep v. 2.5.0
- GTDB-Tk v. 1.7.0
- Barrnap v. 0.9
- BBMask v. 38.79
- SibeliaZ v. 1.0.0
- MAF to FASTA (via Galaxy online tool) v. 1.0.1
- FastTree v. 2.1.11
- Prokka v. 1.14.6
- Minimap2 v. 2.24r1122
- Cactus 2.2.0
- Samtools v. 1.10
- Salmon v. 0.13.1
- SPADES v. 3.13.1
- **FastP v. 0.23.1**
- BBTools v. 38.06
- **Minimap2 v. 0.13.1**

Chapter 3

Results

- The results of testing of a number of taxonomic classification tools are presented
- The taxonomic composition of the microbiota of *Acomys* is presented
- The particulars of metagenomic bins produced from the *Acomys* microbiota are presented
- The particulars of lactic acid bacteria isolated from *Acomys* faecal samples are presented
- The results of halotolerance asessment of a number of lactic acid bacterial isolates from *Acomys* faecal samples are presented

3.1 Processing of read files

3.1.1 Processing for tool testing

Processing of the human mock microbial community saw a 18% reduction in the read count in the read files due to the removed reads either failing the FastP QC check or by matching to the human reference genome used with BBDuk. For the three rodent species with read files used in the testing phase the large difference in read depth between the *M. musculus* files and those from the *Acomys* samples led to a stark difference in the percentage reduction. The *Acomys cahirinus* read files saw a reduction of 21.3% (mean) or 20.5% (median) in the total read count from the raw files after the processing. The *Acomys russatus* read files saw a reduction of 19% (mean) or 21.2% (median) in the total read count from the raw files after the processing. The *M. musculus* read files saw a reduction of 84.2% (mean) or 85.7% (median) after processing.

3.1.2 Processing for analysis

Processing of the sequencing read files saw a reduction in the total read count in the read files due to the removed reads either failing the FastP QC check or by matching to the reference genomes for each *Acomys* species used with BBDuk. A mean of 95% of reads in the raw files were retained in the processed read files used for all subsequent analysis. By host species, the mean percentage of reads retained in the processed *Acomys cahirinus* files was 96.1%, the median was 97.3%, the maximum was 98.8% and the minimum was 78.2%. For *Acomys russatus* the mean percentage of reads retained was 95.7%, the median was 97.4%, the maximum was 98.7% and the minimum was 80.5%.

3.2 Taxonomic classification of read files

3.2.1 Classification for tool testing

Four different taxonomic classification tools were used in this stage of the project, which broadly correspond to three different approaches to classifying sequencing reads for microbiota samples. All four tools were able to classify at least some of the reads in each of the samples though with different degrees of success. To establish whether the processing steps the read files had undergone impacted the ability of the tools to classify the reads the tools were run on the raw files and then the processed files. For the processed files the most permissive settings were used for Kaiju and Kraken 2, being an error allowance of 20 and a minimum required confidence score of 10% respectively.

Figure 3.1 shows the results of this comparison stage for the rodent files. There were no significant differences between the percentage of reads classified in the raw and the processed files for any of the three rodent species using Metaphlan or Kraken 2. Using Kaiju and mOTUs there was a statistically significant difference between the raw and processed files for the *Acomys russatus* files. The mean percentage of reads classified by Kaiju in the raw *Acomys russatus* files was 74.6% compared to 76.2% in the processed files. For mOTUs the percentage of reads classified in the raw *Acomys russatus* files was 0.0055% and in the processed files it was 0.0056%, though this is not abnormal for the standard operation of the tool which uses a small subset of reads from each sample.

The different tools were run on the human mock community which had a known composition. **Table 3.1** shows whether or not each of the 20 species was detected by the classification software in the procesed read file, including for Kaiju and Kraken the parameter altered in the run. All four tools did well at detecting the presence of the twenty species however the very different percentage of reads classified, discussed below, indicate that the ability to detect a known true positive may not be a useful indicator of the suitability of the tool for the full set of *Acomys* data.

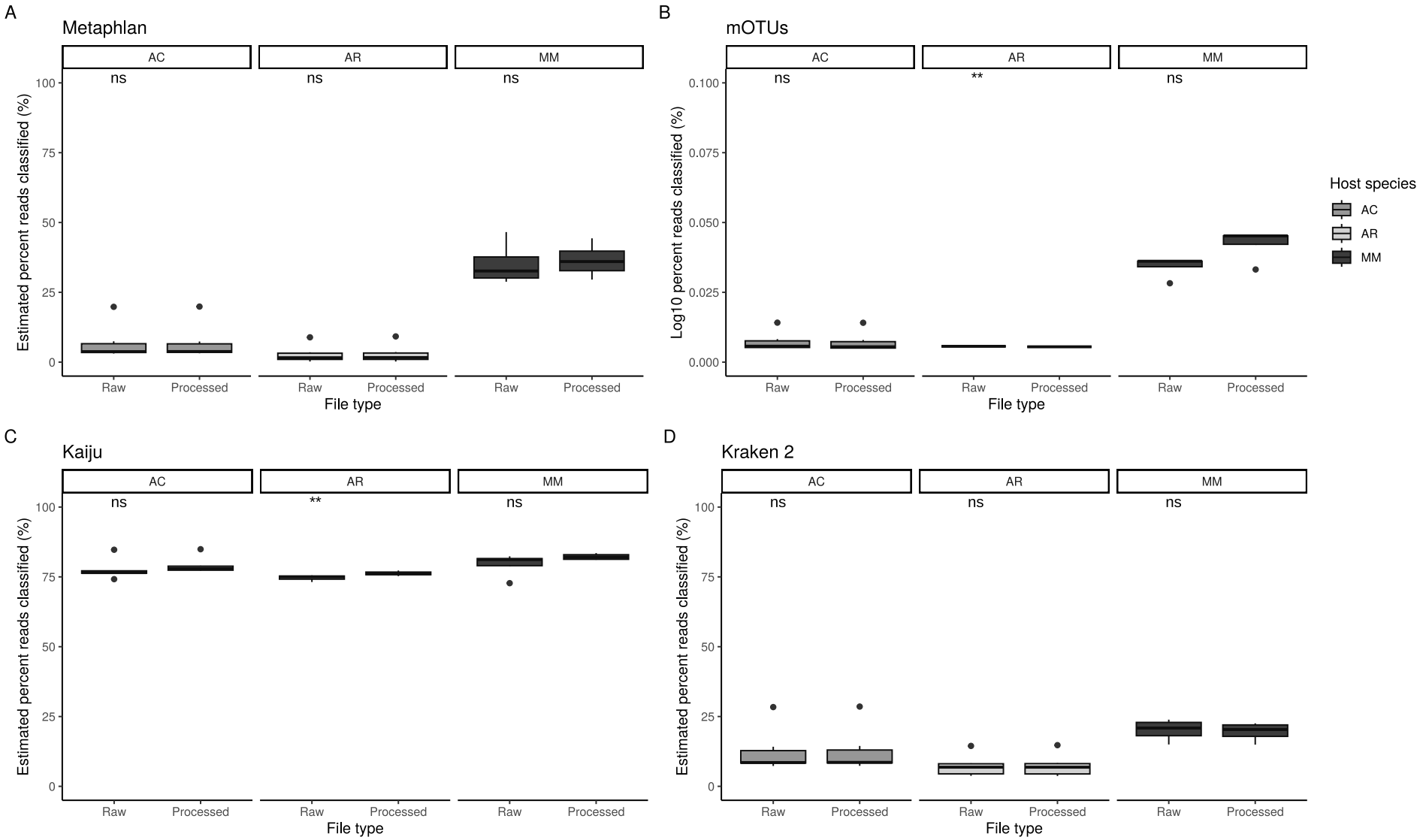

Figure 3.1: 3.1A shows the range of percent of reads classified for each of the files analysed by Metaphlan for the three rodent species with data used in the tools testing, distinguishing from the raw read files and those which had been processed. **3.1B** shows the Log10 values for the range of percent reads classified by mOTUs. **3.1C** shows the range of percent of reads classified by Kraken 2 with a minimum required confidence score of 10%. **3.1D** shows the range of percent of reads classified by Kaiju with an error allowance of 20.

Metaphlan 3

The results for Metaphlan 3 can be seen on a per sample basis in **Figure 3.2A**. The two different sets of *M. musculus* samples analysed had a considerably larger share of their reads classified by Metaphlan than either of the two *Acomys* species' samples. Within the two *Acomys* species samples there appeared to be two outliers, which had a greater percentage of their reads classified than the others of their species; both were November samples. Except for those two samples, Metaphlan could not classify more than 8% of the *Acomys* sample reads; the median percentage of reads classified by Metaphlan was 3.52%. In the *Mus* samples, the minimum percentage of reads in a sample Metaphlan could classify was 29.6%, with the maximum being 44.3%.

Metaphlan classified 115.98% of the reads in the human mock microbial community file. Metaphlan does not directly count the number of reads assigned to each of the taxa it detects and so it cannot be assessed in the same way as the other tools. Looking into the results more closely for the human mock microbial community, it estimates 15,720,250 reads as being within the clade 'Bacteria' - which is greater than the 13,553,639 reads in the actual sample fastq file after pre-processing. Notably however, it also estimates 11,387,028 reads as being 'UNKNOWN' - which is 84.01% of the reads in the submitted file.

mOTUs

Across all processed rodent read files analysed mOTUs assigned at least 1 read to 553 ref-mOTUs, varying from a minimum of 28 in sample AC16N to a maximum of 134 in sample ERR675516 (a *Mus* sample). Across all processed rodent read files mOTUs detected at least one read from a mean of 99.8 ref-mOTUs and a median of 83.5. On a host species by host species basis, the processed *Acomys cahirinus* files had a mean of 70.83 ref-mOTUs with a non-zero relative abundance, the *Acomys russatus* files had a mean of 79.83 ref-mOTUs with a non-zero relative abundance and the *Mus* files had a mean of 141.3 ref-mOTUs with a non-zero relative abundance. The processed *Acomys cahirinus* files had a median of 76.5 ref-mOTUs with a non-zero relative abundance, the *Acomys russatus* a median of 80 ref-mOTUs with a non-zero relative abundance and the *Mus* a median of 138.5 ref-mOTUs with a non-zero relative abundance.

The most (relatively) abundant reference mOTU (ref-mOTU(s)) detected in the *Acomys cahirinus* files was the same in each, `ref_mOTU_v3_01032`. The two other rodent species did not have the same most abundant ref-mOTU in each of their samples. `ref_mOTU_v3_01032` was the most abundant ref-mOTU detected in 4 of the *Acomys russatus* samples however, being the most abundant in samples AR27, AR27N, AR34J and AR34N. The two other *Acomys russatus* samples did not have the same most relatively abundant ref-mOTU detected in each other. 7 different ref-mOTUs were the most relatively abundant across the 16 samples with the *Mus* and the two *Acomys russatus* samples each having a unique most abundant ref-mOTU compared to all other samples. The least abundant ref-mOTU which had a non-zero relative abundance was more diverse, with only two of the files having the same least abundant ref-mOTU. The single shared ref-mOTU was `ref_mOTU_v3_01039`, which was the least common in files AC16J and AR34N.

The percentage of reads which were classified by mOTUs can be seen in **Figure 3.2B**. The authors of the tool believe that their use of a small proportion of reads for analysis and reporting of relative abundances is sufficient to provide accurate classification.

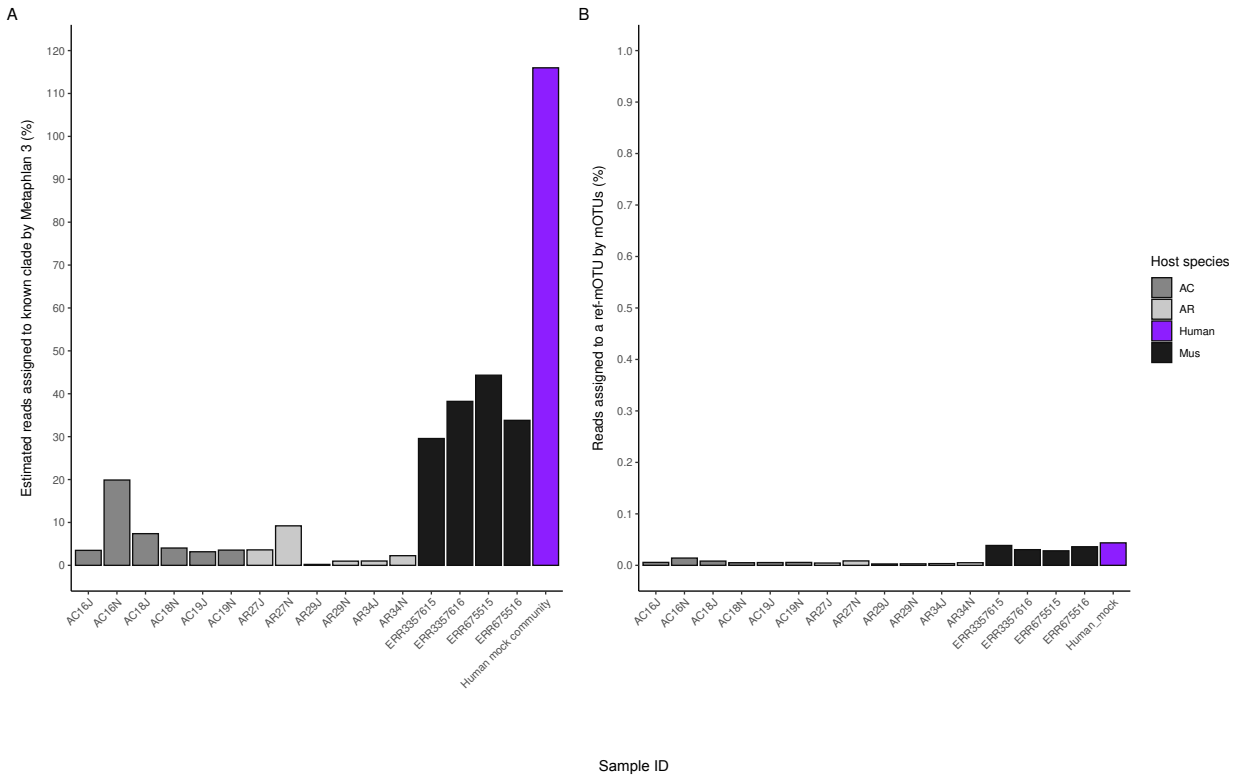

Figure 3.2: A. shows estimated percentage of reads classified by Metaphlan in each processed sample read file. **B.** shows the percentage of reads classified by mOTUs in each processed sample read file.

Looking at the phylum level results for relative abundance for the rodent pilot samples, the phyla which had summed relative abundances (from all samples) of at least 0.1 were Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Verrucomicrobia and Deferribacteres (also called Deferribacterota). The summed relative abundances above were calculated by summing the relative abundances for all reference mOTUs within that phylum and then summing these values. In all individual samples both pre- and post-processing the phylum with the highest relative abundance was Firmicutes, ranging from a relative abundance of 0.96 in AC16N after processing to 0.20 in the *Mus* sample ERR3357615 prior to processing. The summed phylum level relative abundances from mOTUs analysis of the different files used for tool testing can be seen in **Figure 3.3**.

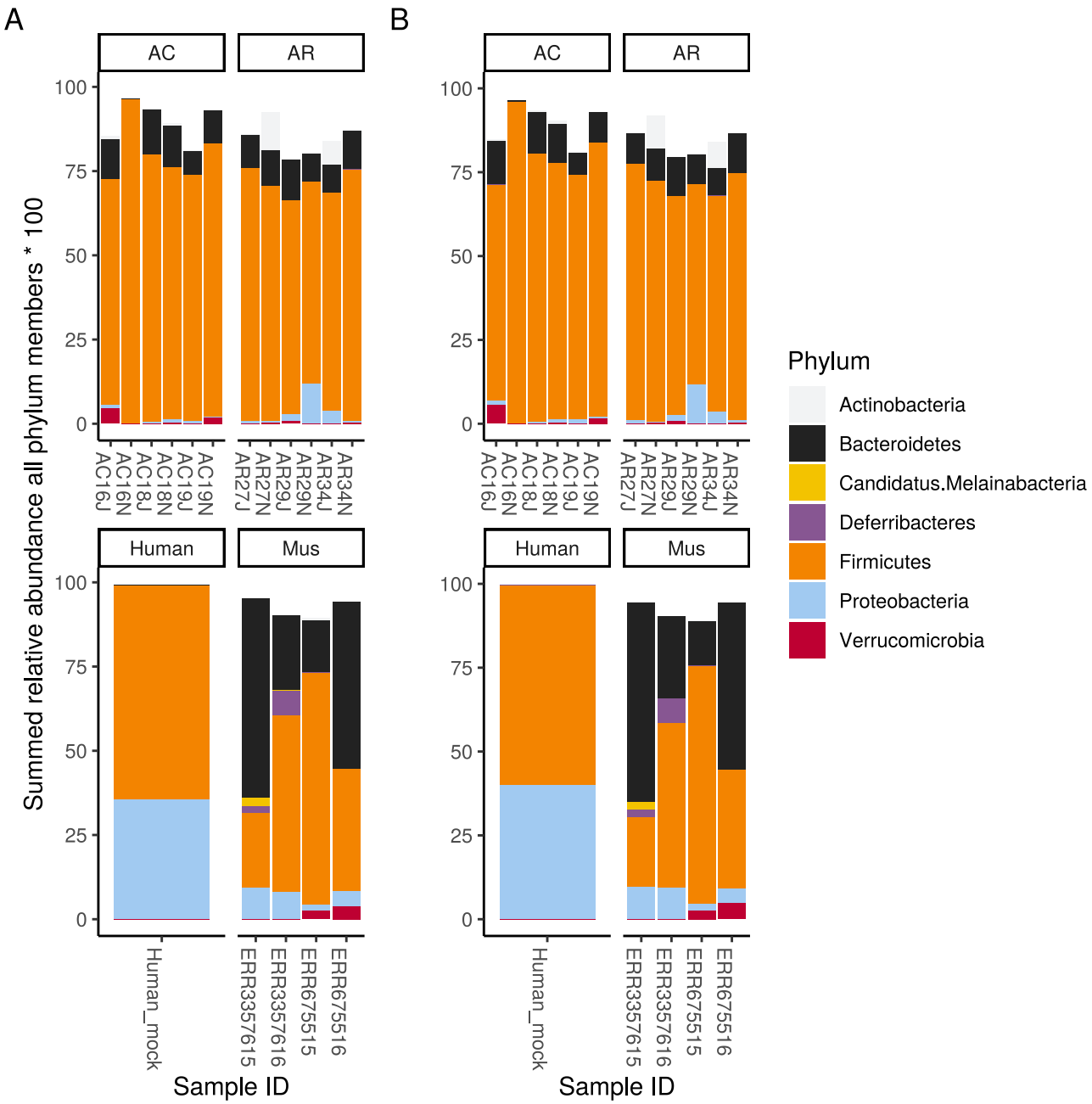

Figure 3.3: Summed relative abundances (multiplied by 100) at the phylum level for mOTUs, showing the 7 phyla which had the greatest combined relative abundance across all samples. Plots are facetted by host species and coloured by phylum. **A.** Processed files. **B.** Raw files.

For the human mock microbial community, mOTUs assigned relative abundance to 27 genera, more than the 17 actually in the community. False positive detections included *Enterobacter*, *Moraxella*, *Promicromonospora*, *Sanguibacter* and *Xylanimonas*. Of the 17 actually present genera, mOTUs assigned relative abundance values to all of them though it failed to detect two of the species in the community. The greatest relative abundance was assigned to the genera *Streptococcus* (0.36), *Staphylococcus* (0.24), *Rhodobacter* (0.22) and *Enterobacter* (0.08). mOTUs found relative abundance for the true positive genera in accordance with the ratios of their known operon counts.

Kraken 2

At the most permissive setting, requiring only a minimum confidence score of 10%, Kraken 2 failed to reach 10% median percentage of reads classified (from all samples) in either of the two *Acomys* species. The laboratory and wild *M. musculus* samples had a median percentage of reads classified above 20% when using the same minimum confidence score; and the human mock community sample had 99.96% of its reads classified by Kraken 2 at 10% minimum confidence. As the required minimum confidence score for a classification increased, the median percentage of reads classified dropped for all host species - though to a greater extent for the two *Acomys* species. Kraken 2 never failed to classify at least 1% of the reads in the *M. musculus* samples at any minimum confidence level, but for some *Acomys* samples it failed to classify even 1% of the reads at the minimum confidence scores. For example, with a minimum confidence score of 50%, Kraken 2 identified only 0.51% of reads in sample AR29J and 0.59% in the AR29N sample. Given that these are from the same individual it might not be surprising that Kraken 2 had similarly low levels of success. When the minimum confidence score was increased to 75%, 7 of the 12 *Acomys* samples failed to reach a minimum of 1% reads classified. The *M. musculus* samples' median percentages of reads classified were below the human mock community level at any given minimum confidence score. At the strictest level, a minimum confidence score of 95%, Kraken 2 only detects a very narrow range of taxa in the pilot samples, irrespective of both host species and month of sampling. Within the *M. musculus* samples there was a difference between the mean percent of reads classified at any given minimum confidence score for the wild samples and the laboratory reared animals. At the 95% minimum confidence score, the mean percentage classification of reads for wild *M. musculus* was 4.82% while for the lab mice samples it was 7.93%. As the strictness of the minimum confidence score decreased both types of samples saw an increasing mean percentage of the reads being classified, until reaching the minimum confidence score of 25% when the two sample types flipped in their relationship. At both 25% minimum confidence score and 10%, the wild mice had a greater mean percentage of reads classified than the lab mice. **Figure 3.4A** provides the median percentage of reads classified by Kraken 2 for each of the three rodent species and highlights the considerable differences between the *M. musculus* samples along with the major impact of changing the minimum required confidence score.

In the human mock community, Kraken 2 classified over 90% of reads at 10, 25 and 50% minimum confidence scores, though it did drop slightly each time. At a 66% minimum confidence score it managed to classify 87.83% of the reads, 74.98% of the reads with minimum confidence score of 75% and then dropped precipitously to only 14.23% of reads being classified when the minimum confidence score was set to 95%. This, though, was more than any of the *M. musculus* samples and considerably more than any *Acomys* sample at this, most stringent level. Kraken 2 always found more NCBI taxIDs in the sample than were in the original community. As can be seen in **Figure 3.5A** the number of false positives decreased with increasingly strict requirements for the minimum confidence level, as there was a drop in the number of unique taxIDs detected. The fact that the number of taxIDs with more than 100 reads remained largely consistent is a sign either that the bulk of the reads which Kraken 2 classified were being assigned to a limited number of taxIDs - rather than it detecting a large number of IDs each of which have a small number of assigned reads - or that it classified the majority of reads into increasingly large taxonomic levels with increasing minimum confidence score.

Even at the most permissive minimum confidence score, 10%, Kraken 2 did not detect all 20 strains in the sample files. However, it did detect all of the overall species, as can be seen in **Table 3.1**. Of the 7 strain IDs not detected by Kraken 2 with a 10% minimum confidence score,

1 was amongst the most abundant in the mock community (*Escherichia coli* str. K-12 substr. MG1655), 1 was amongst the next highest level of abundance (*Pseudomonas aeruginosa* PAO1) and 1 was the next highest level of abundance (*Helicobacter pylori* 26695). The remaining 4 strains not detected had the lowest level of abundance in the initial community. There were multiple representatives of each species in the database though whether or not the particular strains were present cannot be determined from the database itself.

As the minimum required confidence score increased, becoming stricter, the number of exact strain taxIDs detected dropped, until it plateaued at 5 taxIDs with a minimum confidence score of 75% and 95%. Though the number of detections of the overall species also dropped with the increasing minimum confidence score it did not decrease to the same extent. Even at a minimum confidence score of 95% Kraken 2 detected 18 of the 20 overall species taxIDs for the specific strains in the community. At any minimum confidence score above 10% Kraken 2 detected more false positives from genera not belonging to one of the 20 known member strains than within them. The number of false positive genus read classifications rose with increasing minimum confidence score until reaching 75% where the number of reads classified as such started to decrease. At all minimum confidence levels, Kraken 2 classified many more reads with the exact NCBI taxID of the species of one of the 20 known strains than it did the taxID of the strain itself. By the strictest minimum confidence score, 95%, Kraken 2 detected more taxID not belonging to anything in the genera of the 20 known strain members than it did to either the exact species taxID or to any within-genus false positive taxIDs. This suggests that increasing stringency to this level did not necessarily lead to an equivalent reduction in false positives. It is worth noting though that the number of classified reads assigned to taxa outside the genus of one of the 20 known member strains increased with the minimum required confidence score until it exceeded the number of reads classified with the 'exact species' taxID between 66-75% minimum confidence score.

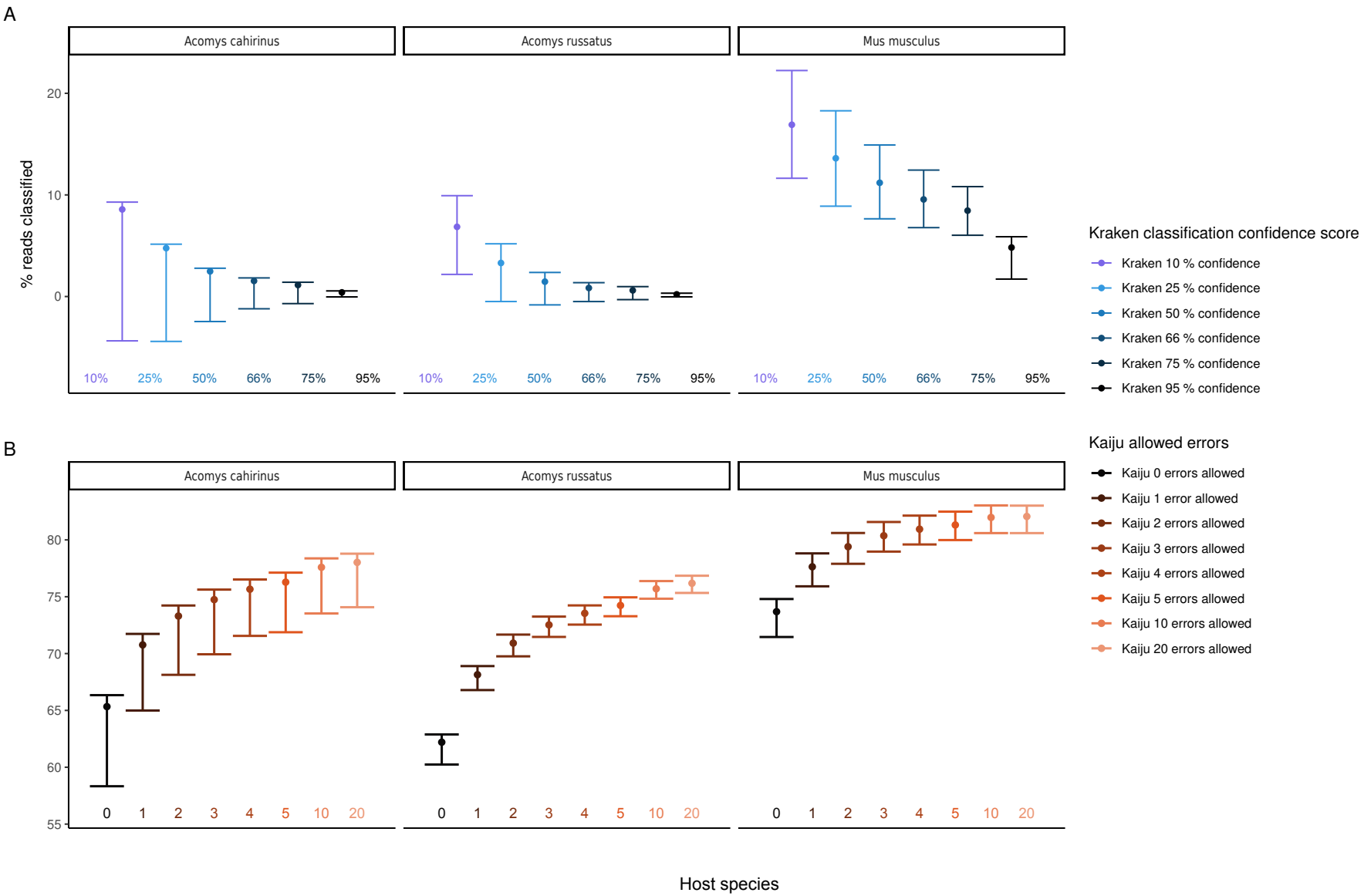

Figure 3.4: 3.4A shows the median percentage of reads classified by Kraken 2 at the indicated minimum required confidence score for processed read files from each of the three rodent species. **3.4B** shows the median percentage of reads classified by Kaiju at the indicated error allowance for processed read files from each of the three rodent species.

Kaiju

Kaiju was the most successful classifier by number of reads classified, as shown in **Figure 3.1** it always classified at least 50% of the reads in the three rodent species samples even at the strictest allowed error number (0). It had median values for percentage reads classified over 60% in all rodent samples at all error allowances. There was greater variance within the median percentage of reads classified by Kaiju in the *A. cahirinus* samples than within either the *A. russatus* or *Mus* samples. Similarly to Kraken 2, Kaiju saw a greater percentage of reads in *Mus* samples classified at all error allowances. Unlike Kraken 2 the strictest setting tested saw notable differences between the two Acomys species sample median percentage reads classified as compared to the Mus samples.

In all rodent samples at any of the tested error allowances, Kaiju achieved a mean percentage of reads classified of at least 62%. The lowest percentage of reads classified in any rodent sample by Kaiju was 61.53%, in sample AR29J, with an error allowance of 0. In the rodent samples, Kaiju had the greatest percentage of classified reads in each of the three species when using an error allowance of 20, as can be seen in **Figure 3.4B**. The mean percentage of reads classified in the *Mus* samples never fell below 70%, even with the 0 error allowance - though the median percentage of reads classified in *Mus* samples did not meaningfully increase with the doubling of permitted errors from 10 to 20. A similar trend can be seen across all the rodent species, in which the doubling of permitted errors from 5 to 10 led to a larger increase in the median percentage of reads classified than the doubling from 10 to 20. Kaiju did not have any appreciable difference in the mean percentage of reads classified at each error allowance tested between the wild and lab reared Mus samples; at its greatest this difference was 2.27% with an error allowance of 1.

With the human mock microbial community, Kaiju classified almost all reads, never falling below 99% reads classified in the processed file. As with the rodent samples the doubling of allowed errors from 10 to 20 did not have as significant an effect on the number of reads classified by Kaiju in the human mock community as the doubling from 5 to 10 allowed errors. Given, though, that Kaiju classified almost all the reads in the mock community the change with the doubling is extremely small in both cases. **Figure 3.5B** shows the results for Kaiju analysis of the human mock microbial community in terms of the number of NCBI taxonomic IDs detected and the number of detected taxIDs assigned at least 100 reads by Kaiju. Kaiju at its strictest error allowance identified over 6,000 unique taxIDs; the sample had a known composition of 20 strains and thus should only show 20 taxIDs at most. The number of unique tax IDs detected and the number of tax IDs with more than 100 reads assigned to them were identical at each error allowance. This indicates that whenever Kaiju, with an increasingly permissive error allowance, detected a new taxID it readily assigned at least 100 reads to it. The number of unique taxIDs detected was never less than 300 times greater than the known number of taxa in the sample, even at the strictest error allowance of 0. With the most generous error allowance Kaiju detected 13,099 unique taxonomic IDs.

Unlike Kraken 2, Kaiju did not fail to detect any of the species taxonomic IDs at any error allowance, as can be seen in **Table 3.1**. While Kaiju failed to detect all of the 20 strain specific taxIDs (only ever managing to detect 19) it maintained the same detection of strains at all error allowances; the number of the 20 strain taxIDs detected remained the same at both the strictest and most permissive error allowances. Kaiju detected an increasing number of non-same genus taxIDs with an increasingly permissive error allowance, though even at the strictest error allowance it detected 1,122 non-same genus taxIDs. At an error allowance of 10 or 20, Kaiju detected more taxIDs from outside the genus of any of the known 20 member strains than either

within the genus but not one of the right species, or of the right species but the wrong strain.

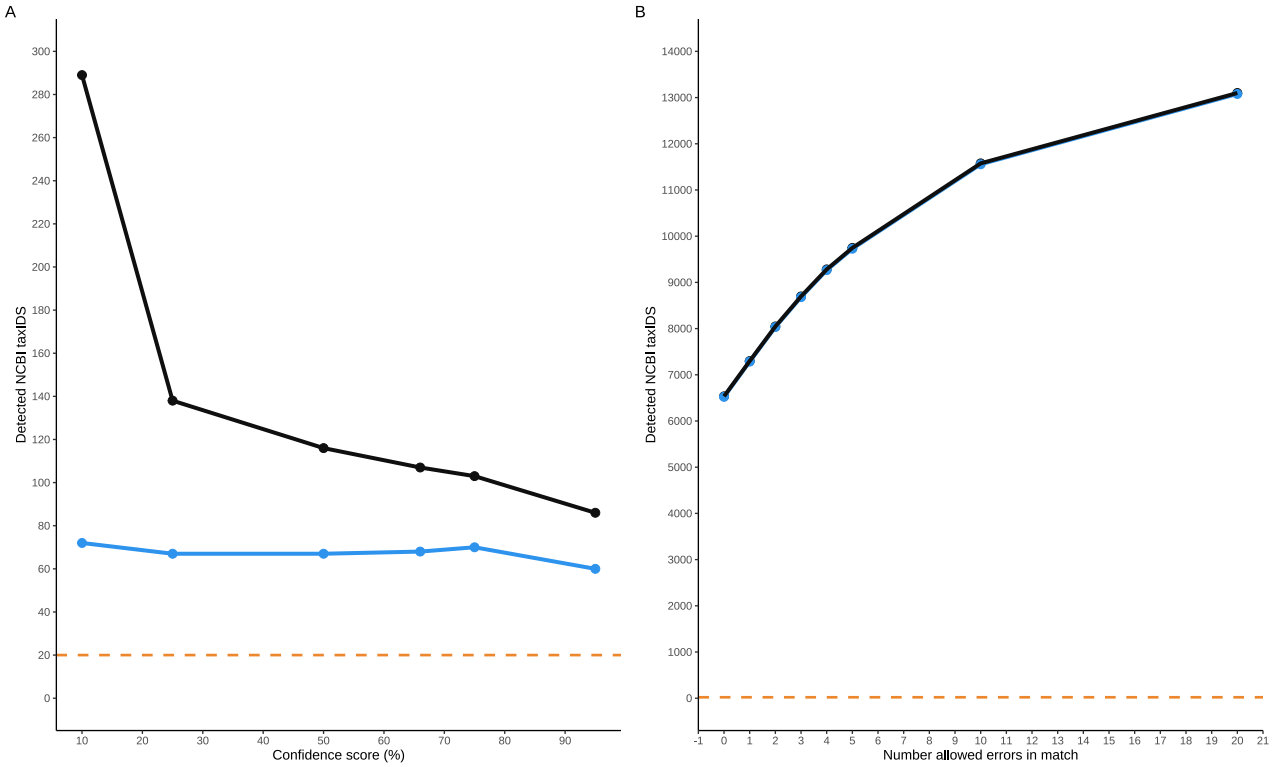

Figure 3.5: A shows in black the number of NCBI taxonomic IDs detected and in blue the number of taxIDs assigned at least 100 reads by Kraken 2 in the processed human mock microbiota reads file with changing minimum required confidence score. **B** shows in black the number of NCBI taxonomic IDs detected and in blue the number of taxIDs assigned at least 100 reads by Kaiju in the processed human mock microbiota reads file with changing error allowance.

Species	NCBI TaxID	Metaphlan	mOTUs	Kaiju 0 errors	Kaiju 1 errors	Kaiju 2 errors	Kaiju 3 errors	Kaiju 4 errors	Kaiju 5 errors	Kaiju 10 errors	Kaiju 20 errors	Kraken 95% conf. score	Kraken 75% conf. score	Kraken 66% conf. score	Kraken 50% conf. score	Kraken 25% conf. score	Kraken 10% conf. score
Acinetobacter baumannii	470	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Schaalia odontolytica	1660	1	**0**	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Bacillus cereus	1396	1	**0**	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Phocaeicola vulgatus	821	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Clostridium beijerinckii	1520	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Deinococcus radiodurans	1299	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Enterococcus faecalis	1351	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Escherichia coli	562	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Helicobacter pylori	210	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Lactobacillus gasseri	1596	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Listeria monocytogenes	1639	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Neisseria meningitidis	487	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cutibacterium acnes	1747	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pseudomonas aeruginosa	287	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cereibacter sphaeroides	1063	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Staphylococcus aureus	1280	1	1	1	1	1	1	1	1	1	1	**0**	1	1	1	1	1
Staphylococcus epidermidis	1282	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Streptococcus agalactiae	1311	1	1	1	1	1	1	1	1	1	1	**0**	**0**	**0**	**0**	1	1
Streptococcus mutans	1309	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Streptococcus pneumoniae	1313	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.1: Lists the 20 species in the human mock community sequenced and used with the different classificationt tools, the NCBI taxonomy ID for the species and then whether the species was detected (1) or not (0) by the indicated tool and parameter variation (where relevant). Failed detections are highlighted in bold.

3.2.2 Classification for analysis

In this stage of the project Kraken 2 and Kaiju were run with multiple parameters, as in the testing stage, however results will only be discussed for a 50% minimum confidence score with Kraken 2 and for 0 errors allowed with Kaiju. The percentage of reads classified by each tool are shown in **Figure 3.6**. As in the tool testing results it is worth noting that by design mOTUs uses a small subset of the reads in the sample files which may make the percentage of reads classified appear distored.

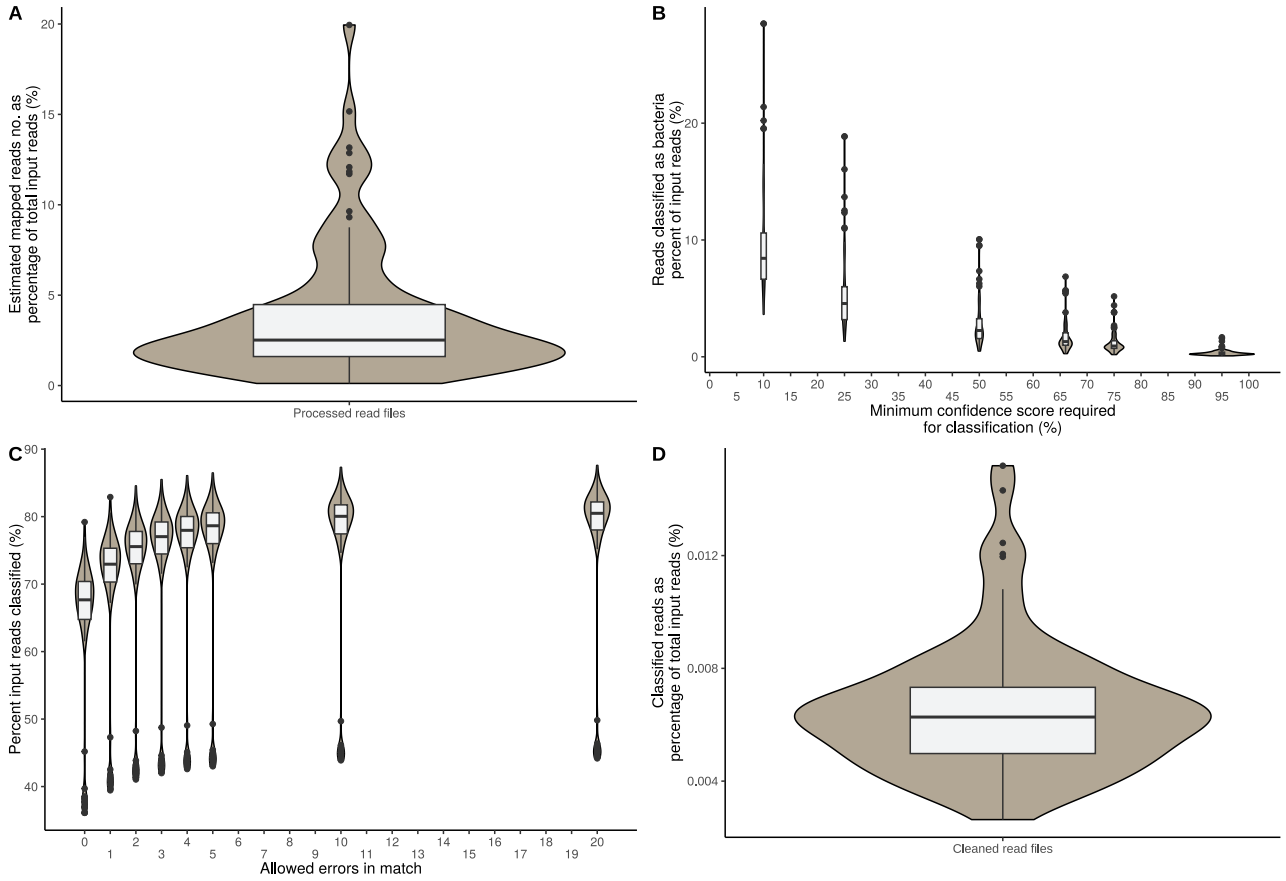

Figure 3.6: A. Box and violin plot of the estimated percentage of reads in all processed input files which could be classified by Metaphlan 3 by mapping to one of the marker sequences. **B.** Box and violin plots of the percentage of reads in all processed input files which could be classified by Kraken 2 at the indicated minimum confidence score required. **C.** Box and violin plots of the percentage of reads in all processed input files which could be classified by Kaiju at the indicated error allowance. **D.** Box and violin plot of the percentage of reads in all partially processed (see **Sub-subsection 2.3.2**) input files which could be classified by mOTUs.

The potential impact of the processing the files underwent on the ability of Metaphlan 3, Kraken 2 and Kaiju to classify them were tested by comparing the results for the raw and processed files, which are shown in **Supplemental Figure 5.2**. Processing did not have a significant impact on the ability of Metaphlan 3 or Kraken 2 to classify reads in any of the samples, it did however lead to a significant increase in the reads classified by Kaiju in both *Acomys* species; an increase in the percent of reads classified after processing in both.

mOTUs

Figure 3.6D. shows the percentage of reads classified by mOTUs, with the caveat that it is designed to use only a small proportion of reads in a sample. mOTUs assigned at least one read

to 144 different taxa across all samples. Across all samples mOTUs detected non-zero relative abundance in at least one sample for 550 unique reference mOTUs. By host species the sample which had the most unique reference mOTUs detected which had a non-zero relative abundance were AC4N for *A. cahirinus* with 177 unique reference mOTUs and for *A. russatus* it was AR42J with 158 unique reference mOTUs. The median number of unique reference mOTUs with a non-zero relative abundances for each host species was 118 for *A. cahirinus* and 116 for *A. russatus*. The reference mOTU which had the highest assigned relative abundance from any sample was `ref_mOTU_v3_01032` in sample AC16N with a relative abundance of 0.78%. This reference mOTU also had the greatest summed relative abundance in both host species, 12.2% from all *A. cahirinus* samples and 4.9% from all *A. russatus* samples. The reference mOTU corresponds to the species *Lactobacillus kefiranofaciens*. Summing all the relative abundances from each sample for the two host species there were three reference mOTUs which had a minimum of 1% summed relative abundance for the *A. cahirinus* samples and 7 for the *A. russatus* samples. The distribution of summed relative abundances for the 550 reference mOTUs by host species can be seen in **Supplemental Figure 5.1** and the particulars of those reference mOTUs which had summed relative abundances of at least 1% in one of the two host species can be found in **Supplemental Table 5.3**.

Looking at the relative abundances reported by mOTUs at the phylum level, summing the relative abundances across all samples there were 5 phyla which had a summed relative abundance of at least 1. These were Firmicutes, Bacteroidetes, Verrucomicrobia, Proteobacteria and Actinobacteria. In all but five samples the phylum with the greatest relative abundance was Firmicutes, the outlying samples were all *A. russatus* samples and all but one of them were June samplings. The range of relative abundances for Firmicutes ranges from 0.96 to 0.26, for Bacteroidetes it ranges from 0.3 to 0 - the lowest non-zero relative abundance for Bacteroidetes was 0.004. Firmicutes was also the only phylum of the five which had assigned relative abundance in all samples. The relative abundances for these 5 phyla can be seen for all samples in **Figure 3.7**. The median relative abundance for Bacteroidetes in *A. cahirinus* samples was 0.079, in *A. russatus* samples it was 0.13. For Verrucomicrobia the median relative abundance was 0.001 for *A. cahirinus* and 0.003 for *A. russatus*. For Proteobacteria the median relative abundance was 0.009 for *A. cahirinus* and 0.011 for *A. russatus*. For Actinobacteria the median relative abundance was 0.0013 for *A. cahirinus* and 0.0013 for *A. russatus*. For Firmicutes the median relative abundance was 0.74 for *A. cahirinus* and 0.66 for *A. russatus*. Of the 134 phyla in the mOTUs database used to analyse these samples, 9 (6.7%) had a summed relative abundance across all samples of greater than 0. 5 of the 134 phyla, 3.7%, had summed relative abundances of 0.1 or greater.

On the genus level the ten genera with the greatest summed relative abundance across all samples were *Lactobacillus*, *Lachnospira*, *Oscillibacter*, *Prevotella*, *Clostridium*, *Akkermansia*, *Alistipes*, *Bacteroides*, *Bifidobacterium* and *Ruminococcus*. *Lactobacillus* accounted for at least 0.5 relative abundance in 16 samples (19.8% of all samples), the genus being the only one which had a recored relative abundance of 0.5 or greater in any sample. The median relative abundance for *Lactobacillus* for *A. cahirinus* samples was 0.3 while for *A. russatus* samples it was 0.22. The median relative abundance for *Lachnospira* was 0.14 in *A. cahirinus* and 0.038 for *A. russatus*, for *Oscillibacter* it was 0.04 for *A. cahirinus* and 0.05 for *A. russatus*. Of the 2,194 genera in the mOTUs reference database used for analysis, 71 (3.2%) had a summed relative abundance across all samples of greater than 0. 23 (1.04%) of the genera had summed relative abundances from across all samples of at least 0.1. The relative abundances for the 10 genera can be seen in **Figure 3.8**. The relative abundances of the 10 genera with the greatest summed relative abundance across all samples split by host species and sampling month can be seen in

Figure 3.9.

For these ten genera there were 2 which showed a statistically significant difference in mean relative abundance between the two sampling months within a species. These were *Bifidobacterium*, which had a statistically significant difference (Kruskal-Wallis test p-value 0.003307) within *A. russatus*, and *Lachnospira*; which had statistically significant differences (*A. cahirinus* p-value 0.02627, *A. russatus* p-value 0.0313) between sampling months in both host species. Looking just at the across host species level, 6 of the 10 genera have a statistically significantly different mean relative abundance in one of the host species. Those which have a significantly greater relative abundance in *A. cahirinus* were *Clostridium*, *Lachnospira* and *Lactobacillus*. Those with a significantly greater relative abundance in *A. russatus* were *Bacteroides*, *Oscillibacter* and *Prevotella*.

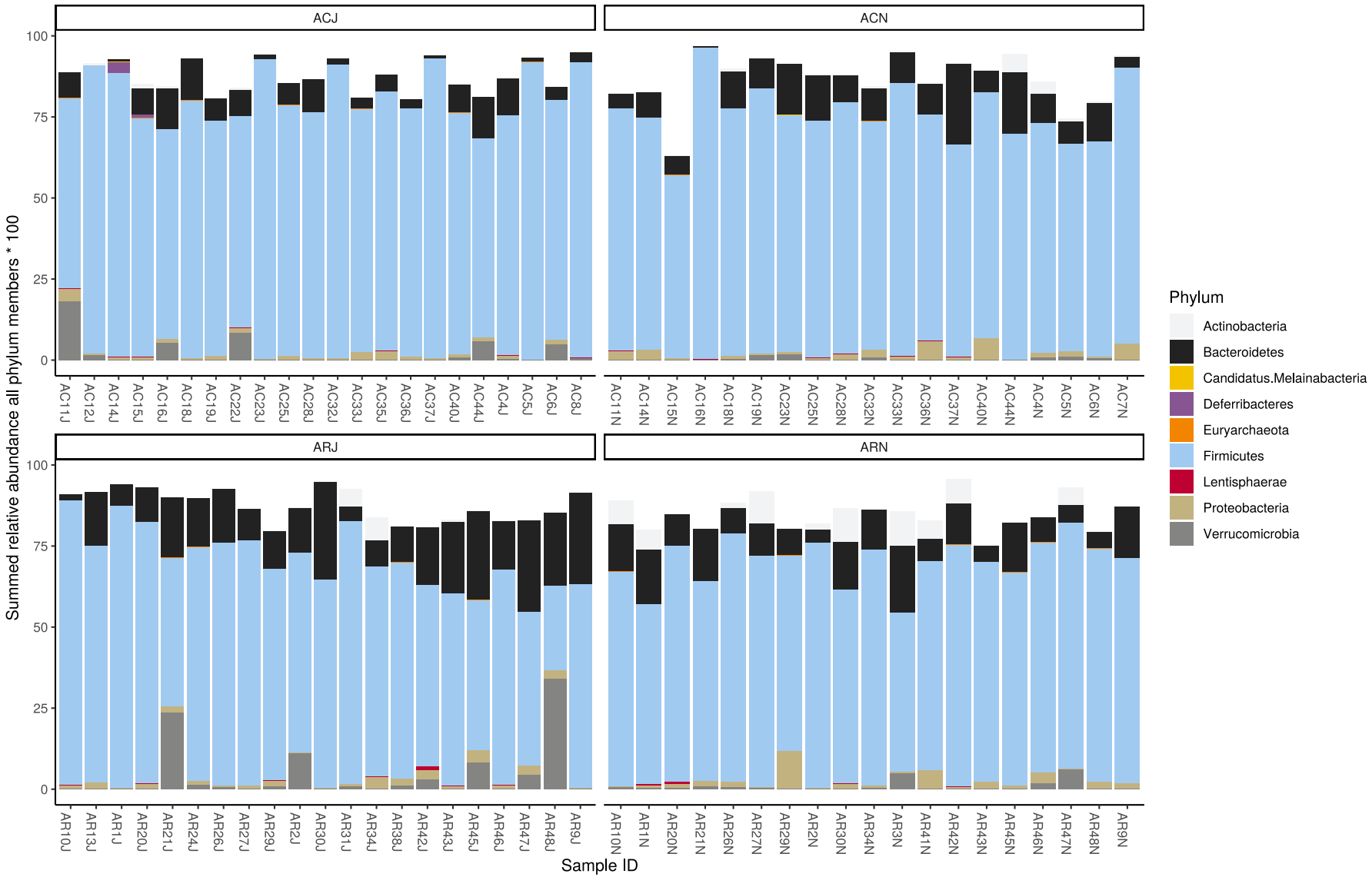

Figure 3.7: Stacked bar charts showing the relative abundances for the 5 phyla which had the greatest summed relative abundance across all samples. Plot is facetted by host species and month and coloured by phylum.

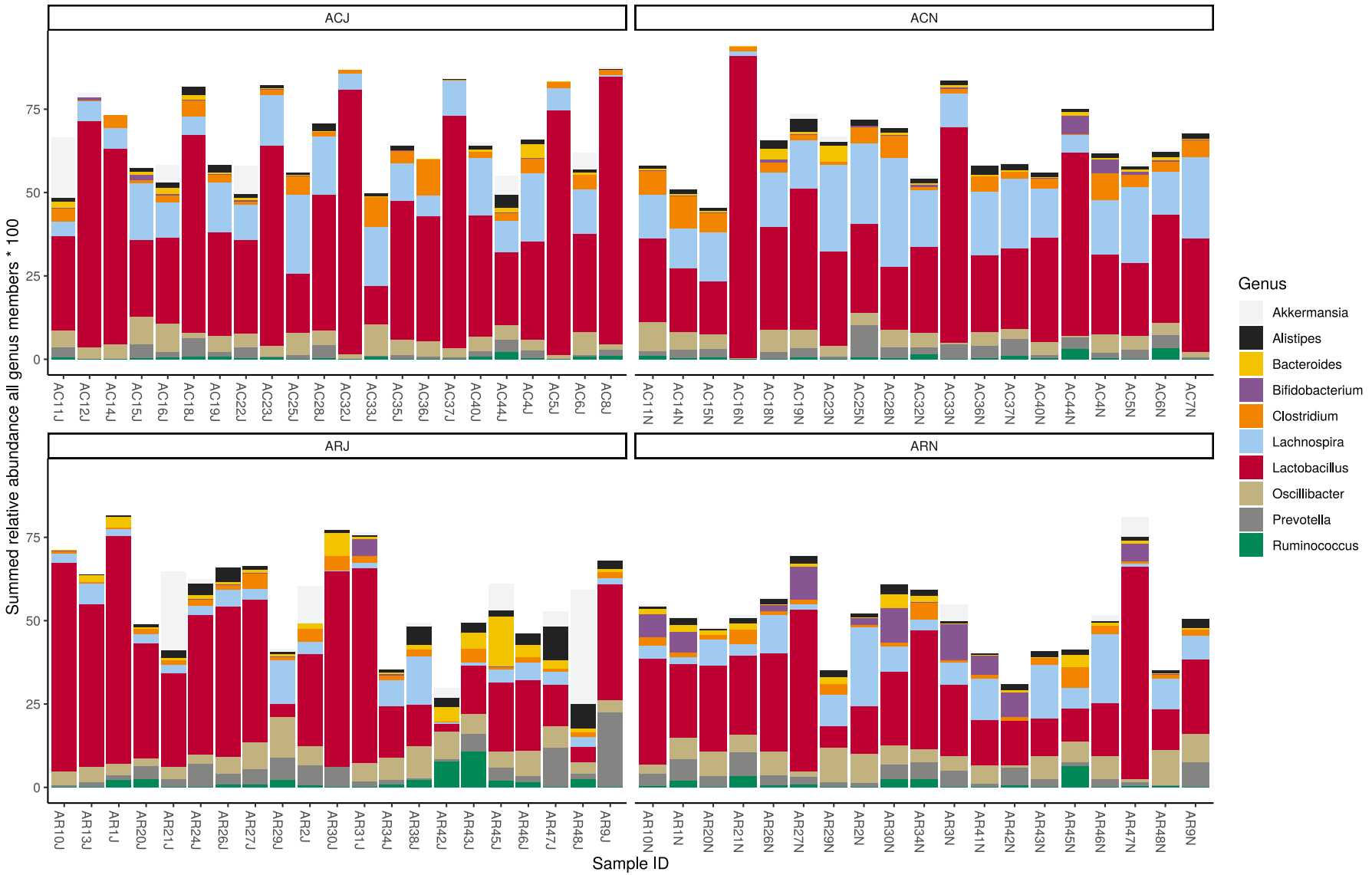

Figure 3.8: Stacked bar charts showing the relative abundances for the 10 genera which had the greatest summed relative abundance across all samples. Plot is facetted by host species and month and coloured by genus.

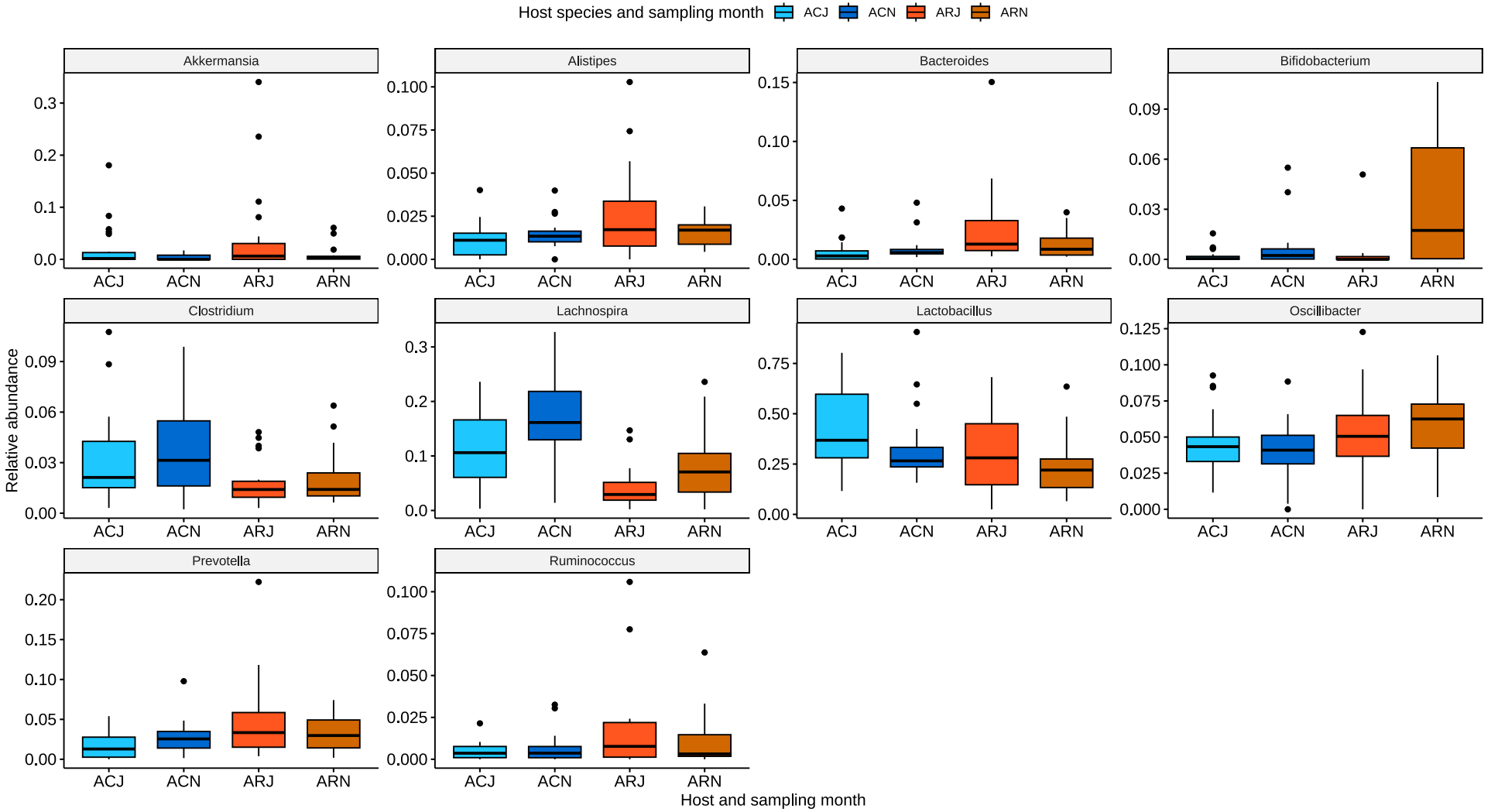

Figure 3.9: Boxplots showing the relative abundances for indicated genera, coloured by host species and sampling month.

Metaphlan 3

Metaphlan left the majority of reads unclassified, it never managed to classify more than 20% of the reads from each sample. The results from Metaphlan are just estimates due to the nature of how Metaphlan works (discussed below) but it still shows a limited ability to classify the reads from the sample. Also supporting this finding is the fact that processing the read files did not make a statistically significant difference in the estimated percentage of reads Metaphlan could classify. The geometric mean for the estimated percent of reads mapped to a known clade by Metaphlan was 2.53%, the median was 2.51% and the range was 0.11% to 19.94%.

Across all subsampled files, Metaphlan assigned reads to a total of 158 unique clades, ranging from the kingdom '`Bacteria`' down taxonomic levels to individual species. From all subsampled read files Metaphlan detected 8 unique phyla, 14 classes, 18 orders, 29 families, 32 genera and 55 species. Importantly, it also detected some reads outside Bacteria which it assigned within Eukaryota (a fungus specifically). The phyla did not evenly account for the estimated abundance provided by Metaphlan. The majority of reads in almost all *A. cahirinus* samples -which could be assigned by Metaphlan at all - were assigned within the Firmicutes phylum. In the samples from *A. russatus* there was a much greater range of relative abundances for the phylum, with the *A. russatus* samples having more reads classified by Metaphlan within the Actinobacteria, Bacteroidetes, Proteobacteria and Verrucomicrobia than the samples from *A. cahirinus*. The single non-bacterial phylum detected was detected in a sample from an *A. russatus* individual collected in June. The difference between the two *Acomys* species seen at the phylum level can also be observed in part at the genus level. Of the 32 genera detected, the vast majority of reads from *A. cahirinus* samples were assigned within the *Lactobacillus*, a large contingent of *A. russatus* samples have the majority of their relative abundance assigned within this genus as can be seen in **Figure 3.10**. A total of 6 *A. russatus* samples have at least 25% of their relative abundance assigned to *Akkermansia*, 9 to *Bacteroides* and 7 *Bifidobacterium*. It is also clear from looking at the genus level that a considerable proportion of the diversity at the genus level is from a very small number, or individual, samples of *A. russatus*. There were 12 genera amongst the 32 detected in which the only assigned reads by Metaphlan were found within a single sample from *A. russatus*; there were a further 3 genera which were found only within a single sample of *A. cahirinus*. The limited number of species detected by Metaphlan across all samples is likely a direct consequence of the low number of reads it was able to classify. For this reason and in light of the findings discussed in the previous chapter the author will not report any more results for Metaphlan analysis of our samples, the author will also not report the results for Metaphlan separated by sequencing technology, date and read depth.

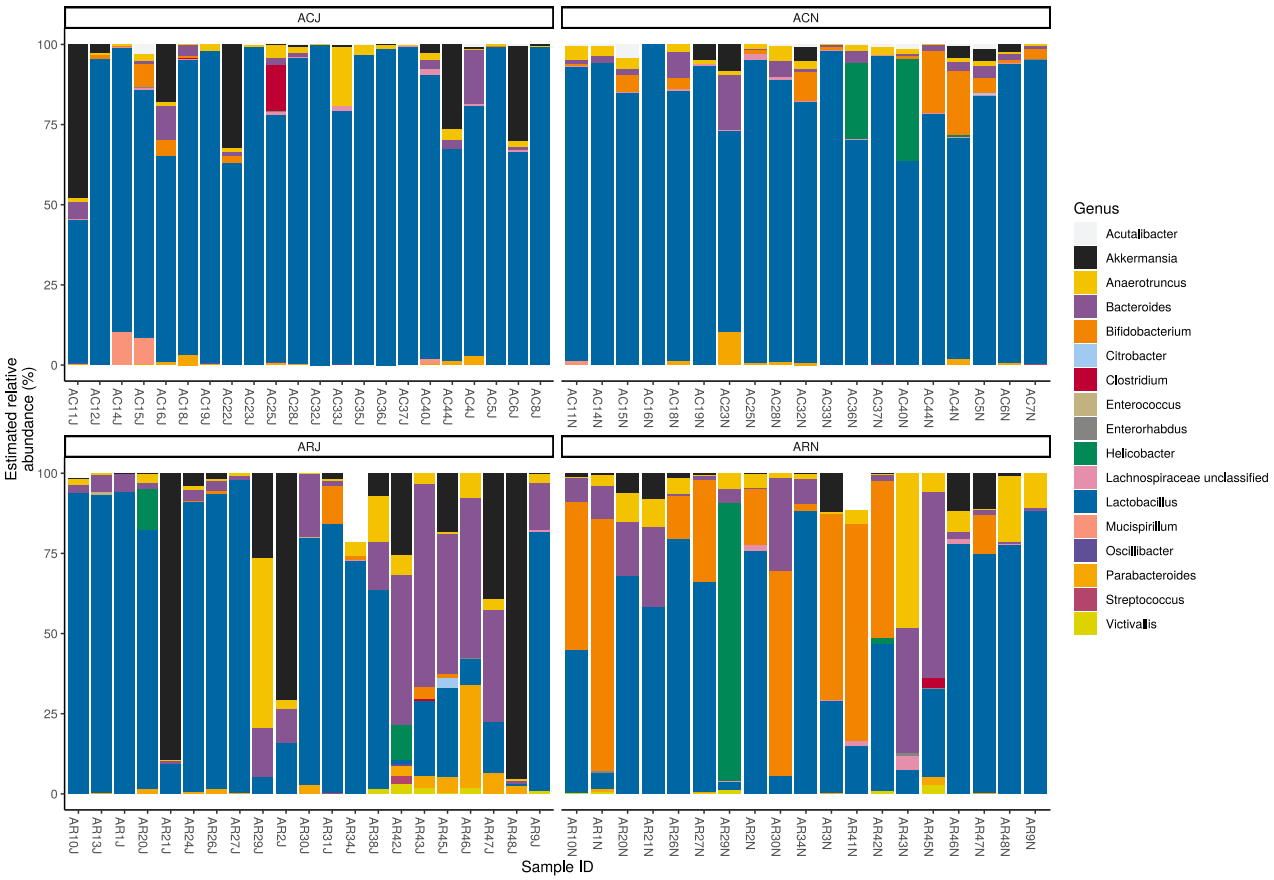

Figure 3.10: Stacked bar charts of the estimated relative abundances of genera detected by Metaphlan 3 from all processed read files faceted by host species and month of collection.

Kraken 2

The results for Kraken 2 were influenced in a large part by the differences in the percentage of reads classified by the different minimum confidence scores required, as seen in **Figure 3.6**. It must be noted however that the overwhelming majority (ranging from 89.9 to 99.5%) of reads in all samples were unclassified by Kraken 2 at a 50% minimum required confidence score; it detected greater diversity than Metaphlan but assigned only a small fraction of the reads to each taxonomic ID it detected. Kraken 2 detected 1,756 taxonomic IDs whereas with the reduced minimum required confidence score of 10% it classified reads to 7,198 taxonomic IDs.

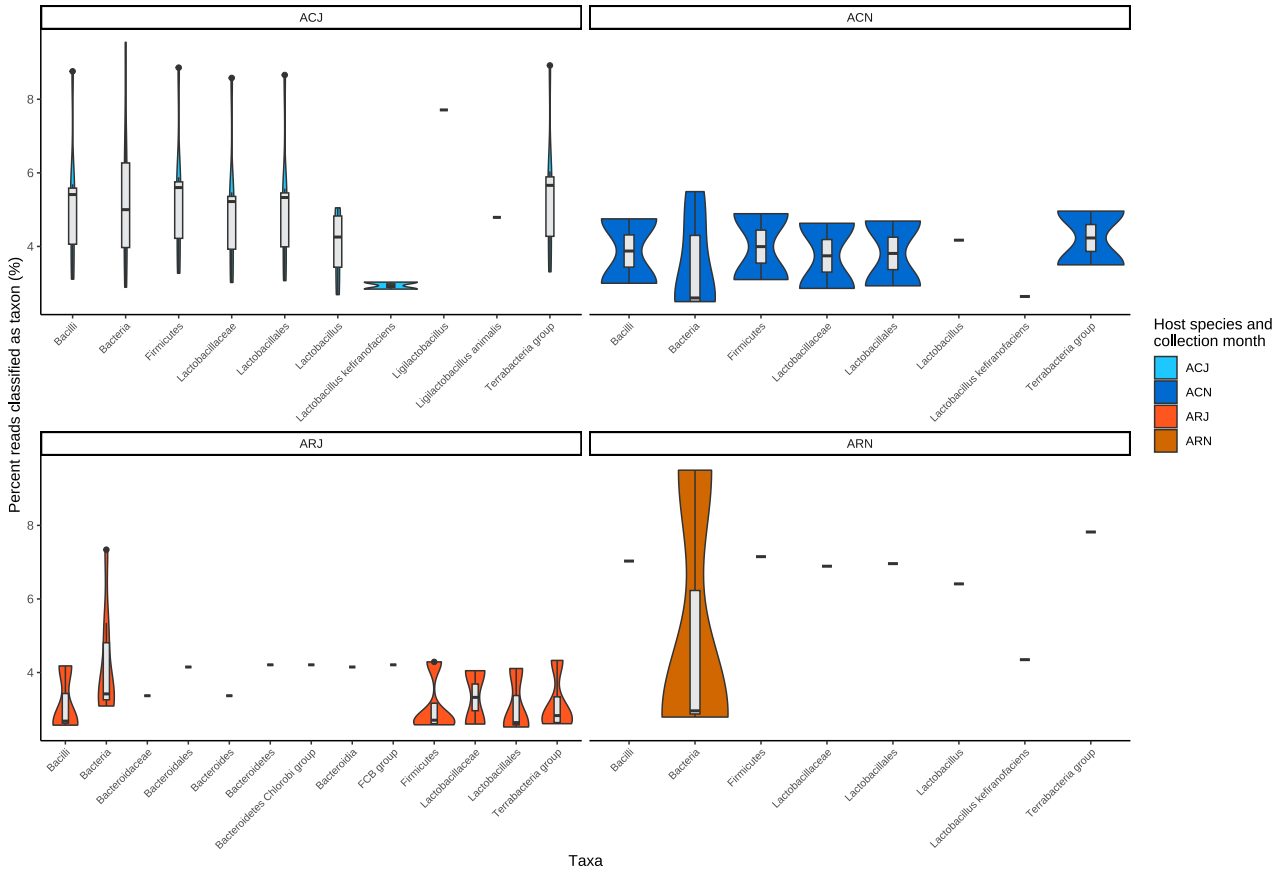

Figure 3.11: Box and violin plots showing the percentage of reads classified by Kraken 2 to each of the taxa which had at least 2.5% of reads classified to it in at least one sample for those read files subsampled to a read depth of 7,600,000. Faceted by host species and month of collection.

Across all taxonomic levels, there were only 10 taxa identified (aside from 'Root' and 'Unclassified') by Kraken 2 which had a geometric mean percentage of reads classified above 0 for all host species and sampling month combinations, when using a minimum required confidence score of 50%. Similar to the Metaphlan results - a single *A. russatus* sample contained diversity not seen in any other samples. This can be seen in **Figure 3.11** with the multiple taxa with a single value measured. Many of the taxa detected by Kraken 2 were from the same lineage, i.e. the species *Lactobacillus kefiranofaciens* and the higher level taxa it is found within. Also apparent in the results was a much lower level of diversity in the pilot sample reads, though a limited number of the detected taxa covered the majority of classified reads; similar to the more deeply sequenced samples.

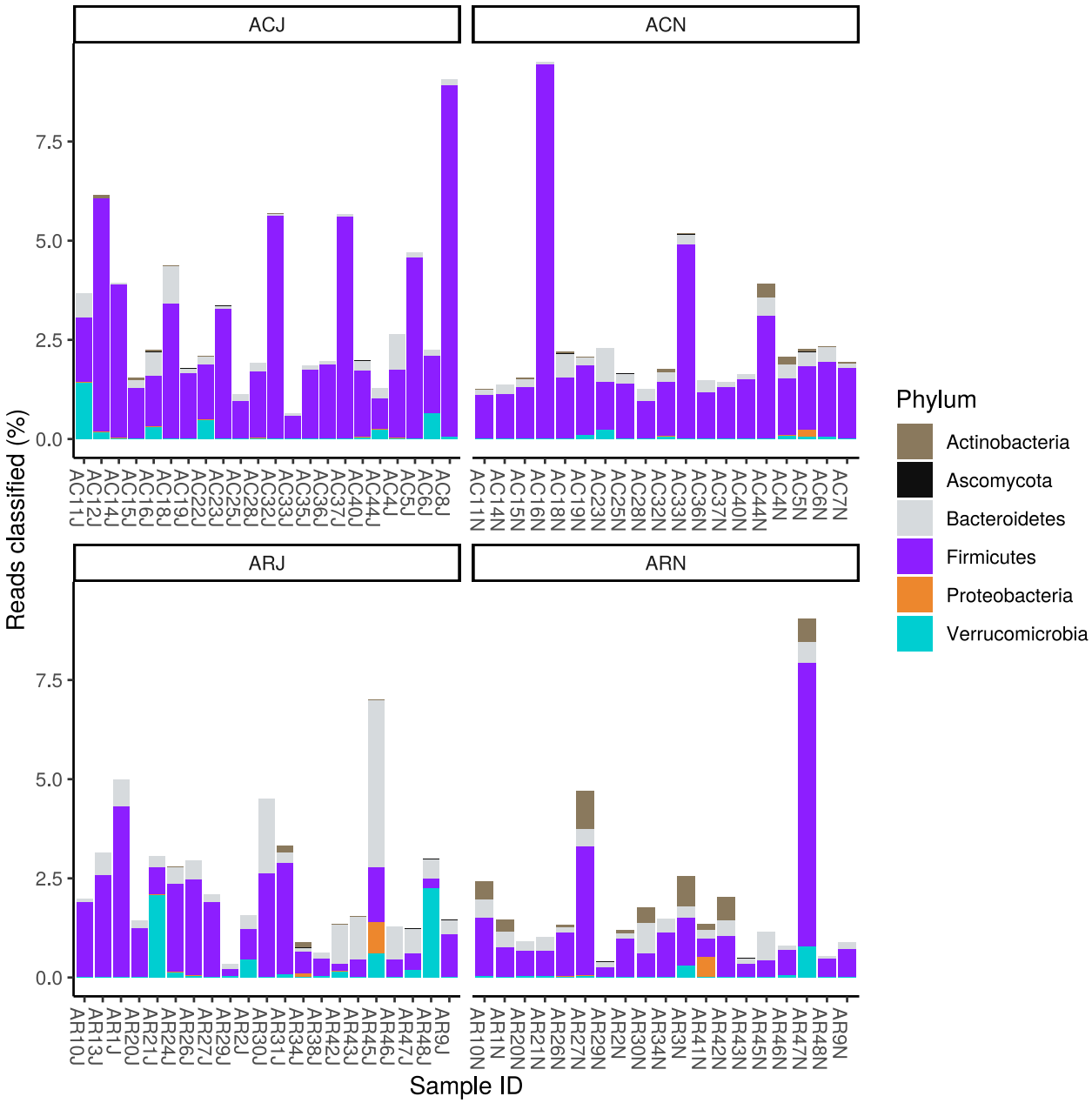

Figure 3.12: Stacked bar charts showing the percentage of reads classified in each sample for phyla detected by Kraken 2. Only those phyla which had a summed percentage of reads classified across all files of $\geq 0.01\%$ are shown.

The low level of reads classified held across both species and across both sampling months when looking only at those samples sequenced more recently and to a greater depth. Kraken 2 found a large percentage of classified reads within the Bacilli, which were distributed across different taxonomic levels. 22 phyla and 345 different genera were detected by Kraken across all processed read files. Only six phyla had a summed percentage of reads classified - summed values for all processed files - of greater than or equal to 0.01%, the percentages for these phyla can be seen in **Figure 3.12**. There were 41 genera which had a summed percentage of reads classified of greater than or equal to 0.01%, the percentages of the ten most abundant of these genera can be seen in **Figure 3.13**. *Lactobacillus*, *Ligilactobacillus*, *Bacteroides*, *Akkermansia* and *ParaBacteroides* were overall the most abundant genera detected by Kraken 2 across all samples. The range of different values for the percentage of reads classified by Kraken 2 into the 20 most abundant - as measured by summed percentage of reads classified from all samples -

genera by host species and collection month can be seen in **Supplemental Figure 5.3**.

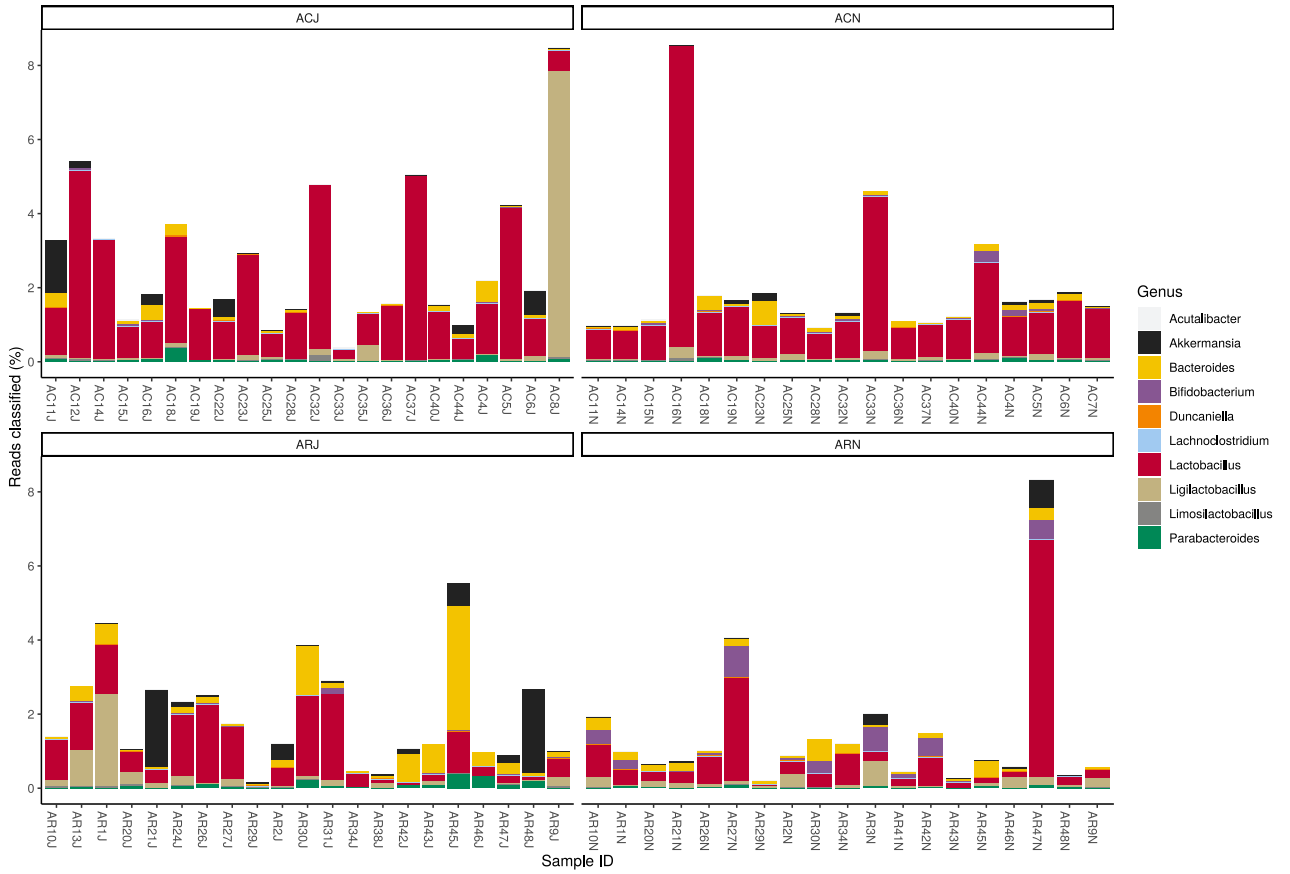

Figure 3.13: Stacked bar charts showing the relative abundance in each sample for genera detected by Kraken 2. Only the ten most abundant (by summed percentage reads classified) of those genera which had a summed percentage of reads classified across all files of $\geq 0.01\%$ are shown.

For *A. cahirinus* samples there was an increase in the geometric mean of taxa classified into the Clostridia (Clostridia itself, the order Eubacteriales and the genus *Lachnospiraceae*) from June to November. There was a drop in value for reads classified within the Firmicutes, with the notable exception of an increase in reads classified as *Lactobacillus kefiranofaciens*. In the *A. russatus* samples there was an increase in the geometric mean percentage of reads classified within the Clostridia (Clostridia itself and Eubacteriales) with the change from June to November, while there was a decrease in the values for Firmicutes and Bacteroidia. The changes from June to November within each host species in those taxa which Kraken 2 assigned the majority of reads it detected to can be seen in **Figure 3.14**.

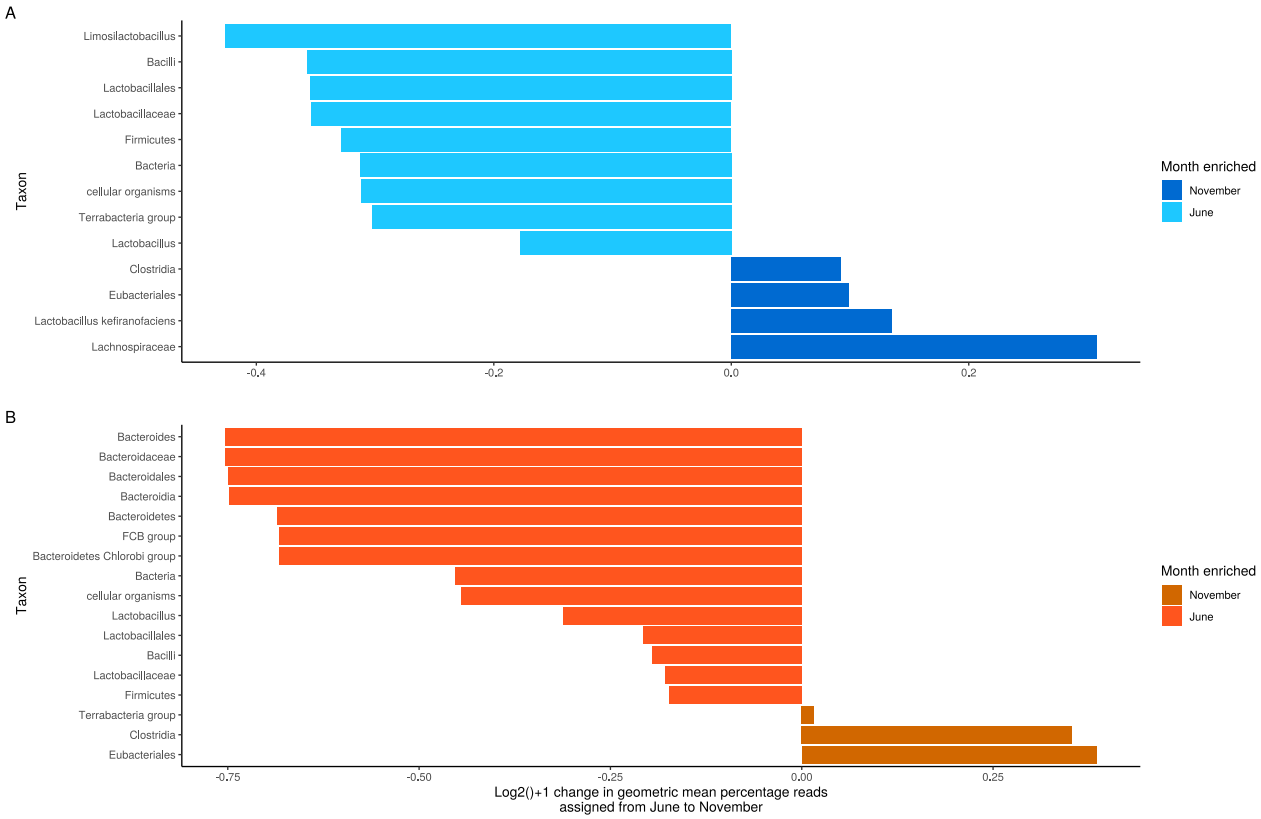

Figure 3.14: Bar charts of log2()+1 change in geometric mean percentage reads assigned to taxa with non-zero geometric mean values, from June to November. Showing only samples sequenced to a read depth of 7,600,000, coloured by sampling month. **A.** *Acomys cahirinus* samples, **A.** *Acomys russatus* samples.

Kaiju

The results discussed here are those obtained when running Kaiju with a maximum error allowance of 0. Relative to both Metaphlan and Kraken 2, Kaiju classified a significantly larger share of the reads in the subsampled files, as shown in **Figure 3.6C**. There were a large number of phyla detected but only 23 phyla had a summed - across all files - percentage reads classified to them of $\geq 1\%$, these 23 had the vast majority of the reads assigned to them. Bacteroidetes, Firmicutes and Proteobacteria were the most commonly classified phyla across both host species and sampling month combinations.

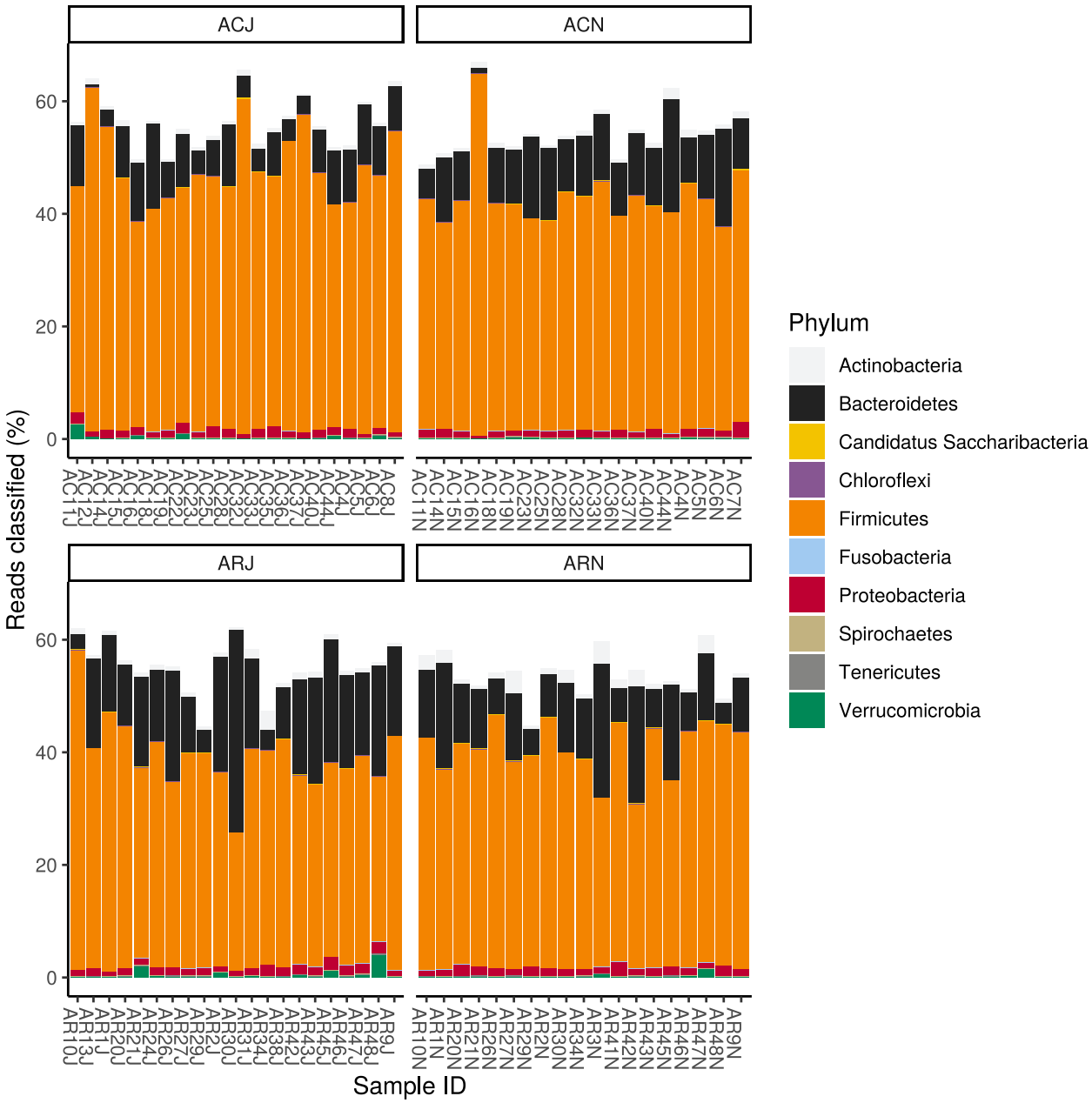

Figure 3.15: Stacked bar charts showing the percent of reads assigned by Kaju with an error allowance of 0 to the 10 most abudant phyla which had a summed percentage of reads classified to them of $\geq 1\%$ across all files, faceted by host species and sampling month.

Firmicutes was assigned the most of all classified reads of the three followed by Bacteroidetes and then Proteobacteria. These three phyla had geometric mean percentages of reads classified >1% across all host species and sample month combinations. The next most commonly classified phyla account for a smaller percentage of reads across the samples and may be false positives generated by Kaiju; especially those phyla detected in a small number of samples. **Figure 3.15** shows the percentage of reads classified to the ten most overall abundant - by summed percentage reads classified - phyla from all processed files.

As Kaiju had much greater success in classifying reads at all, as compared to Metaphlan and Kraken 2, it was possible to investigate the samples to a lower taxonomic level even at the strictest error allowance setting. There was a drop off in the percentage of reads classified with increasing detailed taxonomy, e.g. from phylum to class or family to genus. Kaiju identified a total

of 209 phyla assigned at least one read from across all files analysed. At the genus level, Kaiju identified 4,569 genera with at least one read classified within it across all subsampled read files. Of these 4,569 genera detected, 2,165 had a geometric mean percentage of reads classified above 0 for all four host species and sampling month combinations. Despite the large number of genera detected with a non-zero geometric mean percentage of reads assigned to them by Kaiju, the percentages of reads which were classified into any given genus was typically very low - less than 1% for almost all genera; a pattern which held true across both host species and sampling months. Only 154 genera had at least 1% of the reads from a sample classified to them. The most commonly classified genus detected by Kaiju was *Lactobacillus*, made clear in **Figure 3.16** which shows the percentage of reads classified to the ten most overall abundant - by summed percentage reads classified - genera from all processed files. The range of percentage of reads classified to these twenty most abundant genera can be seen in **Supplemental Figure 5.4**.

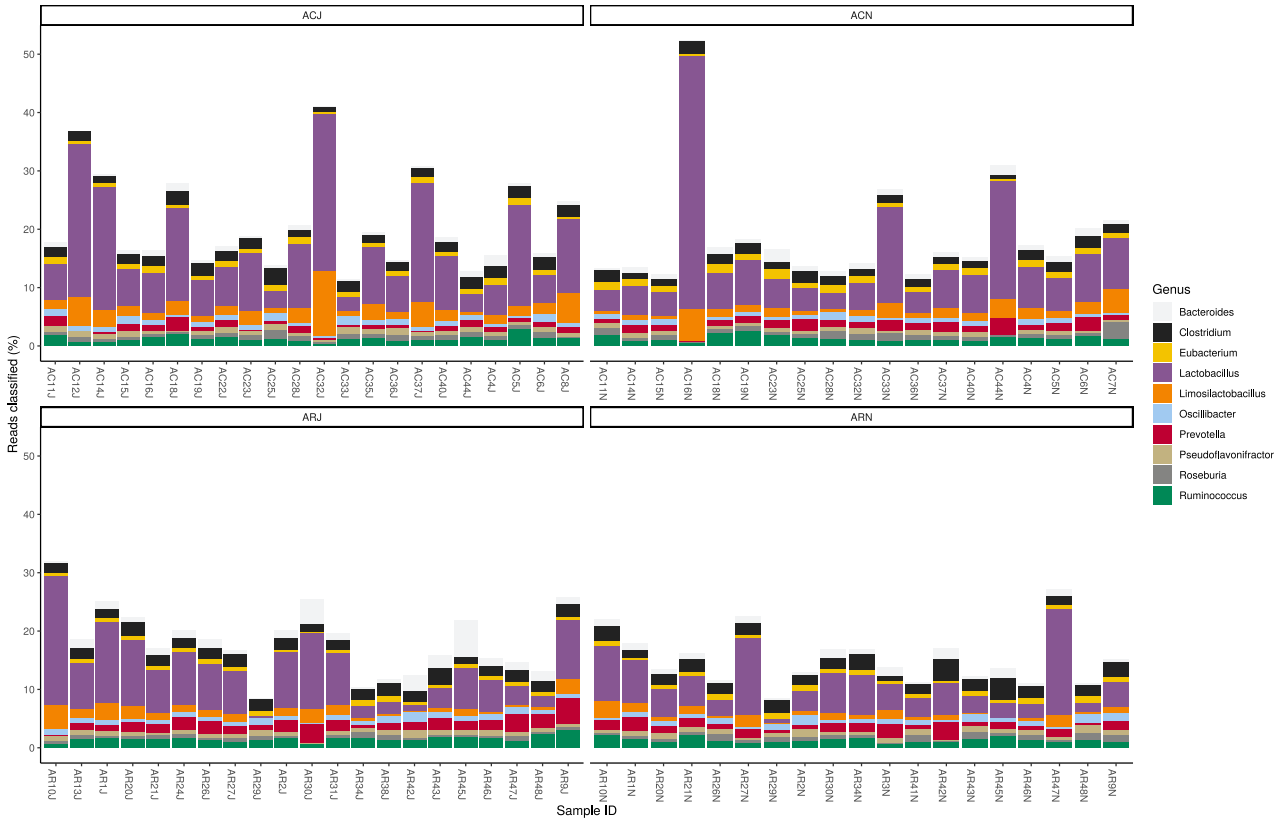

Figure 3.16: Stacked bar charts showing the percent of reads assigned by Kaju with an error allowance of 0 to the 10 most abudant genera which had a summed percentage of reads classified to them of $\geq 1\%$ across all files, faceted by host species and sampling month.

There were some genera which exhibited a change in the geometric mean values within each host species with the change in sampling month of note, which can be seen in **Figure 3.17**. A number of genera exhibited a similar change in geometric mean percentage reads classified from June to November across both host species; including *Helicobacter*, *Akkermansia* and *Thiopseudomonas*.

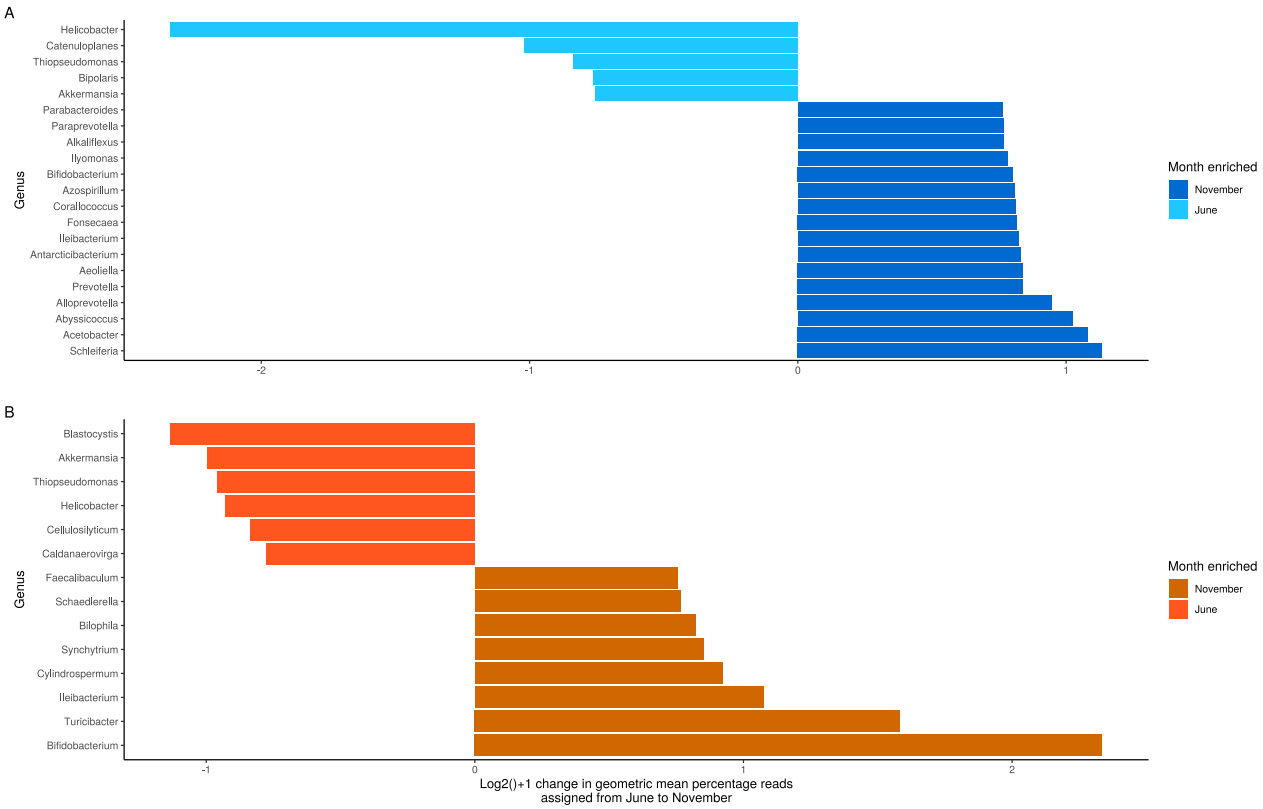

Figure 3.17: Bar charts of log2+1 fold change of geometric mean percentage reads classified for those genera which had a geometric mean percentage read classified above 0 in all host species and sampling month combinations. Showing taxa where the change was larger than +/-0.75. **A.** *A. cahirinus* read files. **B.** *A. russatus* read files.

3.3 Production of metagenomic bins

Due to the low level of classification of results by Metaphlan and Kraken 2, along with the previously observed high rate of false positives from Kaiju, the author elected to produce binned metagenomes from the read files and map our subsampled reads to them directly. These bins were produced by **Professor Chris Quince** and **Dr. Sebastien Raguideau** using the method described earlier in **Section 2.4**.

3.3.1 Bin file processing

In order to determine if the any bin files were identical or extremely similar to each other they were analysed with FastANI to obtain ANI similarity scores. FastANI comparison of all bins to each other did not yield any bins with an ANI similarity score of >84.8% for any comparison, below the 99% threshold suggesting two identical bins had been produced; therefore no bins were removed for being identical. There were 369 ANI estimates which were above 70% but below 84.8%, noting that some bins had multiple hits and some bins did not have any. Looking only at bins which had any ANI estimate hits above 70%, 165 of the 348 bins had at least one ANI estimate above 70% when compared with all other bins. dRep was also run to accomplish the same task, determine if any bins were very similar to each other (95% in this case) and in the event that any it would to pick a best representative of the groups of highly similar bins identified. It did not find any bins matching the threshold of similarity to each other and therefore did not pick any best representatives; as there were no groups identified.

Quality control

In order to determine whether each bin should be used in the mapping, the author carried out some quality checks of the bins. Quality control was a step in the pipeline used to produce the bins however the author used a stricter minimum required completeness of 80% and allowed a maximum contamination of 5%, both measured by CheckM. 348 bins met these criteria and were included for use in later analysis and in the mapping of subsampled reads. **Figure 3.18** shows the distribution of completeness and contamination values for all bins from CheckM analysis.

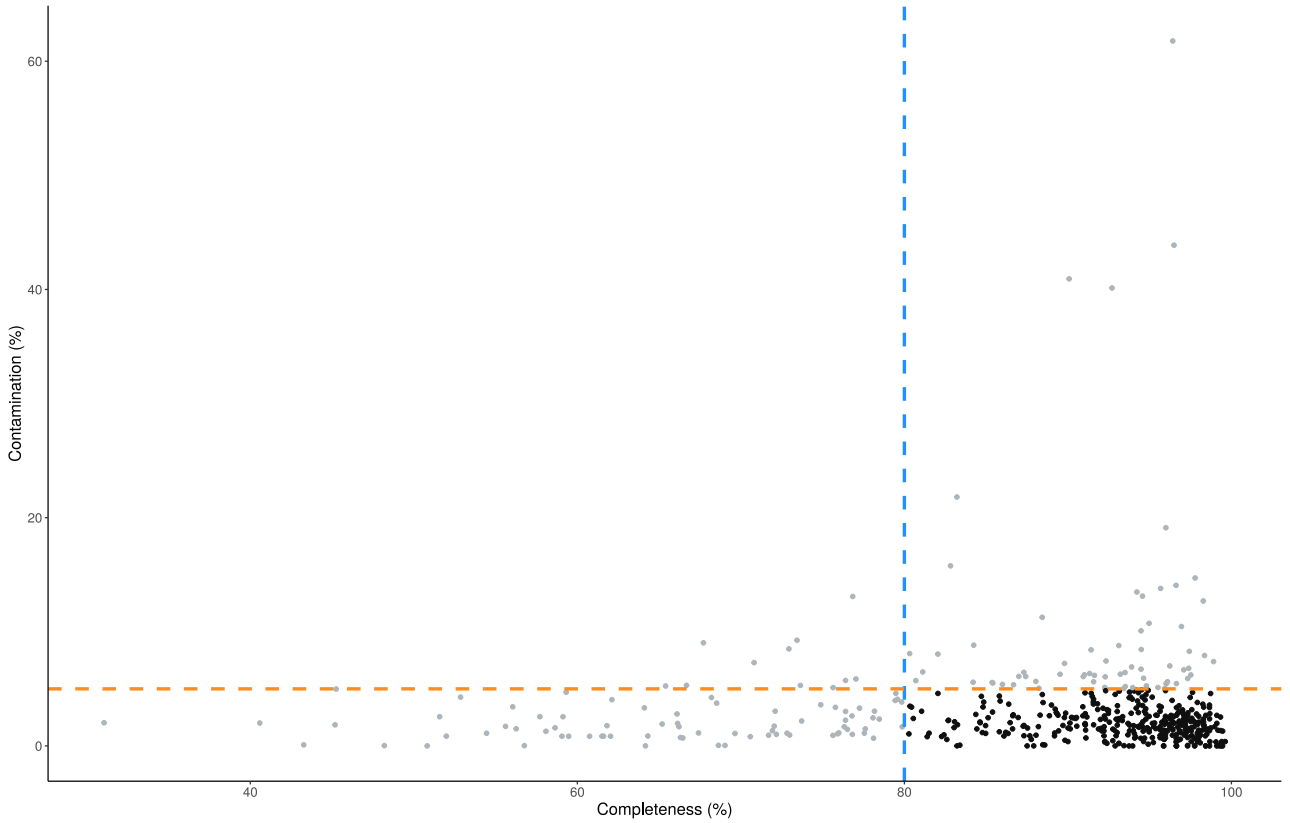

Figure 3.18: Plot showing the values for Completeness (%) and Contamination (%) for all metagenomic bin fasta files from CheckM analysis. The orange line shows the threshold cut-off value for contamination and the blue line the threshold cutoff value for completeness. Point colour indicates those bins which fail or pass either of the cutoff values.

rRNA detection with Barrnap

Barrnap was run to detect any ribosomal RNA genes in the 348 bins which passed the quality control step. 12 bins had an entire 16S rRNA gene detected, 124 bins had an entire 5S rRNA gene detected and 2 bins had an entire 23S rRNA gene detected; note that bins could have multiple rRNA genes detected within them. 90 bins had a partial 16S rRNA gene detected, 42 had a partial 5S rRNA gene detected and 82 had a partial 23S rRNA gene detected; note that some bins had detections of both the partial and complete versions of the same rRNA genes. 31 bins did not yield any Barrnap output and 87 bins which did provide Barrnap output did not have any detections of rRNA genes whether complete or partial. This lack of complete 16S rRNA gene sequences in particular and the lack of any complete or eveb partial rRNA genes which could be found in all bins prevented the use of rRNA seqeuence alignment for phylogenetic analysis [439] of the bins.

3.4 Taxonomic identification of metagenomic bins

It was necessary to attempt to understand the relationships of the bins to each other and gain an idea of what their taxonomic identity might be in order to understand the results of the mapping of the reads to the bins. GTDB-Tk was run on the bin fasta files to attempt to identify potential taxonomic IDs for each bin. The results varied, with some bins having a likely identity down to the genus level. **Table 3.2** shows the number of bins assigned to each phylum detected by GTDB-Tk after analysis of the 348 bins.

Phylum	Number of bins classified
Actinobacteria	9
Bacteroidota	103
Campylobacterota	3
Cyanobacteria	1
Desulfobacterota	12
Elusimicrobiota	1
Firmicutes	19
Firmicutes_A	193
Proteobacteria	6
Spirochaetota	1

Table 3.2: Number of bins classified into each of the listed phyla by GTDB-Tk after analysis of the 348 bins meeting the minimum completeness and maximum contamination thresholds. Phylum name is taken directly from GTDB-Tk output.

At the family level, GTDB-Tk identified 38 different families when attempting to classify the 348 bins. Five of these families accounted for 71% of all bins, in descending order of bins classified these were Lachnospiraceae, Muribaculaceae, Ruminococcaceae, Acutalibacteraceae and Desulfovibrionaceae. Lachnospiraceae and Muribaculaceae each had 104 and 80 bins classified within them respectively thus suggesting that just over 50% of all the bins may be representative of members of these families. All bins had an identified family but this was the lowest taxonomic level GTDB-Tk was able to assign to all bins, with 41 bins lacking an identified genus and 335 lacking an identified species. Of the 307 bins with an identified genus, **Figure 3.19** shows the 7 most commonly classified genera and the number of bins classified within them by GTDB-Tk; the overwhelming majority of bins are not classified into one of these seven genera.

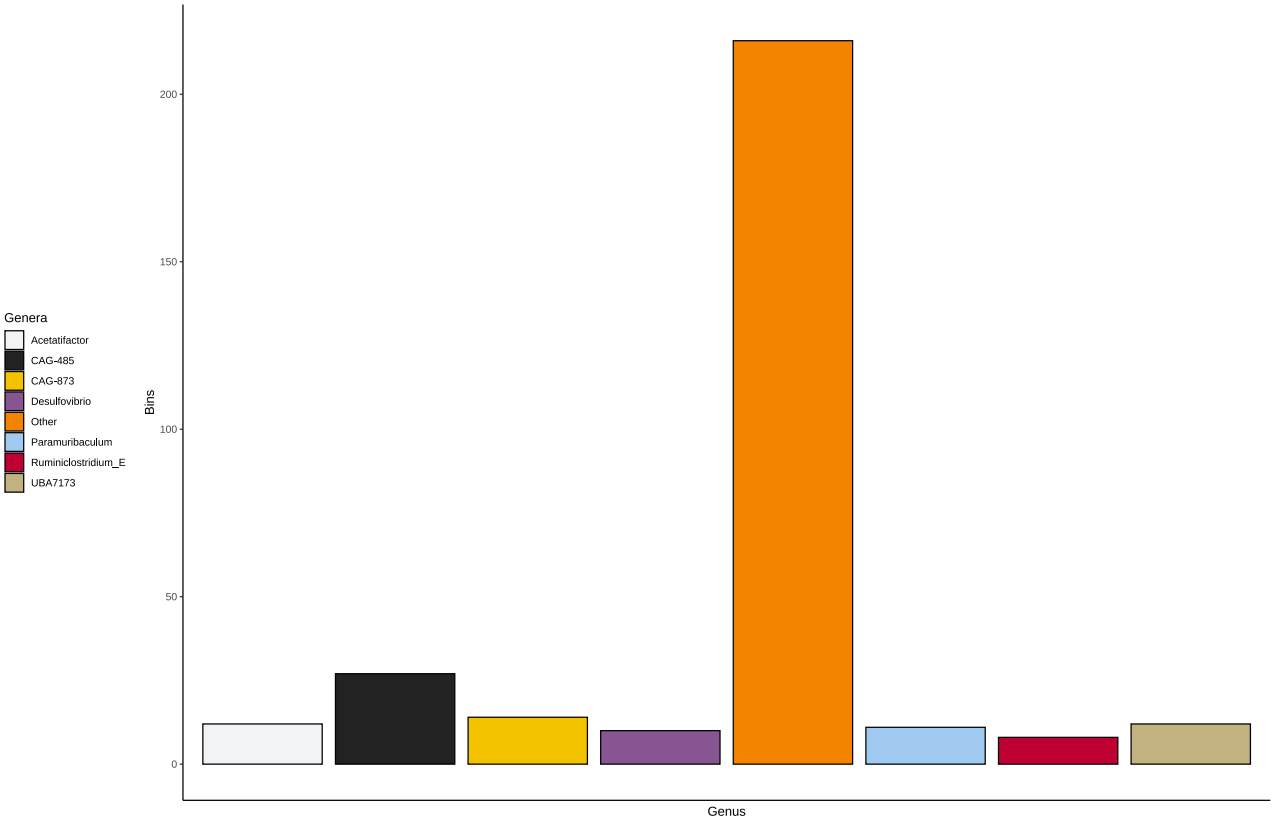

Figure 3.19: Barplot showing the number of bins classified into each of the indicated genera by GTDB-Tk after analysis of the 348 bins, bars coloured by genus. Shown are the 7 most commonly classified genera with remaining genera combined into 'Other'. Genus name is taken directly from GTDB-Tk output.

3.5 Phylogeny of metagenomic bins

As discussed in the methods the phylogenetic tree of the bins was created including the assemblies obtained from the LAB isolates from the *Acomys* and *Apodemus*. Also as discussed Cactus did not provide branch lengths for the tree, though SibeliaZ which was used for creating the tree of the isolate assemblies only does provide branch lengths it would take multiple months to produce a tree of either the bins alone or one including both the bins and the assemblies using the tool. The tree can be seen in **Figure 3.20**.

The tree clearly shows that the greatest degree of separation was between the bins and assemblies overall. Looking at the bins by host enrichment status there is no clear partioning across the tree between those enriched in either host species, nor do those not enriched in either cluster together away from the enriched. Broadly the assemblies form a subtree on their own with only two of the assemblies, 18A_S9 and 39B_S15 not included in it. There are a handful of bins within the assemblies subtree, 6 to be exact. 4 of these are enriched in *Acomys cahirinus* and 2 are not enriched in either host species. The remaining bins fall into two very large and quite wide subtrees, the first occupying the tree from around 11 o'clock (imaging the circular tree as a clock face) at Bin_c801 through to Bin_c82 at around 3 - 4 o'clock. The other subtree is wider and covers the remainder of the tree from Bin_c1492 through to Bin_c513. The single biggest cluster of bins of a similar enrichment status is one of bins not enriched in either host species found at around 12 o'clock on the tree running from Bin_c712 to Bin_m334 - with very closely related branches of non-enriched bins continuing further without interruption until Bin_c354. With few exceptions bins enriched in one host are not found within clusters of bins enriched in the other.

The same tree but this time with the five most commonly classified families for the bins can be seen in **Figure 3.21**. In this instance there is very clear clustering and partioning of the tree by taxonomic family. The Lachnospiraceae and Desulfovibrionaceae are entirely constrained to two subtrees - the Lachnospiraceae being the larger as it was the classification assigned to many more bins. The Ruminococcaceae mostly group together into a subtree with two outliers found instead with the Acutalibacteraceae. The latter is split three ways though all three clusters are found on the same larger subtree branching off from the Lachnospiraceae and along with the Ruminococcaceae. The gap between the two Muribaculaceae subtrees is sizeable though again both are found on the same originating branch from the root and away from the other most abundant families.

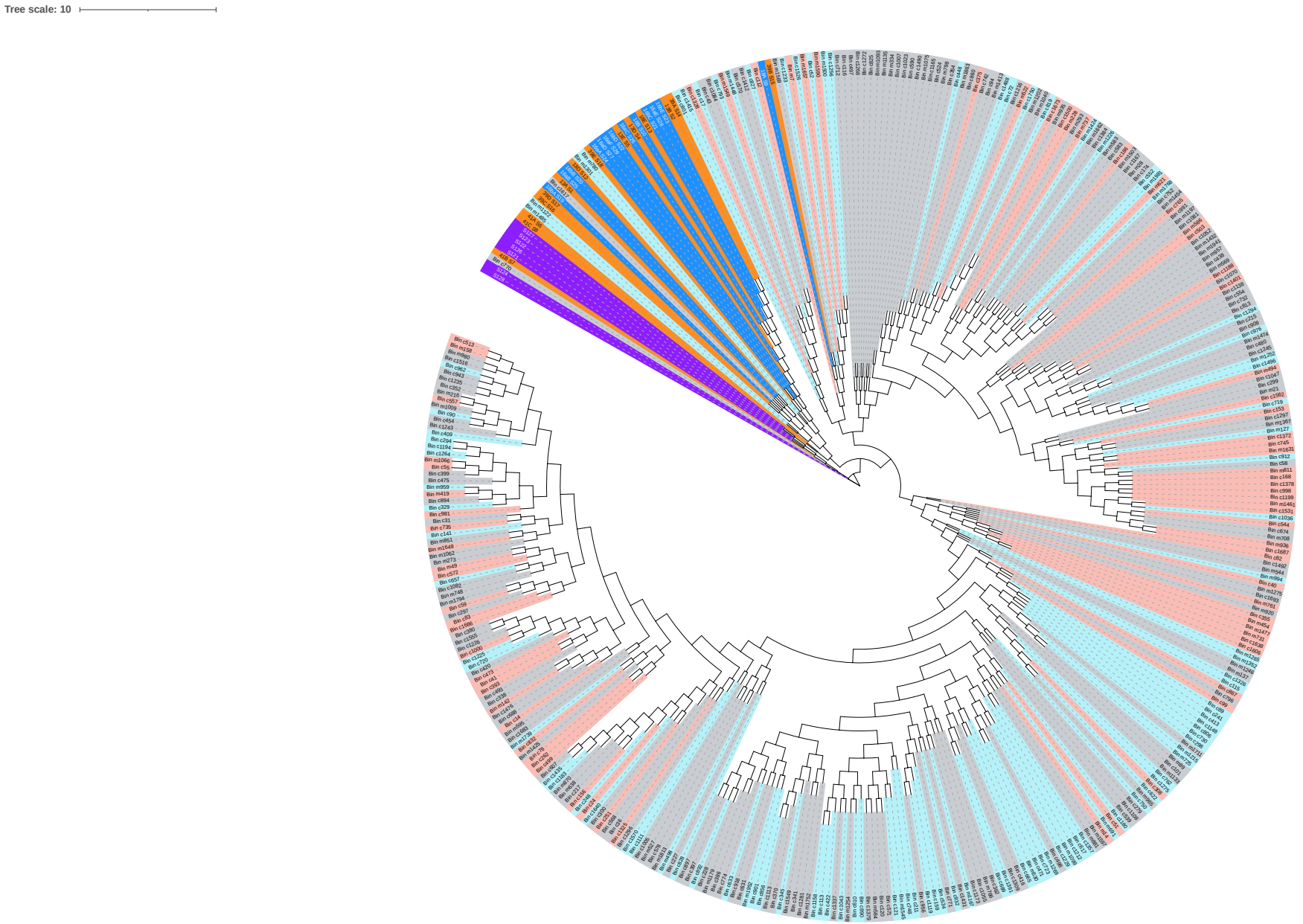

Figure 3.20: Phylogenetic tree of the 348 metagenomic bins and the LAB isolates from the *Acomys* and *Apodemus*, aligned using Cactus and resulting tree visualised using ITOL. Purple tips are assemblies from *Apodemus*, blue assemblies from *Acomys cahirinus*, orange from *Acomys russatus*, grey are bins not enriched in either species, light blue are bins enriched in *Acomys cahirinus* and pink are bins enriched in *Acomys russatus*

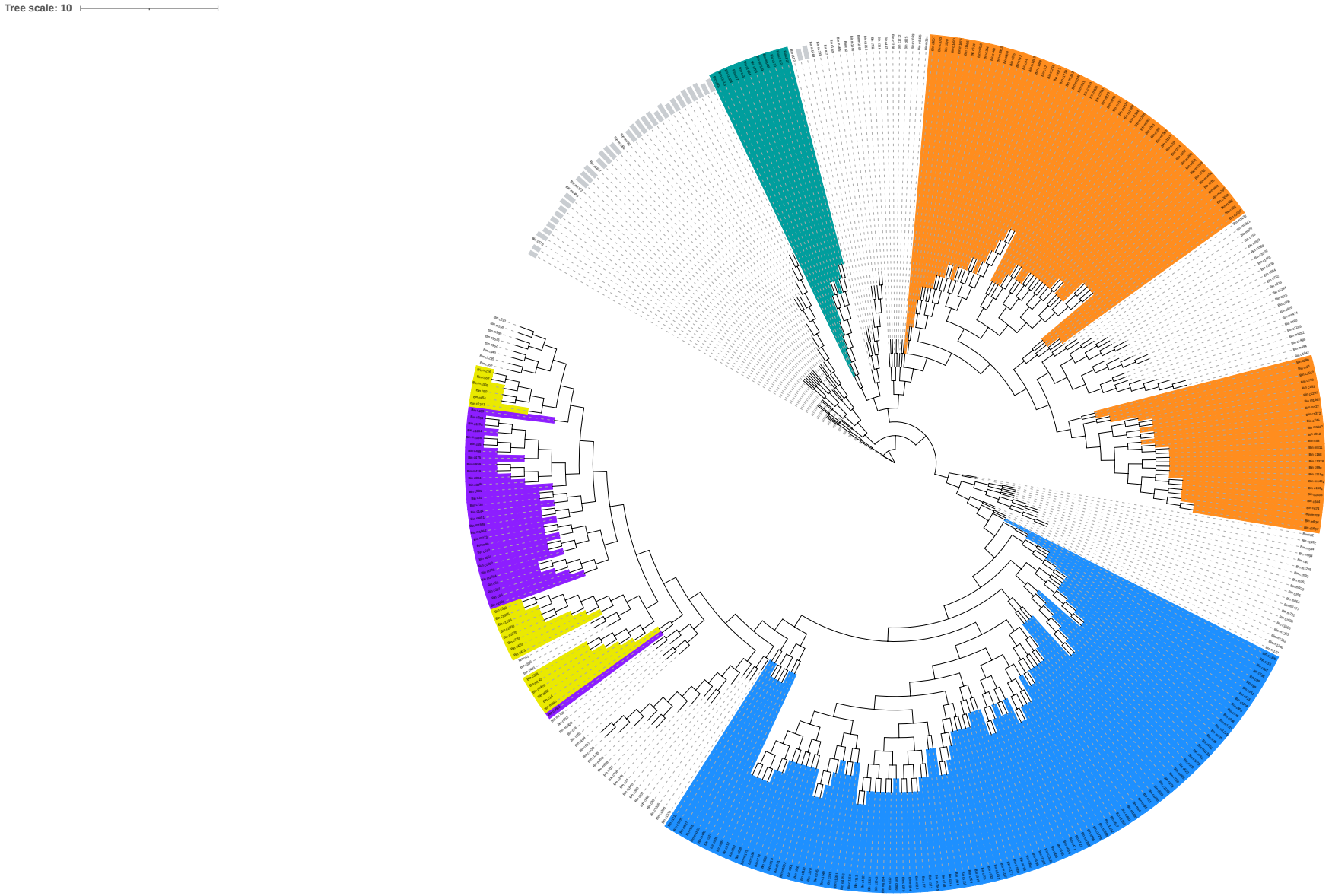

Figure 3.21: Phylogenetic tree of the 348 metagenomic bins and the LAB isolates from the *Acomys* and *Apodemus*, aligned using Cactus and resulting tree visualised using ITOL. Coloured clades are the 5 most common bin GTDB-Tk family level classifications. Blue is Lachnospiraceae, orange is Muribaculaceae, purple is Ruminococcaceae, yellow is Acutalibacteraceae and green is Desulfovibrionaceae. LAB assemblies have been greyed out

3.6 Mapping of *Acomys* faecal sample shotgun reads to metagenomic bins

3.6.1 Mapping of subsampled reads to bins

The mapping of the subsampled reads to a single reference fasta file made by concatenating all the 348 'strict' bin fasta files (after turning these into single, large contig files) produced a range of mapping percentages. The lowest level of mapping was 10.33% from the pilot sample AC16N, the highest was 36.13% for the NovaSeq sequenced sample AR42N. The median percentage of mapped reads for the pilot samples was 22.28%, for NovaSeq sequenced samples it was 24.63%. The range in the pilot study sequenced samples was 10.33% to 29.41% and for the NovaSeq sequenced samples it was 11.70% to 36.13%. The arithmetic mean mapping percentage for the pilot samples was 21.95% and for Novaseq sequenced samples it was 24.76%.

A two sample t-test was carried out to check if the subsampled read depth of the files mapped to the bins reference influenced the percentage of reads mapped. Comparing the arithmetic mean percentage of reads mapped across each of the sequencing runs gave a t test statistic of 1.9136 and a p-value of 0.07495. This is not significant so the subsampled read depth does not cause the difference in the percentage reads mapped. **Supplemental Figure 5.5** shows the range of mapping percentages for all processed sample read files, split by the sequencing run in facets. However, due to concerns that the pilot samples may be influencing the understanding of the results the author elected to set aside those results from further analysis. In order to more thoroughly investigate a potential host or seasonal impact the author restricted all further analysis to those samples which were sequenced on using Novaseq platform and so were subsampled to a greater uniform depth and for which there had been paired monthly samples collected. This dataset consists of 58 samples, made of 15 June and November samples for *A. cahirinus* and 14 June and November samples for *A. russatus*.

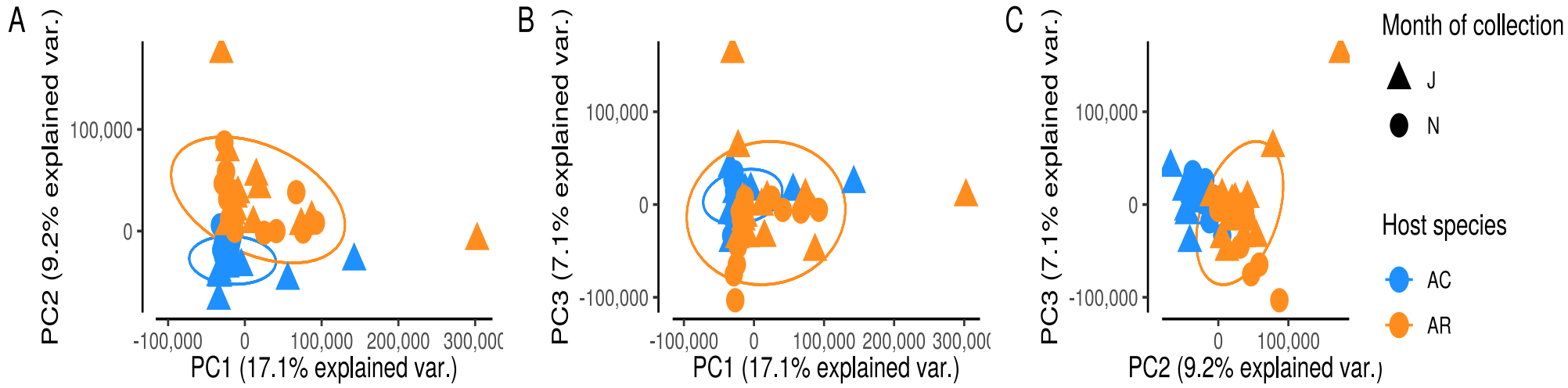

Figure 3.22: PCA plots of reads-per-million (RPMs) from mapping of reads to bins concatenated reference file. Colours distinguish host species and shape distinguishes sampling month. **A.** PCs 1 & 2, **B.** PCs 1 & 3, **C.** PCs 2 & 3.

After this subsetting of the mapping results a Principal Component Analysis (PCA) was carried out using the 'prcomp' command in R to visualise any potential clustering in the results from the host species, sampling month or both. PCs 1 through 5 contributed over 5% to the explanation of variance. Plots of PC1 vs PC2, PC1 vs PC3 and PC2 vs PC3 as seen in **Figure 3.22** show a degree of clustering by host species.

3.6.2 Enriched bins within and across host species

Enriched bins within host species

The PCA analysis of the subset of samples indicated that there is a host effect. To analyse this changes in the geometric means within the host species from June to November were compared. Looking at the samples from *A. cahirinus* there were 9 bins which were significantly different (BH corrected P-value ≤ 0.05) and had an absolute log2 fold change greater than +/-1 between the two sampling months. **Figure 3.23** shows these results. One bin, Bin_m334 (*Helicobacter*_D according to GTDB-Tk), had a geometric mean RPM value above zero for June samples only in *A. cahirinus*. Of the nine bins, 2 were more abundant in the June samples than they were in the November samples - Bin_m1135 (*Helicobacter*_D) and Bin_c1111 (CAG-632). There were no bins which were only present in the November sample; of the 7 bins which were enriched in the November samples 1 was classified as *Muribaculum*, 2 as CAG-873, 1 as *Eubacterium*_R, 1 as *Prevotellamassilia*, 1 as Muribaculacea and 1 as Rikenella.

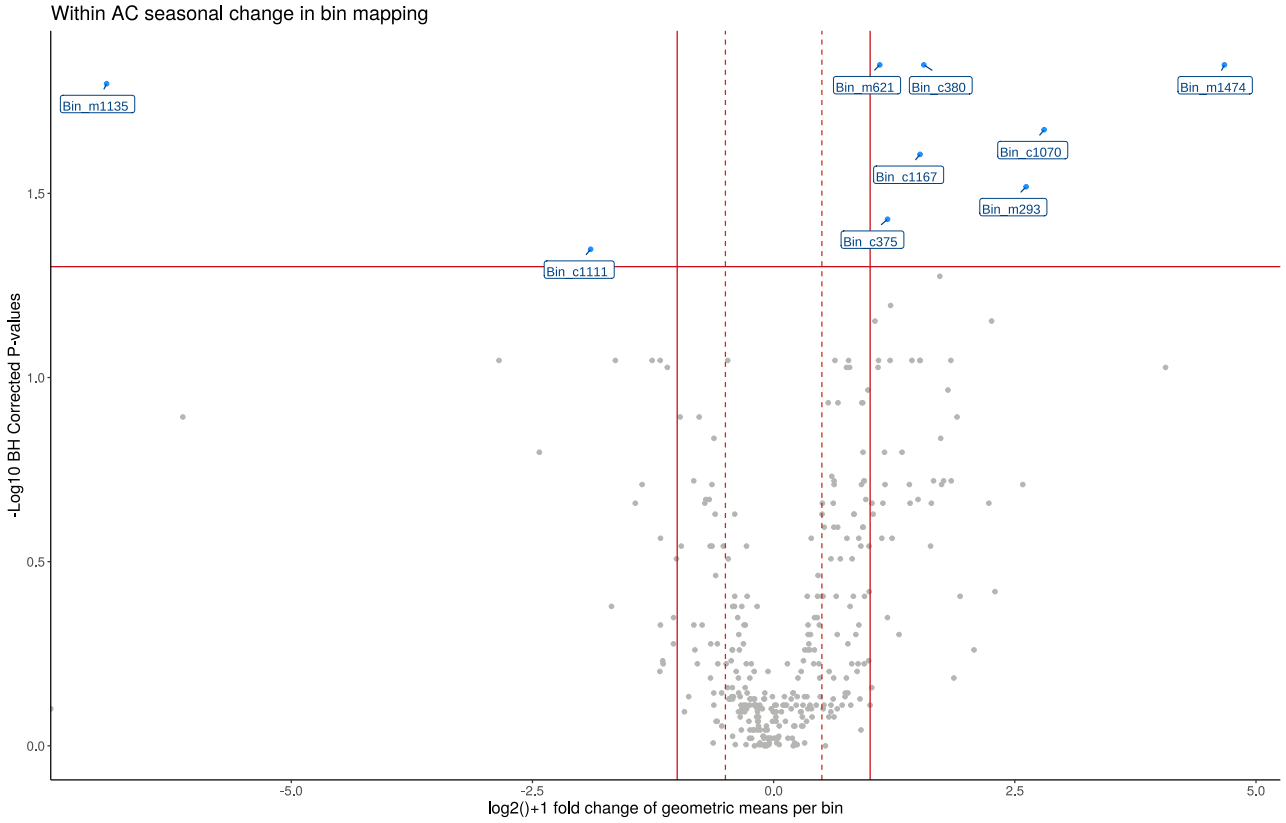

Figure 3.23: Volcano plot of RPMs from mapping of reads from *Acomys cahirinus* samples to bins reference. Solid red lines show p-value <0.05 and log2()+1 fold change of larger than +/-1. Dotted red lines show log2()+1 fold change of greater than +/-0.5. Point colour determined by RPM meeting thresholds for Q-value and fold change.

When looking at within *A. russatus* change from June to November, there was a much smaller number of bins which were differentially represented. The three detected to be differentially repre-

sented between June and November in *A. russatus* were Bin_m137 (*Cellulosilyticum*), Bin_c796 (CAG-590) and Bin_c1435 (CAG-552), as shown in **Figure 3.24**. They met the threshold for significance and magnitude of the fold change (using the same thresholds as with AC). Bin_m137 (*Cellulosilyticum*) was more abundant in the June samples, Bin_c796 (CAG-590) and Bin_c1435 (CAG-552) were more abundant in the November samples. There was also one bin which had a geometric mean RPM value of 0 in the November samples for *A. russatus*, Bin_m1135 (*Helicobacter*_D). Three bins had a geometric mean RPM of 0 in the June samples from *A. russatus*, these are Bin_c217 (Borkfalkiaceae), Bin_m1093 (*Helicobacter*_C) and Bin_c568 (Treponemataceae).

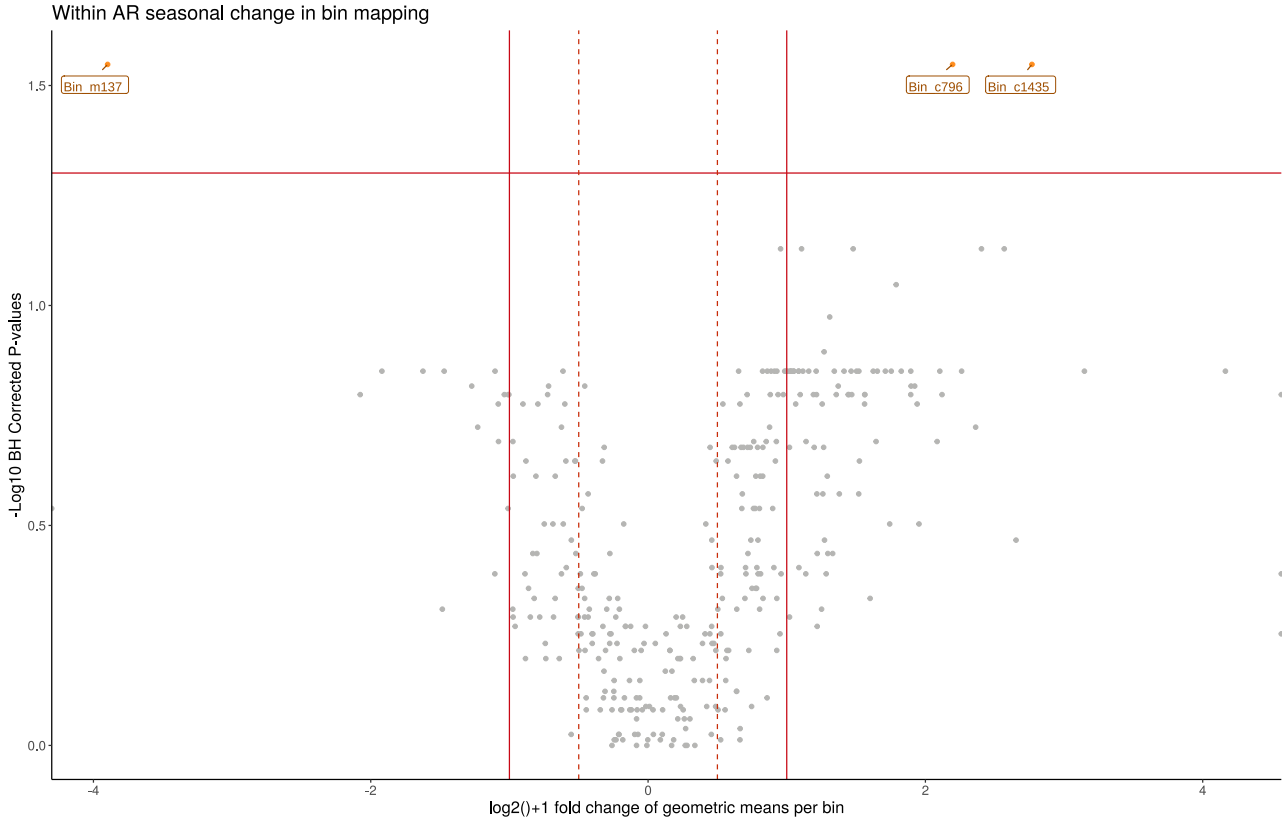

Figure 3.24: Volcano plot of RPMs from mapping of reads from *Acomys russatus* samples to bins reference. Solid red lines show p-value <0.05 and log2()+1 fold change of larger than +/-1. Dotted red lines show log2()+1 fold change of greater than +/-0.5. Point colour determined by RPM meeting thresholds for Q-value and fold change.

Enriched bins across host species and sampling months

The analyses were extended to investigate the entire dataset to compare composition variation between host species and sample months. A number of bins were identified which showed a significant host effect, 172 in total (excluding 4 with infinite fold change values), which can be seen in **Figure 3.25A**. No significant season effect was detected which can be observed in **Figure 3.25B**. A total of 66 bins were found to be significantly enriched in *A. russatus* samples (Q-value of at most 0.05 and log2+1 fold change of at least 1). A total of 106 bins were found to be enriched in the *A. cahirinus* samples (Q-value of at most 0.05 and log2+1 fold change of -1 or less). One bin had a zero value for the geometric mean RPM from all *A. cahirinus* samples for both months, Bin_m334 (*Helicobacter*_D). Four bins had a zero value for the geometric mean RPM from all *A. russatus* samples for both months, these are Bin_c217 (Borkfalkiaceae), Bin_c568 (Treponemataceae), Bin_1093 (*Helicobacter*_C) and Bin_m1135 (*Helicobacter*_D).

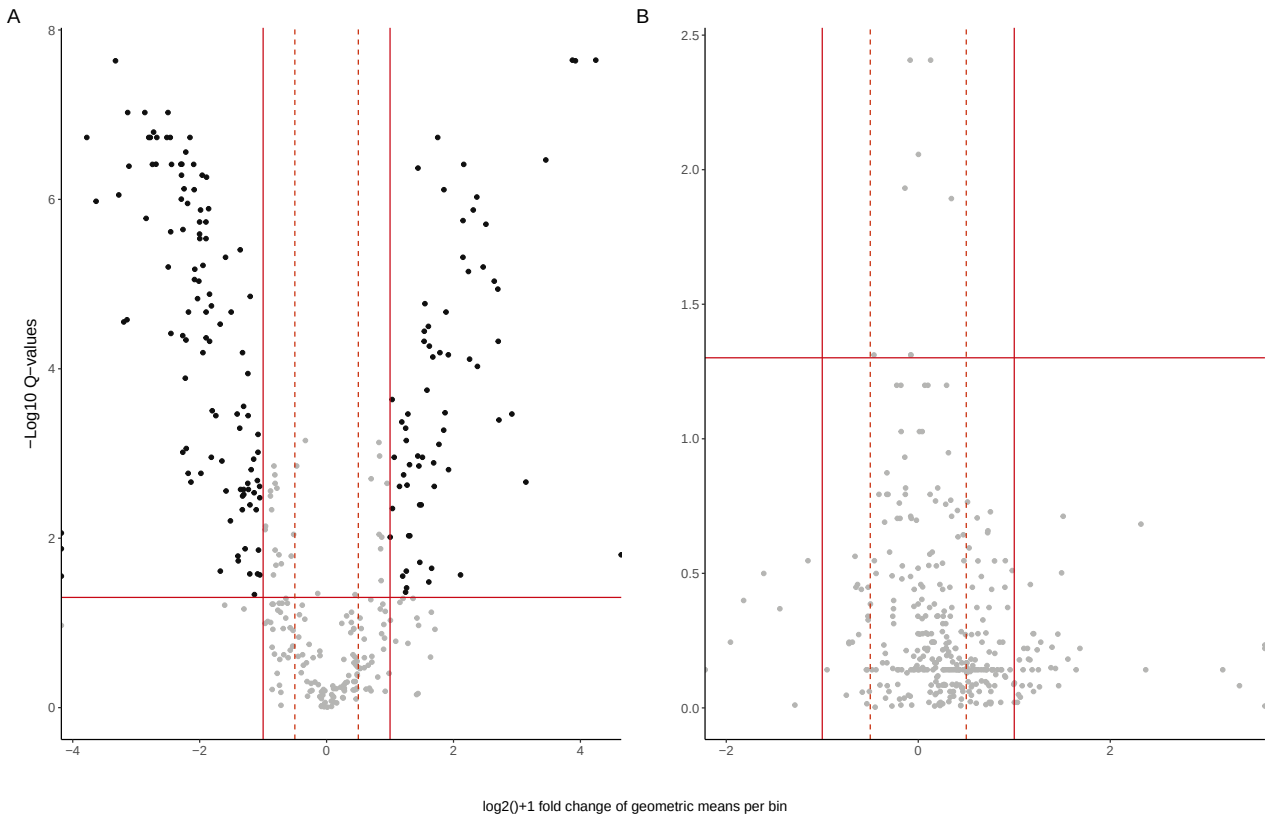

Figure 3.25: Volcano plots showing -log10 Q-values (BH corrected P-values) and log2()+1 fold change for geometric mean RPMs. Solid red lines show threshold values for Q-value (-log10 0.05) and fold change (larger than +/-1). Dotted red lines show fold change (greater than +/-0.5). Points coloured if passing thresholds. **A.** Host species effect results. **B.** Sampling month effect results.

The bins which were enriched in the *A. russatus* samples and met the significance threshold were classified into 19 families by GTDB-Tk. For the bins enriched in the *A. cahirinus* samples which met the significance threshold, GTDB-Tk classified them into 17 families - with 17 bins found within these families. A total of 8 of the families assigned to bins enriched in the *A. russatus* samples were not amongst those assigned to bins enriched in the *A. cahirinus* samples; 6 families were only assigned to bins enriched in the *A. cahirinus* samples and not the bins enriched in the *A. russatus* samples - with 10 bins found within these families (**Figure 3.26**). There were only 6 families which had bins which were statistically significantly different between the two host species but did not meet or exceed the magnitude threshold for the fold change (either positive or negative).

The range of RPMs for the ten most differentially abundant bins in each host species can be seen in **Supplemental Figure 5.6** for *A. cahirinus* and **Supplemental Figure 5.7** for *A. russatus*. Some bins had generally high RPMs in the samples from one host species and low values in the other, other bins had a range of RPM values for both host species but still clustered toward the higher or lower end of the scale .

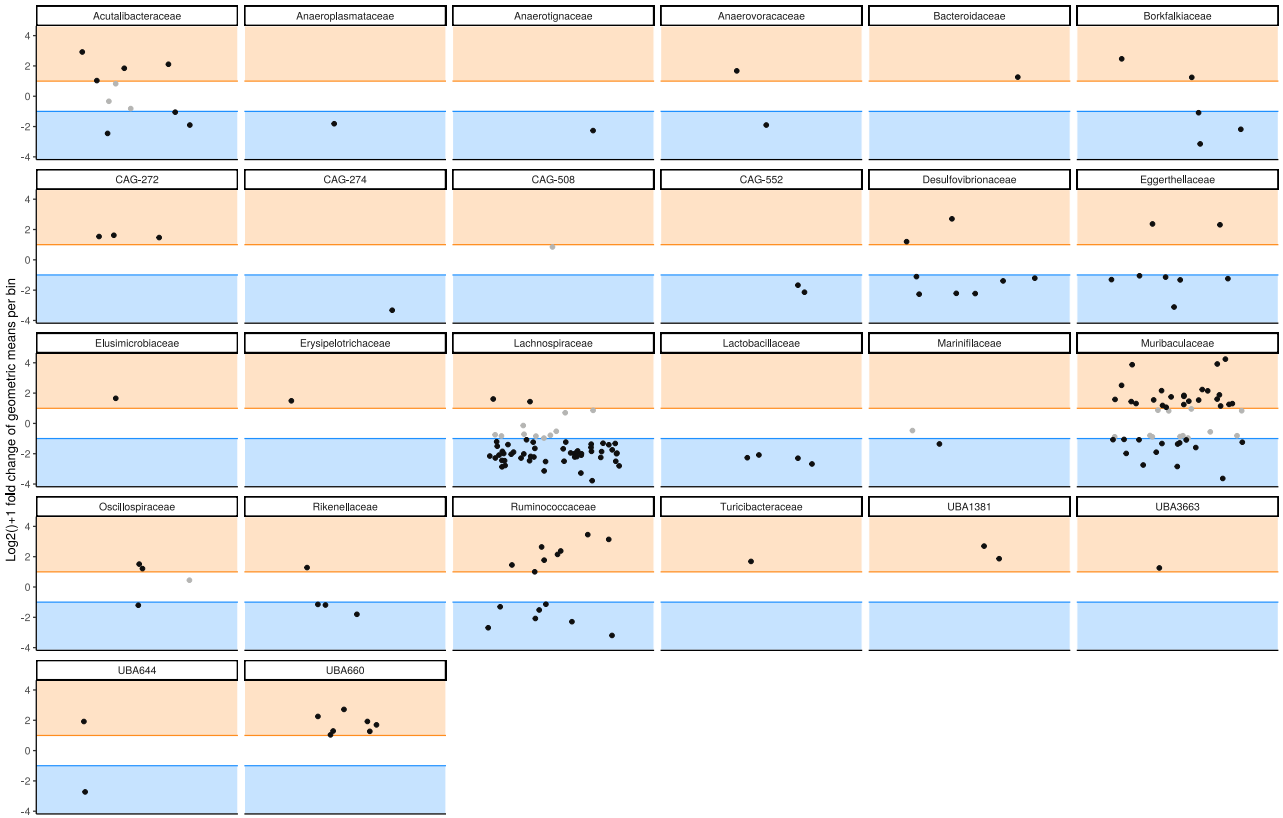

Figure 3.26: Point plots of log2()+1 fold change in geometric mean RPM. Showing bins with a Q-value of ≤ 0.05, grey points had change of lower than +/-1, black had change larger than +/-1. Faceted by taxonomic family assigned to the Bin by GTDB-Tk. Orange coloured field represents geometric mean RPMs enriched in *Acomys russatus* samples, blue field geometric mean RPMs enriched in *A. cahirinus* samples.

3.7 Annotation of metagenomic bins

Prokka annotation of the metagenomic bins led to the detection of 1,993 unique Clusters of Orthologous Genes (COGs) and 12,048 unique genes from all 348 bin files analysed.

3.7.1 COG annotations

The 10 COGs with the highest summed detection counts across all files can be seen in **Table 3.3** along with the summed detection counts across all bins, the bin which had the highest count for detections of the COG across all 348 bin files and a description of the COG from the NCBI Database of COGs.

COG	Count	Bin	Description
COG0745	3327	Bin_c1326	DNA-binding response regulator,OmpR family, contains REC and winged-helix (wHTH) domain
COG1132	2081	Bin_C918	ABC-type multidrug transport system, ATPase and permease component
COG1136	1845	Bin_c771	ABC-type lipoprotein export system, ATPase component
COG1595	1721	Bin_c908	DNA-directed RNA polymerase specialized sigma subunit, sigma24 family
COG0534	1381	Bin_m1179	Na+-driven multidrug efflux pump, DinF/NorM/MATE family
COG0395	1212	Bin_c211	ABC-type glycerol-3-phosphate transport system, permease component
COG1131	1177	Bin_c1431	ABC-type multidrug transport system, ATPase component
COG0488	984	Bin_c1325	ATPase components of ABC transporters with duplicated ATPase domains
COG3279	937	Bin_c771	DNA-binding response regulator, LytR/AlgR family
COG0621	927	Bin_c354	tRNA A37 methylthiotransferase MiaB

Table 3.3: Table giving the 10 most commonly detected COGs across all 348 bin files, showing the summed detection counts across all bins, the bin which had the highest count for detections of the COG from all bin files and a description of the COG from the NCBI Database of COGs.

Looking at the range of detections for all COGs across the bins there were 466 COGs (23.4% of all detected) which had a summed total detection count of 10 or less, 736 (36.9%) with a summed detection count between 11 and 100, 784 (39.3%) with a summed detection count of 101 to 1,000 and 7 (0.35%) with a summed detection count of 1,001 or more. The seven COGs can be seen in more detail in **Table 3.3**. Looking at the range of median detection counts for all COGs, there were 1,525 which had a median detection count of 0, 439 with a median detection count of 1, 13 with a detection count of 2 and 9 with a median detction count of 3 or greater. 302 of the 348 bins were the bin which had the highest detection count of at least one COG from all files analysed, though the number of COGs which any bin had the greatest detection count for

varied quite notably; as can be seen in **Figure 3.27**.

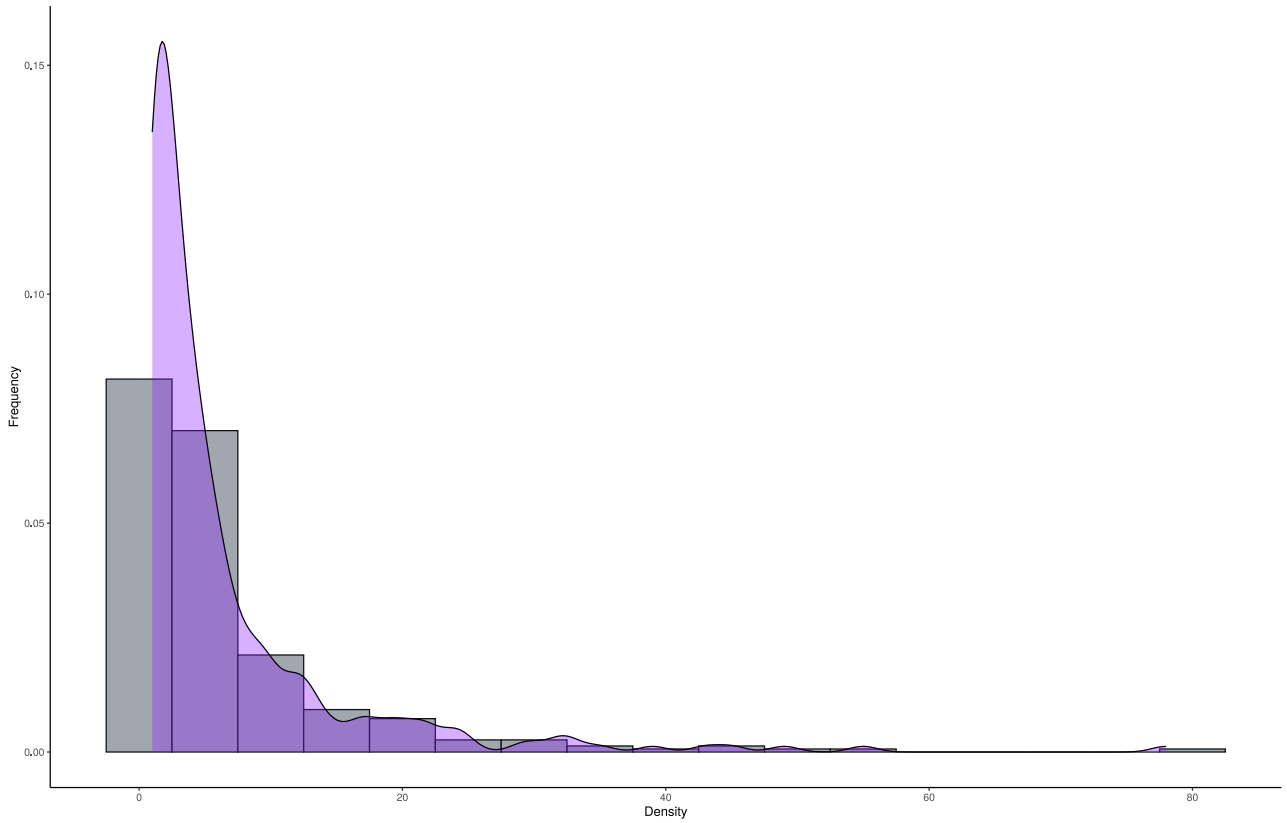

Figure 3.27: Density and frequency plot showing the distribution of the number of bins which had the greatest detection count for a COG across all bins, some bins having the greatest detection counts for multiple COGs

Assessing the distribution of COG detections counts in the context of bin enrichment across or within host species, there were no COGs which had a significant Q-value (≤ 0.05) for distribution in bins which were not differentially abundant, were differentially abundant in *A. cahirinus* against *A. russatus*, were differentially abundant within *A. cahirinus* from June to November or were differentially abundant within *A. russatus* from June to November. There were however a number of COGs which had a statistically significant distribution in those bins which were differentially abundant in *A. russatus* when compared as a whole to *A. cahirinus*; as can be seen in **Figure 3.28**. That is to say that for those bins when the month of sampling was ignored showed a species effect on their relative abundance (as measured by RPMs).

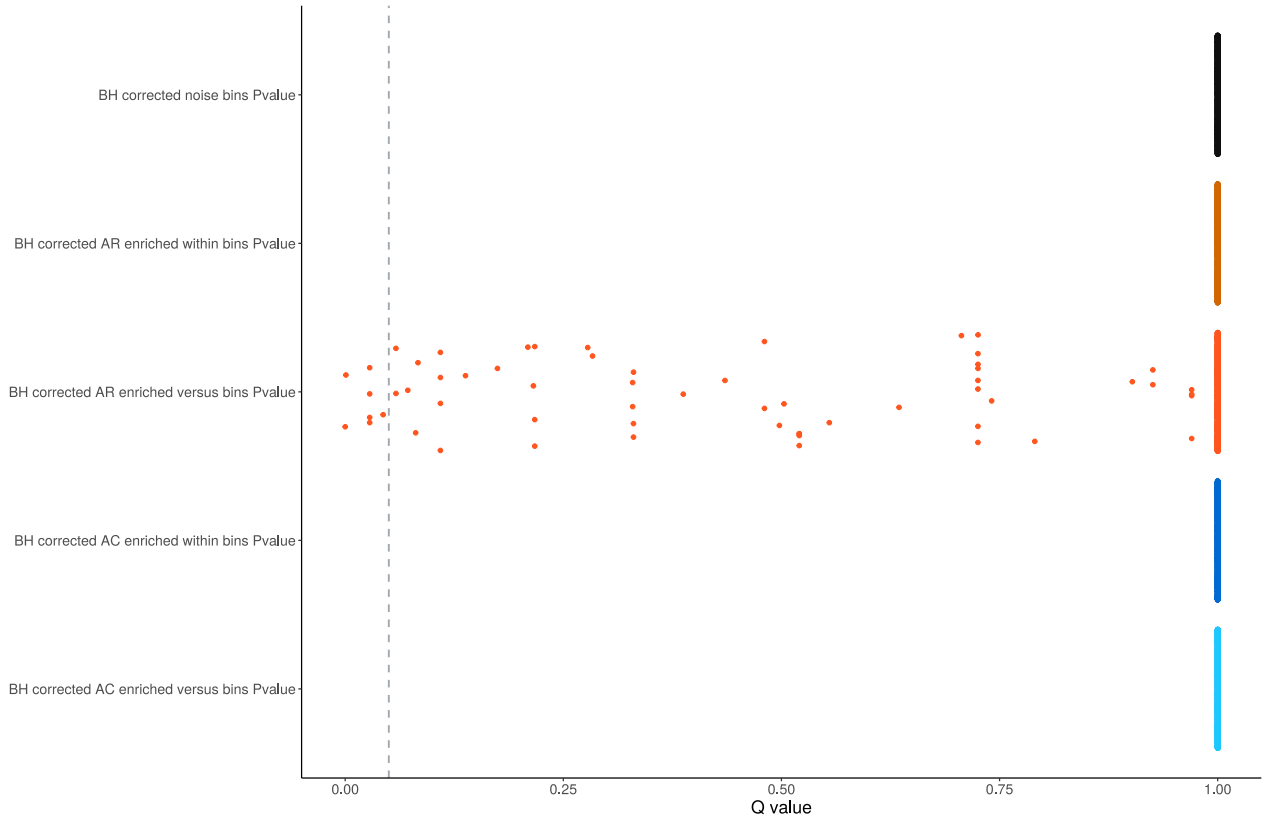

Figure 3.28: Jitter plot showing the Q-values obtained from hypergeometric tests for the distribution of COG detections by Prokka for all 348 bins, distinguishing between bins which showed some manner of differential abundance and those which did not ('Noise' bins). Colours differentiate the various differential abundance statuses of the bins. The dotted line represents a Q-value of 0.05.

The 7 COGs which had a significant Q-value for distribution in bins which were differentially abundant in *A. russatus* when compared to *A. cahirinus* were COG0001, COG0777, COG2176, COG0027, COG0801, COG0853 and COG2317. The Q-values and a description of these COGs is shown in **Table 3.4**.

COG	Q-value	Description
COG0001	0.00003635595	Glutamate-1-semialdehyde aminotransferase
COG0777	0.0008510107	Acetyl-CoA carboxylase beta subunit
COG2176	0.02809097	DNA polymerase III, alpha subunit (gram-positive type)
COG0027	0.02809097	Formate-dependent phosphoribosylglycinamide formyltransferase (GAR transformylase)
COG0801	0.02809097	7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase (folate biosynthesis)
COG0853	0.02809097	Aspartate 1-decarboxylase
COG2317	0.04353106	Zn-dependent carboxypeptidase, M32 family

Table 3.4: Table giving the 7 COGs which had a significant Q-value for distribution in bins which were differentially abundant in *A. russatus* when compared to *A. cahirinus*, showing the Q-values obtained and a description of the COG from the NCBI Database of COGs.

3.7.2 Gene annotations

The 10 genes with the highest summed counts can be seen in **Table 3.5** along with the summed detection counts across all bins, the bin which had the highest count for detections of the gene across all 348 bin files and a description of the gene from the NCBI Database of genes.

Gene	Count	Bin	Description
nusB	336	Bin_c1000	transcription antitermination protein NusB
pth	334	Bin_c1000	peptidyl-tRNA hydrolase
rplD	334	Bin_c1000	50S ribosomal subunit protein L4
rplO	334	Bin_c1005	50S ribosomal subunit protein L15
rplB	332	Bin_c1000	50S ribosomal subunit protein L2
truB	332	Bin_c1000	tRNA pseudouridine(55) synthase
rplW	331	Bin_c1000	50S ribosomal subunit protein L23
sasA_1	331	Bin_c1000	two component system sensor histidine kinase SasA
sasA_2	331	Bin_c1000	two component system sensor histidine kinase SasA
hisS	330	Bin_c1000	histidine–tRNA ligase

Table 3.5: Table giving the 10 most commonly detected genes across all 348 bin files, showing the summed detection counts across all bins, the bin which had the highest count for detections of the gene from all bin files and a description of the gene from the NCBI Database of genes.

Looking at the range of detections for all genes across the bins there were 2,827 genes (23.5% of total detected) which had a summed detection count from all bins of 1, 4,855 genes (40.2%) had a summed detection count of 2 to 10, 3,351 genes (27.8%) had a summed detection count of 10 to 100 and 1,015 genes (8.4%) had a summed detection count of 101 to 1,000. There were no genes detected in every bin. Looking at the range of median detection counts there were 11,523 genes which had a median detection count of 0, 7 genes with a median detection count of 0.5 and 518 genes had a median detection count of 1. 339 of the 348 bins were the bin which had the highest detection count of at least one gene from all files analysed, Bin_c1000 being the bin with the greatest number of genes with the highest detection count in a bin; it was the bin with the highest detection count for 900 genes. **Figure 3.29** shows the density and distribution of genes for which a bin had the highest detection count; the majority having only the greatest

detection count for 10 or fewer genes.

No genes had a statistically significant Q-value for differential distribution in the bins by enrichment status, with correction of P-values obtained from a hypergeometric test leading to none of ≤ 0.05.

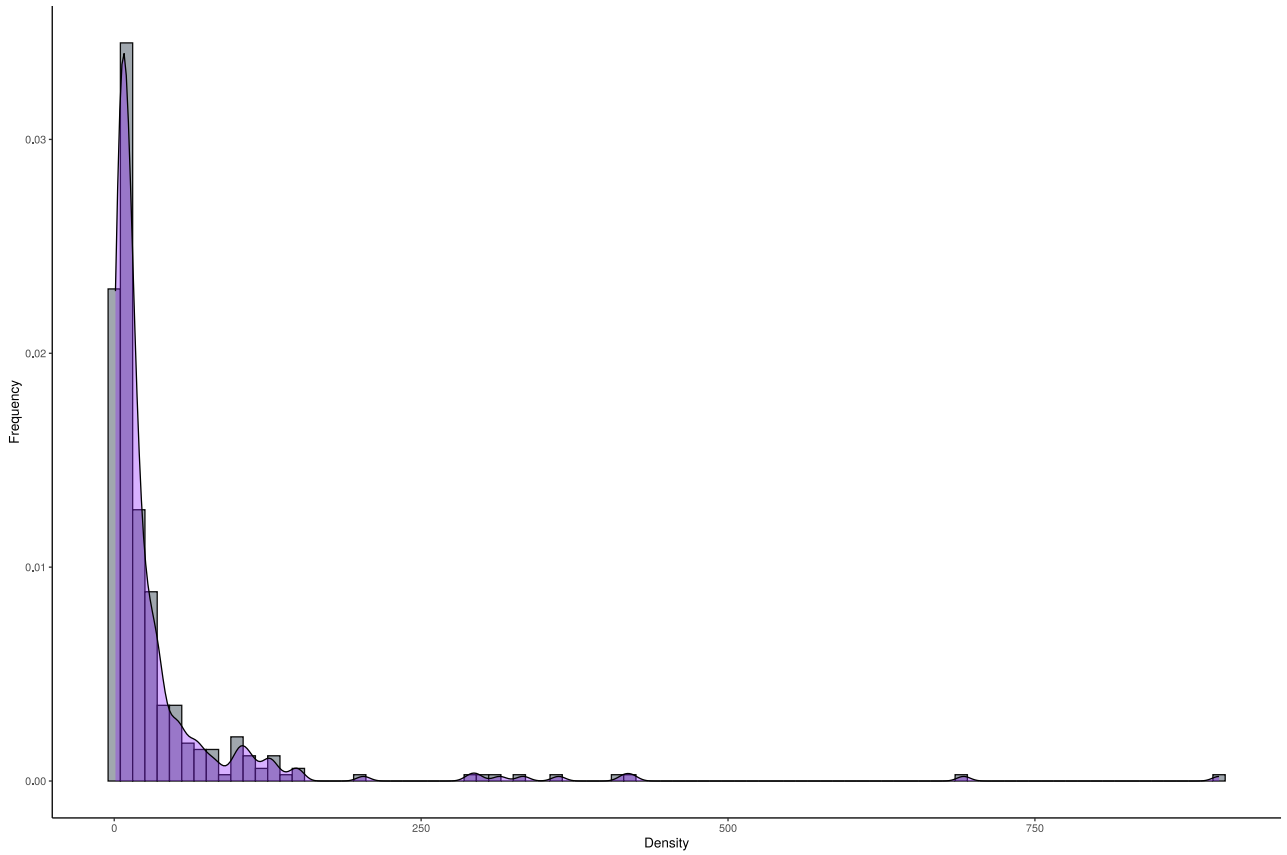

Figure 3.29: Density and frequency plot showing the distribution of the number of bins which had the greatest detection count for a gene across all bins, some bins having the greatest detection counts for multiple genes

3.8 LAB isolation and culturing from *Acomys* faecal samples

3.8.1 Colonies obtained

From the faecal samples a total of 29 isolates were obtained. The three *A. russatus* faecal samples yielded 13 distinct colonies, the *A. cahirinus* samples provided 16 distinct colonies. 5 of the *A. russatus* colonies came from samples from individual AR39, 5 from AR13 and 3 from AR41. For the *A. cahirinus* samples, 11 colonies in total came from the two pellets from individual AC16 - 6 from pellet AC16a and 5 from pellet AC16b. 5 more colonies came from AC18. The isolate IDs and the source *Acomys* organism are shown in **Table 3.6**.

Acomys cahirinus	***Acomys russatus***
AC 16 Pellet A - 16aA_S24	AR 13 - 13A_S1
AC 16 Pellet A - 16aB_S25	AR 13 - 13A_S2
AC 16 Pellet A - 16aC_S26	AR 13 - 13A_S3
AC 16 Pellet A - 16aD_S27	AR 13 - 13A_S4
AC 16 Pellet A - 16aE_S28	AR 13 - 13A_S5
AC 16 Pellet A - 16aF_S24	AR 39 - 39A_S14
AC 16 Pellet B - 16bA_S19	AR 39 - 39B_S15
AC 16 Pellet B - 16bB_S20	AR 39 - 39C_S16
AC 16 Pellet B - 16bC_S21	AR 39 - 39D_S17
AC 16 Pellet B - 16bD_S22	AR 39 - 39E_S18
AC 16 Pellet B - 16bD_S23	AR 41 - 41A_S6
AC 18 - 18A_S9	AR 41 - 41B_S7
AC 18 - 18B_S10	AR 41 - 41C_S8
AC 18 - 18C_S11	
AC 18 - 18D_S12	
AC 18 - 18E_S13	

Table 3.6: IDs for isolates obtained from culturing of faecal samples from *Acomys* individuals along with the host organism the faecal sample originated from. Sample ID is show first then isolate ID

3.8.2 Processing of reads from isolated LAB

During sequencing the negative control yielded sequenced reads, 48 in total. The reads obtained from the negative control were also submitted to the NCBI [440] `blastn` online tool, no hits were found with either `megablast` or `blastn`.

3.9 Assembly of LAB reads

Assemblies were produced from the sequenced reads from the colony isolates following the described methods. After processing to remove any errant contigs which were less than 500 bp in length the resulting assemblies were examined. The *Acomys*isolate assemblies had a size range from 1.7 - 2.5 Mbp, with a median assembly size of 2.2 Mbp and a mean assembly size of 2.2 Mbp. The *Apodemus*isolate assemblies had a size range of 1.9 - 2.7 Mbp, a median size of 2.2 Mbp and a mean size of 2.2 Mbp.

Looking at all the *Acomys* isolate derived assemblies together, irrespective of host species, the number of contigs per assembly ranged from 18 - 146, the median number of contigs per assembly was 60 (73 for *A. cahirinus* and 59 for *A. russatus*). The *Apodemus*isolate assemblies had a much greater range of contigs per assembly, from 36 - 909 with a median of 51. The values for the *Apodemus* isolate assemblies were distorted by the assembly produced from isolate S125, which contained 909 contigs. The variation in the number of contigs for each assembly is shown in **Supplemental Figure 5.8** by faecal sample origin species.

3.9.1 CheckM analysis of isolate assemblies

CheckM provided a value for the completeness and contamination of analysed assemblies. **Figure 3.30** shows the variability of completeness and contamination measurements obtained along with the thresholds used to determine whether the assemblies should be retained for subsequent analysis. One assemblies was discounted for falling outside the thresholds, *Apodemus* isolate S125 which was 17.57% contaminated, as measured by CheckM. All completeness values were >98%, higher than the threshold of 80% the author used, and aside from S125 all assemblies had a contamination value <5%. All assemblies aside from S125 were retained for the next stage of analysis.

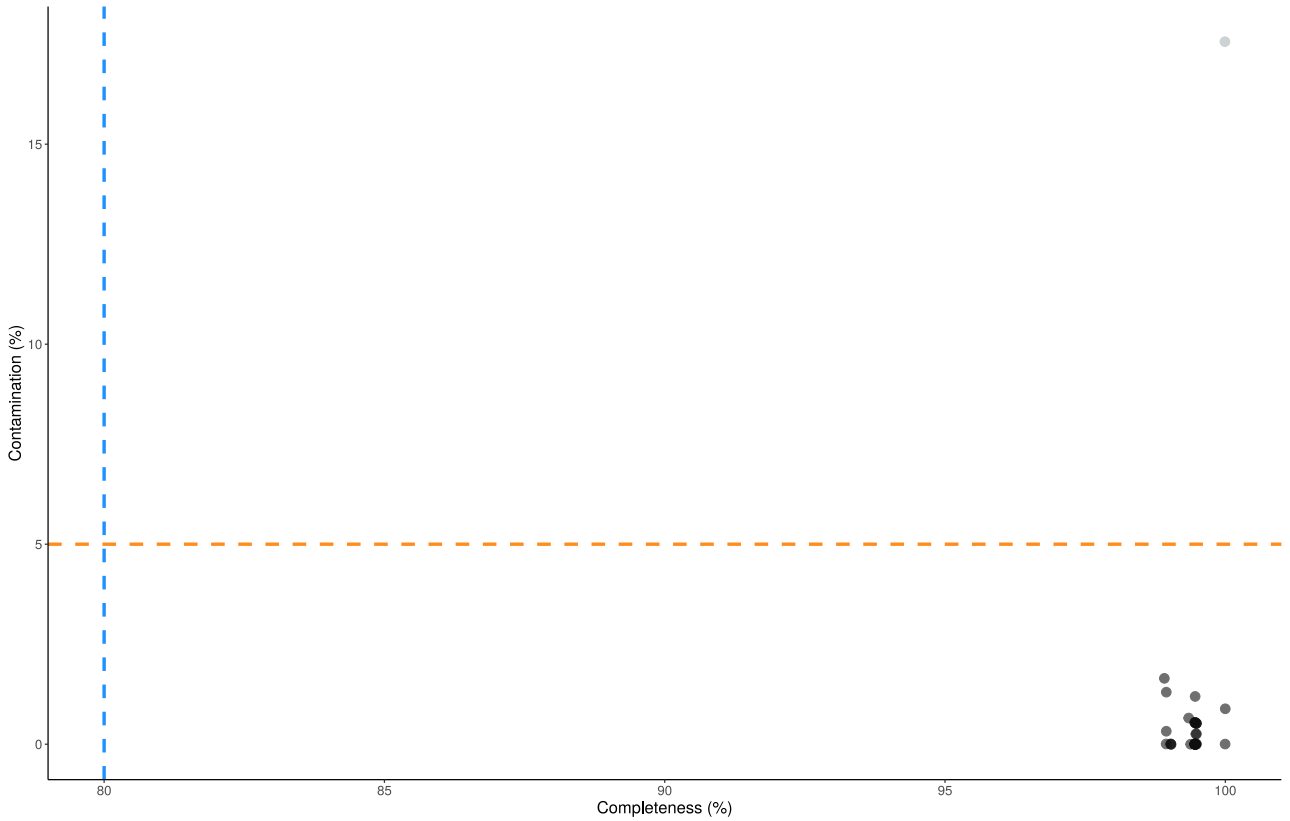

Figure 3.30: Plot showing the range of values measured by CheckM for isolate assemblies contamination and completeness. Coloured, dotted lines show the threshold values used to determine whether assembly should be retained for further analysis. For completeness this was 80%, for contamination this was 5%

3.9.2 ANI analysis of isolate assemblies

To ensure that no analyses of multiple isolates of the same strain was conducted, FastANI was used to compare the ANI similarity between all isolate assemblies. Initial results identified two pairs of assemblies which had an ANI similarity of $\geq 99.9\%$ to each other. These were 13E_S5 to 16aC_S26 and 18E_S13 to 18B_S10. Within the pairs each assembly had the same completeness and contamination values so 13E_S5 and 18B_S10 were randomly selected for further bioinformatics analysis. Isolate 16aC_26 was used in the later halotolerance culturing experimentation and was annotated during that process but was otherwise not used in the subsequent bioinformatic analysis. Masked versions of the assemblies were created using BBMask and then analysed with FastANI again to confirm that the detected high levels of similarity were not the result of highly repetitive regions within the assemblies. Three assemblies did not generate any output with FastANI, nor did they appear in the output from any other assemblies. These were from isolates 16aF_S29, 41A_S6 and 41C_S8; this indicates they did not have an ANI similarity scores of 'close to 80% or higher' [418] to any of the other assemblies. The *Apodemus* isolates when compared to each other all had ANI similarity scores of at least $\approxeq 80\%$. None had values of $\geq 99.9\%$ to each other. Comparing the assemblies to each other from all host species, *Acomys russatus* isolate assembly 41B_S7 also appeared in the results for all *Apodemus* assemblies, indicating a degree of similarity.

3.9.3 rRNA detection in isolate assemblies

Barrnap as a tool searches for ribosomal RNA genes in files submitted to it. Looking at the *Apodemus* assemblies, of the 7 all had a complete 16S rRNA detection by Barrnap - indicating the presence of complete 16S rRNA genes. 12 of the 16 assemblies from *Acomys cahirinus* faecal samples had a complete 16S rRNA detection, as did 10 of the 13 *Acomys russatus* isolate assemblies. One *A. cahirinus* assembly did not have any complete 16S rRNA gene detections, 16aA_S24, as did one *A. russatus* assembly, 13E_S5. Two *A. cahirinus* assemblies had incomplete 16S rRNA gene detections along with complete ones, 18A_S9 and 16aD_S7. Looking at the *A. russatus* isolate assemblies, two isolates had no complete 16S rRNA gene detections - 41B_S7 and 13E_S5 - and one assembly had both a complete and an incomplete 16S rRNA gene detection, 41A_S6. Though they did not have any complete detections, there were large proportions of entire 16S rRNA genes detected in isolate assemblies 13E_S5, 16aA_S24 and 41B_S7.

Barrnap does not only detect 16S rRNA genes. 17 assemblies had a detection of complete 23S rRNA genes, 3 from *Apodemus*, 8 *Acomys cahirinus* and 6 *Acomys russatus*. 37 assemblies had a detection of complete 5S rRNA genes, all 7 from *Apodemus* isolates, 13 from *A. cahirinus* and 11 from *A. russatus*.

3.10 Taxonomic identification of LAB assemblies

The isolates which were found to not be identical to each other either before or after masking and which had not been removed for failing the CheckM thresholds were analysed with GTDB-Tk. The unmasked versions of the assemblies were used to generate a taxonomic identity for the assemblies. All but two isolates were classified into the Lactobacillaceae family, the exceptions being 18A_S9 and 39B_S15 which were classified into the Bifidobacteriaceae. The two outliers were not assigned a species by GTDB-Tk but were placed into the genus *Bifidobacterium*. 7 isolates could be classified to the species level, five were isolates obtained from *Apodemus* samples while the two from *Acomys* samples were 41C_S8 and 41B_S7. 41C_S8 was classified as *Pediococcus pentosaceus* and 41B_S7 was classified as *Ligilactobacillus murinus*. The five *Apodemus* assemblies classified to the species level were also classified as *Ligilactobacillus murinus*. All isolates could be assigned to the genus level, **Figure 3.31** shows the number of assemblies classified to each detected genus.

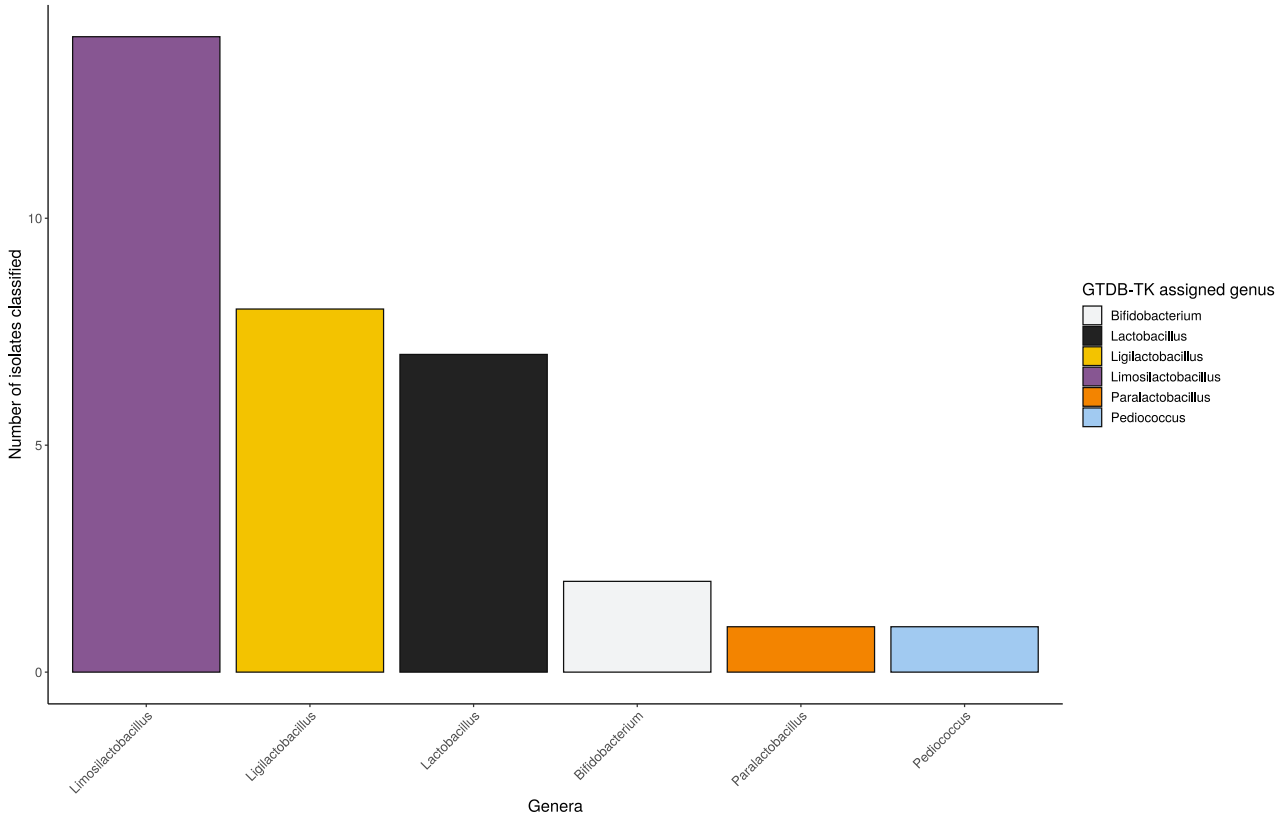

Figure 3.31: Bar plot showing number of assemblies classified to each of the genera detected by GTDB-Tk.

Of the 33 isolates, 14 were classified as *Limosilactobacillus*, 8 as *Ligilactobacillus*, 7 as *Lactobacillus*, 2 as *Bifidobacterium* and 1 each as *Paralactobacillus* and *Pediococcus*. **Table 3.7** gives the lowest level of GTDB-Tk classification for all assemblies.

Isolate assembly	Classification
13A_S1	*Lactobacillus*
13B_S2	*Limosilactobacillus*
13D_S4	*Limosilactobacillus*
13E_S5	*Limosilactobacillus*
16aA_S24	*Limosilactobacillus*
16aB_S25	*Lactobacillus*
16aC_S26	*Limosilactobacillus*
16aD_S27	*Limosilactobacillus*
16aE_S28	*Limosilactobacillus*
16aF_S29	*Limosilactobacillus*
16bA_S19	*Lactobacillus*
16bB_S20	*Lactobacillus*
16bC_S21	*Limosilactobacillus*
16bD_S22	*Limosilactobacillus*
16bE_S23	*Limosilactobacillus*
18A_S9	*Bifidobacterium*
18B_S10	*Limosilactobacillus*
18D_S12	*Lactobacillus*
18E_S13	*Limosilactobacillus*
39A_S14	*Limosilactobacillus*
39B_S15	*Bifidobacterium*
39C_S16	*Lactobacillus*
39D_S17	*Lactobacillus*
39E_S18	*Limosilactobacillus*
41A_S6	*Paralactobacillus*
41B_S7	*Ligilactobacillus murinus*
41C_S8	*Pediococcus pentosaceus*
S121	*Ligilactobacillus murinus*
S122	*Ligilactobacillus murinus*
S123	*Ligilactobacillus murinus*
S124	*Ligilactobacillus*
S125	*Ligilactobacillus*
S126	*Ligilactobacillus murinus*
S127	*Ligilactobacillus murinus*
S128	*Ligilactobacillus*

Table 3.7: Table giving the lowest GTDB-Tk classification for all assemblies including ones which failed QC or were discarded due to ANI similarity to another assembly.

3.11 Phylogeny of LAB assemblies

3.11.1 Relationships between isolates

The phylogenetic tree produced from the alignments of the isolate assemblies can be seen in **Figure 3.32**. Clear from the tree is the clustering of the assemblies from the *Apodemus* faecal samples which can be seen in the bottom of the figure, along with two assemblies originating from *Acomys russatus* faecal samples. The tree shows no clustering of assemblies from *A. cahirinus* with those from *Apodemus*.

Three subtrees of varying levels of cohesion and isolation from each other can be seen in the figure. The clearest is the subtree containing the aforementioned *Apodemus* isolate assemblies and the two *A. russatus* assemblies (41B_S7 and 41A_S6) along with another *A. russatus* assembly, 41C_S8 which marks one extent of the substree and is the most distant from all other members of the subtree. The next most easily distinguished subtree is that which covers the assemblies from 16bA_S19 through 13A_S1. Though this subtree contains both *A. cahirinus* and *A. russatus* assemblies the *A. russatus* assemblies are more closely related to each other than to the *A. cahirinus* assemblies within the subtree. The final subtree is that which contains the assemblies from 16bD_S22 through 13B_S2, it is the least coherent and most difficult to distinguish as a specific subtree from the overal phylogenetic tree. This subtree contains three assemblies from *A. russatus* samples and six from *A. russatus* samples.

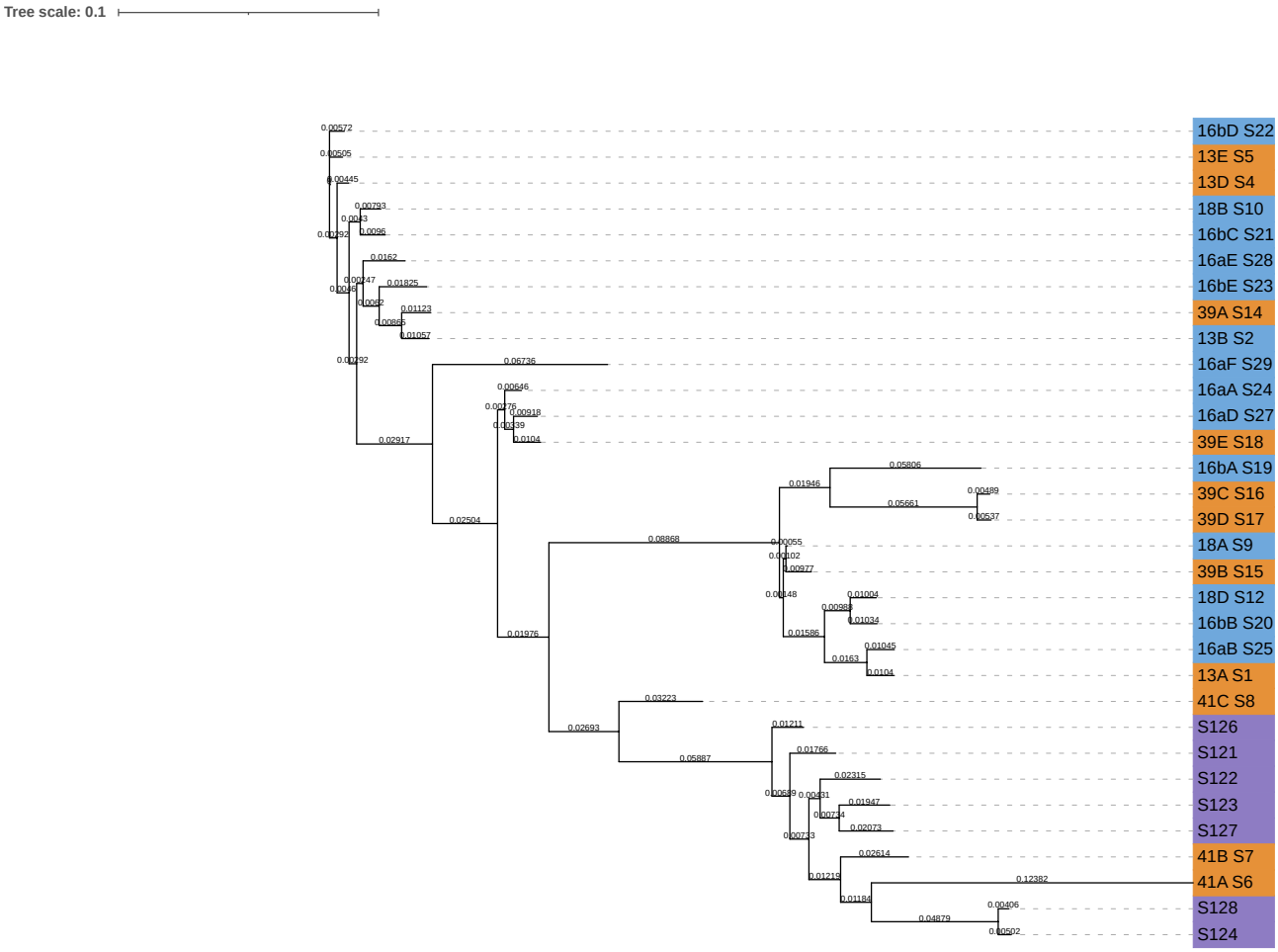

Figure 3.32: Shows a phylogenetic tree of all assemblies which passed quality control metrics. The tree was constructed using FastTree and the alignment was generated using SibeliaZ. Colours are used to distinguish the rodent species or genus the isolate the assembly was derived from, blue for *Acomys cahirinus*, orange for *Acomys russatus* and purple for *Apodemus*

3.11.2 Relationships between isolates, metagenomic bins and reference genomes

The metagenomic bins produced in the project and discussed earlier were analysed alongside the assemblies generated from the isolates to asess the phylogenetic relationships between the assemblies and the bins. This was carried out both through the use of FastANI to compare ANI similarity in an all against all comparison and through the production of a phylogenetic tree which included both the bins and the assemblies.

Isolate assemblies and metagenomic bins ANI similarities

Of the 348 metagenomic bin files there was one, Bin_c770 (classified by GTDB-Tk as a member of the Lactobacillaceae) which had an ANI similarity score above 90% for all but two of the *Apodemus* isolates; for those other two it still had scores above 80%. This bin also had an ANI similarity score of 91% to Acomys isolate 41B_S7, which itself had very high ANI similarity scores to the same *Apodemus* isolates as the bin. For the *Acomys* isolate assemblies, aside from the previously mentioned 41B_S7, there were three bins which had reported (meaning a minimum 80% ANI similarity) matches to some of the isolates. Bin_c1617 (classified as Lactobacillaceae) had ANI similarities scores of between 78 to 80% to isolate assemblies 13A_S1, 16aB_S25,

16aB_S19, 16bB_S20, 18D_S12, 39C_S16 and 39D_S17 - 3 assemblies from *A. russatus* and 4 from *A. cahirinus*. It had marginally higher similarity scores, 80% rather than 78 or 79%, for the assemblies from *A. russatus* samples. Bin_m1569 (classified as Lactobacillaceae) had an ANI similarity score of 99.0% to isolate assembly 18A_S9 and 98.8% to isolate assembly 39B_S15. These assemblies were classified as a *Bifidobacterium* species. Bin_m1485 (classified as Lactobacillaceae) had an ANI similarity score of 78-79% to the *Acomys* isolate assemblies 13A_S1, 16aB_S25, 16bA_S19, 16bB_S20, 18D_S12, 39C_S16 and 39D_S17. These isolate assemblies were classified themselves as *Lactobacillus* members.

Phylogenetic tree of assemblies, bins and reference genomes

Phylogenetic trees of the assemblies and the metagenomic bins can be seen in **Section 3.5** where they are discussed. Here a larger bin was created which included the 348 processed bin files, the processed assembly files and the external bacterial reference genomes described in **Subsection 2.1.7**. **Figure 3.33** shows the phylogenetic tree produced from alignment of all these files with Cactus and visualised with iTOL.

The most immediate observation is the small number of bins which are located near on the same large subtree as the majority of the external reference genomes, 22 of 348. 11 of these bins were classified as Desulfovibrionaceae by GTDB-Tk, which accounted for all but one of the bins classified into the family. In terms of the bins by family classification and relationship to reference genomes, there are only five reference genomes placed in the large subtree which contains all bins classified as Muribaculaceae - two of the external genomes are in fact iMGMC MAGs. Notably none of the assemblies from either the two *Acomys* species or the *Apodemus* are placed outside the large substree containing the majority of the external reference genomes. Looking at assemblies which have an external reference as their nearest neighbour at the end of branch, 39B_S15 has the genome GCF 002289215 (a *Bifidobacterium pseudolongum* genome) as the nearest neighbour. Assembly 16aF_S29 has as its nearest neighbour the genome GCF 009428965 (a *Limosilactobacillus pontis* genome). Assembly 16bA_S19 has as its nearest neighbour the genome GCF 009734005, an *Enterococcus faecium* genome. The tree also shows some bins still being placed far away from others with the same family classification - Bin_c890 was classified as Lachnospiraceae and placed in the large subtree which contains no other bins classified in the family. Three bins classified as Acutalibacteraceae, Bin_m595, Bin_c420 and Bin_c1226 are placed away from all other bins classified within the family. The one Desulfovibrionaceae bin not found with the others was placed at almost the furthest extreme away from them in the tree on a branch which is very close to root.

Looking in more detail at the placement of the reference genomes and distinguishing between those which were iMGMC MAGs, only one of the iMGMC MAGs was placed in the large subtree which contained the majority of external reference genomes and all of the assemblies - running from NZ CP042413 (a *Leuconostoc citreum* genome) through to Bin_m1549. The remaining MAGs were placed across the tree, two in the subtree which contained all of the bins classified as Muribaculaceae. There is a greater proportion of AC-enriched bins (16) in the large subtree containing the majority of the external references and all assemblies than bins enriched in *Acomys russatus* or bins not enriched in either species.

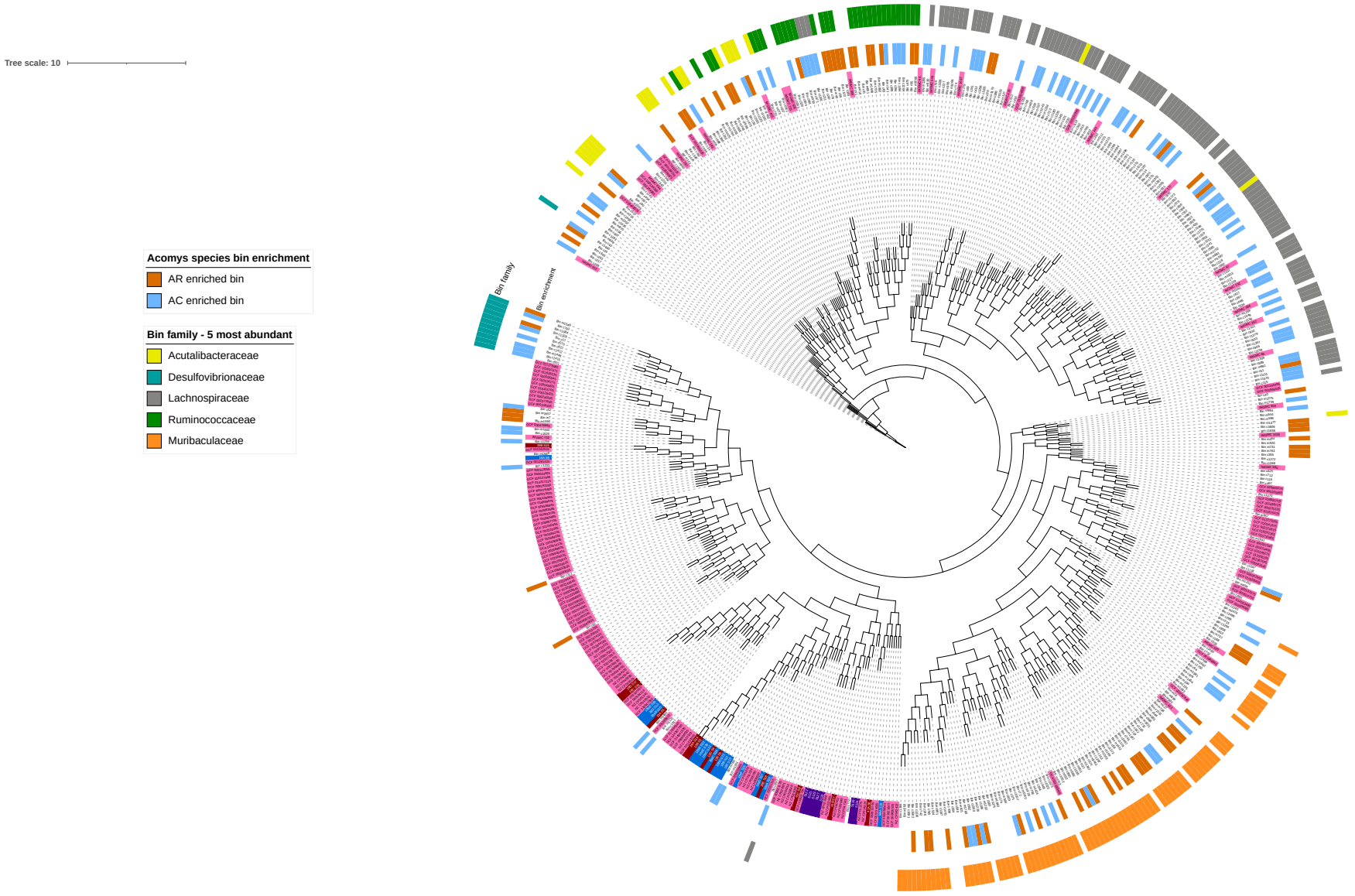

Figure 3.33: Phylogenetic tree of metagenomic bins, *Acomys* isolate assemblies, *Apodemus* isolate assemblies, iMGMC MAGs and downloaded reference genomes. Metagenomic bins are not coloured at the nodes but by external strips where relevant. iMGMC MAGs and downloaded external reference genomes are coloured pink at the nodes, *Acomys russatus* derived isolate assemblies in brown, *Acomys cahirinus* derived isolate assemblies in light blue and *Apodemus* derived isolate assemblies in dark blue. The legends in the figure show the colours used to classify bins by enrichment status on the inner strips and GTDB-Tk family classification for the five most abundant families on the outer strip.

3.12 Mapping of *Acomys* faecal sample shotgun reads to LAB assemblies

Mapping of reads from *Acomys* faecal samples to a concatenated reference fasta file made from the individual assemblies was carried out to establish the likely abundance of the microorganisms which the assemblies originate from within the *Acomys* microbiota. To ensure that the subsampling of the read files did not lead to a major change in the mapping of the reads to the assemblies reference file, the mapping was carried out on the raw read files and the subsampled read files. The results of this can be seen in **Supplemental Figure 5.9**. There were marginal differences in the mapping percentage between the read files when they were subsampled or not. The largest difference was a decrease in mapping percentage of 0.02% after subsampling, seen with reads from sample AC37J. All other sample read files mapped either the same amount whether subsampled or not or had a difference of +/- 0.01% after subsampling. Note that these read files did not include those from the samples which were used to culture the isolates for the assemblies as they had come from samples sequenced on a different date and to a different (lower) initial depth.

The range of mapping to *Acomys* isolate assemblies for reads from *Acomys cahirinus* (from both June and November samplings) was 91.1 to 359,245.89 RPM, the median RPM was 14,246 and the arithmetic mean was 40,000. For reads from *A. russatus* the range of RPM values was 130.2 to 657,155, the median RPM was 7,742 and the arithmetic mean was 40,000. **Figure 3.34** show the different RPM values for all reads to the different isolate assemblies.

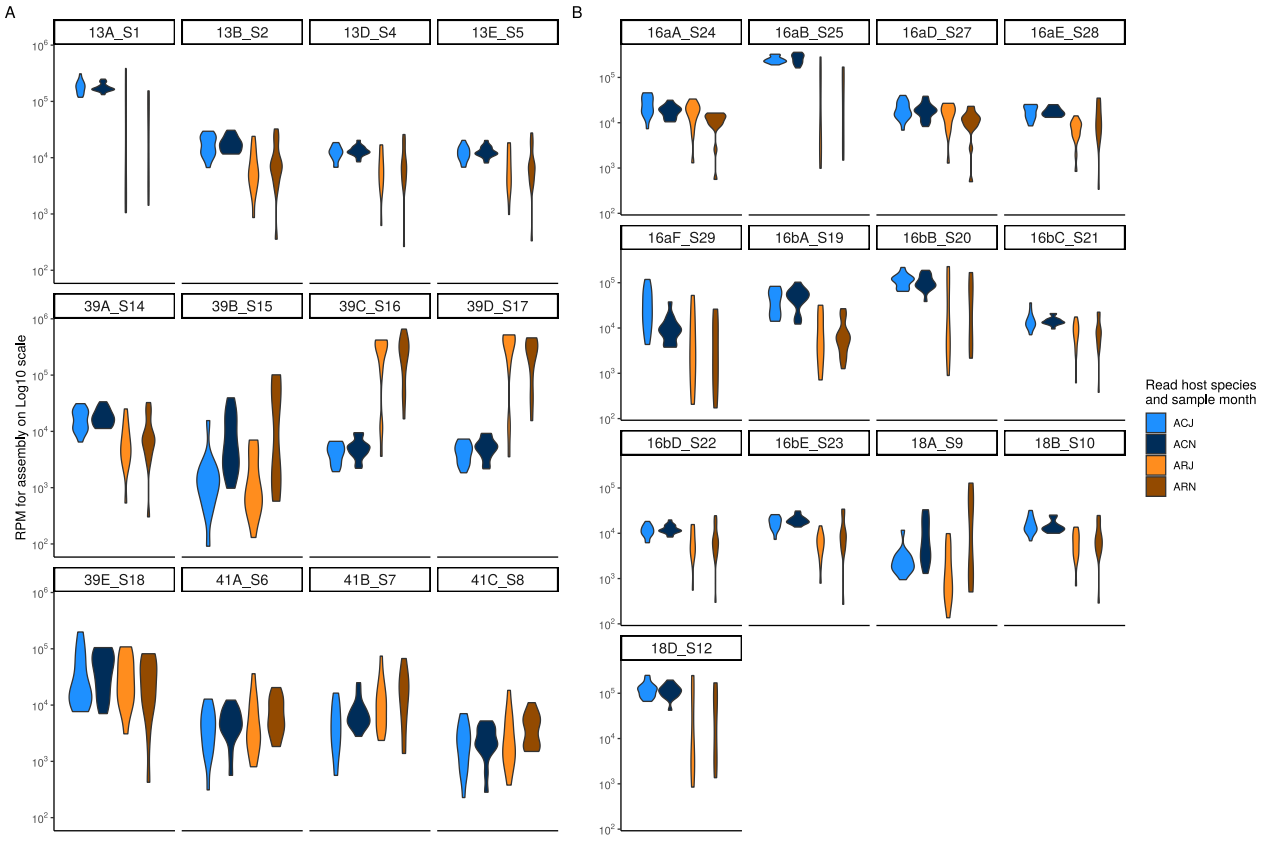

Figure 3.34: Violin plots of reads per-million (RPM) for all subsampled, paired sampling reads from both *Acomys* species to assemblies of isolates from faecal samples from *Acomys cahirinus* **A** and *Acomys russatus* **B**

Figure 3.35 visualises the results of Principal Component Analysis using the RPM data for the

Acomys reads to the assemblies. **Supplemental Figure 5.10** gives the PCA plots for mapping of the reads to the assemblies by the origin species of the faecal sample the assemblies were produced from. This latter step was carried out using the results for the mapping of all reads to the entire combined assemblies reference, not by remapping based on read and assembly host species.

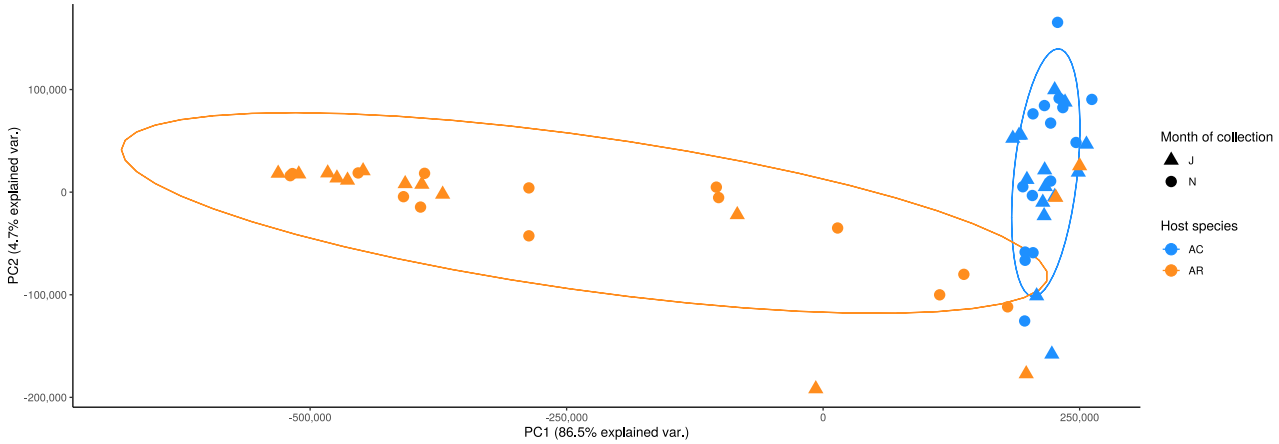

Figure 3.35: Principal Component Analysis of reads per-million (RPM) for all subsampled, paired sampling reads from both Acomys species to assemblies of isolates from faecal samples from *Acomys cahirinus* and *Acomys russatus*

Looking specifically at the mapping of the reads to the assemblies by the origin species of the faeces they were isolated from, the range of RPMs for *A. cahirinus* reads to *A. cahirinus* derived assemblies was 945.3 to 359,245.9, the arithmetic mean RPM was 52,476.3 and the median RPM was 19,139.4. For reads from *A. russatus* mapping to assemblies derived from *A. russatus*, the range of RPMs was 1,303 to 657,155.5, the arithmetic mean RPM was 60,196.63 and the median RPM was 7,785. Looking at the mapping of reads to the assemblies derived from faecal samples from the other *Acomys*, the range of RPMs for *Acomys cahirinus* reads to assemblies derived from *Acomys russatus* faecal samples was 91.1 to 309,244.6, the median RPM was 9,574.6 and the arithmetic mean was 26,484. The range of RPMs for mapping of *Acomys russatus* reads to assemblies derived from *Acomys cahirinus* faecal samples was 136.4 to 282,356.1, the arithmetic mean RPM was 21,357 and the median RPM was 7,700.5. Kruskal-Wallis tests were conducted to see if there was a statistically significant difference between the geometric mean RPM values to each of the isolates for the reads by the host species and sampling month for the faecal sample reads. After correction all assemblies had a statistically significant Q-value (less than 0.05) for a host species difference for mapping by the reads. Only two assemblies had statistically significant Q-values for a sampling month effect, 13A_S1 from *A. russatus* and 16bA_S19 from *A. cahirinus*.

3.13 Annotation of LAB assemblies

Prokka was run on all assemblies, including those excluded from some analysis steps on the basis of ANI similarity $\geq 99.9\%$ another assembly; with the exception of *Apodemus*assembly S125 which had a CheckM contamination score above the threshold of 5%.

3.13.1 COG annotations

A total of 1,143 unique COGs were detected from all assemblies. 695 COGs were detected at least once in assemblies from *Apodemus*samples, 1,036 in assemblies from *A. cahirinus* assemblies and 1,090 in assemblies from *A. russatus* samples. Looking at all COGs within all assemblies for each host species, the mean count of detections by Prokka for *Apodemus*assemblies was 0.737, for *A. cahirinus* assemblies it was 0.674 and for *A. russatus* assemblies it was 0.678. The greatest number of COG detections for an *Apodemus* assembly was 9, for an *A. cahrinus* assembly it was 14 and for an *A. russatus* assembly it was 13. The median COG detection count from all assemblies for each host species was 1 for each. The 10 COGs with the highest summed detection count across all assemblies are detailed in **Table 3.8**.

The most commonly detected COG across all assemblies was COG0531 (Serine transporter YbeC, amino acid:H+ symporter family), this was also the most commonly detected COG in assemblies from both *Acomys* species. The most commonly detected COG from the *Apodemus* assemblies was COG2188 (DNA-binding transcriptional regulator, GntR family). The next four most commonly detected COGs from all assemblies were COG1609 (DNA-binding transcriptional regulator, LacI/PurR family), COG1132 (ABC-type multidrug transport system, ATPase and permease component), COG0745 (DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain) and COG0438 (Glycosyltransferase involved in cell wall bisynthesis). **Figure 3.36A** shows the results of conducting a hypergeometric distribution test in R to assess whether the distribution of each of the 1,143 COGs in assemblies originating from each host species indicated a host effect.

COG	Assembly with most	Summed count	Description
COG0531	16aA_S24	342	Serine transporter YbeC, amino acid:H+ symporter family
COG1609	18A_S9	288	DNA-binding transcriptional regulator, LacI/PurR family
COG1132	13A_S1	184	ABC-type multidrug transport system, ATPase and permease component
COG0745	13A_S1	183	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain
COG0438	16aD_S27	178	Glycosyltransferase involved in cell wall bisynthesis
COG0561	16aB_S25	175	Hydroxymethylpyrimidine pyrophosphatase and other HAD family phosphatases

COG2188	S122	148	DNA-binding transcriptional regulator, GntR family
COG0656	16aF_S29	147	Aldo/keto reductase, related to diketogulonate reductase
COG4690	16aF_S29	146	Dipeptidase
COG1940	16aB_S25	134	Sugar kinase of the NBD/HSP70 family, may contain an N-terminal HTH domain

Table 3.8: Table giving the 10 most commonly detected COGs across all analysed assembly files, showing the summed detection counts across all bins, the assembly which had the highest count for detections of the COG from all assembly files and a description of the COG from the NCBI Database of COGs.

54 unique COGs had a species effect, a Q value of ≤ 0.05, though 13 of the COGs have a significant effect in more than one host species. **Supplemental Table. 5.4** lists the detected COGS which had a Q value of ≤ 0.05, showing each result including both host effects for COGs with a significant result in isolates from more than one species.

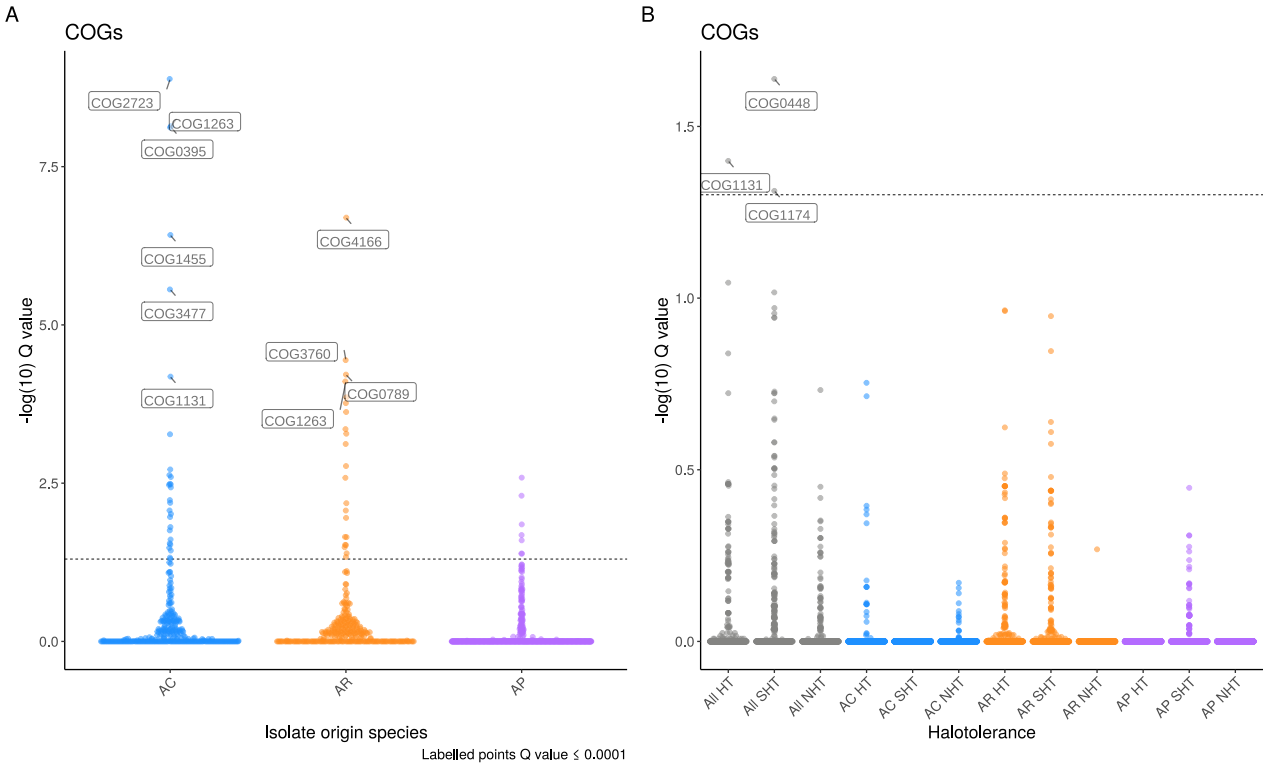

Figure 3.36: Plot showing -log(10) Q values for COGs detected by Prokka from processed assembly files. In both figures dotted line shows 0.05 significance threshold. COGs are not filtered to be unique across all groups. **A.** Results for all assemblies processed, labelled points are those with a Q value ≤ 0.0001. **B.** Results for assemblies from isolates assessed for halotolerance, labelled points are those with a Q value of ≤ 0.05. HT : Halotolerant, SHT : Slightly halotolerant, NHT : Not halotolerant.

Figure 3.36B shows the results for a hypergeometric distribution test by halotolerance of the

assemblies, this being discussed in more detail in **Section 3.14**. Here there were no COGs which had a statistically significant Q-value for any halotolerance when including the origin species as a factor, nor in all the Non-Halotolerant (NHT) isolates. There were two COGs in the Slightly Halotolerant (SHT) isolates which had a significant Q-value, being COG0448 and COG1174. These have the descriptions of 'Glucose-1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase)' and 'ABC-type proline/glycine betaine transport system, permease component' respectively. There was one COG which had a significant Q-value for the Halotolerant (HT) isolates, this being COG1131 which has a description of 'ABC-type multidrug transport system, ATPase component'.

3.13.2 Gene annotations

A total of 3,303 different genes were detected by Prokka from all assemblies. 2,620 were detected at least once in assemblies from *A. cahirinus* isolates, 2,844 from *A. russatus* and 1,491 from *Apodemus* assemblies. **Table 3.9** shows the 10 genes with the greatest summed detection counts across all assemblies along with the assembly which had the greatest detection count for the gene and a description of the gene from the NCBI Database of genes. Looking at all genes within all assemblies for each host species, the mean count of detections by Prokka for *Apodemus*assemblies was 0.35, for *A. cahirinus* assemblies it was 0.32 and for *A. russatus* assemblies it was 0.32. The median gene detection count from all assemblies for each host species was 0 for each. A single gene was detected more than once in an individual assembly, this being the gene *ssrA* which was detected twice in the *A. cahirinus* assembly 16aF_S29.

Gene	Count	Assembly	Description
ssrA	35	16aF_S29	ncRNA
acpS	34	13A_S1	holo-[acyl-carrier-protein] synthase
adk	34	13A_S1	adenylate kinase
alaS	34	13A_S1	alanine–tRNA ligase/DNA-binding transcriptional repressor
alr	34	13A_S1	alanine racemase 1
apt	34	13A_S1	adenine phosphoribosyltransferase
argS	34	13A_S1	arginine–tRNA ligase
artQ	34	13A_S1	arginine ABC transporter permease ArtQ
aspS	34	13A_S1	aspartate–tRNA ligase
atpA	34	13A_S1	ATP synthase F1 complex subunit alpha

Table 3.9: Table giving the 10 most commonly detected genes across all 35 assemblies, showing the summed detection counts across all assemblies, the assembly which had the highest count for detections of the gene from all assembly files and a description of the gene from the NCBI Database of genes.

No genes detected by Prokka had a statistically significant distribution according to the host species the isolate originated from, nor a statistically significant distribution according to the halotolerance of the isolate. This indicates there was no statistically significant enrichment of any detected genes by either sample origin host species or halotolerance of the isolate.

3.14 Halotolerance of select LAB isolates

The growth curves obtained for the assessed isolates can be seen in **Figure 3.37**, showing the results for the 3 replicates of the experiment. The results are broadly consistent and the lack of growth of the negative control indicates that external contamination is not a concern when interpreting these results. The variability of the positive control growth rates along with the failure of the positive control to grow at salinities it was expected to suggest that the media may have inhibited its growth aside from the salinity factor. Growth at 3.5% salinity in at least one of the replicates was considered an indicator that a strain was Halotolerant (HT) with growth at that salinity in two or more replicates being definitive. Growth at 2.5% salinity in at least one of the replicates was considered an indicator that a strain was slightly halotolerant (SHT) with growth at that salinity in two or more replicates being definitive. Isolates which did not show growth at 2.5% salinity were considered Not Halotolerant (NHT).

The results for replicate one, **Figure 3.37A** show a range of halotolerances across the different isolates. Immediately noticeable was the failure of 18D to grow on any salinity percentage media, this is in stark contrast to 41C which grew - even if it then showed a sharp decrease - in 5% salinity media. Interestingly the two *Apodemus* isolates, S122 and S128, show growth at a few salinities up to 2.5%, though the growth rates and stability of the population is not the same at this salinity level between the two isolates. 5 isolates show definite growth at the 3.5% salinity level, 13D, 13E, 16aC, 16bD and 41A; indicating they are capable of growth in media with the same salinity as seawater. 41B appears to have a lower halotolerance than the other isolates which managed to grow at 1% salinity; as the maximum salinity it managed to grow at was 1.75%. 13D, 13E, 16aC and 16bD are from different *Acomys* species, *A. russatus* for the first two and *A. russatus* for the last two, have broadly similar growth curves at the different salinities up to 3.5%. All of the *Acomys* isolates tested with the exception of 41B managed to grow at salinities up to 2.5%.

The results for replicate two, **Figure 3.37B** show some similar results to the first replicate. The positive control appears to show some growth at 0% salinity but this may be a false reading based on the results from the other two replicates. 18D also has a growth curve for the 0% salinity, but again this is an outlier for this isolate compared to the other two replicates; though the very slight and very late uptick at the 1% salinity might indicate that more time would have shown some growth at this or 0% salinity. The two *Apodemus* isolates in this replicate show definite growth up to and including 2.5% salinity though at different rates and with different stable population levels. Both also seem to show growth at 3.5% salinity as well, though to a significantly reduced degree compared to their results for 2.5%. The 41A results are similar to those from the first replicate, with wildly fluctuating readings but seeming to suggest there was some growth, though whether this led to a stable population or not is unclear, at 5% salinity. 41B shows some growth at the slightly higher salinity of 2.5% in this replicate though it takes longer to be established and to start reaching similar levels to that obtained at 1.75% salinity. 13D, 13E, 16aC and 16bD have quite similar results to those from the first replicate, again all managing to grow in the 3.5% salinity media though in the case of 13D this appears to have ended and seen a rapid decrease in the population after initial rapid growth. The three others show rapid growth at 2.5% salinities and the two *A. russatus* isolates maintain a stable population at 2.5% while the two *A. cahirinus* isolates show rapid growth in the population at this salinity followed by a steady decrease. 41C shows similar results to the first replicate, with those salinities it managed to grow at showing rapid growth followed by equally rapid decreases in the OD.

The results for replicate three, **Figure 3.37C** are broadly similar to the other two replicates.

13D in this replicate differs in that it takes much longer to achieve any growth at 3.5% salinity than in either of the two other replicates and at the end of the experiment was showing a much lower growth rate. The positive control does show some slow and late growth at 0% in this replicate, thoguh at a later stage than the result in the second replicate. The *Apodemus* results in this replicate are very similar to those from replicate 2 and quite similar to those from replicate 1, the results for the *A. russatus* isolate 41B are also more similar in this replicate to those from the second replicate than those from the first.

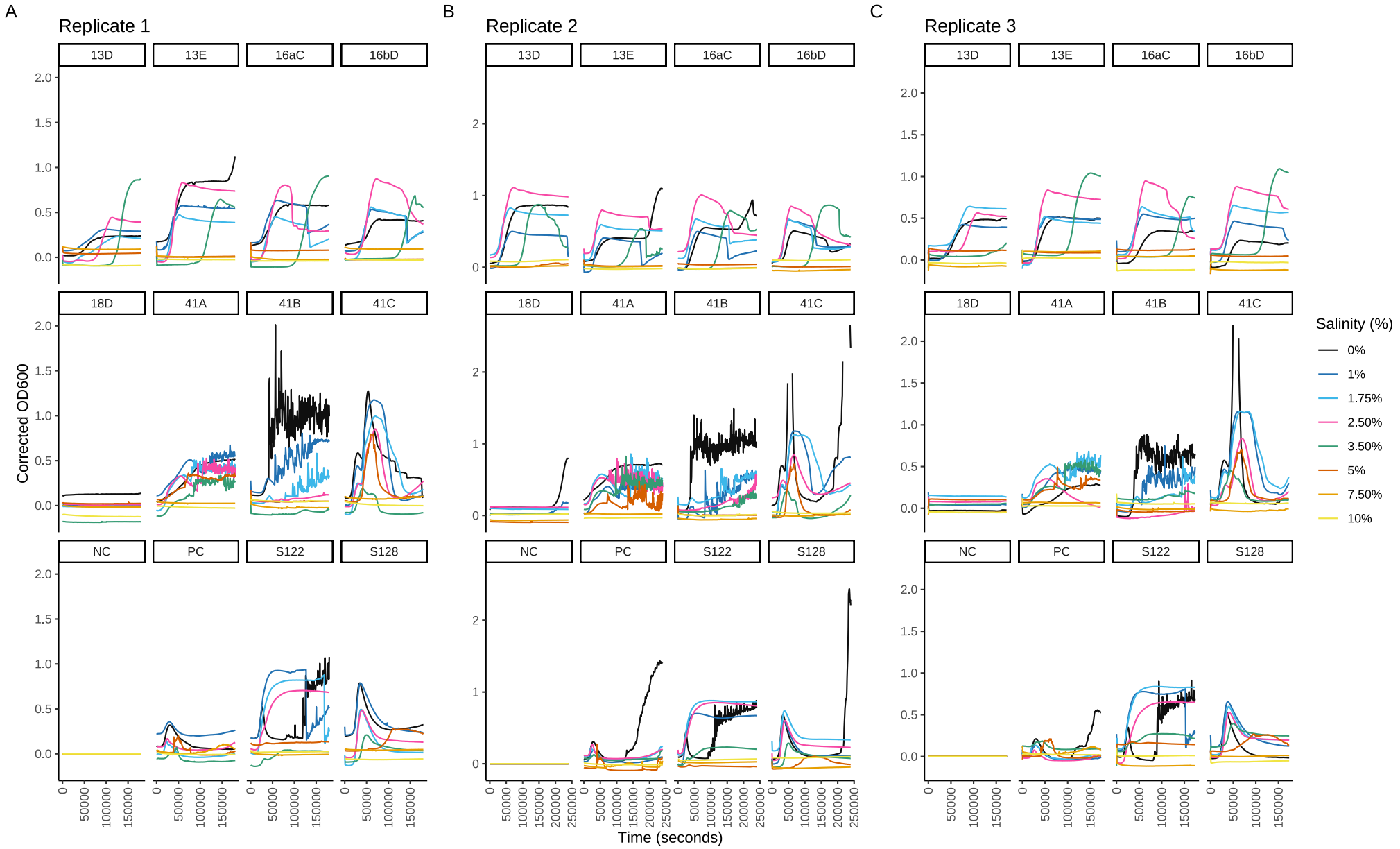

Figure 3.37: Line plot showing change in corrected OD600 reading over 48 hours per strain during growth in MRS+Cys media with variable levels of salinity. Isolate IDs are shown in boxes above each individual plot, NC: negative control, PC: positive control. **A.** Results for replicate 1. **B.** Results for replicate 2. **C.** Results for replicate 3.

Chapter 4

Discussion

- The results of taxonomic classification of the *Acomys* microbiota are discussed
- The taxonomic composition of the microbiota is discussed in the context of host species and sampling month variations
- The identities and nature of metagenomic bins produced from the *Acomys* microbiota are discussed
- Potential links between the microbiota and host tolerance of aridity are suggested from the results of bioinformatic analyses
- The isolation and culturing of Lactic Acid Bacteria from *Acomys* faecal samples is discussed
- The halotolerance of a number of these lactic acid bacteria in the context of host aridity tolerance is highlighted

4.1 Taxonomic classification of read files

4.1.1 Taxonomic classification for tool testing

The results of analysing the raw and processed read files using the four classifiers suggest that there may have been an impact in the ability of Kaiju and mOTUs to classify the reads from the files after processing, though this did not appear to be an issue faced by Metaphlan 3 or Kraken 2. This could indicate that there are concerns in the use of data from Kaiju and mOTUs if the workflow used to prepare the files prior to analysis impacted the capacity of the tools to fulfill their function. Given, though, that the difference in the case of Kaiju was an increase in the percentage of reads classified, this was not believed by the author to be an issue which would preclude its use. Their respective interactions with the project data, discussed later in this chapter, already led to the decision to not use them for meaningful analysis.

The results from the Metaphlan analysis of the human mock microbial community, classifying more reads than are present in the sample indicate some of the additional difficulties when trying to employ bioinformatics tools generally. The authors of Metaphlan make clear that by the nature of its method it cannot give a precise number of reads assigned to each taxa it detects. The 'UNKNOWN' classification or grouping is '...fraction represents the proportion of unclassified reads that I cannot assign to any microbial organisms present in the Metaphlan database, so any other DNA present in your sample, e.g. unknown microbial species, host DNA, food-related DNA, etc... will be included within that fraction' [441]. It is unclear why Metaphlan appeared to categorise taxa and their estimated number of reads multiple times, leading to the estimated number of Bacterial reads exceeding the actual number of reads in the sample while still also appearing to classify 84% of the reads as 'UNKNOWN'. It may potentially be related to Metaphlan's mechanism in using the average genome size of each clade for calculating abundance. Possibly, having a small number of strains - 20 - in the mock microbial community with a limited number of initial abundances issues arose when estimating read counts for the clades. The nature of Metaphlan and the database it employs make it well suited for taxonomic analysis of samples from well-studied, especially human, microbiota. This is evident in the stark difference between the results for the *Acomys* species samples and the *Mus* samples. The need for the markers to be unique results in a more limited database for Metaphlan, which allows the tool to be very precise when working with samples containing well characterised microorganisms. Metaphlan versions have been used in a large number of human studies [442, 443, 444] as they can take advantage of the relatively large amount of human-associated microbial community data already collected; though the tool is not restricted to human-centric investigations [445].

Prior to carrying out the analysis, the author expected Metaphlan 3 to perform similarly with the *Mus* samples as with the human mock, though the results for the human mock community are unreliable due to the assessment metric used it was a surprise that the *Mus* samples had a range of 29-44% of reads classified. However, this result is considerably better than the *Acomys* samples, where it is quite low. Mice (*Mus* species specifically) are one of the most well-studied organisms in metagenomics experiments and the inclusion of lab mice samples in the investigation was intended to provide a non-human reference point for a high level of classification. The samples themselves came from projects which had been specifically chosen as using mouse faecal samples for metagenomics investigations and therefore were not likely to contain a large quantity of previously uncharacterised microorganisms. Independent of the results for the *Mus* and human mock microbial community the poor level of classification of the *Acomys* samples highlights the risks of using a marker-based method of identification with samples from less-studied environments. This risk should reduce with time as databases expand and more microorganisms are

identified, sequenced and isolated. At present a marker based classifier would be a poor fit for an investigation in which it was anticipated that samples would contain a plurality if not significant majority of taxa which are outside the marker databases. This project did not investigate the ability of Metaphlan 3 to distinguish between strains in the human mock community, or the number of false positives it generated as the concerns around the 'UNKNOWN' reads led doubt about the reliability of the results. Nor would Metaphlan 3 be used on the *Acomys* reads in later stages of the project in light of the low classification of reads from those samples so no further testing of Metaphlan 3 took place.

The results for mOTUs are unsurprising in light of the results from Metaphlan 3, as mOTUs operates on a similar principle using matching to particular marker sequences and comparison to a reference database. Looking at the human mock community mOTUs did assign non-zero relative abundance to (i.e. detected members of) all of the 17 genera in the sample, however it also detected an additional 10 genera not in the known mock community and failed to detect two of the species known to be in the mock. It did assign relative abundances to the 17 genera in general accord with the different ratios of the species present in the mock community, indicating a degree of ability to resolve different actual abundance levels in a sample. Looking at the rodent pilot samples it is clear at the phylum level that mOTUs detects a considerable difference between the *Acomys* and *Mus* samples, none of the *Acomys* samples have anything other than Firmicutes as the most abundant phylum whereas in the *Mus* there are two samples with Bacteroidetes as the most abundant phylum. mOTUs has been used successfully by researchers [446, 412, 447] working with a number of sample types but was decided to be inappropriate for the *Acomys* samples in this project.

Kraken 2 had limited success in classifying reads from all rodent samples, both *Mus* and the two species of *Acomys*. Even when using the most permissive minimum confidence score setting, 10%, in the majority of *Acomys* samples Kraken 2 classified less than 10% of the reads. A low level of classification with the *Acomys* samples at higher minimum confidence scores, i.e. 50% and above was not unexpected as Kraken 2 is still dependent on a database of genetic information which suffers from the biases described earlier towards human-associated and environmental samples. The database is unlikely to contain taxa associated with *Acomys* specifically or arid-adapted rodents more generally. The fact that the percentage of reads classified in the *Acomys* samples decreased with the increasing minimum confidence score is therefore not unexpected, however the very low classification level in almost all samples at even 10% minimum required confidence was surprising. It might have been expected that Kraken 2 would struggle to place reads into lower taxonomic levels at any minimum confidence score, even the most permissive level of 10%, but completely failing to classify more than 90% of reads at this level in most samples is was not. Kraken 2 being unable to place the majority of reads in the *Acomys* samples into a phylum suggests that the intestinal microbiota of the *Acomys* is very different from well-known and characterised microbiomes; which comprise the bulk of material in the database it employs.

The observation of low level results for the *Mus* samples was unexpected. As with Metaphlan, the author expected that while Kraken 2 would not be able to classify as many reads in these samples compared to the human mock microbial community it would still achieve relatively high classification percentages. A range of 50-90% of reads being classified to at least the phylum level was anticipated. For the human mock community, the largest difference between the percentage of reads classified between the raw and processed file was with Metaphlan. Given the previously discussed issues with the estimated percentage classification from Metaphlan this result cannot be relied upon. For Kaiju the difference in the percentage of reads classified in the

human mock from raw to processing was 0.055% and for Kraken 2 it was 0.64%. The author does not believe that the processing had a measurable impact on the percentage of reads classified by the classifiers at the most permissive settings. Interestingly, the results of the Wilcoxon tests carried out on the rodent samples indicate a difference between the average percentage of reads Kaiju classified in the *A. russatus* samples with a p-value of 0.0087, with the processed files having slightly more reads classified as a percentage on average. This might be a sign that processing of the read files did lead to the alteration of how many reads could be classified as a proportion of the whole, though the author expects this to be a result of the subsampling rather than the quality control and contaminant removal. This also may not be the case for the stricter settings used on the processed files, during the testing process.

The comparatively similar results between the laboratory and wild *Mus* samples may also suggest the same trend as the low level of classification for the *Acomys* results within the database Kraken 2 uses for its kmer-matching classifications; that it lacks wild rodent derived samples. It would be interesting to determine the overlap between the taxa classified in the *Acomys* and *Mus* samples to see whether these were shared with the typical (i.e. most often studied) human and environmental microbiota. This may explain their being detected by Kraken 2 - especially if these taxa were detected at all confidence levels including the strictest. That the *Acomys* samples for each of the two species which had the highest percentage of reads classified by Metaphlan, AC16N and AR27N, were also those with the greatest percentage of reads classified by Kraken 2 at all minimum confidence scores does somewhat support this. It indicates a component of the microbiota in these individuals was a greater proportion of previously detected and characterised microorganisms. Kraken 2 has been employed in a number of studies looking at a variety of metagenomic samples from animal, human and environmental sources [448, 449, 450] - often it is used alongside other classifiers; potentially including Metaphlan, Kaiju or both. Investigators can make use of custom databases for Kraken 2 (and for many other taxonomic classifiers) which may yield a greater proportion of reads classified. The author did not choose to employ this strategy here, both as it was intended as a benchmarking exercise to establish some thresholds for poor, moderate and good classification of reads from samples and as the author was uncertain what would warrant inclusion in a custom database for the *Acomys* samples due to a lack of prior arid-adapted rodent microbiota studies. Potentially a rodent-centric database could be created by combing the literature for genomes or assemblies isolated from different rodent samples and processing them to produce a unified database of taxa known to be found in other rodents. Alternatively the reads could have been mapped to reference genomes for taxa known to be members of the gut microbiota of other rodents. Whether these approaches would lead to a greater proportion of reads in the *Acomys* samples being classified is debatable, though it could well lead to more reads in the *Mus* samples being successfully classified at all minimum confidence levels. This though does create a concern that the results could be skewed towards whatever fraction of the *Acomys* microbiota was most similar to that of previously studied rodent species.

The human mock microbial community results were much closer to those expected when using Kraken 2. More than 90% of reads classified at minimum confidence scores of 50% and below and even 89% of reads classified at 66% minimum confidence score. The most interesting finding was the large drop in percentage of reads classified from 76% at a minimum confidence score of 75% down to 14% of reads classified with a 95% minimum confidence score. This demonstrates the unsuitability of using this high threshold with the standard database and samples from less-studied environments. The human mock microbial community contained only 20 strains and was based on taxa known to be associated with the human microbiota. Less than a fifth of reads could be classified from what amounts to an almost ideal sample at 95% minimum required con-

fidence. This is an indicator that the low levels of classification of reads in the rodent samples at 95% confidence score may not be due to their composition being primarily novel taxa but instead the result of Kraken 2 behaving conservatively, in accordance with the required confidence level. This is supported by the drop in both the known strains detected and the reads classified with the taxIDs of the overall species themselves with the increasing minimum confidence score required. When considering the investigation, The author determined that both the NCBI taxonomic IDs of the 20 strains and those of the species to which they belong would be considered accurate. This is in the context of the overall project in which the author believed it unlikely there would be species level data for many *Acomys* microbiota members let alone strain level discriminatory data. As such Kraken 2 never detecting all 20 strains is not necessarily a sign of it being unable to classify reads accurately in the human mock community, as it detects the 19 different species at the, more permissive, two lower minimum confidence scores. It is instead an indicator of a lack of precision. Given the kmer-matching and LCA based mechanics of Kraken 2 this is an inevitable feature. If one were to remove the species taxonomic IDs from the database and re-run the analysis it is likely that more of the strains would be detected and there would be significantly more reads assigned to the strains. As this step of the project was in part meant to determine parameters to use with tools in latter stages the author chose a range of minimum confidence scores to assess. The author took advantage of the mock community to compare accuracy and precision of different parameters. A minimum required confidence score of 25% saw Kraken 2 detect all the species type taxonomic IDs of the 20 known strains, a sharp drop in the number of taxIDs detected at all (almost entirely false positives though some true strain level IDs are lost), a drop in the number of incorrect strain IDs detected and only a marginal increase in the number of reads classified outside one of the correct genera or left unclassified. Though this setting could be considered inappropriate for a well-characterised sample, like the human microbiota, it appears a reasonable compromise to use with samples in which there is an expectation of poorly characterised microbial taxa.

Kaiju appeared initially to be a much more capable classifier than either Kraken 2 or Metaphlan, classifying the majority of reads in all samples analysed. This was in line with the author's original expectations as to the ease of classifying reads in the human and *Mus* samples but was quite surprising for the *Acomys* samples. Though a minimum classification of 70% of the Mus reads was lower than might have been expected for samples from frequently studied environments it was still closer to the values anticipated. Given the small size (in terms of taxa) of the human mock community the author was not surprised that Kaiju could classify over 99% of reads at all error allowances. However, the results from the human mock microbial community when considering the number of false positives go some way to explaining the high level of classification in the *Acomys* samples. Even at the strictest error allowance setting Kaiju detected thousands of taxonomic IDs in the mock community, far in excess of the 20 strains, with the number of taxa detected increasing rapidly with an increasing number of permitted errors. The authors did not have a controlled community to compare the results of the *Acomys* classification to, in order to determine exactly what amongst the taxa Kaiju detected in them are likely to be false positives. It appears reasonable that the apparent high number of false positives may be part of the reason it had such seeming success classifying the taxa in the samples. Kaiju is a taxonomic classifier which has been used in studies investigating a wide range of metagenomic sample types [451, 452, 453], it was included in this stage as it offered a different system for classification than the marker or kmer based methods employed by Metaphlan. mOTUs and Kraken 2. As the author used the raw output files rather than any of the reports or summary files for Kaiju and Kraken 2 to measure the classification of reads and the assigned taxIDs the author bypassed methods the tools have for filtering output when generating report or summary files. Kaiju for instance can exclude taxa which do not meet a particular threshold for number of reads from reports. This is a

sensible step to take when working with samples in which the composition can be predicted to an extent. It would be a reasonable filtering step if you had prior expectations that a metagenomic sample contained a small number of taxa in high abundance and few with low abundance; the latter being difficult to distinguish from noise. As the aim of this stage was to compare results using different classifiers with default settings on samples about which there was no prior knowledge of likely composition, it was felt important to use the raw output.

From the Kaiju output, in spite of the large number of false positives it is possible to discern some interesting findings. The percentage of reads classified increases most significantly in the rodent samples from an error allowance of 0 to 1, with diminishing returns from 1 allowed error onwards. This is especially noticeable when looking at the doublings of allowed errors, 1 to 2, 2 to 4, 5 to 10 and 10 to 20. In the *Acomys* samples there is only minor change in the percentage of reads classified in the latter doublings - with this being most apparent in the 10 to 20 doublings. The same holds true in the *Mus* samples where doubling the allowed errors from 10 to 20 does not meaningfully change the median percentage of reads classified in each sample; in the human mock community this is the case as well. This suggests that even with Kaiju finding a huge number of false positives there are a number of reads which cannot be classified, potentially they are different enough from anything in its database or were degraded by the collection and storage steps such that Kaiju cannot classify them even when a relatively large number of errors in the match are permitted. As the author does not have the ability with the *Acomys* and *Mus* samples to distinguish between false positives and true but rare taxa the human mock community was used to try and determine what would be an appropriate error allowance to classify present taxa whilst minimising the number of false positives. Kaiju detected all of the species of the 20 known member strains at all error allowances, and all but one of the exact strain taxIDs with all error allowances - a superior result as compared to Kraken 2 which at its most permissive failed to identify 6 of the known member strains. That the number of unique taxIDs detected is the same as the number of taxIDs with more than 100 assigned reads is both an indicator that the author might have chosen too small a threshold to use and that Kaiju assigned at least 100 reads to all false positive taxa. Looking at the number of reads assigned it is clear that with the exception of the strictest setting, 0 allowed errors, Kaiju did not see major differences in the number of reads assigned to taxIDs aside from the exact strain or species. Halving the number of allowed errors from 20 to 10 did not lead to a halving in the number of reads assigned to taxIDs outside any of the correct genera. Nor did halving it again from 10 to 5. This suggests that Kaiju readily assigns reads to false positive taxa as if it had made an incorrect choice between a false positive taxID and a true positive one then one would have expected that halving the number of allowed errors would lead to an increase in the number of reads assigned to a true positive taxID; of a noticeable amount. Kaiju did not only assign the majority of reads to false positives it also detected far more unique false positive taxIDs as compared to Kraken 2, which indicates that even the strictest error allowance would make it impossible to distinguish between true signal and noise.

4.1.2 Taxonomic classification for analysis

It is apparent from the results of the attempted taxonomic classification of the read files from the two *Acomys* species that they contain microbial diversity not contained within the reference databases employed by the tools which do not employ translation and protein databases. This is not unexpected as this project is, to the author's knowledge, the first investigation of any kind of the intestinal microbiota of the genus *Acomys*. The limitations of the four classification tools employed here have been discussed in greater depth in the preceding subsection. The results from a greater number of *Acomys* samples than the number used for the tools testing stage yielded extremely similar results. Neither the initial sequencing depth, nor the processing of the read

files caused the low percentage of reads classified - though it did seemingly increase the ability of Kaiju to classify the reads. Given the results from the prior chapter with Kaiju and a community with known membership, the author does not believe any of the four taxonomic classification tools used with the faecal samples from either *Acomys* species provided more than a general overview of composition, with Kaiju making false positive classifications and the other three tools both producing false positives and classifying either a limited number of the reads or quite limited diversity of taxa.

The low level of classification by Metaphlan is likely a consequence of both taxa present in the sample for which it has no marker sequences in the reference database and the inevitable impact of various extraction, sequencing or contamination events on the sequencing of the specific markers. For mOTUs it is difficult to determine whether the percentage of reads classified or not is a sign that the data was well suited for it or not. The tool itself uses only a small subset of reads to determine the relative abundance of taxa it detects, however in the absence of prior *Acomys* microbiome studies the author cannot use this metric to assess how suitable mOTUs would have been in this instance. This is unfortunate as one of the primary steps of any investigation into a microbiota is determining what microbial taxa are present in the environment. Indeed, the basic composition may sometimes be the only information reported depending on the nature and available resources of the researchers. At the genus level Kraken 2 and Metaphlan have largely similar detections. Of the 17 genera detected by Metaphlan which it finds in more than one sample, Kraken 2 also detects 13 of the same genera. The similar, high, proportions of reads unclassified between Kraken 2 and Metaphlan is despite the Kraken 2 database taxonomic IDs ranging from the phylum level down to individual species. In theory this taxonomic depth should provide a great range of potential matches for the tool, in combination with its scoring mechanism, to classify reads to. When looking at the taxa assigned at least 2.5% of reads in one sample by Kraken 2, they occupy a large proportion of the reads classified by Kraken 2 despite their low percentage values due to the extremely low percentage of all reads which Kraken 2 could classify in each sample. Though Kraken 2 achieved a very low level of classification, based on the findings from the tools testing stage discussed in the previous subsection the author is satisfied that it does allow the reliable detection in the sample of the taxa it detects barring any biological contraindication. The fact that some reads were only classified to the level of '`cellular.organisms`' does highlight the limits of trying to use the results to conduct anything more than surface level analysis of the very small percentage of reads that Kraken 2 could classify. That said, it allows for an overview of what taxa are changing within the hosts overall, with the caveat that this is only for those taxa it could detect.

Before speaking about the results from the three classifiers which classified at least 1% of the reads from the samples on average it would be useful to discuss the results from mOTUs. As stated on multiple occasions previously, the tool uses a subset of reads in the sample rather than attempting to process all reads in a sample. Just from the individual reference mOTUs, as can be seen in **Supplemental Table 5.3** the individual species or genera associated with the reference mOTUs which had a summed relative abundance of at least 1% in a minimum of 1 sample in either host species align with the detections from Metaphlan, Kraken 2 and Kaiju. That *Lactobacillus kefiranofaciens* had the greatest summed relative abundance from all samples within each species mirrors the results from the other three classifiers for this bacterial species, as do the comparatively high summed relative abundances for the *Muribaculaceae* and *Akkermansia muciniphila*. Looking at the relative abundance results which are provided by mOTUs, the phylum level results are broadly similar across all samples and consistent with the results from the pilot study, Firmicutes being the phylum with the greatest relative abundance in all samples. The mOTUs results also indicate that the non-Firmicutes and non-Bacteroidetes component of the

Acomys microbiome is of restricted diversity, at the phylum level at least, with only Verrucomicrobia and Proteobacteria being detected in more than one sample with a relative abundance of 0.1 or greater. These findings are also similar to previously reported findings for different rodent species gut microbiome phyla compositions [454, 455]. The genus level results are in keeping with those from the three other clasiffication tools, with *Lactobacillus* being the genus most likely to command a plurality, if not majority, of the mOTUs assessed relative abundance in a sample. The genus level results do not show the same pattern of within-host species differences in relative abundance as seen with the equivalent measures from the other classifiers, as mOTUs only finds two genera with statistically significant abundances between sampling months within either host species. *Bifidobacterium* was found to have a sampling month difference which was statistically significant in *A. cahirinus* by Kaiju but this difference is instead found within *A. russatus* by mOTUs. *Lachnospira* was found to have a within host species difference in the relative abundance by mOTUs for both host species but this was not detected by either Kraken or Kaiju; though Kraken did find a within *A. cahirinus* difference for Lachnospiraceae.

The data provided by the three classifiers which managed to classify at least 1% of reads on average do allow some higher level conclusions to be drawn, notwithstanding the inherent limitations of each tool. All three show that there are taxa present in the samples which have previously been reported in the intestinal microbiota of both other mammalian species and from rodents in particular. Of the two tools with limited classification ability, many of the reads they can classify are assigned to taxa corresponding to Lactic Acid Bacteria. It is important to note that the *Lactobacillus* genus has been reorganised since the database used in the Metaphlan execution was produced and some species may no longer be members. The large proportion of reads which could be classified being assigned to Lactobacillaceae informed the decision regarding subsequent isolation and culturing work. Kaiju also assigns a number of reads to different LAB taxa, though distributes them out over many more genera and species. Some of the classifications hint at potential functional roles, i.e. *Lactobacillus* members are detected by all three tools and likely point to the presence of fermentative bacteria in the *Acomys* GIT - and that of the fraction of the microbiota which the tools can classify it is not too dissimilar to other known rodent intestinal microbiota proportions of LAB [456]. These results also could provide some idea of what potential taxonomic diversity is lacking from the reference databases used by Metaphlan and Kraken 2 and therefore represent novel microbial diversity from what has previously been identified and contributes towards the databases. Those reads which could not be classified by Metaphlan and Kraken 2 are the majority for each sample, whereas those which could not be classified by Kaiju are a much smaller fraction - though some reads may only be classified at the phylum or even kingdom level and so not particularly informative - and dedicated work purely with these reads could produce MAGs of novel taxa. In spite of the limitations in the three classification tools, the results produced mean it is possible to look at changes within the two host species in the detection of certain taxa.

The results for Kraken 2 here may be more accurate compared to Kaiju, though they represent a much smaller proportion of all the reads in the sample files these tend to be the same taxa detected. As such any changes observed are more likely to be a reflection of actual biological reality, i.e. the change in geometric mean percentage reads classified for Firmicutes by Kraken 2 is less likely to be due to changes in false positive classifications into the phylum than with Kaiju. Looking within the *A. cahirinus* samples and at Kraken 2 results, when requiring a minimum confidence score of 50% it is apparent that there is an increase in the presence of members of Clostridia from the June to November samplings; both due to the change fold change in the geometric mean percentages classified for Clostridia itself and from the change in both Eubacteriales and Lachnospiraceae which are lower level taxa within Clostridia. The increase in

the geometric mean percentage of reads classified to *Lactobacillus kefiranofaciens* from June to November is at odds with the decrease in the same measure seen for the genus *Lactobacillus* itself, along with the higher level taxa Lactobacillaceae and Lactobacillales which suggests that either a species closely related enough to *L. kefiranofaciens* survives the change while relatives don't or that it is an artifact. The author believes the latter is more likely, that there is a general decrease in the abundance of members of the Lactobacillaceae in *A. cahirinus* from June to November. It is unlikely that a single species within the Lactobacillaceae is sufficiently different from its close relatives to survive a change the others do not. Whether this is a general trend in LAB or purely within this family is impossible to determine from the Kraken 2 results due to the low level of classification of the sample reads overall. Within the samples from *A. russatus* there is a similar pattern of decrease seen in members of the Lactobacillales from June to November, though without an outlying species level result suggesting the opposite. The *A. russatus* samples also suggest an increase in the abundance of members of the Clostridia from June to November, though lacking a family level classification like Lachnospiraceae to buttress this point.

Due to the large number of false positive results obtained when running Kaiju on an artificial sample with known composition described in the previous chapter, only those results from Kaiju analysis with an error allowance of 0 were reported in detail. It is possible that a translation-based method may be better suited to samples such as the ones used in this investigation, from animals which have not previously been the subject of a metagenomic or microbiological investigation and so are poorly represented in taxonomic databases. It is notable that the level of classification drops with Kaiju when looking at lower taxonomic levels, such as genera; which may support the results being true rather than false positives with the majority of reads classified at higher taxonomic levels rather than more informative lower ones. Looking at the genus level results for Kaiju when allowing 0 errors in the match one can observe some possible trends within each of the host species for change over time. In *A. cahirinus* there were a greater number of genera enriched in November as opposed to June, this also being the case in *A. russatus* albeit with fewer genera overall having a statistically significant difference and meeting the magnitude threshold. *Helicobacter* and *Akkermansia* were enriched in the June samples in both host species, as was *Thiopseudomonas*. Both host species saw an enrichment of *Bifidobacterium* from June to November, along with *Ileibacterium* - both genera having been isolated from gut microbiota sampling previously [457, 458]. Due to the low level of total classification obtained by Metaphlan and Kraken 2, along with the previously discussed high error rate of Kaiju and its limited classification levels at lower taxonomic levels, one can only draw higher level conclusions from the data.

It appears likely that there is considerable novel microbial diversity within the faecal microbiota of both *Acomys* species. Whether this is at the family or lower levels, or at higher taxonomic levels like phyla and classes cannot be determined from these results - Metaphlan and Kraken 2 suggest the latter while Kaiju suggests the former. Within the confines of the ability to study the intestinal microbiota using taxonomic classifiers it seems that there is some change within each host species from June to November. This may be related to potential dietary changes, host physiological changes or purely an artifact of the data - a larger sample size would have provided more data which could support or disprove the temporal changes. It is also important to note that there are only two time points, that any changes in the classification of taxa by any of the tools which suggest differences between the two sampling points can be dramatically influenced by variation on the shorter term. It is possible that the individuals were sampled when their intestinal microbiota was in a different state than typical, from illness, dietary composition or other external factor. This then would create a false impression of the typical microbial community composition for that month in the host species. The author believes that the use of the geometric mean

percentages of reads classified helped address this problem, but it is still quite possible that a sample from a given individual is not a true reflection of the normal state of its gut microbiota for that month. It is also possible that the different number of samples within each host species for the two months impacts the general trends discussed above. Due to the limitations of the taxonomic classifiers the author elected not to subset the data to only those sequenced with Novaseq platform (and so with a greater read depth even after subsampling) and for which there were paired June and November samplings; as opposed to the mapping based approach discussed later. It is possible that restricting the analysis to this subset of the data would have eliminated any of the June and November differences discussed earlier even if they are true reflections of biological changes and not consequences of the flaws in the classifiers.

Prior work by Maurice *et. al.* [250] has shown a seasonal change in the mouse *Apodemus sylvaticus* with a change in diet from mostly insects to predominantly seeds. Although the environment in which it lives is quite different from the *Acomys* populations in this study, it is a rodent which has a change over time in the gut microbiota. In Maurice *et. al.*'s case they can directly associate the change with a known and observed change in diet which correlates with the seasonal change; the classification results cannot directly confirm a similar change with the *Acomys* species. The authors also note the large proportion of the *Apodemus sylvaticus* gut microbiota made up of *Lactobacillus* members; detected by all four classifiers in the *Acomys* samples. The authors report a decrease in the relative abundance of *Lactobacillus* from spring and early summer to late summer and autumn; the results from Kraken 2 suggest a potential similar result in the *Acomys* species. Later work on the same species, *Apodemus sylvaticus* by Marsh *et. al.* [459] found similar higher level taxa within their samples as both the earlier work on *Apodemus sylvaticus* and those detected in our *Acomys* species by all four classifiers; namely Lactobacillales, Bacteroidales and Clostridiales. The authors also found a temporal effect on the microbiota, seeing the relative abundance of Ruminococcaceae increase in September to November versus other time points in the year, Muribaculaceae decreased over the same time and Lactobacillaceae fluctuated in variable patterns. The authors also suggest a likely dietary cause associated with the change in season for the microbiota composition altering and note that it was consistent over multiple years of study. The authors note that the families which changed between measurements as the year progressed differed between their two wild populations being studied and propose that this is a sign of functional redundancy in the intestinal microbiota; with the same stimulus triggering responses in different taxa. This may potentially explain the differences seen in the genus level Kaiju results for the two *Acomys* species, for genera with changes in their geometric mean percentage classified from June to November.

4.2 Metagenomic bin identities and phylogeny

4.2.1 Taxonomic identification of metagenomic bins

A total of 348 bins passed the quality threshold of ≤ 5% contamination and ≥ 80% completeness, as measured by CheckM. The author selected these thresholds after review of the literature revealed a range of values often employed with no uniform standards. The 348 bins provide an interesting contrast in their taxonomic classification when compared to the read results from all 4 classifiers. Only 6 bins were classified by GTDB-Tk in the Lactobacillaceae, seemingly at odds with the results from the classification of the reads. This however may be a good indicator that the very low level of classification by Metaphlan and Kraken 2 leads to a focus on the small fraction of taxa they can detect, especially in the case of Metaphlan. It is also worth noting that the simple count of the number of bins assigned to a given genus or family is not a direct measure of how abundant they are in the samples, it is quite possible that there is a family or genus which had only a single bin assigned to it by GTDB-Tk but which is extremely abundant in the microbiome. The phyla detected and the number of bins assigned to them is not unusual, especially the relatively high numbers of bins assigned to Firmicutes (and Firmicutes_A) and Bacteroidota (otherwise referred to as Bacteroidetes) which together contain 90.5% of all the strict bins. These phyla are well known members of the intestinal microbiota of rodents and a wide range of mammals [460], and typically make up the majority of members of the intestinal microbiota in these animals under normal circumstances [461].

The family level GTDB-Tk classifications are informative as they may provide a greater insight into the taxa present in the *Acomys* species. The reads were co-assembled, so some bins may not be a direct equivalent of a genome present in the samples even when discounting mis-assembly or incorrect binning, though the mapping results discussed later will enable better insight into this. At the family level the majority of the bins are classified into a small number of families, Lachnospiraceae, Muribaculaceae, Ruminococcaceae, Acutalibacteraceae and Desulfovibrionaceae. As might be expected given the large number of bins found in phyla commonly associated with other mammalian intestinal microbiomes, most of these families are likewise known to be members of intestinal microbiota from different mammalian species [462, 463, 464]. Both Lachnospiraceae and Ruminococcaceae have been established to contain Short Chain Fatty Acid producers, which have been linked to positive health benefits for the host organism [465] and point towards a place for fermentative bacteria in the *Acomys* intestinal microbiota. Some have been associated with disease and ill health across different studies and host species, including ironically some members of Lachnospiraceae [466]. Lachnospiraceae and Ruminococcaceae have also been found to see increased abundance and diversity in a wild rodent intestinal microbiota, so their larger abundance in the wild rodent samples examined in this project is not out of keeping with previous results [467]. Indeed they were the most commonly detected families in an investigation of wild *Mus* discussed in the introduction [285].

The very low number of bins which could be assigned a species (12, or 3.4%), along with the fact that for some bins the lowest available taxon from the GTDB-Tk classification was at the 'family' level may hint that the bins are lacking in some of the more distinguishing features needed. Though all met the 80% completeness threshold from CheckM analysis this may have been too low to allow for better resolution of the taxonomies - especially in enabling better partitioning into families. Apart from two bins (Bin_c732 / genus RC9 and Bin_c1055 / Lachnospiraceae) the bins which had a species level classification had completeness scores from CheckM over >90% - though some bins which could only be classified to the family level also exceeded 90% completeness.

In case there were a significant difference between the bins which met the quality control thresholds used and those which were rejected for not meeting them, to reiterate these were a maximum contamination of 5% and a minimum completeness of 80% as measured by CheckM, the rejected bins were analysed with GTDB-Tk. They were not processed in any other way as they would not be included in the main analysis in any manner. There were 156 bins which were rejected for failing the QC checks, from these there were 31 families and 71 genera detected by GTDB-Tk. 90 of the bins (57%) were classified into four families, Lachnospiraceae, Sacharimonadaceae, Muribaculaceae and CAG-508. There were no families detected in from the failed bins which were not detected amongst those which passed QC. At the genus level there were no genera detected from the failed bins which were not also detected in those which passed QC. None of the genera were assigned to more than 10 of the bins which failed QC, though as with those which did pass QC not all of those which failed could be assigned a genus level classification. It does not appear that those bins which failed the QC represented novel taxonomic content as compared to those which passed. **Figure 4.1** shows the number of bins classified by GTDB-Tk Core at either the family or genus level which either passed or failed the QC checks; showing only those taxa where there was a bin classified amongst those which failed the QC checks.

mOTUs does seem to capture a slightly larger proportion of the taxonomic diversity represented by the bins than the three other classifiers, as it assigns a comparatively high relative abundance to the genus *Lachnospira* - in more than half of samples it is the genus which has the second greatest relative abundance. These might well correspond to the bins classified by GTDB-Tk into the Lachnospiraceae. The tool does though overwhelmingly give the greatest relative abundance to the genus *Lactobacillus*, which indicates that it still is not accessing the taxonomic diversity represented by the bins; given that only 6 bins were classified into the Lactobacillaceae.

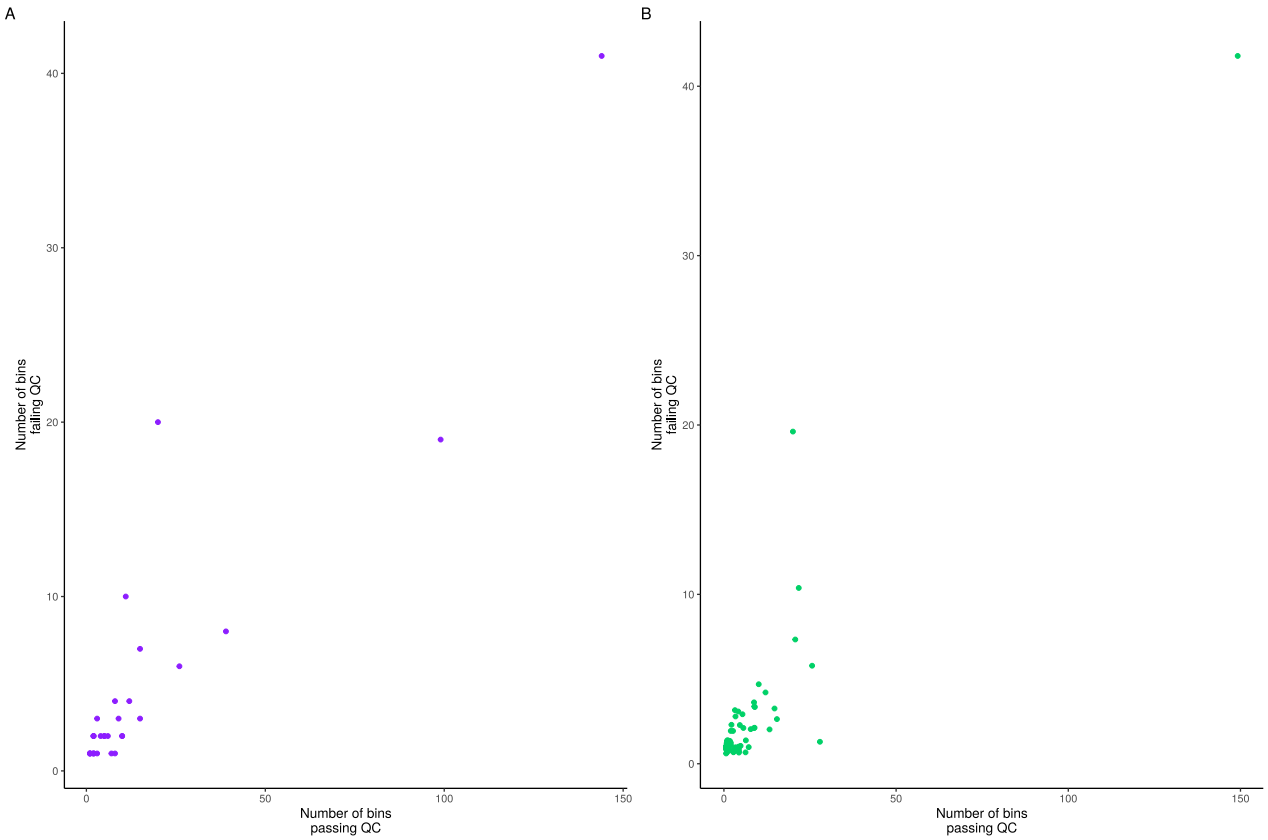

Figure 4.1: Point plots showing the number of bins which either passed of failed QC checks at the **A.** Family and **B.** Genus level from GTDB-Tk classification.

4.2.2 Phylogeny of metagenomic bins

Looking at the first phylogenetic tree it is clear that the bins and the assemblies are very distinct from each other, with only a handful of the bins found amongst the assemblies subtree. It is difficult to know from this tree whether this is a consequence of a fundamental difference between the bins and the assemblies associated with the different means of obtaining them - reads from pure cultures versus reads from a mixture of sources - or is due to real phylogenetic differences between the two data sets. As discussed earlier there are very few bins which were classified as LABs so assemblies produced from LAB sourced from specifically selective media might be phylogenetically distinct from each other in reality. Of the six bins within the assemblies subtree, four enriched in *A. cahirinus* and 2 not enriched in either, all were classified by GTDB-Tk within the Lactobacillaceae. The two assemblies which are apart from the rest were both classified as *Bifidobacterium*, the single bin which is mostly closely placed to them was classified as a member of Bifidobacteriaceae. The expanded tree, including the bins, assemblies and external data may help assess the likelihood of the placement of the assemblies apart from the bins being phylogenetically valid.

That the bins do not mostly cluster by enrichment status is in keeping with bins with enrichment in either host species and those not enriched in either being placed within the same families by GTDB-Tk. Though there are some small groupings of enriched bins within the same host species these are predominantly those enriched in *A. cahirinus* and may reflect more the larger number of bins enriched in the species. There are though only a few instances where a bin enriched in one host has a bin enriched in the other host species as their nearest neighbour. As an example, Bin_m1549 and Bin_c793 are enriched in different host species, are the closest

neighbours to each other being terminal nodes on the same branch and are both classified as Desulfovibrionaceae. Within the bins classified as Muribaculaceae there were two pairs of bins which were enriched in the opposite host species and were the end nodes of the same branch.

The split of the Acutalibacteraceae in the tree into three distinct sections, one quite distant to the other two, is interesting. Especially given that the Ruminococcaceae splits the two sections from the third. This suggests one of two possibilities, either that the classifications of at least some of the bins are incorrect or that there is some overlap between the two families which would be beyond the scope of this project to investigate. Lin *et. al.* [468] found some of their MAGs produced from dariy cow inestinal samples were classified as Acutalibacteraceae by GTDB-Tk so there is precedent for constructed sequence files from intestinal microbiota samples being assigned to the family. The tree might indicate functional overlap for the two families, as both have been shown to interact with host mucosal biofilms [469, 470].

Figure 3.21 also shows a split in the Muribaculaceae in the tree. The two trees are due to the greater number of bins assigned into the family and the split itself is larger, containing multiple subtrees and with bins assigned to at least four different families within the split. Given that the split contains multiple different taxonomic families the author is inclined to believe that the bins of the smaller Muribaculaceae may in fact be a distinct family - especially as they are a distinct and cohesive seperate tree. The closest reference to this potential novel family in the GTDB-Tk database is presumably Muribaculaceae hence the classification. Alternatively this may reflect the previously reported high level of functional diversity within the family [471, 472].

4.3 Mapping of *Acomys* faecal sample shotgun reads to metagenomic bins

The level of mapping of the samples to the concatenated single bins reference was lower than the author had expected, especially as it was the approach decided upon given the difficulties accessing much of the sample content during classification the different tools; which the author consider to be the most reliable of the classifiers used. The arithmetic mean mapping percent for non-pilot samples was higher than the equivalent value for percent reads classified by Kraken 2 at a minimum confidence score of 50% so the author considers it a moderately superior method of analysis. Using Salmon and working with reads per million should help make the most of the results even with the relatively low level of mapping, as does the subsetting of the data. For the mapping based analysis, the author ceased using the entire set of sample reads and instead restricted analysis to only those samples which were sequenced with the NovaSeq platform and for which there were paired June and November samples - so within each host species the comparisons would be between the same individuals in June and November. This means the mapping depth of the read files was uniform after subsampling at a single value which should help make the comparisons more reliable by removing one potential factor varying between each sample - especially as removing the pilot samples allowed for a higher uniform read depth to be subsampled to.

The geometric mean was used to cope with the relatively large differences within the results for host species and sampling month for the bin RPMs calculated by Salmon. For example, the range of RPMs for Bin_m994 (genus UMGS268) just within the *A. cahirinus* June samples was 218 to 16,935. Using the geometric mean as the base point for comparison to look for statistically significant differences hopefully should reduce the influence of outlier samples within the groups being compared. The geometric RPM for Bin_m994 in the *A. cahirinus* June samples was 469.12. The initial step was looking for within host species differences, bins which had statistically different geometric mean RPMs from June to November and then looking for those where the fold change (log2()+1 change specifically) was greater than a given threshold; whether positive or negative. This was to try and identify a potential temporal change within each species individual before comparing the two host species as a group, and before comparing the two sampling months across the species. Coincidentally, there were no bins in either *Acomys* species, when doing the within-host comparisons, which met the Q-value for significance (≤ 0.05) but didn't meet the fold change threshold. Within the *A. russatus* samples there was a decrease in the geometric mean RPM from June to November of Bin_m137 (genus *Cellulosilyticum*), a genus within the Lachnospiraceae known to be able to degrade lignocellulose [473]. As discussed earlier when looking at the taxonomic classifiers run on the reads for potential seasonal differences, there were some differences observed between the June and November samples within each of the two *Acomys* species. It was also touched on then that prior research with wild rodents had reported seasonal differences in wild rodent microbiota and so the author considered it possible that this might be observed in the *Acomys* Without a reason to believe the seasonal change in diet, slight though it may be compared to some other studied rodents [474], would not have a similar impact on the *Acomys* intestinal microbiota the author believed it warranted consideration. The already discussed differences found with the Kaiju, and less reliably Kraken 2, results combined with the mapping based results indicate there was some seeming seasonality to the microbiota; taxonomically at least.

Two bins saw an increase in their geometric mean RPMs from June to November, these were Bin_c796 (genus CAG-590) and Bin_c1435 (family CAG-552). CAG-590 is also a member of the Lachnospiraceae, the family CAG-552 is within the Clostridia as well though in a different order.

The low number of bins which were significantly different and exceed the magnitude threshold for change might be a consequence of the low level of mapping for the strict bins to the reads; though the mapping percentage was greater for *A. russatus* on average than for *A. cahirinus* which saw more bins being above the thresholds for internal comparisons. It may instead be a sign of a very consistent intestinal microbiota in the *A. russatus*. Given that they follow a diurnal activity pattern when sympatric with *A. cahirinus* (as in the sampling site used) it is possible that their dietary patterns are less subject to change and therefore the principal driver of intestinal microbiota composition remains more consistent throughout the year. Potentially *A. russatus* from locations where they are not sympatric with *A. cahirinus* would exhibit a more diverse intestinal microbiome. Following their preferred nocturnal lifestyle might lead to a more varied diet. Metaproteomic analysis would give a good insight into whether the community within *A. russatus* provides a consistent suite of functional services and thus consistency in taxonomic composition could be explained that way. The limited number of samples may also limit the detection of seasonal effects in month-differential bins.

Within the *A. cahirinus* samples there were more bins which saw a statistically significant change from June to November which exceeded the magnitude threshold of greater than 1 (positive or negative), 9 in total. Two bins saw a reduction in their geometric mean RPM, Bin_m1135 (genus *Helicobacter*_D) and Bin_c1111 (genus CAG-632) while 7 others saw their abundance increase. These bins were Bin_m621 and Bin_c1167 (both genus CAG-873), Bin_c375 (genus *Muribaculum*), Bin_c380 (genus *Eubacterium*_R), Bin_c1070 (genus *Prevotellamassilia*), Bin_m293 (family Muribaculaceae) and Bin_m1474 (genus *Rikenella*). Though all but one of the bins which increased in abundance from June to November were members of the Bacteroidetes, this is not likely to be reflective of a seasonal change. Instead the author believes it is a combination of the low mapping percentages and the large proportion of the intestinal microbiota which is likely to be made up of Bacteroidetes. Neither of the two bins with a decrease in RPM from June to November are members of the Firmicutes or Bacteroidetes, which does put them outside the two most common phyla encountered in previously studied rodent intestinal microbiota. Both *Acomys* species do have relatively large ranges for their RPMs, though the maximum value for both in the June samplings exceeds the maximum in the November samples by large amount; suggesting that it is not due to a handful of outlier samples not accounted for by using the geometric mean creating a false positive. *Helicobacter* species can be pathogens but are also commonly reported members of the intestinal microbiota of a range of mammals [475] ; their abundance has been linked in the past in Capuchins to rainfall seasonality [476] though in *A. cahirinus* this is not likely to be the cause. Both *Acomys* species will consume arthropods and these vary in number and accessibility both on a daily cycle (impacting each host differently depending on if they are nocturnal or diurnal) and seasonally [370]. It is possible that there is a greater diversity of food sources available in November than in June, with a decreasing availability of arthropods necessitating greater consumption of seeds and available plant matter; leading to a change in the gut microbiota to bacteria better able to access plant-based energy sources. As neither *Acomys* species stores food it may be the case that in times of reduced food availability their microbiota are selected for those which are associated with obesity - to provide increased energy harvest from their more restricted diet. This is difficult to assess due to the low number samples used in this study, the relatively low mapping percentage of the reads to the strict bins and wide variety in published studies as to what taxa (at all levels from species to phyla) are associated with obesity and increased energy yield from the diet [477].

When comparing all the subset samples from June against all the subset samples from November, grouping them together by month even when the host species differs, reveals no bins which have a statistically significant difference in their geometric mean RPM and a fold change

greater than 1 (positive or negative). There are two bins which have a geometric mean RPM for November of 0, Bin_m1135 (genus *Helicobacter*_D, one of the bins significant in the within - *A. cahirinus* comparison) and Bin_m334 (genus *Helicobacter*_D). Neither of these bins have a statistically significant corrected P-value for the June to November fold change, which is a negative infinite value in any event. Three bins have a geometric mean RPM from all June samples of 0, Bin_c217 (family Borkfalkiaceae), Bin_c568 (family Treponemataceae) and Bin_m1093 (genus *Helicobacter*_C). They all also have a geometric mean RPM of 0 from all *A. russatus* samples, each being absent from at least one *A. russatus* sample. There are 8 bins which meet the Q-value threshold when looking at the June against November changes, though two only just and none manage to exceed a log2()+1 fold change of even 0.5 (positive or negative). The closest bins to meeting the fold change magnitude threshold which do manage to meet the Q-value threshold are Bin_m89 (genus TF01-11) with a June to November log2()+1 fold change in the geometric mean RPM of -0.46 and Bin_c1158 (genus *Acetatifactor*) with a change of 0.34. Of the three bins which meet the Q-value threshold but not the fold change threshold and have a positive fold change, two are members of the Lachnospiraceae family (including Bin_c1158) and the other is within the family CAG-272. Of the five bins which meet the Q-value threshold but not the fold change threshold and have a negative fold change, three are in the Lachnospiraceae (including Bin_m89) and the other two are in the Ruminococcaceae and the Muribaculaceae. Though the changes for the bins within the Lachnospiraceae do not meet the fold change threshold of greater than 1 (positive or negative) or even the lower threshold of 0.5, the observation of bins within the family seeing both positive and negative changes in abundance over time suggests that the family as a whole may be a core member of the microbiota. Different members may provide different functions and so change in abundance as the diet changes with the season. Lachnospiraceae have been observed to be core members of the gut microbiome of Forest Musk Deer (*Moschus berezovskii*) as they remain key members over the year even with dietary changes [478], though they were observed to increase with the increasing availability of fruit in Western Lowland Gorillas (*Gorilla gorilla gorilla*) over the year [479]. The failure of any Bins to meet both the Q-value and fold change thresholds when comparing all June against all November samples may be due to the limited number of samples within the subset being examined (and in the total dataset) but it may be a reflection of a relatively stable microbiome within each species. Though testing within each species individually from June to November did identify bins meeting both significance and magnitude thresholds, a relatively small number were found in each species. These were not amongst those which met the Q-value threshold when comparing both species against each other. These two results suggest that there is a relatively stable intestinal microbiota within each host species which sees some changes seasonally as the composition of the diet changes, with most of these changes being in the relative abundance of members of Lachnospiraceae. Had the mapping of the reads to the bins been more successful it is possible more data would be available to help determine if this latter conjecture is more accurate than simply insufficient data to detect seasonal changes.

Compared to the both species June versus November analysis, the combined *A. cahirinus* results versus the combined *A. russatus* results yielded a large number of bins which met the Q-value and fold change thresholds. 50.5% of all bins (176) met the thresholds and so can be said to be statistically significantly different between the two host species with a fold change of greater than 1 (positive or negative). 109 of these bins were enriched in the *A. cahirinus* samples while 67 were enriched in the *A. russatus* samples. There was one bin, Bin_m334 (genus *Helicobacter*_D) which had a geometric mean RPM for the *A. cahirinus* samples of 0, three bins had a geometric mean RPM of zero for the *A. russatus* samples. These bins were Bin_c217 (family Borkfalkiaceae), Bin_c568 (family Treponemataceae) and Bin_1093 (genus *Helicobacter*_C). The bins with 0 geometric mean RPMs for either of the two host species thus had an infinite

(or negative infinite) fold change in the RPM geometric mean. At the family level the 109 bins enriched in the *A. cahirinus* samples were distributed amongst 19 families whilst the 67 bins enriched in *A. russatus* were spread over 20 families. There were 12 families shared between the two species in their respective enriched bins set: Acutalibacteraceae, Anaerovoracaceae, Borkfalkiaceae, Desulfovibrionaceae, Eggerthellaceae, Helicobacteraceae, Lachnospiraceae, Muribaculaceae, Oscillospiraceae, Rikenellaceae, Ruminococcaceae and UBA644. The seven families enriched in *A. cahirinus* samples only were Anaeroplasmataceae, Anaerotignaceae, CAG-274, CAG-552, Lactobacillaceae, Marinifilaceae and Treponemataceae. The eight families enriched in *A. russatus* only were Bacteroidaceae, CAG-272, Elusimicrobiaceae, Erysipelotrichaceae, Turicibacteraceae, UBA1381, UBA3663 and UBA660. Interestingly, 51 of the bins enriched in *A. cahirinus* were from the Lachnospiraceae while only 2 of the bins enriched in *A. russatus* were; this despite the Lachnospiraceae being the most common classification of a bin by GTDB-Tk at the family level. The Muribaculaceae were the next most common family classification amongst the differential bins, with 14 enriched in *A. cahirinus* and 24 enriched in *A. russatus*. 7 bins were members of the Ruminococcaceae and were enriched in *A. cahirinus* while 8 were enriched in *A. russatus*. The very stark difference in the number of enriched bins between the two species for the Lachnospiraceae is quite interesting, especially as in the Muribaculaceaee and the Ruminococcaceae the numbers of bins enriched in each species is a lot more balanced.

The number of bins assigned to families by GTDB-Tk may be somewhat responsible for the apparent partitioning of enriched bins into the two host species. Looking at those bins which did not have a Q-value meeting the threshold for significantly different geometric RPMs between the two host species, 20 are members of the Lachnospiraceae, 8 are classified within the Ruminococcaceae and 10 within the Muribaculaceae. In the case of the Lachnospiraceae this does in fact support the idea that the Lachnospiraceae are enriched in *A. cahirinus* as there were more bins in the family not significantly different between the two species than there were enriched in *A. russatus*; with 51 bins from the family enriched in *A. cahirinus* as discussed earlier. Though the family was the most commonly classified, the fact that there were more bins which did not significantly differ between the two then were enriched in *A. russatus* does suggest that there is greater abundance of Lachnospiraceae members in *A. cahirinus* then *A. russatus*. Of the 104 bins assigned to the Lachnospiraceae, 51 of them were enriched in *A. cahirinus*, 41 were not significantly different and 2 were enriched in *A. russatus*. 8 had a statistically significant enrichment in *A. cahirinus* but below the threshold of +1 used and 2 had a statistically significant enrichment in *A. russatus* but above the threshold of -1. Muribaculaceaee was the next most commonly classified bin family, with 80 bins assigned. Fewer than half of those bins met both the Q-value and log fold change threshold, 24 were enriched in *A. russatus* and 14 in *A. cahirinus*; a much smaller difference than the Lachnospiraceae. 30 were not significantly different between the two host species, taken together this would suggest that the Muribaculaecae are common to both *Acomys* species and allow the breakdown of complex carbohydrates [480]. Given their very similar diets and the mix of arthropods, seeds, vegetation and snails which can make them up, it is reasonable to infer a benefit for the *Acomys* in harbouring members of the Muribaculaceae. Looking within the family, in both *Acomys* species the genus within Muribaculaceae with the most bins enriched was CAG-485, 5 in *A. cahirinus* and 13 in *A. russatus*.

The vast majority of the bins which were enriched in each of the host species were from families which had enriched bins in both species, the number of enriched bins from families with bins only enriched in one of the two *Acomys* species is very small. In the *A. cahirinus* samples there were 7 families with bins enriched only in that host, but they only accounted for 12 of the 109 enriched bins in *A. cahirinus*. In *A. russatus*, the 8 families with enriched bins exclusive to the host only covered 17 of the 67 enriched bins in the species. This supports the suggestion that

there is, at the family level at least, a relatively consistent 'core' microbiota within both *Acomys* species examined and which remains reasonably constant over the year. The bins which were not statistically significantly different were found within common gut-associated families, 41 in the Lachnospiraceae, 30 in the Muribaculaceae and 16 in the Ruminococcaceae for instance. In their work examining the impact of diet, captivity and reintroduction into the wild of the white-footed mouse (*Peromyscus leucopus*) van Leeuwen *et. al.* [481] found that the bulk of diversity in their study between treatment groups was within the Lachnospiraceae and Muribaculaceaea; with genera within the families changing in abundance. It is possible that the families with a large number of bins in each species, enriched and not, may vary at the lower taxonomic levels as a consequence of social interaction within each species. Shargal *et. al.* [482] established that both *Acomys* species will aggregate and nest together with conspecifics (not interspecifically) in their captive colony and links between social interaction and microbiota composition have been established in a number of species; including in mice by Raulo *et. al.* [483]. Similar diets may set broad constraints on microbiota composition, i.e. at the family and higher levels, but interaction between individuals within the same species may influence lower level membership - especially given the often reported transmission of microorganisms from mothers to offspring [484].

As there were a number of bins which GTDB-Tk could not classify to the genus level the presentation of the results and subsequent discussion have focussed on the family level results both within each host species and when doing cross species and cross sample month comparisons. As mentioned previously, there is considerable variation within the families Lachnospiraceae, Muribaculaceae and Ruminococcaceae. At the genus level *A. cahirinus* had a greater number of genera classifications with more than one bin enriched, 25 out of 46 genus level classifications, than *A. russatus*. 11 genera had more than one bin enriched in *A. russatus*, and 13 of the 67 bins enriched in *A. russatus* were classified into the genus CAG-485 - a common member of the Mouse gut microbiota [485] and placed within the Muribaculaceaee. Amongst the bins enriched in *A. cahirinus*, 5 were in the genera TF01-11, CAG-485 and *Acetatifactor*; CAG-485 is in the Muribaculaceaee while TF01-11 and Acetatifactor are in the Lachnospiraceae. Shared genera enriched in both *Acomys* species were: *Acutalibacter*, *Alistipes*, CAG-115, CAG-485, CAG-873, *Desulfovibrio*, *Duncaniella*, *Eubacterium*_F, *Eubacterium*_R, *Paramuribaculum*, *Ruminiclostridium*_E, UBA11940, UBA7173 and ZJ304; CAG-485 was the most common classification of an enriched bin - 5 in *A. cahirinus* and 13 in *A. russatus*.

Genera enriched in *A. cahirinus* and not in *A. russatus* were: *Acetatifactor*, *Anaerotruncus*, ASF356, *Bilophila*, BX12, CAG-510, CAG-590, CAG-632, CAG-95, D16-63, *Eubacterium*_J, JAAYNV01, *Kineothrix*, *Lachnospira*, *Lactobacillus*, *Limosilactobacillus*, MD308, *Odoribacter*, *Rikenella*, RUG115, *Ruminococcus*_C, TF01-11, UBA3263, UBA3282, UBA7109, UBA7182, UBA9502, UMGS1004, UMGS1872, UMGS268, 1XD42-69 and 14-2.

Genera enriched in *A. russatus* and not *A. cahirinus* were: *Allobaculum*, CAG-353, CAG-41, CAG-460, CAG-582, CAG-710, *Emergencia*, *Eubacterium*_G, *Muribaculum*, *Paraprevotella*, RUG14121, *Ruminococcus*, *Ruminococcus*_E, *Turicibacter*, UBA1436, UBA1777, UBA3855, UBA5578, UBA6857, UBA7057, UMGS1312 and UMGS1815.

4.4 Annotation of metagenomic bins

Bioinformatic annotation of the metagenomic bins provided some degree of insight into the functional potential of the bins, though naturally and unavoidably inferior to dedicated transcriptomic or metabolomic work; these later were not carried out due to a combination of lack of experience on the part of the author and limited access to laboratory facilities caused by the global pandemic.

The COG results were useful as they contained COGs which showed differential distributions in bins which were themselves differentially abundant when comparing the two host *Acomys* species. Broadly the most commonly detected COGs by Prokka are commonly found in bacteria, so while in combination with the CheckM results do help support the idea that the bins are close to the original microbial species in the *Acomys* gut microbiota they do not necessarily provide a great deal of insight into the functional potential of the bins. Of the 10 most commonly detected COGs from all bins there are those which are associated with essential functions in bacteria, such as COG0745 and COG1595. COG0395 was amongst the 10 most commonly detected COGs in the 348 bins and has been found in the genomes of Haloarchaea [486]. More generally the presence in the 10 most commonly detected COGs of multiple COGs associatged with ABC-type transporters may be an indicator that the microorganisms are better capable of responding to changing environmental, in this instance the intestines of the *Acomys*, conditions as the transporters are known to have an array of regulatory functions [487]. They have also been observed to play an important role in osmoadaptation through the movement into the cells of compatible solutes to handle osmotic stress [488, 489]. COG0534, associated with sodium driven efflux pumps, being amongst the 10 most commonly detected COGs could also link to the proposed halotolerance of members of the *Acomys* gut microbiota [490]; or that there is competition amongst the microorganisms which could make it beneficial to have the ability to survive attack by antimicrobial compounds. It is reassuring to see that the distribution of bins which had the greatest single detection count for each COG indicating that the COGs were not most abundant in a single or small handful of bins; had they been this could be a sign that those bins had been misassembled and were not true reflections of a single type of microorganism.

Those COGs which have a statistically significant distribution in those bins which are enriched in *A. russatus* include those which likely have no particular biological meaning, such as COG2176, but also those which may be biologically relevant such as COG0777. The latter is associated with the Acetyl-CoA carboxylase beta subunit which is involved in the production of fatty acids [491]. Short Chain Fatty Acid (SCFAs) production by members of the microbiota have been associated with host health previously [492] and potentially the bins enriched in the *A. russatus* are providing the host with this service. As the results of the differential bin detection and the bin taxonomic classification indicated that there was considerable overlap at the family level between both *Acomys* species' microbiota it is possible that the COGs found in the *A. russatus* enriched bins are simply more abundant in the microorganism which happened to provide the sequences assembled into those bins rather than being a reflection of a true difference between the two *Acomys* species.

The most commonly detected genes are not surprising and are overwhelmingly associated with essential functions, therefore the author does not believe the ten most commonly detected genes reflect anything biologically relevant. sasA_1 and sasA_2 are genes associated with histidine kinase which have been linked previously to salinity tolerance, albeit in Rice (*Oryza sativa*) [493]. Potentially this might be linked to an increased salt tolerance required in the microbiota of both *Acomys* species due to the previously discussed high salinity of their diets. Otherwise the results for the gene annotations are not particularly informative, this is probably due to the

seeming large shared functional potential of the microbiota of both *Acomys* species, the relatively low percentage of reads which were incorporated into the bins, the bins being produced through co-assembly of all sequencing reads and the limited number of bins which were 100% complete according to CheckM.

4.5 LAB isolation and culturing from *Acomys* faecal samples

It was uncertain whether it would be possbile to obtain live cultures from the *Acomys* faecal pellets; they had been stored frozen at -80℃ for multiple years without being frozen in glycerol or any other protective medium. Overall the results are evidence for the general potential viability of bacterial members of a gut microbiota which may survive prolonged periods of cold storage. Given that prior investigations have indicated there is a connection between the storage method of faecal samples and the results of any subsequent metagenomic investigation of these samples [494, 495] it might also be the case that Lactic Acid Bacteria may be more resilient to frozen storage conditions. A way of testing this, and of accessing the much larger proportion of the *Acomys* gut microbiota which the results of the prior chapter suggest are not lactic acid bacteria, would be to attempt to isolate different types of bacteria from *Acomys* faecal samples. The comparatively greater proportion of the *Acomys* gut microbiota which the metagenomic bins analysis indicates may be made up of members of the family Lachnospiraceae might be isolated and then sequenced to obtain entire genomes. The works of Hedberg *et. al.* [496] and Sorbara *et. al.* [497] suggest that blood agar would be a good media to use for untargeted isolation of Lachnospiraceae in particular or other bacteria in the *Acomys* gut microbiota. Given that all of the *Acomys* faecal samples examined bioinformatically contained an overwhelming majority of unclassifiable reads, the application of culturomics or high throughput single-cell sequencing to the investigation of these samples would likely yield novel microbial diversity.

The early results of the taxonomic classification, that all three tools which yielded an above zero average percentage of reads classified included LABs in their detections and placed it as anywhere from 1 - 30% of the reads classified along with the fact that collaborators had developed media explicitly intended to isolate LAB led to the decision to attempt culturing from the stored faecal samples. This decision was reached prior to the outbreak of the global pandemic which prevented the collection of more samples, any (for a prolonged period of time) access to laboratory facilities for non-COVID related research and the delay in resumption of normal research laboratory work thus meant that only those faecal pellets left after extraction and sequencing were available to be used. That they had been stored for years without any protective media and still provided cultures on the selective media confirms that the *Acomys* microbiota harbours LABs, validating their detection by the classification tools if not the different relative abundances they provided. LABs have been detected in multiple rodent [498, 499, 500] and other species [501, 502, 503] and the author elected to investigate this component of the *Acomys* microbiota through bioinformatics-guided culturing. It would have been preferable to take a less restricted approach to the culturomics, to have used non-selective media, to have used different forms of selective media, to have used multiple growth conditions with the various media, to have collected more samples to isolate from and even to have tried different methods to access different culturable sections of the microbiota - these were all discussed and would have been attempted were it not for the restrictions of the global pandemic and associated policies. The results from the taxonomic classification tools may not have suggested any other particular groupings of microorganisms to investigate but the results from the taxonomic classification of the metagenomic bins do suggest some which might be of interest to future investigators.

4.6 Assembly of LAB reads

The genome sizes of the assemblies as provided by CheckM are similar to those provided by Makarova *et. al.* [504] for lactic acid bacteria, as are the predicted number of genes; with a mean of 2,113 (st. dev. 133) predicted genes for *A. cahirinus* isolate assemblies and 2,070 (st. dev. 218) predicted genes for *A. russatus* isolate assemblies. The number of contigs in the assemblies is low, further work would be required to fully refine them into circularised genomes but the assemblies appear to be of good quality based on the size and contig number results. One of the *Apodemus* isolate assemblies, S125, has a much higher contig count than the other *Apodemus* isolates and the *Acomys* isolates - the 17.5% contamination indicated by CheckM explains this extremely high contig count. All *Acomys* isolate assemblies have low levels of contamination and high levels of completeness and so were kept for the subsequent stages of the analysis and highlighting the utility of isolation and then sequencing to investigate bacterial genomes - especially when compared to the metagenomic bins generated in the project. Though the bins are of high quality as well they are not as uniformly high quality as the individually sequenced isolates.

As the isolates were initially picked based on visual inspection to determine whether isolates were likely to be the same or not, it is of credit to Nancy Teng, who carried out the isolation culturing, that only two sets of isolates had an ANI similarity score of greater than 99.9% to each other. One pair of isolates which had a greater than 99.9% ANI similarity score were from within the same host species, and from the same faecal sample which was not surprising. More interesting are the two assemblies which have a 99.98% ANI similarity score to each other from different host species, 13E_S5 from an *A. russatus* sample and 16aC_S26 from an *A. cahirinus* sample. Given the prior chapter indicating that both *Acomys* species contained a relatively low abundance of *Lactobacillus* (the old version of the genus in the reference databases used) species it might be reasonable that in trying to isolate out lactic acid bacteria the same strain from both host species was obtained. Possible future work could include investigating whether the two strains grow in the same manner in media and whether they produce the same metabolites; this would help assess if the two isolates truly are identical. In deciding which member of the two sets of isolates to use in the rest of the analysis, the completeness and contamination values for each of the four did not provide any direction as they were identical within each pair; hence randomly deciding which to proceed with. The value of 99.9% was chosen in accordance with Kujawska *et. al.* [429], which was informed by Olm *et. al.* [505]. Masking the remaining assemblies did not lead to any further detection of ANI similarity scores of 99.9% or greater and so suggests that though they may be closely related, none of the assemblies are from the same strains.

Though there were only these four isolates which were deemed to be too similar to all be included, the same ANI results are also of interest in suggesting which assemblies are closely related species, especially as they can be compared to the subsequent results from the phylogenetic tree creation. The ANI similarity scores, or lack thereof between the *Acomys* and *Apodemus* isolates indicates that identical isolates have not been picked from samples from each of the two genera. The lack of ANI similarity scores of around 80%, so meeting the threshold to be reported, between all but one of the *Acomys* assemblies and those from the *Apodemus* is a point worth noting. However, it cannot alone say anything about differences in microbiota composition between the two genera, the number of samples from both host genera being so low it is likely the project has not exhausted the lactic acid bacteria components of either the *Apodemus* or *Acomys* microbiota. The single *Acomys* isolate assembly which had an ANI similarity score reported for any of the *Apodemus* isolate assemblies was 41B_S7 (an *Acomys russatus* assembly), which had ANI similarity scores ranging from 81.4% for isolate S124 and 96.7% for isolate S121. These results matched with those from GTDB-Tk, which classified 41B_S7 as *Ligilactobacillus murinus* and

classify all the *Apodemus* isolates in the genus *Ligilactobacillus*; the *Apodemus* isolates closest to the *Acomys* isolate were able to be classified to the species level as *Ligilactobacillus murinus*.

4.7 Taxonomic identification of LAB assemblies

The results from GTDB-Tk classification of assembly taxonomies were not surprising, given that the assemblies were obtained from isolates cultured on selective media for Lactic Acid Bacteria. An interesting point which links back to some of the issues around identifying the members of microbial communities from less-studied environments is that GTDB-Tk had greater success classifying the assemblies from the *Apodemus*. Given that these were unintentional byproducts originally from a study investigating *Bifidobacteria* it highlights that more studies are needed to populate databases so investigators are better able to explore microbiota. The faecal microbiota of *Acomys* has already provided a number of novel taxa which can contribute to diversifying databases, a more thorough investigation would likely yield even more. Five of the seven *Apodemus* isolates passing QC could be classified to the species level by GTDB-Tk and all could be classified to the genus level. That all the species identified in the *Apodemus* assemblies are the same, and that the genus identified for them is the same might suggest that members of the *Ligilactobacillus* are prominent members of the gut mirobiota in that genus of rodent. Alternatively they may simply be those which were better able to grow on the selective media intended to select for *Bifidobacteria*.

The project demonstrated the utility of obtaining entire genomes for assembly from culturing and subsequent sequencing of pure isolates in the difference between the level of classification obtainable by GTDB-Tk for the bins and the assemblies from the *Acomys* isolates. All *Acomys* isolates could be classified to the genus level, a much better level of success than that for the metagenomic bins obtained from the *Acomys* faecal samples. That all the isolates were assigned as members of the family Lactobacillaceae is also unsurprising, the media was selective for lactic acid bacteria. The limited diversity in the family is more notable. *Limosilactobacillus* being the most common *Acomys cahirinus* assembly classification may link to their reported antagonism towards other microorganisms, through the production (by some strains) of antimicrobial Polyketides as found by Özçam *et. al.* [506]. Diez-Echave *et. al.* [507] have also shown the potential role of another *Limosilactobacillus* species as a probiotic, which might be true of the species cultured from the *Acomys* faecal samples; and would indicate positive selection for them within the microbiota. The classification of two assemblies as *Bifidobacterium* species, one from each of the two *Acomys* host species may owe much to the selective media having originally been created for isolation of that genus in particular, but might also be a fair reflection of the proportion of the microbiota in each *Acomys* species composed of members of the *Bifidobacterium*. They have been suggested as probiotics and may be considered as indicators of a healthy gut microbiota in both mice (*Mus*) and rats (*Rattus*) meant to be used for research purposes [508].

Assemblies being classified as *Lactobacillus* species may be represented by the reads from the *Acomys* samples classified in the genus by Kraken 2 and Metaphlan. That only seven isolates in total were placed within this genus strengthens the relatively low proportion of the reads classified this way by the read classifiers; though it might also be the case that this particular genus is in low abundance in the individuals which provided the samples used for culturing. The classification of one of the isolates, 41A_S6 from an *Acomys russatus* individual (as *Paralactobacillus*), is a result of the database used by GTDB-Tk including the genus and so being in accordance with Zheng *et. al.* [509]. The single paper on the type species [510] for the genus does not provide any indication of possible connections between the Malaysian food ingredient where it was isolated and the intestines of the *Acomys russatus* so the assembly is likely that of a related but distinct lactic acid bacterium. The author initially thought the classification of one of the *Acomys* assemblies as *Pediococcus pentosaceus* was a result of undetected contamination. This was as it was only assigned to a single assembly in contrast to all other classifications which

had at least two isolates assigned to them. However the author now believes it is a real result and reflects the reported presence of the species in the intestinal microbiota of foals [511] and their beneficial effects in mice [512, 513]. Its detection is likely not a false positive and it may be playing a beneficial role in the *Acomys* intestine.

More culturing, both in terms of picked isolates from the faecal samples used here and a greater number of samples used for culturing may have increased the diversity of detected lactic acid bacteria. However, as mentioned above and in the prior chapter, the low level of detection of the *Lactobacillus* species by the classifiers from the reads and the proportionally low number of metagenomic bins assigned to the Lactobacillaceae by GTDB-Tk may mean that the lactic acid bacteria diversity in the *Acomys* is low. Interestingly, the high number of bins classified within the Lachnospiraceae might be tied to the low number of both bins and limited diversity within the Lactobacillaceae in the *Acomys*; Brownlie *et. al.* [514] report homofermentative Lactobacillaceae inhibiting the growth of Lachnospiraceae. It is possible then to infer that the balance between the two families, as reflected in the number of bins, is between the Lachnospiraceae and heterofermentative Lactobacillaceae. Whether the assemblies obtained from the isolates are of heterofermentative species remains to be seen. The greater diversity of genera from the *Acomys russatus* samples (n=6), as compared to the 3 from the *Acomys cahirinus* samples might be a reflection of greater diversity in the Lactobacillaceae component of the *A. russatus* intestinal microbiome but this project did have enough samples to state this definitively.

The existence of different environmental niches within the intestinal tract of animals which can be colonised by specific communities has been reported across a range of animals. Duncan *et. al.* [515] found a particular community of bacteria living in a biofilm-like state within the mucus layer of the mouse gut lumen. An earlier study by Li *et. al.* [516] recorded bacterial species living in the mucus layer of mouse gut showed distinct resource use and proliferation as compared to the same species residing in the lumen. The ability of a particular bacterial species to live in different niches through different expression patterns and resource use has been reported by Jenior *et. al.* [517] in *Clostridium difficile*. That faecal samples have been found [518] to be unrepresentative of these mucus-associated communities highlights the degree of differentiation according to niche found in microbial communities. Niche exclusion, the driving out of a competitor using a niche by a different competitor through either elimination or adaptation, has been previously reported in microbial communities associated with different environments. The detection of niche exclusion in a microbial community can be accomplished through network analysis [519], as was carried out by Roggenbuck *et. al.* [520] in two New World vulture species; the Black Vulture (*Coragyps atratus*) and the Turkey Vulture (*Cathartes aura*). They found that Clostridia and Fusobacteria outcompeted other bacterial groups in the Vultures' anaerobic hindgut, speculating that their role in the breakdown of carrion was of sufficient benefit for the host to outweigh the production of toxins by these bacteria. Jenior *et. al.* [521] in a more recent investigation using *C. difficile* found that during infection the species 'manipulated the niche landscape of the intestinal tract', in particular it appeared to exclude the rarer members of the caecal microbiota by outcompeting them and driving down their abundance within the different niches *C. difficile* can inhabit. Looking at the same bacterial species, *C. difficile*, H.Foley *et. al.* [522] observed that the toxins produced by the species caused inflammation as a response by the host immune system led to bacteria from the *Bacteroides* - which compete with *C. difficile* for the same resources obtained from degradation of host collagen - decreasing in abundance. Niche modification, in which new niches were created courtesy of the actions of microbes in a different niche, linked with an animal-associated microbiota was reported by Shaani *et. al.* [523] in Cattle. The microbiota can also be modified through direct antagonism between microorganisms through the production of harmful compounds such as Bacteriocins [524, 525, 526], many of which are specific to particu-

lar groups of targetted bacteria [527]; typically close relatives of the producing strain. It is quite possible that the *Acomys* microbiota does experience some niche-based competition and exclusion between the different taxa present and it would be an interesting piece of future research to conduct network analysis with a larger sample set to look for any antagonistic relationships.

4.8 Phylogeny of LAB assemblies

4.8.1 Relationships between isolates

The tree confirms the results from the ANI analysis and GTDB-Tk, that the *Apodemus* isolates are all closely related to each other and likely are within the same genus and some might be even more closely related than that. The placement of the two isolates 41B_S7 and 41A_S6 with these *Apodemus* assemblies and with 41C_S8 on the closest branch outside the subtree demonstrates that there is quite likely to be some shared functional overlap between the two host genera for their microbiota. Whether this is due to the *Acomys* isolates being common to rodents and capable of surviving on the *Acomys* diet or the *Apodemus* isolates reflecting a plasticity in their host's microbiota, such as that discussed by Koziol *et. al.* [528], cannot be determined without further sampling of both hosts.

That the assemblies from isolates from either host species do not form distinct subtrees from each other might be more a reflection of the limited number of isolates and the limited number of hosts. In this case any particular mutations in the individual isolates might be sufficient to overcome a pattern which might be observed with more samples. As it is the tree shows a greater diversity for the isolates originating from *A. russatus* than for *A. cahirinus* though not to any significant extent. This is also likely influenced by the limited detected taxonomic diversity from GTDB-Tk, with all the assemblies from 16bD_S22 to 39E_S18 being classified as *Limosilactobacillus* and the tree suggesting there are likely two if not three species within this genus present across those assemblies. The two assemblies identified as *Bifidobacterium* are also located on their own small branch within a larger subtree, 18A_S9 and 39B_S15, with an extremely short distance between the two; despite them originating from different host species. It may be the case that as with Moraitou *et. al.* [529] the host diet, which in this case is very similar between the two, contributes more to microbiota similarities than the host phylogeny.

The restriction to LABs from the bioinformatically-guided culturing will have limited the potential diversity available and so the tree would be constrained no matter how many samples had been collected and no matter how many host individuals had been sampled as the selective culturing would set boundaries on genetic dissimilarity. Incorporating isolates from putative non-selective culturing would help resolve the relationships between the LABs within the *Acomys* microbiota.

4.8.2 Relationships between isolates, metagenomic bins and reference genomes

Isolate assemblies and metagenomic bins ANI similarities

The low number of assemblies which had ANI similarity scores reported to any of the 348 metagenomic bins is another indicator that the proportion of the *Acomys* microbiota made up of Lactic Acid Bacteria is likely low. For the *Apodemus* assemblies their close relationships are supported by them all having reported ANI similarities to Bin_c770, given that the ANI similarities are mostly above 90% this would indicate the bin itself is likely within the *Ligilactobacillus* genus like all the *Apodemus* assemblies and the *A. russatus* assembly 41B_S7 which was also classified as *Ligilactobacillus*. The three bins which had reported ANI similarity scores to *Acomys* isolates, meaning they had at least approximately 80% ANI similarity, were Bin_c1617, Bin_m1569 and Bin_m1485. All three bins were classified by GTDB-Tk as members of the Lactobacillaceae, which is likely why they have some degree of relationship to the assemblies. Bin_c1617 had ANI

similarity scores of 78-80% to assemblies from both *Acomys* species, 13A_S1, 39C_S16 and 39D_S17 from *A. russatus* and 16aB_S25, 16aB_S19, 16bB_S20 and 18D_S12 from *A. cahirinus*. All of these assemblies were classified as *Lactobacillus* which strongly suggests that the bin is also likely a member of the genus; or a closely related one. Bin_m1569 has an ANI similarity score of above 98% to the assemblies 18A_S9 (from *A. cahirinus*) and 39B_S15 (from *A. russatus*); both of the assemblies were classified as *Bifidobacterium* by GTDB-Tk. As there are only two *Bifidobacterium* isolate assemblies the author would not say there is enough backing from the ANI similarity scores to suggest the bin could be a *Bifidobacterium* species. Bin_m1485 has a 78-79% ANI similarity score to a number of the assemblies, 13A_S1, 39C_S16 and 39D_S17 from *A. russatus* along with 16aB_S25, 16bA_S19, 16bB_S20 and 18D_S12 from *A. cahirinus*. These assemblies were all classified as *Lactobacillus*, and are the same assemblies which had a similar ANI similarity score to the Bin_c1617 - so one could be reasonably confident that Bin_m1485 is either a *Lactobacillus* species or from a closely related genus. These results again support the results from the taxonomic classification of the shotgun reads that the lactic acid bacteria are a proportionally small component of the *Acomys* microbiota. Future investigation would look to try and isolate colonies from those families which the taxonomic classification of the bins indicated were more abundant in the *Acomys* microbiota.

Phylogenetic tree of isolate assemblies, metagenomic bins and reference genomes

The expanded tree shows the addition of MAGs, in this case high quality iMGMC ones from a published reference collection associated with mice, did not particularly alter the split between the assemblies from both *Acomys* species and the *Apodemus* with the metagenomic bins produced from the *Acomys* faecal samples. This, in combination with the majority of the iMGMC MAGs being placed away from the assemblies suggests that there is something fundamental to a genome or bin assembled from sequencing reads originating from disparate sources which distinguishes them from genomes produced from sequencing of pure cultures. The bins used in the project were those which met stricter quality thresholds (min. 80% completeness, max 5% contamination) than are sometimes employed so the author does not believe this is a consequence of contamination in the bins, nor would this explain why the iMGMC MAGs place with the bin files. Using entire sequence alignment with Cactus to produce the trees avoids issues with missing and incomplete marker genes from the bins and is more accurate than kmer-based methods, but it is possible that the maximum 20% which can be missing from the bins causes the split [530]. That the iMGMC MAGs did not cluster away from the bins validates the approach for investigating the *Acomys* microbiota, as they are both ultimately constructs made from shotgun sequencing reads from rodent samples and that they are mixed throughout the tree indicates that the bins are not technically different from the iMGMC MAGs and the different placements within the tree reflect phylogenetic differences. That the assemblies place amongst the reference genomes but, with 3 exceptions, do not have any of the reference genomes as their nearest neighbours supports the conclusion that they represent novel species or potentially higher level taxa though with relatives within other rodent microbiota.

The *Apodemus* isolate assemblies continue to be placed apart from the majority of the *Acomys* isolate assemblies, though the addition of the external reference genomes offers some clarification of the relationships between the seven *Apodemus* asseblies and the four *Acomys* isolate assemblies placed near them. 41C_S8 is surrounded by a *Pediococcus pentosaceus* genome and then between it and the *Apodemus* assemblies by a *Priestia filamentosa* genome. A *Ligilactobacillus murinus* genome sets the limit of the five *Apodemus* assemblies S122 - S126. A different *Ligilactobacillus murinus* genome is placed then next to the assembly 41B_S7 and then the next most closely placed genome is another *Ligilactobacillus murinus* genome, this supports

the GTDB-Tk classifications of the assemblies which were all classified within the *Ligilactobacillus* genus or specifically to *Ligilactobacillus murinus*; with the exception of 41C_S8 which was classified as the same as it's nearest neighbour *Pediococcus pentosaceus*. Assembly 41A_S6 was classified as *Paralactobacillus* and has as its closest neighbour a branch which includes the reference genomes NZ CP014924 and NZ CP012275, genomes for strains of *Lactobacillus paracollinoides* and *Pediococcus damnosus* respectively. The placement of a *Bifidobacterium pseudo- longum* as the nearest neighbour to an isolate genome from *A. russatus* which was classified as a member of the *Bifidobacterium* is reassuring for the accuracy of the GTDB-Tk classifications for the assemblies, relatively close to these within the same smaller subtree but on a different branch can be found both the *A. cahirinus* assembly 18A_S9 and the reference genome GCF 001281425; the latter is a genome of *Bifidobacterium breve* and the former was classified into the *Bifidobacterium*

Reference genome GCA 001689405 is placed in amongst the bins classifed as belonging to the Muribaculaceae and is itself a metagenome assembled genome called '*Candidatus* Homeothermus arabinoxylanisolvens' which was sourced from a *Mus musculus* faecal sample [471], supporting the validity of the bins. Interestingly this large subtree also contains the reference genome GCF 000762845, which is a genome for *Helicobacter japonicus*. This is a member of the Helicobacteraceae rather than the Muribaculaceae - though it is away from any of the bins actually classified into the Muribaculaceae and was itself originally found in mice sourced from three Japanese institutes [531]. The other reference genome found in close relationship to bins classified within the Muribaculaceae is GCF 001688845, which is a genome of *Muribaculum intestinale* - a species originally identified in mice [532]. That a member of a different family is found using the reference genomes amongst the bins classified into one family mirrors the results from the tree with the bins and assemblies alone, in which bins from different families could be found mixed together. The author believes this is likely due to great functional overlap between the genomes of the taxa represented by the bins, or in the case of the reference genomes the actual source microorganism, which overcomes the taxonomic differences when considering the phylogenetics.

The tree also offers some possibility to understand the splitting of bins classified into the same families. The clearest example would be the three bins classified as Acutalibacteraceae which instead are located within subtrees apart from any other bins classified this way. Two of the bins however, Bin_m595 and Bin_c420 cannot be considered in this way as the subtrees containing them contain no reference genomes, but instead iMGMC MAGs. Bin_c1180 however is within a subtree containing a reference genome in addition to the many bins classified as members of the Lachnospiraceae. The reference genome is for a strain of *Ruminococcus gauvreauii*, which is a member of neither the Acutalibacteraceae nor Lachnospiraceae but instead the Oscillospiraceae. It was originally isolated from a faecal sample [533], albeit a human one, but has also been obtained from bile samples [534]. Interestingly, the species has been linked to coronary artery disease through a possible link to diabetes [535] - in captivity both *Acomys* species have been known to become diabetic when fed on the typical laboratory chow. It would be interesting to conduct an examination using *Acomys* in captivity being fed the standard chow and looking for any changes in the abundance (measured by mapping of shotgun reads from faecal samples) of the bins found on this subtree.

That the *Acomys* microbiota is distinct from the five desert rodents sampled by Kohl *et. al.* in their study is suggested by the large number of bins found away from the reference genomes - which include the 45 choosen based on their results. However, given the disparity already observed in this project between the results of the taxonomic classifiers and the taxonomic identities

of the bins the author instead believes that this reflects the different sections of the microbiota reached by the different approaches. It is quite possible that a tree made through a combination of broad culturomics and long read single cell approaches might yield a combination of assemblies, bins and MAGs which would produce a tree with more mixing if the same set of reference genomes was used. The tree is also somewhat skewed by the use of mOTUs results to pick 80 references, given that this is likely impacted by the necessity to be similar to sufficient reference-mOTUs to be classified. Going by the results of both mOTUs and the other classifiers, this is a very different fraction of the reads from those which were assembled and binned together based on the GTDB-Tk classifications of the bins. It might have been more informative to use the results of a different taxonomic classifier however the published benefits of mOTUs with the use of multiple markers provides great certainty that detections were not false positives; hence the author opting to use the results from it to pick a subset of the external references used.

4.9 Mapping of *Acomys* faecal sample shotgun reads to LAB assemblies

The initial result, that mapping with the read files before or after subsampling made only a marginal difference, is not necessarily surprising. Given the low proportion of the *Acomys* microbiota which appears to be made up of Lactobacillaceae members, the likelihood of the subsampling causing a noticeable drop in the mapping percentage is not high. If proportionally few of the reads were from the isolate family then a random reduction in the read number would not be expected to cause a significant change in the number of reads mapping to the assemblies. Using Salmon, in the atypical use presented here, helps interpret the results as it allows the use of a reads-per-million value (RPM) to account for differently sized assembly contigs. This is also why the author combined all the contigs within each assembly into a single one and then concatenated these together into the single mapping reference; it was important to ensure the reads were provided with the entire range of assembly content at the same time for accurate mapping.

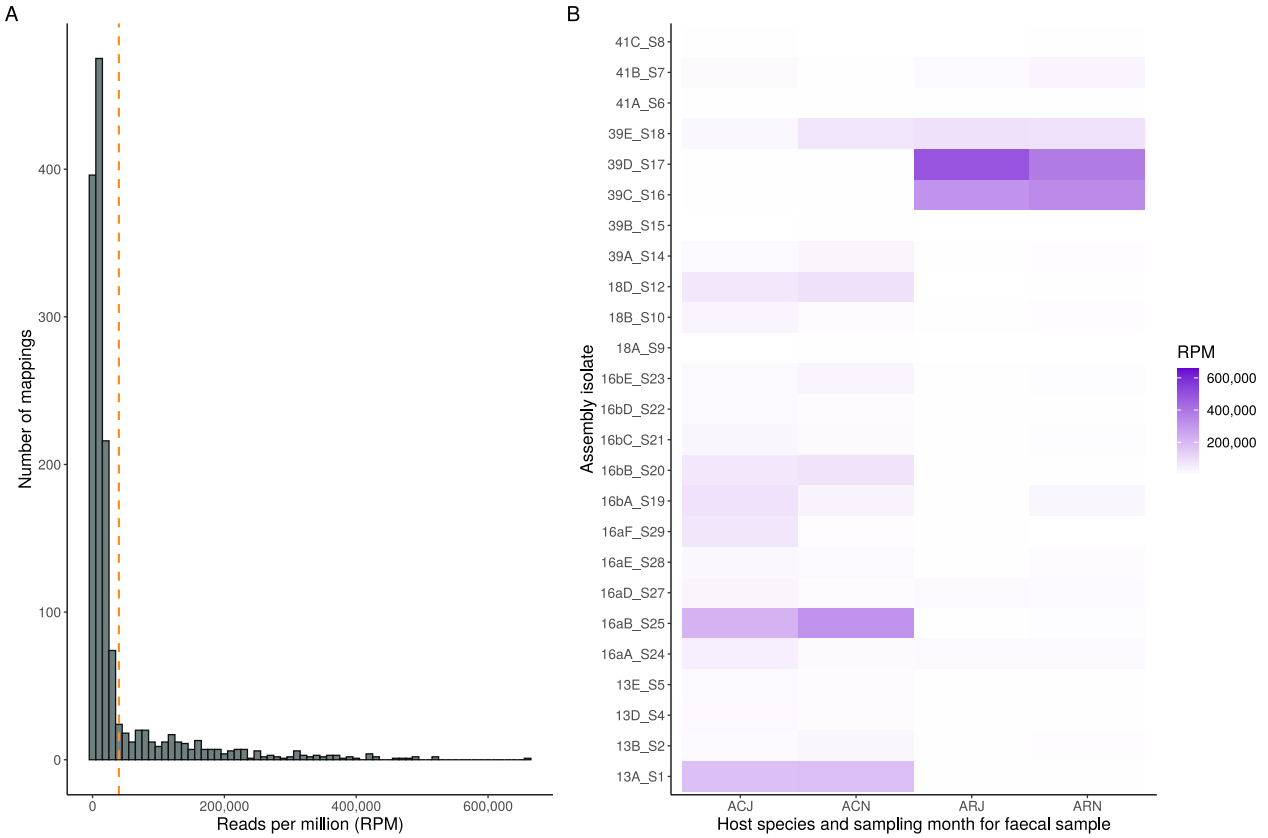

Figure 4.2: Histogram and heatmap. **A.** Histogram of the number of mappings which had the indicated RPM value for mapping of reads to isolate assembly reference, bins have width of 10,000 and dotted orange line shows mean RPM value. **B.** Heatmap showing the RPM values for each of the isolate assemblies by the host species and samping month of the faecal sample the reads originated from.

Figure 4.2A shows the range of RPM values obtained from the mapping. The bulk of RPMs were on the high end, only 77 RPMs were between 100 and 1,000 in contrast to 575 RPMs between 1,000 and 10,000 and 618 RPMs between 10,000 and 100,000. There is a drop in the number of RPMs above this, 179 RPMs were between 100,000 and 1,000,000.The lowest RPM was 91.12 for mapping of subsampled reads from the *A. cahirinus* sample AC32J to the *A.*

russatus derived assembly 39B_S15; this is a standout as the only RPM which was below 100. The highest RPM: 657,155, was for mapping of reads from the *A. russatus* sample AR45N to the *A. russatus* derived assembly 39C_S16. This might be a reflection of the limited number of assemblies that reads could map to, or that the assemblies do actually provide good coverage of the Lactobacillaceae component of the *Acomys* microbiome and the reads mapping are those from these species within the faecal samples. The difference in the RPMs overall between the two *Acomys* species, such as the median RPM to all assemblies for *A. cahirinus* reads being 14,246 while for *A. russatus* reads it was 7,742 might be simply due to chance - potentially the *A. cahirinus* isolates were more common within the *Acomys* intestine than the ones obtained from the *A. russatus* faeces. A larger sampling set might have reduced these differences, both in terms of the number of isolates obtained and the number of sequenced faecal samples to map. **Figure 4.2B** shows the RPM values to the different isolate assemblies by the host species and sampling month of the faecal sample the reads came from. The trend for slightly greater mapping of *Acomys cahirinus* reads to the majority of assemblies is visible. It is interesting when examining the mapping by the *Acomys* species which the reads and the assemblies came from to note the difference between the two. Unsurprisingly, the reads from each of the two *Acomys* species map to a greater level to those assemblies which originate from the same species, the arithmetic mean RPMs for same species read to assembly mapping are 52,476 for *A. cahirinus* and 60,196 for *A. russatus*. These contrast to the mean RPMs for mapping to assemblies from the other *Acomys* species, being 26,484 for *A. cahirinus* reads to *A. russatus* assemblies and 21,357 for *A. russatus* reads to *A. cahirinus* assemblies. That the mean RPMs for the cross-species mapping are not extremely low (i.e. below 1,000) suggests either that the assemblies represent a shared component of the microbiota across both *Acomys* species or that some of the assemblies are actually less common members of the microbiota of their own species. The author suspects the latter is a bit more likely due to the limited number of faecal samples used to obtain the assemblies rather than having obtained isolates which were from taxa found across both *Acomys* species. It is noticeable that the sampling month does not appear to have an impact on the RPMs when compared to the species; in general the mapping level seems to be similar between the two sampling months for either of the host species to any given assembly.

There are some interesting results shown the PCA biplots made from the RPMs. When looking at the mapping to all assemblies there is clear partitioning between the two *Acomys* species - though this is mostly from tighter clustering of the *A. cahirinus* results. The results show that when restricting the analysis to the RPM results for mapping to the assemblies derived from *A. cahirinus* isolates there is still partitioning into the two host species though to a lesser extent. Looking at the PCA plot for mapping to assemblies for the *A. russatus* assemblies there is extreme partitioning caused by the very tight clustering of the *A. cahirinus* read RPMs. Taken together these suggest that there is greater difference in the mapping of *A. russatus* reads to the assemblies derived from either host species from faecal samples within the species than is seen with the *A. cahirinus* faecal samples. This project had too few samples to accurately assess whether this is because the *A. russatus* assemblies represent members of the microbiota of quite different abundances compared to more uniform abundances in the microbiota for the *A. cahirinus* assemblies or because the low number of samples used for the mapping contain very different abundances for the assemblies compared to the majority of members of the species. Based on the above points it is not surprising that there were only two assemblies which had a statistically significant effect (after correction) for sampling month, one from each of the *Acomys* species. Nor is it surprising that all of the assemblies had a statistically significant effect for the host species, though how strong the effect was varied between the assemblies. It is important to bear in mind that the Lactobacillaceae likely make up a small proportion of the *Acomys* intestinal microbiota so these host effects might be significant statistically without necessarily being particularly relevant

biologically. Ideally one would not only have many more samples both for obtaining isolates but also for mapping. The investigation would benefit from experiments involving the addition of isolates from one host species into the other to observe if there were any detectable changes in the recipient.

4.10 Annotation of LAB assemblies

The range in the number of COGs detected in the assemblies varies significantly between the *Apodemus* assemblies and the two *Acomys* assemblies; this may be a consequence of the reduced number of isolates sourced from *Apodemus* but it might also reflect the limited taxonomic diversity found in the GTDB-Tk classification of the assemblies. It is quite possible that the COGs identified from the *Apodemus* assemblies provide a good level of coverage for the COGs which can be found in the *Ligilactobacillus murinus* strains which are in the *Apodemus* microbiota. The ten most commonly detected COGs from all assemblies are associated with essential functions and so do not necessarily indicate any biological significance. As discussed in relation to the annotations of the metagenomic bin files, the number of COGs associated with transporters in the ten most commonly detected may be a sign that the isolates are from microbial organisms which can tolerate changeable environmental conditions. They might also be associated with a required baseline level of halotolerance for members of the *Acomys* microbiota, something supported by the statistically significant distribution of two COGs associated with transporters in those assemblies which are either slightly halotolerant or halotolerant.

54 COGs having a statistically significant host species effect might be a result of the limited number of isolates used, especially as what may be minor differences at the population level are exacerbated by only having isolates obtained from two *A. cahirinus* and three *A. russatus* individuals. It would be useful to see if these COGs were still host differentiated if more isolates could have been obtained and from a wider diversity of individuals of each species. The COGs which were host differentiated had associations with a number of functions, amongst them were transporters, transcriptional regulators and proteins which play a role in carbohydrate metabolism. It would be useful to combine the bioinformatic results obtained here with potential future investigation in the laboratory of the actual functional potential of the assemblies through NMR [536] or high throughput metabolomic profiling [537], an approach outlined by Wang *et. al.* [538]. Having live cells from culturing would be useful for these investigations. Future bioinformatic work could take advantage of the existing COG annotations and follow a similar process to that of Satti *et. al.* [539] and look for COGs associated with probiotic functions and compare their abundances from assemblies originating from the two *Acomys* species. It would also be interesting to assess whether the proteins produced by the isolates had more acidic residues and lower isoelectic points as was found by Mongodin *et. al.* [540] in halophilic bacteria and archaea.

It is perhaps unsurprising that there were no genes which had a statistically significant distribution by degree of halotolerance in those isolates tested for it. The limited number of isolates assessed and the fact that multiple of the isolates were chosen based on their sequence similarity to either each other in general or one other in particular will also have impacted the detection of any differentially abundant genes associated with halotolerance. Less expectedly was that the host species did not have any impact on the distribution of the detected genes; especially given that there were isolates from an entirely different genus host. It may be the case that the use of selective media to isolate out LABs has constrained the range of genetic diversity which might be found, the limited genetic diversity recorded in a number of LAB species and strains [541] suggests that this may be compounding the effect of using a limited number of isolates from an even more limited number of hosts. The gene *adk* which was amongst the 10 most commonly detected genes was also found to be induced in *Lactobacillus plantarum* when cells of the LAB passed through the mouse GIT by Bron *et. al.* [542]. The absence of environmental sampling to obtain isolates means that a similar investigation, of differences in expression and abundance of genes in the wider environment and after transit through the host could not be carried out in this investigation.

There were 355 genes which were not detected in any *Acomys russatus* or *Apodemus* assemblies but were found in at least 1 *Acomys cahirinus* assembly; though none were found in all. There were 355 genes which were not detected in any *Acomys cahirinus* or *Apodemus* assemblies but were found in at least 1 *Acomys russatus* assembly, though again none were found in all assemblies from isolates originating from the species. There were 122 genes found only in *Apodemus* derived assemblies, including 7 found in each assembly. That there were no genes present in all assemblies from an *Acomys* species which were not present in assemblies from the two other rodent species meant that it was not possible to conduct an analysis like that by O'Sullivan *et. al.* [543] to detect niche specific genes; potentially a greater number of isolates from a wider number of hosts might have revealed some host-specific genes. There may be a link between the gene *NoxE* which was found in all *Apodemus* assemblies and none of the *Acomys* assemblies and the diet of the *Apodemus* given the reported [544] role of the gene in decreasing redox potentials through removing oxygen; though the species was not the subject of this project and in any event information on the diets of sampled indivdiuals would be required.

A potential probiotic role for the strains within the *Acomys* microbiota can be suggested through the detection within the assemblies of genes with known probiotic functions. The gene *luxS* has been shown [545] to have probiotic effects through its role in the production of AI-2 and indirect role in the production of AI-3-like agonist molecules; it is found in 22 of the *Acomys* isolate assemblies - and in all the *Apodemus* isolate assemblies. 24 of the *Acomys*, and again all of the *Apodemus*, isolate assemblies contained the gene *dltD* which has been shown [546] to have immunomodulatory effects through d-Alanylation of LTA. Probiotics intended for human use are valued for a lack of antibiotic resistance genes, which will not directly equivalent in the *Acomys* is still a beneficial trait for a probiotic in the wild as it would not be possible for a resistance gene to spread to a pathogen from the microorganism. Bryan *et. al.* [547] investigated tetracycline resistance in *E. coli* strains isolated from a number of animal and human sources looking at multiple tetracycline resistance genes. Of 14 tetracycline genes they tested, only three were found in isolates from *Acomys* in this study; *tetA* in 16, *tetM* in 12 and *tetO* in 11. Riboflavin production is also a trait associated with probiotic bacteria [548] and the gene *ribZ* which is involved in this process was found in 18 of the *Acomys* isolate assemblies and none of the *Apodemus* isolate assemblies.

Production of EPS for adhesion to the intestinal mucosa by forming biofilms is a common trait in probiotic bacteria [549, 550] - as well as some halotolerant ones [551] - and 6 *eps* genes were detected from the isolate assemblies. These were *eps*D found in 2 *Acomys*. isolate assemblies, *eps*E found in 1, *eps*F found in 12, *eps*H found in 6, *eps*J found in 7 and *eps*L which was found in 14. Another set of genes associated with increased survival in the host and so believed to be useful for a possibly probiotic microorganism are the *dlt* genes [552]. Of the *Acomys* isolate assemblies, 25 contained the gene *dltA*, 25 the gene *dltC* and 24 the gene *dltD*. The gene *dgkA* has been linked previously to immunomodulation [553, 554], a useful trait for a probiotic, and was detected in 18 of the *Acomys* isolate assemblies. 10 of the *Acomys* isolate assemblies had the gene *pfkA* detected, which both suggests they are homofermentative LABs [555] and was one of the two genes associated by Brownlie *et. al.* with inhibition of growth of commensal Lachnospiraceae and Muribaculaceae [514]. That there were isolates extracted lacking the gene associated with inhibition of the growth of these families and given the low number of bins classified as members of the Lactobacillaceae suggests that the Lachnospiraceae may outcompete these homofermentative LABs under normal conditions in the *Acomys*

4.11 Halotolerance of select LAB isolates

The halotolerance of the selected isolates was tested on the basis of the well-reported halotolerance of both *Acomys* host species, discussed in more detail in **Sub-subsection 2.1.2**. There have also been multiple investigations into the halotolerance of different LABs [556, 557, 558, 559, 560] so the potential of finding halotolerant bacteria within the microbiota of halotolerant hosts warranted the focus on salinity stress. The regenerative properties of both *Acomys* species, in particular *A. russatus*, have been discussed earlier and are the subject of much ongoing work, however the author did not perceive a way of testing any potential links between this interesting host phenotype and the microbiota without requiring animal experimentation which was beyond the scope of this project though may be of interest to other researchers in the future. Other investigations in the future might consider related stresses to halotolerance such as cryotolerance or even try replicating a wider range of conditions to assess the growth of these isolates under different stress conditions in addition to salt stress. The growth conditions used here were intended to be representative of the *Acomys* intestines, so anaerobic conditions were used. A positive control was included, a bacterial strain reported to be capable of growth in saline conditions and two *Apodemus* faecal sample isolates as representatives of lactic acid bacteria from rodents not adapted to arid conditions. To the best of the author's knowledge this represents the first attempt to test the growth of *Acomys* derived microorganisms under any challenge conditions in the laboratory, and potentially the first halotolerance assessment of any bacteria isolated from arid-adapted rodents.

The results of the three replicates of the halotolerance testing are broadly similar, in most of the isolates there is only slight variation in growth across the three replicates. The first notable result was the failure of 18D to grow at any salinities including the 0% and 10% salinity in two of the three replicates; managing to grow at 0% salinity in replicate 2. The failure to grow at 0% is interesting as this was simply the media initially used for the isolation of all the isolates; it is also not the case that it is an obligate halophile as it fails to grow in any salinity. The isolates were all stored in the same manner after the initial isolation through culturing so it is possible that the 18D stock was damaged or otherwise severely degraded during storage, 18D might also be a strain typically found elsewhere in the *Acomys* GIT which survived until reaching the faecal pellet from which it was isolated. In that case the growth conditions in terms of oxygen level, acidity or otherwise might have been quite different from its typical growth conditions; explaining the lack of growth. The positive control was chosen as published literature indicated the species was capable of growth at most of the salinities used in the experiment, it is possible that the osmoprotectants it typically employs in its normal habitat were absent from the modified MRS media [561].

The two *Apodemus* isolates grew under a wider range of salinities than the author had anitcipiated, given that they are intestinal microbiota LABs from a temperate-adapted rodent species which does not have a published history of halotolerance like the *Acomys* have. Though they did not grow at 3.5% salinity (approximately same as sea water) in all replicates each of the two isolates did show some growth at this level in at least one of the three replicates. In all three replicates S122 managed to grow and achieve a stable population in the 2.5% salinity media which led the author to categorise it as 'Slightly Halotolerant'; though it appears to show some growth in 3.5% salinity media this is to a low enough level that it cannot be reliably distinguished. The growth of S128 across and within the replicates shows an interesting trend, for those salinities where it managed to grow it rapidly peaked and then steadily declined before plateauing at a small but stable population in media with 2.5% salinity; in 3.5% media it seemed to experience rapid growth but then death of the population rather than maintaining a stable population. This suggests that S128 is not halotolerant but is capable of surviving in saline conditions for a short

period of time, or potentially that it is halotolerant but only marginally and after the initial population growth from the provision of nutrients in the media the salinity stress becomes too great for it to maintain a relatively large population. Without more knowledge of the salinity of the diets which *Apodemus* consume along with the availability of water it is hard to say whether this is a meaningful result or not, the isolates are from a single sampling of two individuals and so considerable further sampling in combination with dietary analysis (similar to that by Sato *et. al.* [562]) would be necessary to provide firmer conclusions. It is also quite possible that the diverse and varied intestinal microbiota of the *Apodemus* [528] contained some LABs which were slightly or mildly halotolerant through adaptation to some other factor and conditions of elevated salinity were simply not encountered routinely by the microorganisms.

The considerable differences between the three isolates from *A. russatus* 41 demonstrate that there can be considerable phenotypic variation within closely related microorganisms. Looking at the phylogenetic tree of the isolate assemblies, **Figure 3.32**, shows that the three assemblies are placed close together either in the same subtree as the *Apodemus* isolate assemblies or on a very close branch. All three have different genera both from each other and from the *Apodemus* assemblies in the GTDB-Tk results but still are amongst the most closely related phylogenetically. 41A and 41B both show great fluctuation in the readings across all replicates which suggests it is not a technical issue but still makes it hard to determine anything aside from the general trends which is that both show growth at salinities up to and including 3.5% though do best at either 0% or 1% salinities. 41C on the other hand shows growth at 5% salinity in all three replicates, giving it the greatest salinity any of the isolates managed to grow at, albeit this is rapid and the population is fleeting. It is interesting that it shows a similar pattern though with a lower population achieved and a slower growth rate in the 3.5% salinity media. The lack of any growth at either 7.5% or 10% show that it is not an obligate halotroph which had not yet been provided media with enough salt so the dichotomy between the growth at 3.5% and 5% is unusual. The results for 41C are closer to what the author had anticipated for the positive control given that 41C was classified as the same species as the positive control by GTDB-Tk.

The three replicates all had similar results for isolates 13E, 16aC and 16bD. 13E and 16aC had a greater than 99.9% ANI similarity and so 16aC was not used for mapping, annotation or phylogeny - hence it being absent from the phylogenetic trees and not discussed in those sections and subsections. It was included in the halotolerance testing to assess if the difference in ANI similarity might be reflected in different halotolerances. Both isolates show growth at salinities up to 3.5% in all three replicates, though they have different patterns at this salinity in the replicates. In replicates 1 and 3, 16aC showed a sharp increase in cell count after the halfway point through each replicate but started declining after this initial rapid growth. In replicate 2, with a longer run time it is apparent that this decrease was not returning the population to zero but instead to a plateau which it started to maintain at a stable level. 13E shows a very similar growth pattern at 3.5% in replicates 1 and 3 but in replicate 2 it undergoes a more significant drop in cell population after the initial rapid growth. The similarity in the response to the salt stress, in addition to the ANI similarity and shared GTDB-Tk classification, suggests that the two isolates may indeed be the same strains isolated from the two different *Acomys* host species. 13D shows different results for each of the replicates at 3.5% but similar results for lower salinities across all three replicates. 16bD is interesting in that while it grows at 3.5% salinity it does so quite a while after the time point when it started growing at lower salinities. It is possible that this indicates a halotolerance strategy in which the bacteria need to produce sufficient solutes [563] to survive and begin replicating; or that the salinity of the media impacts the uptake of nutrients from the media [564] which slows growth and reproduction.

Chapter 5

Conclusion

- The main findings of the study are reiterated
- The limitations of the study are discussed
- The future work that could be done to build on the study is presented

5.1 Major findings

5.1.1 Issues arising from limited reference databases

The taxonomic classification results highlight how important it is for metagenomic investigation to have diverse reference databases. The low proportion of reads which were classified by these tools using their marker-based approaches demonstrate the necessity to expand the relevant databases, clearly in the case of Metaphlan and with the caveat that mOTUs in its operation only classifies a small subset of sample reads. Metaphlan did not do as well as the author had anticipated when classifying the mouse (*Mus*) samples meant to provide a well-studied rodent benchmark. This could be as a result of chance, the samples used may happen to contain a greater proportion of microbial taxa which had not been included in the reference databases. Of the four classification tools used, each broadly representing a different approach to classification, Metaphlan is in theory the most dependent on reference databases; as it uses unique single copy marker genes. If the entire genome of a microorganism has been sequenced, save for the specific marker it looks for, Metaphlan will not be able to classify that microbe to the particular level indicated by the marker. It may be able to classify to a higher, less precise phylogenetic level; though this to is dependent on the presence and detection of the necessary marker sequences. The low level of classification for the *Acomys* samples by Metaphlan shows that reliance on markers can cause significant classification failures when taxa present may have never been sequenced previously. In the absence of a reliable method employing a non-marker based approach it is not possible to tell whether mOTUs avoided this issue through the choice of markers for its database, though the intent of the project was not to determine which approach is superior in any event. Increasing the read depth did not noticeably increase the percentage of reads classified by Metaphlan nor change the phyla level relative abundance results significantly for mOTUs. Combined these factors highlight that while marker-based approaches can be useful in specific cases, the data in this project was not one of them.

The kmer-based approach of Kraken 2 did not encounter much more success than the marker-based approaches. The ability to modify the required minimum confidence score does give the tool, and approach, greater versatility when working with well researched sample types. In this project it was clear that the alteration of the minimum required confidence score significantly impacted Kraken 2's ability to classify the reads; visible with the *Acomys* and *Mus* samples in particular. When initially planning the work the author had looked for guidance on the suggested minimum required confidence score and found there was no definitive value. 10% was commonly suggested so the author chose it as the most permissive setting and then used a range of values up to 95%. The median percentage of reads classified by Kraken 2 show that the *Acomys* samples are difficult to classify using a kmer-based approach, though the *Mus* samples also appear difficult to classify using Kraken 2. Given that Kraken 2 could classify the reads at any level from kingdom to subspecies it is clear that the kmer reference database used (the most recent at time of conducting the work) does not contain much of the diversity found in the *Acomys* faecal microbiota, or in the particular *Mus* samples used. The >90% classification of reads in the human mock community using Kraken with minimum confidence scores of 10-50% confirms that there is not a functional issue with the tool or the database itself. If working with samples from a less studied environment and using Kraken 2 it would be prudent to experiment to determine what minimum confidence threshold provides the highest level of classification while minimising the risk of false positives. Notably, Kraken 2 analysis of the read files prior to subsampling did not increase the percentage of reads classified using a 10% minimum confidence score. This strongly suggests that the bulk of the *Acomys* faecal microbiota is absent from the Kraken 2 database to the extent that even with the most permissive setting they cannot be classified

to at any level up to and including 'Bacteria', rather than a large classifiable component being removed by the subsampling. Kmer-based methods theoretically provide greater classification potential for microbiome sequencing data though with some risk of false positive classifications. The most permissive minimum confidence score used with Kraken 2, 10%, did generate a large number of false taxonomic IDs from the human mock community. The strictest setting, 95%, still produced false positives - however this is using a very conservative definition of a false positive result. It seems likely that unless the microbiota being investigated has been studied previously, kmer-based classification of microbiota sample reads will have limited success as the majority of the community will be unclassified and without some biological interpretation false positives cannot be distinguished. Kaiju 3 offered a third approach, using translation of the genomic reads into amino acids and then searching a protein database for matches. It appears to be much more successful at classifying the *Acomys* reads, managing to reach a minimum of 50% reads classified in all *Acomys* samples. However the large number of false positive results from the human mock microbiome suggest that this could also be the case with the *Acomys* samples, but with no easy way of discerning them from any true positives provided by the tool. If working with a more studied microbiome there may be prior studies that provide a guide for distinguishing likely false positives from true positives; which would provide a better use case for Kaiju.

5.1.2 The *Acomys* faecal microbiota

This work was the first investigation of the microbiota of any *Acomys* rodent, and the first of an arid-adapted rodent to the author's knowledge. Through the use of taxonomic classification software along with mapping of metagenomic bins a number of taxa were identified with members likely to be found in the *Acomys* faecal microbiota. The results of the taxonomic classifiers indicate there is a proportion of the *Acomys* microbiota made up of LAB, specifically members of the Lactobacillaceae. These are present in both *Acomys* species and in both June and November, suggesting that whatever their actual abundance as a percentage of the entire microbiota they are consistent (if not core) members. That they are classifiable at all also indicates some similarity between the lactic acid bacteria found in the *Acomys* and those of other rodents, leading to their presence in reference databases and allowing for classification. The taxonomic classifiers indicated that the *A. cahirinus* and *A. russatus* microbiota in November contained fewer *Helicobacter* and *Akkermansia* species than they did in June. Both *Acomys* species also showed increases from June to November in the percent of reads classified within the Clostridia, according to the Kraken 2 results. As mentioned in the discussion there have been prior reported examples of wild rodent microbiomes changing between seasons [250], though the authors linked this to observed dietary changes. The *Acomys* diet can change between June and November though not as significantly as the *Apodemus* in the aforementioned study.

The LAB detected by the classifiers might simply be the only proportion of the microbiota which could be classified, this is buttressed by all four classifiers detecting these kinds of bacteria though varying in relative abundance and specific taxonomic ID assigned. Within the *Acomys* it is possible that the LABs serve a similar probiotic role as they do in human and Mus intestinal systems [252, 251], suggesting that there may be a rodent-wide (or even broader) conservation of Lactobacillaceae which provide benefits for host animals. The creation of metagenomic read bins from *Acomys* faecal samples revealed some other constituents of the *Acomys* faecal mirobiota. It is interesting to note that there was almost no overlap between the taxa detected by the classifiers and the taxonomic assignments to the bins; only 6 of the 348 bins were classified within the Lactobacillaceae. This alone highlights the importance of applying multiple methods to microbiota investigations; especially when dealing with less-studied sample types. This does support the idea that the taxonomic classifiers worked as intended in this case, but that they

were not necessarily appropriate for it. The 6 bins of 348 are around 2% of the total, the LAB classifications from the tools might be detections of these combined with a normal level of false positive classification and inflation of diversity. That many of the bins were classified into the Lachnospiraceae, Muribaculaceae and Ruminococcaceae is not surprising; these families have all been previously identified in mammal microbiome samples. Species which produce Short-Chain Fatty Acids (SCFAs) are found in the Lachnospiraceae and Ruminococcaceae, these have been demonstrated to have beneficial effects for the host organism [492]; within the *Acomys* they may play the same role and thus be selected for. Interestingly, Lachnospiraceae have also been linked to disease states in humans [565] and stress in mice [566] alongside their beneficial traits. The phylogenetic trees produced from analysis of the bins suggests that the major families identified cover a number of genera and species within the *Acomys* microbiome; though given the low mapping of the reads to the bins it is likely that a large proportion of the microbiota is not covered by the trees. The inability to classify all but 12 of the 348 bins to the species level again indicates that the taxonomic classifiers did not fail in their analysis, instead the *Acomys* microbiota contain a large amount of novel diversity.

Mapping of the faecal sample reads to the bins shows some differentiation between the two host species; more so than is seen with the results from the taxonomic classifiers. It was surprising that the level of mapping was comparatively low, between 10-35% depending on the read depth of the file and the host species and collection month of the sample. This is a higher level than the percent of reads classified by some of the classifiers, though it does suggest the bins might still be only a proportion (albeit a larger one) of the *Acomys* faecal microbiota. That the bins were produced from all sample reads co-assembled together prior to binning may explain this, it is possible that the assemblies produced are chimeric and constitute closely related genomes assembled together incorrectly. It is also possible that the majority of the reads obtained from the sequencing of the *Acomys* faecal samples were from rarer taxa which were not abundant enough to be assembled; lacking the requisite coverage of the genome. Perhaps more plausibly the *Acomys* microbiota contains a great deal of taxonomic diversity with some taxa present in greater abundances - those represented by the bins - and the majority being in low abundance in each individual; thus not producing any bins. If considering a similar investigation in the future the author would repeat the process used here but also carry out individual assemblies from the reads; along with aiming for greater sequencing depth to ensure as much of the bacterial diversity present was captured in the assemblies. The actual mapping results themselves point to some host and some slight seasonal differences in the relative abundance of certain bins.

A. russatus had a more consistent microbiota compared to *A. cahirinus* in terms of seasonal change, but both *Acomys* species saw comparatively little change from June to November from the mapping results. This demonstrates the utility of this manner of investigation as one does not need to rely on comparing percentages or abundances of specific taxonomic classifications to look for seasonal differences. It is possible that there are seasonal differences within the two *Acomys* species which could not be detected due to the taxa involved not being represented by any of the bins. Comparing the mapping results across all samples from each of the two species indicates there are a large number of bins which are significantly different between them, 50.5% of the 348 bins. Twelve bacterial families have bins enriched in both *Acomys* species, indicating that there may be diversity at the lower taxonomic level, i.e. genus and species, between the two hosts which is not visible at higher levels. This is bolstered by those families including the ones most commonly assigned to the bins and which have been previously reported in the gut microbiota of mammalian species. There are some differences hidden amongst the larger overall similarities between the two species. Despite the fact that the Lachnospiraceae is the most common classification of a bin by GTDB-Tk at the family level, 51 of the bins enriched in *A. cahirinus*

are from the Lachnospiraceae while only 2 of the bins enriched in *A. russatus* are. A largely shared microbiota, at the family level, between the two species is supported by the finding that there are only a comparatively small number of bins assigned to families which are exclusively enriched in one host. Though there are 109 bins enriched in *A. cahirinus*, only 12 of them are assigned to families enriched exclusively in *A. cahirinus*. In *A. russatus* it is 17 bins of the 67 enriched which are assigned to families exclusively enriched in *A. russatus*. It is possible that the reported predisposition to obesity and diabetes in captivity in *Acomys cahirinus* might be linked to the Lachnospiraceae in their microbiome and links between the bacterial family and diabetes in mice [567].

Taxa which are known to be beneficial to the host organism, as well as known to provide fermentative or digestive functions, are found in the *Acomys* microbiota. It is possible that these taxa are providing the same or similar functions in the arid-adapted rodents, especially in light of their difficult to digest diets. Binning and then mapping allowed the examination of a larger proportion of the microbiota than would otherwise have been possible using the taxonomic classifiers alone and goes some way to showing an approach which can be taken to less-studied microbiota samples. The annotations of both the assemblies and the bins was limited due to the lack of any direct traits to relate detected genes to in the case of the bins and the restriction of only assessing halotolerance in the assemblies.

5.1.3 The utility of culturing microbes

The isolation of 25 lactic acid bacteria from the *Acomys* faecal samples was an interesting and highly complementary aspect of the work. The mouse (*Mus* sp.) gut microbiota has been studied for many years and still routinely has new cultured bacteria added to the reference collection [294] ; demonstrating the importance of culturing microorganisms. The benefit of culturing for exploration of microbial communities even when also employing bioinformatic approaches has been highlighted by Browne *et. al.* [568] for the exploration of complex communities; such as those of the *Acomys*. At the time of planning the isolation the only results available were from the different taxonomic classifiers, with the associated issues discussed above. As the classifiers only really indicated the presence of LABs, especially from *Lactobacillus* itself, it seemed likely that it would be possible to isolate at least one or two strains from the samples. It is notable that despite no particular precautions around storage being taken the faecal samples still provided more than 25 viable colonies; including multiple copies of the same isolate in some cases. This suggests that if working with faecal samples from *Acomys*, provided they were frozen within at least 24 hours of collection, they can still be sources of live microorganisms after 5 years frozen at -80℃ without being kept in protective media. That live LABs were isolated from the faecal samples also confirmed that there are LABs present in the *Acomys* faecal microbiota, supporting the results from both the taxonomic classifiers and the classification of the metagenomic bins. Without using multiple different media types, including non-specific media, the author cannot say how much of the viable microbiome these LAB isolates represent. It is impossible to say from the results here whether the 25 unique isolates cover a large proportion of the LABs in the *Acomys* faecal microbiota, further sampling and isolation would be needed to determine this. The isolates are, naturally, all LAB; being successfully selected by the media. In light of the results from the GTDB-Tk analysis of the metagenomic bins it would have been interesting to try cultivating bacteria from the faecal samples on a wider range of media. It is possible that one might have been able to isolate out and sequence bacteria which one could match to the bins; confirming for certain both the validity of the bins and their presence in the *Acomys* microbiota. The LAB isolated from the *Acomys* samples were, with a couple of notable exceptions, phylogenetically distinct from those obtained from *Apodemus* mice. As can be seen using the phylogenetic trees,

they cluster separately. This matches and supports the different GTDB-Tk genus classifications obtained for them. The classification of a couple of the *Acomys* isolates as the same species as the classification assigned to all of the *Apodemus* isolates suggests that there is at least one *Ligilactobacillus* member spread across quite different rodent species. The *Acomys* and *Apodemus* samples come from sites separated by thousands of miles and of distinctly different types. It is likely then that the two species are not identical but represent different, but closely related species which fulfil some shared function; potentially relating to some dietary similarities regarding chitin or lignocellulose degradation.

The isolation of *Bifidobacterium* species from the faecal samples is interesting, they are also lactic acid bacteria and the intended targets the media was originally developed for, given that they were not suggested by the bioinformatics analysis to be particularly abundant. It is useful to be able to classify the isolates down to the genus level, especially as the bins could only be identified in most cases to the family level and that so few of the bins were classified within the Lactobacillaceae. Culturing in this instance allowed the comparison of different bacteria of the same type, representing the only easily identified section of the microbiota from bioinformatic analysis. It highlighted the diversity which could be found within a small proportion of the sequenced reads independent of any issues with binning or trying to capture enough information for database dependent classifiers. It was beneficial to have the assemblies produced from the isolates when mapping the metagenomic reads to them. The presence of the isolates in the faecal microbiota of at least one individual from either *Acomys* species is confirmed, self-evidently, by isolation from the faecal samples. In contrast to the bins, it is possible to say for certain that any mapping to an isolate assembly is indicative of the presence of either the isolate itself or something very closely related. Looking at the Reads per-million (RPM) for the mapping of the shotgun sequencing reads to the assemblies by the original species the isolates were sourced from was interesting. There were some isolates where the reads from the same *Acomys* species had higher RPMs than reads from the other host, most of those sourced from *A. cahirinus* fell into this category. There were some isolates sourced from *A. russatus* which had similar RPMs across both host species; more so than isolates sourced from *A. cahirinus*. This supported the results suggesting greater Lactobacillaceae content within the *A. cahirinus* faecal microbiota. Though mapping to most isolates was low from all shotgun reads there were more isolates which had appreciable mapping levels at all for the *A. cahirinus* samples. There were some isolates which appeared to show seasonal differences in RPMs across both *Acomys* species, 18A_S9 and 39B_S15 had higher RPMs for reads from November in either species than the reads from June. The mapping results suggested only a minor seasonal effect and not from any of the 6 bins, out of 348 total, classified within the Lactobacillaceae. This highlights a side benefit from the culturing of bacterial isolates to obtain assemblies; a potential seasonal difference is detected which was not found using either of the bioinformatics approaches. Though not conclusive, it is useful to have ANI comparison scores for the bins to the isolate assemblies to provide some support for the validity of the binning. That bins classified as Lactobacillaceae had ANI similarity scores of 78-80% to a number of isolate assemblies is reassuring, it supports both the GTDB-Tk classifications and the content of the bins being accurate.

It would also have been impossible to conclusively test the hypothesis concerning halotolerance in the *Acomys* microbiota without isolates. In retrospect it would have been more useful to lower the upper threshold salinity values used to 5% as there were no isolates which grew at 10% salinity. Even the second highest salinity tested, 7.5% might have been too high to provide any useful datapoints. The comparison between the *Apodemus* and *Acomys* assemblies gave some suggestion that the halotolerance of the *Acomys* might be mirrored in the microbiota; or at least that a non-halotolerant host did not provide halotolerant isolates. Multiple Acomys isolates

did grow at 3.5% salinity, approximately the same value as sea water. This shows those strains possess a degree of halotolerance, capable of growing at a salinity level equivalent to sea water. Conducting the halotolerance experiment with isolates allowed the observation of variability in halotolerance within closely related taxa, with the isolates 41A and 41B having different growth rates at different salinities. In order to assess the halotolerance of other microorganisms found in the *Acomys* faecal microbiota it would be necessary to carry out a similar halotolerance culturing experiment with isolates obtained using other types of media. It is quite possible that the isolates are unusually halotolerant for rodent-associated LAB. Dong *et. al.* [569] found that high salt (1 mL of 10% NaCl solution three times a week) treatment in Wistar rats led to reduced prevalence in *Lactobacillus* in the faecal samples of their treated animals. The isolates obtained from both *Acomys* and *Apodemus* used in the halotolerance culturing might be more halotolerant than other rodent-associated LABs, testing of more isolates would be necessary in order to establish this. It is conceivable that the relatively high salt diets of both *Acomys* species, from salty plant matter, leads to a limited range of LABs being able to survive in the *Acomys* microbiota. This would explain the extremely low number of bins classified within Lactobacillaceae. At the same time this small proportion of the microbiota appears to be the most amenable to detection by the taxonomic classifiers; likely due to reference databases lacking any near relatives of the bulk of *Acomys* microbiota species. Miranda *et. al.* [570] found increased salinity led to decreased abundance of *Lactobacillus* species in specific pathogen-free C57BL/6 mice; they also noted decreased production of SCFAS. The isolates obtained using the LAB selective media might be the only ones present in the *Acomys* microbiota as a result of the highly saline diet - though the level of *in situ* salinity in either *Acomys* species intestinal tract has not been measured to the author's knowledge - and thus capable of growing at salinities up to 3.5%. This might then mean that the Lachnospiraceae, Muribaculaeceae and Ruminococcaceae detected amongst the bins play a greater role in SCFA production (and other benefits to the host) than in other rodent species.

5.2 Challenges and limitations

The parasitome and viriome were not considered in this project, though this was a combination of both the pandmic and the original plan for the work. The author would have been quite interested in examining the viriome, as it has been shown to have a potentially major impact on the microbial communities associated with animal hosts [571, 572, 573, 574, 575]. Research into the viriomes of different wild animals [576, 577] has uncovered both novel viruses [578] and - concerningly in light of the events of recent years - pathogens which pose a threat to humans [579]. The viriome of rodent species have been previously investigated, including wild and domesticated animals [580, 581, 582]. Raghwani *et. al.* [583] recently published the results of their investigation into the viriome of three wild rodent species. They found an interesting transience in a large proportion of the viriome by season and that most of the viruses infected vertebrate or bacteria; in addition though there was some shared viral content between the three species (especially the two more closely related) the majority were host specific at the inferred species level. They also note that these results are not unique to their study [584] and that most of the vertebrate-associated viral genera they detect have been previously reported in other wild rodents from geographically distant samplings [585]. A number of tools have been developed for the study of the viriome, amongst them are VirSorter [586], VirFinder [587] and DeepVirFinder [588]. It is also possible to try and extract viral content from shotgun sequencing of a sample using an approach like that used by Xiong *et. al.* [589] and sequencing approaches meant to increase the sensitity of viral detection have been developed [590] which might be useful for possible investigation of the *Acomys* intestinal viriome. The author did not have prior experience in dealing with viral data and the time constraints of the project meant that it was not possible to investigate the viriome of the *Acomys* samples.

The eukaryotic component of the microbiota was not ignored by the author, however none of the bins produced from the shotgun reads were classified as anything other than bacterial taxa. Both Kaiju and Kraken 2 included eukaryotic taxa in their reference databases and no steps were taken to remove this apart from the use of the host and human genomes for contaminant removal. Despite this, Kraken 2 only detected any eukaryotic reads at all (at the 50% minimum confidence score the author determined to be the most reliable) and this was 0.01% of classified reads in a single sample, AR42N; an *Acomys russatus* sample from November. The sole fungal taxa it classified was *Candida glabrata*. Kaiju when run with an error allowance of 0, the strictest setting and the one the author considered the most reliable, detected fungal taxa in all samples but always assigned < 0.1% reads to them; typically it was < 0.05%. In light of the previously discussed results from the tool testing with Kaiju detecting multiple false positives the author does not believe these results reflect a reliable true positive. In light of this the author believes that either there is no meaningful fungal (or eukaryotic more generally) component of the *Acomys* microbiota, or more likely and in accordance with other studied rodent microbiota, eukaryotic taxa are present but within the majority of reads which could be neither classified nor assembled.

5.2.1 Inherent limitations of the project

The project had some inherent limitations courtesy of the experimental setup as it was originally conceived. As the samples were collected as a side effort of a larger project being conducted by collaborators much of the metadata typically expected in metagenomic studies was not collected; in addition to the lack of an established method for determining the age of wild *Acomys* individuals. This meant it was not possible to assess the potential impact of host sex on the faecal microbiota. An impact of host animal sex on the faecal microbiota has been observed in

a number of species [591, 592], though also has been found absent, playing only a minor role or only present at certain life stages in other studies [593, 594] but could not be measured in this project. Age also has been seen in a number of species to have an influence on the microbiota of different body sites [595, 596, 597, 598] but could not, in wild *Acomys* in the enclosure, be investigated in this experiment as it would be impossible to do anything more than measure age of caught individuals as juveniles or not juveniles; even then this would be a subjective judgement on the part of the individual collecting the data. Though, as with an impact of sex, there have been publications finding more complex interactions between age and the microbiota; some finding it be only a minor influence [599] or instead of clear age-associated communities there is a personalising [600] impact of age on the microbiota composition.

The project as originally designed only made use of short read sequencing for the shotgun metagenomics. In an ideal situation, all samples would have been sequenced using long read sequencing and the author would have limited the use of the short read data to assist with assembly. Long read sequencing allows for much greater accuracy when assigning identities to potential taxa in a metagenomic sample and would ensure the assemblies and bins produced from the reads were of a higher quality [601, 602, 603] than those produced using solely short reads. Long read sequencing would necessitate the use of dedicated taxonomic classification software designed for long read files such as MetaMaps [604] or BugSeq [605]. Whether these tools would have produced similar results to mOTUs, Metaphlan 3 and Kraken 2 from the *Acomys* samples would have been interesting to discover, possibly though the long-read sequencing would have captured a greater proportion of the indiviual genomes present so the taxonomic classification would have been easier for the software.

The planned use of only metagenomic analysis also limited the study. A more detailed investigation of the microbial communities of *Acomys cahirinus* and *Acomys russatus* should also employ metatranscriptomics [606, 607], metaproteomics [608] and metabolomics [609]. This would especially be the case when trying to assess whether the microbiota is providing particular useful functions for the host to allow it to survive in an arid environment. The lack of transcriptomic and proteomic data inherently limits the ability to link any metagenomic findings to adaptation to arid life; the presence and content of microbes can be reported but this projet cannot say definitively what genes they are expressing and what proteins are produced. Some of the major interactions between the two host species and the microbiota can only be inferred without direct metabolomic data to confirm the presence of relevant metabolites.

Samples were not frozen immediately upon collection or until the conclusion of fieldwork; and were stored at -20℃ prior to being shipped to the UK. In the case of the June samples this meant being stored at this temperature for a number of months until the November sample collection. The methods and conditions of storage of faecal samples can have an impact on the measured composition of the microbiota, as reported by Franzosa *et. al.* [610], though the nature and extent of these differences has been observed to be smaller than interindividual differences [611] and given the sampling was carried out as a side component of work by collaborators not focusing on metagenomics it was not possible to use the 'gold standard' approach. Aside from the 12 samples sequenced as a 'pilot project' all samples irrespective of collection date were stored at -80℃ for the same length of time, which was a number of years. Ideally the samples would have been stored securely and safely at the moment of collection and sequenced as soon as possible following collection; had more sampling been carried out this would have been possible.

5.2.2 Consequences of the COVID-19 pandemic

Some other limitiations of the project had been planned to be addressed but this intended fieldwork, sampling, sequencing, culturing and collaboration was prevented by the COVID-19 pandemic and the associated legally imposed restrictions on travel, number of individuals in close proximity and home working both in the UK and in Israel. In addition to these legal obligations, resources and personnel which the author had planned to employ were instead - and entirely correctly - repurposed to work in COVID-19 testing. The author will describe below some steps which were being planned to be carried out during the project to address some of the previously mentioned limitations along with others yet to be discussed.

Restriction to two rodent species

The author and collaborators had discussed the potential of including a third and more distantly related rodent species in the project, the Fat Sand Rat (*Psammomys obesus*). These were also being used in their research by collaborators in Israel and would have been extremely useful for determining whether given taxa were aridity-associated or instead linked to the *Acomys* genus specificaly. This work would have involved taking faecal samples from indivdiuals in a colony maintained by collaborators and sequencing them using either the same methods as the already collected *Acomys* samples or those discussed below. It was prevented from progressing into detailed planning and execution by the pandemic.

Restriction to a single sampling site

The author and collaborators had plans to sample *Acomys cahirinus* along its range within Israel across an aridity gradient, moving from the south where it is sympatric with *Acomys russatus* to the north where there is considerably greater water availability. This would have been a very useful component of the project as it would have directly assessed the impact of aridity on host diet and microbiota within the single species; it would have been possible to link directly the presence or absence of any taxa to the changing aridity levels of the hosts' environment. This work was prevented from progressing into detailed planning and execution by the pandemic.

Restriction to a single sequencing approach

The author had been planning to use cell sorting and single cell techniques in addition to the shotgun sequencing already employed in the project - this not being a novel concept as the approach was demonstrated by Woyke *et. al.* [612] in 2009. It was the author's hope to bypass issues around the assembly and binning of reads, even any potentially sequenced with long read technologies, by using the ability to sort and then individually sequence cells obtained from the faecal samples. This may have been in an approach such as that proposed by Arikawa *et. al.* [613] in which a single cell method is combined with assembly and binning to produce superior draft genomes or like from Yang *et. al.* [614] in which cell sorting is used to produce distinct communities comprised of the rarer taxa in a sample for metagenomic sequencing; though these could also be sequenced at the single cell level instead or in addition. The restriction of access to laboratory facilities and the redirection of personnel and resources to the COVID-19 testing effort prevented this from being pursued.

Restriction to a pair of sequencing dates

The author and collaborators had been planning another round of sample collection from the *Acomys* enclosure to provide more data for the project and reduce the impact which might be

caused by one of the samples being, on the day of collection, not representative of the typical *Acomys* microbiota. This would have involved collection of samples on one day through the same method as described in **Section 2.1** and then collection of more samples after a month had passed. This work was prevented from progressing to execution by the pandemic travel restrictions for international travel and within Israel.

5.3 Future work

More studies into arid-adapted animal microbiota will over time lead to more taxa from them being added to the databases but this is a somewhat random process. It would be beneficial to have samples from the environment in which the *Acomys* species lived as would allow for the addition to existing databases of genomes from the environment being sampled through the possible use of long read sequencing, culturomics and single cell sequencing. This would both be of benefit through the expansion of knowledge of arid environmental microbiota more generally and to compare the faecal microbiota of the *Acomys* themselves to the environmental microbiota of their habitats. There have been large scale projects undertaken previously to investigate the microbiota of particular environments or host organisms, such a project for arid environmental microbiota in general or for arid-adapted animals more specifically might be an effective way of rapidly increasing knowledge of associated microbial communities. On a smaller scale it would be a clear next step to collect more faecal samples from *Acomys* of both studied species and collect more thorough metadata at the same time so as to allow the a more granular exploration of the microbiota and potential links to aridity.

Another potential avenue of investigation would be to create more MAGs from more *Acomys* samples. That a very limited number of samples provided 28 distinct lactic acid bacterial isolates while the binning pipeline produced far fewer suggests that there is a great amount of unobserved taxonomic diversity in the microbiota. Further sampling would provide more genetic material for assembly and binning which could help capture low abundance taxa which likely were missed with the relatively low sample size of this study. MAGs or bins thus generated could be used for read mapping to assess the relative abundance and diversity in the two species as well as be annotated to see if there might be predicted proteins which could tie back to tolerance of the arid conditions the host lives in.

That the Acomys isolates did not display any uniform halotolerance highlights that sometimes there may be no links between an observed host and the microbiota. The isolates might provide other benefits for the host which could be assessed through culturing in different media, degradation of tough cellulose found in the diet, production of immunomodulatory compounds or pathogen resistance through secretion of antimicrobials. It is also possible that there may be halotolerant microorganisms within the *Acomys* microbiota which are not members of the Lactobacillaceae. A repeat of the halotolerance experiment with isolates obtained from non- or differently-selective media could be carried out to investigate this possibility.

This study only uses the faecal samples from two species of Acomys, from a limited number of individuals in one location and collected at only two time points. The considerable taxonomic diversity uncovered highlights how much remains to be discovered, classified and characterised. There are many rodent species which have not been subject to investigation of their microbiota to any extent, even if just a single 16S based study with a handful of samples. A future investigation might look at additional members of the *Acomys* genus to see whether there are any phylogenetic influences on the composition of the microbiota, in which case using non-arid adapted members of the genus would be very informative. Other rodents from outside the genus could also be investigated, both arid-adapted and non-arid adapted, to see if there are any general trends in the microbiota of arid-adapted rodents. Both the discussed in **Sub-subsection 5.2.2** Fat Sand Rat and the Karoo Bush Rat (*Myotomys unisulcatus*), which uses more behavioural adaptations to aridity over physiological ones, are good potential options for expanding investigation of arid-adapted rodent microbiota. Cold desert rodents such as the Gobi Jerboa (*Allactaga bullata*) or those which can be found in places such as the Atacama Desert like Darwin's leaf-eared mouse

(*Phyllotis darwini*) would also be very useful for truly establishing connections between a lack of water in an environment and the resulting impacts on rodent intestinal microbiota. A smaller scale study could restrict itself to samples from *Acomys cahirinus* but taken along an aridity gradient in a region where the species lives along this gradient. This would reduce the complexity of comparing two different species and allow for a more in depth examination of the impact of aridity on the microbiota of a single species. The results of such a study might then be compared to those from any other species, rodent or not, which live along an aridity gradient and have been the subject of a microbiota investigation.

A clear next step would be to investigate the relationship between the *Acomys* intestinal microbiota and increased energy harvest from their difficult to digest diet. This could be accomplished through the well-established technique of faecal microbiota transplant [615]. Faecal samples from either species of *Acomys* could be transplanted into laboratory mice or other rodents and then different metabolic markers, along with bodyweight, tracked to observe whether the recipient animals gave signs of increased energy obtained from their standard diet. Alternatively those taxa which were detected in this project and are culturable could be provided to laboratory animals as either probiotics or following a course of antibiotics to avoid FMT while still assessing energy harvest. The inverse might also be accomplished using *Acomys cahirinus*, individuals from less arid environments could be captured and maintained on a limited diet more similar to that of their counterparts to the south. Then they might undergo FMT using faecal samples from individuals of the same species but from more arid environments, again, monitoring of different metabolic and physical markers would be used to assess whether the recipient animals were able to obtain more energy from their diet.

Supplemental Material

Sample ID	Host	Raw file read counts	QCed file read counts	Cleansed and QCed file read counts
AC4J	AC	11,727,416	11,671,551	11,290,129
AC4N	AC	15,717,505	15,659,694	15,258,604
AC5J	AC	11,409,425	11,375,317	11,031,536
AC5N	AC	10,223,406	10,191,262	9,943,029
AC6J	AC	9,880,005	9,827,392	9,606,631
AC6N	AC	16,343,544	16,296,504	15,963,038
AC7N	AC	11,164,541	11,120,692	10,969,173
AC8J	AC	12,254,093	12,223,691	11,869,263
AC11J	AC	10,573,595	10,546,614	10,350,651
AC11N	AC	10,063,240	10,013,082	9,691,932
AC12J	AC	11,803,371	11,720,325	9,912,282
AC14J	AC	12,144,468	12,064,665	9,495,789
AC14N	AC	10,919,104	10,887,042	10,702,299
AC15J	AC	10,436,818	10,403,146	10,191,353
AC15N	AC	15,174,088	15,121,169	14,761,196
AC16J	AC	5,796,897	5,609,613	5,520,203
AC16N	AC	5,405,200	5,250,073	5,212,835
AC18J	AC	5,363,433	5,187,738	5,044,745
AC18N	AC	5,863,150	5,686,142	5,537,977
AC19J	AC	5,212,061	5,021,304	4,902,726
AC19N	AC	6,441,741	6,266,826	5,967,871
AC22J	AC	10,863,612	10,833,072	10,719,067
AC23J	AC	10,214,591	10,179,793	9,994,986
AC23N	AC	13,493,468	13,432,276	13,266,966
AC25J	AC	10,276,657	10,242,923	10,120,754
AC25N	AC	7,853,266	7,816,193	7,619,692
AC28J	AC	11,480,072	11,448,187	11,254,677
AC28N	AC	12,184,202	12,124,524	11,920,243
AC32J	AC	11,541,134	11,510,891	11,377,431
AC32N	AC	10,844,110	10,782,676	10,521,197
AC33J	AC	11,215,196	11,181,231	10,965,491
AC33N	AC	11,357,703	11,316,412	10,948,568
AC35J	AC	12,279,890	12,235,895	11,969,907
AC36J	AC	12,684,697	12,624,428	12,159,375
AC36N	AC	12,261,837	12,223,467	12,100,748
AC37J	AC	10,231,030	10,195,546	9,225,152
AC37N	AC	12,749,776	12,699,931	12,307,329
AC40J	AC	10,268,035	10,236,710	10,069,261
AC40N	AC	11,137,489	11,094,649	11,005,453
AC44J	AC	10,550,305	10,515,003	10,371,678
AC44N	AC	13,181,319	13,140,756	12,999,506
AR1J	AR	12,264,722	12,193,535	11,574,544

AR1N	AR	13,364,387	13,291,733	12,953,372
AR2J	AR	11,053,365	10,975,110	9,699,929
AR2N	AR	9,922,328	9,895,653	9,763,831
AR3N	AR	14,341,441	14,290,978	14,155,217
AR9J	AR	11,830,642	11,786,161	11,607,001
AR9N	AR	10,001,408	9,975,501	9,860,074
AR10J	AR	11,008,145	10,939,596	9,425,939
AR10N	AR	12,276,995	12,240,845	11,784,266
AR13J	AR	11,567,591	11,513,561	11,075,258
AR20J	AR	12,079,940	12,045,102	11,849,293
AR20N	AR	10,520,845	10,481,947	10,297,523
AR21J	AR	10,290,580	10,259,151	10,085,079
AR21N	AR	12,112,516	12,068,484	11,904,111
AR24J	AR	11,359,830	11,333,829	11,184,808
AR26J	AR	9,983,104	9,959,589	9,808,533
AR26N	AR	12,873,047	12,822,233	12,696,889
AR27J	AR	5,527,574	5,333,921	5,240,857
AR27N	AR	5,767,979	5,562,624	5,384,643
AR29J	AR	5,325,115	5,140,792	5,046,617
AR29N	AR	5,812,516	5,624,582	5,564,290
AR30J	AR	12,097,770	12,061,048	11,925,088
AR30N	AR	12,930,571	12,880,249	12,648,152
AR31J	AR	10,698,989	10,643,023	8,618,768
AR34J	AR	4,830,100	4,667,227	4,464,759
AR34N	AR	5,826,643	5,646,727	5,445,810
AR38J	AR	16,911,155	16,834,912	16,437,044
AR41N	AR	13,139,684	13,091,220	12,833,489
AR42J	AR	13,060,085	13,012,309	12,323,643
AR42N	AR	13,539,806	13,490,361	13,241,744
AR43J	AR	9,931,372	9,904,226	9,621,839
AR43N	AR	11,721,155	11,668,609	11,465,039
AR45J	AR	13,377,506	13,322,930	11,668,536
AR45N	AR	11,540,914	11,466,531	10,754,657
AR46J	AR	12,022,981	11,984,661	11,682,493
AR46N	AR	13,190,360	13,132,985	12,913,459
AR47J	AR	11,233,487	11,196,712	11,009,219
AR47N	AR	14,409,732	14,346,323	13,814,780
AR48J	AR	12,130,662	12,095,429	11,923,422
AR48N	AR	12,992,043	12,953,161	12,808,291

Table 5.1: Information about Acomys faecal samples, shows Sample ID, host species (AC - *Acomys cahirinus*, AR - *Acomys russatus*), count of reads in raw file, count of reads in read files after QC and count of reads in read file after QC and contaminant filtering.

Name	Determinant
Aneurinibacillus aneurinilyticus	Kohl et. al. [403] publication results
Arthrobacter koreensis	Kohl et. al. publication results
Bacillus amyloliquefaciens	Kohl et. al. publication results
Bacillus badius	Kohl et. al. publication results
Bacillus cereus	Kohl et. al. publication results
Bacillus gaemokensis	Kohl et. al. publication results
Bacillus haynesii	Kohl et. al. publication results
Bacillus inaquosorum	Kohl et. al. publication results
Bacillus licheniformis	Kohl et. al. publication results
Bacillus nakamurai	Kohl et. al. publication results
Bacillus paramycoides	Kohl et. al. publication results
Bacillus proteolyticus	Kohl et. al. publication results
Bacillus safensis	Kohl et. al. publication results
Bacillus subtilis	Kohl et. al. publication results
Bacillus tequilensis	Kohl et. al. publication results
Bacillus tropicus	Kohl et. al. publication results
Bacillus velezensis	Kohl et. al. publication results
Bacillus wiedmannii	Kohl et. al. publication results
Bacillus zanthoxyli	Kohl et. al. publication results
Bacillus zhangzhouensis	Kohl et. al. publication results
Cronobacter turicensis	Kohl et. al. publication results
Enterobacter hormaechei	Kohl et. al. publication results
Enterobacter kobei	Kohl et. al. publication results
Enterococcus mundtii	Kohl et. al. publication results
Escherichia coli	Kohl et. al. publication results
Escherichia fergusonii	Kohl et. al. publication results
Klebsiella pneumoniae	Kohl et. al. publication results
Klebsiella quasipneumoniae	Kohl et. al. publication results
Lysinibacillus boronitolerans	Kohl et. al. publication results
Mammaliicoccus sciuri	Kohl et. al. publication results
Paenibacillus amylolyticus	Kohl et. al. publication results
Peribacillus simplex	Kohl et. al. publication results
Priestia aryabhattai	Kohl et. al. publication results
Priestia filamentosa	Kohl et. al. publication results

Priestia flexa	Kohl et. al. publication results
Priestia megaterium	Kohl et. al. publication results
Rossellomorea marisflavi	Kohl et. al. publication results
Rothia nasimurium	Kohl et. al. publication results
Rummeliibacillus stabekisii	Kohl et. al. publication results
Staphylococcus capitis	Kohl et. al. publication results
Staphylococcus gallinarum	Kohl et. al. publication results
Staphylococcus hominis	Kohl et. al. publication results
Staphylococcus saprophyticus	Kohl et. al. publication results
Staphylococcus ureilyticus	Kohl et. al. publication results
Brevibacterium frigoritolerans	Kohl et. al. publication results
Acinetobacter baumannii	mOTUs results from Acomys reads
Actinomyces oris	mOTUs results from Acomys reads
Adlercreutzia equolifaciens	mOTUs results from Acomys reads
Akkermansia muciniphila	mOTUs results from Acomys reads
Alistipes finegoldii	mOTUs results from Acomys reads
Alistipes finegoldii/onderdonkii	mOTUs results from Acomys reads
Alistipes indistinctus	mOTUs results from Acomys reads
Alistipes obesi	mOTUs results from Acomys reads
Alistipes putredinis	mOTUs results from Acomys reads
Alistipes senegalensis	mOTUs results from Acomys reads
Alistipes shahii	mOTUs results from Acomys reads
Alistipes timonensis	mOTUs results from Acomys reads
Bacteroidales bacterium	mOTUs results from Acomys reads
Bacteroides acidifaciens	mOTUs results from Acomys reads
Bacteroides bouchesdurhonensis/faecichinchillae	mOTUs results from Acomys reads
Bacteroides caecimuris	mOTUs results from Acomys reads
Bacteroides congonensis	mOTUs results from Acomys reads
Bacteroides dorei/vulgatus	mOTUs results from Acomys reads
Bacteroides intestinalis	mOTUs results from Acomys reads
Bacteroides oleiciplenus/stercorirosoris	mOTUs results from Acomys reads
Bacteroides sartorii	mOTUs results from Acomys reads
Bacteroides uniformis	mOTUs results from Acomys reads
Bifidobacterium breve	mOTUs results from Acomys reads

Bifidobacterium pseudolongum	mOTUs results from Acomys reads
Blautia schinkii	mOTUs results from Acomys reads
Candidatus Gastranaerophilus	mOTUs results from Acomys reads
Candidatus Homeothermus	mOTUs results from Acomys reads
Citrobacter rodentium	mOTUs results from Acomys reads
Citrobacter sedlakii	mOTUs results from Acomys reads
Clostridium botulinum/paraputrificum	mOTUs results from Acomys reads
Clostridium perfringens	mOTUs results from Acomys reads
Clostridium phoceensis	mOTUs results from Acomys reads
Comamonas testosteroni	mOTUs results from Acomys reads
Enterococcus faecalis	mOTUs results from Acomys reads
Enterococcus faecium	mOTUs results from Acomys reads
Faecalibacterium prausnitzii	mOTUs results from Acomys reads
Flavonifractor plautii	mOTUs results from Acomys reads
Helicobacter japonicus	mOTUs results from Acomys reads
Hungateiclostridiaceae bacterium	mOTUs results from Acomys reads
Ileibacterium valens	mOTUs results from Acomys reads
Intestinimonas massiliensis	mOTUs results from Acomys reads
Kocuria polaris	mOTUs results from Acomys reads
Lactobacillus acidophilus	mOTUs results from Acomys reads
Lactobacillus amylovorus/kitasatonis	mOTUs results from Acomys reads
Lactobacillus gallinarum	mOTUs results from Acomys reads
Lactobacillus gasseri	mOTUs results from Acomys reads
Lactobacillus intestinalis	mOTUs results from Acomys reads
Lactobacillus johnsonii	mOTUs results from Acomys reads
Lactobacillus kalixensis	mOTUs results from Acomys reads
Lactobacillus pontis	mOTUs results from Acomys reads
Lactobacillus taiwanensis	mOTUs results from Acomys reads
Lactobacillus ultunensis	mOTUs results from Acomys reads
Lactobacillus vaginalis	mOTUs results from Acomys reads
Methanobrevibacter smithii	mOTUs results from Acomys reads
Mobilicoccus massiliensis	mOTUs results from Acomys reads
Muribaculum intestinale	mOTUs results from Acomys reads
Paeniclostridium sordellii	mOTUs results from Acomys reads

Pantoea latae/septica	mOTUs results from Acomys reads
Parabacteroides distasonis	mOTUs results from Acomys reads
Parabacteroides goldsteinii	mOTUs results from Acomys reads
Parabacteroides johnsonii	mOTUs results from Acomys reads
Parabacteroides merdae	mOTUs results from Acomys reads
Pseudoflavonifractor capillosus	mOTUs results from Acomys reads
Pseudomonas xanthomarina/stutzeri	mOTUs results from Acomys reads
Romboutsia timonensis	mOTUs results from Acomys reads
Ruminococcaceae bacterium	mOTUs results from Acomys reads
Ruminococcus gauvreauii	mOTUs results from Acomys reads
Sanguibacter keddieii	mOTUs results from Acomys reads
Sanguibacter marinus	mOTUs results from Acomys reads
Sphingobium yanoikuyae	mOTUs results from Acomys reads
Staphylococcus aureus	mOTUs results from Acomys reads
Staphylococcus microti	mOTUs results from Acomys reads
Stenotrophomonas maltophilia	mOTUs results from Acomys reads
Streptococcus acidominimus	mOTUs results from Acomys reads
Streptococcus cuniculi	mOTUs results from Acomys reads
Turicibacter sanguinis	mOTUs results from Acomys reads
uncultured Flavonifractor	mOTUs results from Acomys reads
Streptococcus acidominimus	mOTUs results from Acomys reads
Streptococcus cuniculi	mOTUs results from Acomys reads
Turicibacter sanguinis	mOTUs results from Acomys reads
iMGMC_64	iMGMC high quality MAG
iMGMC_66	iMGMC high quality MAG
iMGMC_87	iMGMC high quality MAG
iMGMC_177	iMGMC high quality MAG
iMGMC_193	iMGMC high quality MAG
iMGMC_232	iMGMC high quality MAG
iMGMC_262	iMGMC high quality MAG
iMGMC_297	iMGMC high quality MAG
iMGMC_386	iMGMC high quality MAG
iMGMC_413	iMGMC high quality MAG
iMGMC_435	iMGMC high quality MAG

iMGMC_444	iMGMC high quality MAG
iMGMC_505	iMGMC high quality MAG
iMGMC_520	iMGMC high quality MAG
iMGMC_554	iMGMC high quality MAG
iMGMC_580	iMGMC high quality MAG
iMGMC_609	iMGMC high quality MAG
iMGMC_613	iMGMC high quality MAG
iMGMC_694	iMGMC high quality MAG
iMGMC_715	iMGMC high quality MAG
iMGMC_732	iMGMC high quality MAG
iMGMC_804	iMGMC high quality MAG
iMGMC_912	iMGMC high quality MAG
iMGMC_1015	iMGMC high quality MAG
iMGMC_1047	iMGMC high quality MAG
Lactobacillus kefiri strain DH5 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus kefiranciens subsp. kefiranciens strain LKK75 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus kefiranciens strain 1207 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus helveticus strain LZ-R-5 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus helveticus strain TK-J7A chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus helveticus strain D75 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus reuteri complete genome strain ATCC 53608	GTDB-Tk classifications of Acomys bins
Limosilactobacillus reuteri strain ATG-F4 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Limosilactobacillus reuteri strain LL7 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Ligilactobacillus murinus strain CR147 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Ligilactobacillus murinus strain CR1	GTDB-Tk classifications of Acomys bins
Ligilactobacillus murinus strain V10 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Ligilactobacillus animalis strain L	GTDB-Tk classifications of Acomys bins
Ligilactobacillus animalis strain P38 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Pediococcus pentosaceus strain FDAARGOS 1009 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Limosilactobacillus reuteri strain IRT chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus rhamnosus strain JCM1553 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus zymae strain ACA-DC 34	GTDB-Tk classifications of Acomys bins
Lactobacillus paracasei subsp. tolerans strain MGB0747 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Pediococcus damnosus strain TMW 2.1533 chromosome complete genome	GTDB-Tk classifications of Acomys bins

Leuconostoc citreum strain CBA3624 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lacticaseibacillus paracasei strain EG9 chromosome complete genome	GTDB-Tk classifications of Acomys bins
Lactobacillus paracollinoides strain TMW 1.1995 chromosome complete genome	GTDB-Tk classifications of Acomys bins

Table 5.2: List of the identities of downloaded genomes, assemblies or MAGs used in the creation of phylogenetic trees during the project. Determinant refers to the rationale for including the file as outlined in subsection 2.1.7.

Reference mOTU	***A. cahirinus* summed relative abundance (%)**	***A. russatus* summed relative abundance (%)**	**Identity**
ref_mOTU_v3_01032	12.1	4.90	*Lactobacillus kefiranofaciens*
ref_mOTU_v3_03349	0.93	2.42	*Lactobacillus helveticus*
ref_mOTU_v3_03348	0.22	1.59	*Lactobacillus gallinarum*
ext_mOTU_v3_15367	0.12	1.38	*Erysipelotrichaceae* species incertae sedis
ext_mOTU_v3_17736	0.61	1.16	*Muribaculaceae* species incertae sedis
ext_mOTU_v3_18740	0.63	1.16	*Prevotella* species incertae sedis
ref_mOTU_v3_03591	0.53	1.06	*Akkermansia muciniphila*
ref_mOTU_v3_04416	1.26	0.82	*Lactobacillus vaginalis*

Table 5.3: Table giving the reference mOTUs which had a summed relative abundance of at least 1% in at least one of the two host species, the summed totals themselves and the taxonomic identity the reference corresponds to.

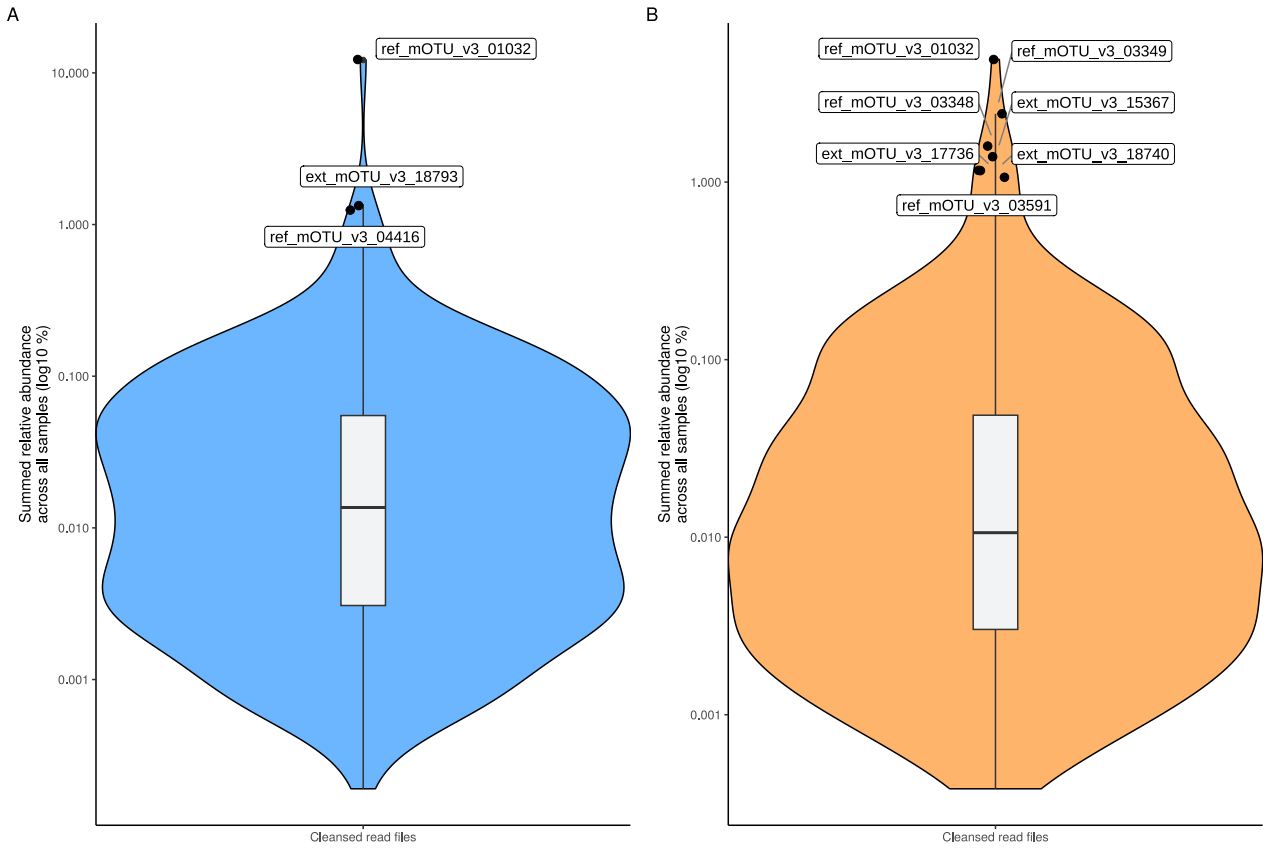

Figure 5.1: Box and violin plots of summed relative abundances for all detected reference mOTUs from cleaned read files from samples from **A.** *Acomys cahirinus* and **B.** *Acomys russatus*. Labelled points are those where the summed relative abundance for the host species was at least 1%

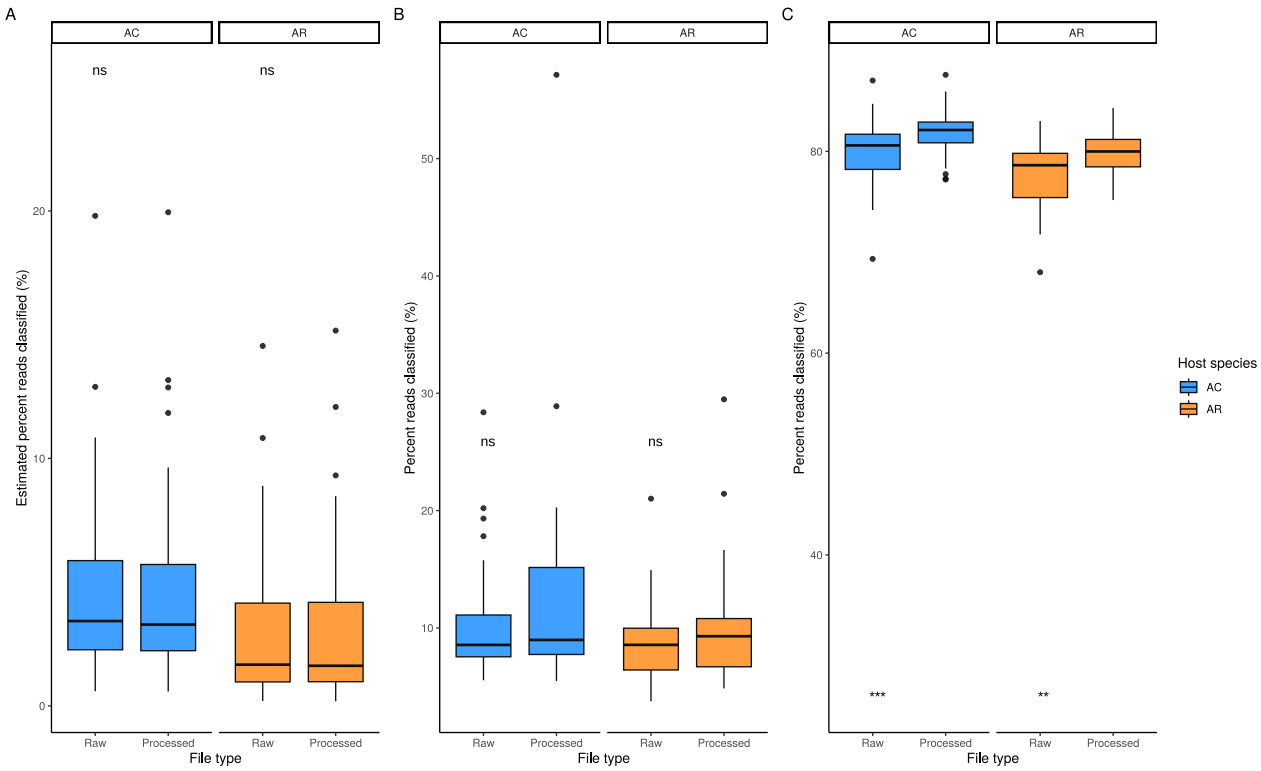

Figure 5.2: Boxplots showing the percentage of reads classified by **A.** Metaphlan 3, **B.** Kraken 2 and **C.** Kaiju for the raw and processed read files. Also shown are results of Wilcoxon signed-rank test for difference as a result of processing with significance marked by asterisks. *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$.

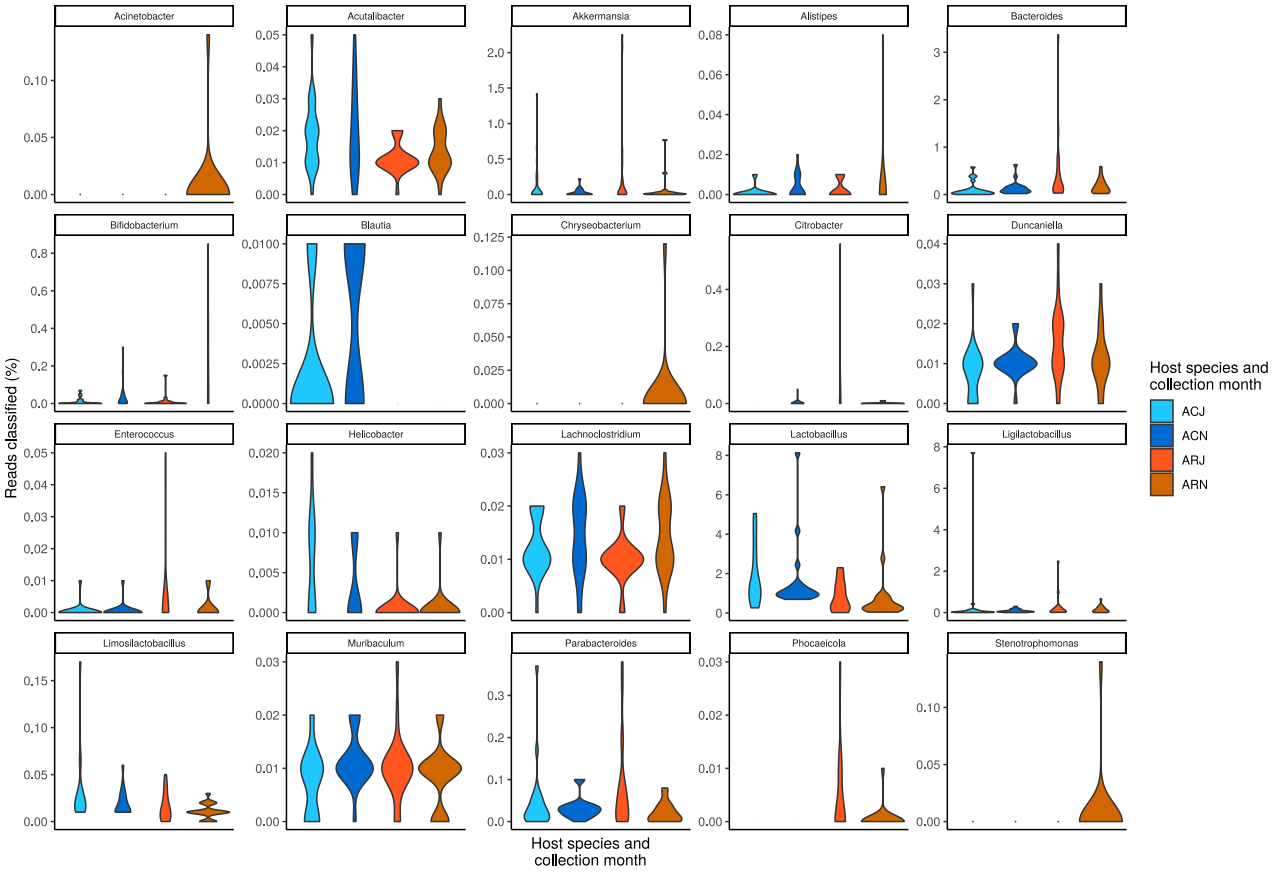

Figure 5.3: Violin plots showing the range of percentages of reads classified by Kraken 2 at a minimum confidence score of 50% for the 20 genera with the highest summed percentage of reads classified across all samples, coloured by host species and month of collection and faceted by genus.

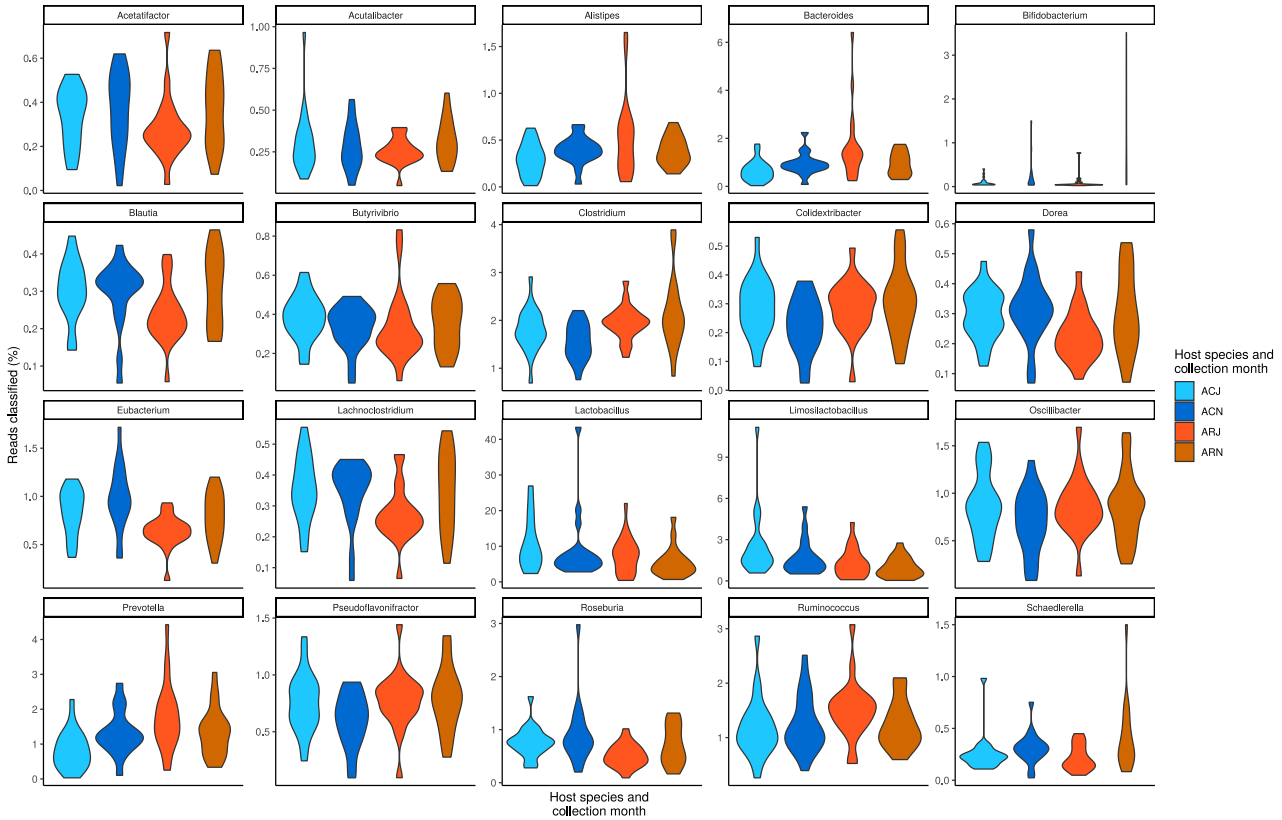

Figure 5.4: Violin plots showing the range of percentages of reads classified by Kaiju with an error allowance of 0 for the 20 genera with the highest summed percentage of reads classified across all samples, coloured by host species and month of collection and faceted by genus.

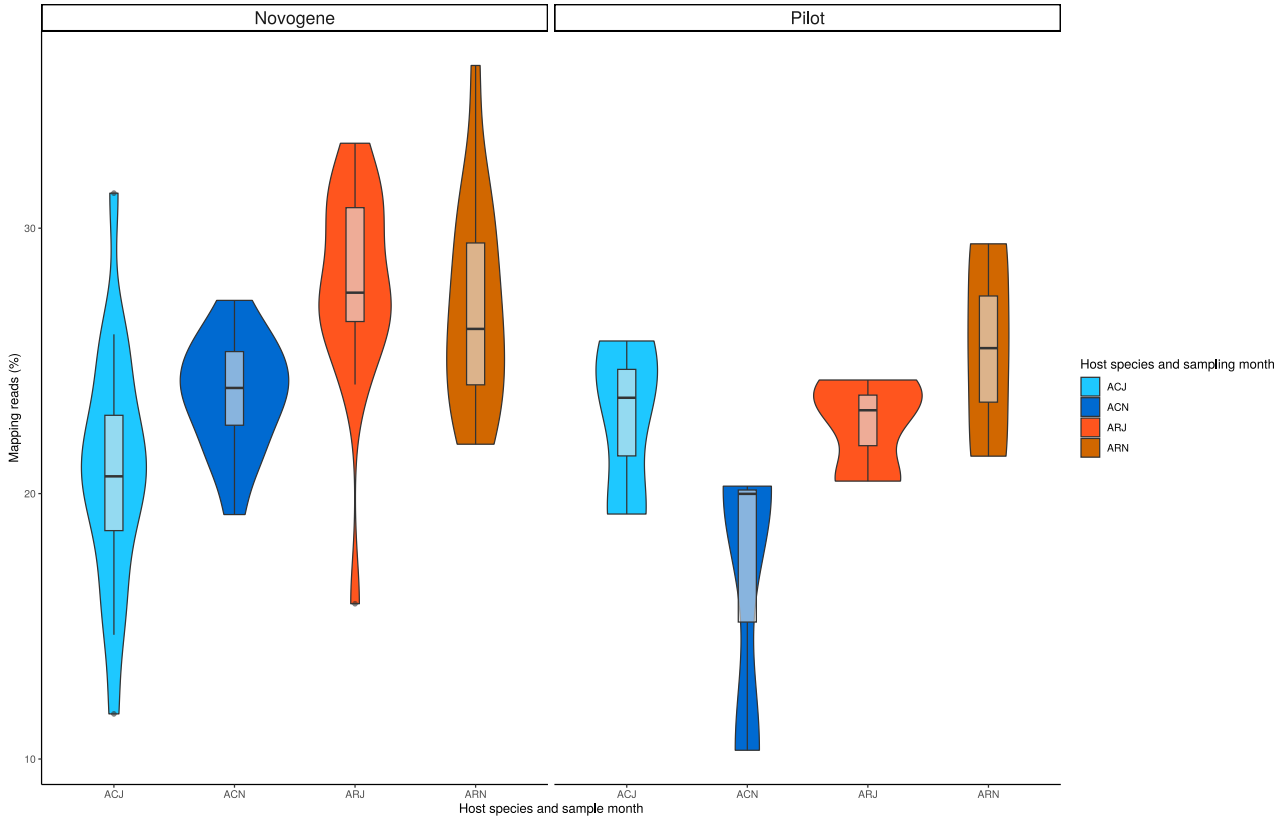

Figure 5.5: Box and violin plots showing the range of percentage of reads mapping from *Acomys* faecal sample sequencing to reference file of 348 concatenated metagenomic bins, coloured by host species and sampling month and faceted by read depth.

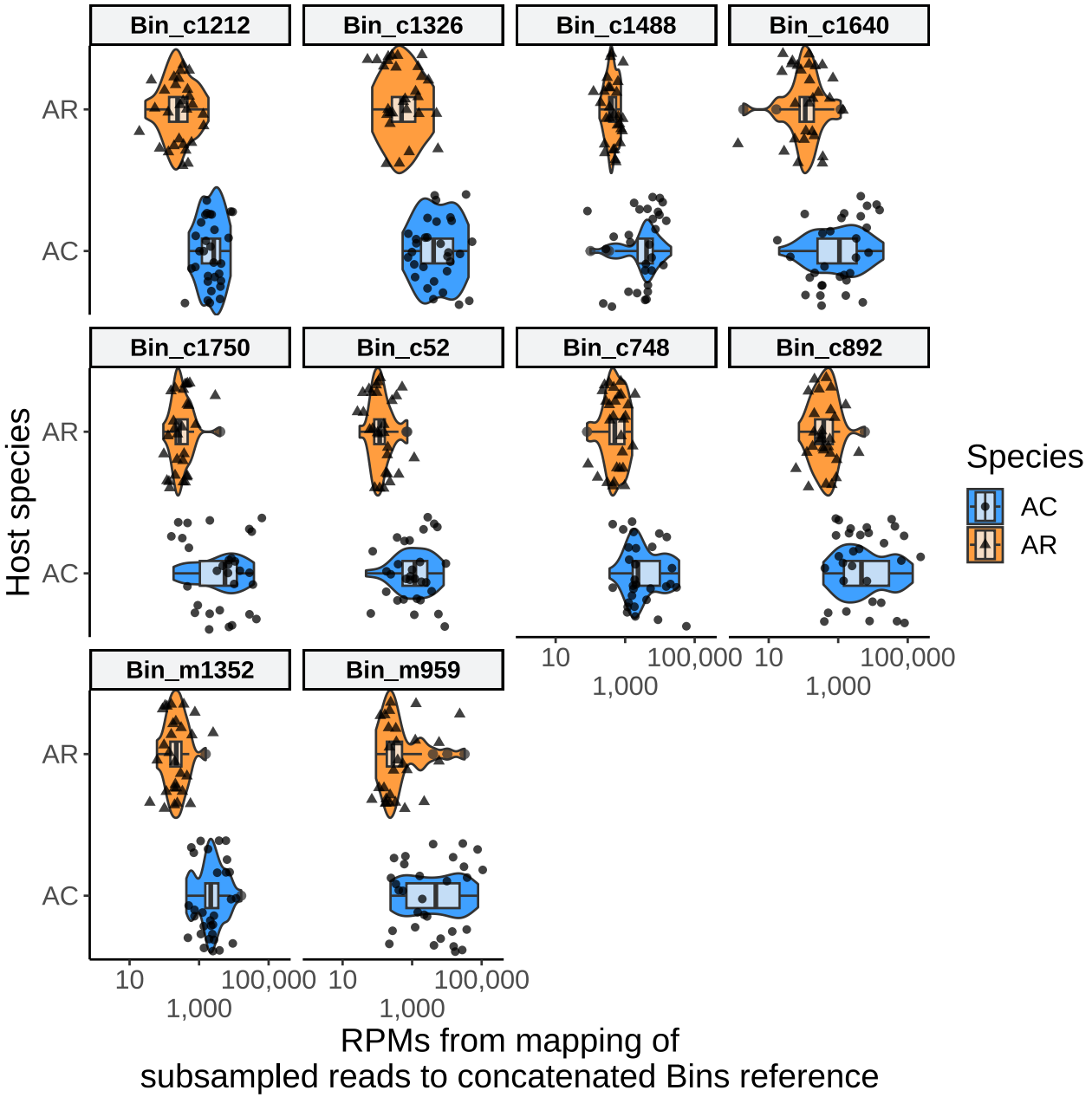

Figure 5.6: Box and violin plots showing the range of RPMs for the 10 most enriched bins in *Acomys cahirinus* on a log10 scale

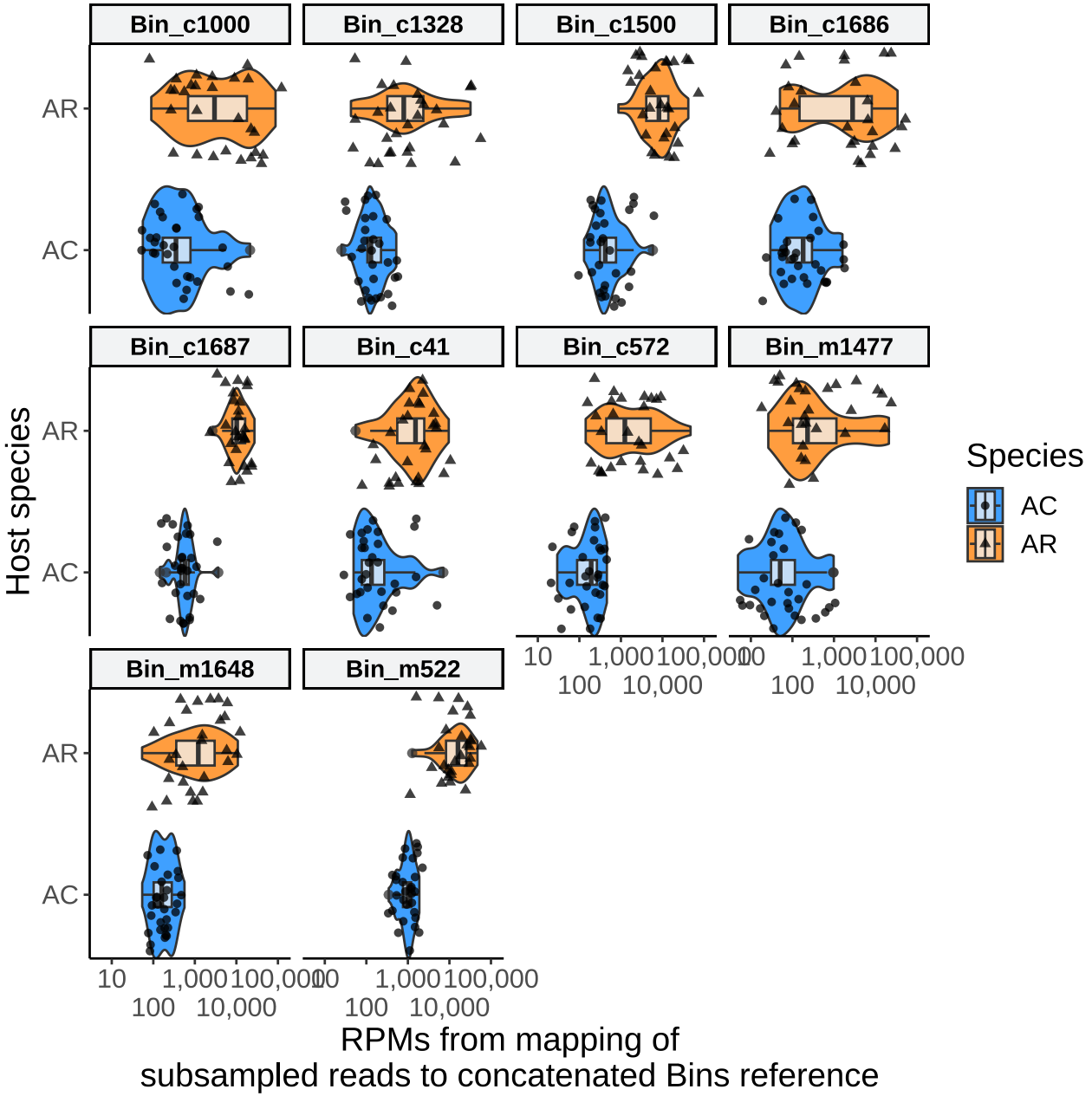

Figure 5.7: Box and violin plots showing the range of RPMs for the 10 most enriched bins in *Acomys russatus* on a log10 scale

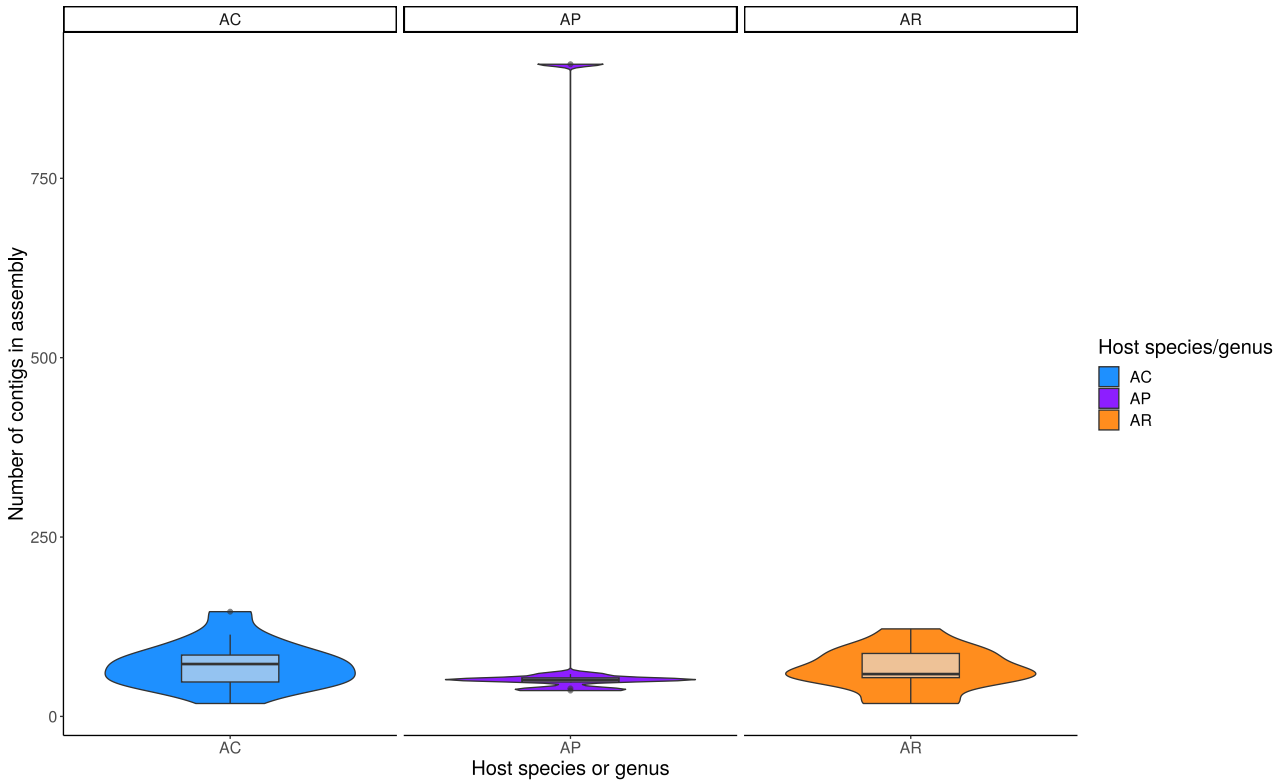

Figure 5.8: Box and violin plot of the number of contigs per assembly separated by facet and fill colour into host species (for *Acomys* species) or genus (for *Apodemus*).

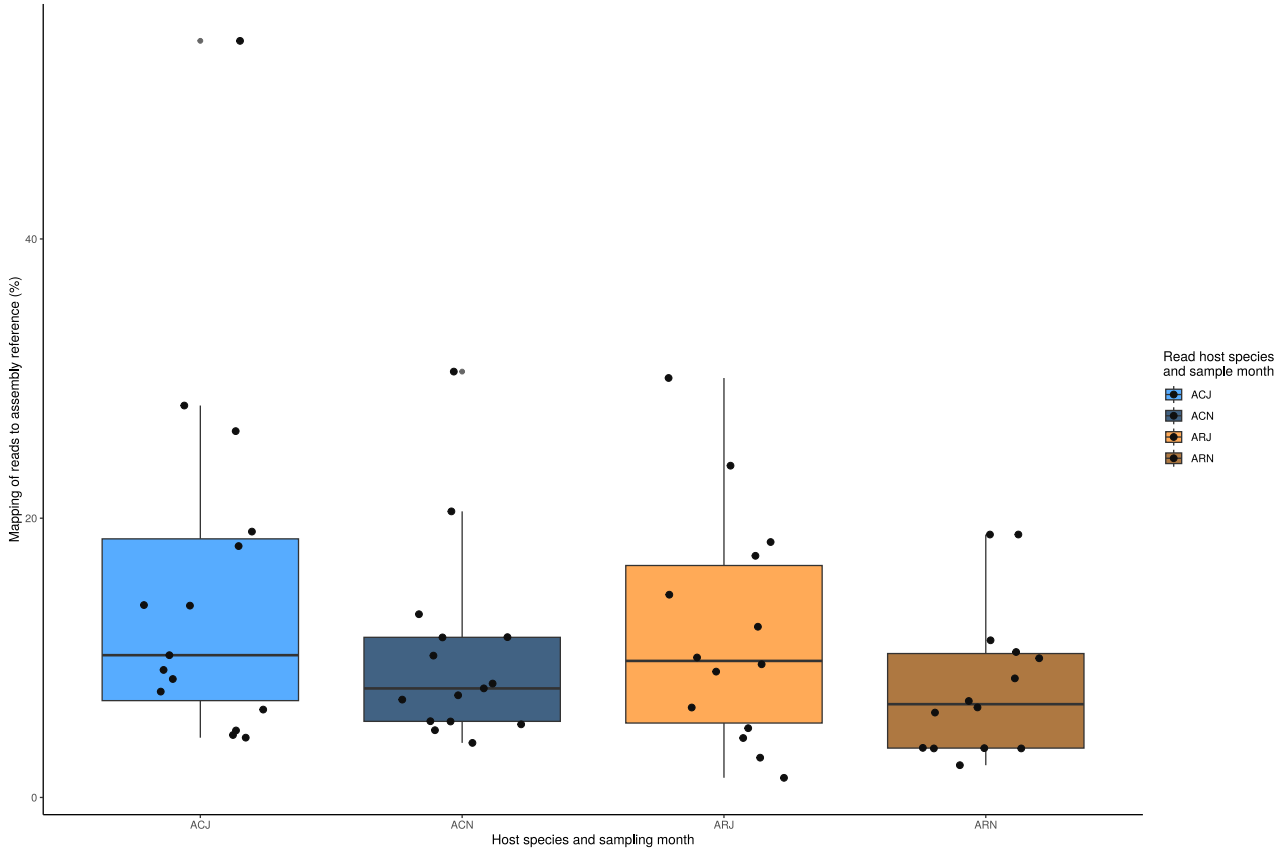

Figure 5.9: Box and jitter plots of the percent of reads mapping with Minimap2 to a combined reference file made from masked assembly files. **A.** The percentages for read files not subsampled to a uniform depth. **B.** Percentages for read files subsampled to a uniform depth of 7,600,000 reads per file.

COG ID	Species with host effect	Q value
COG3477	AC	0.00000
COG1250	AC	0.00053
COG2153	AC	0.00236
COG2096	AC	0.00330
COG2326	AC	0.02955
COG0108	AC	0.03669
COG3760	AC	0.00253
COG2211	AC	0.00585
COG0395	AC	0.00000
COG0246	AC	0.00337
COG0596	AC	0.00192
COG2814	AC	0.02817
COG2390	AC	0.01083
COG2723	AC	0.00000
COG1455	AC	0.00000
COG2826	AC	0.00969
COG1316	AC	0.00369
COG2071	AC	0.03325
COG1263	AC	0.00000

COG0735	AC	0.00323
COG1117	AC	0.01761
COG0366	AC	0.00854
COG1959	AC	0.01562
COG1674	AC	0.02451
COG2217	AC	0.04999
COG0577	AC	0.00646
COG0235	AC	0.04739
COG1131	AC	0.00007
COG3760	AR	0.00004
COG3385	AR	0.00014
COG2211	AR	0.00017
COG3839	AR	0.00044
COG0395	AR	0.00052
COG0246	AR	0.02232
COG0095	AR	0.01118
COG2337	AR	0.03005
COG2814	AR	0.03252
COG2723	AR	0.00076
COG2148	AR	0.00856
COG1455	AR	0.04063
COG1737	AR	0.00170
COG0764	AR	0.00260
COG2071	AR	0.02251
COG4166	AR	0.00000
COG0601	AR	0.00024
COG1263	AR	0.00008
COG0735	AR	0.00655
COG1959	AR	0.04634
COG0789	AR	0.00006
COG0577	AR	0.02981
COG1680	AP	0.00259
COG3579	AP	0.00499
COG2271	AP	0.01416
COG1120	AP	0.02088
COG0095	AP	0.02529
COG0782	AP	0.04095
COG4690	AP	0.04104

Table 5.4: Table of COGs which had a statistically significant, ≤ 0.05 Q value for a host effect on distribution in assemblies by the origin species of the faecal sample the isolate was obtained from. Shows the COGs, species of origin and Q-value.

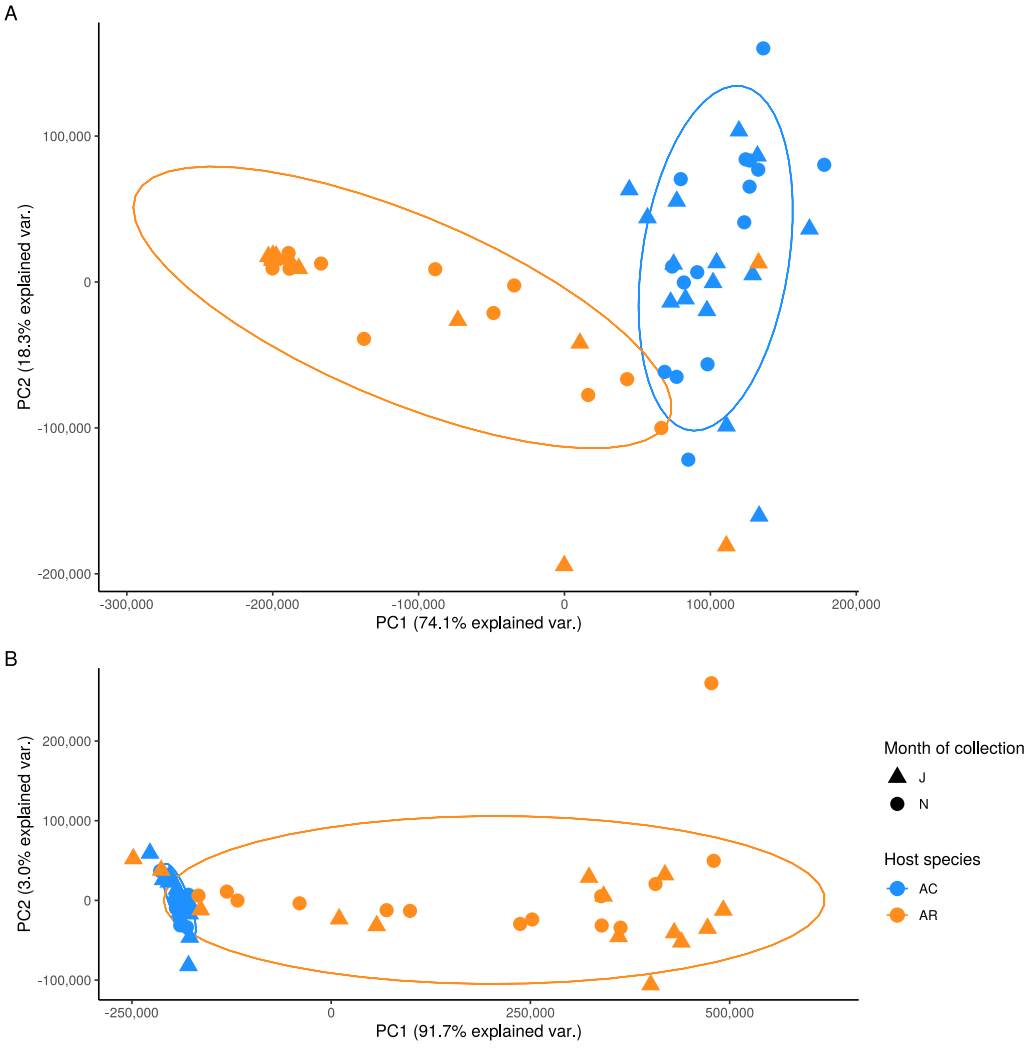

Figure 5.10: Principal Component Analysis of reads per-million (RPM) for all subsampled, paired sampling reads from both Acomys species to assemblies of isolates from faecal samples from **A.** *Acomys cahirinus* only and **B.** *Acomys russatus* only.

Glossary of Terms

16S Sequencing Sequencing of the 16S rRNA gene for the purposes of taxonomic classification

AC enriched vs In context, bins which were enriched in *Acomys cahirinus* versus *Acomys russatus*

AC within In context, bins which were enriched within *Acomys cahirinus* in either June or November versus the other month

Acomys Genus of rodents

Acomys cahirinus A rodent species in the genus *Acomys*

Acomys russatus A rodent species in the genus *Acomys*

Apodemus Genus of rodents

Apodemus agrarius A rodent species in the genus *Apodemus*

Apodemus flavicollis A rodent species in the genus *Apodemus*

Apodemus sylvaticus A rodent species in the genus *Apodemus*

AR enriched vs In context, bins which were enriched in *Acomys russatus* versus *Acomys cahirinus*

AR within In context, bins which were enriched within *Acomys russatus* in either June or November versus the other month

Assembly Collected sequence reads arranged into a longer, contiguous piece of sequence composed of contigs or scaffolds

Amplicon Sequence Variant Sequence variants which are the product of denoising such that amplification and sequencing errors should be removed

Barrnap Software tool for detection of ribosomal RNA genes in sequence read files

BBDuk Software for removal of host or human contamination from nucleotide sequencing files

CheckM Software for assessing quality of genomic assemblies

Cleansed See either 'Host cleansed' or 'Contaminant cleansed'

Contaminant cleansed Author's term for nucleotide sequencing files which have been subjected to some method for removing foreign reads

Ein Gedi Location in the Judean Desert, Israel, nearest notable location to the *Acomys.* enclosure

FastANI Software for measuring ANI similarity between nucleotide sequence files

FastP Software for quality control of nucleotide sequencing files

Genomics Study of all the genes of an organism

Genus false positive Author's term for NCBI taxonomic IDs assigned by a classifier to a read from a mock microbial community. The taxonomic ID is from the genus of a known member species in the mock community but is not any of the subspecies/strain taxonomic IDs or the higher level taxonomic ID for the type species itself

Halophilic Organisms which require conditions of elevated salinity to grow

Halotolerant Organisms capable of growth in conditions of elevated salinity

Host cleansed Author's term for nucleotide sequencing files which have been subjected to some procedure meant to remove host organism reads

Hydric Environment with very high levels of moisture

Illumina MiSeq A DNA sequencing technology offered by Illumina, Inc.

JolyTree Reference-free software for creating phylogenetic trees

Kaiju Taxonomic classification software

Kmer In context, a section of nucleotides of variable length

Kraken 2 Taxonomic classification software

Mesic Environment with a balance supply of moisture

Metabolomics Study of small molecules within an environment from a single cell to entire organism

Metagenome The collection of genomes and genes from the members of a microbiota. [19]

Metaphlan 2 Taxonomic classification software

Microbiome the entire habitat, including the microorganisms (bacteria, archaea, lower and higher eurkaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions. [19]

Microbiota The assemblage of microorganisms present in a defined environment

Minimap2 Nucleotide sequence alignment software

mOTUs Taxonomic classification software

Novogene In context, the remaining Acomys samples sequenced to a higher depth by Novogene Co. using Illumina NovaSeq technology in 2020 or pertaining to the results of analysis of these samples

Operational Taxonomic Unit A group of sequences which are considered to be closely related based on some mathematical threshold and are typically considered to be funtionally similar

Phylotype In context, a DNA sequence or group of sequences which are sufficiently similar in the composition of marker sequence that they are grouped together

Pilot In context, 12 Acomys samples sequenced to a lower depth using Illumina MiSeq technology in 2017 or pertaining to the results of analysis of these samples

Precocial Young of a precocial species are very developed at birth

Proteomics Study of all proteins produced by an organism

Salmon In context, software for quantifying expression of transcripts using RNA sequencing data. Used in this project with DNA sequencing data to measure reads per-million

Samtools Set of software for interacting with and manipulating sequence read files

Seqtk Software for the the manipulation of nucleotide sequencing files

Shotgun sequencing In context, sequencing of DNA by splitting it randomly into fragments and sequencing these separately

TaxID Taxonomic identity number for a taxon from the NCBI taxonomy database

Transcriptomics Study of all the RNA transcripts produced by an organism

Xeric Arid environment with low annual rainfall and little moisture

Bibliography

[1] H. Gest. “The Discovery of Microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society”. In: *Notes and Records of the Royal Society of London* 58.2 (May 22, 2004), pp. 187–201. DOI: 10.1098/rsnr.2004.0055. URL: https://royalsocietypublishing.org/doi/10.1098/rsnr.2004.0055 (visited on 11/20/2022).

[2] E. Klein. “Pathogenic Microbes in Milk”. In: *Epidemiology & Infection* 1.1 (Jan. 1901), pp. 78–95. ISSN: 0022-1724. DOI: 10.1017/S0022172400000061. URL: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/pathogenic-microbes-in-milk/BA9F830763161E3289F783B232666F30 (visited on 11/20/2022).

[3] Paul H. De Kruif. “Dissociation of Microbic Species. III. Differentiation of Microbes D and G by Acid Agglutination”. In: *Proceedings of the Society for Experimental Biology and Medicine* 19.1 (Oct. 1, 1921), pp. 37–38. ISSN: 0037-9727. DOI: 10.3181/00379727-19-18. URL: https://journals.sagepub.com/doi/abs/10.3181/00379727-19-18 (visited on 11/20/2022).

[4] Paula Watnick and Roberto Kolter. “Biofilm, City of Microbes”. In: *Journal of Bacteriology* 182.10 (May 15, 2000), pp. 2675–2679. DOI: 10.1128/JB.182.10.2675-2679.2000. URL: https://journals.asm.org/doi/10.1128/JB.182.10.2675-2679.2000 (visited on 11/20/2022).

[5] Natalia Garcia-Gonzalez et al. “Health-Promoting Role of Lactiplantibacillus Plantarum Isolated from Fermented Foods”. In: *Microorganisms* 9.2 (2 Feb. 2021), p. 349. ISSN: 2076-2607. DOI: 10.3390/microorganisms9020349. URL: https://www.mdpi.com/2076-2607/9/2/349 (visited on 11/20/2022).

[6] J. D. Watson and F. H. C. Crick. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”. In: *Nature* 171.4356 (4356 Apr. 1953), pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0. URL: https://www.nature.com/articles/171737a0 (visited on 11/20/2022).

[7] C Picard et al. “Detection and Enumeration of Bacteria in Soil by Direct DNA Extraction and Polymerase Chain Reaction”. In: *Applied and Environmental Microbiology* 58.9 (Sept. 1992), pp. 2717–2722. DOI: 10.1128/aem.58.9.2717-2722.1992. URL: https://journals.asm.org/doi/10.1128/aem.58.9.2717-2722.1992 (visited on 11/20/2022).

[8] Y. K Lee et al. “A Simple Method for DNA Extraction from Marine Bacteria That Produce Extracellular Materials”. In: *Journal of Microbiological Methods* 52.2 (Feb. 1, 2003), pp. 245–250. ISSN: 0167-7012. DOI: 10.1016/S0167-7012(02)00180-X. URL: https://www.sciencedirect.com/science/article/pii/S016770120200180X (visited on 11/20/2022).

[9] Hai-Rong Cheng and Ning Jiang. “Extremely Rapid Extraction of DNA from Bacteria and Yeasts”. In: *Biotechnology Letters* 28.1 (Jan. 1, 2006), pp. 55–59. ISSN: 1573-6776. DOI: 10.1007/s10529-005-4688-z. URL: https://doi.org/10.1007/s10529-005-4688-z (visited on 11/20/2022).

[10] A M Maxam and W Gilbert. “A New Method for Sequencing DNA.” In: *Proceedings of the National Academy of Sciences* 74.2 (Feb. 1977), pp. 560–564. DOI: 10.1073/pnas.74.2.560. URL: https://www.pnas.org/doi/abs/10.1073/pnas.74.2.560 (visited on 11/20/2022).

[11] F. Sanger, S. Nicklen, and A. R. Coulson. “DNA Sequencing with Chain-Terminating Inhibitors”. In: *Proceedings of the National Academy of Sciences* 74.12 (Dec. 1977), pp. 5463–5467. DOI: 10.1073/pnas.74.12.5463. URL: https://www.pnas.org/doi/abs/10.1073/pnas.74.12.5463 (visited on 11/20/2022).

[12] Lloyd M. Smith et al. “Fluorescence Detection in Automated DNA Sequence Analysis”. In: *Nature* 321.6071 (6071 June 1986), pp. 674–679. ISSN: 1476-4687. DOI: 10.1038/321674a0. URL: https://www.nature.com/articles/321674a0 (visited on 11/20/2022).

[13] Andrew H. Laszlo et al. “Decoding Long Nanopore Sequencing Reads of Natural DNA”. In: *Nature Biotechnology* 32.8 (8 Aug. 2014), pp. 829–833. ISSN: 1546-1696. DOI: 10.1038/nbt.2950. URL: https://www.nature.com/articles/nbt.2950 (visited on 11/20/2022).

[14] Frederick R. Blattner et al. “The Complete Genome Sequence of Escherichia Coli K-12”. In: *Science* 277.5331 (Sept. 5, 1997), pp. 1453–1462. DOI: 10.1126/science.277.5331.1453. URL: https://www.science.org/doi/full/10.1126/science.277.5331.1453 (visited on 11/20/2022).

[15] J. Craig Venter et al. “The Sequence of the Human Genome”. In: *Science* 291.5507 (Feb. 16, 2001), pp. 1304–1351. DOI: 10.1126/science.1058040. URL: https://www.science.org/doi/full/10.1126/science.1058040 (visited on 11/20/2022).

[16] Peter J. Turnbaugh and Jeffrey I. Gordon. “The Core Gut Microbiome, Energy Balance and Obesity”. In: *The Journal of Physiology* 587.17 (2009), pp. 4153–4158. ISSN: 1469-7793. DOI: 10.1113/jphysiol.2009.174136. URL: https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.2009.174136 (visited on 11/20/2022).

[17] Dana Willner, Rebecca Vega Thurber, and Forest Rohwer. “Metagenomic Signatures of 86 Microbial and Viral Metagenomes”. In: *Environmental Microbiology* 11.7 (2009), pp. 1752–1766. ISSN: 1462-2920. DOI: 10.1111/j.1462-2920.2009.01901.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2009.01901.x (visited on 11/20/2022).

[18] Lucas W. Mendes et al. “Soil-Borne Microbiome: Linking Diversity to Function”. In: *Microbial Ecology* 70.1 (July 1, 2015), pp. 255–265. ISSN: 1432-184X. DOI: 10.1007/s00248-014-0559-2. URL: https://doi.org/10.1007/s00248-014-0559-2 (visited on 11/20/2022).

[19] Julian R. Marchesi and Jacques Ravel. “The vocabulary of microbiome research: a proposal”. In: *Microbiome* 3.1 (2015), p. 31. DOI: 10.1186/s40168-015-0094-5.

[20] Garret Suen et al. “An Insect Herbivore Microbiome with High Plant Biomass-Degrading Capacity”. In: *PLOS Genetics* 6.9 (Sept. 23, 2010), e1001129. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1001129. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001129 (visited on 11/20/2022).

[21] Charlie G. Buffie and Eric G. Pamer. “Microbiota-Mediated Colonization Resistance against Intestinal Pathogens”. In: *Nature Reviews Immunology* 13.11 (11 Nov. 2013), pp. 790–801. ISSN: 1474-1741. DOI: 10.1038/nri3535. URL: https://www.nature.com/articles/nri3535 (visited on 11/20/2022).

[22] Seth Rakoff-Nahoum et al. “Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis”. In: *Cell* 118.2 (July 23, 2004), pp. 229–241. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2004.07.002. pmid: 15260992. URL: https://www.cell.com/cell/abstract/S0092-8674(04)00661-0 (visited on 11/20/2022).

[23] Sajad Ali et al. “Harnessing Plant Microbiome for Mitigating Arsenic Toxicity in Sustainable Agriculture”. In: *Environmental Pollution* 300 (May 1, 2022), p. 118940. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2022.118940. URL: https://www.sciencedirect.com/science/article/pii/S0269749122001543 (visited on 11/21/2022).

[24] Marina Lleal et al. “A Single Faecal Microbiota Transplantation Modulates the Microbiome and Improves Clinical Manifestations in a Rat Model of Colitis”. In: *EBioMedicine* 48 (Oct. 1, 2019), pp. 630–641. ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2019.10.002. URL: https://www.sciencedirect.com/science/article/pii/S2352396419306668 (visited on 11/21/2022).

[25] Fanli Kong et al. “Identification of Gut Microbiome Signatures Associated with Longevity Provides a Promising Modulation Target for Healthy Aging”. In: *Gut Microbes* 10.2 (Mar. 4, 2019), pp. 210–215. ISSN: 1949-0976. DOI: 10.1080/19490976.2018.1494102. pmid: 30142010. URL: https://doi.org/10.1080/19490976.2018.1494102 (visited on 11/21/2022).

[26] Yuanyuan Cheng et al. “The Tasmanian Devil Microbiome—Implications for Conservation and Management”. In: *Microbiome* 3.1 (Dec. 21, 2015), p. 76. ISSN: 2049-2618. DOI: 10.1186/s40168-015-0143-0. URL: https://doi.org/10.1186/s40168-015-0143-0 (visited on 11/21/2022).

[27] Syed Mohsin Bukhari et al. “Metagenomics analysis of the fecal microbiota in Ring-necked pheasants (Phasianus colchicus) and Green pheasants (Phasianus versicolor) using next generation sequencing”. In: *Saudi Journal of Biological Sciences* 29.3 (2022), pp. 1781–1788.

[28] Luke Stevenson. “Discovery and Biosynthesis of Natural Products from New Zealand Soil Metagenome Libraries”. Victoria University of Wellington, 2020. URL: http://researcharchive.vuw.ac.nz/handle/10063/9004 (visited on 06/01/2021).

[29] Luen-Luen Li et al. “Bioprospecting Metagenomes: Glycosyl Hydrolases for Converting Biomass”. In: *Biotechnology for Biofuels* 2.1 (May 18, 2009), p. 10. ISSN: 1754-6834. DOI: 10.1186/1754-6834-2-10. URL: https://doi.org/10.1186/1754-6834-2-10 (visited on 11/21/2022).

[30] David Dubnau et al. “Gene conservation in Bacillus species. I. Conserved genetic and nucleic acid base sequence homologies.” In: *Proceedings of the National Academy of Sciences* 54.2 (1965), pp. 491–498.

[31] C. R. Woese. “BACTERIAL EVOLUTION”. In: *Microbiological Reviews* 51.2 (1987), pp. 221–271. ISSN: 0146-0749.

[32] Thomas P. Niedringhaus et al. “Landscape of Next-Generation Sequencing Technologies”. In: *Analytical Chemistry* 83.12 (June 15, 2011), pp. 4327–4341. ISSN: 0003-2700. DOI: 10.1021/ac2010857. URL: https://doi.org/10.1021/ac2010857 (visited on 04/12/2022).

[33] Sang Tae Park and Jayoung Kim. “Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing”. In: *International Neurourology Journal* 20 (Suppl 2 Nov. 22, 2016), S76–83. ISSN: 2093-4777, 2093-6931. DOI: 10.5213/inj.1632742.371. URL: http://www.einj.org/journal/view.php?doi=10.5213/inj.1632742.371 (visited on 04/12/2022).

[34] J B Patel. “16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory.” eng. In: *Mol Diagn* 6.4 (2001), pp. 313–321. DOI: 10.1054/modi.2001.29158.

[35] L Barth Reller, Melvin P Weinstein, and Cathy A Petti. “Detection and identification of microorganisms by gene amplification and sequencing”. In: *Clinical infectious diseases* 44.8 (2007), pp. 1108–1114.

[36] Motoo Kimura. “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences”. In: *Journal of molecular evolution* 16 (1980), pp. 111–120.

[37] Benjamin Hillmann et al. “Evaluating the information content of shallow shotgun metagenomics”. In: *Msystems* 3.6 (2018), e00069–18.

[38] Nurnabila Syafiqah Muhamad Rizal et al. “Advantages and limitations of 16S rRNA next-generation sequencing for pathogen identification in the diagnostic microbiology laboratory: perspectives from a middle-income country”. In: *Diagnostics* 10.10 (2020), p. 816.

[39] Telleasha L Greay et al. “Evaluation of 16S next-generation sequencing of hypervariable region 4 in wastewater samples: An unsuitable approach for bacterial enteric pathogen identification”. In: *Science of the total environment* 670 (2019), pp. 1111–1124.

[40] Jean Pierre Rutanga et al. “16S metagenomics for diagnosis of bloodstream infections: opportunities and pitfalls”. In: *Expert review of molecular diagnostics* 18.8 (2018), pp. 749–759.

[41] Lauren V Alteio et al. “A critical perspective on interpreting amplicon sequencing data in soil ecological research”. In: *Soil Biology and Biochemistry* 160 (2021), p. 108357.

[42] Rachel Poretsky et al. “Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics”. In: *PloS one* 9.4 (2014), e93827.

[43] Jill E Clarridge III. “Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases”. In: *Clinical microbiology reviews* 17.4 (2004), pp. 840–862.

[44] Ren-Mao Tian et al. “Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene”. In: *Genome biology and Evolution* 7.8 (2015), pp. 2310–2320.

[45] Silvia G Acinas et al. “Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons”. In: *Journal of bacteriology* 186.9 (2004), pp. 2629–2635.

[46] Hayley B Hassler et al. “Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies”. In: *Microbiome* 10.1 (2022), p. 104.

[47] Susan M Huse et al. “Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing”. In: *PLoS genetics* 4.11 (2008), e1000255.

[48] Yong Wang and Pei-Yuan Qian. “Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies”. In: *PloS one* 4.10 (2009), e7401.

[49] M Guembe et al. “Use of universal 16S rRNA gene PCR as a diagnostic tool for venous access port-related bloodstream infections”. In: *Journal of clinical microbiology* 51.3 (2013), pp. 799–804.

[50] Ramya Srinivasan et al. “Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens”. In: *PloS one* 10.2 (2015).

[51] CA Petti, CR Polage, and P Schreckenberger. “The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods”. In: *Journal of clinical microbiology* 43.12 (2005), pp. 6123–6125.

[52] JL Flanagan et al. “Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with Pseudomonas aeruginosa”. In: *Journal of clinical microbiology* 45.6 (2007), pp. 1954–1962.

[53] GC Baker, Jacques J Smith, and Donald A Cowan. “Review and re-analysis of domain-specific 16S primers”. In: *Journal of microbiological methods* 55.3 (2003), pp. 541–555.

[54] Erwin G Zoetendal, Elaine E Vaughan, and Willem M De Vos. “A microbial world within us”. In: *Molecular microbiology* 59.6 (2006), pp. 1639–1650.

[55] Shinichi Kai et al. “Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer”. In: *FEBS open bio* 9.3 (2019), pp. 548–557.

[56] Anne K Dunn and Eric V Stabb. “Culture-independent characterization of the microbiota of the ant lion Myrmeleon mobilis (Neuroptera: Myrmeleontidae)”. In: *Applied and Environmental Microbiology* 71.12 (2005), pp. 8784–8794.

[57] Fabrice Armougom and Didier Raoult. “Exploring microbial diversity using 16S rRNA high-throughput methods”. In: *J Comput Sci Syst Biol* 2.1 (2009), pp. 74–92.

[58] Jethro S. Johnson et al. “Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis”. In: *Nature Communications* 10.1 (2019), pp. 1–11. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13036-1. URL: https://www.nature.com/articles/s41467-019-13036-1 (visited on 12/04/2019).

[59] Raf Winand et al. “Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies”. In: *International journal of molecular sciences* 21.1 (2019), p. 298.

[60] Patrick D Schloss and Sarah L Westcott. “Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis”. In: *Applied and environmental microbiology* 77.10 (2011), pp. 3219–3226.

[61] Ruth E Ley et al. “Obesity alters gut microbial ecology”. In: *Proceedings of the national academy of sciences* 102.31 (2005).

[62] Samuel Dupont et al. “First insights into the microbiome of a carnivorous sponge”. In: *FEMS Microbiology Ecology* 86.3 (2013), pp. 520–531.

[63] Kevin Panke-Buisse et al. “Selection on soil microbiomes reveals reproducible impacts on plant function”. In: *The ISME journal* 9.4 (2015), pp. 980–989.

[64] Viggó Thór Marteinsson et al. “Microbial communities in the subglacial waters of the Vatnajökull ice cap, Iceland”. In: *The ISME journal* 7.2 (2013), pp. 427–437.

[65] Ryan J Newton et al. “A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan”. In: *Microbial ecology* 65 (2013), pp. 1011–1023.

[66] Nur Syafika Mohd-Yusof et al. “First report on metagenomic analysis of gut microbiome in Island Flying Fox (Pteropus hypomelanus) revealing latitudinal correlation as opposed to host phylogeny in island populations of Malaysia”. In: *Authorea Preprints* (2020).

[67] Marion Leclerc, Jean-Philippe Delgènes, and Jean-Jacques Godon. “Diversity of the Archaeal Community in 44 Anaerobic Digesters as Determined by Single Strand Conformation Polymorphism Analysis and 16S rDNA Sequencing”. In: *Environmental Microbiology* 6.8 (2004), pp. 809–819. ISSN: 1462-2920. DOI: 10.1111/j.1462-2920.2004.00616.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2004.00616.x (visited on 11/25/2022).

[68] Tairacan Augusto Pereira da Fonseca, Rodrigo Pessôa, and Sabri Saeed Sanabani. “Molecular Analysis of Bacterial Microbiota on Brazilian Currency Note Surfaces”. In: *International Journal of Environmental Research and Public Health* 12.10 (10 Oct. 2015), pp. 13276–13288. ISSN: 1660-4601. DOI: 10.3390/ijerph121013276. URL: https://www.mdpi.com/1660-4601/12/10/13276 (visited on 11/25/2022).

[69] Isabel S Cunha et al. “Bacteria and Archaea community structure in the rumen microbiome of goats (Capra hircus) from the semiarid region of Brazil”. In: *Anaerobe* 17.3 (2011), pp. 118–124.

[70] Alexander J Probst, Anna K Auerbach, and Christine Moissl-Eichinger. “Archaea on human skin”. In: *PloS one* 8.6 (2013), e65388.

[71] Alexander K Umbach, Ashley A Stegelmeier, and Josh D Neufeld. “Archaea are rare and uncommon members of the mammalian skin microbiome”. In: *Msystems* 6.4 (2021), e00642–21.

[72] Rolf Henrik Nilsson et al. “The ITS region as a target for characterization of fungal communities using emerging sequencing technologies”. In: *FEMS Microbiology Letters* 296.1 (2009), pp. 97–101.

[73] Hanna Miettinen et al. “Microbiome composition and geochemical characteristics of deep subsurface high-pressure environment, Pyhäsalmi mine Finland”. In: *Frontiers in Microbiology* 6 (2015), p. 1203.

[74] Mallory J Suhr, Nabaraj Banjara, and Heather E Hallen-Adams. “Sequence-based methods for detecting and evaluating the human gut mycobiome”. In: *Letters in applied microbiology* 62.3 (2016), pp. 209–215.

[75] William R. Rittenour et al. “Internal Transcribed Spacer rRNA Gene Sequencing Analysis of Fungal Diversity in Kansas City Indoor Environments”. In: *Environmental Science: Processes & Impacts* 16.1 (Dec. 19, 2013), pp. 33–43. ISSN: 2050-7895. DOI: 10.1039/C3EM00441D. URL: https://pubs.rsc.org/en/content/articlelanding/2014/em/c3em00441d (visited on 11/25/2022).

[76] Seth Commichaux et al. “taxaTarget: Fast, sensitive, and precise classification of microeukaryotes in metagenomic data”. In: *Research Square* (2022).

[77] Joann L Cloud et al. “Evaluation of partial 16S ribosomal DNA sequencing for identification of Nocardia species by using the MicroSeq 500 system with an expanded database”. In: *Journal of Clinical Microbiology* 42.2 (2004), pp. 578–584.

[78] CA Petti et al. “Interpretive criteria for identification of bacteria and fungi by DNA target sequencing; approved guideline”. In: *Clinical and Laboratory Standards Institute (CLSI) Documents* 28 (2008), pp. 19087–1898.

[79] Huma Siddiqui et al. “Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons”. In: *BMC microbiology* 11 (2011), pp. 1–12.

[80] Salvador Lladó Fernández, Tomáš Větrovskỳ, and Petr Baldrian. “The concept of operational taxonomic units revisited: genomes of bacteria that are regarded as closely related are often highly dissimilar”. In: *Folia microbiologica* 64 (2019), pp. 19–23.

[81] J Gregory Caporaso et al. “QIIME allows analysis of high-throughput community sequencing data”. In: *Nature methods* 7.5 (2010), pp. 335–336.

[82] Evan Bolyen et al. “Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2”. In: *Nature biotechnology* 37.8 (2019), pp. 852–857.

[83] Patrick D. Schloss et al. “Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities”. In: *Applied and Environmental Microbiology* 75.23 (Dec. 1, 2009), pp. 7537–7541. ISSN: 0099-2240, 1098-5336. DOI: 10.1128/AEM.01541-09. pmid: 19801464. URL: https://aem.asm.org/content/75/23/7537 (visited on 12/09/2019).

[84] Mahesh S Desai et al. “A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility”. In: *Cell* 167.5 (2016), pp. 1339–1353.

[85] Benjamin J Callahan et al. “DADA2: High-resolution sample inference from Illumina amplicon data”. In: *Nature methods* 13.7 (2016), pp. 581–583.

[86] Robert C. Edgar. “Search and clustering orders of magnitude faster than BLAST”. In: *Bioinformatics* 26.19 (Aug. 2010), pp. 2460–2461. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq461. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/19/2460/48857155/bioinformatics_26_19_2460.pdf. URL: https://doi.org/10.1093/bioinformatics/btq461.

[87] Robert C Edgar. “UPARSE: highly accurate OTU sequences from microbial amplicon reads”. In: *Nature methods* 10.10 (2013), pp. 996–998.

[88] Robert C Edgar. “UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing”. In: *BioRxiv* (2016), p. 081257.

[89] Andrei Prodan et al. “Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing”. In: *PLoS One* 15.1 (2020), e0227434.

[90] Jari Oksanen et al. *vegan: Community Ecology Package*. R package version 1.15-1. 2008. URL: http://cran.r-project.org/,%20http://vegan.r-forge.r-project.org/.

[91] Paul J. McMurdie and Susan Holmes. “phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data”. In: *PLOS ONE* 8.4 (Apr. 2013), pp. 1–11. DOI: 10.1371/journal.pone.0061217. URL: https://doi.org/10.1371/journal.pone.0061217.

[92] Morgan G. I. Langille et al. “Predictive Functional Profiling of Microbial Communities Using 16S rRNA Marker Gene Sequences”. In: *Nature Biotechnology* 31.9 (Sept. 2013), pp. 814–821. ISSN: 1546-1696. DOI: 10.1038/nbt.2676. URL: https://www.nature.com/articles/nbt.2676 (visited on 12/09/2019).

[93] Gavin M. Douglas et al. “PICRUSt2 for Prediction of Metagenome Functions”. In: *Nature Biotechnology* 38.6 (6 June 2020), pp. 685–688. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0548-6. URL: https://www.nature.com/articles/s41587-020-0548-6 (visited on 12/18/2022).

[94] Kathrin P. Aßhauer et al. “Tax4Fun: Predicting Functional Profiles from Metagenomic 16S rRNA Data”. In: *Bioinformatics* 31.17 (Sept. 1, 2015), pp. 2882–2884. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv287. URL: https://academic.oup.com/bioinformatics/article/31/17/2882/183768 (visited on 12/19/2019).

[95] S. Iwai et al. “Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes”. In: *Plos One* 11.11 (Nov. 2016), p. 18. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0166104.

[96] Elmar Pruesse et al. “SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB”. In: *Nucleic Acids Research* 35.21 (Oct. 2007), pp. 7188–7196. ISSN: 0305-1048. DOI: 10.1093/nar/gkm864.

[97] C. Quast et al. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools”. In: *Nucleic Acids Research* 41.D1 (Jan. 2013), pp. D590–D596. ISSN: 0305-1048. DOI: 10.1093/nar/gks1219.

[98] J. R. Cole et al. “Ribosomal Database Project: Data and Tools for High Throughput rRNA Analysis”. In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. D633–D642. ISSN: 0305-1048. DOI: 10.1093/nar/gkt1244.

[99] Daniel McDonald et al. “An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea”. In: *The ISME Journal* 6.3 (2012), pp. 610–618.

[100] Robert Edgar. “Taxonomy annotation and guide tree errors in 16S rRNA databases”. In: *PeerJ* 6 (2018). DOI: 10.7717/peerj.5030.

[101] Ann L. Griffen et al. “CORE: A Phylogenetically-Curated 16S rDNA Database of the Core Oral Microbiome”. In: *PLOS ONE* 6.4 (Apr. 2011), pp. 1–10. DOI: 10.1371/journal.pone.0019051. URL: https://doi.org/10.1371/journal.pone.0019051.

[102] Marco Meola et al. “DAIRYdb: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products”. In: *BMC Genomics* 20.1 (2019), p. 560.

[103] Simon Jon McIlroy et al. “MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups”. In: *Database* 2017 (Mar. 2017). bax016. ISSN: 1758-0463. DOI: 10.1093/database/bax016. eprint: https://academic.oup.com/database/article-pdf/doi/10.1093/database/bax016/19231478/bax016.pdf. URL: https://doi.org/10.1093/database/bax016.

[104] Morten Simonsen Dueholm et al. “Generation of comprehensive ecosystem-specific reference databases with species-level resolution by high-throughput full-length 16S rRNA gene sequencing and automated taxonomy assignment (AutoTax)”. In: *MBio* 11.5 (2020), e01557–20.

[105] Benjamin J Callahan, Paul J McMurdie, and Susan P Holmes. “Exact sequence variants should replace operational taxonomic units in marker-gene data analysis”. In: *The ISME journal* 11.12 (2017), pp. 2639–2643.

[106] Nilay Peker et al. “A comparison of three different bioinformatics analyses of the 16S–23S rRNA encoding region for bacterial identification”. In: *Frontiers in microbiology* 10 (2019), p. 620.

[107] Julien Tremblay et al. “Primer and platform effects on 16S rRNA tag sequencing”. In: *Frontiers in microbiology* 6 (2015), p. 771.

[108] Marlène Chiarello et al. “Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold”. In: *PLoS One* 17.2 (2022), e0264443.

[109] Dong-Lei Sun et al. “Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity”. In: *Applied and environmental microbiology* 79.19 (2013), pp. 5962–5969.

[110] Marcel Martinez-Porchas et al. “How conserved are the conserved 16S-rRNA regions?” In: *PeerJ* 5 (2017), e3036.

[111] Heiko Nacke et al. “Identification and Characterization of Novel Cellulolytic and Hemicellulolytic Genes and Enzymes Derived from German Grassland Soil Metagenomes”. In: *Biotechnology Letters* 34.4 (Apr. 1, 2012), pp. 663–675. ISSN: 1573-6776. DOI: 10.1007/s10529-011-0830-2. URL: https://doi.org/10.1007/s10529-011-0830-2 (visited on 12/18/2022).

[112] Lin-Xing Chen et al. “Accurate and Complete Genomes from Metagenomes”. In: *Genome Research* 30.3 (Jan. 3, 2020), pp. 315–333. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.258640.119. pmid: 32188701. URL: https://genome.cshlp.org/content/30/3/315 (visited on 06/01/2021).

[113] Matthias Scholz et al. “Strain-Level Microbial Epidemiology and Population Genomics from Shotgun Metagenomics”. In: *Nature Methods* 13.5 (5 May 2016), pp. 435–438. ISSN: 1548-7105. DOI: 10.1038/nmeth.3802. URL: https://www.nature.com/articles/nmeth.3802 (visited on 12/18/2022).

[114] R. Ranjan et al. “Analysis of the Microbiome: Advantages of Whole Genome Shotgun versus 16S Amplicon Sequencing”. In: *Biochemical and Biophysical Research Communications* 469.4 (Jan. 2016), pp. 967–977. ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2015.12.083.

[115] Sara Saheb Kashaf et al. “Recovering prokaryotic genomes from host-associated, short-read shotgun metagenomic sequencing data”. In: *Nature Protocols* 16.5 (2021), pp. 2520–2541.

[116] Michael Tessler et al. “Large-Scale Differences in Microbial Biodiversity Discovery between 16S Amplicon and Shotgun Sequencing”. In: *Scientific Reports* 7.1 (July 31, 2017), pp. 1–14. ISSN: 2045-2322. DOI: 10.1038/s41598-017-06665-3. URL: https://www.nature.com/articles/s41598-017-06665-3 (visited on 12/19/2019).

[117] Adam G. Clooney et al. “Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis”. In: *PLOS ONE* 11.2 (Feb. 5, 2016), e0148028. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0148028. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148028 (visited on 12/18/2022).

[118] Duy Tin Truong et al. “MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling”. In: *Nature Methods* 12.10 (10 Oct. 2015), pp. 902–903. ISSN: 1548-7105. DOI: 10.1038/nmeth.3589. URL: https://www.nature.com/articles/nmeth.3589 (visited on 06/02/2021).

[119] D. E. Wood and S. L. Salzberg. “Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments”. In: *Genome Biology* 15.3 (2014), p. 12. ISSN: 1465-6906. DOI: 10.1186/gb-2014-15-3-r46.

[120] Juan Jovel et al. “Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics”. In: *Frontiers in Microbiology* 7 (2016). ISSN: 1664-302X. DOI: 10.3389/fmicb.2016.00459. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2016.00459/full (visited on 12/19/2019).

[121] H. Soon Gweon et al. “The Impact of Sequencing Depth on the Inferred Taxonomic Composition and AMR Gene Content of Metagenomic Samples”. In: *Environmental Microbiome* 14 (Oct. 24, 2019), p. 7. ISSN: 2524-6372. DOI: 10.1186/s40793-019-0347-1. pmid: 33902704. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8204541/ (visited on 09/27/2022).

[122] Donovan H. Parks et al. “Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life”. In: *Nature Microbiology* 2.11 (Nov. 2017), pp. 1533–1542. ISSN: 2058-5276. DOI: 10.1038/s41564-017-0012-7. URL: https://www.nature.com/articles/s41564-017-0012-7/ (visited on 12/19/2019).

[123] Taylor E. Reiter and C. Titus Brown. “MAGs Achieve Lineage Resolution”. In: *Nature Microbiology* 7.2 (2 Feb. 2022), pp. 193–194. ISSN: 2058-5276. DOI: 10.1038/s41564-021-01027-2. URL: https://www.nature.com/articles/s41564-021-01027-2 (visited on 04/12/2022).

[124] Derek M. Bickhart et al. “Generating Lineage-Resolved, Complete Metagenome-Assembled Genomes from Complex Microbial Communities”. In: *Nature Biotechnology* (Jan. 3, 2022), pp. 1–9. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01130-z. URL: https://www.nature.com/articles/s41587-021-01130-z (visited on 04/12/2022).

[125] Adriana Rego et al. “Secondary Metabolite Biosynthetic Diversity in Arctic Ocean Metagenomes”. In: *Microbial Genomics* 7.12 (), p. 000731. ISSN: 2057-5858, DOI: 10.1099/mgen.0.000731. URL: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000731 (visited on 04/12/2022).

[126] Maria J. Soto-Giron et al. “The Edible Plant Microbiome Represents a Diverse Genetic Reservoir with Functional Potential in the Human Host”. In: *Scientific Reports* 11.1 (1 Dec. 15, 2021), p. 24017. ISSN: 2045-2322. DOI: 10.1038/s41598-021-03334-4. URL: https://www.nature.com/articles/s41598-021-03334-4 (visited on 04/12/2022).

[127] Dylan G. Maghini et al. “Improved High-Molecular-Weight DNA Extraction, Nanopore Sequencing and Metagenomic Assembly from the Human Gut Microbiome”. In: *Nature Protocols* 16.1 (1 Jan. 2021), pp. 458–471. ISSN: 1750-2799. DOI: 10.1038/s41596-020-00424-x. URL: https://www.nature.com/articles/s41596-020-00424-x (visited on 09/27/2022).

[128] Liang Chen et al. “Short- and Long-Read Metagenomics Expand Individualized Structural Variations in Gut Microbiomes”. In: *Nature Communications* 13.1 (1 June 8, 2022), p. 3175. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30857-9. URL: https://www.nature.com/articles/s41467-022-30857-9 (visited on 09/27/2022).

[129] Peter Menzel, Kim Lee Ng, and Anders Krogh. “Fast and Sensitive Taxonomic Classification for Metagenomics with Kaiju”. In: *Nature Communications* 7.1 (1 Apr. 13, 2016), p. 11257. ISSN: 2041-1723. DOI: 10.1038/ncomms11257. URL: https://www.nature.com/articles/ncomms11257 (visited on 06/03/2021).

[130] Andrew J. McArdle and Myrsini Kaforou. “Sensitivity of Shotgun Metagenomics to Host DNA: Abundance Estimates Depend on Bioinformatic Tools and Contamination Is the Main Issue”. In: *Access Microbiology* 2.4 (Feb. 17, 2020), acmi000104. ISSN: 2516-8290. DOI: 10.1099/acmi.0.000104. pmid: 33005868. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7523627/ (visited on 10/05/2022).

[131] Simon H. Ye et al. “Benchmarking Metagenomics Tools for Taxonomic Classification”. In: *Cell* 178.4 (Aug. 8, 2019), pp. 779–794. ISSN: 0092-8674. DOI: 10.1016/j.cell.2019.07.010. URL: https://www.sciencedirect.com/science/article/pii/S0092867419307755 (visited on 04/12/2022).

[132] Florian P Breitwieser, Jennifer Lu, and Steven L Salzberg. “A Review of Methods and Databases for Metagenomic Classification and Assembly”. In: *Briefings in Bioinformatics* 20.4 (July 19, 2019), pp. 1125–1136. ISSN: 1477-4054. DOI: 10.1093/bib/bbx120. URL: https://doi.org/10.1093/bib/bbx120 (visited on 10/05/2022).

[133] C. Quince et al. “Shotgun Metagenomics, from Sampling to Analysis (Vol 35, Pg 833, 2017)”. In: *Nature Biotechnology* 35.12 (Dec. 2017), pp. 1211–1211. ISSN: 1087-0156. DOI: 10.1038/nbt1217-1211b.

[134] Nicola Segata et al. “Metagenomic Microbial Community Profiling Using Unique Clade-Specific Marker Genes”. In: *Nature Methods* 9.8 (8 Aug. 2012), pp. 811–814. ISSN: 1548-7105. DOI: 10.1038/nmeth.2066. URL: https://www.nature.com/articles/nmeth.2066 (visited on 06/19/2020).

[135] Aitor Blanco-Miguez et al. *Extending and Improving Metagenomic Taxonomic Profiling with Uncharacterized Species with MetaPhlAn 4*. Aug. 22, 2022. DOI: 10.1101/2022.08.22.504593. URL: https://www.biorxiv.org/content/10.1101/2022.08.22.504593v1 (visited on 10/05/2022). preprint.

[136] Francesco Beghini et al. “Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with bioBakery 3”. In: *eLife* 10 (May 4, 2021). Ed. by Peter Turnbaugh, Eduardo Franco, and C Titus Brown, e65088. ISSN: 2050-084X. DOI: 10.7554/eLife.65088. URL: https://doi.org/10.7554/eLife.65088 (visited on 03/23/2022).

[137] Shinichi Sunagawa et al. “Metagenomic species profiling using universal phylogenetic marker genes”. In: *Nature methods* 10.12 (2013), pp. 1196–1199.

[138] Daniel R Mende et al. “Accurate and universal delineation of prokaryotic species”. In: *Nature methods* 10.9 (2013), pp. 881–884.

[139] Alessio Milanese et al. “Microbial Abundance, Activity and Population Genomic Profiling with mOTUs2”. In: *Nature Communications* 10.1 (1 Mar. 4, 2019), p. 1014. ISSN: 2041-1723. DOI: 10.1038/s41467-019-08844-4. URL: https://www.nature.com/articles/s41467-019-08844-4 (visited on 03/19/2023).

[140] Bo Liu et al. “Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences”. In: *Genome biology* 12 (2011), pp. 1–27.

[141] Nicole M Davis et al. “Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data”. In: *Microbiome* 6 (2018), pp. 1–14.

[142] Pierre Pericard et al. “MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes”. In: *Bioinformatics* 34.4 (2018), pp. 585–591.

[143] David A Fitzpatrick et al. “A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis”. In: *BMC evolutionary biology* 6.1 (2006), pp. 1–15.

[144] Francesco Asnicar et al. “Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0”. In: *Nature communications* 11.1 (2020), p. 2500.

[145] Alexis Criscuolo. “A fast alignment-free bioinformatics procedure to infer accurate distance-based phylogenetic trees from genome assemblies”. In: *Research Ideas and Outcomes* 5 (2019), e36178.

[146] Peter Kämpfer et al. “Devosia equisanguinis sp. nov., isolated from horse blood”. In: *International Journal of Systematic and Evolutionary Microbiology* 71.11 (2021), p. 005090.

[147] Edgar Badell et al. “Corynebacterium rouxii sp. nov., a novel member of the diphtheriae species complex”. In: *Research in microbiology* 171.3-4 (2020), pp. 122–127.

[148] Stefan E Heiden et al. “A Klebsiella pneumoniae ST307 outbreak clone from Germany demonstrates features of extensive drug resistance, hypermucoviscosity, and enhanced iron acquisition”. In: *Genome Medicine* 12 (2020), pp. 1–15.

[149] Peter Kämpfer et al. “Paenibacillus allorhizosphaerae sp. nov., from soil of the rhizosphere of Zea mays”. In: *International Journal of Systematic and Evolutionary Microbiology* 71.10 (2021), p. 005051.

[150] Metin Balaban et al. “Genome-wide alignment-free phylogenetic distance estimation under a no strand-bias model”. In: *Bioinformatics Advances* 2.1 (2022), vbac055.

[151] Samuel V Angiuoli and Steven L Salzberg. “Mugsy: fast multiple alignment of closely related whole genomes”. In: *Bioinformatics* 27.3 (2011), pp. 334–342.

[152] Aaron C.E. Darling et al. “Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements”. In: *Genome Research* 14.7 (July 2004), pp. 1394–1403. ISSN: 1088-9051. DOI: 10.1101/gr.2289704. pmid: 15231754. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC442156/ (visited on 12/07/2022).

[153] Jiao Pan et al. “The insect-killing bacterium Photorhabdus luminescens has the lowest mutation rate among bacteria”. In: *Marine life science & technology* 3 (2021), pp. 20–27.

[154] Anne M Estes et al. “Comparative genomics of the Erwinia and Enterobacter olive fly endosymbionts”. In: *Scientific Reports* 8.1 (2018), p. 15936.

[155] Mohammad Tarequl Islam et al. “Draft genome sequences of nine Vibrio sp. isolates from across the United States closely related to Vibrio cholerae”. In: *Microbiology Resource Announcements* 7.21 (2018), e00965–18.

[156] Jean-Claude Ogier et al. “rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing”. In: *BMC microbiology* 19.1 (2019), pp. 1–16.

[157] Li C Xia et al. “Accurate genome relative abundance estimation based on shotgun metagenomic reads”. In: *PloS one* 6.12 (2011), e27992.

[158] Xin Bai, Jie Ren, and Fengzhu Sun. “MLR-OOD: A Markov Chain Based Likelihood Ratio Method for Out-Of-Distribution Detection of Genomic Sequences”. In: *Journal of Molecular Biology* 434.15 (2022), p. 167586.

[159] Vladimir I Ulyantsev et al. “MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data”. In: *Bioinformatics* 32.18 (2016), pp. 2760–2767.

[160] Gherman V Uritskiy, Jocelyne DiRuggiero, and James Taylor. “MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis”. In: *Microbiome* 6.1 (2018), pp. 1–13.

[161] Carlo Ferravante et al. “HOME-BIO (sHOtgun MEtagenomic analysis of BIOlogical entities): a specific and comprehensive pipeline for metagenomic shotgun sequencing data analysis”. In: *BMC bioinformatics* 22.7 (2021), pp. 1–10.

[162] Dinghua Li et al. “MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph”. In: *Bioinformatics* 31.10 (May 15, 2015), pp. 1674–1676. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv033.

[163] Sergey Nurk et al. “metaSPAdes: a new versatile metagenomic assembler”. In: *Genome research* 27.5 (2017), pp. 824–834.

[164] Hsin-Hung Lin and Yu-Chieh Liao. “Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes”. In: *Scientific reports* 6.1 (2016), p. 24175.

[165] Johannes Alneberg et al. “Binning Metagenomic Contigs by Coverage and Composition”. In: *Nature Methods* 11.11 (11 Nov. 2014), pp. 1144–1146. ISSN: 1548-7105. DOI: 10.1038/nmeth.3103. URL: https://www.nature.com/articles/nmeth.3103 (visited on 06/03/2021).

[166] Vijini Mallawaarachchi, Anuradha Wickramarachchi, and Yu Lin. “GraphBin: Refined Binning of Metagenomic Contigs Using Assembly Graphs”. In: *Bioinformatics* 36.11 (June 1, 2020), pp. 3307–3313. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa180. URL: https://academic.oup.com/bioinformatics/article/36/11/3307/5804980 (visited on 06/30/2020).

[167] H Bjørn Nielsen et al. “Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes”. In: *Nature biotechnology* 32.8 (2014), pp. 822–828.

[168] Paul D Donovan et al. “Identification of fungi in shotgun metagenomics datasets”. In: *PLoS One* 13.2 (2018), e0192898.

[169] Zixuan Xie and Chaysavanh Manichanh. “FunOMIC: Pipeline with built-in fungal taxonomic and functional databases for human mycobiome profiling”. In: *Computational and Structural Biotechnology Journal* 20 (2022), pp. 3685–3694.

[170] Matteo Soverini et al. “HumanMycobiomeScan: a new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples”. In: *BMC genomics* 20 (2019), pp. 1–7.

[171] Vanessa R. Marcelino, Edward C. Holmes, and Tania C. Sorrell. “The Use of Taxon-Specific Reference Databases Compromises Metagenomic Classification”. In: *BMC Genomics* 21.1 (Feb. 27, 2020), p. 184. ISSN: 1471-2164. DOI: 10.1186/s12864-020-6592-2. URL: https://doi.org/10.1186/s12864-020-6592-2 (visited on 06/14/2020).

[172] Shuai Wang, Yiqi Jiang, and Shuaicheng Li. “PStrain: an iterative microbial strains profiling algorithm for shotgun metagenomic sequencing data”. In: *Bioinformatics* 36.22-23 (2020), pp. 5499–5506.

[173] Xin Li, Haiyan Hu, and Xiaoman Li. “mixtureS: a novel tool for bacterial strain genome reconstruction from reads”. In: *Bioinformatics* 37.4 (2021), pp. 575–577.

[174] Christopher S Smillie et al. “Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation”. In: *Cell host & microbe* 23.2 (2018), pp. 229–240.

[175] Manimozhiyan Arumugam et al. “SmashCommunity: a metagenomic annotation and analysis tool”. In: *Bioinformatics* 26.23 (2010), pp. 2977–2978.

[176] Genivaldo Gueiros Z Silva et al. “SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data”. In: *Bioinformatics* 32.3 (2016), pp. 354–361.

[177] Jens Roat Kultima et al. “MOCAT2: a metagenomic assembly, annotation and profiling framework”. In: *Bioinformatics* 32.16 (2016), pp. 2520–2523.

[178] Jens Roat Kultima et al. “MOCAT: a metagenomics assembly and gene prediction toolkit”. In: (2012).

[179] Richard J Randle-Boggis et al. “Evaluating techniques for metagenome annotation using simulated sequence data”. In: *FEMS microbiology ecology* 92.7 (2016).

[180] Kevin Chen and Lior Pachter. “Bioinformatics for whole-genome shotgun sequencing of microbial communities”. In: *PLoS computational biology* 1.2 (2005), e24.

[181] Victor Kunin et al. “A bioinformatician’s guide to metagenomics”. In: *Microbiology and molecular biology reviews* 72.4 (2008), pp. 557–578.

[182] Mincheol Kim et al. “Analytical tools and databases for metagenomics in the next-generation sequencing era”. In: *Genomics & informatics* 11.3 (2013), pp. 102–113.

[183] Vincent J Henry et al. “OMICtools: an informative directory for multi-omic data analysis”. In: *Database* 2014 (2014).

[184] Matthew Thoendel et al. “Comparison of three commercial tools for metagenomic shotgun sequencing analysis”. In: *Journal of clinical microbiology* 58.3 (2020), e00981–19.

[185] Ana Elena Pérez-Cobas, Laura Gomez-Valero, and Carmen Buchrieser. “Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses”. In: *Microbial genomics* 6.8 (2020).

[186] Bas E Dutilh et al. “Perspective on taxonomic classification of uncultivated viruses”. In: *Current opinion in virology* 51 (2021), pp. 207–215.

[187] M. Bonnet et al. “Bacterial Culture through Selective and Non-Selective Conditions: The Evolution of Culture Media in Clinical Microbiology”. In: *New Microbes and New Infections* 34 (Mar. 1, 2020), p. 100622. ISSN: 2052-2975. DOI: 10.1016/j.nmni.2019.100622. URL: https://www.sciencedirect.com/science/article/pii/S2052297519301192 (visited on 11/20/2022).

[188] S. Strugger. “Fluorescence Microscope Examination of Bacteria in Soil”. In: *Canadian Journal of Research* 26c.2 (Apr. 1948), pp. 188–193. ISSN: 1923-4287. DOI: 10.1139/cjr48c-019. URL: https://cdnsciencepub.com/doi/10.1139/cjr48c-019 (visited on 12/29/2022).

[189] Alexander Fleming. “On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzæ”. In: *British journal of experimental pathology* 10.3 (June 1929), pp. 226–236. ISSN: 0007-1021. pmid: null. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2048009/ (visited on 11/25/2022).

[190] BarryJ Marshall and J. Robin Warren. “UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC ULCERATION”. In: *The Lancet*. Originally Published as Volume 1, Issue 8390 323.8390 (June 16, 1984), pp. 1311–1315. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(84)91816-6. URL: https://www.sciencedirect.com/science/article/pii/S0140673684918166 (visited on 11/25/2022).

[191] Robert Koch. “Die Äetiologie der Tuberkulose”. In: *Berliner Klinische Wochenschrift* 15 (1882), pp. 221–230.

[192] Jean-Christophe Lagier et al. “Current and Past Strategies for Bacterial Culture in Clinical Microbiology”. In: *Clinical Microbiology Reviews* 28.1 (Jan. 1, 2015), pp. 208–236. ISSN: 0893-8512, 1098-6618. DOI: 10.1128/CMR.00110-14. pmid: 25567228. URL: https://cmr.asm.org/content/28/1/208 (visited on 12/20/2020).

[193] Ramzi Ghodbane, Didier Raoult, and Michel Drancourt. “Dramatic Reduction of Culture Time of Mycobacterium Tuberculosis”. In: *Scientific Reports* 4.1 (1 Feb. 28, 2014), p. 4236. ISSN: 2045-2322. DOI: 10.1038/srep04236. URL: https://www.nature.com/articles/srep04236 (visited on 11/25/2022).

[194] R. Phillip Dellinger et al. “Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2008”. In: *Intensive Care Medicine* 34.1 (Jan. 1, 2008), pp. 17–60. ISSN: 1432-1238. DOI: 10.1007/s00134-007-0934-2. URL: https://doi.org/10.1007/s00134-007-0934-2 (visited on 11/25/2022).

[195] K C Mathura et al. “Study of Clinical Profile and Antibiotic Sensitivity Pattern in Culture Positive Typhoid Fever Cases”. In: *Kathmandu University medical journal (KUMJ)* 3.4 (Oct. 1, 2005), pp. 376–379. ISSN: 1812-2078. pmid: 16449839.

[196] Sara Cuadros-Orellana, Metchild Pohlschröder, and Lucia R. Durrant. “Isolation and Characterization of Halophilic Archaea Able to Grow in Aromatic Compounds”. In: *International Biodeterioration & Biodegradation* 57.3 (Apr. 1, 2006), pp. 151–154. ISSN: 0964-8305. DOI: 10.1016/j.ibiod.2005.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0964830506000205 (visited on 11/25/2022).

[197] B. Ozcan et al. “Characterization of Extremely Halophilic Archaea Isolated from Saline Environment in Different Parts of Turkey”. In: *Microbiology* 75.6 (Dec. 1, 2006), pp. 739–746. ISSN: 1608-3237. DOI: 10.1134/S002626170606018X. URL: https://doi.org/10.1134/S002626170606018X (visited on 11/25/2022).

[198] Wayne L. Duck et al. “Isolation of Flagellated Bacteria Implicated in Crohn’s Disease”. In: *Inflammatory Bowel Diseases* 13.10 (Oct. 1, 2007), pp. 1191–1201. ISSN: 1078-0998. DOI: 10.1002/ibd.20237. URL: https://doi.org/10.1002/ibd.20237 (visited on 11/25/2022).

[199] Joshua W. Gatson et al. “Bacillus Tequilensis Sp. Nov., Isolated from a 2000-Year-Old Mexican Shaft-Tomb, Is Closely Related to Bacillus Subtilis”. In: *International Journal of Systematic and Evolutionary Microbiology* 56.7 (), pp. 1475–1484. ISSN: 1466-5034, DOI: 10.1099/ijs.0.63946-0. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.63946-0 (visited on 11/25/2022).

[200] V.A. Vicente et al. “Environmental Isolation of Black Yeast-like Fungi Involved in Human Infection”. In: *Studies in Mycology* 61.1 (June 1, 2008), pp. 137–144. DOI: 10.3114/sim.2008.61.14.

[201] Tom D’Elia et al. “Isolation of Fungi from Lake Vostok Accretion Ice”. In: *Mycologia* 101.6 (Nov. 1, 2009), pp. 751–763. ISSN: 0027-5514. DOI: 10.3852/08-184. URL: https://doi.org/10.3852/08-184 (visited on 11/25/2022).

[202] J. -C. Lagier et al. “Microbial Culturomics: Paradigm Shift in the Human Gut Microbiome Study”. In: *Clinical Microbiology and Infection* 18.12 (Dec. 1, 2012), pp. 1185–1193. ISSN: 1198-743X. DOI: 10.1111/1469-0691.12023. URL: https://www.sciencedirect.com/science/article/pii/S1198743X14608041 (visited on 11/25/2022).

[203] G. Greub. “Culturomics: A New Approach to Study the Human Microbiome”. In: *Clinical Microbiology and Infection* 18.12 (Dec. 1, 2012), pp. 1157–1159. ISSN: 1198-743X. DOI: 10.1111/1469-0691.12032. pmid: 23148445. URL: https://www.clinicalmicrobiologyandinfec
com/article/S1198-743X(14)60799-0/fulltext (visited on 11/25/2022).

[204] Gabriele Andrea Lugli et al. “A Microbiome Reality Check: Limitations of in Silico-Based Metagenomic Approaches to Study Complex Bacterial Communities”. In: *Environmental Microbiology Reports* 11.6 (2019), pp. 840–847. ISSN: 1758-2229. DOI: 10.1111/1758-2229.12805. URL: https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12805 (visited on 11/22/2019).

[205] A. Pfleiderer et al. “Culturomics Identified 11 New Bacterial Species from a Single Anorexia Nervosa Stool Sample”. In: *European Journal of Clinical Microbiology & Infectious Diseases* 32.11 (Nov. 1, 2013), pp. 1471–1481. ISSN: 1435-4373. DOI: 10.1007/s10096-013-1900-2. URL: https://doi.org/10.1007/s10096-013-1900-2 (visited on 11/25/2022).

[206] Awa Diop et al. “Microbial Culturomics Unravels the Halophilic Microbiota Repertoire of Table Salt: Description of Gracilibacillus Massiliensis Sp. Nov.” In: *Microbial Ecology in Health and Disease* 27.1 (Jan. 1, 2016), p. 32049. ISSN: null. DOI: 10.3402/mehd.v27.32049. pmid: 27760679. URL: https://www.tandfonline.com/doi/abs/10.3402/mehd.v27.32049 (visited on 11/25/2022).

[207] Emmanouil Angelakis et al. “MALDI-TOF Mass Spectrometry and Identification of New Bacteria Species in Air Samples from Makkah, Saudi Arabia”. In: *BMC Research Notes* 7.1 (Dec. 9, 2014), p. 892. ISSN: 1756-0500. DOI: 10.1186/1756-0500-7-892. URL: https://doi.org/10.1186/1756-0500-7-892 (visited on 11/25/2022).

[208] Karissa L. Cross et al. “Targeted Isolation and Cultivation of Uncultivated Bacteria by Reverse Genomics”. In: *Nature Biotechnology* 37.11 (11 Nov. 2019), pp. 1314–1321. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0260-6. URL: https://www.nature.com/articles/s41587-019-0260-6 (visited on 11/25/2022).

[209] Karsten Zengler et al. “High-Throughput Cultivation of Microorganisms Using Microcapsules”. In: *Methods in Enzymology*. Vol. 397. Environmental Microbiology. Academic Press, Jan. 1, 2005, pp. 124–130. DOI: 10.1016/S0076-6879(05)97007-9. URL: https://www.sciencedirect.com/science/article/pii/S0076687905970079 (visited on 11/26/2022).

[210] Stanislav S. Terekhov et al. “Microfluidic Droplet Platform for Ultrahigh-Throughput Single-Cell Screening of Biodiversity”. In: *Proceedings of the National Academy of Sciences* 114.10 (Mar. 7, 2017), pp. 2550–2555. DOI: 10.1073/pnas.1621226114. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1621226114 (visited on 11/26/2022).

[211] William H. Lewis et al. “Innovations to Culturing the Uncultured Microbial Majority”. In: *Nature Reviews Microbiology* 19.4 (4 Apr. 2021), pp. 225–240. ISSN: 1740-1534. DOI: 10.1038/s41579-020-00458-8. URL: https://www.nature.com/articles/s41579-020-00458-8 (visited on 11/25/2022).

[212] Lindsey Bomar et al. “Directed Culturing of Microorganisms Using Metatranscriptomics”. In: *mBio* 2.2 (Apr. 5, 2011), e00012–11. DOI: 10.1128/mBio.00012-11. URL: https://journals.asm.org/doi/full/10.1128/mBio.00012-11 (visited on 09/27/2022).

[213] Gene W. Tyson et al. “Genome-Directed Isolation of the Key Nitrogen Fixer Leptospirillum Ferrodiazotrophum Sp. Nov. from an Acidophilic Microbial Community”. In: *Applied and Environmental Microbiology* 71.10 (Oct. 2005), pp. 6319–6324. DOI: 10.1128/AEM.71.10.6319-6324.2005. URL: https://journals.asm.org/doi/full/10.1128/AEM.71.10.6319-6324.2005 (visited on 11/25/2022).

[214] Maneesh Dave et al. “The Human Gut Microbiome: Current Knowledge, Challenges, and Future Directions”. In: *Translational Research*. In-Depth Review: Of Microbes and Men: Challenges of the Human Microbiome 160.4 (Oct. 1, 2012), pp. 246–257. ISSN: 1931-5244. DOI: 10.1016/j.trsl.2012.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1931524412001624 (visited on 11/21/2022).

[215] P. D. Cani. “Human Gut Microbiome: Hopes, Threats and Promises”. In: *Gut* 67.9 (Sept. 2018), pp. 1716–1725. ISSN: 0017-5749. DOI: 10.1136/gutjnl-2018-316723.

[216] Allyson L. Byrd, Yasmine Belkaid, and Julia A. Segre. “The Human Skin Microbiome”. In: *Nature Reviews Microbiology* 16.3 (3 Mar. 2018), pp. 143–155. ISSN: 1740-1534. DOI: 10.1038/nrmicro.2017.157. URL: https://www.nature.com/articles/nrmicro.2017.157 (visited on 11/21/2022).

[217] Linda Abou Chacra and Florence Fenollar. “Exploring the Global Vaginal Microbiome and Its Impact on Human Health”. In: *Microbial Pathogenesis* 160 (Nov. 1, 2021), p. 105172. ISSN: 0882-4010. DOI: 10.1016/j.micpath.2021.105172. URL: https://www.sciencedirect.com/science/article/pii/S0882401021004460 (visited on 11/21/2022).

[218] Kazuma Yagi et al. “The Lung Microbiome during Health and Disease”. In: *International Journal of Molecular Sciences* 22.19 (19 Jan. 2021), p. 10872. ISSN: 1422-0067. DOI: 10.3390/ijms221910872. URL: https://www.mdpi.com/1422-0067/22/19/10872 (visited on 11/21/2022).

[219] R. E. Ley et al. “Microbial Ecology - Human Gut Microbes Associated with Obesity”. In: *Nature* 444.7122 (Dec. 2006), pp. 1022–1023. ISSN: 0028-0836. DOI: 10.1038/nature4441022a.

[220] Vancheswaran Gopalakrishnan et al. “The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy”. In: *Cancer cell* 33.4 (2018), pp. 570–580.

[221] Ibrahim Hamad et al. “Metabarcoding Analysis of Eukaryotic Microbiota in the Gut of HIV-infected Patients”. In: *PLOS ONE* 13.1 (Jan. 31, 2018), e0191913. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0191913. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191913 (visited on 05/31/2020).

[222] Khanh Phuong S Tong et al. “Association between hemoglobin A1c, Vitamin C, and microbiome in diabetic foot ulcers and intact skin: A cross-sectional study”. In: *Health Science Reports* 5.5 (2022), e718.

[223] Nicola Vitulo et al. “Bark and Grape Microbiome of Vitis Vinifera: Influence of Geographic Patterns and Agronomic Management on Bacterial Diversity”. In: *Frontiers in Microbiology* 9 (2019). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2018.03203 (visited on 12/09/2022).

[224] Pankaj Trivedi et al. “Plant–microbiome interactions: from community assembly to plant health”. In: *Nature reviews microbiology* 18.11 (2020), pp. 607–621.

[225] Susan R. Whitehead et al. “The Apple Microbiome: Structure, Function, and Manipulation for Improved Plant Health”. In: *The Apple Genome*. Ed. by Schuyler S. Korban. Compendium of Plant Genomes. Cham: Springer International Publishing, 2021, pp. 341–382. ISBN: 978-3-030-74682-7. DOI: 10.1007/978-3-030-74682-7_16. URL: https://doi.org/10.1007/978-3-030-74682-7_16 (visited on 12/09/2022).

[226] Unyarat Ritpitakphong et al. “The Microbiome of the Leaf Surface of Arabidopsis Protects against a Fungal Pathogen”. In: *New Phytologist* 210.3 (2016), pp. 1033–1043. ISSN: 1469-8137. DOI: 10.1111/nph.13808. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.13808 (visited on 12/09/2022).

[227] Ana Pineda et al. “Conditioning the Soil Microbiome through Plant–Soil Feedbacks Suppresses an Aboveground Insect Pest”. In: *New Phytologist* 226.2 (2020), pp. 595–608. ISSN: 1469-8137. DOI: 10.1111/nph.16385. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.16385 (visited on 12/09/2022).

[228] Samy Selim et al. “Selection of Newly Identified Growth-Promoting Archaea Haloferax Species With a Potential Action on Cobalt Resistance in Maize Plants”. In: *Frontiers in Plant Science* 13 (2022).

[229] David Johnston-Monje, Janneth P Gutiérrez, and Luis Augusto Becerra Lopez-Lavalle. “Seed-transmitted bacteria and fungi dominate juvenile plant microbiomes”. In: *Frontiers in Microbiology* 12 (2021), p. 737616.

[230] Eleanor R Kirkman et al. “Diversity and ecological guild analysis of the oil palm fungal microbiome across root, rhizosphere, and soil compartments”. In: *Frontiers in Microbiology* 13 (2022), p. 209.

[231] Ling Xu et al. “Holo-Omics for Deciphering Plant-Microbiome Interactions”. In: *Microbiome* 9.1 (Mar. 24, 2021), p. 69. ISSN: 2049-2618. DOI: 10.1186/s40168-021-01014-z. URL: https://doi.org/10.1186/s40168-021-01014-z (visited on 03/25/2021).

[232] Posy E Busby et al. “Facilitating reforestation through the plant microbiome: Perspectives from the phyllosphere”. In: *Annual Review of Phytopathology* 60 (2022), pp. 337–356.

[233] Philipp Dirksen et al. “CeMbio - The Caenorhabditis Elegans Microbiome Resource”. In: *G3 Genes|Genomes|Genetics* 10.9 (Sept. 1, 2020), pp. 3025–3039. ISSN: 2160-1836. DOI: 10.1534/g3.120.401309. URL: https://doi.org/10.1534/g3.120.401309 (visited on 12/10/2022).

[234] Fan Zhang et al. “Caenorhabditis Elegans as a Model for Microbiome Research”. In: *Frontiers in Microbiology* 8 (2017). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00485 (visited on 12/11/2022).

[235] M.m. O’ Donnell et al. “The Core Faecal Bacterial Microbiome of Irish Thoroughbred Racehorses”. In: *Letters in Applied Microbiology* 57.6 (2013), pp. 492–501. ISSN: 1472-765X. DOI: 10.1111/lam.12137. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/lam.12137 (visited on 12/09/2022).

[236] Rachel Gilroy et al. “Metagenomic Investigation of the Equine Faecal Microbiome Reveals Extensive Taxonomic Diversity”. In: *PeerJ* 10 (Mar. 23, 2022), e13084. ISSN: 2167-8359. DOI: 10.7717/peerj.13084. URL: https://peerj.com/articles/13084 (visited on 12/09/2022).

[237] James E. McDonald et al. “Characterising the Canine Oral Microbiome by Direct Sequencing of Reverse-Transcribed rRNA Molecules”. In: *PLOS ONE* 11.6 (June 8, 2016), e0157046. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0157046. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157046 (visited on 12/10/2022).

[238] Hillary A. Craddock et al. “Phenotypic Correlates of the Working Dog Microbiome”. In: *npj Biofilms and Microbiomes* 8.1 (1 Aug. 22, 2022), pp. 1–9. ISSN: 2055-5008. DOI: 10.1038/s41522-022-00329-5. URL: https://www.nature.com/articles/s41522-022-00329-5 (visited on 12/10/2022).

[239] Guiding Li et al. “Comparative and Functional Analyses of Fecal Microbiome in Asian Elephants”. In: *Antonie van Leeuwenhoek* 115.9 (Sept. 1, 2022), pp. 1187–1202. ISSN: 1572-9699. DOI: 10.1007/s10482-022-01757-1. URL: https://doi.org/10.1007/s10482-022-01757-1 (visited on 12/10/2022).

[240] Parul Mittal et al. “The Gene Catalog and Comparative Analysis of Gut Microbiome of Big Cats Provide New Insights on Panthera Species”. In: *Frontiers in Microbiology* 11 (2020). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01012 (visited on 12/10/2022).

[241] Alice Michel et al. “Isolated Grauer’s Gorilla Populations Differ in Diet and Gut Microbiome”. In: *Molecular Ecology* n/a.n/a (). ISSN: 1365-294X. DOI: 10.1111/mec.16663. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16663 (visited on 12/11/2022).

[242] Vanessa A. Leone et al. “Atypical Behavioral and Thermoregulatory Circadian Rhythms in Mice Lacking a Microbiome”. In: *Scientific Reports* 12.1 (1 Aug. 25, 2022), p. 14491. ISSN: 2045-2322. DOI: 10.1038/s41598-022-18291-9. URL: https://www.nature.com/articles/s41598-022-18291-9 (visited on 12/11/2022).

[243] Jian Ran, Qiu-Hong Wan, and Sheng-Guo Fang. “Gut Microbiota of Endangered Crested Ibis: Establishment, Diversity, and Association with Reproductive Output”. In: *PLOS ONE* 16.4 (Apr. 23, 2021), e0250075. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0250075. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0250075 (visited on 01/17/2022).

[244] Maureen Berg et al. “Assembly of the Caenorhabditis Elegans Gut Microbiota from Diverse Soil Microbial Environments”. In: *Isme Journal* 10.8 (Aug. 2016), pp. 1998–2009. ISSN: 1751-7362. DOI: 10.1038/ismej.2015.253.

[245] Jinmei Ding et al. “Inheritance and Establishment of Gut Microbiota in Chickens”. In: *Frontiers in Microbiology* 8 (2017). ISSN: 1664-302X. DOI: 10.3389/fmicb.2017.01967. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01967/full (visited on 06/01/2020).

[246] W. Jin et al. “The Bacterial and Archaeal Community Structures and Methanogenic Potential of the Cecal Microbiota of Goats Fed with Hay and High-Grain Diets”. In: *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 111.11 (Nov. 2018), pp. 2037–2049. ISSN: 0003-6072. DOI: 10.1007/s10482-018-1096-7.

[247] Y. Zheng et al. “Gut Microbiota Analysis of Juvenile Genetically Improved Farmed Tilapia (Oreochromis Niloticus) by Dietary Supplementation of Different Resveratrol Concentrations”. In: *Fish & Shellfish Immunology* 77 (June 2018), pp. 200–207. ISSN: 1050-4648. DOI: 10.1016/j.fsi.2018.03.040.

[248] M. a. A. Mamun et al. “The Composition and Stability of the Faecal Microbiota of Merino Sheep”. In: *Journal of Applied Microbiology* n/a.n/a (). ISSN: 1365-2672. DOI: 10.1111/jam.14468. URL: https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/jam.14468 (visited on 12/03/2019).

[249] J. H. Campbell et al. “Host Genetic and Environmental Effects on Mouse Intestinal Microbiota”. In: *Isme Journal* 6.11 (Nov. 2012), pp. 2033–2044. ISSN: 1751-7362. DOI: 10.1038/ismej.2012.54.

[250] C. F. Maurice et al. “Marked Seasonal Variation in the Wild Mouse Gut Microbiota”. In: *Isme Journal* 9.11 (Nov. 2015), pp. 2423–2434. ISSN: 1751-7362. DOI: 10.1038/ismej.2015.53.

[251] Mirna Čoklo, Dina Rešetar Maslov, and Sandra Kraljević Pavelić. “Modulation of Gut Microbiota in Healthy Rats after Exposure to Nutritional Supplements”. In: *Gut Microbes* 12.1 (Nov. 9, 2020), p. 1779002. ISSN: 1949-0976. DOI: 10.1080/19490976.2020.1779002. pmid: 32845788. URL: https://doi.org/10.1080/19490976.2020.1779002 (visited on 12/27/2022).

[252] J. Killer et al. “Lactobacillus Rodentium Sp. Nov., from the Digestive Tract of Wild Rodents”. In: *International Journal of Systematic and Evolutionary Microbiology* 64 (Pt_5), pp. 1526–1533. ISSN: 1466-5034, DOI: 10.1099/ijs.0.054924-0. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.054924-0 (visited on 12/19/2022).

[253] S. Andrés-Lasheras et al. “Presence of Clostridium Difficile in Pig Faecal Samples and Wild Animal Species Associated with Pig Farms”. In: *Journal of Applied Microbiology* 122.2 (2017), pp. 462–472. ISSN: 1365-2672. DOI: 10.1111/jam.13343. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jam.13343 (visited on 11/25/2022).

[254] Anthony C. Hilton, Richard J. Willis, and Samantha J. Hickie. “Isolation of Salmonella from Urban Wild Brown Rats (Rattus Norvegicus) in the West Midlands, UK”. In: *International Journal of Environmental Health Research* 12.2 (June 1, 2002), pp. 163–168. ISSN: 0960-3123. DOI: 10.1080/09603120220129328. pmid: 12396533. URL: https://doi.org/10.1080/09603120220129328 (visited on 11/25/2022).

[255] Dong-Ho Chang et al. “Faecalibaculum Rodentium Gen. Nov., Sp. Nov., Isolated from the Faeces of a Laboratory Mouse”. In: *Antonie van Leeuwenhoek* 108.6 (Dec. 1, 2015), pp. 1309–1318. ISSN: 1572-9699. DOI: 10.1007/s10482-015-0583-3. URL: https://doi.org/10.1007/s10482-015-0583-3 (visited on 11/25/2022).

[256] Kar Hui Ong et al. “Occurrence and Antimicrobial Resistance Traits of Escherichia Coli from Wild Birds and Rodents in Singapore”. In: *International Journal of Environmental Research and Public Health* 17.15 (15 Jan. 2020), p. 5606. ISSN: 1660-4601. DOI: 10.3390/ijerph17155606. URL: https://www.mdpi.com/1660-4601/17/15/5606 (visited on 11/25/2022).

[257] Shoukui Hu et al. “Helicobacter Himalayensis Sp. Nov. Isolated from Gastric Mucosa of Marmota Himalayana”. In: *International Journal of Systematic and Evolutionary Microbiology* 65 (Pt_6), pp. 1719–1725. ISSN: 1466-5034, DOI: 10.1099/ijs.0.000163. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.000163 (visited on 11/25/2022).

[258] Lina Niu et al. “Streptococcus Halotolerans Sp. Nov. Isolated from the Respiratory Tract of Marmota Himalayana in Qinghai-Tibet Plateau of China”. In: *International Journal of Systematic and Evolutionary Microbiology* 66.10 (), pp. 4211–4217. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.001337. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.001337 (visited on 11/25/2022).

[259] Lina Niu et al. “Streptococcusmarmotae Sp. Nov., Isolated from the Respiratory Tract of Marmota Himalayana”. In: *International Journal of Systematic and Evolutionary Microbiology* 66.11 (), pp. 4315–4322. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.001350. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.001350 (visited on 11/25/2022).

[260] Lina Niu et al. “Streptococcus Himalayensis Sp. Nov., Isolated from the Respiratory Tract of Marmota Himalayana”. In: *International Journal of Systematic and Evolutionary Microbiology* 67.2 (), pp. 256–261. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.001609. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.001609 (visited on 11/25/2022).

[261] Jiajia Meng et al. “Lactobacillus Xujianguonis Sp. Nov., Isolated from Faeces of Marmota Himalayana”. In: *International Journal of Systematic and Evolutionary Microbiology* 70.1 (), pp. 11–15. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.003598. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.003598 (visited on 11/25/2022).

[262] Melissa Soh et al. “Schaedlerella Arabinosiphila Gen. Nov., Sp. Nov., a D-arabinose-utilizing Bacterium Isolated from Faeces of C57BL/6J Mice That Is a Close Relative of Clostridium Species ASF 502”. In: *International Journal of Systematic and Evolutionary Microbiology* 69.11 (), pp. 3616–3622. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.003671. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.003671 (visited on 11/25/2022).

[263] Maja Lukac et al. “Bacterial and Fungal Flora in Faecal Samples from the Balkan Snow Vole (Dinaromys Bogdanovi) in Captivity”. In: *Journal of Zoo and Aquarium Research* 5.4 (4 Oct. 31, 2017), pp. 167–171. ISSN: 2214-7594. DOI: 10.19227/jzar.v5i4.293. URL: https://jzar.org/jzar/article/view/293 (visited on 11/25/2022).

[264] Chen Ben-Tzvi. “The Ecological Factors Affecting the Occurrence and Abundance of Escherichia Coli and Its Colicins in Wild Rodents”. MA thesis. Albert Katz International School of Desert Research: Ben-Gurion University of the Negev, 2019. 34 pp.

[265] Anthony P. Neumann, Caroline A. McCormick, and Garret Suen. “Fibrobacter Communities in the Gastrointestinal Tracts of Diverse Hindgut-Fermenting Herbivores Are Distinct from Those of the Rumen”. In: *Environmental Microbiology* 19.9 (2017), pp. 3768–3783. ISSN: 1462-2920. DOI: 10.1111/1462-2920.13878. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.13878 (visited on 11/25/2022).

[266] Jens Walter et al. “Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents”. In: *Cell* 180.2 (Jan. 23, 2020), pp. 221–232. ISSN: 0092-8674. DOI: 10.1016/j.cell.2019.12.025. URL: http://www.sciencedirect.com/science/article/pii/S009286741931387X (visited on 01/31/2020).

[267] Michael A Pellizzon and Matthew R Ricci. “Effects of rodent diet choice and fiber type on data interpretation of gut microbiome and metabolic disease research”. In: *Current protocols in toxicology* 77.1 (2018), e55.

[268] J. Z. Hu et al. “Effect of Postnatal Low-Dose Exposure to Environmental Chemicals on the Gut Microbiome in a Rodent Model”. In: *Microbiome* 4 (June 2016), p. 11. ISSN: 2049-2618. DOI: 10.1186/s40168-016-0173-2.

[269] Maria Isabel Queipo-Ortuno et al. “Gut microbiota composition in male rat models under different nutritional status and physical activity and its association with serum leptin and ghrelin levels”. In: *PloS one* 8.5 (2013), e65465.

[270] Peng Zheng et al. “Gut microbiome remodeling induces depressive like behaviors through a pathway mediated by the host’s metabolism”. In: *Molecular psychiatry* 21.6 (2016), pp. 786–796.

[271] John R Kelly et al. “Transferring the blues: depression-associated gut microbiota induces neurobehavioural changes in the rat”. In: *Journal of psychiatric research* 82 (2016), pp. 109–118.

[272] Deepak K Rajpal et al. “Selective spectrum antibiotic modulation of the gut microbiome in obesity and diabetes rodent models”. In: *PLoS One* 10.12 (2015), e0145499.

[273] P. J. Turnbaugh et al. “Diet Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome”. In: *Cell host & microbe* 3.4 (2008), pp. 213–223.

[274] Amandine Everard et al. “Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity”. In: *The ISME journal* 8.10 (2014), pp. 2116–2130.

[275] Serguei O Fetissov. “Role of the gut microbiota in host appetite control: bacterial growth to animal feeding behaviour”. In: *Nature Reviews Endocrinology* 13.1 (2017), pp. 11–25.

[276] Claire Barbier de La Serre et al. “Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation”. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 299.2 (2010), pp. 440–448.

[277] Kyung-Ah Kim et al. “High fat diet-induced gut microbiota exacerbates inflammation and obesity in mice via the TLR4 signaling pathway”. In: (2012).

[278] Wannipa Tunapong et al. “Chronic treatment with prebiotics, probiotics and synbiotics attenuated cardiac dysfunction by improving cardiac mitochondrial dysfunction in male obese insulin-resistant rats”. In: *European journal of nutrition* 57 (2018), pp. 2091–2104.

[279] Chun Cui et al. “Vancomycin pretreatment on MPTP-induced parkinson’s disease mice exerts neuroprotection by suppressing inflammation both in brain and gut”. In: *Journal of Neuroimmune Pharmacology* (2022), pp. 1–18.

[280] Ye Tu et al. “The associations of gut microbiota, endocrine system and bone metabolism”. In: *Frontiers in Microbiology* 14 (2023).

[281] Mengyang Xu et al. “Difference in post-stress recovery of the gut microbiome and its altered metabolism after chronic adolescent stress in rats”. In: *Scientific Reports* 10.1 (2020), pp. 1–10.

[282] Joël W Jameson, Denis Réale, and Steven W Kembel. “Gut microbiome modulates behaviour and life history in two wild rodents”. In: *BioRxiv* (2020), pp. 2020–02.

[283] Mustafa Sibai et al. “Microbiome and longevity: high abundance of longevity-linked muribaculaceae in the gut of the long-living rodent spalax leucodon”. In: *OMICS: A Journal of Integrative Biology* 24.10 (2020), pp. 592–601.

[284] Miriam Linnenbrink et al. “The role of biogeography in shaping diversity of the intestinal microbiota in house mice”. In: *Molecular ecology* 22.7 (2013), pp. 1904–1916.

[285] Laura Weldon et al. “The gut microbiota of wild mice”. In: *PLoS One* 10.8 (2015), e0134643.

[286] Simon H Williams et al. “New York City house mice (Mus musculus) as potential reservoirs for pathogenic bacteria and antimicrobial resistance determinants”. In: *MBio* 9.2 (2018), e00624–18.

[287] Moira A Gilliver et al. “Antibiotic resistance found in wild rodents”. In: *Nature* 401.6750 (1999), pp. 233–234.

[288] Taichi A Suzuki and Michael W Nachman. “Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice”. In: *PloS one* 11.9 (2016), e0163720.

[289] K. D. Kohl et al. “Herbivorous Rodents (Neotoma Spp.) Harbour Abundant and Active Foregut Microbiota”. In: *Environmental Microbiology* 16.9 (Sept. 2014), pp. 2869–2878. ISSN: 1462-2912. DOI: 10.1111/1462-2920.12376.

[290] Jakub Kreisinger et al. “Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing”. In: *Molecular Ecology* 23.20 (2014), pp. 5048–5060.

[291] Jacqueline M Leung et al. “Rapid environmental effects on gut nematode susceptibility in rewilded mice”. In: *PLoS biology* 16.3 (2018), e2004108.

[292] Stephan P. Rosshart et al. “Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance”. In: *Cell* 171.5 (Nov. 16, 2017), 1015–1028.e13. ISSN: 0092-8674. DOI: 10.1016/j.cell.2017.09.016. URL: http://www.sciencedirect.com/science/article/pii/S0092867417310656 (visited on 09/28/2020).

[293] Benedikt Hild et al. “Neonatal exposure to a wild-derived microbiome protects mice against diet-induced obesity”. In: *Nature metabolism* 3.8 (2021), pp. 1042–1057.

[294] Afrizal Afrizal et al. “Enhanced Cultured Diversity of the Mouse Gut Microbiota Enables Custom-Made Synthetic Communities”. In: *Cell Host & Microbe* 30.11 (Nov. 9, 2022), 1630–1645.e25. ISSN: 1931-3128. DOI: 10.1016/j.chom.2022.09.011. pmid: 36208631. URL: https://www.cell.com/cell-host-microbe/abstract/S1931-3128(22)00467-X (visited on 12/27/2022).

[295] Peter Osborne et al. “A Rather Dry Subject; Investigating the Study of Arid-Associated Microbial Communities”. In: *Environmental Microbiome* 15.1 (Dec. 1, 2020), p. 20. ISSN: 2524-6372. DOI: 10.1186/s40793-020-00367-6. URL: https://doi.org/10.1186/s40793-020-00367-6 (visited on 09/04/2022).

[296] M. Coryell et al. “The Gut Microbiome Is Required for Full Protection against Acute Arsenic Toxicity in Mouse Models”. In: *Nature Communications* 9 (Dec. 2018), p. 9. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07803-9.

[297] Victor T. Schmidt et al. “Community Assembly of a Euryhaline Fish Microbiome during Salinity Acclimation”. In: *Molecular Ecology* 24.10 (2015), pp. 2537–2550. ISSN: 1365-294X. DOI: 10.1111/mec.13177. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13177 (visited on 12/11/2022).

[298] Bachar Cheaib et al. “The Yellow Perch (Perca Flavescens) Microbiome Revealed Resistance to Colonisation Mostly Associated with Neutralism Driven by Rare Taxa under Cadmium Disturbance”. In: *Animal Microbiome* 3.1 (Jan. 5, 2021), p. 3. ISSN: 2524-4671. DOI: 10.1186/s42523-020-00063-3. URL: https://doi.org/10.1186/s42523-020-00063-3 (visited on 12/11/2022).

[299] Anton Lavrinienko et al. “Skin and Gut Microbiomes of a Wild Mammal Respond to Different Environmental Cues”. In: *Microbiome* 6.1 (Nov. 26, 2018), p. 209. ISSN: 2049-2618. DOI: 10.1186/s40168-018-0595-0. URL: https://doi.org/10.1186/s40168-018-0595-0 (visited on 11/22/2019).

[300] Gary R. Graves et al. “Does Solar Irradiation Drive Community Assembly of Vulture Plumage Microbiotas?” In: *Animal Microbiome* 2.1 (July 14, 2020), p. 24. ISSN: 2524-4671. DOI: 10.1186/s42523-020-00043-7. URL: https://doi.org/10.1186/s42523-020-00043-7 (visited on 08/06/2020).

[301] Claudia Barelli et al. “The Gut Microbiota Communities of Wild Arboreal and Ground-Feeding Tropical Primates Are Affected Differently by Habitat Disturbance”. In: *mSystems* 5.3 (June 30, 2020). ISSN: 2379-5077. DOI: 10.1128/mSystems.00061-20. pmid: 32457237. URL: https://msystems.asm.org/content/5/3/e00061-20 (visited on 10/28/2020).

[302] Rachel A. Arango et al. “Experimental Warming Reduces Survival, Cold Tolerance, and Gut Prokaryotic Diversity of the Eastern Subterranean Termite, Reticulitermes Flavipes (Kollar)”. In: *Frontiers in Microbiology* 12 (2021). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2021.632715 (visited on 10/30/2022).

[303] Claire Chevalier et al. “Gut Microbiota Orchestrates Energy Homeostasis during Cold”. In: *Cell* 163.6 (Dec. 3, 2015), pp. 1360–1374. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.11.004. URL: http://www.sciencedirect.com/science/article/pii/S0092867415014841 (visited on 12/16/2019).

[304] Liang Chi et al. “The Effects of an Environmentally Relevant Level of Arsenic on the Gut Microbiome and Its Functional Metagenome”. In: *Toxicological Sciences* 160.2 (Dec. 1, 2017), pp. 193–204. ISSN: 1096-6080. DOI: 10.1093/toxsci/kfx174. URL: https://doi.org/10.1093/toxsci/kfx174 (visited on 10/30/2022).

[305] Qian Wang, Timothy R McDermott, and Seth T Walk. “A Single Microbiome Gene Alters Murine Susceptibility to Acute Arsenic Exposure”. In: *Toxicological Sciences* 181.1 (May 1, 2021), pp. 105–114. ISSN: 1096-6080. DOI: 10.1093/toxsci/kfab017. URL: https://doi.org/10.1093/toxsci/kfab017 (visited on 10/30/2022).

[306] Angélica Jaramillo and Luis E. Castañeda. “Gut Microbiota of Drosophila Subobscura Contributes to Its Heat Tolerance and Is Sensitive to Transient Thermal Stress”. In: *Frontiers in Microbiology* 12 (2021). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2021.654108 (visited on 10/30/2022).

[307] Juan Sepulveda and Andrew H. Moeller. “The Effects of Temperature on Animal Gut Microbiomes”. In: *Frontiers in Microbiology* 11 (2020). ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.00384. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00384/full (visited on 05/06/2020).

[308] David Casero et al. “Space-Type Radiation Induces Multimodal Responses in the Mouse Gut Microbiome and Metabolome”. In: *Microbiome* 5.1 (Aug. 18, 2017), p. 105. ISSN: 2049-2618. DOI: 10.1186/s40168-017-0325-z. URL: https://doi.org/10.1186/s40168-017-0325-z (visited on 10/30/2022).

[309] Young Suk Kim, Jinu Kim, and Soo-Je Park. “High-Throughput 16S rRNA Gene Sequencing Reveals Alterations of Mouse Intestinal Microbiota after Radiotherapy”. In: *Anaerobe* 33 (June 1, 2015), pp. 1–7. ISSN: 1075-9964. DOI: 10.1016/j.anaerobe.2015.01.004. URL: https://www.sciencedirect.com/science/article/pii/S1075996415000050 (visited on 10/30/2022).

[310] Jérôme Breton et al. “Ecotoxicology inside the Gut: Impact of Heavy Metals on the Mouse Microbiome”. In: *BMC Pharmacology and Toxicology* 14.1 (Dec. 11, 2013), p. 62. ISSN: 2050-6511. DOI: 10.1186/2050-6511-14-62. URL: https://doi.org/10.1186/2050-6511-14-62 (visited on 10/30/2022).

[311] Maja Šrut et al. “Earthworms and Cadmium – Heavy Metal Resistant Gut Bacteria as Indicators for Heavy Metal Pollution in Soils?” In: *Ecotoxicology and Environmental Safety* 171 (Apr. 30, 2019), pp. 843–853. ISSN: 0147-6513. DOI: 10.1016/j.ecoenv.2018.12.102. URL: https://www.sciencedirect.com/science/article/pii/S0147651318314003 (visited on 10/30/2022).

[312] M. M. Abdelsattar et al. “Impacts of Saline Water Stress on Livestock Production: A Review”. In: *SVU-International Journal of Agricultural Sciences* 2.1 (Jan. 1, 2020), pp. 1–12. ISSN: 2636-3801. DOI: 10.21608/svuijas.2020.67635. URL: https://svuijas.journals.ekb.eg/article_67635.html (visited on 12/13/2022).

[313] Tyler G. Evans and Dietmar Kültz. “The Cellular Stress Response in Fish Exposed to Salinity Fluctuations”. In: *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology* 333.6 (2020), pp. 421–435. ISSN: 2471-5646. DOI: 10.1002/jez.2350. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jez.2350 (visited on 12/13/2022).

[314] Tomasz Dulski et al. “Effect of Salinity on the Gut Microbiome of Pike Fry (Esox Lucius)”. In: *Applied Sciences* 10.7 (7 Jan. 2020), p. 2506. ISSN: 2076-3417. DOI: 10.3390/app10072506. URL: https://www.mdpi.com/2076-3417/10/7/2506 (visited on 10/30/2022).

[315] Anakena M. Castillo et al. “Salinity Effects on the Microbiome of a Neotropical Water Strider”. In: *Hydrobiologia* (Nov. 16, 2021). ISSN: 1573-5117. DOI: 10.1007/s10750-021-04732-5. URL: https://doi.org/10.1007/s10750-021-04732-5 (visited on 10/30/2022).

[316] Daniela Rosado et al. “Longitudinal Sampling of External Mucosae in Farmed European Seabass Reveals the Impact of Water Temperature on Bacterial Dynamics”. In: *ISME Communications* 1.1 (1 June 21, 2021), pp. 1–11. ISSN: 2730-6151. DOI: 10.1038/s43705-021-00019-x. URL: https://www.nature.com/articles/s43705-021-00019-x (visited on 10/30/2022).

[317] Samantha S. Fontaine, Patrick M. Mineo, and Kevin D. Kohl. “Experimental Manipulation of Microbiota Reduces Host Thermal Tolerance and Fitness under Heat Stress in a Vertebrate Ectotherm”. In: *Nature Ecology & Evolution* 6.4 (4 Apr. 2022), pp. 405–417. ISSN: 2397-334X. DOI: 10.1038/s41559-022-01686-2. URL: https://www.nature.com/articles/s41559-022-01686-2 (visited on 10/30/2022).

[318] Fotini Kokou et al. “Host Genetic Selection for Cold Tolerance Shapes Microbiome Composition and Modulates Its Response to Temperature”. In: *eLife* 7 (Nov. 20, 2018). Ed. by Wendy S Garrett and Rob Knight, e36398. ISSN: 2050-084X. DOI: 10.7554/eLife.36398. URL: https://doi.org/10.7554/eLife.36398 (visited on 12/16/2019).

[319] Yunhua Zhang et al. “Decline in Symbiont-Dependent Host Detoxification Metabolism Contributes to Increased Insecticide Susceptibility of Insects under High Temperature”. In: *The ISME Journal* 15.12 (12 Dec. 2021), pp. 3693–3703. ISSN: 1751-7370. DOI: 10.1038/s41396-021-01046-1. URL: https://www.nature.com/articles/s41396-021-01046-1 (visited on 10/30/2022).

[320] Bin Wang et al. “Ambient Temperature Structures the Gut Microbiota of Zebrafish to Impact the Response to Radioactive Pollution”. In: *Environmental Pollution* 293 (Jan. 15, 2022), p. 118539. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2021.118539. URL: https://www.sciencedirect.com/science/article/pii/S0269749121021217 (visited on 10/30/2022).

[321] C. J. Barrow. “World Atlas of Desertification”. In: *Land Degradation & Development* 3.4 (1992), pp. 249–249. ISSN: 1099-145X. DOI: 10.1002/ldr.3400030407. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ldr.3400030407 (visited on 12/09/2019).

[322] D Häussinger. “The Role of Cellular Hydration in the Regulation of Cell Function.” In: *Biochemical Journal* 313 (Pt 3 Feb. 1, 1996), pp. 697–710. ISSN: 0264-6021. pmid: 8611144. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1216967/ (visited on 12/09/2019).

[323] J. Yu and Y. Steinberger. “Vertical Distribution of Soil Microbial Biomass and Its Association with Shrubs from the Negev Desert”. In: *Journal of Arid Environments* 78 (Mar. 1, 2012), pp. 110–118. ISSN: 0140-1963. DOI: 10.1016/j.jaridenv.2011.11.012. URL: https://www.sciencedirect.com/science/article/pii/S0140196311003466 (visited on 12/16/2022).

[324] Arnoldus Schytte Blix. “Adaptations to Polar Life in Mammals and Birds”. In: *Journal of Experimental Biology* 219.8 (Apr. 15, 2016), pp. 1093–1105. ISSN: 0022-0949, 1477-9145. DOI: 10.1242/jeb.120477. pmid: 27103673. URL: https://jeb.biologists.org/content/219/8/1093 (visited on 12/09/2019).

[325] N. P. Taylor and D. C. Zappi. “An Alternative View of Generic Delimitation and Relationships in Tribe Cereeae (Cactaceae)”. In: *Bradleya* 1989.7 (Dec. 1989), pp. 13–40. ISSN: 0265-086X. DOI: 10.25223/brad.n7.1989.a2. URL: https://bioone.org/journals/Bradleya/volume-1989/issue-7/brad.n7.1989.a2/An-alternative-view-of-generic-delimitation-and-relationships-in-tribe/10.25223/brad.n7.1989.a2.full (visited on 12/09/2019).

[326] Muhammad Nadeem et al. “Research Progress and Perspective on Drought Stress in Legumes: A Review”. In: *International Journal of Molecular Sciences* 20.10 (10 Jan. 2019), p. 2541. DOI: 10.3390/ijms20102541. URL: https://www.mdpi.com/1422-0067/20/10/2541 (visited on 05/22/2020).

[327] R. N. B. Kay. “Responses of African Livestock and Wild Herbivores to Drought”. In: *Journal of Arid Environments* 37.4 (Dec. 1, 1997), pp. 683–694. ISSN: 0140-1963. DOI: 10.1006/jare.1997.0299. URL: http://www.sciencedirect.com/science/article/pii/S0140196397902998 (visited on 02/19/2020).

[328] Clara Ivette Rincón-Molina et al. “Structure and Diversity of the Bacterial Communities in the Acid and Thermophilic Crater-Lake of the Volcano “El Chichón”, Mexico”. In: *Geomicrobiology Journal* 36.2 (Feb. 7, 2019), pp. 97–109. ISSN: 0149-0451. DOI: 10.1080/01490451.2018.1509158. URL: https://doi.org/10.1080/01490451.2018.1509158 (visited on 11/21/2022).

[329] Jiwen Liu et al. “Proliferation of Hydrocarbon-Degrading Microbes at the Bottom of the Mariana Trench”. In: *Microbiome* 7.1 (Apr. 12, 2019), p. 47. ISSN: 2049-2618. DOI: 10.1186/s40168-019-0652-3. URL: https://doi.org/10.1186/s40168-019-0652-3 (visited on 11/21/2022).

[330] Robert M. Bowers et al. “Characterization of Airborne Microbial Communities at a High-Elevation Site and Their Potential To Act as Atmospheric Ice Nuclei”. In: *Applied and Environmental Microbiology* 75.15 (Aug. 2009), pp. 5121–5130. DOI: 10.1128/AEM.00447-09. URL: https://journals.asm.org/doi/full/10.1128/AEM.00447-09 (visited on 11/21/2022).

[331] Julianne L. Baron et al. “Shift in the Microbial Ecology of a Hospital Hot Water System Following the Introduction of an On-Site Monochloramine Disinfection System”. In: *PLOS ONE* 9.7 (July 17, 2014), e102679. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0102679. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0102679 (visited on 11/21/2022).

[332] Marina Fomina, Ji Won Hong, and Geoffrey Michael Gadd. “Effect of Depleted Uranium on a Soil Microcosm Fungal Community and Influence of a Plant-Ectomycorrhizal Association”. In: *Fungal Biology*. Fungal Adaption to Stress 124.5 (May 1, 2020), pp. 289–296. ISSN: 1878-6146. DOI: 10.1016/j.funbio.2019.08.001. URL: https://www.sciencedirect.com/science/article/pii/S1878614619301047 (visited on 11/21/2022).

[333] A. Almatroudi et al. “Staphylococcus Aureus Dry-Surface Biofilms Are More Resistant to Heat Treatment than Traditional Hydrated Biofilms”. In: *Journal of Hospital Infection* 98.2 (Feb. 1, 2018), pp. 161–167. ISSN: 0195-6701. DOI: 10.1016/j.jhin.2017.09.007. URL: https://www.sciencedirect.com/science/article/pii/S0195670117305042 (visited on 11/21/2022).

[334] Salem, B. B. “Arid Zone Forestry: A Guide for Field Technicians”. In: *FAO Conservation Guide*. Vol. 20. Rome, Italy: FAO, 1989. ISBN: 92-5-102809-5.

[335] Omer Lavy et al. “Microbiome-Related Aspects of Locust Density-Dependent Phase Transition”. In: *Environmental Microbiology* 24.1 (2022), pp. 507–516. ISSN: 1462-2920. DOI: 10.1111/1462-2920.15883. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.15883 (visited on 09/04/2022).

[336] Aarón Barraza et al. “Characterization of Microbial Communities from Rumen and Large Intestine of Lactating Creole Goats Grazing in Arid Plant Communities”. In: *Microbiology* 167.10 (Nov. 30, 2021), p. 001092. ISSN: 1350-0872, 1465-2080. DOI: 10.1099/mic.0.001092. URL: https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.001092 (visited on 09/04/2022).

[337] Claire E. Couch et al. “Bighorn Sheep Gut Microbiomes Associate with Genetic and Spatial Structure across a Metapopulation”. In: *Scientific Reports* 10.1 (1 Apr. 20, 2020), p. 6582. ISSN: 2045-2322. DOI: 10.1038/s41598-020-63401-0. URL: https://www.nature.com/articles/s41598-020-63401-0 (visited on 08/19/2020).

[338] Javad Gharechahi et al. “In-Depth Diversity Analysis of the Bacterial Community Resident in the Camel Rumen”. In: *Systematic and Applied Microbiology* 38.1 (Feb. 1, 2015), pp. 67–76. ISSN: 0723-2020. DOI: 10.1016/j.syapm.2014.09.004. URL: http://www.sciencedirect.com/science/article/pii/S072320201400143X (visited on 11/22/2019).

[339] Samantha Bird et al. “Geography, Seasonality, and Host-Associated Population Structure Influence the Fecal Microbiome of a Genetically Depauparate Arctic Mammal”. In: *Ecology and Evolution* 9.23 (2019), pp. 13202–13217. ISSN: 2045-7758. DOI: 10.1002/ece3.5768. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5768 (visited on 04/01/2020).

[340] Jing He et al. “Characterizing the Bacterial Microbiota in Different Gastrointestinal Tract Segments of the Bactrian Camel”. In: *Scientific Reports* 8.1 (1 Jan. 12, 2018), p. 654. ISSN: 2045-2322. DOI: 10.1038/s41598-017-18298-7. URL: https://www.nature.com/articles/s41598-017-18298-7 (visited on 09/04/2020).

[341] Kathrin Engel et al. “Family Matters: Skin Microbiome Reflects the Social Group and Spatial Proximity in Wild Zebra Finches”. In: *BMC Ecology* 20.1 (Nov. 13, 2020), p. 58. ISSN: 1472-6785. DOI: 10.1186/s12898-020-00326-2. URL: https://doi.org/10.1186/s12898-020-00326-2 (visited on 12/19/2022).

[342] Javad Gharechahi and Ghasem Hosseini Salekdeh. “A Metagenomic Analysis of the Camel Rumen’s Microbiome Identifies the Major Microbes Responsible for Lignocellulose Degradation and Fermentation”. In: *Biotechnology for Biofuels* 11.1 (Aug. 2, 2018), p. 216. ISSN: 1754-6834. DOI: 10.1186/s13068-018-1214-9. URL: https://doi.org/10.1186/s13068-018-1214-9 (visited on 03/31/2020).

[343] Zahra Nouri et al. “The Microbiota-Gut-Kidney Axis Mediates Host Osmoregulation in a Small Desert Mammal”. In: *npj Biofilms and Microbiomes* 8.1 (1 Apr. 4, 2022), pp. 1–10. ISSN: 2055-5008. DOI: 10.1038/s41522-022-00280-5. URL: https://www.nature.com/articles/s41522-022-00280-5 (visited on 09/04/2022).

[344] M. Delia Basanta et al. “Comparative Analysis of Skin Bacterial Diversity and Its Potential Antifungal Function Between Desert and Pine Forest Populations of Boreal Toads Anaxyrus Boreas”. In: *Microbial Ecology* 84.1 (July 1, 2022), pp. 257–266. ISSN: 1432-184X. DOI: 10.1007/s00248-021-01845-1. URL: https://doi.org/10.1007/s00248-021-01845-1 (visited on 09/04/2022).

[345] Han Aricha et al. “Comparative Analysis of Fecal Microbiota of Grazing Mongolian Cattle from Different Regions in Inner Mongolia, China”. In: *Animals* 11.7 (7 July 2021), p. 1938. ISSN: 2076-2615. DOI: 10.3390/ani11071938. URL: https://www.mdpi.com/2076-2615/11/7/1938 (visited on 09/04/2022).

[346] Raphael Eisenhofer, Kristofer M. Helgen, and David Taggart. “Signatures of Landscape and Captivity in the Gut Microbiota of Southern Hairy-nosed Wombats (Lasiorhinus Latifrons)”. In: *Animal Microbiome* 3.1 (Jan. 6, 2021), p. 4. ISSN: 2524-4671. DOI: 10.1186/s42523-020-00068-y. URL: https://doi.org/10.1186/s42523-020-00068-y (visited on 01/14/2021).

[347] Ângela M. Ribeiro et al. “31° South: The Physiology of Adaptation to Arid Conditions in a Passerine Bird”. In: *Molecular Ecology* 28.16 (2019), pp. 3709–3721. ISSN: 1365-294X. DOI: 10.1111/mec.15176. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15176 (visited on 12/03/2019).

[348] Chao Li et al. “Diet-Induced Microbiome Shifts of Sympatric Overwintering Birds”. In: *Applied Microbiology and Biotechnology* 105.14 (Aug. 1, 2021), pp. 5993–6005. ISSN: 1432-0614. DOI: 10.1007/s00253-021-11448-y. URL: https://doi.org/10.1007/s00253-021-11448-y (visited on 09/03/2021).

[349] Amanda C. Perofsky, Rebecca J. Lewis, and Lauren Ancel Meyers. “Terrestriality and Bacterial Transfer: A Comparative Study of Gut Microbiomes in Sympatric Malagasy Mammals”. In: *The ISME Journal* 13.1 (Jan. 2019), pp. 50–63. ISSN: 1751-7370. DOI: 10.1038/s41396-018-0251-5. URL: https://www.nature.com/articles/s41396-018-0251-5 (visited on 11/29/2019).

[350] Somasundhari Shanmuganandam et al. “Uncovering the Microbiome of Invasive Sympatric European Brown Hares and European Rabbits in Australia”. In: *PeerJ* 8 (Aug. 18, 2020), e9564. ISSN: 2167-8359. DOI: 10.7717/peerj.9564. URL: https://peerj.com/articles/9564 (visited on 12/13/2022).

[351] Sebastian Menke et al. “Oligotyping Reveals Differences between Gut Microbiomes of Free-Ranging Sympatric Namibian Carnivores (Acinonyx Jubatusl Canis Mesomelas) on a Bacterial Species-like Level”. In: *Frontiers in Microbiology* 5 (Oct. 14, 2014). ISSN: 1664-302X. DOI: 10.3389/fmicb.2014.00526.

[352] Ho-Kyung Song et al. “Environmental Filtering of Bacterial Functional Diversity along an Aridity Gradient”. In: *Scientific Reports* 9.1 (Jan. 29, 2019), pp. 1–10. ISSN: 2045-2322. DOI: 10.1038/s41598-018-37565-9. URL: https://www.nature.com/articles/s41598-018-37565-9 (visited on 11/20/2019).

[353] Ivo Sedláček et al. “Hymenobacter Amundsenii Sp. Nov. Resistant to Ultraviolet Radiation, Isolated from Regoliths in Antarctica”. In: *Systematic and Applied Microbiology* 42.3 (May 1, 2019), pp. 284–290. ISSN: 0723-2020. DOI: 10.1016/j.syapm.2018.12.004. URL: http://www.sciencedirect.com/science/article/pii/S0723202018300596 (visited on 09/01/2020).

[354] Arame Ndiaye et al. “Evolutionary Systematics and Biogeography of the Arid Habitat-Adapted Rodent Genus Gerbillus (Rodentia, Muridae): A Mostly Plio-Pleistocene African History”. In: *Journal of Zoological Systematics and Evolutionary Research* 54.4 (2016), pp. 299–317. ISSN: 1439-0469. DOI: 10.1111/jzs.12143. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jzs.12143 (visited on 12/24/2022).

[355] G. B. Diaz, R. A. Ojeda, and E. L. Rezende. “Renal Morphology, Phylogenetic History and Desert Adaptation of South American Hystricognath Rodents”. In: *Functional Ecology* 20.4 (2006), pp. 609–620. ISSN: 1365-2435. DOI: 10.1111/j.1365-2435.2006.01144.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2435.2006.01144.x (visited on 12/24/2022).

[356] A. Haim. “Thermoregulation and Metabolism of Wagner’s Gerbil (Gerbillus Dasyurus): A Rock Dwelling Rodent Adapted to Arid and Mesic Environments”. In: *Journal of Thermal Biology* 12.1 (June 1, 1987), pp. 45–48. ISSN: 0306-4565. DOI: 10.1016/0306-4565(87)90022-2. URL: https://www.sciencedirect.com/science/article/pii/0306456587900222 (visited on 12/24/2022).

[357] Richard E. MacMillen. “ADAPTIVE PHYSIOLOGY OF HETEROMYID RODENTS”. In: *Great Basin Naturalist Memoirs* 7 (1983), pp. 65–76. ISSN: 0160239X, 23312742. JSTOR: 23378268. URL: http://www.jstor.org/stable/23378268 (visited on 12/23/2022).

[358] B. G. Lovegrove. “The Metabolism of Social Subterranean Rodents: Adaptation to Aridity”. In: *Oecologia* 69.4 (July 1, 1986), pp. 551–555. ISSN: 1432-1939. DOI: 10.1007/BF00410361. URL: https://doi.org/10.1007/BF00410361 (visited on 12/24/2022).

[359] Rochelle Buffenstein. “Aspects of the Energetics and Renal Physiology of Some African Arid-Adapted Rodents”. University of Cape Town, 1984. URL: http://hdl.handle.net/11427/7612.

[360] T. Aghova et al. “Multiple Radiations of Spiny Mice (Rodentia: Acomys) in Dry Open Habitats of Afro-Arabia: Evidence from a Multi-Locus Phylogeny”. In: *Bmc Evolutionary Biology* 19 (Mar. 2019), p. 22. ISSN: 1471-2148. DOI: 10.1186/s12862-019-1380-9.

[361] Amiram Shkolnik and Arieh Borut. “Temperature and Water Relations in Two Species of Spiny Mice (Acomys)”. In: *Journal of Mammalogy* 50.2 (May 12, 1969), pp. 245–255. ISSN: 0022-2372. DOI: 10.2307/1378340. URL: https://academic.oup.com/jmammal/article/50/2/245/901099 (visited on 05/03/2020).

[362] Don E. Wilson, Russell A. Mittermeier, and Jr Thomas E. Lacher. “Muridae”. In: *Handbook of the Mammals of the World – Volume 7 Rodents II*. Barcelona: Lynx Edicions, Nov. 30, 2017, pp. 536–884. ISBN: 978-84-16728-04-6. DOI: 10.5281/zenodo.6887260. URL: https://zenodo.org/record/6887260 (visited on 12/24/2022).

[363] Nowak, R.M. *Walker's Mammals of the World, Fifth Edition*. 5th ed. Vol. 2. Baltimore: The Johns Hopkins University Press, 1991.

[364] Heinrich Mendelssohn. *Mammalia of Israel*. Fauna Palaestina. Israel Academy of Sciences and Humanities, 1999. ISBN: 978-965-208-013-4.

[365] Eleazar Shafrir and Jonathan H. Adler. “Enzymatic and Metabolic Responses to Affluent Diet of Two Diabetes-Prone Species of Spiny Mice: Acomys Cahirinus and Acomys Russatus”. In: *International Journal of Biochemistry* 15.12 (Jan. 1, 1983), pp. 1439–1446. ISSN: 0020-711X. DOI: 10.1016/0020-711X(83)90076-9. URL: https://www.sciencedirect.com/science/article/pii/0020711X83900769 (visited on 12/24/2022).

[366] Nava Zisapel et al. “Daily Scheduling of the Golden Spiny Mouse under Photoperiodic and Social Cues”. In: *Journal of Experimental Zoology* 284.1 (1999), pp. 100–106. ISSN: 1097-010X. DOI: 10.1002/(SICI)1097-010X(19990615)284:1<100::AID-JEZ13>3.0.CO;2-5. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-010X%2819990615%29284%3A1%3C100%3A%3AAID-JEZ13%3E3.0.CO%3B2-5 (visited on 12/24/2022).

[367] Georgy Shenbrot. “IUCN Red List of Threatened Species: Acomys Russatus”. In: *IUCN Red List of Threatened Species* (Feb. 2, 2016). URL: https://www.iucnredlist.org/en (visited on 12/24/2022).

[368] Noga Kronfeld et al. “COEXISTING POPULATIONS OF ACOMYS CAHIRINUS AND A. RUSSATUS: A PRELIMINARY REPORT”. In: *Israel Journal of Ecology and Evolution* 40.2 (May 18, 1994), pp. 177–183. ISSN: 2224-4662, 1565-9801. DOI: 10.1080/00212210.1994.10688746. URL: https://brill.com/view/journals/ijee/40/2/article-p177_9.xml (visited on 12/24/2022).

[369] Menna Jones and Tamar Dayan. “Foraging Behavior and Microhabitat Use by Spiny Mice, Acomys Cahirinus and A. Russatus, in the Presence of Blanford's Fox (Vulpes Cana) Odor”. In: *Journal of Chemical Ecology* 26.2 (Feb. 1, 2000), pp. 455–469. ISSN: 1573-1561. DOI: 10.1023/A:1005417707588. URL: https://doi.org/10.1023/A:1005417707588 (visited on 12/24/2022).

[370] N. Kronfeld-Schor and T. Dayan. “The Dietary Basis for Temporal Partitioning: Food Habits of Coexisting Acomys Species”. In: *Oecologia* 121.1 (Oct. 1999), pp. 123–128. ISSN: 0029-8549. DOI: 10.1007/s004420050913.

[371] Michael Scantlebury et al. “Comparative Seasonal Acclimatization of Food and Energy Consumption in Adjacent Populations of Common Spiny Mice (Acomys Cahirinus)”. In: *Journal of Zoology* 267.3 (2005), pp. 323–328. ISSN: 1469-7998. DOI: 10.1017/S095283690500751X. URL: https://onlinelibrary.wiley.com/doi/abs/10.1017/S095283690500751X (visited on 12/24/2022).

[372] H. Dickinson and D. W. Walker. “Managing a Colony of Spiny Mice (Acomys Cahirinus) for Perinatal Research”. In: *ANZCCART News* 20.1 (2007), pp. 4–10. ISSN: 1039-9089.

[373] Nadia Bellofiore et al. “First Evidence of a Menstruating Rodent: The Spiny Mouse (Acomys Cahirinus)”. In: *American Journal of Obstetrics and Gynecology* 216.1 (Jan. 1, 2017), 40.e1–40.e11. ISSN: 0002-9378. DOI: 10.1016/j.ajog.2016.07.041. URL: https://www.sciencedirect.com/science/article/pii/S0002937816304768 (visited on 12/24/2022).

[374] Richard H. Porter, John A. Matochik, and Jennifer W. Makin. “The Role of Familiarity in the Development of Social Preferences in Spiny Mice”. In: *Behavioural Processes* 9.2 (Apr. 1, 1984), pp. 241–254. ISSN: 0376-6357. DOI: 10.1016/0376-6357(84)90044-5. URL: https://www.sciencedirect.com/science/article/pii/0376635784900445 (visited on 12/24/2022).

[375] Richard H. Porter, John A. Matochik, and Jennifer W. Makin. “Discrimination between Full-Sibling Spiny Mice (Acomys Cahirinus) by Olfactory Signatures”. In: *Animal Behaviour* 34.4 (Aug. 1, 1986), pp. 1182–1188. ISSN: 0003-3472. DOI: 10.1016/S0003-3472(86)80178-6. URL: https://www.sciencedirect.com/science/article/pii/S0003347286801786 (visited on 12/24/2022).

[376] Jennifer Wheeler Makin and Richard H. Porter. “Paternal Behavior in the Spiny Mouse (Acomys Cahirinus)”. In: *Behavioral and Neural Biology* 41.2 (July 1, 1984), pp. 135–151. ISSN: 0163-1047. DOI: 10.1016/S0163-1047(84)90513-2. URL: https://www.sciencedirect.com/science/article/pii/S0163104784905132 (visited on 12/24/2022).

[377] Noa Pinter-Wollman et al. “Can Aggression Be the Force Driving Temporal Separation between Competing Common and Golden Spiny Mice?” In: *Journal of Mammalogy* 87.1 (Feb. 20, 2006), pp. 48–53. ISSN: 0022-2372. DOI: 10.1644/04-MAMM-A-194R2.1. URL: https://doi.org/10.1644/04-MAMM-A-194R2.1 (visited on 12/24/2022).

[378] Claude Baudoin, Abraham Haim, and Jean-Luc Durand. “Effect of Conspecific and Heterospecific Urine Odors on the Foraging Behavior of the Golden Spiny Mouse”. In: *Integrative Zoology* 8.s1 (2013), pp. 1–8. ISSN: 1749-4877. DOI: 10.1111/j.1749-4877.2012.00291.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-4877.2012.00291.x (visited on 12/24/2022).

[379] Francesca Cassola (Global Mammal Assessment). “IUCN Red List of Threatened Species: Acomys Cahirinus”. In: *IUCN Red List of Threatened Species* (Sept. 7, 2016). URL: https://www.iucnredlist.org/en (visited on 12/24/2022).

[380] Ruirui Hu et al. “A database of animal metagenomes”. In: *Scientific Data* 9.1 (2022), p. 312.

[381] Ran Yao et al. “Evaluation of the function of wild animal gut microbiomes using next-generation sequencing and bioinformatics and its relevance to animal conservation”. In: *Evolutionary Bioinformatics* 15 (2019), p. 1176934319848438.

[382] Nicholas D. Youngblut et al. “Large Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity”. In: *bioRxiv* (June 5, 2020), p. 2020.06.05.135962. DOI: 10.1101/2020.06.05.135962. URL: https://www.biorxiv.org/content/10.1101/2020.06.05.135962v1 (visited on 10/21/2020).

[383] Doron Levin et al. “Diversity and Functional Landscapes in the Microbiota of Animals in the Wild”. In: *Science* 372.6539 (Apr. 16, 2021). ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb5352. pmid: 33766942. URL: https://science.sciencemag.org/content/372/6539/eabb5352 (visited on 05/03/2021).

[384] Magdalena Skarżyńska et al. “A metagenomic glimpse into the gut of wild and domestic animals: Quantification of antimicrobial resistance and more”. In: *PLoS One* 15.12 (2020), e0242987.

[385] Michael Scantlebury et al. “Non-Shivering Thermogenesis in Common Spiny Mice Acomys Cahirinus from Adjacent Habitats: Response to Seasonal Acclimatization and Salinity Acclimation”. In: *Journal of Thermal Biology* 28.4 (May 1, 2003), pp. 287–293. ISSN: 0306-4565. DOI: 10.1016/S0306-4565(03)00005-6. URL: https://www.sciencedirect.com/science/article/pii/S0306456503000056 (visited on 11/26/2022).

[386] Tilaye Wube, Abraham Haim, and Fuad Fares. “Effect of Increased Dietary Salinity on the Reproductive Status and Energy Intake of Xeric and Mesic Populations of the Spiny Mouse, Acomys”. In: *Physiology & Behavior* 96.1 (Jan. 8, 2009), pp. 122–127. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2008.09.006. URL: https://www.sciencedirect.com/science/article/pii/S0031938408002813 (visited on 04/18/2022).

[387] Uri Shanas et al. “The Effects of Season and Dietary Salt Content on Body Temperature Daily Rhythms of Common Spiny Mice from Different Micro-Habitats”. In: *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 132.2 (June 1, 2002), pp. 287–295. ISSN: 1095-6433. DOI: 10.1016/S1095-6433(02)00033-8. URL: https://www.sciencedirect.com/science/article/pii/S1095643302000338 (visited on 11/26/2022).

[388] Uri Shanas and Abraham Haim. “Diet Salinity and Vasopressin as Reproduction Modulators in the Desert-Dwelling Golden Spiny Mouse (Acomys Russatus)”. In: *Physiology & Behavior* 81.4 (June 1, 2004), pp. 645–650. ISSN: 0031-9384. DOI: 10.1016/j.physbeh.2004.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0031938404000745 (visited on 11/26/2022).

[389] Udi Ron and Abraham Haim. “HOW DEHYDRATION AFFECTS THE THERMOREGULATORY ANDOSMOREGULATORY ABILITIES OF THE GOLDEN SPINY MOUSE ACOMYS RUSSATUS”. In: *Israel Journal of Ecology and Evolution* 47.1 (May 6, 2001), pp. 15–20. ISSN: 2224-4662, 1565-9801. DOI: 10.1560/Y6AH-6JWH-NNR8-DHVP. URL: https://brill.com/view/journals/ijee/47/1/article-p15_3.xml (visited on 11/26/2022).

[390] Arthur E. Harriman. “Preferences by Egyptian Spiny Mice for Solutions of Sugars, Salts, and Acids in Richter-Type Drinking Tests”. In: *Perceptual and Motor Skills* 50 (3_suppl June 1, 1980), pp. 1075–1081. ISSN: 0031-5125. DOI: 10.2466/pms.1980.50.3c.1075. URL: https://doi.org/10.2466/pms.1980.50.3c.1075 (visited on 11/26/2022).

[391] J. D. Rhoades, A. Kandiah, and A. M. Mashali. *The Use of Saline Waters for Crop Production*. FAO Irrigation and Drainage Papers 48. Rome, Italy: Food and Agriculture Organization of the United Nations, 1992. 134 pp. ISBN: 92-5-103237-8. URL: https://www.ars.usda.gov/arsuserfiles/20361500/pdf_pubs/P1313.pdf.

[392] Tahsin Ferdous et al. “The rise to power of the microbiome: power and sample size calculation for microbiome studies”. In: *Mucosal Immunology* 15.6 (2022), pp. 1060–1070.

[393] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022. URL: https://www.R-project.org/.

[394] J Allaire. “RStudio: integrated development environment for R”. In: *Boston, MA* 770.394 (2012), pp. 165–171.

[395] Till R. Lesker et al. “An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome”. In: *Cell Reports* 30.9 (Mar. 3, 2020), 2909–2922.e6. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2020.02.036. URL: https://www.sciencedirect.com/science/article/pii/S2211124720301972 (visited on 05/29/2022).

[396] Liang Xiao et al. “A Catalog of the Mouse Gut Metagenome”. In: *Nature Biotechnology* 33.10 (Oct. 2015), pp. 1103–1108. ISSN: 1546-1696. DOI: 10.1038/nbt.3353. URL: https://www.nature.com/articles/nbt.3353 (visited on 12/19/2019).

[397] Hiroshi Mori et al. “VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing”. In: *BMC Systems Biology* 12.2 (2018), pp. 47–58.

[398] Denise M O’Sullivan et al. “An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities”. In: *Scientific Reports* 11.1 (2021), p. 10590.

[399] Volkan Sevim et al. “Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies”. In: *Scientific data* 6.1 (2019), p. 285.

[400] Wenyi Xu et al. “Characterization of shallow whole-metagenome shotgun sequencing as a high-accuracy and low-cost method by complicated mock microbiomes”. In: *Frontiers in Microbiology* 12 (2021), p. 678319.

[401] Dieter M Tourlousse et al. “Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements”. In: *Microbiology spectrum* 10.2 (2022), e01915–21.

[402] Yu Hu et al. “Implications of error-prone long-read whole-genome shotgun sequencing on characterizing reference microbiomes”. In: *IScience* 23.6 (2020), p. 101223.

[403] Kevin D Kohl et al. “Gut Microbial Ecology of Five Species of Sympatric Desert Rodents in Relation to Herbivorous and Insectivorous Feeding Strategies”. In: *Integrative and Comparative Biology* 62.2 (Aug. 1, 2022), pp. 237–251. ISSN: 1540-7063. DOI: 10.1093/icb/icac045. URL: https://doi.org/10.1093/icb/icac045 (visited on 09/04/2022).

[404] Shifu Chen et al. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor”. In: *Bioinformatics* 34.17 (Sept. 1, 2018), pp. i884–i890. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty560. URL: https://doi.org/10.1093/bioinformatics/bty560 (visited on 06/03/2021).

[405] Brian Bushnell. *BBMap: A Fast, Accurate, Splice-Aware Aligner.* United States: Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), Mar. 17, 2014. URL: https://www.osti.gov/servlets/purl/1241166.

[406] Heng Li. *Seqtk*. Version 1.0. Available on Github, July 8, 2013. URL: https://github.com/lh3/seqtk/archive/refs/tags/1.0.tar.gz.

[407] Derrick E. Wood, Jennifer Lu, and Ben Langmead. “Improved Metagenomic Analysis with Kraken 2”. In: *Genome Biology* 20.1 (Nov. 28, 2019), p. 257. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1891-0. URL: https://doi.org/10.1186/s13059-019-1891-0 (visited on 06/19/2020).

[408] Hans-Joachim Ruscheweyh et al. “mOTUs: Profiling Taxonomic Composition, Transcriptional Activity and Strain Populations of Microbial Communities”. In: *Current Protocols* 1.8 (2021), e218. ISSN: 2691-1299. DOI: 10.1002/cpz1.218. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpz1.218 (visited on 03/19/2023).

[409] Marcel van de Wouw et al. “Distinct Actions of the Fermented Beverage Kefir on Host Behaviour, Immunity and Microbiome Gut-Brain Modules in the Mouse”. In: *Microbiome* 8.1 (May 18, 2020), p. 67. ISSN: 2049-2618. DOI: 10.1186/s40168-020-00846-5. URL: https://doi.org/10.1186/s40168-020-00846-5 (visited on 05/29/2020).

[410] Sara B Weinstein et al. “Wild herbivorous mammals (genus Neotoma) host a diverse but transient assemblage of fungi”. In: *Symbiosis* 87.1 (2022), pp. 45–58.

[411] Lucelia Cabral et al. “Gut Microbiome of Capybara, the Amazon Master of the Grasses, Harbors Unprecedented Enzymatic Strategies for Plant Glycans Breakdown”. In: (Mar. 3, 2022). ISSN: 2693-5015. DOI: 10.21203/rs.3.rs-456076/v1. URL: https://www.researchsquare.com/article/rs-456076/v1 (visited on 03/03/2022).

[412] Tao Zhang et al. “Stronger gut microbiome modulatory effects by postbiotics than probiotics in a mouse colitis model”. In: *npj Science of Food* 6.1 (2022), p. 53.

[413] Sebastien-Raguideau. *Metahood*. Nov. 9, 2022. URL: https://github.com/Sebastien-Raguideau/Metahood (visited on 12/13/2022).

[414] S. Andrews. *FastQC: A Quality Control Tool for High Throughput Sequence Data - Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*. 2010.

[415] Felix Krueger et al. *FelixKrueger/TrimGalore: V0.6.7 - DOI via Zenodo*. Version 0.6.7. Zenodo, July 2021. DOI: 10.5281/zenodo.5127899. URL: https://doi.org/10.5281/zenodo.5127899.

[416] Dongwan D. Kang et al. “MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies”. In: *PeerJ* 7 (July 26, 2019), e7359. ISSN: 2167-8359. DOI: 10.7717/peerj.7359. URL: https://peerj.com/articles/7359 (visited on 11/18/2022).

[417] Donovan H. Parks et al. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes”. In: *Genome Research* 25.7 (Jan. 7, 2015), pp. 1043–1055. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.186072.114. pmid: 25977477. URL: https://genome.cshlp.org/content/25/7/1043 (visited on 10/17/2022).

[418] Chirag Jain et al. “High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries”. In: *Nature Communications* 9.1 (1 Nov. 30, 2018), p. 5114. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07641-9. URL: https://www.nature.com/articles/s41467-018-07641-9 (visited on 11/10/2022).

[419] Matthew R. Olm et al. “dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication”. In: *The ISME Journal* 11 (2017), pp. 2864–2868. DOI: 10.1038/ismej.2017.126.

[420] Pierre-Alain Chaumeil et al. “GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database”. In: *Bioinformatics* 36.6 (Mar. 1, 2020), pp. 1925–1927. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz848. URL: https://doi.org/10.1093/bioinformatics/btz848 (visited on 10/17/2022).

[421] T. Seemann. *Barrnap 0.9 : Rapid Ribosomal RNA Prediction*. Version 0.9. URL: https://github.com/tseemann/barrnap.

[422] Torsten Seemann. “Prokka: rapid prokaryotic genome annotation”. In: *Bioinformatics* 30.14 (Mar. 2014), pp. 2068–2069. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu153. eprint: https://academic.oup.com/bioinformatics/article-pdf/30/14/2068/48924770/bioinformatics_30_14_2068.pdf. URL: https://doi.org/10.1093/bioinformatics/btu153.

[423] Heng Li. “Minimap2: Pairwise Alignment for Nucleotide Sequences”. In: *Bioinformatics* 34.18 (Sept. 15, 2018), pp. 3094–3100. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty191. URL: https://doi.org/10.1093/bioinformatics/bty191 (visited on 10/19/2022).

[424] Petr Danecek et al. “Twelve Years of SAMtools and BCFtools”. In: *GigaScience* 10.2 (Feb. 1, 2021), giab008. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. URL: https://doi.org/10.1093/gigascience/giab008 (visited on 10/19/2022).

[425] Rob Patro et al. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression”. In: *Nature Methods* 14.4 (4 Apr. 2017), pp. 417–419. ISSN: 1548-7105. DOI: 10.1038/nmeth.4197. URL: https://www.nature.com/articles/nmeth.4197 (visited on 10/19/2022).

[426] Frank Wilcoxon. “Individual Comparisons by Ranking Methods.” In: *Biometrics Bulletin* 1.6 (1945), pp. 80–83. URL: https://doi.org/10.2307/3001968.

[427] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (1995), pp. 289–300. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1995.tb02031.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x (visited on 10/19/2022).

[428] William H. Kruskal and W. Allen Wallis. “Use of Ranks in One-Criterion Variance Analysis”. In: *Journal of the American Statistical Association* 47.260 (Dec. 1, 1952), pp. 583–621. ISSN: 0162-1459. DOI: 10.1080/01621459.1952.10483441. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441 (visited on 10/19/2022).

[429] Magdalena Kujawska et al. “Bifidobacterium Castoris Strains Isolated from Wild Mice Show Evidence of Frequent Host Switching and Diverse Carbohydrate Metabolism Potential”. In: *ISME Communications* 2.1 (1 Feb. 25, 2022), pp. 1–14. ISSN: 2730-6151. DOI: 10.1038/s43705-022-00102-x. URL: https://www.nature.com/articles/s43705-022-00102-x (visited on 11/22/2022).

[430] Anton Bankevich et al. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing”. In: *Journal of Computational Biology* 19.5 (May 2012), pp. 455–477. DOI: 10.1089/cmb.2012.0021. URL: https://www.liebertpub.com/doi/10.1089/cmb.2012.0021 (visited on 11/23/2022).

[431] Alexis Criscuolo. “A Fast Alignment-Free Bioinformatics Procedure to Infer Accurate Distance-Based Phylogenetic Trees from Genome Assemblies”. In: *Research Ideas and Outcomes* 5 (June 10, 2019), e36178. ISSN: 2367-7163. DOI: 10.3897/rio.5.e36178. URL: https://riojournal.com/article/36178/ (visited on 11/08/2022).

[432] Joel Armstrong et al. “Progressive Cactus Is a Multiple-Genome Aligner for the Thousand-Genome Era”. In: *Nature* 587.7833 (7833 Nov. 2020), pp. 246–251. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2871-y. URL: https://www.nature.com/articles/s41586-020-2871-y (visited on 12/07/2022).

[433] Ilia Minkin and Paul Medvedev. “Scalable multiple whole-genome alignment and locally collinear block construction with SibeliaZ”. In: *Nature communications* 11.1 (2020), p. 6327.

[434] Morgan N Price, Paramvir S Dehal, and Adam P Arkin. “FastTree 2–approximately maximum-likelihood trees for large alignments”. In: *PloS one* 5.3 (2010), e9490.

[435] Ivica Letunic and Peer Bork. “Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation”. In: *Nucleic Acids Research* 49.W1 (July 2, 2021), W293–W296. ISSN: 0305-1048. DOI: 10.1093/nar/gkab301. URL: https://doi.org/10.1093/nar/gkab301 (visited on 11/07/2022).

[436] Warapond Wanna et al. “Evaluation of Probiotic Characteristics and Whole Genome Analysis of Pediococcus Pentosaceus MR001 for Use as Probiotic Bacteria in Shrimp Aquaculture”. In: *Scientific Reports* 11.1 (1 Sept. 15, 2021), p. 18334. ISSN: 2045-2322. DOI: 10.1038/s41598-021-96780-z. URL: https://www.nature.com/articles/s41598-021-96780-z (visited on 05/30/2022).

[437] Fredy Morales et al. “Isolation and Partial Characterization of Halotolerant Lactic Acid Bacteria from Two Mexican Cheeses”. In: *Applied Biochemistry and Biotechnology* 164.6 (July 1, 2011), pp. 889–905. ISSN: 1559-0291. DOI: 10.1007/s12010-011-9182-6. URL: https://doi.org/10.1007/s12010-011-9182-6 (visited on 11/25/2022).

[438] Alexander A. Bachmanov, Gary K. Beauchamp, and Michael G. Tordoff. “Voluntary Consumption of NaCl, KCl, CaCl2, and NH4Cl Solutions by 28 Mouse Strains”. In: *Behavior Genetics* 32.6 (Nov. 1, 2002), pp. 445–457. ISSN: 1573-3297. DOI: 10.1023/A:1020832327983. URL: https://doi.org/10.1023/A:1020832327983 (visited on 11/25/2022).

[439] Matthew R Olm et al. “Consistent metagenome-derived metrics verify and delineate bacterial species boundaries”. In: *Msystems* 5.1 (2020), e00731–19.

[440] Eric W Sayers et al. “Database Resources of the National Center for Biotechnology Information”. In: *Nucleic Acids Research* 50.D1 (Jan. 7, 2022), pp. D20–D26. ISSN: 0305-1048. DOI: 10.1093/nar/gkab1112. URL: https://doi.org/10.1093/nar/gkab1112 (visited on 11/27/2022).

[441] aitor.blancomiguez. *How Is UNKNOWN Calculated?* Dec. 6, 2021. URL: https://forum.biobakery.org/t/how-is-unknown-calculated/558/15.

[442] Jian Sun et al. “Environmental Remodeling of Human Gut Microbiota and Antibiotic Resistome in Livestock Farms”. In: *Nature Communications* 11.1 (1 Mar. 18, 2020), p. 1427. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15222-y. URL: https://www.nature.com/articles/s41467-020-15222-y (visited on 06/02/2022).

[443] Lianmin Chen et al. “Genetic and Microbial Associations to Plasma and Fecal Bile Acids in Obesity Relate to Plasma Lipids and Liver Fat Content”. In: *Cell Reports* 33.1 (Oct. 6, 2020), p. 108212. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2020.108212. URL: https://www.sciencedirect.com/science/article/pii/S2211124720312018 (visited on 06/02/2022).

[444] Yun Kit Yeoh et al. “Gut Microbiota Composition Reflects Disease Severity and Dysfunctional Immune Responses in Patients with COVID-19”. In: *Gut* 70.4 (Apr. 1, 2021), pp. 698–706. ISSN: 0017-5749, 1468-3288. DOI: 10.1136/gutjnl-2020-323020. pmid: 33431578. URL: https://gut.bmj.com/content/70/4/698 (visited on 06/02/2022).

[445] Caroline Isabel Kothe, Nacer Mohellibi, and Pierre Renault. “Revealing the Microbial Heritage of Traditional Brazilian Cheeses through Metagenomics”. In: *Food Research International* 157 (July 1, 2022), p. 111265. ISSN: 0963-9969. DOI: 10.1016/j.foodres.2022.111265. URL: https://www.sciencedirect.com/science/article/pii/S0963996922003222 (visited on 06/02/2022).

[446] Xiaochen Wang et al. “Captivity influences the gut microbiome of Rhinopithecus roxellana”. In: *Frontiers in Microbiology* 12 (2021), p. 3841.

[447] Naoyoshi Nagata et al. “Population-level metagenomics uncovers distinct effects of multiple medications on the human gut microbiome”. In: *Gastroenterology* 163.4 (2022), pp. 1038–1052.

[448] Alexandre Almeida et al. “A unified catalog of 204,938 reference genomes from the human gut microbiome”. In: *Nature biotechnology* 39.1 (2021), pp. 105–114.

[449] Margaret MC Lam et al. “A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex”. In: *Nature communications* 12.1 (2021), p. 4188.

[450] Marcela França Dias et al. “Exploring the resistome, virulome and microbiome of drinking water in environmental and clinical settings”. In: *Water Research* 174 (2020), p. 115630.

[451] Reiner Jumpertz von Schwartzenberg et al. “Caloric Restriction Disrupts the Microbiota and Colonization Resistance”. In: *Nature* 595.7866 (7866 July 2021), pp. 272–277. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03663-4. URL: https://www.nature.com/articles/s41586-021-03663-4 (visited on 06/03/2022).

[452] Steven J. Biller et al. “Marine Microbial Metagenomes Sampled across Space and Time”. In: *Scientific Data* 5.1 (1 Sept. 4, 2018), p. 180176. ISSN: 2052-4463. DOI: 10.1038/sdata.2018.176. URL: https://www.nature.com/articles/sdata2018176 (visited on 06/03/2022).

[453] Sully Márquez et al. “Metagenome of a Bronchoalveolar Lavage Fluid Sample from a Confirmed COVID-19 Case in Quito, Ecuador, Obtained Using Oxford Nanopore MinION Technology”. In: *Microbiology Resource Announcements* 9.41 (2020), e00996–20. DOI: 10.1128/MRA.00996-20. URL: https://journals.asm.org/doi/full/10.1128/MRA.00996-20.

[454] Nusrat A Jahan et al. “Nanopore-based surveillance of zoonotic bacterial pathogens in farm-dwelling Peridomestic rodents”. In: *pathogens* 10.9 (2021), p. 1183.

[455] Kah-Ooi Chua et al. “Bacterial Microbiome of Faecal Samples of Naked Mole-Rat Collected from the Toilet Chamber”. In: *BMC Research Notes* 15.1 (Mar. 18, 2022), p. 107. ISSN: 1756-0500. DOI: 10.1186/s13104-022-06000-8. URL: https://doi.org/10.1186/s13104-022-06000-8 (visited on 05/29/2022).

[456] Dongyao Li et al. “Microbial Biogeography and Core Microbiota of the Rat Digestive Tract”. In: *Scientific Reports* 7.1 (1 Apr. 4, 2017), p. 45840. ISSN: 2045-2322. DOI: 10.1038/srep45840. URL: https://www.nature.com/articles/srep45840 (visited on 11/13/2022).

[457] M. Mailhe et al. “‘Ileibacterium Massiliense’ Gen. Nov., Sp. Nov., a New Bacterial Species Isolated from Human Ileum of a Patient with Crohn Disease”. In: *New Microbes and New Infections* 17 (May 1, 2017), pp. 25–26. ISSN: 2052-2975. DOI: 10.1016/j.nmni.2016.11.022. URL: https://www.sciencedirect.com/science/article/pii/S2052297516301342 (visited on 12/28/2022).

[458] Francesca Turroni et al. “Human Gut Microbiota and Bifidobacteria: From Composition to Functionality”. In: *Antonie van Leeuwenhoek* 94.1 (June 1, 2008), pp. 35–50. ISSN: 1572-9699. DOI: 10.1007/s10482-008-9232-4. URL: https://doi.org/10.1007/s10482-008-9232-4 (visited on 12/28/2022).

[459] Kirsty J. Marsh et al. “Synchronous Seasonality in the Gut Microbiota of Wild Mouse Populations”. In: *Frontiers in Microbiology* 13 (2022). ISSN: 1664-302X. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2022.809735 (visited on 11/13/2022).

[460] Rodrigo Mendes and Jos M. Raaijmakers. “Cross-Kingdom Similarities in Microbiome Functions”. In: *The ISME Journal* 9.9 (9 Sept. 2015), pp. 1905–1907. ISSN: 1751-7370. DOI: 10.1038/ismej.2015.7. URL: https://www.nature.com/articles/ismej20157 (visited on 11/14/2022).

[461] Elizabeth L. Johnson et al. “Microbiome and Metabolic Disease: Revisiting the Bacterial Phylum Bacteroidetes”. In: *Journal of Molecular Medicine* 95.1 (Jan. 1, 2017), pp. 1–8. ISSN: 1432-1440. DOI: 10.1007/s00109-016-1492-2. URL: https://doi.org/10.1007/s00109-016-1492-2 (visited on 12/23/2022).

[462] Conor J. Meehan and Robert G. Beiko. “A Phylogenomic View of Ecological Specialization in the Lachnospiraceae, a Family of Digestive Tract-Associated Bacteria”. In: *Genome Biology and Evolution* 6.3 (Mar. 1, 2014), pp. 703–713. ISSN: 1759-6653. DOI: 10.1093/gbe/evu050. URL: https://doi.org/10.1093/gbe/evu050 (visited on 11/14/2022).

[463] Saisai Zhou et al. “Characterization of Metagenome-Assembled Genomes and Carbohydrate-Degrading Genes in the Gut Microbiota of Tibetan Pig”. In: *Frontiers in Microbiology* 11 (2020). ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.595066. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2020.595066/full (visited on 12/28/2020).

[464] Jesse C. Thomas et al. “Unveiling the Gut Microbiota and Resistome of Wild Cotton Mice, Peromyscus Gossypinus, from Heavy Metal- and Radionuclide-Contaminated Sites in the Southeastern United States”. In: *Microbiology Spectrum* 9.1 (Aug. 25, 2021), e00097–21. DOI: 10.1128/Spectrum.00097-21. URL: https://journals.asm.org/doi/full/10.1128/Spectrum.00097-21 (visited on 11/18/2022).

[465] Angelica P. Ahrens et al. “A Six-Day, Lifestyle-Based Immersion Program Mitigates Cardiovascular Risk Factors and Induces Shifts in Gut Microbiota, Specifically Lachnospiraceae, Ruminococcaceae, Faecalibacterium Prausnitzii: A Pilot Study”. In: *Nutrients* 13.10 (10 Oct. 2021), p. 3459. ISSN: 2072-6643. DOI: 10.3390/nu13103459. URL: https://www.mdpi.com/2072-6643/13/10/3459 (visited on 11/14/2022).

[466] Mirco Vacca et al. “The Controversial Role of Human Gut Lachnospiraceae”. In: *Microorganisms* 8.4 (4 Apr. 2020), p. 573. ISSN: 2076-2607. DOI: 10.3390/microorganisms8040573. URL: https://www.mdpi.com/2076-2607/8/4/573 (visited on 11/14/2022).

[467] Elliott Schmidt, Nadia Mykytczuk, and Albrecht I. Schulte-Hostedde. “Effects of the Captive and Wild Environment on Diversity of the Gut Microbiome of Deer Mice (Peromyscus Maniculatus)”. In: *The ISME Journal* 13.5 (5 May 2019), pp. 1293–1305. ISSN: 1751-7370. DOI: 10.1038/s41396-019-0345-8. URL: https://www.nature.com/articles/s41396-019-0345-8 (visited on 08/19/2020).

[468] Limei Lin et al. “Genome-centric investigation of bile acid metabolizing microbiota of dairy cows and associated diet-induced functional implications”. In: *The ISME Journal* 17.1 (2023), pp. 172–184.

[469] Rosemarie De Weirdt and Tom Van de Wiele. “Micromanagement in the gut: microenvironmental factors govern colon mucosal biofilm structure and functionality”. In: *npj Biofilms and Microbiomes* 1.1 (2015), pp. 1–6.

[470] Xiaoqiong Gu et al. “Gut Ruminococcaceae levels at baseline correlate with risk of antibiotic-associated diarrhea”. In: *Iscience* 25.1 (2022), p. 103644.

[471] Kate L Ormerod et al. “Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals”. In: *Microbiome* 4 (2016), pp. 1–17.

[472] Byron J Smith, Richard A Miller, and Thomas M Schmidt. “Muribaculaceae genomes assembled from metagenomes suggest genetic drivers of differential response to acarbose treatment in mice”. In: *Msphere* 6.6 (2021), e00851–21.

[473] Ning Zhu et al. “Metagenomic and Metaproteomic Analyses of a Corn Stover-Adapted Microbial Consortium EMSD5 Reveal Its Taxonomic and Enzymatic Basis for Degrading Lignocellulose”. In: *Biotechnology for Biofuels* 9.1 (Nov. 9, 2016), p. 243. ISSN: 1754-6834. DOI: 10.1186/s13068-016-0658-z. URL: https://doi.org/10.1186/s13068-016-0658-z (visited on 11/16/2022).

[474] Tiantian Ren et al. “Seasonal, Spatial, and Maternal Effects on Gut Microbiome in Wild Red Squirrels”. In: *Microbiome* 5.1 (Dec. 21, 2017), p. 163. ISSN: 2049-2618. DOI: 10.1186/s40168-017-0382-3. URL: https://doi.org/10.1186/s40168-017-0382-3 (visited on 11/28/2019).

[475] N. T. Baxter et al. “Intra- and Interindividual Variations Mask Interspecies Variation in the Microbiota of Sympatric Peromyscus Populations”. In: *Applied and Environmental Microbiology* 81.1 (Jan. 2015), pp. 396–404. ISSN: 0099-2240. DOI: 10.1128/aem.02303-14.

[476] J. D. Orkin et al. “Seasonality of the Gut Microbiota of Free-Ranging White-Faced Capuchins in a Tropical Dry Forest”. In: *Isme Journal* 13.1 (Jan. 2019), pp. 183–196. ISSN: 1751-7362. DOI: 10.1038/541396-018-0256-0.

[477] Philippe Gérard. “Gut Microbiota and Obesity”. In: *Cellular and Molecular Life Sciences* 73.1 (Jan. 1, 2016), pp. 147–162. ISSN: 1420-9071. DOI: 10.1007/s00018-015-2061-5. URL: https://doi.org/10.1007/s00018-015-2061-5 (visited on 11/16/2022).

[478] Xiaolong Hu et al. “High-Throughput Analysis Reveals Seasonal Variation of the Gut Microbiota Composition Within Forest Musk Deer (Moschus Berezovskii)”. In: *Frontiers in Microbiology* 9 (July 26, 2018), p. 1674. ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.01674.

[479] Allison L. Hicks et al. “Gut Microbiomes of Wild Great Apes Fluctuate Seasonally in Response to Diet”. In: *Nature Communications* 9.1 (1 May 3, 2018), p. 1786. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04204-w. URL: https://www.nature.com/articles/s41467-018-04204-w (visited on 11/17/2022).

[480] Ilias Lagkouvardos et al. “Sequence and Cultivation Study of Muribaculaceae Reveals Novel Species, Host Preference, and Functional Potential of This yet Undescribed Family”. In: *Microbiome* 7.1 (Feb. 19, 2019), p. 28. ISSN: 2049-2618. DOI: 10.1186/s40168-019-0637-2. URL: https://doi.org/10.1186/s40168-019-0637-2 (visited on 05/31/2022).

[481] Pauline van Leeuwen et al. “Effects of Captivity, Diet, and Relocation on the Gut Bacterial Communities of White-Footed Mice”. In: *Ecology and Evolution* 10.11 (2020), pp. 4677–4690. ISSN: 2045-7758. DOI: 10.1002/ece3.6221. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6221 (visited on 08/19/2020).

[482] E. Shargal, N. Kronfeld-Schor, and T. Dayan. “Population Biology and Spatial Relationships of Coexisting Spiny Mice (Acomys) in Israel”. In: *Journal of Mammalogy* 81.4 (Nov. 2000), pp. 1046–1052. ISSN: 0022-2372. DOI: 10.1644/1545-1542(2000)081<1046:pbasro>2.0.co;2.

[483] Aura Raulo et al. “Social Networks Strongly Predict the Gut Microbiota of Wild Mice”. In: *The ISME Journal* 15.9 (9 Sept. 2021), pp. 2601–2613. ISSN: 1751-7370. DOI: 10.1038/s41396-021-00949-3. URL: https://www.nature.com/articles/s41396-021-00949-3 (visited on 11/17/2021).

[484] Pamela Ferretti et al. “Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome”. In: *Cell Host & Microbe* 24.1 (July 11, 2018), 133–145.e5. ISSN: 1931-3128. DOI: 10.1016/j.chom.2018.06.005. URL: https://www.sciencedirect.com/science/article/pii/S1931312818303172 (visited on 11/17/2022).

[485] Samuel C. Forster et al. “Novel Gut Pathobionts Confound Results in a Widely Used Mouse Model of Human Inflammatory Disease”. In: *bioRxiv* (Feb. 9, 2021), p. 2021.02.09.430393. DOI: 10.1101/2021.02.09.430393. URL: https://www.biorxiv.org/content/10.1101/2021.02.09.430393v1 (visited on 03/23/2021).

[486] Amber L Hartman et al. “The complete genome sequence of Haloferax volcanii DS2, a model archaeon”. In: *PloS one* 5.3 (2010), e9605.

[487] Josef Deutscher, Christof Francke, and Pieter W Postma. “How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria”. In: *Microbiology and molecular biology Reviews* 70.4 (2006), pp. 939–1031.

[488] Gwendolyn J Gregory and E Fidelma Boyd. “Stressed out: Bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from Vibrionaceae”. In: *Computational and structural biotechnology journal* 19 (2021), pp. 1014–1027.

[489] Laura Teichmann et al. “From substrate specificity to promiscuity: hybrid ABC transporters for osmoprotectants”. In: *Molecular Microbiology* 104.5 (2017), pp. 761–780.

[490] Matthias Kurz, Anika NS Brünig, and Erwin A Galinski. “NhaD type sodium/proton-antiporter of Halomonas elongata: a salt stress response mechanism in marine habitats?” In: *Saline Systems* 2.1 (2006), pp. 1–12.

[491] P ROY VAGELOS. “Regulation of fatty acid biosynthesis”. In: *Current topics in cellular regulation*. Vol. 4. Elsevier, 1971, pp. 119–166.

[492] Weronika Ratajczak et al. “Immunomodulatory Potential of Gut Microbiome-Derived Short-Chain Fatty Acids (SCFAs)”. In: *Acta Biochimica Polonica* 66.1 (1 Mar. 4, 2019), pp. 1–12. ISSN: 1734-154X. DOI: 10.18388/abp.2018_2648. URL: https://ojs.ptbioch.edu.pl/index.php/abp/article/view/2648 (visited on 12/26/2022).

[493] Ratna Karan, Sneh L Singla-Pareek, and Ashwani Pareek. “Histidine kinase and response regulator genes as they relate to salinity tolerance in rice”. In: *Functional & integrative genomics* 9 (2009), pp. 411–417.

[494] Jocelyn M. Choo, Lex EX Leong, and Geraint B. Rogers. “Sample Storage Conditions Significantly Influence Faecal Microbiome Profiles”. In: *Scientific Reports* 5.1 (1 Nov. 17, 2015), p. 16350. ISSN: 2045-2322. DOI: 10.1038/srep16350. URL: https://www.nature.com/articles/srep16350 (visited on 12/01/2022).

[495] Alena L. Pribyl et al. “Critical Evaluation of Faecal Microbiome Preservation Using Metagenomic Analysis”. In: *ISME Communications* 1.1 (1 May 5, 2021), pp. 1–10. ISSN: 2730-6151. DOI: 10.1038/s43705-021-00014-2. URL: https://www.nature.com/articles/s43705-021-00014-2 (visited on 12/01/2022).

[496] Maria E. Hedberg et al. “Lachnoanaerobaculum Gen. Nov., a New Genus in the Lachnospiraceae: Characterization of Lachnoanaerobaculum Umeaense Gen. Nov., Sp. Nov., Isolated from the Human Small Intestine, and Lachnoanaerobaculum Orale Sp. Nov., Isolated from Saliva, and Reclassification of Eubacterium Saburreum (Prévot 1966) Holdeman and Moore 1970 as Lachnoanaerobaculum Saburreum Comb. Nov.” In: *International Journal of Systematic and Evolutionary Microbiology* 62 (Pt_11), pp. 2685–2690. ISSN: 1466-5034, DOI: 10.1099/ijs.0.033613-0. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.033613-0 (visited on 12/01/2022).

[497] Matthew T. Sorbara et al. “Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity”. In: *Cell Host & Microbe* 28.1 (July 8, 2020), 134–146.e4. ISSN: 1931-3128. DOI: 10.1016/j.chom.2020.05.005. URL: https://www.sciencedirect.com/science/article/pii/S1931312820302870 (visited on 12/01/2022).

[498] Alba Cortés et al. “The gut microbial metabolic capacity of microbiome-humanized vs. wild type rodents reveals a likely dual role of intestinal bacteria in hepato-intestinal schistosomiasis”. In: *PLoS neglected tropical diseases* 16.10 (2022), e0010878.

[499] Rafig Gurbanov, Uygar Kabaoğlu, and Tuba Yağcı. “Metagenomic analysis of intestinal microbiota in wild rats living in urban and rural habitats”. In: *Folia Microbiologica* 67.3 (2022), pp. 469–477.

[500] Xiaoying Yang et al. “Seasonal changes in the distinct taxonomy and function of the gut microbiota in the wild ground squirrel (Spermophilus dauricus)”. In: *Animals* 11.9 (2021), p. 2685.

[501] Akio Shinohara et al. “Comparison of the gut microbiotas of laboratory and wild Asian house shrews (Suncus murinus) based on cloned 16S rRNA sequences”. In: *Experimental animals* 68.4 (2019), pp. 531–539.

[502] Alice Baniel et al. “Seasonal shifts in the gut microbiome indicate plastic responses to diet in wild geladas”. In: *Microbiome* 9.1 (2021), pp. 1–20.

[503] Meredith K Tavenner, Sue M McDonnell, and Amy S Biddle. “Development of the equine hindgut microbiome in semi-feral and domestic conventionally-managed foals”. In: *Animal Microbiome* 2 (2020), pp. 1–17.

[504] K. Makarova et al. “Comparative Genomics of the Lactic Acid Bacteria”. In: *Proceedings of the National Academy of Sciences* 103.42 (Oct. 17, 2006), pp. 15611–15616. DOI: 10.1073/pnas.0607117103. URL: https://www.pnas.org/doi/full/10.1073/pnas.0607117103 (visited on 12/01/2022).

[505] Matthew R. Olm et al. “Identical Bacterial Populations Colonize Premature Infant Gut, Skin, and Oral Microbiomes and Exhibit Different in Situ Growth Rates”. In: *Genome Research* 27.4 (Jan. 4, 2017), pp. 601–612. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.213256.116. pmid: 28073918. URL: https://genome.cshlp.org/content/27/4/601 (visited on 12/01/2022).

[506] Mustafa Özçam et al. “A Secondary Metabolite Drives Intraspecies Antagonism in a Gut Symbiont That Is Inhibited by Cell-Wall Acetylation”. In: *Cell Host & Microbe* 30.6 (June 8, 2022), 824–835.e6. ISSN: 1931-3128. DOI: 10.1016/j.chom.2022.03.033. URL: https://www.sciencedirect.com/science/article/pii/S1931312822001639 (visited on 12/02/2022).

[507] Patricia Diez-Echave et al. “Probiotic and Functional Properties of Limosilactobacillus Reuteri INIA P572”. In: *Nutrients* 13.6 (6 June 2021), p. 1860. ISSN: 2072-6643. DOI: 10.3390/nu13061860. URL: https://www.mdpi.com/2072-6643/13/6/1860 (visited on 12/02/2022).

[508] Axel Kornerup Hansen et al. “Bacterial Species to Be Considered in Quality Assurance of Mice and Rats”. In: *Laboratory Animals* 53.3 (June 1, 2019), pp. 281–291. ISSN: 0023-6772. DOI: 10.1177/0023677219834324. URL: https://doi.org/10.1177/0023677219834324 (visited on 12/02/2022).

[509] Jinshui Zheng 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”. In: *International Journal of Systematic and Evolutionary Microbiology* 70.4 (), pp. 2782–2858. ISSN: 1466-5034, DOI: 10.1099/ijsem.0.004107. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.004107 (visited on 12/02/2022).

[510] J J Leisner et al. “Description of Paralactobacillus Selangorensis Gen. Nov., Sp. Nov., a New Lactic Acid Bacterium Isolated from Chili Bo, a Malaysian Food Ingredient.” In: *International Journal of Systematic and Evolutionary Microbiology* 50.1 (), pp. 19–24. ISSN: 1466-5034, DOI: 10.1099/00207713-50-1-19. URL: https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-50-1-19 (visited on 12/02/2022).

[511] B.c. Silva et al. “Selection of a Candidate Probiotic Strain of Pediococcus Pentosaceus from the Faecal Microbiota of Horses by in Vitro Testing and Health Claims in a Mouse Model of Salmonella Infection”. In: *Journal of Applied Microbiology* 122.1 (2017), pp. 225–238. ISSN: 1365-2672. DOI: 10.1111/jam.13339. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jam.13339 (visited on 12/02/2022).

[512] Xiaoyuan Bian et al. “Pediococcus Pentosaceus LI05 Alleviates DSS-induced Colitis by Modulating Immunological Profiles, the Gut Microbiota and Short-Chain Fatty Acid Levels in a Mouse Model”. In: *Microbial Biotechnology* 13.4 (2020), pp. 1228–1244. ISSN: 1751-7915. DOI: 10.1111/1751-7915.13583. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1751-7915.13583 (visited on 12/02/2022).

[513] Rundong Wang et al. “Modulation of Intestinal Barrier, Inflammatory Response, and Gut Microbiota by Pediococcus Pentosaceus Zy-B Alleviates Vibrio Parahaemolyticus Infection in C57BL/6J Mice”. In: *Journal of Agricultural and Food Chemistry* 70.6 (Feb. 16, 2022), pp. 1865–1877. ISSN: 0021-8561. DOI: 10.1021/acs.jafc.1c07450. URL: https://doi.org/10.1021/acs.jafc.1c07450 (visited on 12/02/2022).

[514] Emma J. E. Brownlie et al. “Acids Produced by Lactobacilli Inhibit the Growth of Commensal Lachnospiraceae and S24-7 Bacteria”. In: *Gut Microbes* 14.1 (Dec. 31, 2022), p. 2046452. ISSN: 1949-0976. DOI: 10.1080/19490976.2022.2046452. pmid: 35266847. URL: https://doi.org/10.1080/19490976.2022.2046452 (visited on 12/02/2022).

[515] Kellyanne Duncan, Kelly Carey-Ewend, and Shipra Vaishnava. “Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer”. In: *Gut Microbes* 13.1 (2021), p. 1874815.

[516] Hai Li et al. “The outer mucus layer hosts a distinct intestinal microbial niche”. In: *Nature communications* 6.1 (2015), p. 8292.

[517] Matthew L Jenior et al. “Clostridium difficile colonizes alternative nutrient niches during infection across distinct murine gut microbiomes”. In: *MSystems* 2.4 (2017), e00063–17.

[518] Niv Zmora et al. “Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features”. In: *Cell* 174.6 (2018), pp. 1388–1405.

[519] Albert Barberán et al. “Using network analysis to explore co-occurrence patterns in soil microbial communities”. In: *The ISME journal* 6.2 (2012), pp. 343–351.

[520] Michael Roggenbuck et al. “The Microbiome of New World Vultures”. In: *Nature Communications* 5.1 (1 Nov. 25, 2014), p. 5498. ISSN: 2041-1723. DOI: 10.1038/ncomms6498. URL: https://www.nature.com/articles/ncomms6498 (visited on 08/16/2020).

[521] Matthew L Jenior et al. “Clostridium difficile alters the structure and metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization”. In: *Msphere* 3.3 (2018), e00261–18.

[522] Joshua R Fletcher et al. “Clostridioides difficile exploits toxin-mediated inflammation to alter the host nutritional landscape and exclude competitors from the gut microbiota”. In: *Nature communications* 12.1 (2021), p. 462.

[523] Zhimin Zhang et al. “Microbiome of Co-cultured Fish Exhibits Host Selection and Niche Differentiation at the Organ Scale”. In: *Frontiers in Microbiology* 10 (2019). ISSN: 1664-302X. DOI: 10.3389/fmicb.2019.02576. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02576/full (visited on 12/23/2019).

[524] Simon Heilbronner et al. “The microbiome-shaping roles of bacteriocins”. In: *Nature Reviews Microbiology* 19.11 (2021), pp. 726–739.

[525] Lubov Petkova Nedialkova et al. “Inflammation fuels colicin Ib-dependent competition of Salmonella serovar Typhimurium and E. coli in enterobacterial blooms”. In: *PLoS pathogens* 10.1 (2014), e1003844.

[526] S Brook Peterson, Savannah K Bertolli, and Joseph D Mougous. “The central role of interbacterial antagonism in bacterial life”. In: *Current Biology* 30.19 (2020), R1203–R1214.

[527] Mary C Rea et al. “Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile”. In: *Proceedings of the National Academy of Sciences* 107.20 (2010), pp. 9352–9357.

[528] Adam Koziol et al. “Gut microbiomes of mammal species show differential responses to identical series of environmental stressors”. In: (2023).

[529] Markella Moraitou et al. “Ecology, not host phylogeny, shapes the oral microbiome in closely related species”. In: *Molecular Biology and Evolution* 39.12 (2022), msac263.

[530] Finlay Maguire et al. “Metagenome-Assembled Genome Binning Methods with Short Reads Disproportionately Fail for Plasmids and Genomic Islands”. In: *bioRxiv* (Apr. 28, 2020), p. 2020.03.31.997171. DOI: 10.1101/2020.03.31.997171. URL: https://www.biorxiv.org/content/10.1101/2020.03.31.997171v2 (visited on 07/01/2020).

[531] Zeli Shen et al. “Novel Helicobacter species H. japonicum isolated from laboratory mice from Japan induces typhlocolitis and lower bowel carcinoma in C57BL/129 IL10-/- mice”. In: *Carcinogenesis* 37.12 (2016), pp. 1190–1198.

[532] Ilias Lagkouvardos et al. “The Mouse Intestinal Bacterial Collection (miBC) Provides Host-Specific Insight into Cultured Diversity and Functional Potential of the Gut Microbiota”. In: *Nature Microbiology* 1.10 (10 Aug. 8, 2016), pp. 1–15. ISSN: 2058-5276. DOI: 10.1038/nmicrobiol.2016.131. URL: https://www.nature.com/articles/nmicrobiol2016131 (visited on 06/23/2020).

[533] M-C Domingo et al. “Ruminococcus gauvreauii sp. nov., a glycopeptide-resistant species isolated from a human faecal specimen”. In: *International journal of systematic and evolutionary microbiology* 58.6 (2008), pp. 1393–1397.

[534] Natalia Molinero et al. “Survival strategies and metabolic interactions between Ruminococcus gauvreauii and Ruminococcoides bili, isolated from human bile”. In: *Microbiology spectrum* 10.4 (2022), e02776–21.

[535] Takumi Toya et al. “Coronary artery disease is associated with an altered gut microbiome composition”. In: *PloS one* 15.1 (2020), e0227147.

[536] Pamela Vernocchi, Federica Del Chierico, and Lorenza Putignani. “Gut microbiota profiling: metabolomics based approach to unravel compounds affecting human health”. In: *Frontiers in microbiology* 7 (2016), p. 1144.

[537] Fahad Abdullah M Al-Dhabaan and Ali Hassan Bakhali. “Analysis of the bacterial strains using Biolog plates in the contaminated soil from Riyadh community”. In: *Saudi journal of biological sciences* 24.4 (2017), pp. 901–906.

[538] Tong Wang et al. “Pairing Metagenomics and Metaproteomics to Pinpoint Ecological Niches and Metabolic Essentiality of Microbial Communities”. In: *bioRxiv* (2022), pp. 2022–11.

[539] Maria Satti et al. “Comparative analysis of probiotic bacteria based on a new definition of core genome”. In: *Journal of bioinformatics and computational biology* 16.03 (2018), p. 1840012.

[540] Emmanuel F Mongodin et al. “The genome of Salinibacter ruber: convergence and gene exchange among hyperhalophilic bacteria and archaea”. In: *Proceedings of the National Academy of Sciences* 102.50 (2005), pp. 18147–18152.

[541] Laure Diancourt et al. “Multilocus sequence typing of Lactobacillus casei reveals a clonal population structure with low levels of homologous recombination”. In: *Applied and environmental microbiology* 73.20 (2007), pp. 6601–6611.

[542] Peter A Bron et al. “Identification of Lactobacillus plantarum genes that are induced in the gastrointestinal tract of mice”. In: *Journal of bacteriology* 186.17 (2004), pp. 5721–5729.

[543] Orla O’sullivan et al. “Comparative genomics of lactic acid bacteria reveals a niche-specific gene set”. In: *Bmc Microbiology* 9 (2009), pp. 1–9.

[544] Sybille Tachon, Johannes Bernhard Brandsma, and Mireille Yvon. “NoxE NADH oxidase and the electron transport chain are responsible for the ability of Lactococcus lactis to decrease the redox potential of milk”. In: *Applied and environmental microbiology* 76.5 (2010), pp. 1311–1319.

[545] Gerald W Tannock et al. “Ecological behavior of Lactobacillus reuteri 100-23 is affected by mutation of the luxS gene”. In: *Applied and Environmental Microbiology* 71.12 (2005), pp. 8419–8425.

[546] Mónica Perea Vélez et al. “Functional analysis of D-alanylation of lipoteichoic acid in the probiotic strain Lactobacillus rhamnosus GG”. In: *Applied and environmental microbiology* 73.11 (2007), pp. 3595–3604.

[547] Andrew Bryan, Nir Shapir, and Michael J Sadowsky. “Frequency and distribution of tetracycline resistance genes in genetically diverse, nonselected, and nonclinical Escherichia coli strains isolated from diverse human and animal sources”. In: *Applied and environmental microbiology* 70.4 (2004), pp. 2503–2507.

[548] Ngoc Tung Quach et al. “Phenotypic features and analysis of genes supporting probiotic action unravel underlying perspectives of Bacillus velezensis VTX9 as a potential feed additive for swine”. In: *Annals of Microbiology* 71.1 (2021), pp. 1–14.

[549] Graziano Caggianiello, Michiel Kleerebezem, and Giuseppe Spano. “Exopolysaccharides produced by lactic acid bacteria: from health-promoting benefits to stress tolerance mechanisms”. In: *Applied microbiology and biotechnology* 100 (2016), pp. 3877–3886.

[550] Ana Agustina Bengoa et al. “Exopolysaccharides From Lactobacillus Paracasei Isolated From Kefir as Potential Bioactive Compounds for Microbiota Modulation”. In: *Frontiers in Microbiology* 11 (2020). ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.583254. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2020.583254/full (visited on 11/30/2020).

[551] Diego Chambi et al. “Exopolysaccharides Production by Cultivating a Bacterial Isolate from the Hypersaline Environment of Salar de Uyuni (Bolivia) in Pretreatment Liquids of Steam-Exploded Quinoa Stalks and Enzymatic Hydrolysates of Curupaú Sawdust”. In: *Fermentation* 7.1 (1 Mar. 2021), p. 33. ISSN: 2311-5637. DOI: 10.3390/fermentation7010033. URL: https://www.mdpi.com/2311-5637/7/1/33 (visited on 04/28/2022).

[552] P Reimundo et al. “dltA gene mutation in the teichoic acids alanylation system of Lactococcus garvieae results in diminished proliferation in its natural host”. In: *Veterinary microbiology* 143.2-4 (2010), pp. 434–439.

[553] Bhanu Priya Gsnesh et al. “Crosstalk between Lactobacillus reuteri and mammalian intestinal epithelium”. In: *The FASEB Journal* 31 (2017), pp. 652–14.

[554] Bhanu Priya Ganesh et al. “Diacylglycerol kinase synthesized by commensal Lactobacillus reuteri diminishes protein kinase C phosphorylation and histamine-mediated signaling in the mammalian intestinal epithelium”. In: *Mucosal immunology* 11.2 (2018), pp. 380–393.

[555] Gemma Buron-Moles et al. “Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes”. In: *Applied microbiology and biotechnology* 103.7 (2019), pp. 3135–3152.

[556] Shiro Itoi et al. “Isolation of halotolerant Lactococcus lactis subsp. lactis from intestinal tract of coastal fish”. In: *International journal of food microbiology* 121.1 (2008), pp. 116–121.

[557] Takao Iino, Ken-ichiro Suzuki, and Shigeaki Harayama. “Lacticigenium naphtae gen. nov., sp. nov., a halotolerant and motile lactic acid bacterium isolated from crude oil”. In: *International journal of systematic and evolutionary microbiology* 59.4 (2009), pp. 775–780.

[558] Trinh Thi Phuong Thao et al. “Characterization halotolerant lactic acid bacteria Pediococcus pentosaceus HN10 and in vivo evaluation for bacterial pathogens inhibition”. In: *Chemical Engineering and Processing-Process Intensification* 168 (2021), p. 108576.

[559] Fredy Morales et al. “Isolation and partial characterization of halotolerant lactic acid bacteria from two Mexican cheeses”. In: *Applied biochemistry and biotechnology* 164 (2011), pp. 889–905.

[560] Yi-sheng Chen et al. “Leuconostoc litchii sp. nov., a novel lactic acid bacterium isolated from lychee”. In: *International Journal of Systematic and Evolutionary Microbiology* 70.3 (2020), pp. 1585–1590.

[561] Aurélie Baliarda et al. “Potential osmoprotectants for the lactic acid bacteria Pediococcus pentosaceus and Tetragenococcus halophila”. In: *International journal of food microbiology* 84.1 (2003), pp. 13–20.

[562] Jun J Sato et al. “Dietary niche partitioning between sympatric wood mouse species (Muridae: Apodemus) revealed by DNA meta-barcoding analysis”. In: *Journal of Mammalogy* 99.4 (2018), pp. 952–964.

[563] Aharon Oren et al. “Life at high salt concentrations”. In: *The prokaryotes* 3 (2006), pp. 263–282.

[564] E Peleg, A Tietz, and I Friedberg. “Effects of salts and ionophores on proline transport in a moderately halophilic halotolerant bacterium”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 596.1 (1980), pp. 118–128.

[565] Feng Shen et al. “Gut Microbiota Dysbiosis in Patients with Non-Alcoholic Fatty Liver Disease”. In: *Hepatobiliary & Pancreatic Diseases International* 16.4 (Aug. 15, 2017), pp. 375–381. ISSN: 1499-3872. DOI: 10.1016/S1499-3872(17)60019-5. URL: https://www.sciencedirect.com/science/article/pii/S1499387217600195 (visited on 12/27/2022).

[566] Shiyin Li et al. “Lachnospiraceae Shift in the Microbial Community of Mice Faecal Sample Effects on Water Immersion Restraint Stress”. In: *AMB Express* 7 (Apr. 17, 2017), p. 82. ISSN: 2191-0855. DOI: 10.1186/s13568-017-0383-4. pmid: 28417435. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5393979/ (visited on 12/27/2022).

[567] Keishi Kameyama and Kikuji Itoh. “Intestinal Colonization by a *Lachnospiraceae* Bacterium Contributes to the Development of Diabetes in Obese Mice”. In: *Microbes and Environments* advpub (2014), ME14054. DOI: 10.1264/jsme2.ME14054.

[568] Hilary P. Browne et al. “Culturing of ‘Unculturable’ Human Microbiota Reveals Novel Taxa and Extensive Sporulation”. In: *Nature* 533.7604 (7604 May 2016), pp. 543–546. ISSN: 1476-4687. DOI: 10.1038/nature17645. URL: https://www.nature.com/articles/nature17645 (visited on 04/03/2020).

[569] Zhaogang Dong et al. “The Effects of High-Salt Gastric Intake on the Composition of the Intestinal Microbiota in Wistar Rats”. In: *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research* 26 (June 6, 2020), e922160-1-e922160–11. ISSN: 1234-1010. DOI: 10.12659/MSM.922160. pmid: 32504527. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7297027/ (visited on 12/26/2022).

[570] Pedro M. Miranda et al. “High Salt Diet Exacerbates Colitis in Mice by Decreasing Lactobacillus Levels and Butyrate Production”. In: *Microbiome* 6.1 (Mar. 22, 2018), p. 57. ISSN: 2049-2618. DOI: 10.1186/s40168-018-0433-4. URL: https://doi.org/10.1186/s40168-018-0433-4 (visited on 12/26/2022).

[571] Zhirui Cao et al. “The gut virome: A new microbiome component in health and disease”. In: *EBioMedicine* 81 (2022), p. 104113.

[572] Seesandra V Rajagopala et al. “Metatranscriptomics to characterize respiratory virome, microbiome, and host response directly from clinical samples”. In: *Cell reports methods* 1.6 (2021), p. 100091.

[573] Melissa A Fernandes et al. “Enteric virome and bacterial microbiota in children with ulcerative colitis and Crohn’s disease”. In: *Journal of pediatric gastroenterology and nutrition* 68.1 (2019), p. 30.

[574] Efrem S Lim, David Wang, and Lori R Holtz. “The bacterial microbiome and virome milestones of infant development”. In: *Trends in microbiology* 24.10 (2016), pp. 801–810.

[575] Fen Zhang et al. “Longitudinal dynamics of gut bacteriome, mycobiome and virome after fecal microbiota transplantation in graft-versus-host disease”. In: *Nature communications* 12.1 (2021), p. 65.

[576] Tongling Shan et al. “Virome in the cloaca of wild and breeding birds revealed a diversity of significant viruses”. In: *Microbiome* 10.1 (2022), pp. 1–21.

[577] Wan-Ting He et al. “Virome characterization of game animals in China reveals a spectrum of emerging pathogens”. In: *Cell* 185.7 (2022), pp. 1117–1129.

[578] Matheus A. Duarte et al. “Faecal virome analysis of wild animals from Brazil”. In: *Viruses* 11.9 (2019), p. 803.

[579] Ina Smith and Lin-Fa Wang. “Bats and their virome: an important source of emerging viruses capable of infecting humans”. In: *Current opinion in virology* 3.1 (2013), pp. 84–91.

[580] Cadhla Firth et al. “Detection of zoonotic pathogens and characterization of novel viruses carried by commensal Rattus norvegicus in New York City”. In: *MBio* 5.5 (2014), e01933–14.

[581] Tung G Phan et al. “The fecal viral flora of wild rodents”. In: *PLoS pathogens* 7.9 (2011), e1002218.

[582] Simon H Williams et al. “Viral diversity of house mice in New York City”. In: *MBio* 9.2 (2018), e01354–17.

[583] Jayna Raghwani et al. “Seasonal dynamics of the wild rodent faecal virome”. In: *Molecular Ecology* (2022).

[584] Bert Vanmechelen et al. “Discovery and genome characterization of three new Jeilongviruses, a lineage of paramyxoviruses characterized by their unique membrane proteins”. In: *BMC genomics* 19 (2018), pp. 1–11.

[585] Zhiqiang Wu et al. “Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases”. In: *Microbiome* 6 (2018), pp. 1–14.

[586] Jiarong Guo et al. “VirSorter2: A Multi-Classifier, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses”. In: *Microbiome* 9.1 (Feb. 1, 2021), p. 37. ISSN: 2049-2618. DOI: 10.1186/s40168-020-00990-y. URL: https://doi.org/10.1186/s40168-020-00990-y (visited on 03/16/2023).

[587] Jie Ren et al. “VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data”. In: *Microbiome* 5 (2017), pp. 1–20.

[588] Jie Ren et al. “Identifying viruses from metagenomic data using deep learning”. In: *Quantitative Biology* 8 (2020), pp. 64–77.

[589] Xinwei Xiong et al. “Identifying Biomarkers of the Gut Bacteria, Bacteriophages and Serum Metabolites Associated with Three Weaning Periods in Piglets”. In: *BMC Veterinary Research* 18.1 (Mar. 17, 2022), p. 104. ISSN: 1746-6148. DOI: 10.1186/s12917-022-03203-w. URL: https://doi.org/10.1186/s12917-022-03203-w (visited on 01/30/2023).

[590] Thomas Briese et al. “Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis”. In: *MBio* 6.5 (2015), e01491–15.

[591] Andres Gomez, David Luckey, and Veena Taneja. “The Gut Microbiome in Autoimmunity: Sex Matters”. In: *Clinical Immunology*. Microbiome and Immune Diseases 159.2 (Aug. 1, 2015), pp. 154–162. ISSN: 1521-6616. DOI: 10.1016/j.clim.2015.04.016. URL: https://www.sciencedirect.com/science/article/pii/S1521661615001576 (visited on 12/19/2022).

[592] Eldin Jašarević, Kathleen E. Morrison, and Tracy L. Bale. “Sex Differences in the Gut Microbiome–Brain Axis across the Lifespan”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1688 (Feb. 19, 2016), p. 20150122. DOI: 10.1098/rstb.2015.0122. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2015.0122 (visited on 12/19/2022).

[593] Yong Sung Kim et al. “Sex Differences in Gut Microbiota”. In: *The World Journal of Men’s Health* 38.1 (Jan. 1, 2020), pp. 48–60. ISSN: 2287-4208. DOI: 10.5534/wjmh.190009. URL: https://doi.org/10.5534/wjmh.190009 (visited on 12/19/2022).

[594] Meng-Yuan Hu et al. “Sexual Dimorphism of the Gut Microbiota in the Chinese Alligator and Its Convergence in the Wild Environment”. In: *International Journal of Molecular Sciences* 23.20 (20 Jan. 2022), p. 12140. ISSN: 1422-0067. DOI: 10.3390/ijms232012140. URL: https://www.mdpi.com/1422-0067/23/20/12140 (visited on 12/19/2022).

[595] T. Odamaki et al. “Age-Related Changes in Gut Microbiota Composition from Newborn to Centenarian: A Cross-Sectional Study”. In: *Bmc Microbiology* 16 (May 2016), p. 12. ISSN: 1471-2180. DOI: 10.1186/s12866-016-0708-5.

[596] Lara Parata et al. “Age, Gut Location and Diet Impact the Gut Microbiome of a Tropical Herbivorous Surgeonfish”. In: *FEMS Microbiology Ecology* (). DOI: 10.1093/femsec/fiz179. URL: https://academic.oup.com/femsec/advance-article/doi/10.1093/femsec/fiz179/5632104 (visited on 12/04/2019).

[597] Mareike C. Janiak et al. “Age and Sex-Associated Variation in the Multi-Site Microbiome of an Entire Social Group of Free-Ranging Rhesus Macaques”. In: *Microbiome* 9.1 (Mar. 22, 2021), p. 68. ISSN: 2049-2618. DOI: 10.1186/s40168-021-01009-w. URL: https://doi.org/10.1186/s40168-021-01009-w (visited on 03/23/2021).

[598] Sage D Rohrer et al. “Composition and function of the Galapagos penguin gut microbiome vary with age, location, and a putative bacterial pathogen”. In: *Scientific Reports* 13.1 (2023), pp. 1–13.

[599] Keijiro Mizukami et al. “Age-related analysis of the gut microbiome in a purebred dog colony”. In: *FEMS microbiology letters* 366.8 (2019), fnz095.

[600] Baptiste Sadoughi et al. “Aging Gut Microbiota of Wild Macaques Are Equally Diverse, Less Stable, but Progressively Personalized”. In: *Microbiome* 10.1 (June 19, 2022), p. 95. ISSN: 2049-2618. DOI: 10.1186/s40168-022-01283-2. URL: https://doi.org/10.1186/s40168-022-01283-2 (visited on 06/20/2022).

[601] Anna Cuscó et al. “Long-Read Metagenomics Retrieves Complete Single-Contig Bacterial Genomes from Canine Feces”. In: *BMC Genomics* 22.1 (May 6, 2021), p. 330. ISSN: 1471-2164. DOI: 10.1186/s12864-021-07607-0. URL: https://doi.org/10.1186/s12864-021-07607-0 (visited on 12/25/2022).

[602] Daniel H. Huson et al. “MEGAN-LR: New Algorithms Allow Accurate Binning and Easy Interactive Exploration of Metagenomic Long Reads and Contigs”. In: *Biology Direct* 13.1 (Apr. 20, 2018), p. 6. ISSN: 1745-6150. DOI: 10.1186/s13062-018-0208-7. URL: https://doi.org/10.1186/s13062-018-0208-7 (visited on 12/25/2022).

[603] J. A. Frank et al. “Improved Metagenome Assemblies and Taxonomic Binning Using Long-Read Circular Consensus Sequence Data”. In: *Scientific Reports* 6.1 (1 May 9, 2016), p. 25373. ISSN: 2045-2322. DOI: 10.1038/srep25373. URL: https://www.nature.com/articles/srep25373 (visited on 12/25/2022).

[604] Alexander T. Dilthey et al. “Strain-Level Metagenomic Assignment and Compositional Estimation for Long Reads with MetaMaps”. In: *Nature Communications* 10.1 (1 July 11, 2019), p. 3066. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10934-2. URL: https://www.nature.com/articles/s41467-019-10934-2 (visited on 01/17/2021).

[605] Jeremy Fan, Steven Huang, and Samuel D. Chorlton. “BugSeq: A Highly Accurate Cloud Platform for Long-Read Metagenomic Analyses”. In: *BMC Bioinformatics* 22.1 (Mar. 25, 2021), p. 160. ISSN: 1471-2105. DOI: 10.1186/s12859-021-04089-5. URL: https://doi.org/10.1186/s12859-021-04089-5 (visited on 12/25/2022).

[606] Jasmine J. Hatton et al. “Diet Affects Arctic Ground Squirrel Gut Microbial Metatranscriptome Independent of Community Structure”. In: *Environmental Microbiology* 19.4 (2017), pp. 1518–1535. ISSN: 1462-2920. DOI: 10.1111/1462-2920.13712. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.13712 (visited on 12/25/2022).

[607] Yue Jiang et al. “Metatranscriptomic Analysis of Diverse Microbial Communities Reveals Core Metabolic Pathways and Microbiome-Specific Functionality”. In: *Microbiome* 4.1 (Jan. 12, 2016), p. 2. ISSN: 2049-2618. DOI: 10.1186/s40168-015-0146-x. URL: https://doi.org/10.1186/s40168-015-0146-x (visited on 12/25/2022).

[608] *Five Key Aspects of Metaproteomics as a Tool to Understand Functional Interactions in Host Associated Microbiomes*. Jenny Stanford Publishing, Dec. 22, 2021, pp. 647–660. ISBN: 978-1-00-318043-2. DOI: 10.1201/9781003180432-35. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003180432-35/five-key-aspects-metaproteomics-tool-understand-functional-interactions-host-associated-microbiomes-fernanda-salvato-robert-hettich-manuel-kleiner (visited on 12/25/2022).

[609] Guoliang Li et al. “Host-Microbiota Interaction Helps to Explain the Bottom-up Effects of Climate Change on a Small Rodent Species”. In: *The ISME Journal* 14.7 (7 July 2020), pp. 1795–1808. ISSN: 1751-7370. DOI: 10.1038/s41396-020-0646-y. URL: https://www.nature.com/articles/s41396-020-0646-y (visited on 02/07/2022).

[610] Eric A Franzosa et al. “Relating the metatranscriptome and metagenome of the human gut”. In: *Proceedings of the National Academy of Sciences* 111.22 (2014), E2329–E2338.

[611] Doratha A Byrd et al. “Comparison of methods to collect fecal samples for microbiome studies using whole-genome shotgun metagenomic sequencing”. In: *Msphere* 5.1 (2020), e00827–19.

[612] Tanja Woyke et al. “Assembling the marine metagenome, one cell at a time”. In: *PloS one* 4.4 (2009), e5299.

[613] Koji Arikawa et al. “Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics”. In: *Microbiome* 9 (2021), pp. 1–16.

[614] Yang Yang et al. “Single-cell metagenomics and metagenomics approaches reveal extracellular electron transfer of psychrophilic electroactive biofilms”. In: *Science of The Total Environment* 836 (2022), p. 155606.

[615] Faming Zhang et al. “Microbiota transplantation: concept, methodology and strategy for its modernization”. In: *Protein & cell* 9.5 (2018), pp. 462–473.