Alma Mater Studiorum · Università di Bologna

DIPARTIMENTO DI FARMACIA E BIOTECHNOLOGIE - FABIT CORSO DI LAUREA IN GENOMICS

Exploring the microbiome of large carnivores and related evolutionary trajectories.

Presentata da:
Gabriel Haw

Matricola n°:
0000922060

Relatore:
Simone Rampelli

Correlatore:
Marco Candella

A.A.
2021-2022 VI session

Abstract

The focus of this thesis was to explore and study the evolutionary relationships between mammalian lineages belonging to the order Carnivora and their respective gut microbiomes. Currently it is well documented that both diet as well as host phylogeny influence the gut microbiome composition, however there remains uncertainty about the exact microbial players making up the ancient microbiome and in particular which microbial species are preserved throughout the evolution of Carnivora. A greater understanding of these players and their preservation was achieved within this work through the comparison of microbial taxa originating from present-day mammalian hosts with those coming from an extinct paleosubspecies of spotted Hyena (*Crocuta crocuta spelaea*) sampled within Italy. By applying the bioinformatic tool Count, we were able to detect signals of phylosymbiosis at common ancestral nodes along with the preservation of microbial commensals. In particular, the genera *Lachnoclostridium*, *Clostridium* and *Peptoclostridium* were shown to be preserved over evolutionary timescales with further presence also observed within the ancient hyena species. Such bacteria are involved in animal fat and protein degradation and for this reason are associated with a carnivorous diet. Furthermore, both *Lachnoclostridium* and *Clostridium* have the ability to synthesise short-chain fatty acids (SCFA) which are known beneficial metabolites correlated to the intestinal health of animals. Together, such results can be utilised to increase our understanding of the evolutionary importance of gut microbiome components, providing further and unique insights into their relevance for host biology.

Contents

1 Introduction

1.1 The Gut Microbiome (GM)

Microbioata that shape our gut are indispensable and have become increasingly researched overtime. The microbiome acts as a complex microbial ecosystem, with bacteria constituting the largest percentage of the microbial biomass, and other players including archaea, fungi and protozoa (Wu et al., 2022; Turnbaugh et al., 2007). The term "holobiont" characterizes the co-existence of both host and its microbial symbionts, including those that are considered core constant and to the lesser extent variable members (Gilbert et al., 2012). These symbionts can be introduced as a result of vertical or horizontal transmission, or through environmental acquisition (speciation of symbiosis). The significance and ubiquity of microbial interactions with respect to mammalian hosts has become apparent thanks to a multitude of studies focusing on host development. The GM aids primarily in the digestion of complex food items and therefore participates at the level of nutrient and energy uptake (Stevens and Hume, 1998; de Jonge et al., 2022). Moreover, it plays both a prominent role in the mediation of the central nervous system and efficient functioning of the immune system, protecting the host from the threat of various potential pathogens (Gill et al., 2006; Zhu et al., 2011; Pickard et al., 2017; Kogut et al., 2020). The relationship between host and microbiome is one of the utmost importance and symbiotic relationships of this nature undergo extensive generations of co-evolution alongside the host gastrointestinal system (Stevens and Hume, 1998). Holobiont co-evolution has driven microbial specialization so that microbiota may better to meet the various nutritional needs of host species.

1.1.1 Core mammalian bacterial taxa

The acquisition of gut bacterial communities with respect to mammalian hosts occurs during their movement through the birth canal and the subsequent maternal, social and environmental transmissions (Nelson et al., 2013). The gut microbiomes of healthy mammalian animals are largely constituted by the bacterial phyla Firmicutes, Proteobacteria, and Bacteroidetes (Wu et al., 2022). Firmicutes and Bacteroidetes are found to be responsible for carbohydrate metabolism and are vital for their role in energy production and conversion, amino acid transport and metabolism, and production of short-chain fatty acids (Magne et al., 2020). Proteobacteria on the other hand are thought to play a key role in gut preparation (Shin et al., 2015). The colonization of the gut by necessary anaerobes allows for the consumption of oxgyen and lowering of redox potential within the environment of the gut, therefore being an important driver for healthy gut functioning.

1.1.2 GM evolution of various dietary niches

When positioning the spotlight on both the microbiome diversity and traits linked to host phylogeny, the two appear to exist in an interdependent setting with host diet acting as the cardinal factor (Gregor et al., 2022). Diet has been shown to predominate in the acquisition of ancient and large microbial lineages (Groussin et al., 2017), supporting the notion that the introduction of these lineages at different evolutionary timescales is attributable to adaptions in host diet (i.e herbivourous and carnivorous). As a result of this inter-dependency and the accumulation of specific microbes driven by dietary specialization, microbial taxa are capable of providing some insight into the evolutionary trajectories belonging to a wide range of organisms.

The importance of the GM is most noticeable at the level of herbivores. Early mammals hosted a simplistic gut morphology, one that is still observed in many of the carnivore (Van Valkenburgh, 2007) and omnivore representatives of today (Stevens and Hume, 1998). As animals are not able to produce the endogenous cellulolytic enzymes required to digest plant-based materials, herbivorous mammals are therefore heavily dependent on their microbial symbionts (Smant et al., 1998). This has generated a more complex gastrointestinal tract development, compared to their carnivorous and omnivorous counterparts (Ley et al., 2008; Bayané and Guiot, 2011).

Microbial lineages corresponding to herbivorous mammals seem to have appeared earlier on in the bacterial evolution whilst those allowing for the specialization necessary to transition to a carnivorous diet emerged much later (Groussin et al., 2017; Sperling et al., 2013). Furthermore, when studying both microbial composition and diversity between these two dietary groups, appreciable differences arise at microbial species level. Microbiota associated to Carnivora tend to show a larger degree of variability both within and between species, and taken together with research conducted on herbivores it has been shown that their microbial diversity is significantly lower compared to that coming from their herbivorous counterpart (Zoelzer et al., 2021). Interestingly, omnivorous mammals have been found to host no omnivore-specific bacterial taxa but rather a combination of herbivorous and carnivorous bacterial groups, consistent with the idea that mammals constituting this dietary group evolved from either herbivorous or carnivorous ancestors (Groussin et al., 2017).

1.1.3 Patterns of phylosymbiosis

Phylosymbiosis can be defined as "microbial community relationships that recapitulate the phylogeny of their host" (Lim and Bordenstein, 2020). Phylosymbiosis is a powerful tool for understanding evolutionary processes that are key drivers for host-microbial structure. Ultimately, phylosymbiotic signals can be studied through the identification

of significantly recurring microbial taxa. Taxa that are said to recur in a non-random manner allow us to study symbiotic relationships between host and microbes. Microbial symbionts seen in successive host generations can be utilised to capture evolutionary changes or adaptions taking place at the level of host phylogeny. Therefore, by better understanding the extent to which gut microbial composition and function are correlated to host phylogeny we may better assess drivers of phylogenetic evolution and whether these events are caused by changes occurring at the level of diet or can be explained with greater resolution at the level of phylogenetic background.

Phylosymbiosis is often influenced by the richness of microbial taxa, a richness generated by the plethora of animals constituting the animal kingdom (Mallott and Amato, 2021). This means as we examine various gut microbiomes, belonging to different taxonomic clades, we tend to see less and less bacterial phyla advocating for the presence of phylosymbiosis. This pattern is believed to the result of a non-uniform assembly process taking place at the level of individual taxa. A pattern mammals appear to escape, evident by their frequent exhibition of phylosymbiosis (Mallott and Amato, 2021). This exception maybe the result of a unique combination of mammalian traits ie. viviparity, lactation and parental care, that allow for a more uniform uniform process when acquiring individual taxa (Mallott and Amato, 2021).

Both vertical and horizontal transmission are drivers of the acquisition of various microbial communities at the level of the host (Groussin et al., 2017; Mallott and Amato, 2021). Vertical transmission has the ability to participate in both restricted and large scale microbial acquisition. The aquistion of a select few microbial lineages can occur at the level of germ-line colonization by intracellular symbiotic microorganisms (Mallott and Amato, 2021). In contrast, large scale microbial community acquisition can be driven by various physiological and behavioural pathways in both invertebrates and vertebrates, as is the case when looking at animals that exhibit parental care, ultimately allowing for contact mediated transmission. Horizontal transmission, much like that of large scale vertical transmission, is predominately seen at larger scales and occurs as a result of contact mediated interactions both with conspecifics and the nearby environment (Mallott and Amato, 2021). Therefore, so to does host behaviour and geographical setting influence the various microbes that can be encountered.

1.2 Carnivora

Carnivores have fascinated mankind for Millennia despite constituting only 10% of all mammalian genera and only 2% of all mammalian biomass, thus making them a relatively small order (Gittleman, 2013). As an order they can be differentiated into two main evolutionary lineages, namely Feliformia and Caniformia with the divergence occurring between the two lineages at a minimum age of ∼43 Ma (Mega annum). The suborder

Feliformia is constituted by cats, mongooses, civets, hyenas, meerkats and other 'cat-like' carnivorans, whereas Caniformia contains both terrestrial families such as wolves, dogs, coyotes, foxes and bears, as well as marine carnivores belonging to the taxanomic clade Pinnipedia (Seals) (Eizirik et al., 2010). This taxonomic classification was first put forward by Flower (1869) and has since been consistently backed by multiple phylogenetic studies (Wesley-Hunt and Flynn, 2005).

1.3 Hyaenidae

Despite their canine-resemblance Hyaenidae bear closer genetic and molecular relation to Feliform taxa, in particular the family Herpestidae (Holekamp et al., 2007). A taxonomic clade containing both mongooses and the Madagascan fossa, with fossil data suggesting the split from their last common Feliform ancestor having occurred in the Oligocene, roughly 25–29 Ma (Holekamp et al., 2007).

The emergence of hyaenids occurred at the start of the the Miocene epoch (23.03–5.33 Ma), a period of global warmth and the introduction of many fauna and flora, some of which evolving into predecessors still seen today (Steinthorsdottir et al., 2021). Following this emergence hyenas then diversified into what can be considered two subgroups, smaller dog-like hyenas and larger more muscular bone-crushing hyenas. However, dog-like hyenas fell victim to climate change and coupled with the appearance of canids, can only be traced to the genus Proteles, a present day insectivorous relative, commonly known as the aardwolf. In contrast, their bone-crushing relatives thrived and due to their efficient scavenging abilities, expanded their geographical range eventually crossing the Gomphothere land bridge connecting Eurasia to Afro-Arabia approximately 15-19 Ma (Harzhauser et al., 2007; Holekamp et al., 2007). Currently, it is believed that all four extant hyaenid species emerged from within Africa, with the Hyaenidae family representing one of the smallest carnivora families and being made up of four species of hyena, namely the striped hyena (*Hyaena hyaena*), the brown hyena (*Parahyaena brunnea*), the spotted hyena (*Crocuta crocuta*), and lastly the aardwolf (*Proteles cristata*) (Hu et al., 2021).

1.3.1 Genus Crocuta

The genus Crocuta, having attracted considerable evolutionary and systematic interest as result of social systems similar to that of many primates, is constituted by both several closely related extinct and extant lineages (Holekamp et al., 2007). The genus is believed to have evolved from within Africa no later than 4 Ma, and thanks to superior scavenging capabilities was able to expand and disperse from the continent 2-2.5 Ma (Lewis and Werdelin, 2022). Towards the end of the Pleistocene, the genus had an extensive geographical range (Westbury et al., 2020), occupying a significant portion of

Eurasia. Fossil records indicate that cave hyenas had shorter distal limbs in comparison to there African analogue, an indication coupled with changes in tooth morphology, that they predominated as scavengers rather than active hunters (Westbury et al., 2020). These and other morphological changes brought on by the environmental pressures attributed to their geographical expansion led to the hypothesised taxonomic distinction between the Eurasian and African lineages. However, in recent years with the use of short fragments of mtDNA, African spotted hyenas were found to be more intermingled with respect to the mitochondrial haplogroups coming from Eurasian cave hyenas (Rohland et al., 2005) than originally thought, suggesting that the two are not in fact distinct taxa.

Today the genus is continued by the Spotted hyena (*Crocuta crocuta*). Originating in sub-Saharan Africa this species displays complex social behaviors (Westbury et al., 2020), typically living within large clans headed by a matriarch and cubs inheriting social ranks just below those of their mothers (Holekamp et al., 2007). As a highly adaptable species Spotted hyenas are best described as opportunistic feeders participating in both hunting and scavenging practices. Currently, they represent the most abundant large carnivore within Africa (Westbury et al., 2020).

1.3.2 Eurasian cave hyena and the genus Homo.

"As far back in human evolutionary history as evidence permits us to see, the presence of hyenas in ancestral human landscapes is always implied, if not positively attested to by their bones and teeth marks" (Baynes-Rock, 2015).

The co-occurrence between ancient hyenas and early hominins is evident thanks to the identification of bone deposits found within caves dating back to the Middle and Late Pleistocene, specifically those attributed to both Neanderthals (*Homo neanderthalensis*) and spotted hyenas (*Crocuta crocuta*) (Baquedano et al., 2016). Furthermore, bone remains indicating prey item overlaps between both spotted hyenas and Paleolithic man within Europe have been identified, reiterating the dietary similarities between these groups at the time (Baquedano et al., 2016). The co-existence between hyenas and our human ancestors wasn't one of mutualism and more and more often hominins began taking control of the adaptive niches previously dominated by hyenas (Baynes-Rock, 2015). Moreover, it can be assumed that initially spotted hyenas following their dispersion into Eurasia, likely considered these bipedal primates as prey items themselves and would actively hunt them if the opportunity arose (Baquedano et al., 2016).

However, these large terrestrial carnivores would came up short, and declines in spotted hyena populations roughly 20 Kya (Thousand years) would mark the start to an end for the Eurasian subspecies, with their complete disappearance taking place between 14-11 Kya (Stiner, 2021). Their extinction within Europe has been linked to a collapse in prey

item numbers as a result of spikes in hominin and wolf populations, associated carcass scarcity and lastly the disappearance of Pleistocene grasslands and semi-open woodlands habitats (Stiner, 2021).

The interaction between that of Homo and Crocuta having spanned millennia, has and is considered a major driver of the evolutionary trajectories of each Genus, a relationship that still exists today in many parts of Africa.

1.4 The analysis of ancient microbiomes

Next generation sequencing (NGS) is technology that relies on the parallel sequencing of millions of small fragments of DNA. These fragments or reads are subsequently joined together through their mapping and alignment onto a reference genome through multiple sequencing iterations increasing alignment accuracy (Behjati and Tarpey, 2013). As a result of technological developments within NGS platforms microbial exploration and identification has grown in both resolution and throughput (Warinner et al., 2015). Coupled with this has been the advancements in complementary bioinformatic tools, providing researchers with the ability to analyze larger amounts of sequencing data more efficiently. A consequence of this progress has been the development in fields such as paleomicrobiology, where we can now gain insight into the microbial ecologies of organisms that once called us home.

16S rRNA sequencing has been the mainstay when it comes to both the identification and classification of bacterial taxa for decades. The 16S rRNA gene represents the most established genetic marker and is made up by both highly conserved and hypervaribale regions, which have the ability to serve as both universal primer binding sites and specific discriminatory sequences allowing for the differentiation of microbial taxa. Through the use of public reference databases such as SILVA 16S rRNA gene sequencing can provide insight into whole bacterial, archaeal and fungal communities of both present day and extinct hosts. However, 16S rRNA sequencing can only identify bacteria up until the taxonomic rank genus level as a result of the high sequence similarity between 16S rRNA genes belonging to closely related bacterial species (Gupta et al., 2019).

Whole shotgun sequencing (WGS) unlike amplicon sequencing is able to capture metagenomic sequences from a much larger microbial species pool as a result of the technique not being constrained to the 16S rRNA gene region. Here small DNA fragments originating from the entire genome are assembled into larger contigs and eventually built into chromosomes. Therefore, shotgun sequencing has more discriminatory power in terms of bacterial classification as it is able to provide a much larger sequence pool to distinguish between closely related bacterial species (Durazzi et al., 2021). WGS as a deep sequencing technique is becoming more and more common in microbiome studies

is used to differentiate between bacterial strains, reveal presence and absence of genes and pathways as well as identify rare microbial taxa (Lind and Pollard, 2021).

2 Methods and Materials.

The relevant sources and computational packages used within this study can be found in the Annex within Table 5 and Table 6 respectively, with the code utilized in order to generate graphs or run permutational testing found within the section: Code used within the thesis.

2.1 Ancient microbiome analysis

The coprolite sample belonging to C. spelaea (ca. 40,000 - 50,000 years ago) was found at the Uluzzo C rock shelter in Apulia, southern Italy, an important archaeological site due to the preservation in rock successions (Spinapolice et al., 2022).

2.1.1 aDNA extraction protocol and library preparation

Ancient DNA analysis and extraction were carried out externally from the microbial characterization conducted within this thesis where both extraction and incubation of the samples was performed using the protocol described in detail within (Rampelli et al., 2021) and as a result were therefore specified to coprolite-materials.

Library preparation was conducted following identical preparation methods detailed within (Rampelli et al., 2021), where the method "BEST" (Blunt-End-Single-Tube)(Carøe et al., 2018) was used to construct the shotgun sequencing library. The final library generated using the protocol utilized in (Rampelli et al., 2021) was then sequenced using the Illumina NextSeq 500 platform with 2 X 75 bp paired-end chemistry (Rampelli et al., 2021). Sequencing data was then processed whereby merged reads where kept if they matched the forward and reverse bar codes using AdapterRemoval

2.1.2 Bioinformatic analysis of C. spelaea

We applied PhyloFlash (Gruber-Vodicka et al., 2020) on the metagenomic sequences belonging to C. spelaea in order to have an initial understanding of microbial structure. Further characterization was then performed using the HOPs pipeline (Hübler et al., 2019) where previously filtered reads were aligned to a reference database using MALT (MEGAN ALignment Tool) (Vågene et al., 2018). This alignment tool assigns reads to taxonomic nodes based on the naïve Lowest Common Ancestor (LCA) algorithm. Reads assigned to bacterial genomes were retained using MaltExtract and subsequently realigned to their respective genomes. Additional filters within MaltExtract were set

to retain only the top one percent of assigned bacterial taxa. These taxa were used as representative bacterial taxa in subsequent downstream analysis.

2.2 Mammalian microbiome data set

The data set utilized within this study contains gut microbial taxa originating from a total of 19 mammalian species (15 carnivores, 2 herbivores and 2 omnivores), further dividable into 4 distinct orders. The extant animal samples were procured from two separate external sources which can be located in with the Annex. Whilst this study focuses on carnivore-specific microbial taxa, outliers in the form of both herbivorous and omnivorous mammals were included for comparative purposes. The extant species sequence data was generated using both the Illumina MiSeq and Shotgun sequencing platforms, with 16S rRNA amplicons focusing on the V3 region in particular (Milani et al., 2020).

As the majority of extant species data was taken from one study, we can limit or reduce the potential bias at the level of DNA extraction protocol, 16S primers and the method of sequencing (Groussin et al., 2017). Moreover, samples coming from that particular data set, were collected from animals that had not received pharmaceutical treatment in the six months prior to fecal collection (Milani et al., 2020) further limiting any biases generated at the level of identifiable bacterial taxa.

2.2.1 Quality filtering of 16S rRNA and shotgun sequences

The raw fastq files belonging to the extant species data set as well as the individual African spotted hyena sample (*Crocuta crocuta*) were downloaded from the European Nucleotide Archive (ENA). There after, the fastq files along with that belonging to C. spelaea were then processed using FastQC which ultimately performs quality control checks on raw sequence data. The sequences were then pruned for adapters and the subsequent trimming of stretches of mixed low-quality 5' and 3' termini bases was carried out using the package AdapterRemoval (Lindgreen, 2012). Quality filtering of the sequences was then performed, where only sequences having a mean quality score >20 were retained. This was achieved using BBDuk, a data-quality-related trimming tool. The quality filtered sequences were then used in all downstream analysis.

2.3 Microbial analysis of the mammalian hosts

Microbial profiling of sequences originating from both 16S rRNA and shotgun methods was performed using PhyloFlash, a tool designed for short-read taxonomic profiling of metagenomic data. 16S rRNA microbial profiles were evaluated through Phyloflash whereby metagenomic sequences were assigned to 16S rRNA gene references within the

SILVA database. From this, sequences that matched to that of a reference sequence within the database produced a corresponding bacterial hit. These hits were then stored for the 16s rRNA dataset.

The microbial profiles originating from the shotgun dataset were subsequently evaluated through Phyloflash whereby the shotgun metagenomic sequences were also assigned to gene references within the SILVA database. Again shotgun metagenomic sequences that matched to that of a reference sequence within the database produced a corresponding bacterial hit which we then stored in a separate shotgun metagenomics dataset.

In order to produce a unique amalgamated 16s rRNA and shotgun bacterial dataset the python package merge_metaphlan_tables.py was used to merge both the host samples and their respective bacterial hits, allowing for the construction of a deduplicated OTU abundance table. Initial filtering implemented in RStudio was performed to limit the number of bacterial taxa introduced through environmental means e.g. sampling spillover, and to reduce unwanted taxa. This amounted to the retention of reads if they were present in at least one other extant mammalian species and were identified as non-Chloroplast and non-Eukaryota taxa, ultimately resulting in total of 8,685 bacterial taxa.

2.3.1 Exploring alpha and beta diversities

In order explore host-microbial relationships between and within the studied mammalian lineages we made use of the α and β diversity metrics, which allowed for the assessment of gut microbial components belonging to the respective hosts.

To achieve this we built a bacterial monophyletic tree that allowed us to assess the different host-microbial diversity metrics. Here both the Shotgun and 16S rRNA reads were aligned using SINA v 1.7.2 (Pruesse et al., 2012) to reference sequences within the SILVA database (release 138.1). The trimAl package (Capella-Gutiérrez et al., 2009) was then used for further filtering of the aligned reads, where sites having gap positions $> 95\%$ were removed. The OTU abundance table was filtered for Proteobacteria, Actinobacteria and Firmicutes necessary to generate the monophyletic tree. Constraints were also applied by generating unique 3-digit binary identifiers necessary for the superimposition of the monophyletic tree topology and to avoid the mixing of bacterial phyla. FastTree (Price et al., 2009) was then used to build the constrained monophyletic tree containing all 8,685 identified individual bacterial taxa, and both the CAT approximation and GTR model were applied to further evaluate rates of heterogeneity across individual sites.

Read count within the abundance table was then assessed in order to identify the sample with the lowest number of counts. The resultant statistics can be seen in Table 1 with

the minimum and maximum number of reads being shown. Here one can observe a large difference between counts as well as the standard deviation which can introduce a bias in the host-bacterial profiling downstream. In order to simulate an even number reads per sample and perform more accurate microbial profiling, sample read count rarefaction was employed and was performed using the python package single_rarefaction.py. Rarefaction involves the random subsampling of metagenomic reads from the library entire read library in order to select a normalised library size and represents a widely used normalisation technique (Cameron et al., 2021). Here reads were rarefied to a read count of 5000 which helps reduced read count associated biases seen at the level of reads originating from the different sampling methods i.e., Shotgun and 16S rRNA. This normalised count table was then used for subsequent downstream analysis.

Num samples: 19
Num observations: 14399
Total count: 1179082
Table density (fraction of non-zero values): 0.125
Counts/sample summary:
Min: 5135.000
Max: 101240.000
Median: 50486.000
Mean: 49128.417
Std. dev.: 31603.950
Sample Metadata Categories: None provided
Observation Metadata Categories: taxonomy

Table 1: output generated from single_rarefaction.py

2.3.2 Biostatistics and graphical representation

The normalised count table was then used along with the monophyletic tree to generate both phylogenetic alpha and beta diversity metrics using the respective python packages alpha_diversity.py and beta_diversity.py implemented in QIIME, an open-source bioinformatics pipeline that can be used for performing microbiome analysis. The alpha_diversity.py package granted us the ability to measure microbial diversities within individual samples and through the use of the observed species metric, which counts of the number of unique OTUs within each mammalian sample.

The beta_diversity.py package was used to generate both weighted and unweighted Unifrac tables which allowed the measurement of mammalian phylogenetic distances using bacterial taxa within the monophyletic tree. Distances here are calculated using the Unifrac metric, a phylogenetically aware metric where distances are measured as a fraction of

branch length that is said to determine whether or not taxa are inherited from one branch or the other (Lozupone and Knight, 2005). These Unifrac tables were converted into their respective dataframes within R whereby we could graphically report the results through the use of Principal Coordinate Analysis (PCoA) implemented with the R library vegan.

2.4 Analysis of phylosymbiotic taxa

2.4.1 Mitochondrial reconstruction of the C. spelaea sample

Reconstruction of the mitochondrial DNA belonging to C. spelaea was carried out using Schmutzi (Renaud et al., 2015), an iterative approach that allows for estimated reconstruction of endogenous mitochondrial genomes. Here reads were trimmed and mapped using the Burrows-Wheeler aligner (bwa) (Li and Durbin, 2009) to the mitochondrial reference sequence taken directly from the NCBI database and belonging to that of C. crocuta. The bwa mapping options included the maximum accepted edit distance being set to 1%, the maximum number of gap openings set to 2 and the seed length and long gap options being disabled in accordance with sub-commands used in (Rampelli et al., 2021). Furthermore, alignment filtering using Samtools was also carried out and alignments having a quality score > 30 were retained. Deamination rates were then assessed using both bam2prof and the endoCaller program, a sub-package of Schmutzi that allows for the consensus calling of mitochondrial data. The reconstructed sequence was then used for subsequent downstream phylogenetic analysis.

2.4.2 Mitochondrial sequences of extant mammalian species

Here a file containing all the mitochondrial dna sequences pertaining to the samples used within the study was generated. The mitochondrial sequences belonging to the extant mammalian hosts were taken from the NCBI database (see Annex), where as the mitochondrial sequence of C. spelaea was generated within this study, see subsection: Mitochondrial reconstruction of the C. spelaea sample. Initial alignment of the sequences was carried out using MUSCLE and subsequent gap position trimming was achieved through the use of trimAl.

2.4.3 Mitochondrial tree construction

Tree construction was performed using FastTreeMP (Price et al., 2009), were an additional sub-command was used to make the maximum-likelihood nearest neighbour interchanges more exhaustive. Figtree was then used to root the tree on the node that allowed for the separation of carnivorous and herbivorous taxa, with omnivorous taxa being allocated to the carnivorous split.

2.4.4 The Count pipeline

In order to determine the presence of phylosymbiotic microbial taxa we employed the software package Count which granted us the ability to explore bacterial counts along the respective mammalian phylogenetic profiles (Groussin et al., 2017). The normalised count table containing the bacterial phyla was investigated at the taxonomic level Genus throughout this step in order to obtain a more resolved phylosymbiotic assessment. Further filtering of the table was performed through the phyloseq package within RStudio according to minimum sample prevalence (>= 5%) and minimum sample abundance (>=15) where minimum sample abundance retains taxa that have at least this many read counts within one or more samples. The resulting 563 bacterial taxa, which we considered core, were further converted in Count into their respective binary profiles indicating bacterial presence or absence. Optimization of the newly generated binary table in context of the phylogenetic tree produced in subsection: Mitochondrial sequences of extant mammalian species was then performed. Here optimization criteria were set for both the model type and rate variation across families, with model type set to gain and loss and the penalty for rate variation across families set to 4 for all parameters i.e., gain, loss, and edge (Groussin et al., 2017). Lastly the number of optimization rounds adjusted to 20. From this we extracted the posterior probabilities which give an idea of the probability of observing a bacterial taxon at a particular node within the phylogenetic tree. The posterior probabilities belonging to each individual taxa were then summed across all nodes within the tree to ultimately generate the total bacterial prevalence at each phylogentic node for the observed data.

2.4.5 Statistical evaluation of microbial taxa

To further determine whether or not the observed bacterial profiles represented spurious or authentic phylosymbiotic relationships we generated 20 null tables for comparative purposes using the permatswap function contained within the RStudio library vegan. This function allowed us to generate null tables which allow for the degradation of observed phylogenetic relationships whilst preservation of characteristics such as species frequency, richness or rarity at each site. Inner options within the permatswap function were as follows: The method was set to "swap", which sequentially changes the structure of original presence absence matrix but does not not influence marginal sums, the number of permuted matrices was set to 100 and lastly the number of discarded permuted matrices between two sequential generations within the "swap" method was set to 50000 to better increase the randomization of the generated tables.

Theses tables were then converted into their respective binary profiles within Count using the before-mentioned optimization procedures. The posterior probabilities associ-

ated to the bacterial phyla at each node were then summed accordingly for all generated null tables. We then ran permutational tests whereby null nodal values greater or equal to the observed nodal value were summed and divided by the number of permutations (number of permutations was set to 999). This allowed us to compare bacterial profiles at each of the 18 nodes generated by the null tables with the corresponding observed bacterial profiles at the identical nodes. From this we were able to determine the significance of the association using the statistical P-value metric, whereby bacterial significance was granted on the basis of a P-value < 0.05.

Following the methodology implemented in (Groussin et al., 2017) we measured the Standard effect size (SES), which allowed us to make inference on the strength of phylosymbiotic profiles observed between mammalian Family and to the lesser extent Order lineages. The Standard effect size was measured by computing the difference between the bacterial nodal posterior probabilities extracted from the observed OTU nodes with the mean of the respective values originating from the null tables. This value was then standardized using the standard deviation generated by the null values for the corresponding phylogenetic node and is better illustrated using:

$$\mathbf{SES} = \frac{\mathbf{observed\ nodal\ OTU\ value\ -\ mean(Null\ nodal\ OTU\ value)}}{\mathbf{standard\ dev(Null\ nodal\ OTU\ value)}}$$

The R library vegan was then used to generate the graphical representation of the phylogenetic tree through the use of the package ape whereby nodes were colored according to the calculated standard effect size and significant phylosmbiotic signals were indicated with a black dot.

3 Results

Initially within this section we identify microbial taxa originating from the C. spelaea sample. We then briefly explore the microbial components of the extant carnivorous mammalian hosts, allowing us to identify and explore bacterial patterns at the level of the gut microbiome, and potentially identify shared taxa between extinct and extant hosts. Subsequently, we assess both the α and β diversities of the mammalian subjects studied within this thesis, potentially providing us with insight into dietary and phylogenic features of both extinct and extant lineages. Lastly, we use a subset of microbial taxa to study phylosymbiotic interactions within and between host mammalian lineages, potentially characterizing mammalian phylogenies. Taken together we can identify core microbial taxa that can be traced through evolutionary space illustrating their symbiotic importance and providing an evolutionary metric outside of that provided by anatomical and mitochondrial level classification.

3.1 Metagenomic profiling of the gut microbiome belonging to C. spelaea

Bacterial Species	Reads
Arthrobacter crystallopoietes	716
Archangium gephyra	698
Pseudarthrobacter sulfonivorans	673
Paenibacillus prosopidis	669
Pseudarthrobacter equi	638
Arthrobacter oryzae	611
Nitrososphaera viennensis EN76	596
Gemmatirosa kalamazoonesis	566
Geodermatophilus sabuli	557
Geodermatophilus ruber	535
Ramlibacter tataouinensis	505
Luteitalea pratensis	501
Geodermatophilus africanus	482
Paenibacillus harenae DSM 16969	445
Microvirga ossetica	441
Bacillus weihaiensis	440
Sphingomonas indica	435
Conexibacter woesei DSM 14684	433
Modestobacter marinus	428
Solirubrobacter soli DSM 22325	422

Table 2: The 20 most abundant microbial taxa identified within the C. spelaea coprolite samples identified using the HOPs pipeline (see Bioinformatic analysis of C. spelaea)

Through the use of the HOPs pipeline (Hübler et al., 2019), we were able to identify several C. spelaea associated microorganisms originating from the corresponding coprolite samples. Associated microbial taxa where then studied for abundance where we

retained the taxa displaying the greatest number of read hits at the taxonomic level species. Here, 20 out of the total 4,818 identified bacterial taxa specific to C. spelaea are shown in Table 2 along with their corresponding recorded read counts. In order to have taxonomic classification across several taxonomic ranks we then used PhyloFlash (Gruber-Vodicka et al., 2020) whereby microbial components of the mammalian GM were identified including members of the families *Bacillaceae*, *Lachnospiraceae*, *Streptococcaceae* and *Moraxellaceae* along with potential environmental contaminants such as the family *Streptomycetaceae* (Rampelli et al., 2021; Zoelzer et al., 2021).

When exploring the microbial composition of C. spelaea in terms of mean relative abundance (RB) the phyla *Firmicutes* (42%), *Proteobacteria* (18%), *Actinobacteria* (16%) and *Bacteroidota* (12%) were seen to be the most abundant. These identified phyla are consistently seen within present day mammalian lineages (Wu et al., 2022) as a result of the various advantageous functions they provide at host level and are considered dominant core phyla. The mean RB at both phylum and family taxonomic level are shown in Figure 1 (left), where RB is measure as the sum of a particular taxa relative to the total identified bacterial taxa.

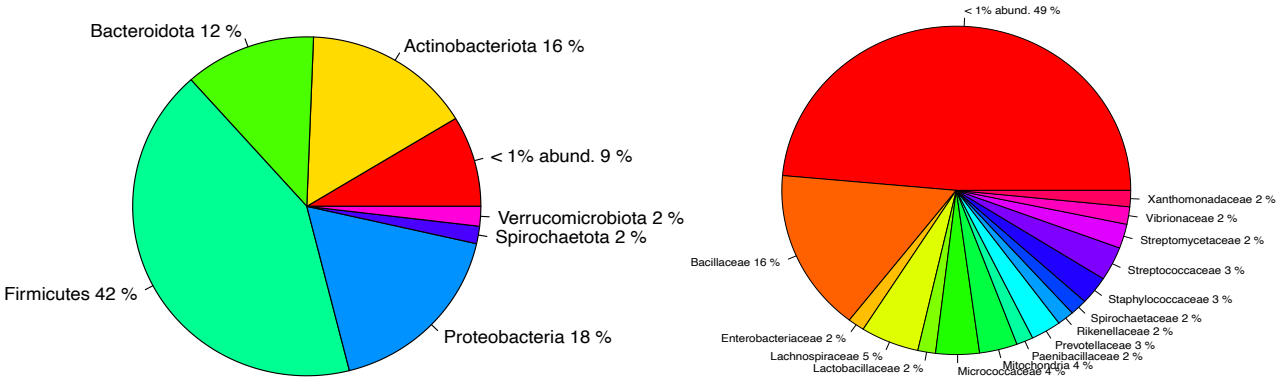

Figure 1: Mean relative abundance(%) at taxonomic level Phylum (left) and Family (right) of abundant taxa. Bacterial taxa were filtered using a relative abundance >= 0.015 where taxa falling below this threshold are indicated as "1% abund."

Microbial characterization at the taxonomic level family indicated in Figure 1 (right) allowed us to see a more resolved image of the bacterial taxa belonging to C. spelaea. Interestingly, the highest relative abundance at family level is awarded to *Bacillaceae* (16%) a gram-positive bacterial family constituting the genus *Bacillus*, which at species level (*Bacillus weihaiensis*) was identified at high read counts (see Table 2). We also saw a high relative abundance of the family *Lachnospiraceae* (5%) considered a core gut microbial family as a result of their ability to degrade complex polysaccharides into short-chain fatty acids (Meehan and Beiko, 2014).

3.2 The microbial taxa characterising mammalian samples

Initially to have a general idea of the all the mammalian microbial components, exploration of bacterial taxa across all samples was carried out at the taxonomic level family. The taxa were filtered using their mean relative abundance (RB >=0.004 in over %20 of the samples) resulting in the retention of 20 bacterial families.

When looking at Figure 2 it becomes apparent that all mammalian lineages harbour the bacterial family *Lachnospiraceae* a dominant gut bacterial family found typically within mammalian digestive tracts (Meehan and Beiko, 2014). Many of the species that make up this family show production of butyric acid, an acid shown to be important for epithelial cell growth of both microbes and host (Meehan and Beiko, 2014). Here it also is apparent that Carnivora exclusively harbor bacterial lineages such as *Fusobacteriaceae*, *Coriobacteriaceae* and *Clostridiaceae* as all families appear to be absent within herbivorous lineages and appear in only one omnivorous species (Meerkat or *Suricata suricata*), a species known to predate on small mammals. These microbial composition patterns have all has also been recorded in external literature (Zoelzer et al., 2021). We also observed high amounts of the family *Bacteriodaceae* across many of the mammalian samples where it appeared to be found at an extremely high relative abundance within the human sample (63.9%), which is also consistently seen (King et al., 2019). Generally, at family level, it appeared as though carnivorous mammals retain a degree of similarity with respect to their microbial components and interestingly weren't vastly different to the herbivorous and omnivorous counterparts.

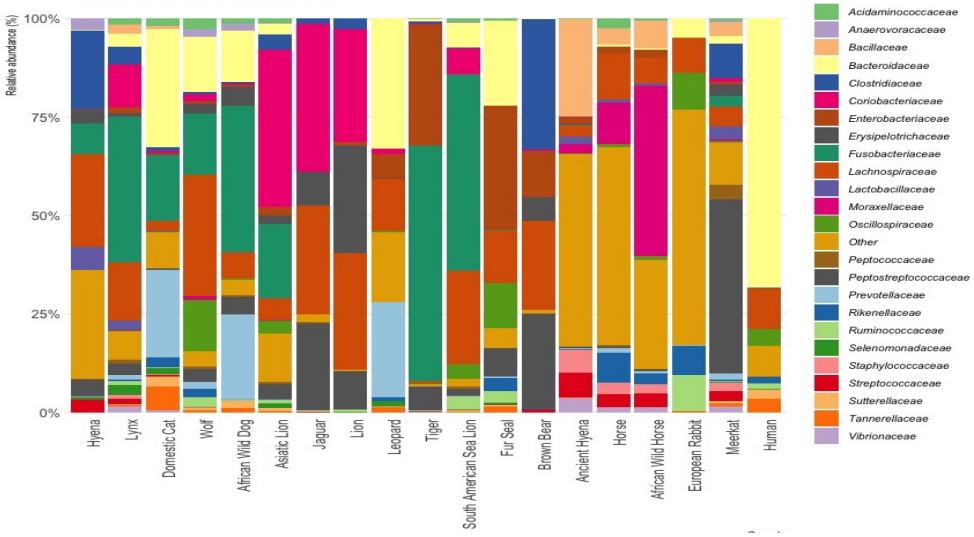

Figure 2: Bacterial families (25) filtered using a relative abundance threshold of (>=0.004) with prevalence (>=% 20) in the host species samples.

In order to have a more resolved image of the differences between the microbial components belonging to each diet we then studied bacterial composition at the taxonomic level Genus. Here we show 39 bacterial OTUs that were seen to display the highest relative abundance whereby the remaining genera were condensed to "Other". Furthermore, we noticed that some bacterial OTUs were only able to be classified to the taxonomic level Order or Family and therefore we refrain from using the term genera unless Genus level resolution was achieved.

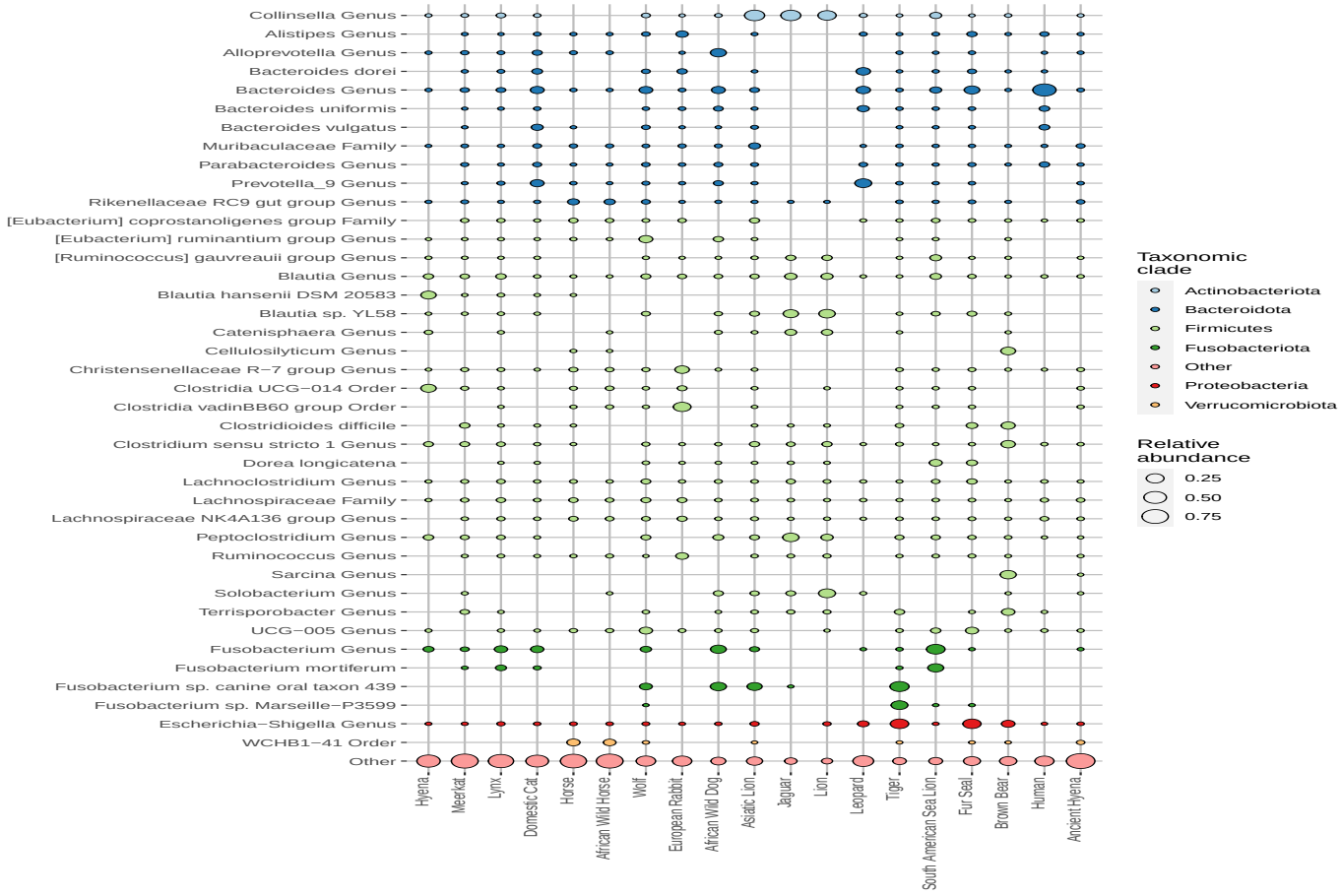

Figure 3: The top 40 bacterial OTUs according to relative abundance across samples, where taxa outside of this range have been condensed into "Other".

When assessing microbial composition at Genus level it was evident that the Order Carnivora harbored the Genus *Fusobacterium* at high relative abundances reiterating. Interestingly, whilst it appeared to be correlated to carnivorous diets it remained absent within African spotted hyena, Polar bear and Lion species despite all three being considered exclusive meat eaters. Carnivorous lineages when compared to the herbivorous outliers displayed more variation with respect to the bacterial OTUs seen within these

lineages as well as the relative abundances at which these OTUs were occurred. The ancient C. spelaea sample appeared to host a blended microbial composition whereby it harbored both bacterial OTUs that appeared to be either carnivorous i.e., *Fusobacterium* or herbivorous *WCHB1-41* in nature and the relative abundances at which these taxa occurred was seen to be relatively low. Overall when studying the bacterial components at genus level it was apparent that despite microbial discrepancies between the diets, there appears to be a sort of fundamental microbial basis or composition, where there are less bacterial taxa said to belong exclusively to one particular diet or the other, as can be seen within Figure 3. Furthermore, differences appeared to be at the level of relative abundances rather than presence microbial absence-profiles.

3.3 Microbial diversity across the mammalian GM

The content of mammalian gut microbiomes or microbial constituents can vary greatly between two communities or hosts (Groussin et al., 2017). Within and throughout this section we look at these compositional dissimilarities whereby we make use of the reconstructed bacterial phylogenetic tree, generated within subsection: Exploring alpha and beta diversities, in order to separate taxa through principal coordinates analysis (PCoA) using the unweighted UniFrac β-diversities. We then explore compositional diversity within hosts or α-diversity so that we may better understand microbial richness or the the number of bacterial species seen within an individual mammalian sample.

3.3.1 Exploring microbial richness within animal subjects

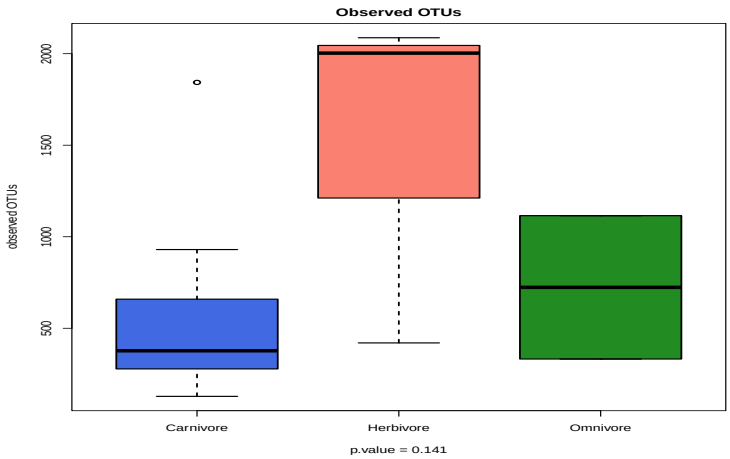

Figure 4: Alpha diversity explored using an observed species metric, providing insight into the microbial richness belonging to the different dietary groups and determined by the number of identified bacterial operational taxonomic units (OTUs). Here we can see that herbivores possess more microbial diversity with respect to the other dietary groups.

Exploration of α-diversity was implemented through the use of the observed species metric (see Exploring microbial richness within animal subjects) where subjects were grouped according to diet. This method allowed us to estimate the number of bacterial taxa within the individual host species, whereby we then checked for significant differences between groups using the Kruskal-Wallis H test. Usually statistically significant differences are recorded between the different diets (Nishida and Ochman, 2018), however we recorded no statistically significant differences between groups indicated by a p-value of P=0.22. This result can possibly be explained by the small number of study subjects used within this study when compared with other studies exploring this diversity measure. Although no significant differences were observed, herbivorous mammals were shown to exhibit more microbial richness at individual level compared to that of the omnivorous and carnivorous animals, a result that is consistent with external literature (Nishida and Ochman, 2018).

3.3.2 Exploring microbial diversity between animal host subjects

Throughout this section we explore microbial diversity between the host lineages using the weighted Unifrac distance metric, a relative abundance sensitive metric (Lozupone and Knight, 2005). This dissimilarity metric allowed us to take into consideration phylogenetic relatedness between bacterial taxa, providing a more accurate assessment of host species similarity at the level of their individual gut microbiomes.

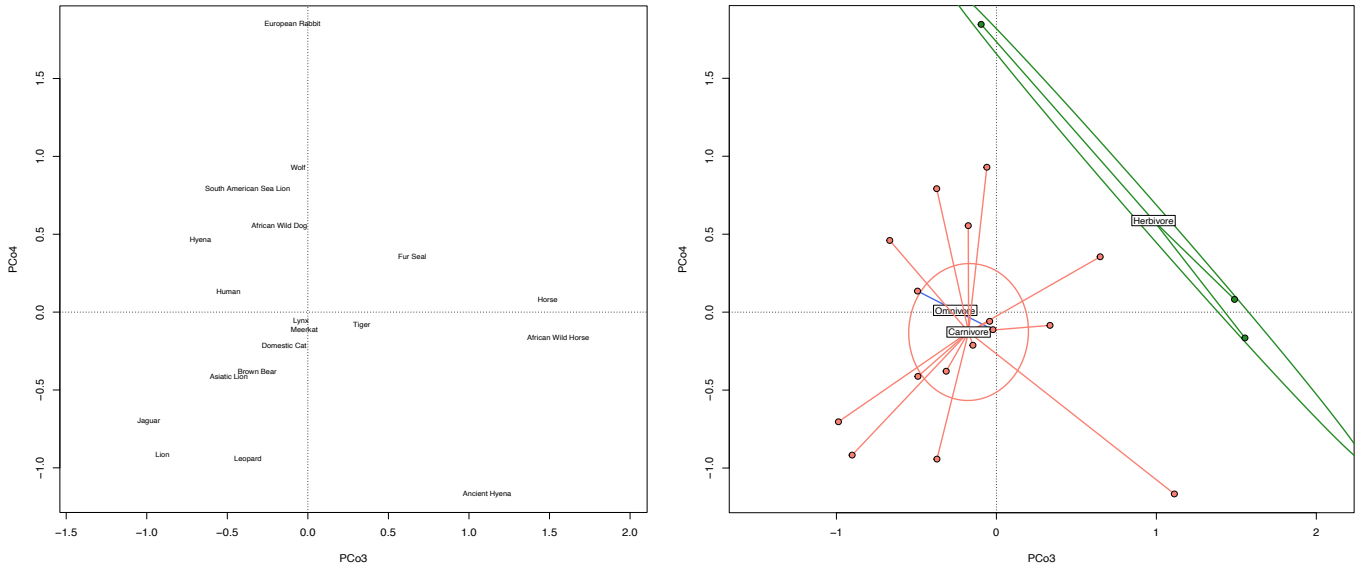

Figure 5: PCoA based on weighted UniFrac distances using the bacterial phylogenetic tree (contains 8,685 16S rRNA gene references aligned across all host samples used within this study). In the plot on the right the various diets can be seen with the corresponding animal host shown in the plot on the left. The ellipses indicate the regions from the centroids having 95% confidence level. The separation between the dietary clusters was seen to be non-significant (P=0.073 for a permutational test using pseudo-F ratio).

Analysing the host species within PCoA space in Figure 5 it appeared that was diet-related clustering whereby carnivorous and omnivorous taxa appear to cluster together whilst herbivorous taxa remained separate. However, when assessing the significance of this clustering or separation it was shown to be non-significant (P=0.073 for a permutational test using pseudo-F ratio) when omnivorous and carnivorous diets were treated as equivalent and appeared to be less significant when diets were treated as unique (P-value of P=0.149 for a permutational test using pseudo-F ratio). A result potentially attributed to shared microbial taxa between the various animal samples therefore reducing the degree of separation seen between the dietary groups.

At individual host species level, the extinct paleosubspecies C. spelaea appeared to be positioned between the herbivorous and carnivorous-omnivorous clusters rather than within or in proximity to the carnivorous-omnivorous cluster as was expected. The lack of observed clustering similarity to the dietary groups coupled with its location within PCoA space suggests that C. spelaea hosts microbial components that are blurred between the diets rather than having bacterial taxa largely associated to one diet or the other. A feature that was observed when looking at the bacterial OTUs present within the extant lineage both in Figure 2 and Figure 3.

The carnivorous lineages showed a large degree of variation represented by their positioning outside of their diet related confidence interval (95%) suggesting a large degree of gut microbial variation between the different lineages, a feature further explored in (Zoelzer et al., 2021), and reiterated by the dissimilarities seen within Figure 3. The large carnivorous cat lineages i.e., Jaguar, Lion, Leopard and Asiatic lion are seen to cluster together (bottom left corner of the left plot) in Figure 5 whereas the smaller cat lineages i.e., Lynx and the Domestic cat appear to cluster more with the omnivorous Meerkat sample than with the larger Feliformia of the suborder. Interestingly, the African Spotted Hyena appeared to cluster closer to the African wild dog than with the large carnivorous cats with whom it shares suborder, and was seen to be distantly positioned from its ancestral paleosubspecies, recapitulating gut microbial dissimilarity (see Figure 2). Interestingly, the carnivorous taxa (Wolf, African Wild Dog) clustered together with seals, thus reiterating the existence of collaborating behaviors between these mammalian taxa (Milani et al., 2020). The omnivorous taxa were seen to cluster closer to the carnivorous animal lineages rather than that of the herbivorous mammals an did not appear to be situated between the herbivore-carnivore cluster which would've suggested that they harbored a blend of carnivorous and herbivorous taxa. However, this was believed to be attributed to the choice of omnivores i.e., Meerkat, Human, rather than a general feature associated to the omnivorous gut microbiome.

To better understand the clustering observed between the various dietary groups (see Figure 5) and the associated lack of significant separation, we then looked at the microbial taxa that are shared across all three host diets. By identifying the extent to which bacterial OTUs are shared between the dietary groups we hoped to have a better understanding of the observed clustering features. Within Figure 6 the venn diagram contains all identified 8,685 bacterial taxa.

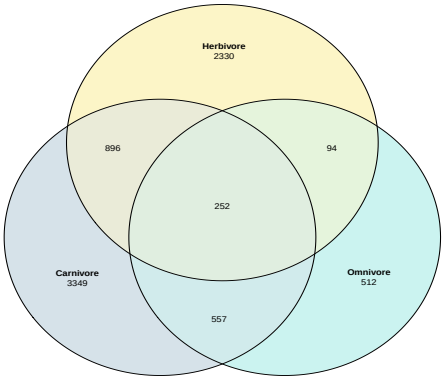

Figure 6: Venn diagram containing bacterial OTUs shared between the diets where total bacterial counts for each diet were seen to be carnivores: 5054, herbivores: 3572 and omnivores: 1415 respectively.

It became apparent that there was a relatively high number of shared bacterial taxa between the herbviorous and carnivorous animals (27% of bacterial taxa found in carnivores was shown to be found in herbivores) as well as the omnivorous and carnivorous animals (17% of bacterial taxa found in carnivores was shown to be found in omnivores) (see Figure 6), which can potentially explain the lack of significant clustering as a result of the high number of shared bacterial taxa between the groups. The number of shared bacterial taxa seen between the omnivorous and herbivorous lineages was significantly lower and reiterates the lack of clustering seen between the omnivorous and herbivorous mammals (see Figure 5), a result further indicated in Figure 2 when particularly looking at the Human and to a lesser extent Meerkat samples.

3.4 Phylosymbiotic relationships seen between seen between the animal hosts studied in the scope of this thesis

Within this section we made use of the program Count (see Annex), using a filtered table whereby a total of 563 bacterial taxa were retained (see The Count pipeline), in order to estimate the ancestral gut community compositions within a phylogenetic and probabilistic setting. This in turn allowed us to assess phylosymbiotic profiles belonging to the carnivorous species lineages used within this study, whereby higher microbial compositional similarity appears to exist between more closely related hosts vs the microbial composition seen in more distantly related hosts.

Here we graphically represent the reconstructed host phylogenetic tree generated using the mitochondrial DNA (mtDNA) belonging to each host (see Mitochondrial tree construction), whereby host lineages were clustered according to mtDNA similarity, meaning animal hosts that are positioned closer together are said to share a larger degree of phylogenetic relation. However, as mtDNA tree reconstruction was not time calibrated we were unable to make assumptions about the decay of phylosymbiotic signal over time. When we looked at the reconstructed mtDNA tree (Figure 7) there appeared to be a distinct partitioning of the mammalian lineages whereby one branch denoted the herbivorous mammalian clades whilst the other branch denoted the carnivorous clades. Furthermore, we noticed additional partitioning within the carnivorous split where further separation appeared to take place at the level of the mammalian suborders Feliformia and Caniformia. The partitioning was consistent with literature regarding the mammalian lineages belonging to the order Carnivora and its respective suborders (Eizirik et al., 2010).

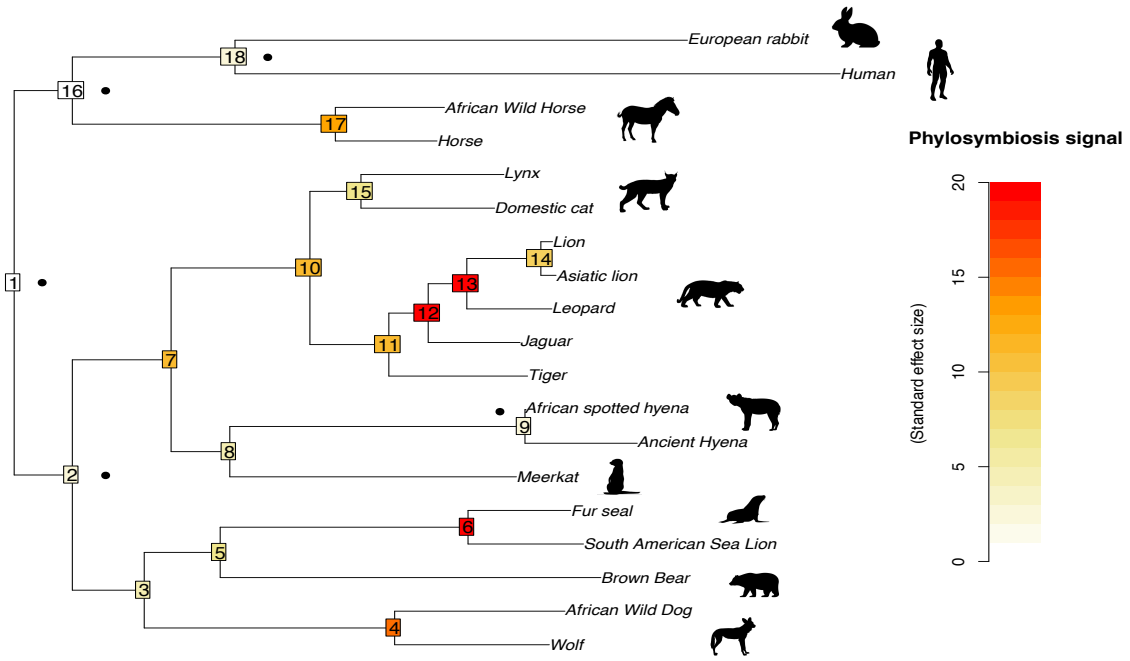

Figure 7: Reconstructed host mtDNA tree where mammalian phylogeny with internal nodes are coloured according to the degree with which they possess compositions of clade-specific bacteria (see Statistical evaluation of microbial taxa). Here the black dots denote mammalian clades that are shown to harbour no significant phylosymbiotic signal ($P>0.05$, Permutation test, 999 permutations). Standard effect size (SES) for a given ancestor is a measure of the corresponding phylosymbiotic signal

Generally it appeared as though the mammalian clades exhibited a large degree of compositional similarity with respect to their clade-specific microbial components, seen by the high standard effect size (SES) (Figure 7) along the ancestral nodes. Here all

mammalian clades that did not possess a black dot were said to host significant phylosymbiotic signals ($P<0.05$, permutation test, 999 permutations). As expected when moving backwards from the leaves of the tree we saw a reduction in the displayed phylosymbiotic strength at ancestral nodes, a pattern consistent with external literature (Groussin et al., 2017). This means that as we progress towards more phylogentically distant hosts we see a complementary drop in the associated phylosymbiotic signal. When focusing on the suborders Caniformia and Feliformia it was evident that the ancestral node shared between these two suborders (Node 2) lacked a significant phylosymbiotic signal, whereas the ancestral nodes within the two distinct suborder branches seemed to display significant phylosymbiotic strength (Nodes: 4,6,7,8,11,10,12,13,14,15).

Interestingly, the Ancient hyena sample was shown to display no significant phylosymbiotic signal at the common ancestral node (Node 9) shared with the African spotted hyena sample. However, the ancestral node (Node 8) that the two Crocuta subspecies shared with the Meerkat sample was seen to host a significant phylosymbiotic signal (see Figure 7), and as the tree suggests the Meerkat species (family *Herpestidae*) has been shown to be a sister family of the *Hyaenidae* family (Koepfli et al., 2006) and also predates on small mammalian animals. Moreover, a greater SES value, indicated by the appropriate colour shift, was displayed at the ancestral node (Node 7) separating the large carnivorous cats from the Hyena and Meerkat samples, ultimately indicating greater compositional similarity with respect to the microbial components shared between these families. In light of this phylosymbiotic signal it appears as though there is significant preservation of gut microbial components within the carnivorous lineages being maintained over evolutionary time-scales.

Within Table 3 we present 4 OTUs per ancestral node observed to have a posterior probability (>0.5) at the respective node. As result of their apparent preservation throughout the evolution of carnivorous species these OTUs potentially play a fundamental symbiotic role with respect to their host. Therefore they allow for a more informative gut microbial characterization of the host species used within this study. Assessing the significant phylosymbiotic ancestral nodes it became apparent there were bacterial genera and species that were seen to reoccur especially within the Carnivora lineages. At Genus level the genera *Peptoclostridium*, *Escherichia-Shigella*, *Clostridium sensu stricto 1* and *Lachnoclostridium* were seen to occur across several ancestral nodes and appear to be very prominent with high posterior probabilities (see Table 3). Furthermore, at species level we saw the recurrence of *Blautia sp. YL58* a species belonging to the family *Lachnospiraceae*, a dominant bacterial family (see Figure 2), which was observed at considerable frequency across the nodes. The bacterial OTUs and the recurrence across several ancestral carnivorous nodes potentially suggests they provide advantageous functionality at host-level and therefore are maintained across lineages.

Phylogenetic node	Bacterial OTUs	post prob
1	Lachnospiraceae NK4A136 group Genus	0.79172
1	Escherichia-Shigella Genus	0.67604
2	Blautia sp. YL58	0.99992
2	Escherichia-Shigella Genus	0.99989
2	Clostridium sensu stricto 1 Genus	0.99937
2	Peptoclostridium Genus	0.99861
3	Blautia sp. YL58	0.99992
3	Escherichia-Shigella Genus	0.99989
3	Clostridium sensu stricto 1 Genus	0.99937
3	Peptoclostridium Genus	0.99861
4	Escherichia-Shigella Genus	0.99901
4	CP015405.4239136.4240677	0.99863
4	Blautia sp. YL58	0.99818
4	Peptoclostridium Genus	0.99610
5	Blautia sp. YL58	0.99992
5	Escherichia-Shigella Genus	0.99989
5	Clostridium sensu stricto 1 Genus	0.99937
5	Peptoclostridium Genus	0.99861
6	Escherichia-Shigella Genus	0.99953
6	Blautia sp. YL58	0.99935
6	Peptoclostridium Genus	0.99813
6	Lachnospiraceae NK4A136 group Genus	0.99430
7	Blautia sp. YL58	0.99992
7	Escherichia-Shigella Genus	0.99989
7	Clostridium sensu stricto 1 Genus	0.99937
7	Peptoclostridium Genus	0.99861
8	Escherichia-Shigella Genus	0.99948
8	Blautia sp. YL58	0.99919
8	Clostridium sensu stricto 1 Genus	0.99853
8	Clostridium sp. Ade.TY	0.97521
9	Escherichia-Shigella Genus	0.99948
9	Blautia sp. YL58	0.99919
9	Clostridium sensu stricto 1 Genus	0.99853
9	Clostridium sp. Ade.TY	0.97521
10	Blautia sp. YL58	0.99992
10	Escherichia-Shigella Genus	0.99989
10	Clostridium sensu stricto 1 Genus	0.99940
10	Peptoclostridium Genus	0.99865
11	Blautia sp. YL58	0.99992
11	Escherichia-Shigella Genus	0.99989
11	Clostridium sensu stricto 1 Genus	0.99940
11	Peptoclostridium Genus	0.99865
12	Clostridium sensu stricto 1 Genus	0.99997
12	Blautia sp. YL58	0.99998
12	Peptoclostridium Genus	0.99997
12	Lachnoclostridium Genus	0.99985
13	Clostridium sensu stricto 1 Genus	0.99999
13	Blautia sp. YL58	0.99996
13	Peptoclostridium Genus	0.99997
13	Lachnoclostridium Genus	0.99985
14	Clostridium sensu stricto 1 Genus	0.99995
14	Blautia sp. YL58	0.99999
14	Peptoclostridium Genus	0.99997
14	Lachnoclostridium Genus	0.99984
15	Escherichia-Shigella Genus	0.99999
15	Blautia sp. YL58	0.99999
15	Clostridium sensu stricto 1 Genus	0.99999
15	Bacteroides ovatus	0.99999
16	Lachnospiraceae NK4A136 group Genus	0.79173
16	Escherichia-Shigella Genus	0.67604
17	Escherichia-Shigella Genus	0.99843
17	Bacteroides ovatus	0.99745
17	Prevotella Genus	0.99735
17	Lachnospiraceae NK3A20 group Genus	0.99735
18	Lachnospiraceae NK4A136 group Genus	0.94897
18	UCG-003 Genus	0.78736
18	Parabacteroides sp. CT06	0.77962
18	Odoribacter Genus	0.77962

Table 3: Bacterial OTUs having posterior probabilities >=0.5 across the various ancestral nodes (see Figure 5) where posterior probabilities were generated through the use of the Count program (see The Count pipeline).

As expected when comparing the SES values seen within Figure 7 we saw higher posterior probabilities at nodes associated with higher phylosymbiotic signals, a feature indicating more resemblance between host lineages within gut microbial context. Reiterating this pattern was the observed decrease in posterior probabilities found at more distant ancestral nodes i.e., Node 1 and 16, where only two bacterial OTUs were identified with posterior probabilities >0.5.

The inclusion of the ancestral node belonging to the herbivorous species is also shown within Table 3 (Node 17) in order to see if there was any similarity with the bacterial taxa observed within the carnivorous nodes as well as to identify potentially important bacterial OTUs in the herbivorous context. In accordance with what was previously observed (see Figure 3), we observed different bacterial taxa apart from the genus *Escherichia-Shigella*, which seems to be a common bacterial taxa found within the majority of nodes. The different bacterial OTUs observed between the carnivorous and herbivorous hosts reiterates the diversity we expected as a result of differing environmental, phylogenic and or dietary features. Assessment of the bacterial taxa at the ancestral node shared between the hyena subspecies and Meerkat sample (Node 8 within Figure 7) showed a high presence of the bacterial genus *Clostridium* which was shown to be present within the top 40 bacterial taxa (see Figure 3) when using relative abundance across these three samples as a metric. This genus is a member of the family *Clostridiaceae* which was seen to be particularly dominant with the African Spotted Hyena sample (see Figure 2) and therefore potentially suggests its importance within this carnivorous clade.

These bacterial OTUs along with the measures explored within the results of this thesis can provide insight into core microbial components belonging to carnivorous mammalian lineages, that are actively preserved throughout evolutionary timescales as a result of mutualistic relationship at host-microbe level. These OTUs are discussed (see Discussion) for potential functionality and significance within the carnivorous lineages.

4 Discussion

Within this thesis were able to bioinformatically assess the metagenome of extracted samples originating from the extinct paleosubspecies *Crocuta crocuta spelaea* (ca. 40,000 - 50,000 years ago). We reconstructed the ancient mitochondrial DNA reads using an mtDNA reference sequence belonging to the distal extant African Spotted Hyena species (*Crocuta crocuta*). The ancient reconstructed mtDNA along with the various mtDNA sequences belonging to the studied host species were then utilised to reconstruct a phylogenetic tree. Through the analysis of bacterial reads extracted from the respective host species coupled with the reconstructed phylogenetic tree, we were able to perform microbial characterisation in the context of important bacterial OTUs shown to be pre-

served over macro-evolutionary timescales. Furthermore, we identified important bacterial families seen consistently across the studied animal lineages and further classified these bacterial OTUs at higher taxonomic ranks providing potential insight into their functionality within their carnivorous animal hosts.

Through the family level microbial characterisation of the host species studied within the scope of this thesis we were able to identify bacterial families that were observed with high frequencies across the various carnivores, namely *Lachnospiraceae*, *Fusobacteriacea*, *Clostridiaceae*, *Bacteroidaceae*, *Peptostreptococcaceae* and *Coriobacteriaceae* (see Figure 2). In accordance with this, members of the bacterial family *Fusobacteriacea* are often seen within animal species that rely on a rich protein and high-fat based diet (Zoelzer et al., 2021) with the genus *Fusobacterium* thought to be a carnivore-specific bacteria (Milani et al., 2020; Zoelzer et al., 2021)(see Figure 3). This appears to be consistent with the results obtained within this thesis whereby herbivorous taxa were shown to harbor no members of the family and the only omnivore shown to possess this family is known to predate on small mammals (Doolan and Macdonald, 1996). The bacterial families *Clostridiaceae* and *Bacteroidaceae* were shown to be present across the carnivorous lineages (see Figure 2) and have been previously recorded to exist within the gastrointestinal tract of various carnivores (Milani et al., 2020; Zoelzer et al., 2021). The family *Bacteroidaceae* appear to coincide with fiber-rich diets and have be shown to be decoupled from protein intake, whilst the family *Clostridiaceae* seem to have an important role in protein metabolism (Kerr et al., 2013; Zoelzer et al., 2021). These bacterial families appear to be important constituents of the carnivorous gut microbiomes studied within this thesis and therefore can act as potential proxies for further microbial analysis.

The assessment of the bacterial OTUs that were observed to be significantly preserved across carnivorous lineages (posterior probability >0.5) allowed us to identify host-bacterial relationships that have previously been recorded, and therefore support previous findings (Vital et al., 2015). The characterisation of butyrate-synthesis pathways carried out in (Vital et al., 2015), allowed for further distinction between herbivorous and carnivorous animals at an enzymatic level whereby butyrate kinase (buk) was shown to be more abundant within many of the carnivorous animals. The Buk gene communities have been shown to be to linked to host species that that are correlated with protein-rich environments with the correlated adaptability of microbes that are able to synthesise butyrate to a carnivorous diet cite(Vital et al., 2015).

OTU	Family	Genus	Species	Posteriors
CCPS01000022.154.1916	Enterobacteriaceae	Escherichia-Shigella	Escherichia-Shigella Genus	17.041692
CP015405.4239136.4240677	Lachnospiraceae	Blautia	Blautia sp. YL58	13.995900
KC245386.1.1440	Peptostreptococcaceae	Peptoclostridium	Peptoclostridium Genus	13.930563
EU118957.1.1472	Clostridiaceae	Clostridium sensu stricto 1	Clostridium sensu stricto 1 Genus	12.991548
AVSV01000024.80.1573	Clostridiaceae	Clostridium sensu stricto 1	Clostridium sp. Ade.TY	12.703726
FMGA01000006.679.2193	Lachnospiraceae	Lachnoclostridium	Lachnoclostridium	12.130587
KC259922.1.1338	Clostridiaceae	Clostridium sensu stricto 1	Clostridium sensu stricto 1 Genus	11.734195
AM982561.1.1486	Peptostreptococcaceae	Terrisporobacter	Terrisporobacter Genus	11.658863
JFOF01000052.9721.11268	Peptostreptococcaceae	Clostridioides	Clostridioides difficile	11.499636
EU774619.1.1204	Peptostreptococcaceae	Romboutsia	Romboutsia Genus	11.411869
AB563237.1.1541	Streptococcaceae	Streptococcus	Streptococcus gallolyticus subsp. macedonicus	11.347924
AY581822.1.1517	Erysipelotrichaceae	Catenisphaera	Catenisphaera Genus	11.317032
KC245215.1.1442	Coriobacteriaceae	Collinsella	Collinsella Genus	10.692515
JQ207307.1.1339	Peptostreptococcaceae	Peptostreptococcus	Peptostreptococcus Genus	9.641696
CP015244.2920395.2921952	Enterobacteriaceae	Escherichia-Shigella	Escherichia coli O91 str. RM7190	9.598562
EU773733.1.1430	Peptococcaceae	Peptococcus	Peptococcus Genus	9.030022

Table 4: Prevalent bacterial OTUs observed across the ancestral nodes (see Figure 5) where bacterial OTUs were retained if their summed posterior probability was >=9 (which is equivalent to an OTU having a posterior probability of 0.5 across all 18 ancestral nodes). Posterior probabilities were generated through the use of the Count program (see The Count pipeline).

In accordance with (Vital et al., 2015), the bacterial genera *Lachnoclostridium*, *Clostridium*, which were seen to be both extremely prevalent across the carnivorous host species (see Figure 3) and highly preserved (see Table 4), have been shown to synthesise short-chain fatty acids (SCFAs), i.e., butyrate (Chen et al., 2021; Appert et al., 2020). The genus *Clostridium* belongs to the family *Clostridiaceae* which have been recorded to aid in the breakdown of both proteins and fats and therefore are consistently seen in carnivorous animals (Bermingham et al., 2017). Moreover, the genus *Clostridium sensu stricto 1* includes members that have been recorded to employ the buk-pathway (Vital et al., 2015; Appert et al., 2020). Furthermore, this genus is considered a beneficial bacteria and is suggested to have the capacity to provide energy to the intestinal cells and assist in the protection of the gut barrier (Kong et al., 2019). The prevalence with which we see these bacterial OTUs potentially reiterates the significance of the butyrate-synthesis pathways observed within carnivorous lineages. Another genus observed to grow within a carnivorous dietary setting was *Peptoclostridium*, which is known proteolytic bacteria and has been shown to increase in diets that consumed red meats (Becker et al., 2021; Xiao et al., 2021). Both *Peptoclostridium* and *Clostridium sensu stricto 1* belong to the group *Clostridia* and take part in amino acid fermentation, and therefore aid in the digestion of meat (Zhu et al., 2021). In addition to this it is believed that *Peptostreptococcaceae*, the family to which *Peptoclostridium* belongs, is thought to play a role in

gut homeostasis advocated by an observed higher presence within healthier animals (Zhu et al., 2021).

Interestingly, it has been shown that the relative abundance of many species belonging to the genus *Blautia* tend to increase as a result of higher SCFAs concentration levels (Liu et al.), and this feature is potentially supported by the consistent coupling with the before mentioned butyrate producers (see Table 3). Furthermore, the *Blautia sp. YL58* strain, which was observed in multiple carnivorous lineages within this study (see Figure 3 and Table 3) has been reported as a protective bacterial agent in the context of various cancers (Flemer et al., 2018; Serino, 2019; Shi et al., 2023), therefore providing advantageous functionality with respect to its colonized hosts. It appears as though bacteria that have the capacity for SCFA synthesis are vital within the GM of carnivorous animals and have further been linked to a number of health benefits, such as the protection of cardiovascular, anti-inflammatory and immuno-regulatory systems (Xiong et al., 2022), reemphasising their importance. Lastly, the species *Escherichia coli* which has been recorded to be highly correlated with butyrate-kinase taxa (Vital et al., 2015) is a known sister species to those constituting the genus *Shigella* and therefore this study might indicate that there is additional correlation between *Shigella* and buk-related taxa, due to the extreme prevalence with which we see it, however additional research is required.

The work conducted within this thesis highlights specific bacterial OTUs that appear to be maintained as a result of their ability to synthesis short-chain fatty acids (SCFAs), particularly butyrate which can then undergo rapid oxidisation after which it can be utilized for the production of cellular energy (Salvi and Cowles, 2021). This thesis reiterates the importance of these bacterial OTUs by the prevalence and relative abundance with which we see them across the various carnivorous animal lineages studied with the scope of this work. Furthermore, we have identified significantly recurring taxa (see Table 4) that can be further analysed for their relationship within or outside of the metabolic pathways shaping the plethora of animal species. Their presence within the ancient C. spelaea sample suggests (see Figure 3) that there has been evolutionary preservation of both butyrate producing and proteolytic bacteria that have been shown to share a correlation with protein rich diets (Vital et al., 2015), and therefore coupled with the presence of these bacteria within extant carnivorous lineages, provides insight into the microbial components that are being conserved over evolutionary timescales.

4.0.1 Limitations

Previous microbial characterisation of animals samples taken from captivity showed higher relative abundance levels with respect to certain bacterial genera (*Clostridium* and *Streptococcus luteciae*) (McKenzie et al., 2017). As all animal species within this

work (excluding both extinct and extant hyena samples) were captive, this work should studied within the context of captive animals. Furthermore, metagenomic samples were generated using different techniques i.e., 16S rRNA and shotgun, where shotgun has been known to detect identify less abundant taxa (Durazzi et al., 2021) compared to 16S rRNA, however when assessing core microbial taxa the two methods provide sufficient microbial resolution.

References

O. Appert, A. R. Garcia, R. Frei, C. Roduit, F. Constancias, V. Neuzil-Bunesova, R. Ferstl, J. Zhang, C. Akdis, R. Lauener, C. Lacroix, and C. Schwab. Initial butyrate producers during infant gut microbiota development are endospore formers. *Environmental Microbiology*, 22(9):3909–3921, 2020. ISSN 1462-2920. doi: 10.1111/1462-2920.15167. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.15167. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.15167.

E. Baquedano, C. Laplana, J. L. Arsuaga, R. Huguet Pàmies, B. Márquez, and A. Pérez-González. Selection of cave shelter by Neanderthals (Homo neanderthalensis) and spotted hyaenas (Crocuta crocuta) at the Calvero de la Higuera sites (Pinilla del Valle, Madrid Region, Spain). 2016. ISSN 2341-2496. URL https://cir.cenieh.es/handle/20.500.12136/1301. Accepted: 2019-06 Publisher: Universidad de Alcalá.

A. Bayané and S. R. Guiot. Animal digestive strategies versus anaerobic digestion bioprocesses for biogas production from lignocellulosic biomass. *Reviews in Environmental Science and Bio/Technology*, 10(1):43–62, Mar. 2011. ISSN 1572-9826. doi: 10.1007/s11157-010-9209-4. URL https://doi.org/10.1007/s11157-010-9209-4.

M. Baynes-Rock. Converging on Ancient Bones: A Review of the Evidence for the Close Relatedness of Humans (Homo sapiens) and Spotted Hyenas (Crocuta crocuta). *Humanimalia*, 7(1):1–22, Oct. 2015. ISSN 2151-8645. doi: 10.52537/humanimalia.9980. URL https://humanimalia.org/article/view/9980. Number: 1.

A. A. M. J. Becker, K. C. Hill, and P. Butaye. Unraveling the Gut Microbiome of the Invasive Small Indian Mongoose (Urva auropunctata) in the Caribbean. *Microorganisms*, 9(3):465, Mar. 2021. ISSN 2076-2607. doi: 10.3390/microorganisms9030465. URL https://www.mdpi.com/2076-2607/9/3/465. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

S. Behjati and P. S. Tarpey. What is next generation sequencing? *Archives of Disease in Childhood. Education and Practice Edition*, 98(6):236–238, Dec. 2013. ISSN 1743-0585. doi: 10.1136/archdischild-2013-304340. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3841808/.

E. N. Bermingham, P. Maclean, D. G. Thomas, N. J. Cave, and W. Young. Key bacterial families (Clostridiaceae, Erysipelotrichaceae and Bacteroidaceae) are related to the digestion of protein and energy in dogs. *PeerJ*, 5:e3019, Mar. 2017. ISSN 2167-8359. doi: 10.7717/peerj.3019. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5337088/.

E. S. Cameron, P. J. Schmidt, B. J.-M. Tremblay, M. B. Emelko, and K. M. Müller. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Scientific Reports*, 11(1):22302, Nov. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-01636-1. URL https://www.nature.com/articles/s41598-021-01636-1. Number: 1 Publisher: Nature Publishing Group.

S. Capella-Gutiérrez, J. M. Silla-Martínez, and T. Gabaldón. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25 (15):1972–1973, Aug. 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp348. URL https://doi.org/10.1093/bioinformatics/btp348.

C. Carøe, S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, and M. T. P. Gilbert. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, 9 (2):410–419, 2018. ISSN 2041-210X. doi: 10.1111/2041-210X.12871. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12871. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12871.

L. Chen, X. Zhou, Y. Wang, D. Wang, Y. Ke, and X. Zeng. Propionate and Butyrate Produced by Gut Microbiota after Probiotic Supplementation Attenuate Lung Metastasis of Melanoma Cells in Mice. *Molecular Nutrition & Food Research*, 65(15):2100096, 2021. ISSN 1613-4133. doi: 10.1002/mnfr.202100096. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mnfr.202100096. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mnfr.202100096.

N. de Jonge, B. Carlsen, M. H. Christensen, C. Pertoldi, and J. L. Nielsen. The Gut Microbiome of 54 Mammalian Species. *Frontiers in Microbiology*, 13, 2022. ISSN 1664-302X. URL https://www.frontiersin.org/articles/10.3389/fmicb.2022.886252.

S. P. Doolan and D. W. Macdonald. Diet and foraging behaviour of group-living meerkats, Suricata suricatta, in the southern Kalahari. *Journal of Zoology*, 239(4):697–716, 1996. ISSN 1469-7998. doi: 10.1111/j.1469-7998.1996.tb05472.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7998.1996.tb05472.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-7998.1996.tb05472.x.

F. Durazzi, C. Sala, G. Castellani, G. Manfreda, D. Remondini, and A. De Cesare. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, 11(1):3030, Feb. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-82726-y. URL https://www.nature.com/articles/s41598-021-82726-y. Number: 1 Publisher: Nature Publishing Group.

E. Eizirik, W. J. Murphy, K.-P. Koepfli, W. E. Johnson, J. W. Dragoo, R. K. Wayne, and S. J. O'Brien. Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, 56(1):49–63, July 2010. ISSN 1055-7903. doi: 10.1016/j.ympev.2010.01.033. URL https://www.sciencedirect.com/science/article/pii/S1055790310000448.

B. Flemer, R. D. Warren, M. P. Barrett, K. Cisek, A. Das, I. B. Jeffery, E. Hurley, M. O'Riordain, F. Shanahan, and P. W. O'Toole. The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*, 67(8):1454–1463, Aug. 2018. ISSN 0017-5749, 1468-3288. doi: 10.1136/gutjnl-2017-314814. URL https://gut.bmj.com/content/67/8/1454. Publisher: BMJ Publishing Group Section: Gut microbiota.

S. F. Gilbert, J. Sapp, and A. I. Tauber. A Symbiotic View of Life: We Have Never Been Individuals. *The Quarterly Review of Biology*, 87(4):325–341, 2012. doi: 10.1086/668166. URL https://doi.org/10.1086/668166. _eprint: https://doi.org/10.1086/668166.

S. R. Gill, M. Pop, R. T. DeBoy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, and K. E. Nelson. Metagenomic Analysis of the Human Distal Gut Microbiome. *Science (New York, N.Y.)*, 312(5778):1355–1359, June 2006. ISSN 0036-8075. doi: 10.1126/science.1124234. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3027896/.

J. L. Gittleman. *Carnivore Behavior, Ecology, and Evolution*. Springer Science & Business Media, Mar. 2013. ISBN 978-1-4757-4716-4. Google-Books-ID: 7OfgBwAAQBAJ.

R. Gregor, M. Probst, S. Eyal, A. Aksenov, G. Sasson, I. Horovitz, P. C. Dorrestein, M. M. Meijler, and I. Mizrahi. Mammalian gut metabolomes mirror microbiome composition and host phylogeny. *The ISME Journal*, 16(5):1262–1274, May 2022. ISSN 1751-7370. doi: 10.1038/s41396-021-01152-0. URL https://www.nature.com/articles/s41396-021-01152-0.

M. Groussin, F. Mazel, J. G. Sanders, C. S. Smillie, S. Lavergne, W. Thuiller, and E. J. Alm. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, 8(1):14319, Feb. 2017. ISSN 2041-1723. doi: 10.1038/ncomms14319. URL https://www.nature.com/articles/ncomms14319.

H. R. Gruber-Vodicka, B. K. B. Seah, and E. Pruesse. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*, 5(5):e00920–20, Oct. 2020. doi: 10.1128/mSystems.00920-20. URL https://journals.asm.org/doi/full/10.1128/mSystems.00920-20. Publisher: American Society for Microbiology.

S. Gupta, M. S. Mortensen, S. Schjørring, U. Trivedi, G. Vestergaard, J. Stokholm, H. Bisgaard, K. A. Krogfelt, and S. J. Sørensen. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Communications Biology*, 2(1):1–7, Aug. 2019. ISSN 2399-3642. doi: 10.1038/s42003-019-0540-1. URL https://www.nature.com/articles/s42003-019-0540-1. Number: 1 Publisher: Nature Publishing Group.

M. Harzhauser, A. Kroh, O. Mandic, W. E. Piller, U. Göhlich, M. Reuter, and B. Berning. Biogeographic responses to geodynamics: A key study all around the Oligo–Miocene Tethyan Seaway. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, 246(4): 241–256, Dec. 2007. ISSN 00445231. doi: 10.1016/j.jcz.2007.05.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0044523107000186.

K. E. Holekamp, S. T. Sakai, and B. L. Lundrigan. Social intelligence in the spotted hyena (Crocuta crocuta). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):523–538, Feb. 2007. doi: 10.1098/rstb.2006.1993. URL https://royalsocietypublishing.org/doi/full/10.1098/rstb.2006.1993. Publisher: Royal Society.

J. Hu, M. V. Westbury, J. Yuan, Z. Zhang, S. Chen, B. Xiao, X. Hou, H. Ji, X. Lai, M. Hofreiter, and G. Sheng. Ancient mitochondrial genomes from Chinese cave hyenas provide insights into the evolutionary history of the genus Crocuta. *Proceedings of the Royal Society B: Biological Sciences*, 288(1943):20202934, Jan. 2021. doi: 10.1098/rspb.2020.2934. URL https://royalsocietypublishing.org/doi/full/10.1098/rspb.2020.2934. Publisher: Royal Society.

R. Hübler, F. Key, C. Warinner, K. Bos, J. Krause, and A. Herbig. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biology*, 20(1), 2019. ISSN 1474-7596. doi: 10.1186/s13059-019-1903-0.

K. R. Kerr, G. Forster, S. E. Dowd, E. P. Ryan, and K. S. Swanson. Effects of Dietary Cooked Navy Bean on the Fecal Microbiome of Healthy Companion Dogs. *PLOS ONE*, 8(9):e74998, Sept. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0074998. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074998. Publisher: Public Library of Science.

C. H. King, H. Desai, A. C. Sylvetsky, J. LoTempio, S. Ayanyan, J. Carrie, K. A. Crandall, B. C. Fochtman, L. Gasparyan, N. Gulzar, P. Howell, N. Issa, K. Krampis, L. Mishra, H. Morizono, J. R. Pisegna, S. Rao, Y. Ren, V. Simonyan, K. Smith, S. VedBrat, M. D. Yao, and R. Mazumder. Baseline human gut microbiota profile in healthy people and standard reporting template. *PLOS ONE*, 14(9):e0206484, Sept. 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0206484. URL https://journals.

plos.org/plosone/article?id=10.1371/journal.pone.0206484. Publisher: Public Library of Science.

K.-P. Koepfli, S. M. Jenks, E. Eizirik, T. Zahirpour, B. Van Valkenburgh, and R. K. Wayne. Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Molecular Phylogenetics and Evolution*, 38(3): 603–620, Mar. 2006. ISSN 1055-7903. doi: 10.1016/j.ympev.2005.10.017.

M. H. Kogut, A. Lee, and E. Santin. Microbiome and pathogen interaction with the immune system. *Poultry Science*, 99(4):1906–1913, Apr. 2020. ISSN 1525-3171. doi: 10.1016/j.psj.2019.12.011.

C. Kong, R. Gao, X. Yan, L. Huang, and H. Qin. Probiotics improve gut microbiota dysbiosis in obese mice fed a high-fat or high-sucrose diet. *Nutrition*, 60:175–184, Apr. 2019. ISSN 0899-9007. doi: 10.1016/j.nut.2018.10.002. URL https://www.sciencedirect.com/science/article/pii/S0899900718306324.

M. E. Lewis and L. Werdelin. A revision of the genus Crocuta (Mammalia, Hyaenidae). *Palaeontographica Abteilung A*, pages 1–115, Apr. 2022. ISSN ,. doi: 10.1127/pala/2022/0120. URL https://www.schweizerbart.de/papers/pala/detail/322/101138/A_revision_of_the_genus_Crocuta_Mammalia_Hyaenidae?af=crossref. Publisher: Schweizerbart'sche Verlagsbuchhandlung.

R. E. Ley, M. Hamady, C. Lozupone, P. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon. Evolution of mammals and their gut microbes. *Science (New York, N.Y.)*, 320(5883):1647–1651, June 2008. ISSN 0036-8075. doi: 10.1126/science.1155725. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2649005/.

H. Li and R. Durbin. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760, July 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp324. URL https://doi.org/10.1093/bioinformatics/btp324.

S. J. Lim and S. R. Bordenstein. An introduction to phylosymbiosis. *Proceedings of the Royal Society B: Biological Sciences*, 287(1922):20192900, Mar. 2020. doi: 10.1098/rspb.2019.2900. URL https://royalsocietypublishing.org/doi/10.1098/rspb.2019.2900.

A. L. Lind and K. S. Pollard. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome*, 9(1):58, Mar. 2021. ISSN 2049-2618. doi: 10.1186/s40168-021-01015-y. URL https://doi.org/10.1186/s40168-021-01015-y.

S. Lindgreen. AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Research Notes*, 5, 2012. ISSN 1756-0500. doi: 10.1186/1756-0500-5-337.

X. Liu, B. Mao, J. Gu, J. Wu, S. Cui, G. Wang, J. Zhao, H. Zhang, and W. Chen. Blautia—a new functional genus with potential probiotic properties? *Gut Microbes*, 13(1):1875796. ISSN 1949-0976. doi: 10.1080/19490976.2021.1875796. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7872077/`.

C. Lozupone and R. Knight. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, 71(12):8228–8235, Dec. 2005. doi: 10.1128/AEM.71.12.8228-8235.2005. URL `https://journals.asm.org/doi/10.1128/AEM.71.12.8228-8235.2005`. Publisher: American Society for Microbiology.

F. Magne, M. Gotteland, L. Gauthier, A. Zazueta, S. Pesoa, P. Navarrete, and R. Balamurugan. The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in Obese Patients? *Nutrients*, 12(5):1474, May 2020. ISSN 2072-6643. doi: 10.3390/nu12051474.

E. K. Mallott and K. R. Amato. Host specificity of the gut microbiome. *Nature Reviews Microbiology*, 19(10):639–653, Oct. 2021. ISSN 1740-1534. doi: 10.1038/s41579-021-00562-3. URL `https://www.nature.com/articles/s41579-021-00562-3`. Number: 10 Publisher: Nature Publishing Group.

V. J. McKenzie, S. J. Song, F. Delsuc, T. L. Prest, A. M. Oliverio, T. M. Korpita, A. Alexiev, K. R. Amato, J. L. Metcalf, M. Kowalewski, N. L. Avenant, A. Link, A. Di Fiore, A. Seguin-Orlando, C. Feh, L. Orlando, J. R. Mendelson, J. Sanders, and R. Knight. The Effects of Captivity on the Mammalian Gut Microbiome. *Integrative and Comparative Biology*, 57(4):690–704, Oct. 2017. ISSN 1540-7063. doi: 10.1093/icb/icx090. URL `https://doi.org/10.1093/icb/icx090`.

C. J. Meehan and R. G. Beiko. A Phylogenomic View of Ecological Specialization in the Lachnospiraceae, a Family of Digestive Tract-Associated Bacteria. *Genome Biology and Evolution*, 6(3):703–713, Mar. 2014. ISSN 1759-6653. doi: 10.1093/gbe/evu050. URL `https://doi.org/10.1093/gbe/evu050`.

C. Milani, G. Alessandri, L. Mancabelli, M. Mangifesta, G. A. Lugli, A. Viappiani, G. Longhi, R. Anzalone, S. Duranti, F. Turroni, M. C. Ossiprandi, D. van Sinderen, and M. Ventura. Multi-omics Approaches To Decipher the Impact of Diet and Host Physiology on the Mammalian Gut Microbiome. *Applied and Environmental Microbiology*, 86(23):e01864–20, Nov. 2020. ISSN 0099-2240. doi: 10.1128/AEM.01864-20. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7657629/`.

T. M. Nelson, T. L. Rogers, and M. V. Brown. The Gut Bacterial Community of Mammals from Marine and Terrestrial Habitats. *PLOS ONE*, 8(12):e83655, Dec. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0083655. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083655. Publisher: Public Library of Science.

A. H. Nishida and H. Ochman. Rates of Gut Microbiome Divergence in Mammals. *Molecular ecology*, 27(8):1884, Apr. 2018. doi: 10.1111/mec.14473. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5935551/. Publisher: NIH Public Access.

J. M. Pickard, M. Y. Zeng, R. Caruso, and G. Núñez. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunological Reviews*, 279(1):70–89, Sept. 2017. ISSN 1600-065X. doi: 10.1111/imr.12567.

M. N. Price, P. S. Dehal, and A. P. Arkin. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26(7):1641–1650, July 2009. ISSN 0737-4038. doi: 10.1093/molbev/msp077. URL https://doi.org/10.1093/molbev/msp077.

E. Pruesse, J. Peplies, and F. O. Glöckner. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28(14):1823–1829, July 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts252. URL https://doi.org/10.1093/bioinformatics/bts252.

S. Rampelli, S. Turroni, F. Debandi, A. Alberdi, S. L. Schnorr, C. A. Hofman, A. Taddia, R. Helg, E. Biagi, P. Brigidi, F. D'Amico, M. Cattani, and M. Candela. The gut microbiome buffers dietary adaptation in Bronze Age domesticated dogs. *iScience*, 24 (8):102816, Aug. 2021. ISSN 2589-0042. doi: 10.1016/j.isci.2021.102816.

G. Renaud, V. Slon, A. Duggan, and J. Kelso. Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*, 16(1), 2015. ISSN 1474-7596. doi: 10.1186/s13059-015-0776-0.

N. Rohland, J. L. Pollack, D. Nagel, C. Beauval, J. Airvaux, S. Pääbo, and M. Hofreiter. The Population History of Extant and Extinct Hyenas. *Molecular Biology and Evolution*, 22(12):2435–2443, Dec. 2005. ISSN 0737-4038. doi: 10.1093/molbev/msi244. URL https://doi.org/10.1093/molbev/msi244.

C. A. Rojas, K. E. Holekamp, M. Viladomat Jasso, V. Souza, J. A. Eisen, and K. R. Theis. Taxonomic, Genomic, and Functional Variation in the Gut Microbiomes of Wild Spotted Hyenas Across 2 Decades of Study. *mSystems*, 0(0):e00965–22, Dec. 2022. doi: 10.1128/msystems.00965-22. URL https://journals.asm.org/doi/full/10.1128/msystems.00965-22. Publisher: American Society for Microbiology.

P. S. Salvi and R. A. Cowles. Butyrate and the Intestinal Epithelium: Modulation of Proliferation and Inflammation in Homeostasis and Disease. *Cells*, 10(7):1775, July 2021. ISSN 2073-4409. doi: 10.3390/cells10071775. URL https://www.mdpi.com/2073-4409/10/7/1775. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

M. Serino. SCFAs — the thin microbial metabolic line between good and bad. *Nature Reviews Endocrinology*, 15(6):318–319, June 2019. ISSN 1759-5037. doi: 10.1038/s41574-019-0205-7. URL https://www.nature.com/articles/s41574-019-0205-7. Number: 6 Publisher: Nature Publishing Group.

F. Shi, G. Liu, Y. Lin, C. l. Guo, J. Han, E. S. H. Chu, C. Shi, Y. Li, H. Zhang, C. Hu, R. Liu, S. He, G. Guo, Y. Chen, X. Zhang, O. O. Coker, S. H. Wong, J. Yu, and J. She. Altered gut microbiome composition by appendectomy contributes to colorectal cancer. *Oncogene*, 42(7):530–540, Feb. 2023. ISSN 1476-5594. doi: 10.1038/s41388-022-02569-3. URL https://www.nature.com/articles/s41388-022-02569-3. Number: 7 Publisher: Nature Publishing Group.

N.-R. Shin, T. W. Whon, and J.-W. Bae. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology*, 33(9):496–503, Sept. 2015. ISSN 0167-7799. doi: 10.1016/j.tibtech.2015.06.011. URL https://www.sciencedirect.com/science/article/pii/S0167779915001390.

G. Smant, J. P. W. G. Stokkermans, Y. Yan, J. M. de Boer, T. J. Baum, X. Wang, R. S. Hussey, F. J. Gommers, B. Henrissat, E. L. Davis, J. Helder, A. Schots, and J. Bakker. Endogenous cellulases in animals: Isolation of -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(9):4906–4911, Apr. 1998. ISSN 0027-8424. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC20186/.

E. A. Sperling, C. A. Frieder, A. V. Raman, P. R. Girguis, L. A. Levin, and A. H. Knoll. Oxygen, ecology, and the Cambrian radiation of animals. *Proceedings of the National Academy of Sciences*, 110(33):13446–13451, Aug. 2013. doi: 10.1073/pnas.1312778110. URL https://www.pnas.org/doi/10.1073/pnas.1312778110. Publisher: Proceedings of the National Academy of Sciences.

E. E. Spinapolice, A. Zerboni, M. C. Meyer, S. Talamo, G. S. Mariani, L. A. Gliganic, L. Buti, M. Fusco, M. P. Maiorano, S. Silvestrini, R. Sorrentino, A. Vazzana, M. Romandini, A. Fiorini, A. Curci, and S. Benazzi. Back to Uluzzo – archaeological, palaeoenvironmental and chronological context of the Mid–Upper Palaeolithic sequence at Uluzzo C Rock Shelter (Apulia, southern Italy). *Journal of Quaternary Science*, 37(2):217–234, 2022. ISSN 1099-1417. doi: 10.1002/

jqs.3349. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jqs.3349. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jqs.3349.

M. Steinthorsdottir, H. K. Coxall, A. M. de Boer, M. Huber, N. Barbolini, C. D. Bradshaw, N. J. Burls, S. J. Feakins, E. Gasson, J. Henderiks, A. E. Holbourn, S. Kiel, M. J. Kohn, G. Knorr, W. M. Kürschner, C. H. Lear, D. Liebrand, D. J. Lunt, T. Mörs, P. N. Pearson, M. J. Pound, H. Stoll, and C. a. E. Strömberg. The Miocene: The Future of the Past. *Paleoceanography and Paleoclimatology*, 36 (4):e2020PA004037, 2021. ISSN 2572-4525. doi: 10.1029/2020PA004037. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2020PA004037. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020PA004037.

C. E. Stevens and I. D. Hume. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological Reviews*, 78(2):393–427, Apr. 1998. ISSN 0031-9333. doi: 10.1152/physrev.1998.78.2.393.

M. C. Stiner. Comparative ecology and taphonomy of spotted hyenas, humans, and wolves in Pleistocene Italy. 2021.

P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. The human microbiome project. *Nature*, 449(7164):804–810, Oct. 2007. ISSN 1476-4687. doi: 10.1038/nature06244.

B. Van Valkenburgh. Déjà vu: the evolution of feeding morphologies in the Carnivora. *Integrative and Comparative Biology*, 47(1):147–163, July 2007. ISSN 1540-7063. doi: 10.1093/icb/icm016. URL https://doi.org/10.1093/icb/icm016.

M. Vital, J. Gao, M. Rizzo, T. Harrison, and J. M. Tiedje. Diet is a major factor governing the fecal butyrate-producing community structure across Mammalia, Aves and Reptilia. *The ISME Journal*, 9(4):832–843, Apr. 2015. ISSN 1751-7362. doi: 10.1038/ismej.2014.179. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4817703/.

J. Vågene, A. Herbig, M. G. Campana, N. M. Robles García, C. Warinner, S. Sabin, M. A. Spyrou, A. Andrades Valtueña, D. Huson, N. Tuross, K. I. Bos, and J. Krause. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, 2(3):520–528, Mar. 2018. ISSN 2397-334X. doi: 10.1038/s41559-017-0446-6. URL https://www.nature.com/articles/s41559-017-0446-6. Number: 3 Publisher: Nature Publishing Group.

C. Warinner, C. Speller, M. J. Collins, and C. M. Lewis. Ancient human microbiomes. *Journal of human evolution*, 0:125–136, Feb. 2015. ISSN 0047-2484. doi: 10.1016/j.jhevol.2014.10.016. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4312737/.

G. D. Wesley-Hunt and J. J. Flynn. Phylogeny of the carnivora: Basal relationships among the carnivoramorphans, and assessment of the position of 'miacoidea' relative to carnivora. *Journal of Systematic Palaeontology*, 3(1):1–28, Jan. 2005. ISSN 1477-2019. doi: 10.1017/S1477201904001518. URL https://doi.org/10.1017/S1477201904001518. Publisher: Taylor & Francis _eprint: https://doi.org/10.1017/S1477201904001518.

M. V. Westbury, S. Hartmann, A. Barlow, M. Preick, B. Ridush, D. Nagel, T. Rathgeber, R. Ziegler, G. Baryshnikov, G. Sheng, A. Ludwig, I. Wiesel, L. Dalen, F. Bibi, L. Werdelin, R. Heller, and M. Hofreiter. Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Science Advances*, 6(11):eaay0456, Mar. 2020. doi: 10.1126/sciadv.aay0456. URL https://www.science.org/doi/10.1126/sciadv.aay0456. Publisher: American Association for the Advancement of Science.

X. Wu, Q. Wei, X. Wang, Y. Shang, and H. Zhang. Evolutionary and dietary relationships of wild mammals based on the gut microbiome. *Gene*, 808:145999, Jan. 2022. ISSN 0378-1119. doi: 10.1016/j.gene.2021.145999. URL https://www.sciencedirect.com/science/article/pii/S0378111921005941.

T. Xiao, T. Liang, D.-H. Geng, L. Wang, L. Liu, X. Zhou, H. Pu, J. Huang, S. Zhou, and L.-T. Tong. Dietary Proteins Alter Fermentation Characteristics of Human Gut Microbiota In Vitro. *Plant Foods for Human Nutrition*, 76(4):419–426, Dec. 2021. ISSN 1573-9104. doi: 10.1007/s11130-020-00836-w. URL https://doi.org/10.1007/s11130-020-00836-w.

R.-G. Xiong, D.-D. Zhou, S.-X. Wu, S.-Y. Huang, A. Saimaiti, Z.-J. Yang, A. Shang, C.-N. Zhao, R.-Y. Gan, and H.-B. Li. Health Benefits and Side Effects of Short-Chain Fatty Acids. *Foods (Basel, Switzerland)*, 11(18):2863, Sept. 2022. ISSN 2304-8158. doi: 10.3390/foods11182863.

L. Zhu, Q. Wu, J. Dai, S. Zhang, and F. Wei. Evidence of cellulose metabolism by the giant panda gut microbiome. *Proceedings of the National Academy of Sciences*, 108(43):17714–17719, Oct. 2011. doi: 10.1073/pnas.1017956108. URL https://www.pnas.org/doi/10.1073/pnas.1017956108. Publisher: Proceedings of the National Academy of Sciences.

Y. Zhu, Z. Han, H. Wang, C. Liu, H. Si, and C. Xu. Adaptation of the Gut Microbiota of Amur Tigers to a Special Diet. *Current Microbiology*, 78(4):1628–1635, Apr. 2021. ISSN 1432-0991. doi: 10.1007/s00284-021-02399-8. URL https://doi.org/10.1007/s00284-021-02399-8.

F. Zoelzer, A. L. Burger, and P. W. Dierkes. Unraveling differences in fecal microbiota stability in mammals: from high variable carnivores and consistently stable herbivores. *Animal Microbiome*, 3(1):77, Nov. 2021. ISSN 2524-4671. doi: 10.1186/s42523-021-00141-0. URL https://doi.org/10.1186/s42523-021-00141-0.

5 Annex

Sequence data	Source	Study accession	Run accession
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161891
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161900
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161935
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161960
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161961
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161962
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161968
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161969
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161994
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9161999
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9162002
16S rRNA Data	(Milani et al., 2020)	PRJNA545289	SRR9162046
Shotgun Data	(Milani et al., 2020)	PRJNA545214	SRR9161834
Shotgun Data	(Milani et al., 2020)	PRJNA545214	SRR9161836
Shotgun Data	(Milani et al., 2020)	PRJNA545214	SRR9161840
Shotgun Data	(Milani et al., 2020)	PRJNA545214	SRR9161841
Shotgun Data	(Milani et al., 2020)	PRJNA545214	SRR9161846
16S rRNA Data (C. crocuta)	(Rojas et al., 2022)	PRJNA733503	SRR14978404
Shotgun Data (C. spelaea)	This Study	-	-

Table 5: Resources used to obtain sequence data.

Libraries and Packages	
Bwa	`http://bio-bwa.sourceforge.net/bwa.shtml`
Schmutzi	`https://github.com/grenaud/schmutzi`
PhyloFlash	`http://hrgv.github.io/phyloFlash/`
Trimal	`https://github.com/scapella/trimal`
HOPS	`https://github.com/rhuebler/HOPS`
AdapterRemoval	`https://adapterremoval.readthedocs.io/en/latest/`
MALT	`https://github.com/husonlab/malt`
SINA v 1.7.2	`https://github.com/epruesse/SINA`
Samtools	`https://github.com/samtools`
FastTree	`https://github.com/PavelTorgashov/FastTree`
Bubbleplot (see figure Figure 3)	`https://github.com/alex-bagnoud/OTU-table-to-bubble-plot.git`
Phyloseq	`https://github.com/joey711/phyloseq.git`

Table 6: Various computational tools used to analyse data.

6 Code used within the thesis

Listing 1: Code for Pie chart.

```
### plotting pie chart ####
phy <- transform_sample_counts(pseq, function(x) x/sum(x)) # relative abundance
glom <- tax_glom(phy, taxrank = 'Phylum') # list taxa at taxonomic level "Phylum"
data_glom<- psmelt(glom) # create dataframe from phyloseq object
data_glom$Phylum <- as.character(data_glom$Phylum) # convert to character
data_glom$Phylum[data_glom$Abundance < 0.015] <- "< 1% abund." # simple way to rename phyla with < 1% abundance
agg_df <- aggregate(data_glom$Abundance, by=list(data_glom$Phylum), FUN=sum) # aggregates taxa that have same name, summing abundance
pct <- round(100*agg_df$Abundance/sum(agg_df$Abundance)) # converts to a percentage
pie(agg_df$Abundance,
    labels = paste(agg_df$Phylum, sep = " ", pct, "%"),
    col = rainbow(length(agg_df$Phylum)))
```

Listing 2: Code for relative abundance plot.

```
### relative abundance/compositional plot
phy <- transform_sample_counts(pseq, function(x) x/sum(x)) # relative abundance
filtered <- aggregate_rare(phy, level = "Family", detection = "relative abundance", prevalence = "min number of samples") # filters for rare taxa
n <- 25 # number of colors we want to generate
qual_col_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual_col_pals)))

p <- plot_composition(filtered,
                      taxonomic.level = "Family",
                      sample.sort = "Diet",
                      x.label = "Animal") +
  scale_fill_manual(values = col_vector) +
  guides(fill = guide_legend(ncol = 1)) +
  scale_y_percent() +
  labs(y = "Relative abundance (%)",
       title = "Relative abundance data") +
  theme_ipsum(grid="Y") +
  theme(axis.text.x = element_text(angle=90, hjust=1),
        legend.text = element_text(face = "italic"))
```

Listing 3: Code for PCoA plot.

```
library(vegan)
uni <-  "weighted_unifrac_otu_table_beta.txt"
pDataFile <- "env file"
x.beta1 <- read.delim(uni, header = T, row.names = 1, sep="\t")
env <-read_excel(pDataFile)
length(env$Species)

rownames(x.beta1)=as.character(env$Animal)
colnames(x.beta1)=as.character(env$Animal)
group=env$Diet

vare.x1 <- capscale(as.dist(x.beta1) ~ 1)
vare.x1; vare.x1$CA
coords.x1= scores(vare.x1, display = "sites", choices=c(3,4))
coords.x1=as.data.frame(coords.x1)
par(mar = c(2, 2, 2, 2))
group=as.factor(group)
plot(vare.x1,type="n",cex = 6.5,choices=c(3,4),xlab="PCo3",ylab=paste("
    PCo4"))
ordiellipse(vare.x1,choices = c(3,4), group, kind = "se",conf = 0.95,show.
    groups=c("Herbivore"),col=c('forestgreen'),lwd=2)
ordiellipse(vare.x1,choices = c(3, 4), group, kind = "se",conf = 0.95,show
    .groups=c("Omnivore"),col=c('royalblue'),lwd=2)
ordiellipse(vare.x1,choices = c(3,4), group, kind = "se",conf = 0.95,show.
    groups=c("Carnivore"),col=c('salmon'),lwd=2)
ordispider(vare.x1,choices = c(3,4), group, col=c('forestgreen'), show.
    groups=c("Herbivore"),label=T,lwd=2)
ordispider(vare.x1,choices = c(3,4), group, col=c('royalblue'), show.
    groups=c("Omnivore"),label=T,lwd=2)
ordispider(vare.x1,choices = c(3,4), group, col=c('salmon'), show.groups=c
    ("Carnivore"),label=T,lwd=2)
with(coords.x1, points(MDS3,MDS4,pch=21,bg=colore,cex=1.2))
# significance test for separation of clusters
adonis(x.beta1~group, perm=999)
adonis(formula = x.beta1~group, data = x.beta1, permutations = 999)
```

Listing 4: Code for Venn diagram.

```
#### shared bacterial taxa between diets ####
diet_states <- unique(as.character(meta(carbom)$Diet))
list_core <- c() # an empty object to store information
par(mfrow=c(2,1))
for (n in diet_states){ # for each variable n in DiseaseState
  #print(paste0("Identifying Core Taxa for ", n))

  ps.sub <- subset_samples(pseq, Diet == n) # Choose sample from Diet by n
  core_m <- taxa_names(ps.sub)
  print(paste0("No. of core taxa in ", n, " : ", length(core_m))) # print
```

```
      core taxa identified in each DiseaseState.
  list_core[[n]] <- core_m
}

mycols <- c(nonCRC="#d6e2e9", CRC="#cbf3f0", H="#fcf5c7")
list <- c()
plot(venn(list_core),
     fills = mycols)
```

Listing 5: Code for phylosymbiotic tree.

```
library(vegan)
library(ape)
# observed ancestral nodes
observedA = c()
# null ancetral nodes
ancestorN=c()
#Permutation test for single values
permutation.test <- function(data1,observed_value, n){
  permutations=c()
  result=0
  for(i in 1:n){
    permutations[i]=sample(data1, length(data1), FALSE)[1]
  }
  result=sum(abs(permutations) >= abs(observed_value))/(n)
  return(list(result, permutations))
}

#permutation test
permutation.test(ancestorN1,observedA[1],999)[[1]]

# standardized error size values
ses <- (observedA[1]-mean(ancestorN1))/sd(ancestorN1)
# tree
dfin<- ape::read.tree(text="tree")
ses <- c(n1,n2,n3,n4,n5,n6,n7,n8,n9,n10,n11,n12,n13,n14,n15,n16,n17,n18)
colors1 <- colorRampPalette(c("white","khaki","orange","red"))
colors  <- colors1(10)[as.numeric(cut(ses,breaks = 10))]
plot.phylo(dfin)
nodelabels(bg = colors, cex=1.2,text=1:dfin$Nnode)
colori=colorRampPalette(colors = c("white","khaki","orange","red"))
colori1=colori(20)
# legend
barplot(cbind(rep(1,20)),col=colori1,horiz = F,border = NA, main = "
    Phylosymbiosis signal", ylab  = "(Standard effect size))
```

Acknowledgements

The work conducted throughout this thesis would not have been possible without the assistance and guidance of Professor Simone Rampelli of the University of Bologna. I'd also like to thank both family and friends that lent me a helping hand along the way, to you all I am forever grateful.