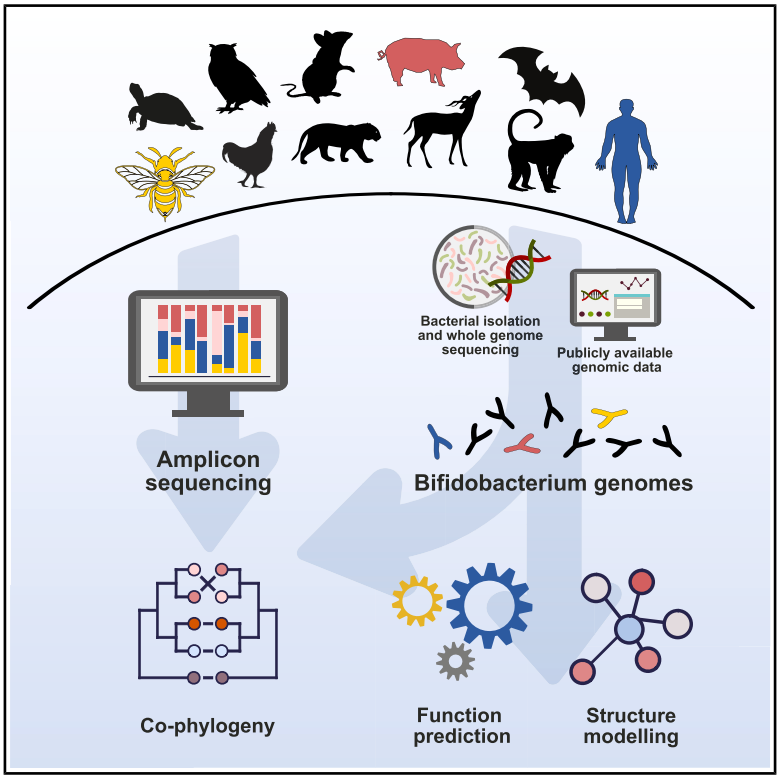
Article

Cell Host & Microbe

Host-specific microbiome and genomic signatures in *Bifidobacterium* reveal co-evolutionary and functional adaptations across diverse animal hosts

Graphical abstract



Authors

Magdalena Kujawska, David Seki, Lisa Chalklen, ..., Suparna Mitra, Lucy Crouch, Lindsay J. Hall

Correspondence

l.hall.3@bham.ac.uk

In brief

Bifidobacterium are beneficial members of the microbiota across animal hosts. Kujawska et al. demonstrate that bifidobacteria have likely co-evolved more closely with mammals, particularly primates, and that species associated with different hosts can share similar carbohydrate metabolism traits, highlighting diet as a major driver of bifidobacterial evolution and function.

Highlights

- Gut microbiome composition is influenced by host phylogeny, particularly in mammals
- *Bifidobacterium* displays host-specific co-phylogenetic associations at a strain level
- Diet appears to be a major force shaping bifidobacterial evolution and function

Kujawska et al., 2025, Cell Host & Microbe *33*, 1502–1517
September 10, 2025
https://doi.org/10.1016/j.chom.2025.08.008

CellPress

CellPress
OPEN ACCESS

Cell Host & Microbe

Article

Host-specific microbiome and genomic signatures in *Bifidobacterium* reveal co-evolutionary and functional adaptations across diverse animal hosts

Magdalena Kujawska,[1,2,3] David Seki,[1,4] Lisa Chalklen,[5] Jennifer Malsom,[5] Raymond Kiu,[2,3] Sara Goatcher,[6,7] Ioulios Christoforou,[8] Suparna Mitra,[9] Lucy Crouch,[2,3] and Lindsay J. Hall[1,2,3,10,11,*]

[1]Intestinal Microbiome, ZIEL - Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany
[2]Department of Microbes, Infection, & Microbiomes, School of Infection, Inflammation and Immunology, College of Medicine & Health, University of Birmingham, Birmingham B15 42TT, UK
[3]Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 42TT, UK
[4]Centre for Microbiology and Environmental Systems Science, Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, University of Vienna, 1030 Vienna, Austria
[5]Food, Microbiome, & Health, Quadram Institute Bioscience. Norwich Research Park, Norwich NR4 7UQ, UK
[6]Banham Zoo, Kenninghall Road, Banham, Norwich NR16 2HE, UK
[7]Africa Alive Zoological Reserve, Whites Lane, Kessingland, Lowestoft NR33 7TF, UK
[8]Pafos Zoo, Xylomantrou Street, Agiou Georgiou, 8650 Peyia, Paphos, Cyprus
[9]Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK
[10]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK
[11]Lead contact
*Correspondence: l.hall.3@bham.ac.uk
https://doi.org/10.1016/j.chom.2025.08.008

SUMMARY

Animals harbor divergent microbiota, including various *Bifidobacterium* species, yet their evolutionary relationships and functional adaptations remain understudied. Using samples from insects, reptiles, birds, and mammals, we integrated taxonomic, genomic, and predicted functional annotations to uncover how *Bifidobacterium* adapts to host-specific environments. Host phylogeny is a major determinant of gut microbial composition. Distinct microbiota in mammalian and avian hosts reflect evolutionary adaptations to dietary niches, such as carnivory, and ecological pressures. At a strain-resolved level, *Bifidobacterium* and their hosts exhibit strong co-phylogenetic associations, driven by vertical transmission and dietary selection. Functional analyses highlight striking host-specific adaptations in *Bifidobacterium*, particularly in carbohydrate metabolism and oxidative stress responses. In mammals, *Bifidobacterium* strains are enriched in glycoside hydrolases tailored to complex carbohydrate-rich diets, including multi-domain GH13_28 α-amylases associated with degradation of resistant starch. Together, these findings deepen our understanding of host-microbe co-evolution and the critical role of microbiota in shaping animal health and adaptation.

INTRODUCTION

Members of the genus *Bifidobacterium* have long been associated with beneficial health outcomes, particularly due to their probiotic properties. Human-derived bifidobacteria, especially those associated with infancy, have received considerable research attention, leading to extensive efforts in their isolation and characterization.[1–3] However, *Bifidobacterium* species are not confined to humans, and they are widely distributed across the animal kingdom, associating with a diverse array of hosts belonging to different classes, including mammals, birds, and social insects.[4,5] Increased availability and affordability of next-generation sequencing (NGS) have significantly expanded our understanding of *Bifidobacterium* diversity, revealing novel species and elucidating inter- and intra-species variability.

The distribution of *Bifidobacterium* across diverse hosts has led to the evolutionary hypotheses regarding their host specificity. Some species, like *Bifidobacterium animalis* and *Bifidobacterium pseudolongum*, appear to be generalists found across diverse animal classes. By contrast, other species demonstrate stronger host specificity, particularly within primates.[6,7] However, research has been disproportionately dominated by observations from primates, with approximately half of the recognized *Bifidobacterium* species type strains recovered from non-human primates, particularly members of the family *Callitrichidae*.[8,9] This shaped our understanding of bifidobacterial host range and ecological specialization, even though other mammalian hosts, and indeed other classes of animals, likely harbor a wealth of undiscovered *Bifidobacterium* species and subspecies.[10–12]

Central to evolutionary studies is the concept of phylosymbiosis, which refers to the evolutionary concordance between host

phylogeny and microbiome composition.[13] This phenomenon suggests that closely related host species tend to harbor more similar microbial communities, likely due to co-evolutionary processes. The existence of such processes between two or more taxa can be assessed by co-phylogeny, which evaluates the congruence of phylogenetic relationships between different groups of organisms due to their long-term interaction.[14–16] Despite the growing availability of genomic data, robust studies investigating evolutionary and functional relationships between *Bifidobacterium* and their animal hosts at the strain level remain scarce. The pan-genome of the genus and its phylogenetic relationship are yet to be fully resolved. Co-phylogenetic analyses have suggested a close evolutionary relationship between certain bifidobacteria and their primate hosts, particularly those within the *Hominidae* family.[17,18] Previous large-scale studies have often focused on intra-species features or a limited number of isolates, leaving much to be explored about the genomic and functional diversity of *Bifidobacterium* across the animal kingdom.[19,20]

In the context of *Bifidobacterium*, phylosymbiosis could play a crucial role in shaping the distribution and functionality of these microbes across different hosts. Additionally, diet plays a critical role in the shaping of the microbiome, with glycoside hydrolases (GHs)—enzymes that break down complex carbohydrates—particularly important.[21,22] The diversity and evolution of GH enzymes within *Bifidobacterium* genomes offer insights into how diet and microbial function co-evolve in a host-specific manner, driving adaptation and ecological flexibility.[23]

To address existing knowledge gaps and expand our understanding of *Bifidobacterium* diversity, we subjected 393 fecal samples from 175 diverse animal hosts to genomic DNA extraction and *Bifidobacterium* isolation. This effort resulted in the generation of the 16S rRNA amplicon sequencing data for a total of 219 fecal samples from 126 animal hosts, including representatives from the classes Aves, Reptilia, Insecta, and Mammalia. In addition, we isolated and sequenced 96 unduplicated *Bifidobacterium* strains from 37 animal hosts, forming the basis of a robust genomic dataset that we further supplemented with high-quality publicly available genomes. To assess the evolutionary links between diet, microbial function, and host phylogeny, we conducted a comprehensive analysis of the host gut microbiota composition, with a particular focus on the associations of *Bifidobacterium* with certain host lineages, and examined *Bifidobacterium* diversity and functions at the genomic level, particularly traits related to carbohydrate metabolism and the most abundant GH13 family. This study aims to provide a more detailed understanding of the adaptive strategies that enable *Bifidobacterium* to thrive across diverse host environments, offering insights into the co-evolution of hosts and their microbiomes.

RESULTS

Host-specific microbial diversity and enrichment of *Bifidobacterium* in animal gut microbiomes

Out of the 393 animal fecal samples collected from 175 diverse hosts in this study, we successfully performed 16S rRNA amplicon sequencing for a total of 219 samples from 126 diverse hosts across four taxonomic classes—Insecta, Reptilia, Aves, and Mammalia (Table S1). We used the generated data to explore broad-scale patterns of microbial diversity in these host groups and, more specifically, to examine the distribution of the genus *Bifidobacterium* across host lineages. We focused on *Bifidobacterium* due to its role as a keystone genus across multiple host taxa, where it contributes to microbiome development, gut homeostasis, and host-microbe interactions.

The dataset was rarefied to 10,000 reads per sample, and low-abundance amplicon sequencing variants (ASVs) occurring at <1% relative abundance in any sample were filtered out, with the resulting matrices used for subsequent analyses. The investigation of the taxonomic composition of the microbiota across the four diverse animal classes revealed notable prevalence and dominance of *Escherichia coli* in the majority of avian host orders, except for Coraciiformes and Struthioniformes (Figure 1A). Mammalian hosts exhibited an overall higher relative abundance of bacterial genera belonging to the phylum Bacteroidota, for example, *Bacteroides* and *Prevotella*. Similarly, *Bifidobacterium* ASVs were enriched in this host class compared with Aves, particularly in host orders Edentata and Primates (Figure 1A). These observations underscore the potential evolutionary and ecological significance of these microbiota members in mammalian hosts.

Principal-component analysis (PCA) indicated a distinct separation of Aves and non-carnivorous Mammalia along axis-PC1, while the difference along axis-PC2 was mostly driven by the formation of a distinct cluster by samples from mammalian host orders Marsupialia and Artiodactyla (Figure 1B). We also observed that microbial richness was significantly higher in Mammalia compared with Aves (t test, $p.adj < 0.0001$; Figure 1C). Furthermore, microbiota composition varied significantly according to host class (PERMANOVA, $p.adj < 0.05$), with significant differences in pairwise comparisons of Euclidean distances between the taxonomic classes Mammalia vs. Aves, Reptilia vs. Mammalia, and Mammalia vs. Insecta (adonis, $p.adj = 0.006$, 0.024, and 0.038, respectively). At the host taxonomic level of order, we only identified several pairwise comparisons among members of Aves and Mammalia that were significantly different (adonis, $p.adj < 0.05$) (Figure 1D), with Strigiformes highly dissimilar to Psittaciformes (average Euclidean distance = 50.67) and Rodentia highly dissimilar to Carnivora (average Euclidean distance = 50.54).

To test for homogeneity of dispersion across taxonomic classes, we performed permutational analysis of multivariate dispersions (PERMDISP). Overall, PERMDISP was statistically significant ($p.adj < 0.0001$), suggesting that the observed differences in microbial diversity may be driven by differences in variance rather than differences in group centroids (average distance to centroid: Aves = 40.23, Insecta = 32.63, Mammalia = 29.57, and Reptilia = 30.53), likely confounding previous PERMANOVA results.

Next, we conducted pairwise comparisons between taxonomic classes using Jaccard distances based on presence-absence matrices and calculated checkerboard units (checkerboard score [C-scores]) to measure the extent to which pairs of microbial species avoid co-occurrence in host habitats (Figure 1E). For all host classes, except for Insecta, C-score distributions differed significantly from randomized null models ($p.adj < 0.05$), indicating that patterns of segregation may be meaningful in Reptilia, Aves, and Mammalia. C-score distributions differed significantly

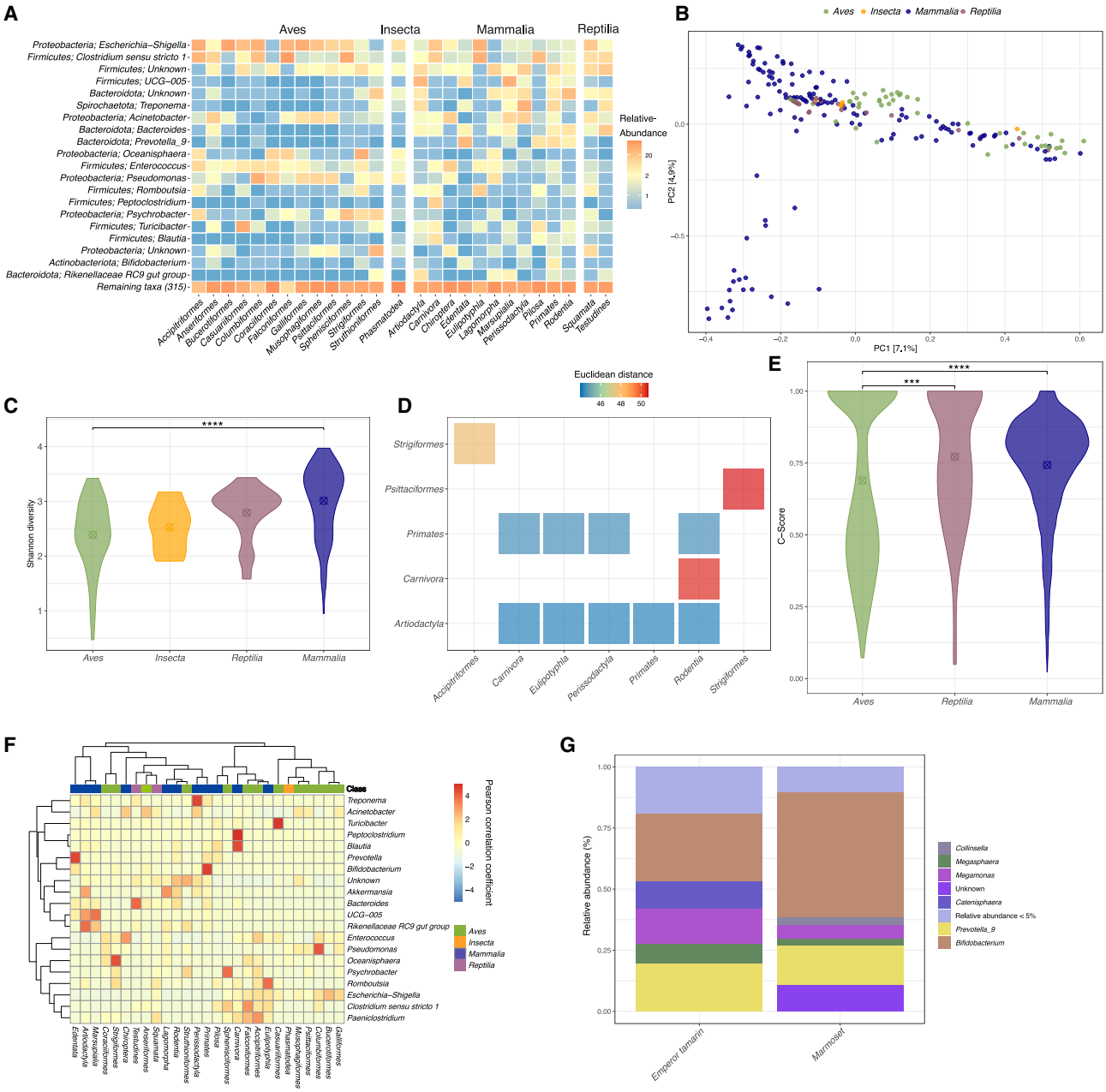


Figure 1. Characteristics of gut microbiota in insects, reptiles, birds, and mammals

(A) Mean average relative abundance of most abundant microbial genera (y axis) in given host orders (x axis).

(B) Principal-component analysis (PCA) of relative species counts over Euclidean distance.

(C) Microbial richness per taxonomic class, measured by Shannon's index.

(D) Euclidean distances between host pairs that significantly differ according to microbial compositionality.

(E) Checkerboard scores (C-scores) of microbial communities across host taxonomic classes. Pairwise C-scores were calculated using Jaccard-based presence-absence matrices to assess the degree of species segregation (checkerboard structure) within each class.

(F) Hierarchical clustering of twenty most abundant bacterial genera and host orders based on Pearson correlation of genus-level microbial profiles.

(G) Relative abundance of microbial genera in emperor tamarins and marmosets.

Violin plots in the figure indicate the group median, and asterisks represent adjusted p values: *** $p < 0.001$ and **** $p < 0.0001$.

between Aves and both Reptilia and Mammalia (p.adj < 0.05), suggesting that Aves exhibit lower checkerboard structure. These results potentially indicate reduced niche separation or host filtering relative to other vertebrates.

To explore patterns in microbiota profiles at the lower host and bacterial taxonomic levels, we performed hierarchical clustering of host orders and bacterial genera using Pearson correlation as the similarity measure (Figure 1F). This analysis revealed notable

CellPress
OPEN ACCESS

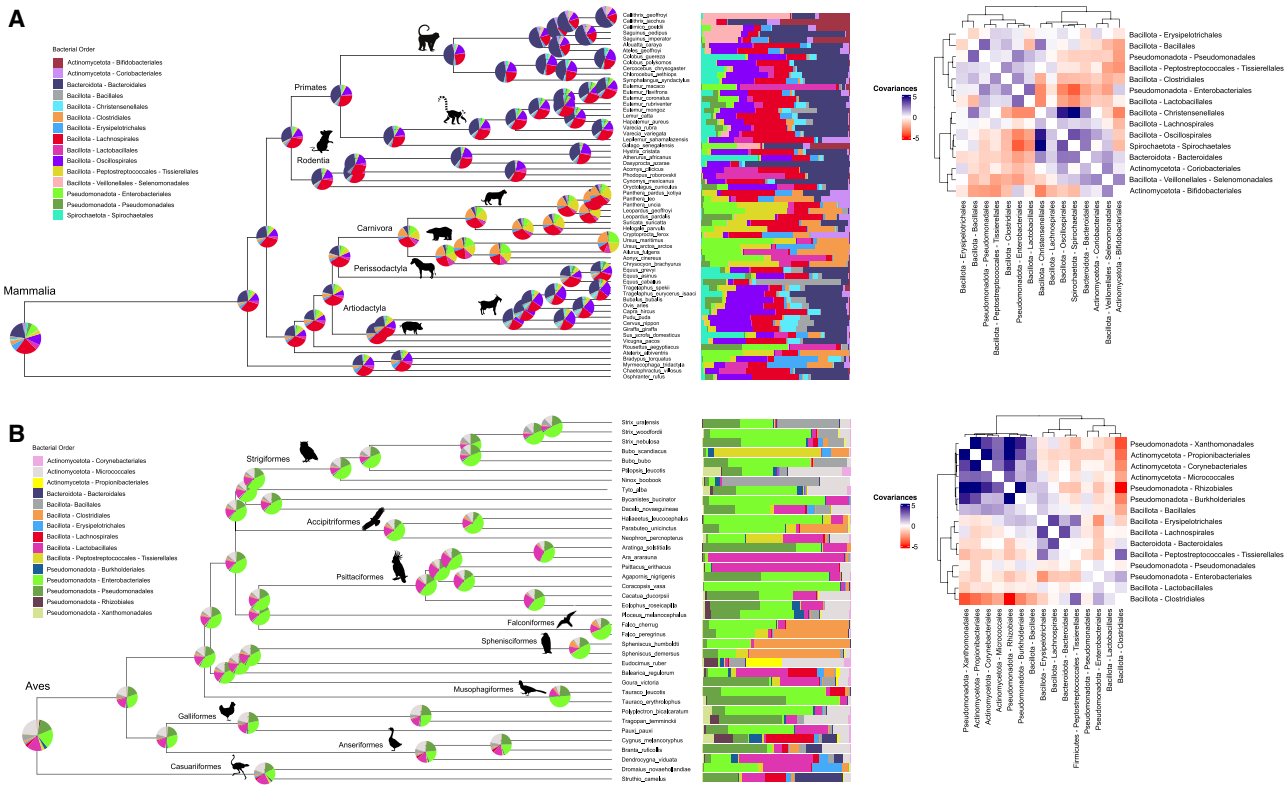


Figure 2. Ancestral reconstruction of the host gut microbiota

Ancestral reconstruction of (A) mammalian and (B) avian gut microbiota with estimated covariances between the main bacterial taxa (heatmaps on the right). Pie charts at the root and nodes of the dated phylogenetic trees represent estimated ancestral microbiota compositions, while relative abundances (averaged) of the 15 most abundant bacterial orders based on the 16S rRNA amplicon sequencing performed in this study are represented by bar charts to the right of phylogenies and ancestral estimations. For each covariance matrix, we represented negative covariances in red and positive covariances in blue.

observations. At the host level, mammalian carnivores clustered closely with birds of prey, indicating similarity of gut microbial composition among phylogenetically distinct predatory groups. At the bacterial level, the genera *Prevotella* and *Bifidobacterium* clustered together and exhibited a strong positive correlation with Edentata and Primates—although the correlation pattern was reversed in these two host orders—which suggested potential host-specific enrichment. The pronounced association between *Bifidobacterium* and their primate hosts appears to be driven by the very high abundance of bifidobacteria in tamarins and marmosets (family *Callitrichidae*) (Figure 1G). These findings suggest that certain primate species may have evolved specialized gut environments that support *Bifidobacterium* persistence, possibly linked to their diet and unique ecological niches.

Phylosymbiosis signals and microbial covariance patterns across mammalian and avian lineages

Microbiota composition can vary significantly within and between host species,[24,25] yet more closely related hosts often harbor a more similar microbiota,[13,26] a phenomenon known as phylosymbiosis. To investigate phylosymbiosis patterns in our dataset, we applied a multivariate Brownian motion model[27] to the gut bacterial microbiota of 62 mammalian and 38 bird species, for which both the dated phylogenetic trees and the corresponding 16S rRNA amplicon data were available. Our analysis revealed a stronger phylosymbiosis signal in mammals compared with birds ($\lambda \simeq 0.41$ and $\lambda \simeq 0.24$, respectively). Despite the low and uneven number of mammalian vs. avian hosts included in the analysis, our results were consistent with previous data for >200 mammalian and >300 bird host species[27] and suggest that the gut microbiota of mammals is more tightly linked to host phylogeny than that of birds.

Ancestral reconstruction indicated that members of the bacterial phylum Pseudomonadota, particularly the orders Enterobacteriales and Pseudomonadales, were more abundant in the ancestral gut microbiota of birds than mammals (Figures 2A and 2B). This distinction likely reflects differences in the evolutionary history and dietary habits of these two host groups. This analysis also revealed notable shifts in bacterial community composition in the ancestors of various mammalian and avian orders. In mammals, the largest shift was detected in the ancestor of the host order Carnivora, characterized by an increased proportion of bacterial orders Clostridiales (Bacillota), Peptostreptococcales-Tissierellales (Bacillota), and Enterobacteriales (Pseudomonadota), along with a decreased proportion of members of the order Bacteroidales (Bacteroidota). Another notable shift occurred in the ancestor of the Primate suborder Simiiformes (encompassing both New and Old World monkeys), which exhibited an increased proportion of bacterial taxa belonging to the order Bacteroidales. Additionally, an increased

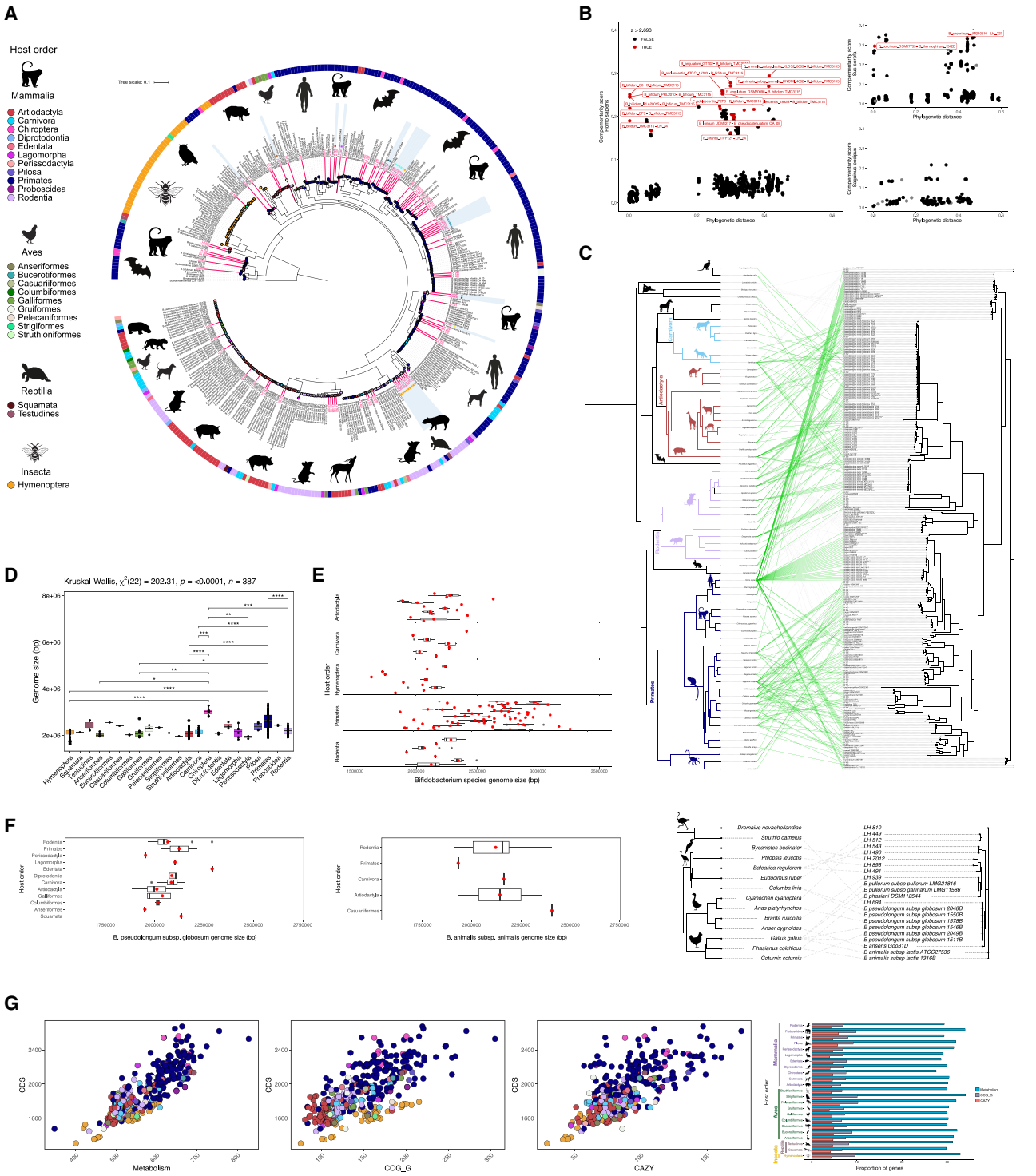


Figure 3. General features and co-phylogeny of bifidobacterial genomes used in this study

(A) Maximum-likelihood tree of 387 animal-associated bifidobacterial genomes. Isolation source (animal host order) is marked with a colored strip and node symbols on the tree. Isolates recovered in this study are labeled in pink. Light blue shading marks strains with average nucleotide identity <95% to known *Bifidobacterium* type strains, and colored squares within the shading denote individual species.

(B) Analysis of interactions between *Bifidobacterium* in selected animal hosts based on metabolic complementarity scores calculated for pairs of strains. Significant complementarity outliers are colored in red based on *Z* score threshold of ±2.698. See also Table S4.

(C) Tanglegram reflecting co-phylogeny of bifidobacterial strains and their hosts. Significant links ($p < 0.01$) are marked in green. See also Table S5.

(legend continued on next page)

proportion of members of the bacterial order Veillonellales-Selenomonadales was detected in the ancestor of New World monkeys. In birds, a major shift was observed in the ancestor of the host clade Neoaves, marked by an increased proportion of the members of the bacterial order Enterobacteriales (Pseudomonadota). In addition, the ancestors of host orders Falconiformes and Sphenisciformes were characterized by a larger proportion of bacterial taxa belonging to the order Clostridiales (Bacillota), highlighting a possible microbial adaptation to carnivorous diets and unique ecological niches.

Consistent with previous reports, we detected both positive and negative covariances among various bacterial taxa in mammals and birds. Visual inspection of the covariance matrices revealed the clustering of bacterial orders into subsets, with taxa within these subsets tending to covary interdependently. In both mammals and birds, positive covariances were observed between bacterial members of Pseudomonadota (Pseudomonadales and Enterobacteriales) and Bacilllota (Clostridiales and Peptostreptococcales-Tissierellales), with this covariance profile particularly dominant in the microbiota of Carnivora and carnivorous birds (Figures 2A and 2B). In mammals, notably, orders Bacteroidales and Coriobacteriales covaried with orders Bifidobacteriales and Veillonellales-Selenomonadales in New World monkeys alone, suggesting a unique microbiota profile in these primates (Figure 2A). In birds, we detected strong covariance between members of Actinomycetota and Pseudomonadota (orders Propionibacteriales, Corynebacteriales, Micrococcales and Xanthomonadales, Micrococcales, and Rhizobiales, respectively), with this subset exhibiting a strong negative covariance with the order Clostridiales (Figure 2B). These results indicate the existence of distinct microbial communities across mammalian and avian hosts that may reflect their evolutionary adaptations, possibly driven by host diet and the collaborative nature of microbial interactions.

Genomic diversity, phylogenetic congruence, and general functional features of animal-associated *Bifidobacterium*

To expand current understanding of genomic features of animal-associated bifidobacteria, we isolated and whole-genome sequenced strains from a broad range of hosts across the animal tree of life. Alongside DNA extraction for 16S rRNA amplicon sequencing, we subjected the collected 393 animal fecal samples from 175 diverse hosts to bacterial isolation and successfully recovered *Bifidobacterium* from 37 (21.1%) host species, with the highest isolate recovery rate observed in mammals (27 out of 87 hosts, 31%), followed by birds (8 out of 59 hosts, 13.6%) and reptiles (2 out of 15 hosts, 13.3%). Where applicable, our isolation efforts aligned with the 16S rRNA amplicon data, capturing *Bifidobacterium* from hosts with high average relative abundance of *Bifidobacterium* ASVs—particularly within the orders Edentata and Primates, including marmosets and tamarins (Table S2). Notably, we also isolated *Bifidobacterium* from hosts for which 16S rRNA amplicon data were not available.

Culturing and subsequent whole-genome sequencing of isolates resulted in the final collection of 96 unduplicated *Bifidobacterium* strains from mammals ($n = 84$; with $n = 54$ strains isolated from marmosets and tamarins), birds ($n = 10$), and reptiles ($n = 2$). The genome sizes of these isolates ranged from 1.85 Mb (*Bifidobacterium pseudolongum* subsp. *pseudolongum*) to 3.26 Mb (*Bifidobacterium simiventris*), corresponding to 1,529 and 2,674 protein-coding open reading frames (ORFs), respectively. These values fall within the range previously reported for *Bifidobacterium* species.[4,23]

To explore the diversity within the *Bifidobacterium* genus, we expanded our dataset by including a further 291 publicly available animal-associated bifidobacterial genomes, which resulted in a final dataset of 387 strains. Using the Genome Taxonomy Database classification (GTDB-tk) and the comparison of average nucleotide identity (ANI) values between the recovered 96 isolates and 105 *Bifidobacterium* type strains retrieved from the NCBI as part of the public dataset, we assigned 66 genomes (69%) to 32 *Bifidobacterium* species and subspecies (ANI > 95%) (Table S3). The predominant species and subspecies identified included *Bifidobacterium pseudolongum* subsp. *globosum* (7 isolates, 11% of assigned genomes), *Bifidobacterium reuteri* (6 isolates, 10% of assigned genomes), *Bifidobacterium animalis* subsp. *animalis* (4 isolates, 6% of assigned genomes), and *Bifidobacterium callitrichidarum* (4 isolates, 6% of assigned genomes). Surprisingly, 30 isolates (31%) showed ANI values below 95% compared with known *Bifidobacterium* type strains, suggesting they may represent previously unreported species. These isolates were primarily recovered from primates—tamarins and marmosets (63%), with other hosts including a South American rodent—*Dasyprocta azarae* (Azara's agouti), various bird species, and a red-footed tortoise (*Chelonoidis carbonaria*). ANI comparisons between these isolates indicated that they belonged to 11 individual species.

Phylogenetic analysis revealed moderate clustering of *Bifidobacterium* strains according to host order, supported by ANOSIM statistics ($R = 0.5037$, $p = 0.0001$) (Figure 3A). Strains isolated from the hosts belonging to the same order tended to cluster together within bifidobacterial species, reflecting the notion that certain species, such as *Bifidobacterium asteroides*, are host specific (e.g., only found in insects), while others, like *B. animalis* or *B. pseudolongum*, exhibit more cosmopolitan distributions. The comparison of the ANI values between *Bifidobacterium* strains revealed that similar strains (ANI > 99%) are not

(D) Comparison of genome sizes between 387 *Bifidobacterium* isolates grouped according to host order. Significant differences between groups based on the Kruskal-Wallis statistics ($p < 0.0001$) are denoted with asterisks.
(E) Distribution of genome sizes of *Bifidobacterium* species in hosts for which we had more than 20 bifidobacterial genomes available: Hymenoptera, Artiodactyla, Carnivora, Rodentia, and Primates. Average genome sizes of individual *Bifidobacterium* species are marked with a red dot.
(F) Distribution of genome sizes of *Bifidobacterium* species found in various hosts, *Bifidobacterium pseudolongum* subsp. *globosum* and *Bifidobacterium animalis* subsp. *animalis*. Average size of genomes associated with particular host groups are marked with a red dot.
(G) Number of coding sequences as a function of number of genes involved in overall metabolism (left), carbohydrate metabolism (COG_G) (middle left), and representing CAZyme abundance (middle right), and proportional representation of these three functional features in bifidobacterial genomes grouped according to host order (right). See also Table S6.

only present within the same host species but also across different hosts and, in some cases, across distinct host taxonomic orders, as is the case of *Bifidobacterium animalis* subsp. *lactis* (Table S3). At the same time, relatively distantly related bifidobacterial species are conserved in certain hosts, for example, in humans (order Primates) and pigs (order Artiodactyla) (Figure 3A).

To explore the relationship between *Bifidobacterium* strains within shared environments (i.e., individual hosts), we used PhyloMInt to compute pairwise competition and complementarity scores based on predicted metabolic networks. This analysis focused on strains associated with three representative hosts: *Sus scrofa* (pig), *Homo sapiens*, and *Saguinus oedipus* (cotton-top tamarin). According to niche differentiation theory, closely related species occupying the same ecological niche are more likely to compete due to overlapping functional traits and shared resource demands, reducing the likelihood of stable coexistence, unless they diverge along at least one niche axis (e.g., time, space, or nutrient use).[28] Our analysis revealed high competition scores between strains belonging to closely related *Bifidobacterium* species across all three host species, with a general trend of decreasing competition as phylogenetic distance increased (Table S4). This pattern may help explain observed host-specific strain conservation. Notably, we also identified outlier strain pairs with significant elevated complementarity scores (*Z* score threshold of ±2.698) relative to their phylogenetic distance (Figure 3B; Table S4), particularly in humans. These findings support previous reports of coexistence among diverse bifidobacterial strains and species within the same host ecosystem.[3]

Next, we tested for co-phylogeny signal between mammalian and avian host phylogenies and those of their associated bifidobacterial strains. Significant phylogenetic congruence was detected in Mammalia (ParaFitGlobal = 10,071,739, mean p_{global} = 0.001) but not in Aves (ParaFitGlobal = 21,600.47, mean p_{global} = 0.122) (Figure 3C). Most mammals displayed significant links to associated strains (Table S5). Notable exceptions included the African elephant, sloth, domestic cat, fruit bat, and certain primate species. Interestingly, for the order Artiodactyla, phylogenetic congruence with associated strains seemed to be dependent on the specific *Bifidobacterium* species, with significant associations found between "cosmopolitan" species like *B. pseudolongum* and their porcine host.

Comparative analysis of genome sizes across *Bifidobacterium* isolates grouped by host order indicated significant differences (Kruskal-Wallis x^2 = 202.31, $p < 0.0001$, df = 22). For example, *Bifidobacterium* isolates (Figure 3D) from Chiroptera (bats) and Primates had the largest genomes (3.01 ± 0.16 and 2.60 ± 0.33 Mb [mean ± SD], respectively), while those associated with Perissodactyla (horses, rhinoceroses) and Struthioniformes (ostriches) had the smallest genomes (1.92 ± 0.1 and 1.96 ± 0 Mb, respectively). These findings align with previous analyses of 129 publicly available *Bifidobacterium* strains.[20] The visualization of the distribution of *Bifidobacterium* species genome sizes across host groups for which we had more than 20 *Bifidobacterium* genomes available, namely Hymenoptera (bees and bumblebees), Artiodactyla (even-toed ungulates), Carnivora, Rodentia, and Primates, revealed a clear variation (Figure 3E). *Bifidobacterium* species associated with Primates show an overall high variability in genome sizes. Genome sizes of species associated with Rodentia display a larger spread, while those of species associated with Hymenoptera tend to be smaller and less variable. The visualization of the distribution of genome sizes of cosmopolitan *B. pseudolongum* subsp. *globosum* across its various host groups indicated that mammalian hosts, especially Rodentia and Primates, tend to harbor strains with generally larger genomes compared with bird hosts (i.e., Galliformes, Columbiformes, and Anseriformes) (Figure 3F). By contrast, the distribution of genome sizes of another cosmopolitan species, *B. animalis* subsp. *animalis*, across its hosts revealed a reversed trend, with strains isolated from Casuariiformes displaying the largest genome sizes. However, valid conclusions are limited by an insufficient number of data points (Figure 3F).

Functional classification of gene content assigned 80.9% of ORFs to clusters of orthologous groups (COG) categories, with 19.1% categorized as proteins of unknown function (Table S6). As expected for *Bifidobacterium*, carbohydrate transport and metabolism (COG_G) was the second most abundant category (after function unknown) and constituted 9.5% of assigned functions, reflecting the saccharolytic lifestyle of this genus. Additionally, genes involved in amino acid metabolism (COG_E) constituted 8.6% of assigned functions. Carbohydrate active enzymes (CAZymes) made up 59.1% of ORFs in the COG_G category, with GHs being the predominant class across all host orders. On average, overall metabolism, COG_G, and CAZymes, respectively, constituted 32.87%, 9.64%, and 4.25% of ORFs in insect-associated *Bifidobacterium* genomes; 30.78%, 7.73%, and 4.45% in *Bifidobacterium* from reptiles; 30.66%, 7.58%, and 4.54% in those associated with birds; and 29.53%, 7.54%, and 4.59% in bifidobacterial genomes associated with mammals. Notably, the abundance of putative metabolism-associated genes varied between host orders, with an expected tendency for strains with larger genomes to harbor more carbohydrate metabolism genes (Figure 3G).

Host-specific carbohydrate metabolism adaptations in *Bifidobacterium*

Given that taxon-specific functional properties can reflect broad-scale differences between host groups, we conducted a focused analysis on the functions of *Bifidobacterium*, particularly those related to carbohydrate metabolism. We concentrated on host order groups for which we had more than 20 *Bifidobacterium* genomes available: Hymenoptera, Artiodactyla, Carnivora, Rodentia, and Primates.

We analyzed 1,555 KEGG orthologs[29] (KOs) to identify abundance differences between *Bifidobacterium* from Hymenoptera vs. those associated with mammals, both Primates and non-primates. The selection of these categories was dictated by the natural biological variation in the distribution and diversity of bifidobacteria across host orders and their high abundance in Primates. We found significant abundance differences in 377 (24.2%) and 292 (18.7%) KOs, respectively (Qbonferroni < 0.05; Figure 4A; Table S7). Specifically, bifidobacterial genomes associated with insect hosts (Hymenoptera) exhibited higher abundances of KOs related to oxidative phosphorylation, particularly components of the cytochrome *bd* system. By contrast, mammal-associated *Bifidobacterium* showed higher abundance

CellPress
OPEN ACCESS

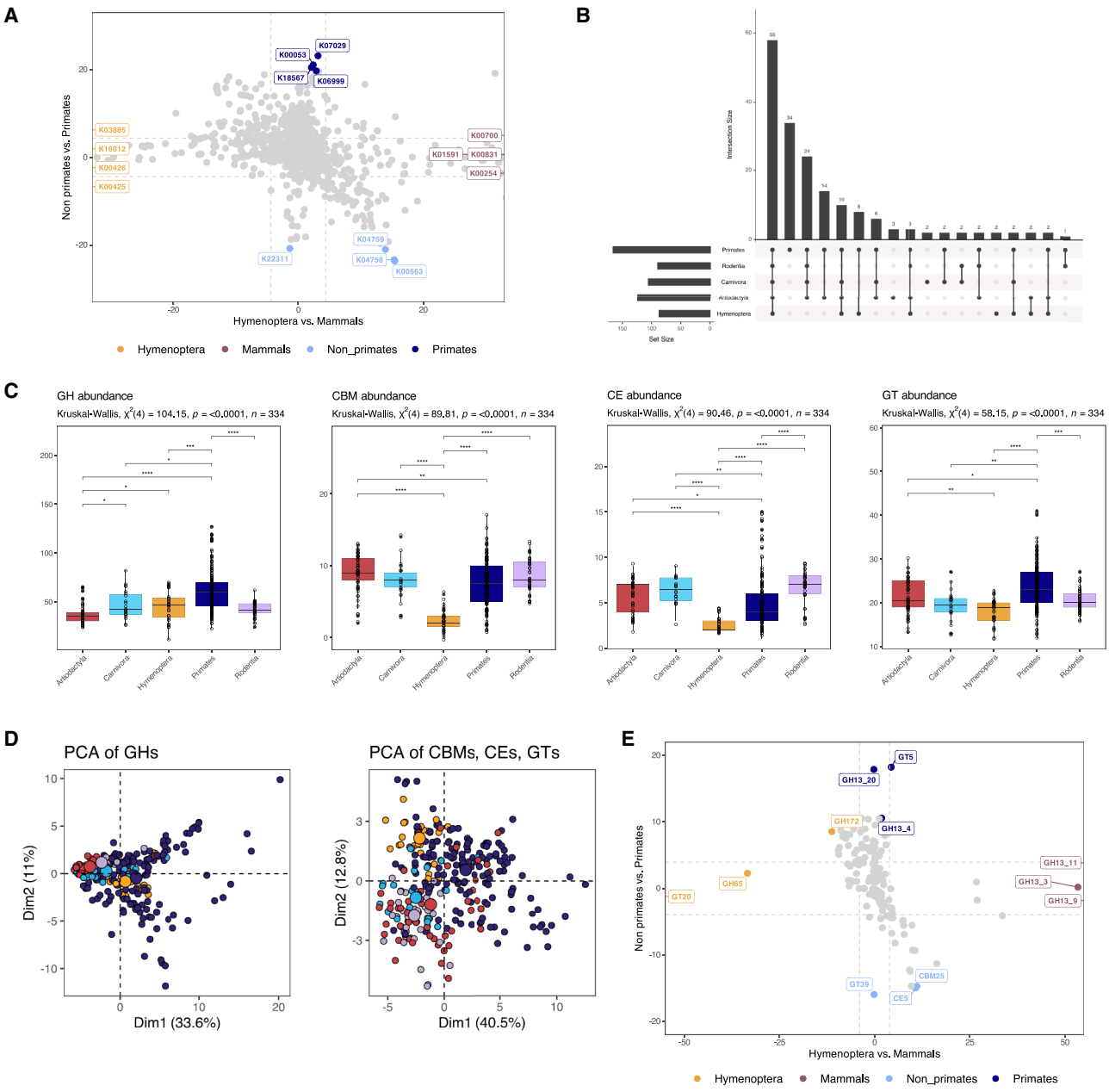


Figure 4. Functional analysis of bifidobacterial genomes from the perspective of the host taxonomic level of order

(A) Effect sizes (t-values from univariate linear regression) of the abundance differences of 1,555 KEGG orthologs in *Bifidobacterium* from Hymenoptera vs. those associated with mammals (x axis) and Primates and non-primates (y axis). Top features with the highest t-value in each group are colored and labeled. Horizontal and vertical lines depict the t-value threshold (| t-value | > 4.38) for statistical significance after Bonferroni correction (Q < 0.05, two sided).

(B) Distribution of CAZyme classes in bifidobacteria grouped according to host order groups.

(C) Abundance of CAZyme classes in bifidobacterial genomes grouped according to host order. GH, glycoside hydrolase; CBM, carbohydrate-binding module; CE, carbohydrate esterase; GT, glycosyltransferase.

(D) PCA analysis based on the CAZyme abundance matrix, with data for GHs analyzed separately.

(E) Effect sizes (t-values from univariate linear regression) of the abundance differences of 189 CAZyme classes in *Bifidobacterium* from Hymenoptera vs. those associated with mammals (x axis) and primates and non-primates (y axis). Top features with the highest t-value in each group are colored and labeled. Horizontal and vertical lines depict the t-value threshold (| t-value | > 3.95) for statistical significance after Bonferroni correction (Q < 0.05, two sided).

See also Table S7.

in categories linked to ferrous iron transport, cofactor and vitamin biosynthesis, and carbohydrate metabolism, notably including enzymes belonging to GH13_9 subfamily (K00700), where the current members have been characterized as hydrolyzing the α-1,6-branches in amylopectin and glycogen. Primate-associated genomes were enriched in categories covering

major facilitator superfamily transporters, glycerophospholipid metabolism, and metabolism of cofactors and vitamins.

We further examined the distribution of CAZyme families and subfamilies across host order groups. Our analysis revealed 58 CAZyme (sub)families shared between all groups, with the GH13 family displaying the highest overall abundance (Figure 4B). The GH13 family is relatively well studied, with substrate specificities and structures solved in many of the subfamilies. Notably for this study, this family solely has activity against α-linked glucose polysaccharides. The most prevalent of these polymers is plant-derived starch, which is likely why the GH13 family is so abundant, but others include glycogen (animal derived), dextran (bacteria derived), and pullulan (fungal derived).

The *Bifidobacterium* genomes from Primates exhibited the greatest CAZyme diversity, encoding a total of 166 classes encompassing 34 unique families. The most abundant unique GH families in this group, represented by GH59, GH88, GH123, and GH151 (Figure 4B), are associated with the breakdown of host glycans, including glycosphingolipids, O-glycans, and glycosaminoglycans. Conversely, *Bifidobacterium* genomes from Hymenoptera were the least diverse, with 87 classes and only 2 unique (sub)families—GH43_31 and GH144, which are associated with β-galactofuranosidase and β-glucanase activities, respectively. When analyzing the abundance of specific CAZyme families, we observed distinct differences between host groups (Figure 4C). Primate-associated *Bifidobacterium* encoded the highest average number of GH enzymes and glycosyl transferases (GTs) (61.2 ± 20.7 and 23.7 ± 5.14 [mean ± SD], respectively), while the lowest abundances of these enzyme classes were recorded in *Bifidobacterium* from Artiodactyla (36.7 ± 9.08 GH enzymes) and Hymenoptera (18.1 ± 2.86 GTs), respectively. Notably, carbohydrate-binding module (CBM) families were the most abundant in *Bifidobacterium* genomes associated with Artiodactyla (9.05 ± 2.56), while Rodentia and Carnivora showed similar numbers for carbohydrate esterases (CEs) (6.62 ± 1.87 and 6.45 ± 1.44, respectively).

To further assess the diversity of bifidobacterial CAZymes across host groups, we performed PCA on the CAZyme abundance matrix. The PCA did not reveal clear separation between different host groups (Figure 4D), which was supported by ANOSIM statistics (GH: R = 0.2975, p = 0.001, and ANOSIM CBM, CE, GT R: 0.3521, p = 0.001). However, additional analysis of the 189 CAZyme families and subfamilies in *Bifidobacterium* from Hymenoptera (insect-associated) vs. those from mammals, Primates, and non-primates (Qbonferroni < 0.05; Figure 4E; Table S7) indicated significant association of certain CAZyme classes with particular host groups. Compared with insects, the GH13 family was more prevalent in mammals. Within the GH13 family in mammals, subfamilies GH13_3 (glucan biosynthesis in bacteria), GH13_9 (glycogen branch synthesis), and GH13_11 (glycogen metabolism) were widely represented, with GH13_4 (amylosucrases) and GH13_20 significantly more abundant in Primates.

Consistent with our association analysis, a representation of the GH family abundance matrix, normalized to the number of genomes per host group, revealed distinct patterns in the distribution of particular GH families across host groups (Figure S1A) and indicated a consistent difference between the putative CAZy GH families present in the insect-associated bifidobacterial genomes compared with the other groups. For example, the GH5_18 (β-mannose), GH29 (α-fucose), GH32 (levanases), GH38 (α-mannosidases), GH43_22 (α-arabinofuranosidases and β-xylosidases), GH65 (α-glucose), GH78 (α-rhamnosidase), and GH146 (β-arabinofuranosidases) families are relatively more abundant in insects. Interestingly, many of these families are associated with breaking down plant glycans and polysaccharides. These findings highlight the functional adaptations of *Bifidobacterium* to different host environments, particularly in relation to carbohydrate metabolism.

Structural and functional diversity of GH13 enzymes in *Bifidobacterium* across diverse animal hosts

GH13 family sequences were the most abundantly represented in our dataset and showed associations with specific animal host groups. Due to these unique attributes, we postulated that this GH13 enzyme could be used as a sentinel family for assessing evolutionary functional processes. We utilized 4,303 amino acid sequences to construct a maximum-likelihood evolutionary tree (Figure S1B), which revealed well-defined clusters within GH13 subfamilies. The sequence diversity within these subfamilies was evident, as reflected by the presence of distinct subclusters, consistent with previous reports.[30]

To understand how this sequence diversity translates into functional potential, we compared the predicted structures of representative proteins from GH13 subfamilies that were particularly abundant in bifidobacteria associated with mammals in comparison to insects. Specifically, we analyzed the α-maltosyltransferase GH13_3, the α-glucan branching GH13_9, and the α-glucan debranching GH13_11 structures, revealing obvious differences in the presence of non-catalytic domains among enzymes from Primate-, Rodentia-, and Artiodactyla-associated species (Figure 5A). Superimposition of these structures with reference models complexed with their respective identified ligands (4U3C with maltohexaose, 5GQX with maltoheptaose, and 7U3B with acarbose, respectively) indicated that bifidobacterial enzymes likely have similar substrate affinities, with conserved ligand orientation and positioning across subfamilies—likely due to their function in glycogen metabolism.

We also compared the structures of GH13 subfamilies significantly associated with Primates vs. other mammals, including the GH13_4 amylosucrases and the GH13_20 (variety of activities), with a range of reference models complexed with their ligands (Figure S2). Our analysis suggested that bifidobacterial enzymes from these subfamilies might bind multiple substrates. For instance, the pocket-shaped active site of GH13_4 amylosucrases could accommodate both sucrose and maltohexaose, while the open cleft-shaped active sites of GH13_20 enzymes likely prefer cyclodextrins but may also bind other ligands, such as maltohexaose (*B. psychraerophilum* DSM 22366) and short-chain oligosaccharides (*B. longum* subsp. *longum* LH_12).

Notably, concerning the breakdown of α-linked glucose polysaccharides, mammalian-associated bifidobacteria are recognized as primary degraders of resistant starch (RS), with species such as *Bifidobacterium adolescentis*, *Bifidobacterium choerinum*, and *B. pseudolongum* exhibiting RS-degrading phenotypes.[31–33] RS has a compact crystalline structure, so the enzymes and binding modules that act on it have binding sites

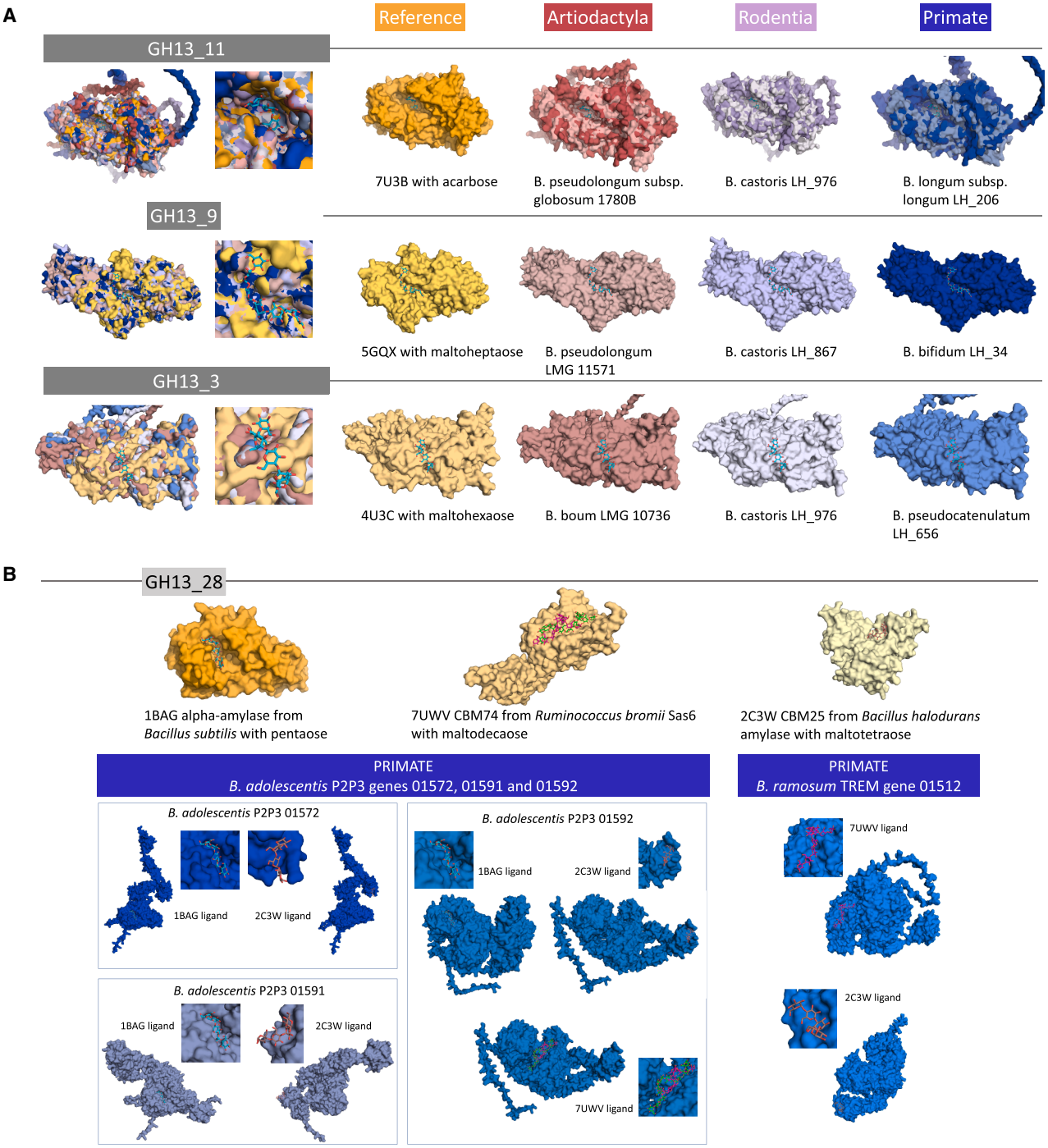


Figure 5. Analysis of selected bifidobacterial GH13 glycoside hydrolases

(A) Comparison of AlphaFold models generated for selected bifidobacterial sequences with structures of representative proteins from subfamilies indicated as particularly abundant in bifidobacteria associated with mammals: GH13_11 (glycogen metabolism) and GH13_9, and GH13_3 (glycogen synthesis).

(B) Comparison of AlphaFold models generated for GH13_28 sequences from selected primate bifidobacterial strains linked to the degradation of resistant starches, *Bifidobacterium adolescentis* P2P3 and *Bifidobacterium ramosum* TREM, with solved structures of selected GH13_28 proteins (1BAG), and specifically, starch-associated carbohydrate-binding modules CBM25 (2C3W) and CBM74 (7UWV) coupled with their respective ligands. Protein models from primate-associated strains are colored in the shades of blue, those from bifidobacteria isolated from hosts belonging to the order Artiodactyla in the shades of red, and those from rodent-associated *Bifidobacterium* in the shades of purple.

See also Figure S3.

with very different conformations to those that have specificity for soluble substrates. This capacity is linked to GH13_28 α-amylases containing specific CBMs, namely CBM25, CBM26, and CBM74.[33] We identified GH13_28 enzymes containing CBM74 and either CBM25 and/or the CBM26 modules in 57 out of the 387 genomes (14.7%), spanning 20 bifidobacterial species and subspecies. Most of these genomes (84.2%, n = 48) also contained additional α-amylases with either CBM25 or CBM26 modules alone. Further analysis revealed that some species, such as *Bifidobacterium castoris*, consistently possessed genes encoding GH13_28 α-amylases, while in other species, their presence was strain dependent.

We overlaid the structures of identified GH13_28 proteins from *Bifidobacterium* strains with reported RS-degrading phenotypes (*B. adolescentis* P2P3, *B. choerinum* FMB-1, and an in-house strain *B. castoris* LH_775) onto selected models of α-amylases (1BAG) and starch-recognizing modules CBM25 (2C3W) and CBM74 (7UWV) complexed with α-glucans (Figures 5B and S3). Successful overlay of different ligands onto the same structure allowed for the identification of the presence of both the catalytic domain and the CBMs in several of the selected bifidobacterial GH13_28 enzymes, confirming the presence of multiple carbohydrate-binding domains with specific substrate preferences. Binding sites for the 2C3W CBM25 ligand from *Bacillus halodurans* (maltotetraose) were conserved across analyzed *Bifidobacterium* GH13_28 structures. Similarly, we identified binding sites for either double or single helical starch (model 7UWV from *Ruminococcus bromii*) in bifidobacterial structures containing a predicted CBM74 domain. These results were replicated for other animal-associated species identified as potential RS degraders based on genomic predictions, namely *Bifidobacterium ramosum*, *Bifidobacterium thermophilum*, and *Bifidobacterium tsurumiense* (Figures 5B and S3). Our findings suggest that CAZy annotations coupled with structural modeling can provide valuable insights into the RS-degrading phenotype of bifidobacterial strains.

DISCUSSION

By combining taxonomic profiling, genomic, phylogenetic, and functional analyses, we demonstrate how host phylogeny and ecology shape the composition, function, and evolutionary trajectories of *Bifidobacterium* at a strain-resolved level.

Our data indicate that gut microbiome composition is significantly influenced by host phylogeny, particularly in mammals. At the strain level, *Bifidobacterium* displayed strong host-specific co-phylogenetic associations, likely driven by vertical transmission and dietary factors. These results align with previous studies that have suggested strong signals of phylosymbiosis between mammalian hosts and their gut microbiota,[34,35] yet our dataset extends these findings by showing that associations of *Bifidobacterium* are stronger in hosts with specific dietary niches. For instance, high *Bifidobacterium* abundance and strong phylogenetic signals in Primates, particularly marmosets and tamarins, are likely driven by diets rich in complex carbohydrates and fibrous plant material,[36] resulting in persistence of bifidobacterial species adept at glycan utilization. Additionally, the rise of bacterial order Veillonellales-Selenomonadales in the ancestors of the New World monkeys, positively covarying with bacterial order Bifidobacteriales, suggests potential cross-feeding interactions. This mirrors interactions observed in humans, where scavenging *Veillonella* and *Megasphaera* species ferment lactate produced by bifidobacteria to acetate/propionate and butyrate, respectively, indicating conserved microbial interaction networks across hosts.[37,38]

A key finding was the robust co-phylogenetic signal between rodents and their *Bifidobacterium* strains, despite our earlier studies suggesting no such relationship in wild mice (genus *Apodemus*).[39] This shift in observations likely stems from the expanded dataset and different methodological approach in this study, which included a host timetree, a broader range of *Bifidobacterium* strains, and the GTDB-tk bac120 marker set, and highlights the importance of comprehensive genomic datasets and robust phylogenetic methods in revealing patterns of host-microbe co-evolution.[40] These results also underscore the importance of vertical transmission in maintaining phylogenetic congruence at the strain level and suggest that co-evolution between *Bifidobacterium* and their hosts is nuanced and complex. The evolutionary stability of these host-microbe relationships at the strain level suggests that vertical transmission potentially overshadows ecological factors such as cross-species transmission and raises important questions about the mechanisms driving phylosymbiosis, including the potential role of additional factors, such as microbial interactions, with the host immune system, strain retention due to nutrient availability, or the role of host genotype in host-microbe adaptations. For example, in humans, the lactase non-persistent host genotype, characterized by the lower expression of the *LCT* gene, has previously been associated with stronger selection for lactose-degrading microbiota members.[41] The increased relative abundance of *Bifidobacterium* in lactase non-persistent hosts has been linked to the particular *LCT* locus (rs4988235) and milk consumption, supporting the notion that the mammalian lactase enzyme and bifidobacterial β-galactosidase enzyme are in direct competition for lactose.[42]

In contrast to mammalian hosts, we observed a weaker phylosymbiosis signal in birds, likely reflecting the more variable nature of their gut microbiota, previously suggested to result from the evolution of powered flight.[43,44] Although the smaller sample size limits our conclusions here, the ancestral reconstruction of microbiota shifts in bird lineages, particularly the increased abundance of bacterial orders Enterobacteriales and Clostridiales in Neoaves, mirrors the dietary shifts toward carnivory in Carnivora. These dietary changes appear to be a driving force behind the microbiota shifts we observed in both avian and mammalian lineages.[45] Interestingly, a correlation between species of *Clostridium* and animal-based diets has also been observed in humans.[46] Members of this genus have previously been identified as important degraders of amino acids, including lysine, alanine, and glycine, particularly abundant in meat and meat products.[47,48] Overall, our findings point to a complex interplay between host phylogeny, diet, and vertical transmission, rather than purely ecological dynamics driving these patterns.

Our functional analyses revealed striking host-specific metabolic adaptations in *Bifidobacterium*, particularly in carbohydrate metabolism and oxidative stress responses. Insect-associated *Bifidobacterium* were enriched in oxidative phosphorylation

genes. This finding is particularly intriguing given the anoxic nature of insect guts,[49] suggesting that bifidobacterial cytochrome *bd* oxidases might play a yet unclear role. One possibility is defense against reactive oxygen species (ROS) released into the insect gut environment as part of the innate immune system, a mechanism that has also been seen in *E. coli* and *Mycobacterium smegmatis*.[50,51] These findings raise the possibility that *Bifidobacterium* has evolved unique mechanisms to mitigate oxidative stress, an adaptation that could represent a previously unappreciated aspect of insect gut symbiosis.

In mammals, the diversity and abundance of bifidobacterial GH enzymes, particularly those involved in glycogen metabolism (GH13_9 and GH13_11), reflect adaptations to complex carbohydrate-rich diets. Consistent with previous findings, our study also shows the absence of these functions in insect-associated *Bifidobacterium*, likely reflecting the nectar- and pollen-based diets of their hosts, which lack complex α-glucoside linkages.[52,53] The absence of glycogen metabolism suggests an evolutionary loss, potentially due to reductive genome evolution, a phenomenon previously observed in obligate intracellular symbionts.[54,55] However, since insect-associated *Bifidobacterium* are not obligate symbionts, these strains may rely on glycogen metabolism by the host or other microbiota members, highlighting an example of microbial interdependence within the gut ecosystem.

Moreover, the identification of the presence of multi-domain GH13_28 α-amylases containing CBM25, CBM26, and the CBM74 module in *Bifidobacterium* across various hosts highlights the genus's ecological flexibility. As recently demonstrated in *R. bromii*,[56] enzymes containing these modules are crucial for the efficient degradation of both short- and long-chain starch molecules, including RS, which are prevalent in herbivorous diets. Indeed, our structural modeling confirms the affinity of these carbohydrate-binding domains for their respective ligands, highlighting the metabolic versatility of *Bifidobacterium*. This underscores the adaptive potential of *Bifidobacterium* to diverse dietary environments and suggests potential targets for enhancing their starch-degrading capabilities in animal and human probiotic applications.

Our study has limitations. Sample transport and storage may have impacted bacterial isolation and DNA yield and quality.[57] While ideal conditions involve immediate sample freezing at −80°C, this is not always possible due to logistical constraints.[58] Nevertheless, all samples were frozen at −20°C or −80°C within 48 h. Previous studies show that fecal microbiota remains stable at room temperature for 24–72 h and under long-term −80°C storage, with minimal effects on major taxa.[59–61] Although freeze-thaw cycles can affect community profiles,[62] these were minimized in our study.

Additionally, differences in sequencing protocols may have introduced bias.[63,64] Samples collected in Cyprus and the UK were sequenced using different hypervariable regions (V1-V2 vs. V4). Although different regions may affect taxonomic resolution, studies suggest these particular regions provide qualitatively similar representations of core microbial communities.[65] Indeed, a recent study of 168,000 human microbiome samples found only 15 genera with significant differences across V1-V6 hypervariable regions, mostly within V4 and V3-V4.[66] Thus, comparisons remain valid, particularly for broad community patterns.

In our genomic analyses, we acknowledge overrepresentation of certain *Bifidobacterium* species. However, this reflects known host-specific variation in bifidobacterial diversity, e.g., primates often harbor more diverse *Bifidobacterium* communities than rodents.[8,23] Additionally, strains with high genomic similarity (ANI > 99%) were found across host species and taxonomic orders, underscoring natural ecological overlap. We therefore consider these patterns to be biologically meaningful rather than artifacts of sampling and/or statistical bias.

In conclusion, this research significantly broadens our understanding of the evolutionary biology and functional ecology of *Bifidobacterium*. By integrating diverse approaches, we have revealed how these microbes adapt to their hosts through a balance of vertical transmission, dietary influences, and host-specific interactions. This work provides a foundation for targeted therapeutic interventions tailored to specific host diets or ecological niches. Future research should explore how dietary and ecological components influence microbiota across hosts, enhancing our understanding of gut microbiota dynamics and their functional roles in health and disease.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Lindsay J. Hall (l.hall.3@bham.ac.uk).

Materials availability

Bacterial strains isolated as part of the study can be requested from the lead contact with a completed materials transfer agreement.

Data and code availability

- 16S rRNA amplicon sequencing data analyzed in this study have been deposited to the NCBI SRA: PRJNA1200941. The draft genomes of 96 *Bifidobacterium* isolates sequenced here have been deposited to the NCBI Genomes :PRJNA1200594. The code and associated files have been deposited to Figshare: https://doi.org/10.6084/m9.figshare.29635103.
- The study does not make use of unpublished data or software.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We would like to thank the zoo staff at Banham Zoo (UK), Africa Alive! (UK), and Pafos Zoo (Cyprus) for their assistance in collecting samples for this study. This work was funded by Wellcome Trust Investigator Awards 100/974/C/13/Z and 220540/Z/20/A, a BBSRC Norwich Research Park Bioscience Doctoral Training grant no. BB/M011216/1 (supervisor, L.J.H.; student, M.K.), an Institute Strategic Programme Gut Microbes and Health grant no. BB/R012490/1 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356, and a BBSRC Institute Strategic Programme Food Microbiome and Health BB/X011054/1 and its constituent project BBS/E/QU/230001B to L.J.H.

AUTHOR CONTRIBUTIONS

L.J.H. and M.K. designed the overall study. S.G. and I.C. were responsible for sample collection. J.M. and L. Chalklen processed fecal samples and extracted DNA for the 16S rRNA amplicon sequencing. M.K. pre-processed the 16S rRNA sequencing reads. S.M. performed initial analyses of the 16S rRNA sequencing data. Taxonomic analyses were performed by D.S., and phylosymbiosis analysis was performed by M.K. J.M., R.K., and M.K. processed fecal samples for bacterial isolations, isolated *Bifidobacterium* strains,

and extracted genomic DNA for WGS. M.K. performed all genomic, evolutionary, and functional prediction analysis and visualized the data. M.K. and L.J.H. drafted the manuscript, and D.S. and L. Crouch provided further edits and co-writing of the final version. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Faecal sample collection
- METHOD DETAILS
  - DNA extraction from faecal samples and 16S rRNA amplicon sequencing
  - 16S rRNA amplicon sequencing data analysis
  - Isolation of bifidobacteria from faecal samples
  - Genomic DNA extraction from bacterial isolates
  - Whole genome sequencing of bacterial isolates
  - Genomic data processing and compilation of the final dataset
  - Phylogenetic analysis
  - Analysis of *Bifidobacterium* strain interactions in hosts
  - Co-phylogeny analysis
  - Genomic and functional annotation
  - Assessment of between-group differences in abundance of functional features
  - Prediction of protein structure and ligand binding of selected glycoside hydrolases
  - Data visualisation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2025.08.008.

Received: February 10, 2025
Revised: May 29, 2025
Accepted: August 8, 2025
Published: September 10, 2025

REFERENCES

1. Gotoh, A., Katoh, T., Sakanaka, M., Ling, Y.W., Yamada, C., Asakuma, S., Urashima, T., Tomabechi, Y., Katayama-Ikegami, A., Kurihara, S., et al. (2018). Sharing of human milk oligosaccharides degradants within bifidobacterial communities in faecal cultures supplemented with Bifidobacterium bifidum. Sci. Rep. *8*, 13958. https://doi.org/10.1038/s41598-018-32080-3.
2. Thomson, P., Medina, D.A., and Garrido, D. (2018). Human milk oligosaccharides and infant gut bifidobacteria: Molecular strategies for their utilization. Food Microbiol. *75*, 37–46. https://doi.org/10.1016/j.fm.2017.09.001.
3. Lawson, M.A.E., O'Neill, I.J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z., Chalklen, L., and Hall, L.J. (2020). Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. ISME J. *14*, 635–648. https://doi.org/10.1038/s41396-019-0553-2.
4. Milani, C., Turroni, F., Duranti, S., Lugli, G.A., Mancabelli, L., Ferrario, C., van Sinderen, D., and Ventura, M. (2016). Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. Appl. Environ. Microbiol. *82*, 980–991. https://doi.org/10.1128/AEM.03500-15.
5. Lugli, G.A., Milani, C., Turroni, F., Duranti, S., Mancabelli, L., Mangifesta, M., Ferrario, C., Modesto, M., Mattarelli, P., Jiří, K., et al. (2017). Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family. BMC Genomics *18*, 568. https://doi.org/10.1186/s12864-017-3955-4.
6. Lugli, G.A., Mancino, W., Milani, C., Duranti, S., Mancabelli, L., Napoli, S., Mangifesta, M., Viappiani, A., Anzalone, R., Longhi, G., et al. (2019). Dissecting the Evolutionary Development of the Species Bifidobacterium animalis through Comparative Genomics Analyses. Appl. Environ. Microbiol. *85*, e02806-18. https://doi.org/10.1128/AEM.02806-18.
7. Lugli, G.A., Duranti, S., Albert, K., Mancabelli, L., Napoli, S., Viappiani, A., Anzalone, R., Longhi, G., Milani, C., Turroni, F., et al. (2019). Unveiling Genomic Diversity among Members of the Species Bifidobacterium pseudolongum, a Widely Distributed Gut Commensal of the Animal Kingdom. Appl. Environ. Microbiol. *85*, e03065-18. https://doi.org/10.1128/AEM.03065-18.
8. Milani, C., Mangifesta, M., Mancabelli, L., Lugli, G.A., James, K., Duranti, S., Turroni, F., Ferrario, C., Ossiprandi, M.C., van Sinderen, D., et al. (2017). Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. ISME J. *11*, 2834–2847. https://doi.org/10.1038/ismej.2017.138.
9. Sayers, E.W., Beck, J., Brister, J.R., Bolton, E.E., Canese, K., Comeau, D.C., Funk, K., Ketter, A., Kim, S., Kimchi, A., et al. (2020). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *48*, D9–D16. https://doi.org/10.1093/nar/gkz899.
10. Orkin, J.D., Campos, F.A., Myers, M.S., Cheves Hernandez, S.E., Guadamuz, A., and Melin, A.D. (2019). Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. ISME J. *13*, 183–196. https://doi.org/10.1038/s41396-018-0256-0.
11. Modrackova, N., Stovicek, A., Burtscher, J., Bolechova, P., Killer, J., Domig, K.J., and Neuzil-Bunesova, V. (2021). The bifidobacterial distribution in the microbiome of captive primates reflects parvorder and feed specialization of the host. Sci. Rep. *11*, 15273. https://doi.org/10.1038/s41598-021-94824-y.
12. Malukiewicz, J., D'Arc, M., Dias, C.A., Cartwright, R.A., Grativol, A.D., Moreira, S.B., Souza, A.R., Tavares, M.C.H., Pissinatti, A., Ruiz-Miranda, C.R., et al. (2023). Bifidobacteria define gut microbiome profiles of golden lion tamarin (Leontopithecus rosalia) and marmoset (Callithrix sp.) metagenomic shotgun pools. Sci. Rep. *13*, 15679. https://doi.org/10.1038/s41598-023-42059-4.
13. Lim, S.J., and Bordenstein, S.R. (2020). An introduction to phylosymbiosis. Proc. Biol. Sci. *287*, 20192900. https://doi.org/10.1098/rspb.2019.2900.
14. Legendre, P., Desdevises, Y., and Bazin, E. (2002). A statistical test for host-parasite coevolution. Syst. Biol. *51*, 217–234. https://doi.org/10.1080/10635150252899734.
15. Page, R.D.M. (2003). Tangled Trees: Phylogeny, Cospeciation, and Coevolution (University of Chicago Press).
16. Balbuena, J.A., Míguez-Lozano, R., and Blasco-Costa, I. (2013). PACo: A Novel Procrustes Application to Cophylogenetic Analysis. PLoS One *8*, e61048. https://doi.org/10.1371/journal.pone.0061048.
17. Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., Muller, M.N., Pusey, A.E., Peeters, M., Hahn, B.H., and Ochman, H. (2016). Cospeciation of gut microbiota with hominids. Science *353*, 380–382. https://doi.org/10.1126/science.aaf3951.
18. Lugli, G.A., Alessandri, G., Milani, C., Mancabelli, L., Ruiz, L., Fontana, F., Borragán, S., González, A., Turroni, F., Ossiprandi, M.C., et al. (2020). Evolutionary development and co-phylogeny of primate-associated bifidobacteria. Environ. Microbiol. *22*, 3375–3393. https://doi.org/10.1111/1462-2920.15108.
19. Lugli, G.A., Milani, C., Duranti, S., Mancabelli, L., Mangifesta, M., Turroni, F., Viappiani, A., van Sinderen, D., and Ventura, M. (2018). Tracking the

taxonomy of the genus Bifidobacterium based on a phylogenomic approach. Appl. Environ. Microbiol. *84*, e02249-17. https://doi.org/10.1128/AEM.02249-17.

20. Rodriguez, C.I., and Martiny, J.B.H. (2020). Evolutionary relationships among bifidobacteria and their hosts and environments. BMC Genomics *21*, 26. https://doi.org/10.1186/s12864-019-6435-1.
21. Berkhout, M.D., Plugge, C.M., and Belzer, C. (2022). How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. Glycobiology *32*, 182–200. https://doi.org/10.1093/glycob/cwab105.
22. La Rosa, S.L., Ostrowski, M.P., Vera-Ponce de León, A., McKee, L.S., Larsbrink, J., Eijsink, V.G., Lowe, E.C., Martens, E.C., and Pope, P.B. (2022). Glycan processing in gut microbiomes. Curr. Opin. Microbiol. *67*, 102143. https://doi.org/10.1016/j.mib.2022.102143.
23. Alessandri, G., van Sinderen, D., and Ventura, M. (2021). The genus bifidobacterium: From genomics to functionality of an important component of the mammalian gut microbiota running title: Bifidobacterial adaptation to and interaction with the host. Comput. Struct. Biotechnol. J. *19*, 1472–1487. https://doi.org/10.1016/j.csbj.2021.03.006.
24. Hird, S.M., Sánchez, C., Carstens, B.C., and Brumfield, R.T. (2015). Comparative gut microbiota of 59 neotropical bird species. Front. Microbiol. *6*, 1403. https://doi.org/10.3389/fmicb.2015.01403.
25. Amato, K.R., G Sanders, J., Song, S.J., Nute, M., Metcalf, J.L., Thompson, L.R., Morton, J.T., Amir, A., J McKenzie, V., Humphrey, G., et al. (2019). Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. ISME J. *13*, 576–587. https://doi.org/10.1038/s41396-018-0175-0.
26. Brooks, A.W., Kohl, K.D., Brucker, R.M., van Opstal, E.J., and Bordenstein, S.R. (2016). Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. PLoS Biol. *14*, e2000225. https://doi.org/10.1371/journal.pbio.2000225.
27. Perez-Lamarque, B., Sommeria-Klein, G., Duret, L., and Morlon, H. (2023). Phylogenetic Comparative Approach Reveals Evolutionary Conservatism, Ancestral Composition, and Integration of Vertebrate Gut Microbiota. Mol. Biol. Evol. *40*, msad144. https://doi.org/10.1093/molbev/msad144.
28. Zhang, L., Liu, Z., Sun, K., Jin, L., Yu, J., and Wang, H. (2024). Multi-dimensional niche differentiation of two sympatric breeding secondary cave-nesting birds in Northeast China using DNA metabarcoding. Ecol. Evol. *14*, e11709. https://doi.org/10.1002/ece3.11709.
29. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. *44*, D457–D462. https://doi.org/10.1093/nar/gkv1070.
30. Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng. Des. Sel. *19*, 555–562. https://doi.org/10.1093/protein/gzl044.
31. Jung, D.-H., Seo, D.-H., Kim, G.-Y., Nam, Y.-D., Song, E.-J., Yoon, S., and Park, C.-S. (2018). The effect of resistant starch (RS) on the bovine rumen microflora and isolation of RS-degrading bacteria. Appl. Microbiol. Biotechnol. *102*, 4927–4936. https://doi.org/10.1007/s00253-018-8971-z.
32. Jung, D.-H., Kim, G.-Y., Kim, I.-Y., Seo, D.-H., Nam, Y.-D., Kang, H., Song, Y., and Park, C.-S. (2019). Bifidobacterium adolescentis P2P3, a Human Gut Bacterium Having Strong Non-Gelatinized Resistant Starch-Degrading Activity. J. Microbiol. Biotechnol. *29*, 1904–1915. https://doi.org/10.4014/jmb.1909.09010.
33. Jung, D.-H., and Park, C.-S. (2023). Resistant starch utilization by Bifidobacterium, the beneficial human gut bacteria. Food Sci. Biotechnol. *32*, 441–452. https://doi.org/10.1007/s10068-023-01253-w.
34. Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., and Alm, E.J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat. Commun. *8*, 14319. https://doi.org/10.1038/ncomms14319.
35. Nishida, A.H., and Ochman, H. (2018). Rates of Gut Microbiome Divergence in Mammals. Mol. Ecol. *27*, 1884–1897. https://doi.org/10.1111/mec.14473.
36. Susanne, R., and Ann-Kathrin, O. (2007). Husbandry and Management of New World Species: Marmosets and Tamarins. Lab. Primate *145*. https://doi.org/10.1016/B978-012080261-6/50010-6.
37. Button, J.E., Cosetta, C.M., Reens, A.L., Brooker, S.L., Rowan-Nash, A. D., Lavin, R.C., Saur, R., Zheng, S., Autran, C.A., Lee, M.L., et al. (2023). Precision modulation of dysbiotic adult microbiomes with a human-milk-derived synbiotic reshapes gut microbial composition and metabolites. Cell Host Microbe *31*, 1523–1538.e10. https://doi.org/10.1016/j.chom.2023.08.004.
38. Zhao, S., Lau, R., Zhong, Y., and Chen, M.-H. (2023). Lactate cross-feeding between Bifidobacterium species and Megasphaera indica contributes to butyrate formation in the human colonic environment. Appl. Environ. Microbiol. *90*, e01019-23. https://doi.org/10.1128/aem.01019-23.
39. Kujawska, M., Raulo, A., Millar, M., Warren, F., Baltrūnaitė, L., Knowles, S.C.L., and Hall, L.J. (2022). Bifidobacterium castoris strains isolated from wild mice show evidence of frequent host switching and diverse carbohydrate metabolism potential. ISME Commun. *2*, 20. https://doi.org/10.1038/s43705-022-00102-x.
40. Hayward, A., Poulin, R., and Nakagawa, S. (2021). A broadscale analysis of host-symbiont cophylogeny reveals the drivers of phylogenetic congruence. Ecol. Lett. *24*, 1681–1696. https://doi.org/10.1111/ele.13757.
41. Suzuki, T.A., and Ley, R.E. (2020). The role of the microbiota in human genetic adaptation. Science *370*, eaaz6827. https://doi.org/10.1126/science.aaz6827.
42. Goodrich, J.K., Davenport, E.R., Clark, A.G., and Ley, R.E. (2017). The relationship between the human genome and microbiome comes into view. Annu. Rev. Genet. *51*, 413–433. https://doi.org/10.1146/annurev-genet-110711-155532.
43. Grond, K., Sandercock, B.K., Jumpponen, A., and Zeglin, L.H. (2018). The avian gut microbiota: community, physiology and function in wild birds. J. Avian Biol. *49*, e01788. https://doi.org/10.1111/jav.01788.
44. Song, S.J., Sanders, J.G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., Mazel, F., Lutz, H.L., Winker, K., Graves, G.R., et al. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. mBio *11*, e02901-19. https://doi.org/10.1128/mBio.02901-19.
45. Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R.E., and Farnleitner, A.H. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nat. Commun. *10*, 2200. https://doi.org/10.1038/s41467-019-10191-3.
46. Asnicar, F., Berry, S.E., Valdes, A.M., Nguyen, L.H., Piccinno, G., Drew, D.A., Leeming, E., Gibson, R., Le Roy, C.L., Khatib, H.A., et al. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat. Med. *27*, 321–332. https://doi.org/10.1038/s41591-020-01183-8.
47. Smith, E.A., and Macfarlane, G.T. (1998). Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. FEMS Microbiol. Ecol. *25*, 355–368. https://doi.org/10.1111/j.1574-6941.1998.tb00487.x.
48. Dai, Z., Zheng, W., and Locasale, J.W. (2022). Amino acid variability, tradeoffs and optimality in human diet. Nat. Commun. *13*, 6683. https://doi.org/10.1038/s41467-022-34486-0.
49. Zheng, H., Powell, J.E., Steele, M.I., Dietrich, C., and Moran, N.A. (2017). Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. Proc. Natl. Acad. Sci. USA *114*, 4775–4780. https://doi.org/10.1073/pnas.1701819114.
50. Lindqvist, A., Membrillo-Hernández, J., Poole, R.K., and Cook, G.M. (2000). Roles of respiratory oxidases in protecting Escherichia coli K12

from oxidative stress. Antonie Leeuwenhoek *78*, 23–31. https://doi.org/10.1023/A:1002779201379.

51. Lu, P., Heineke, M.H., Koul, A., Andries, K., Cook, G.M., Lill, H., van Spanning, R., and Bald, D. (2015). The cytochrome bd-type quinol oxidase is important for survival of Mycobacterium smegmatis under peroxide and antibiotic-induced stress. Sci. Rep. *5*, 10333. https://doi.org/10.1038/srep10333.
52. Bottacini, F., Milani, C., Turroni, F., Sánchez, B., Foroni, E., Duranti, S., Serafini, F., Viappiani, A., Strati, F., Ferrarini, A., et al. (2012). Bifidobacterium asteroides PRL2011 Genome Analysis Reveals Clues for Colonization of the Insect Gut. PLoS One *7*, e44229. https://doi.org/10.1371/journal.pone.0044229.
53. Sun, Z., Zhang, W., Guo, C., Yang, X., Liu, W., Wu, Y., Song, Y., Kwok, L. Y., Cui, Y., Menghe, B., et al. (2015). Comparative genomic analysis of 45 type strains of the genus Bifidobacterium: a snapshot of its genetic diversity and evolution. PLoS One *10*, e0117912. https://doi.org/10.1371/journal.pone.0117912.
54. Henrissat, B., Deleury, E., and Coutinho, P.M. (2002). Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? Trends Genet. *18*, 437–440. https://doi.org/10.1016/S0168-9525(02)02734-8.
55. Esteban-Torres, M., Ruiz, L., Rossini, V., Nally, K., and van Sinderen, D. (2023). Intracellular glycogen accumulation by human gut commensals as a niche adaptation trait. Gut Microbes *15*, 2235067. https://doi.org/10.1080/19490976.2023.2235067.
56. Photenhauer, A.L., Villafuerte-Vega, R.C., Cerqueira, F.M., Armbruster, K.M., Mareček, F., Chen, T., Wawrzak, Z., Hopkins, J.B., Vander Kooi, C.W., Janeček, Š., et al. (2024). The Ruminococcus bromii amylosome protein Sas6 binds single and double helical α-glucan structures in starch. Nat. Struct. Mol. Biol. *31*, 255–265. https://doi.org/10.1038/s41594-023-01166-6.
57. Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R., and Ley, R.E. (2014). Conducting a Microbiome Study. Cell *158*, 250–262. https://doi.org/10.1016/j.cell.2014.06.037.
58. Human; Microbiome; Project Consortium (2012). A framework for human microbiome research. Nature *486*, 215–221. https://doi.org/10.1038/nature11209.
59. Carroll, I.M., Ringel-Kulka, T., Siddle, J.P., Klaenhammer, T.R., and Ringel, Y. (2012). Characterization of the Fecal Microbiota Using High-Throughput Sequencing Reveals a Stable Microbial Community during Storage. PLoS One *7*, e46953. https://doi.org/10.1371/journal.pone.0046953.
60. Dominianni, C., Wu, J., Hayes, R.B., and Ahn, J. (2014). Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol. *14*, 103. https://doi.org/10.1186/1471-2180-14-103.
61. Lauber, C.L., Zhou, N., Gordon, J.I., Knight, R., and Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol. Lett. *307*, 80–86. https://doi.org/10.1111/j.1574-6968.2010.01965.x.
62. Sergeant, M.J., Constantinidou, C., Cogan, T., Penn, C.W., and Pallen, M.J. (2012). High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length and annealing temperature. PLoS One *7*, e38094. https://doi.org/10.1371/journal.pone.0038094.
63. Kuczynski, J., Lauber, C.L., Walters, W.A., Parfrey, L.W., Clemente, J.C., Gevers, D., and Knight, R. (2011). Experimental and analytical tools for studying the human microbiome. Nat. Rev. Genet. *13*, 47–58. https://doi.org/10.1038/nrg3129.
64. Yang, B., Wang, Y., and Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics *17*, 135. https://doi.org/10.1186/s12859-016-0992-y.
65. Ibarbalz, F.M., Pérez, M.V., Figuerola, E.L.M., and Erijman, L. (2014). The Bias Associated with Amplicon Sequencing Does Not Affect the Quantitative Assessment of Bacterial Community Dynamics. PLoS One *9*, e99722. https://doi.org/10.1371/journal.pone.0099722.
66. Abdill, R.J., Graham, S.P., Rubinetti, V., Ahmadian, M., Hicks, P., Chetty, A., McDonald, D., Ferretti, P., Gibbons, E., Rossi, M., et al. (2025). Integration of 168,000 samples reveals global patterns of the human gut microbiome. Cell *188*, 1100–1118.e17. https://doi.org/10.1016/j.cell.2024.12.017.
67. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
68. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.
69. Andersen, K.S., Kirkegaard, R.H., Karst, S.M., and Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. Preprint at bioRxiv. https://doi.org/10.1101/299537.
70. R Core Team. (2017) R: A Language and Environment for Statistical Computing (The R Foundation).
71. Kolde, R. (2019) Pheatmap: Pretty Heatmaps. https://doi.org/10.32614/CRAN.package.pheatmap.
72. Smith, S.D. (2019). phylosmith: an R-package for reproducible and efficient microbiome analysis with phyloseq-objects. JOSS *4*. https://doi.org/10.21105/joss.01442.
73. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., T. van den Brand, , et al. (2024) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. https://doi.org/10.32614/CRAN.package.ggplot2.
74. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol. *34*, 1812–1819. https://doi.org/10.1093/molbev/msx116.
75. Guo, J., Gabry, J., Goodrich, B., Johnson, A., Weber, S., Badr, H.S., Lee, D., Sakrejda, K., Martin, M., University, T. of C., and et al. (2024). Rstan: R Interface to Stan. https://doi.org/10.32614/CRAN.package.rstan.
76. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. https://doi.org/10.1101/gr.186072.114.
77. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.
78. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics *38*, 5315–5316. https://doi.org/10.1093/bioinformatics/btac672.
79. Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., and Toth, I.K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods *8*, 12–24. https://doi.org/10.1039/C5AY02550H.
80. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
81. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. *52*, W78–W82. https://doi.org/10.1093/nar/gkae268.
82. Lam, T.J., Stamboulian, M., Han, W., and Ye, Y. (2020). Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. PLoS Comput. Biol. *16*, e1007951. https://doi.org/10.1371/journal.pcbi.1007951.
83. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

84. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

85. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L. J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122. https://doi.org/10.1093/molbev/msx148.

86. Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., and Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. *51*, W115–W121. https://doi.org/10.1093/nar/gkad328.

87. Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Posit Software, PBC, and Wickham, H. (2023) Dplyr: A Grammar of Data Manipulation. https://doi.org/10.32614/CRAN.package.dplyr.

88. Wilke, C.O. (2024). Cowplot: streamlined Plot Theme and Plot Annotations for "ggplot2.". https://doi.org/10.32614/CRAN.package.cowplot.

89. Conway, J., and Gehlenborg, N. (2019) UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. https://doi.org/10.32614/CRAN.package.UpSetR.

90. Kassambara, A., and Mundt, F. (2020) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. https://doi.org/10.32614/CRAN.package.factoextra.

91. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

92. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D *66*, 486–501. https://doi.org/10.1107/S0907444910007493.

93. DeLano, W.L. (2024). The PyMOL Molecular Graphics System (Schrödinger, LLC).

94. Bibekar, P., Krapp, L., and Peraro, M.D. (2024). PeSTo-Carbs: Geometric Deep Learning for Prediction of Protein–Carbohydrate Binding Interfaces. J. Chem. Theor. Comput. *20*, 2985–2991. https://doi.org/10.1021/acs.jctc.3c01145.

95. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

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

97. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, r., Legendre, P., McGillin, D., Minchin, P.R., O'Hara, R.B., Simpson, G.B., Solymos, P., et al. (2019) Vegan: Community Ecology Package. .https://doi.org/10.32614/CRAN.package.vegan

98. Kassambara, A. (2023) Rstatix: Pipe-Friendly Framework for Basic Statistical Tests. https://doi.org/10.32614/CRAN.package.rstatix.

99. Paradis, E., and Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

100. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596. https://doi.org/10.1093/nar/gks1219.

101. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L. J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

102. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

103. Walker, A.W., Martin, J.C., Scott, P., Parkhill, J., Flint, H.J., and Scott, K. P. (2015). 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. Microbiome *3*, 26. https://doi.org/10.1186/s40168-015-0087-4.

104. Suzuki, M.T., and Giovannoni, S.J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. *62*, 625–630. https://doi.org/10.1128/aem.62.2.625-630.1996.

105. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Appl. Environ. Microbiol. *79*, 5112–5120. https://doi.org/10.1128/AEM.01043-13.

106. Page, A.J., De Silva, N., Hunt, M., Quail, M.A., Parkhill, J., Harris, S.R., Otto, T.D., and Keane, J.A. (2016). Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb. Genom. *2*, e000083. https://doi.org/10.1099/mgen.0.000083.

107. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

108. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput. Biol. *13*, e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

109. Oren, A., Arahal, D.R., Göker, M., Moore, E.R.B., Rossello-Mora, R., and Sutcliffe, I.C. (2023). International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). Int. J. Syst. Evol. Microbiol. *73*. https://doi.org/10.1099/ijsem.0.005585.

110. Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.-W., De Meyer, S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol. *68*, 461–466. https://doi.org/10.1099/ijsem.0.002516.

111. Rühlemann, M.C., Bang, C., Gogarten, J.F., Hermes, B.M., Groussin, M., Waschina, S., Poyet, M., Ulrich, M., Akoua-Koffi, C., Deschner, T., et al. (2024). Functional host-specific adaptation of the intestinal microbiome in hominids. Nat. Commun. *15*, 326. https://doi.org/10.1038/s41467-023-44636-7.

112. Inkscape Developers (2024). Inkscape (Inkscape Project).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Bifidobacterium spp. isolates	This paper	Listed in Table S2
Biological samples		
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z1
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z2
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z3
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z4
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z5
Pogona vitticeps faecal sample	Banham Zoo / Africa Alive	Z6
Genetta sp. faecal sample	Banham Zoo / Africa Alive	Z7
Eurycantha calcarata faecal sample	Banham Zoo / Africa Alive	Z8
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z9
Platalea sp. faecal sample	Banham Zoo / Africa Alive	Z10
Eulemur rubriventer faecal sample	Banham Zoo / Africa Alive	Z11
Chrysocyon brachyurus faecal sample	Banham Zoo / Africa Alive	Z12
Vicugna pacos faecal sample	Banham Zoo / Africa Alive	Z13
Dasyprocta sp. faecal sample	Banham Zoo / Africa Alive	Z14
Osphranter rufus faecal sample	Banham Zoo / Africa Alive	Z15
Saguinus Oedipus faecal sample	Banham Zoo / Africa Alive	Z16
Gromphadorhina portentosa faecal sample	Banham Zoo / Africa Alive	Z17
Pudu sp. faecal sample	Banham Zoo / Africa Alive	Z18
Chelonoidis carbonaria faecal sample	Banham Zoo / Africa Alive	Z19
Ailurus fulgens faecal sample	Banham Zoo / Africa Alive	Z20
Epibolus pulchripes faecal sample	Banham Zoo / Africa Alive	Z21
Colobus sp. faecal sample	Banham Zoo / Africa Alive	Z22
Aonyx cinerea faecal sample	Banham Zoo / Africa Alive	Z23
Panthera tigris altaica faecal sample	Banham Zoo / Africa Alive	Z24
Achatina fulica faecal sample	Banham Zoo / Africa Alive	Z25
Hapalemur aureus faecal sample	Banham Zoo / Africa Alive	Z26
Cynomys sp. faecal sample	Banham Zoo / Africa Alive	Z27
Branta ruficollis faecal sample	Banham Zoo / Africa Alive	Z28
Chaetophractus villosus faecal sample	Banham Zoo / Africa Alive	Z29
Suricata suricatta faecal sample	Banham Zoo / Africa Alive	Z30
Leopardus geoffroyi faecal sample	Banham Zoo / Africa Alive	Z31
Cervus nippon faecal sample	Banham Zoo / Africa Alive	Z32
Giraffa sp. faecal sample	Banham Zoo / Africa Alive	Z33
Helogale parvula faecal sample	Banham Zoo / Africa Alive	Z34
Varecia variegata faecal sample	Banham Zoo / Africa Alive	Z35
Spheniscus demersus faecal sample	Banham Zoo / Africa Alive	Z36
Haliaeetus leucocephalus faecal sample	Banham Zoo / Africa Alive	Z37
Pogona vitticeps faecal sample	Banham Zoo / Africa Alive	Z38
Helogale parvula faecal sample	Banham Zoo / Africa Alive	Z39
Chelonoidis carbonaria faecal sample	Banham Zoo / Africa Alive	Z40
Dasyprocta sp. faecal sample	Banham Zoo / Africa Alive	Z41

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cervus nippon faecal sample	Banham Zoo / Africa Alive	Z42
Ailurus fulgens faecal sample	Banham Zoo / Africa Alive	Z43
Leopardus geoffroyi faecal sample	Banham Zoo / Africa Alive	Z44
Colobus sp. faecal sample	Banham Zoo / Africa Alive	Z45
Epibolus pulchripes faecal sample	Banham Zoo / Africa Alive	Z46
Aonyx cinerea faecal sample	Banham Zoo / Africa Alive	Z47
Osphranter rufus faecal sample	Banham Zoo / Africa Alive	Z48
Hapalemur aureus faecal sample	Banham Zoo / Africa Alive	Z49
Eurycantha calcarata faecal sample	Banham Zoo / Africa Alive	Z50
Branta ruficollis faecal sample	Banham Zoo / Africa Alive	Z51
Ovis aries faecal sample	Banham Zoo / Africa Alive	Z52
Suricata suricatta faecal sample	Banham Zoo / Africa Alive	Z53
Cynomys sp. faecal sample	Banham Zoo / Africa Alive	Z54
Platalea sp. faecal sample	Banham Zoo / Africa Alive	Z55
Achatina fulica faecal sample	Banham Zoo / Africa Alive	Z56
Spheniscus demersus faecal sample	Banham Zoo / Africa Alive	Z57
Genetta sp. faecal sample	Banham Zoo / Africa Alive	Z58
Saguinus oedipus faecal sample	Banham Zoo / Africa Alive	Z59
Haliaeetus leucocephalus faecal sample	Banham Zoo / Africa Alive	Z60
Pudu sp. faecal sample	Banham Zoo / Africa Alive	Z61
Chrysocyon brachyuru faecal sample	Banham Zoo / Africa Alive	Z62
Panthera tigris altaica faecal sample	Banham Zoo / Africa Alive	Z63
Giraffa sp. faecal sample	Banham Zoo / Africa Alive	Z64
Gromphadorhina portentosa faecal sample	Banham Zoo / Africa Alive	Z65
Chaetophractus villosus faecal sample	Banham Zoo / Africa Alive	Z66
Varecia variegata faecal sample	Banham Zoo / Africa Alive	Z67
Eulemur rubriventer faecal sample	Banham Zoo / Africa Alive	Z68
Callithrix geoffroyi faecal sample	Banham Zoo / Africa Alive	Z69
Equus grevyi faecal sample	Banham Zoo / Africa Alive	Z70
Symphalangus syndactylus faecal sample	Banham Zoo / Africa Alive	Z71
Panthera pardus kotiya faecal sample	Banham Zoo / Africa Alive	Z72
Varecia rubra faecal sample	Banham Zoo / Africa Alive	Z73
Lemur catta faecal sample	Banham Zoo / Africa Alive	Z74
Otocolobus manul faecal sample	Banham Zoo / Africa Alive	Z75
Leopardus pardalis faecal sample	Banham Zoo / Africa Alive	Z76
Rousettus aegyptiacus faecal sample	Banham Zoo / Africa Alive	Z77
Panthera uncia faecal sample	Banham Zoo / Africa Alive	Z78
Goura victoria faecal sample	Banham Zoo / Africa Alive	Z79
Saguinus imperator faecal sample	Banham Zoo / Africa Alive	Z80
Ateles sp. faecal sample	Banham Zoo / Africa Alive	Z81
Colobus sp. faecal sample	Banham Zoo / Africa Alive	Z82
Dasyprocta sp. faecal sample	Banham Zoo / Africa Alive	Z83
Saguinus oedipus faecal sample	Banham Zoo / Africa Alive	Z84
Aonyx cinerea faecal sample	Banham Zoo / Africa Alive	Z85
Suricata suricatta faecal sample	Banham Zoo / Africa Alive	Z86
Cervus nippon faecal sample	Banham Zoo / Africa Alive	Z87
Platalea sp. faecal sample	Banham Zoo / Africa Alive	Z88
Osphranter rufus faecal sample	Banham Zoo / Africa Alive	Z89
Chrysocyon brachyurus faecal sample	Banham Zoo / Africa Alive	Z90

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Hapalemur aureus faecal sample	Banham Zoo / Africa Alive	Z91
Eulemur rubriventer faecal sample	Banham Zoo / Africa Alive	Z92
Branta ruficollis faecal sample	Banham Zoo / Africa Alive	Z93
Chelonoidis carbonaria faecal sample	Banham Zoo / Africa Alive	Z94
Leopardus geoffroyi faecal sample	Banham Zoo / Africa Alive	Z95
Varecia variegata faecal sample	Banham Zoo / Africa Alive	Z96
Helogale parvula faecal sample	Banham Zoo / Africa Alive	Z97
Ailurus fulgens faecal sample	Banham Zoo / Africa Alive	Z98
Haliaeetus leucocephalus faecal sample	Banham Zoo / Africa Alive	Z99
Giraffa sp. faecal sample	Banham Zoo / Africa Alive	Z100
Genetta sp. faecal sample	Banham Zoo / Africa Alive	Z101
Spheniscus demersus faecal sample	Banham Zoo / Africa Alive	Z102
Pudu sp. faecal sample	Banham Zoo / Africa Alive	Z103
Cynomys sp. faecal sample	Banham Zoo / Africa Alive	Z104
Panthera tigris altaica faecal sample	Banham Zoo / Africa Alive	Z105
Chaetophractus villosus faecal sample	Banham Zoo / Africa Alive	Z106
Cariama cristata faecal sample	Banham Zoo / Africa Alive	Z107
Leontopithecus rosalia faecal sample	Banham Zoo / Africa Alive	Z108
Cebuella pygmaea faecal sample	Banham Zoo / Africa Alive	Z109
Eudocimus ruber faecal sample	Banham Zoo / Africa Alive	Z110
Cygnus melancoryphus faecal sample	Banham Zoo / Africa Alive	Z111
Alouatta caraya faecal sample	Banham Zoo / Africa Alive	Z112
Callimico goeldii faecal sample	Banham Zoo / Africa Alive	Z113
Pithecia pithecia faecal sample	Banham Zoo / Africa Alive	Z114
Leontopithecus chrysomelas faecal sample	Banham Zoo / Africa Alive	Z115
Eudocimus ruber faecal sample	Banham Zoo / Africa Alive	Z116
Pithecia pithecia faecal sample	Banham Zoo / Africa Alive	Z117
Leontopithecus chrysomelas faecal sample	Banham Zoo / Africa Alive	Z118
Alouatta sp. faecal sample	Banham Zoo / Africa Alive	Z119
Leontopithecus rosalia faecal sample	Banham Zoo / Africa Alive	Z120
Ateles sp. faecal sample	Banham Zoo / Africa Alive	Z121
Cebuella pygmaea faecal sample	Banham Zoo / Africa Alive	Z122
Cygnus melancoryphus faecal sample	Banham Zoo / Africa Alive	Z123
Saguinus imperator faecal sample	Banham Zoo / Africa Alive	Z124
Cariama cristata faecal sample	Banham Zoo / Africa Alive	Z125
Callithrix geoffroyi faecal sample	Banham Zoo / Africa Alive	Z126
Pudu sp. faecal sample	Banham Zoo / Africa Alive	Z127
Ovis aries aries Cameroon faecal sample	Banham Zoo / Africa Alive	Z128
Tragelaphus eurycerus isaaci faecal sample	Banham Zoo / Africa Alive	Z129
Acinonyx jubatus jubatus faecal sample	Banham Zoo / Africa Alive	Z130
Orycteropus afer faecal sample	Banham Zoo / Africa Alive	Z131
Chlorocebus aethiops faecal sample	Banham Zoo / Africa Alive	Z132
Balearica regulorum gibbericeps faecal sample	Banham Zoo / Africa Alive	Z133
Struthio camelus faecal sample	Banham Zoo / Africa Alive	Z134
Eulemur mongoz faecal sample	Banham Zoo / Africa Alive	Z135
Panthera leo faecal sample	Banham Zoo / Africa Alive	Z136
Tragelaphus angasii faecal sample	Banham Zoo / Africa Alive	Z137
Procavia capensis faecal sample	Banham Zoo / Africa Alive	Z138
Cyanochen cyanopterus faecal sample	Banham Zoo / Africa Alive	Z139

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ovis aries aries somali faecal sample	Banham Zoo / Africa Alive	Z140
Hystrix cristata faecal sample	Banham Zoo / Africa Alive	Z141
Equus grevyi faecal sample	Banham Zoo / Africa Alive	Z142
Ammotragus lervia faecal sample	Banham Zoo / Africa Alive	Z143
Coracopsis vasa faecal sample	Banham Zoo / Africa Alive	Z144
Addax nasomaculatus faecal sample	Banham Zoo / Africa Alive	Z145
Stigmochelys pardalis faecal sample	Banham Zoo / Africa Alive	Z146
Hypogeomys antimena faecal sample	Banham Zoo / Africa Alive	Z147
Cynictis penicillata faecal sample	Banham Zoo / Africa Alive	Z148
Tragelaphus spekii gratus faecal sample	Banham Zoo / Africa Alive	Z149
Ceratotherium simum simum faecal sample	Banham Zoo / Africa Alive	Z150
Epibolus pulchripes faecal sample	Banham Zoo / Africa Alive	Z151
Malacochersus tornieri faecal sample	Banham Zoo / Africa Alive	Z152
Gromphadorhina portentosa faecal sample	Banham Zoo / Africa Alive	Z153
Acomys cilicicus faecal sample	Banham Zoo / Africa Alive	Z154
Thallomys paedulcus faecal sample	Banham Zoo / Africa Alive	Z155
Chlorocebus aethiops faecal sample	Banham Zoo / Africa Alive	Z156
Orycteropus afer faecal sample	Banham Zoo / Africa Alive	Z157
Stigmochelys pardalis faecal sample	Banham Zoo / Africa Alive	Z158
Hystrix cristata faecal sample	Banham Zoo / Africa Alive	Z159
Eulemur coronatus faecal sample	Banham Zoo / Africa Alive	Z160
Bubalus bubalis faecal sample	Banham Zoo / Africa Alive	Z161
Addax nasomaculatus faecal sample	Banham Zoo / Africa Alive	Z162
Ovis aries aries cameroon faecal sample	Banham Zoo / Africa Alive	Z163
Acinonyx jubatus jubatus faecal sample	Banham Zoo / Africa Alive	Z164
Balearica regulorum gibbericeps faecal sample	Banham Zoo / Africa Alive	Z165
Tragelaphus angasii faecal sample	Banham Zoo / Africa Alive	Z166
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z167
Panthera leo faecal sample	Banham Zoo / Africa Alive	Z168
Lemur catta faecal sample	Banham Zoo / Africa Alive	Z169
Syncerus caffer nanus faecal sample	Banham Zoo / Africa Alive	Z170
Balearica pavonina faecal sample	Banham Zoo / Africa Alive	Z171
Cynictis penicillata faecal sample	Banham Zoo / Africa Alive	Z172
Colobus polykomos faecal sample	Banham Zoo / Africa Alive	Z173
Eulemur macaco faecal sample	Banham Zoo / Africa Alive	Z174
Epibolus pulchripes faecal sample	Banham Zoo / Africa Alive	Z175
Panthera tigris altaica faecal sample	Banham Zoo / Africa Alive	Z176
Helogale parvula faecal sample	Banham Zoo / Africa Alive	Z177
Otocolobus manul faecal sample	Banham Zoo / Africa Alive	Z178
Dasyprocta azarae faecal sample	Banham Zoo / Africa Alive	Z179
Suricata suricatta faecal sample	Banham Zoo / Africa Alive	Z180
Ailurus fulgens faecal sample	Banham Zoo / Africa Alive	Z181
Varecia variegata faecal sample	Banham Zoo / Africa Alive	Z182
Aonyx cinerea faecal sample	Banham Zoo / Africa Alive	Z183
Pudu sp. faecal sample	Banham Zoo / Africa Alive	Z184
Gromphadorhina portentosa faecal sample	Banham Zoo / Africa Alive	Z185
Chelonoidis carbonaria faecal sample	Banham Zoo / Africa Alive	Z186
Eulemur rubriventer faecal sample	Banham Zoo / Africa Alive	Z187
Goura victoria faecal sample	Banham Zoo / Africa Alive	Z188

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Colobus sp. faecal sample	Banham Zoo / Africa Alive	Z189
Atelerix albiventris faecal sample	Banham Zoo / Africa Alive	Z190
Oryctolagus cuniculus faecal sample	Banham Zoo / Africa Alive	Z191
Thallomys paedulcus faecal sample	Banham Zoo / Africa Alive	Z192
Callithrix_jacchus faecal sample	Banham Zoo / Africa Alive	Z193
Ursus arctos arctos faecal sample	Banham Zoo / Africa Alive	Z194
Ursus arctos arctos faecal sample	Banham Zoo / Africa Alive	Z195
Phodopus roborovskii faecal sample	Banham Zoo / Africa Alive	Z196
Myrmecophaga tridactyla faecal sample	Banham Zoo / Africa Alive	Z197
Syncerus caffer nanus faecal sample	Banham Zoo / Africa Alive	Z198
Eulemur macaco faecal sample	Banham Zoo / Africa Alive	Z199
Sus scrofa domesticus faecal sample	Banham Zoo / Africa Alive	Z200
Panthera leo faecal sample	Banham Zoo / Africa Alive	Z201
Ateles sp. faecal sample	Banham Zoo / Africa Alive	Z202
Acomys cilicicus faecal sample	Banham Zoo / Africa Alive	Z203
Chlorocebus aethiops faecal sample	Banham Zoo / Africa Alive	Z204
Galago sp. faecal sample	Banham Zoo / Africa Alive	Z205
Symphalangus syndactylus faecal sample	Banham Zoo / Africa Alive	Z206
Saguinus imperator faecal sample	Banham Zoo / Africa Alive	Z207
Panthera uncia faecal sample	Banham Zoo / Africa Alive	Z208
Hystrix cristata faecal sample	Banham Zoo / Africa Alive	Z209
Colobus sp. faecal sample	Banham Zoo / Africa Alive	Z210
Dasyprocta sp. faecal sample	Banham Zoo / Africa Alive	Z211
Tragelaphus spekii gratus faecal sample	Banham Zoo / Africa Alive	Z212
Atherurus africanus faecal sample	Banham Zoo / Africa Alive	Z213
Myrmecophaga tridactyla faecal sample	Banham Zoo / Africa Alive	Z214
Callithrix_jacchus faecal sample	Banham Zoo / Africa Alive	Z215
Ursus arctos arctos faecal sample	Banham Zoo / Africa Alive	Z216
Ursus arctos arctos faecal sample	Banham Zoo / Africa Alive	Z217
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z218
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z219
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z220
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z221
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z222
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z223
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z224
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z225
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z226
Lepilemur sahamalazensis faecal sample	Banham Zoo / Africa Alive	Z227
Scatophagus sp. faecal sample	Banham Zoo / Africa Alive	Z228
Scatophagus sp. faecal sample	Banham Zoo / Africa Alive	Z228W
Spheniscus humboldti faecal sample	Banham Zoo / Africa Alive	Z229
Spheniscus humboldti faecal sample	Banham Zoo / Africa Alive	Z229W
Hippocampus sp. faecal sample	Banham Zoo / Africa Alive	Z230
Hippocampus sp. faecal sample	Banham Zoo / Africa Alive	Z230W
Raja clavata faecal sample	Banham Zoo / Africa Alive	Z231
Raja clavata faecal sample	Banham Zoo / Africa Alive	Z231W
Chelonia mydas faecal sample	Banham Zoo / Africa Alive	Z232
Chelonia mydas faecal sample	Banham Zoo / Africa Alive	Z232W

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Astyanax mexicanus faecal sample	Banham Zoo / Africa Alive	Z233
Astyanax mexicanus faecal sample	Banham Zoo / Africa Alive	Z233W
Ursus maritimus faecal sample	Banham Zoo / Africa Alive	Z234
Ursus maritimus faecal sample	Banham Zoo / Africa Alive	Z235
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z236A
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z236B
Lemur catta faecal sample	Banham Zoo / Africa Alive	Z237A
Lemur catta faecal sample	Banham Zoo / Africa Alive	Z237B
Varecia rubra faecal sample	Banham Zoo / Africa Alive	Z238
Eulemur coronatus faecal sample	Banham Zoo / Africa Alive	Z239A
Eulemur coronatus faecal sample	Banham Zoo / Africa Alive	Z239B
Eulemur mongoz faecal sample	Banham Zoo / Africa Alive	Z240A
Eulemur mongoz faecal sample	Banham Zoo / Africa Alive	Z240B
Dasyprocta azarae faecal sample	Banham Zoo / Africa Alive	Z241A
Dasyprocta azarae faecal sample	Banham Zoo / Africa Alive	Z241B
Rousettus aegyptiacus faecal sample	Banham Zoo / Africa Alive	Z242
Leopardus pardalis faecal sample	Banham Zoo / Africa Alive	Z243
Alouatta caraya faecal sample	Banham Zoo / Africa Alive	Z244
Callimico goeldii faecal sample	Banham Zoo / Africa Alive	Z245
Sus scrofa domesticus faecal sample	Banham Zoo / Africa Alive	Z246
Panthera pardus kotiya faecal sample	Banham Zoo / Africa Alive	Z247
Oryctolagus cuniculus faecal sample	Banham Zoo / Africa Alive	Z248
Cyclura cornuta faecal sample	Banham Zoo / Africa Alive	Z249
Atelerix albiventris faecal sample	Banham Zoo / Africa Alive	Z250
Epibolus pulchripes faecal sample	Banham Zoo / Africa Alive	Z251
Eublepharis macularius faecal sample	Banham Zoo / Africa Alive	Z252
Lampropeltis triangulum hondurensis faecal sample	Banham Zoo / Africa Alive	Z253
Achatina fulica faecal sample	Banham Zoo / Africa Alive	Z254
Phryganistria heusii yentuensis faecal sample	Banham Zoo / Africa Alive	Z255
Atherurus africanus faecal sample	Banham Zoo / Africa Alive	Z256
Bubalus bubalis faecal sample	Banham Zoo / Africa Alive	Z257
Ceratotherium simum simum faecal sample	Banham Zoo / Africa Alive	Z258
Colobus polykomos faecal sample	Banham Zoo / Africa Alive	Z259
Galago senegalensis faecal sample	Banham Zoo / Africa Alive	Z260
Hystrix cristata faecal sample	Banham Zoo / Africa Alive	Z261
Oryctolagus cuniculus faecal sample	Banham Zoo / Africa Alive	Z262
Ovis aries aries somali faecal sample	Banham Zoo / Africa Alive	Z263
Panthera leo faecal sample	Banham Zoo / Africa Alive	Z264
Cercocebus chrysogaster faecal sample	Banham Zoo / Africa Alive	Z265
Procavia capensis faecal sample	Banham Zoo / Africa Alive	Z266
Stigmochelys pardalis faecal sample	Banham Zoo / Africa Alive	Z267
Struthio camelus faecal sample	Banham Zoo / Africa Alive	Z268
Tragelaphus eurycerus isaaci faecal sample	Banham Zoo / Africa Alive	Z269
Gerrhosaurus major faecal sample	Banham Zoo / Africa Alive	Z270
Gromphadorhina portentosa faecal sample	Banham Zoo / Africa Alive	Z271
Malacochersus tornieri faecal sample	Banham Zoo / Africa Alive	Z272
Python regius faecal sample	Banham Zoo / Africa Alive	Z273
Uromastyx acanthurina faecal sample	Banham Zoo / Africa Alive	Z274
Leopardus pardalis faecal sample	Banham Zoo / Africa Alive	Z275

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rousettus aegyptiacus faecal sample	Banham Zoo / Africa Alive	Z276
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z277A
Eulemur flavifrons faecal sample	Banham Zoo / Africa Alive	Z277B
Dendrocygna viduata faecal sample	Banham Zoo / Africa Alive	Z278
Coracopsis vasa faecal sample	Banham Zoo / Africa Alive	Z279
Hemitheconyx caudicinctus faecal sample	Banham Zoo / Africa Alive	Z280
Neophron percnopterus faecal sample	Banham Zoo / Africa Alive	Z281
Psittacus erithacus erithacus faecal sample	Banham Zoo / Africa Alive	Z282
Capra hircus faecal sample	Banham Zoo / Africa Alive	Z283
Phryganistria heusii yentuensis faecal sample	Banham Zoo / Africa Alive	Z284
Lampropeltis triangulum sinaloae faecal sample	Banham Zoo / Africa Alive	Z285
Dendrobates auratus faecal sample	Banham Zoo / Africa Alive	Z286
Carausius morosus faecal sample	Banham Zoo / Africa Alive	Z287
Atelerix sp. faecal sample	Banham Zoo / Africa Alive	Z288
Polyplectron bicalcaratum faecal sample	Banham Zoo / Africa Alive	Z289
Dromaius novaehollandiae faecal sample	Banham Zoo / Africa Alive	Z290
Equus asinus faecal sample	Banham Zoo / Africa Alive	Z291
Pauxi pauxi faecal sample	Banham Zoo / Africa Alive	Z292
Tragopan temminckii faecal sample	Banham Zoo / Africa Alive	Z293
Bycanistes bucinator faecal sample	Banham Zoo / Africa Alive	Z294
Capra hircus faecal sample	Banham Zoo / Africa Alive	Z295
Cacatua ducorpsi and Cacatua galerita faecal sample	Banham Zoo / Africa Alive	Z296
Tauraco erythrolophus faecal sample	Banham Zoo / Africa Alive	Z297
Cacatua ducorpsii faecal sample	Banham Zoo / Africa Alive	Z298
Tockus deckeni faecal sample	Banham Zoo / Africa Alive	Z299
Schistocerca gregaria faecal sample	Banham Zoo / Africa Alive	Z300
Leucopsar rothschildi faecal sample	Banham Zoo / Africa Alive	Z301
Strix uralensis faecal sample	Banham Zoo / Africa Alive	Z302
Strix woodfordii faecal sample	Banham Zoo / Africa Alive	Z303
Strix rufipes faecal sample	Banham Zoo / Africa Alive	Z304
Bubo lacteus faecal sample	Banham Zoo / Africa Alive	Z305
Strix nebulosa faecal sample	Banham Zoo / Africa Alive	Z306
Ploceus melanocephalus faecal sample	Banham Zoo / Africa Alive	Z307
Strix leptogrammica faecal sample	Banham Zoo / Africa Alive	Z308
Falco cherrug faecal sample	Banham Zoo / Africa Alive	Z309
Pulsatrix perspicillata faecal sample	Banham Zoo / Africa Alive	Z310
Ptilopsis leucotis faecal sample	Banham Zoo / Africa Alive	Z311
Equus africanus somaliensis faecal sample	Banham Zoo / Africa Alive	Z312
Bubo scandiacus faecal sample	Banham Zoo / Africa Alive	Z313
Bradypus_torquatus faecal sample	Banham Zoo / Africa Alive	Z314
Ninox boobook faecal sample	Banham Zoo / Africa Alive	Z315
Cryptoprocta ferox faecal sample	Banham Zoo / Africa Alive	Z316
Eurycantha calcarata faecal sample	Banham Zoo / Africa Alive	Z317
Lamprotornis superbus faecal sample	Banham Zoo / Africa Alive	Z318
Achatina achatina faecal sample	Banham Zoo / Africa Alive	Z319
Acheta domesticus faecal sample	Banham Zoo / Africa Alive	Z320
Agapornis nigrigenis faecal sample	Banham Zoo / Africa Alive	Z321
Boa constrictor faecal sample	Banham Zoo / Africa Alive	Z322
Pantherophis guttatus faecal sample	Banham Zoo / Africa Alive	Z323

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Parabuteo unicinctus faecal sample	Banham Zoo / Africa Alive	Z324
Eolophus roseicapilla faecal sample	Banham Zoo / Africa Alive	Z325
Python regius faecal sample	Banham Zoo / Africa Alive	Z326
Carausius morosus faecal sample	Banham Zoo / Africa Alive	Z327
Ara ararauna faecal sample	Banham Zoo / Africa Alive	Z328
Cyclura cornuta faecal sample	Banham Zoo / Africa Alive	Z329
Falco peregrinus faecal sample	Banham Zoo / Africa Alive	Z330
Polyboroides typus faecal sample	Banham Zoo / Africa Alive	Z331
Dacelo sp. faecal sample	Banham Zoo / Africa Alive	Z332
Aratinga solstitialis faecal sample	Banham Zoo / Africa Alive	Z333
Tauraco leucotis faecal sample	Banham Zoo / Africa Alive	Z334
Bubo bubo faecal sample	Banham Zoo / Africa Alive	Z335
Tyto alba faecal sample	Banham Zoo / Africa Alive	Z336
Equus ferus caballus faecal sample	Banham Zoo / Africa Alive	Z337
Falco jugger faecal sample	Banham Zoo / Africa Alive	Z338
Aquila nipalensis faecal sample	Banham Zoo / Africa Alive	Z339
Falco tinnunculus faecal sample	Banham Zoo / Africa Alive	Z340
Falco sparverius faecal sample	Banham Zoo / Africa Alive	Z341
Falco peregrinus faecal sample	Banham Zoo / Africa Alive	Z342
Falco biarmicus faecal sample	Banham Zoo / Africa Alive	Z343
Buteo buteo faecal sample	Banham Zoo / Africa Alive	Z344
Falco rupicoloides faecal sample	Banham Zoo / Africa Alive	Z345
Buteo jamaicensis faecal sample	Banham Zoo / Africa Alive	Z346
Strix aluco faecal sample	Banham Zoo / Africa Alive	Z347
Tyto alba faecal sample	Pafos Zoo	Z348
Dromaius novaehollandiae faecal sample	Pafos Zoo	Z349
Bubo bubo faecal sample	Pafos Zoo	Z350
Goura victoria faecal sample	Pafos Zoo	Z351
Eolophus roseicapilla faecal sample	Pafos Zoo	Z352
Cacatua ducorpsii faecal sample	Pafos Zoo	Z353
Bycanistes bucinator faecal sample	Pafos Zoo	Z354
Cygnus melancoryphus faecal sample	Pafos Zoo	Z355
Eudocimus ruber faecal sample	Pafos Zoo	Z356
Strix nebulosa faecal sample	Pafos Zoo	Z357
Dacelo sp. faecal sample	Pafos Zoo	Z358
Haliaeetus leucocephalus faecal sample	Pafos Zoo	Z359
Struthio camelus faecal sample	Pafos Zoo	Z360
Polyplectron bicalcaratum faecal sample	Pafos Zoo	Z361
Tauraco erythrolophus faecal sample	Pafos Zoo	Z362
Ara ararauna faecal sample	Pafos Zoo	Z363
Spheniscus humboldti faecal sample	Pafos Zoo	Z364
Aratinga solstitialis faecal sample	Pafos Zoo	Z365
Chelonoidis carbonaria faecal sample	Pafos Zoo	Z366
Lampropeltis triangulum hondurensis faecal sample	Pafos Zoo	Z367
Eublepharis macularius faecal sample	Pafos Zoo	Z368
Tauraco leucotis faecal sample	Pafos Zoo	Z369
Python regius faecal sample	Pafos Zoo	Z370
Gromphadorhina portentosa faecal sample	Pafos Zoo	Z371
Boa constrictor faecal sample	Pafos Zoo	Z372

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pogona vitticeps faecal sample	Pafos Zoo	Z373
Chelonoidis carbonaria faecal sample	Pafos Zoo	Z374
Balearica regulorum gibbericeps faecal sample	Pafos Zoo	Z375
Lampropeltis triangulum sinaloae faecal sample	Pafos Zoo	Z376
Leucopsar rothschildi faecal sample	Pafos Zoo	Z377
Gromphadorhina portentosa faecal sample	Pafos Zoo	Z378
Pantherophis guttatus faecal sample	Pafos Zoo	Z379
Psittacus erithacus erithacus faecal sample	Pafos Zoo	Z380
Acheta domesticus faecal sample	Pafos Zoo	Z381
Deposited data		
16S rRNA amplicon sequencing data	This paper	NCBI SRA: PRJNA1200941
Draft genomes of 96 *Bifidobacterium* isolates	This paper	NCBI Genomes: PRJNA1200594
R code and associated files	This paper	Figshare: https://doi.org/10.6084/m9.figshare.29635103
Software and algorithms		
Cutadapt	Martin M[67]	https://doi.org/10.14806/ej.17.1.200
DADA2	Callahan et al.[68]	https://doi.org/10.1038/nmeth.3869
ampvis2	Andersen et al.[69]	https://doi.org/10.1101/299537
R	R Core Team[70]	http://www.r-project.org/index.html
pheatmap	Kolde R[71]	https://cran.r-project.org/web/packages/pheatmap/index.html
phylosmith	Smith SD[72]	https://doi.org/10.21105/joss.01442
ggplot2	Wickham et al.[73]	https://cran.r-project.org/web/packages/ggplot2/index.html
TimeTree	Kumar et al.[74]	https://doi.org/10.1093/molbev/msx116
ABDOMEN	Perez-Lamarque et al.[27]	https://doi.org/10.1093/molbev/msad144
rstan	Guo et al.[75]	https://cran.r-project.org/web/packages/rstan/index.html
CheckM	Parks et al.[76]	https://doi.org/10.1101/gr.186072.114
dRep	Olm et al.[77]	https://doi.org/10.1038/ismej.2017.126
GTDB-Tk	Chaumeil et al.[78]	https://doi.org/10.1093/bioinformatics/btac672
pyANI	Pritchard et al.[79]	https://doi.org/10.1039/c5ay02550h
IQ-TREE 2	Minh et al.[80]	https://doi.org/10.1093/molbev/msaa015
iTOL	Letunic et al.[81]	https://doi.org/10.1093/nar/gkae268
PhyloMInt	Lam et al.[82]	https://doi.org/10.1371/journal.pcbi.1007951
phytools	Revell LJ[83]	https://doi.org/10.1111/j.2041-210X.2011.00169.x
prokka	Seeman T[84]	https://doi.org/10.1093/bioinformatics/btu153
eggNOG-Mapper	Huerta-Cepas et al.[85]	https://doi.org/10.1093/molbev/msx148
dbCAN3	Zheng et al.[86]	https://doi.org/10.1093/nar/gkad328
dplyr	Wickham et al.[87]	https://cran.r-project.org/web/packages/dplyr/index.html
cowplot	Wilke CO[88]	https://cran.r-project.org/web/packages/cowplot/index.html
UpSetR	Conway et al.[89]	https://cran.r-project.org/web/packages/UpSetR/index.html
factoextra	Kassambara et al.[90]	http://www.sthda.com/english/rpkgs/factoextra.
AlphaFold3	Abramson et al.[91]	https://doi.org/10.1038/s41586-024-07487-w
Coot	Emsley et al.[92]	https://doi.org/10.1107/S0907444910007493
PyMOL	DeLano W[93]	https://www.pymol.org/
PeSTo-Carbs	Bibekar et al.[94]	https://doi.org/10.1021/acs.jctc.3c01145
MAFFT	Katoh et al.[95]	https://doi.org/10.1093/molbev/mst010

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
trimAL	Capella-Gutierrez et al.[96]	https://doi.org/10.1093/bioinformatics/btp348
vegan	Okansen et al.[97]	https://CRAN.R-project.org/package=vegan
rstatix	Kassambara A[98]	https://cran.r-project.org/web/packages/rstatix/index.html
ape	Paradis et al.[99]	https://doi.org/10.1093/bioinformatics/bty633
Other		
SILVA database	Quast et al.[100]	https://doi.org/10.1093/nar/gks1219
eggNOG database	Huerta-Cepas et al.[101]	https://doi.org/10.1093/nar/gky1085
Protein Data Bank	Berman et al.[102]	https://doi.org/10.1093/nar/28.1.235

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

A total of 393 animal faecal samples from 175 diverse animal hosts were collected by animal-care staff at Banham Zoo (UK) and Africa Alive! (UK) (total n = 359; n = 225 mammalian samples, 77 = avian samples, 25 = reptilian samples, 14 = insect samples, n = 18 amphibian and diplopode samples) and Pafos Zoo (Cyprus) (total n = 34; n = 22 avian samples; n = 9 reptilian samples; n = 3 insect samples). Samples were collected opportunistically from animals under routine care; animals were not part of experimental studies, and sampling was pragmatic, non-invasive, and involved no direct handling. All animals were maintained under the zoos' standard species-specific husbandry protocols, which meet or exceed national animal welfare guidelines. Detailed individual data (species strain, age, developmental stage, sex, or specific maintenance parameters) were not systematically recorded and are unavailable.

Faecal sample collection

All faecal samples were collected into sterile Sterilin (Thermo Scientific) specimen containers with a spoon. Samples from Banham Zoo and Africa Alive! were stored at 4°C under anaerobic conditions using Oxoid AnaeroGen 2.5L sachets and transported to our laboratory within 48h. Samples from Pafos Zoo were kept in -20°C until shipped to the UK. At our laboratory, all samples were stored at -80°C. Samples from Pafos Zoo were collected from avian, reptilian and insect hosts, for which we had corresponding samples from Banham Zoo (UK) and Africa Alive! (UK). All 393 faecal samples collected in this study were subjected to genomic DNA extraction and bacterial isolation.

METHOD DETAILS

DNA extraction from faecal samples and 16S rRNA amplicon sequencing

FastDNA Spin Kit for Soil (MP Biomedicals) was used to extract DNA from ~200mg of animal faecal material following manufacturer instructions, with the bead-beating time extended to 3 min. DNA concentration and quality were quantified using Qubit dsDNA BR Assay Kit (Invitrogen) in Qubit 2.0 Fluorometer according to the manufacturer's protocol (Invitrogen).

Due to project duration and technical capacity, samples from Pafos Zoo (n = 34) were sequenced using the V1-V2 16S rRNA gene region at Novogene (Cambridge, UK), and samples from Banham Zoo and Africa Alive! (n = 359) were sequenced using the V4 16S rRNA gene region at Earlham Institute (Norwich, UK). PCR amplicons for the V1-V2 region of the 16S rRNA gene were generated with primers: Modified 27F 5'- AGMGTTYGATYMTGGCTCAG-3'[103] and 338R 5'- GCTGCCTCCCGTAGGAGT-3'[104] PCR amplicons for the V4 region were generated with primers SA501-SA508(F) and SA701-SA712(R).[105] Out of the 393 animal faecal samples collected from 175 diverse hosts in this study, we successfully performed 16S rRNA amplicon sequencing for a total of 219 samples from 126 diverse hosts: n = 21 samples from Pafos Zoo (n = 17 avian samples (17 hosts) and n = 4 reptilian samples (4 hosts)) and n = 198 samples from Banham Zoo and Africa Alive! (n = 135 mammalian samples (64 hosts), n = 42 avian samples (40 hosts), n = 14 reptilian samples (10 hosts), n = 7 insect samples (5 hosts)) (Table S1).

16S rRNA amplicon sequencing data analysis

The datasets sequenced using different 16S rRNA hypervariable regions were pre-processed separately until the taxonomy assignment step. Cutadapt v.1.18[67] was used to remove primers from each dataset containing primers. Reads below the quality score of 20 (minQ=20) and maximum expected error of 1 (maxEE=1) were filtered out from each dataset using DADA2,[68] followed by the inference of amplicon sequence variants (ASV). Chimeras were removed from each dataset separately, after which the datasets were merged.

Taxonomy was assigned using SILVA database v.138.[100] 'Ampvis2'[69] package implemented in R v.4.2.3[70] was used to rarefy the ASV table at 10,000 reads per sample and to analyse and visualise the 16S rRNA amplicon data from hosts belonging to classes Insecta, Reptilia, Aves and Mammalia. We performed PERMANOVA to test for differences in microbiota composition between host groups and PERMDISP to test for homogeneity of dispersion across taxonomic classes. Hierarchical clustering was performed

with 'pheatmap'[71]. Co-occurrence analysis of microbial species was performed using the function 'co_occurrence()' in R package 'phylosmith'[72]. All p-values were adjusted using Bonferroni's method. Data was visualized using the R package 'ggplot2'[73].

To perform phylosymbiosis analysis using the 16S rRNA amplicon data, we first used the TimeTree v.5.0 resource[74] to acquire separate dated phylogenetic trees encompassing as many mammalian and avian host species from out dataset as possible. Out of 126 diverse hosts for which we had the corresponding 16S rRNA amplicon data, dated phylogeny was available for 62 mammalian and 38 avian hosts. We obtained the microbiota composition of each of these hosts by averaging the corresponding rarefied sample 16S rRNA amplicon data per host species and extracted the relative abundances of the main bacterial orders. We only took the 15 most abundant orders under consideration. We then used the 'ABDOMEN'[27] R suite to apply the multivariate Brownian motion model for variations in microbiota composition over host evolutionary time to both mammalian and avian datasets separately, with 3000 STAN[75] iterations permutated 10 times.

Isolation of bifidobacteria from faecal samples

Depending on the amount of available faecal material, samples (~50mg or ~100mg) were re-suspended in either 450µl or 900µl of sterile Phosphate Buffer Saline, vortexed for 30s, mixed on a shaker at 1600rpm, and used to produce serial dilutions (neat to 10^{-4}). The dilutions were plated onto either de Man-Rogosa-Sharpe (MRS) agar (BD Biosciences) supplemented with mupirocin (50mg/l) (AppliChem) and L-cysteine hydrochloride monohydrate (50mg/l) (Sigma-Aldrich) or Brain Heart Infusion (BHI) agar (Oxoid) supplemented with mupirocin (50mg/l), L-cysteine hydrochloride monohydrate (50mg/l) and sodium iodoacetate (7.5mg/l) (Sigma-Aldrich). Plates were incubated in an anaerobic cabinet (Don Whitley Scientific) for 48-72 hours. Three colonies from each dilution were randomly selected and streaked to purity on either BHI agar or MRS agar supplemented with L-cysteine hydrochloride monohydrate (50mg/l). Pure cultures were stored in cryogenic tubes at -80°C.

Genomic DNA extraction from bacterial isolates

Bacterial pellets were re-suspended in 2ml of 25% sucrose in 10mM Tris (Sigma-Aldrich) and 1mM EDTA (Sigma-Aldrich) at pH8.0. Cells were then treated using 50µl of 100mg/ml lysozyme (Roche). Further, 100µl of 20mg/ml Proteinase K (Roche), 30µl of 10mg/ml RNase A (Roche), 400µl of 0.5 M EDTA and 250µl of 10% Sarkosyl NL30 (Sigma-Aldrich) were added into the lysed bacterial suspension. The samples were then incubated on ice for 2 hours, followed by 50°C overnight water bath.

Next, samples were subject to three rounds of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) (Sigma-Aldrich) extraction using Qiagen MaXtract High Density tubes. Further two rounds of extractions with Chloroform:Isoamyl Alcohol (24:1) (Sigma-Aldrich) were then performed to remove residual phenol, followed by ethanol precipitation and 70% ethanol wash, after which DNA pellets were resuspended in 300µl of 10mM Tris (pH8.0). Sample DNA concentration was quantified using Qubit dsDNA BR Assay Kit in Qubit 2.0 Fluorometer. Extracted DNA was stored in -20°C until further analysis.

Whole genome sequencing of bacterial isolates

This work was performed at the Wellcome Trust Sanger Institute (Hinxton, UK) and at the Quadram Institute Bioscience (Norwich, UK). At Hinxton, DNA was sequenced using 96-plex Illumina HiSeq 2500 platform as described previously.[3] At Quadram Institute Bioscience, genomic DNA was normalised to 0.5ng/µl with EB (10mM Tris-HCl). 0.9µl of TD Tagment DNA Buffer (Illumina Catalogue No. 15027866) was mixed with 0.09µl TDE1, Tagment DNA Enzyme (Illumina Catalogue No. 15027865) and 2.01µl PCR grade water in a master mix and 3µl added to a chilled 96 well plate. 2µl of normalised DNA (1ng total) was pipette mixed with the 3µl of the tagmentation mix and heated to 55 ºC for 10 minutes in a PCR block. A PCR master mix was made up using 4µl kapa2G buffer, 0.4µl dNTPs, 0.08µl Polymerase and 6.52µl PCR grade water, contained in the Kap2G Robust PCR kit (Sigma Catalogue No. KK5005) per sample and 11µl added to each well need to be used in a 96-well plate. 2µl of each P7 and P5 of Nextera XT Index Kit v2 index primers (Illumina Catalogue No. FC-131-2001 to 2004) were added to each well. Finally, the 5µl of tagmentation mix was added and mixed. The PCR was run with 72ºC for 3 minutes, 95ºC for 1 minute, 14 cycles of 95ºC for 10s, 55ºC for 20s and 72ºC for 3 minutes. Following the PCR reaction, the libraries were quantified using the Quant-iT dsDNA Assay Kit, high sensitivity kit (Catalogue No. 10164582) and run on a FLUOstar Optima plate reader. Libraries were pooled following quantification in equal quantities. The final pool was double-SPRI size selected between 0.5 and 0.7X bead volumes using KAPA Pure Beads (Roche Catalogue No. 07983298001). The final pool was quantified on a Qubit 3.0 instrument and run on a High Sensitivity D1000 ScreenTape (Agilent Catalogue No. 5067-5579) using the Agilent Tapestation 4200 to calculate the final library pool molarity. The pool was run at a final concentration of 10pM on an Illumina MiSeq instrument.

Genomic data processing and compilation of the final dataset

For data generated at the Wellcome Trust Sanger Institute, genome assemblies were performed by the sequencing provider using the assembly pipeline described by Page et al.[106] Sequencing reads generated at the Quadram Institute Bioscience were pre-processed with fastp v.0.23.2[107] and assembled using Unicycler v.0.4.9[108] with the "–mode conservative" option, after which contigs below 1000 bp were filtered out. We estimated completeness and contamination of all genomes sequenced as part of this study using CheckM v.1.2.0,[76] and retained sequences with completeness >99% and contamination <1%. Additionally, genomes of 105 *Bifidobacterium* type strains - strains made up of living cultures of an organism, which are descended from strains designated as the

nomenclatural types[109] - and those of animal-associated *Bifidobacterium* isolates, were identified and downloaded from the NCBI. dRep v.2.5.0 was used to dereplicate the dataset at 99.9% identity threshold[77] to retain bifidobacterial representation across distinct host taxonomic orders.

After such pre-processing, the final dataset comprised of 387 *Bifidobacterium* sequences, with 96 genomes resulting from isolation work and whole genome sequencing performed in this study and 291 genomes downloaded from the public datasets (Table S2).

GTDB-Tk v.2.1.0[78] was used to classify all genomic sequences to the strain level and produce a multiple sequence alignment of the single copy gene markers. Python3 module pyANI v.0.2.10 with default settings was used to calculate the average nucleotide identity values (ANI).[79] Species delineation cut-off was set at 95% identity.[110]

Phylogenetic analysis

IQ-Tree v.2.0.5[80] was used to test for the best substitution model fitting the GTDB-Tk-created alignment and produce a global *Bifidobacterium* phylogenetic tree, which was visualised with ITOL v.6.9.[81] ANOSIM statistics with 9,999 permutations of phylogenetic distance matrix was used to assess the clustering of *Bifidobacterium* strains according to host order.

Analysis of *Bifidobacterium* strain interactions in hosts

To explore potential cooperative nature of closely related strains in selected hosts, we used the pipeline PhyloMInt to compute metabolic interaction indices (competition and complementarity) for pairs of *Bifidobacterium* strains from *Sus scrofa* (pig), *Homo sapiens* (human), and *Saguinus oedipus* (cotton-top tamarin).[82] The hosts were chosen based on the number of available *Bifidobacterium* strains. Obtained pairwise complementarity index values between *Bifidobacterium* pairs were plotted against their phylogenetic distance. To identify significant complementarity outliers defined pairs of strains with high complementarity indices relative to their phylogenetic distance and to limit outlier detection to localised values, we grouped pairwise observations into bins based on a fixed phylogenetic distance interval of 0.01. We then calculated the Z-score within each bin, respectively. We used the Z-score threshold of ±2.698 to identify significant outliers within bins.[82]

Co-phylogeny analysis

TimeTree v.5.0 resource was used to retrieve dated trees for mammalian and avian hosts harbouring bifidobacteria from our dataset. GTDB-Tk and IQ-Tree were then used to produce alignments and phylogenies for bifidobacteria associated with these mammalian and avian hosts, respectively. 'Parafit' function with the p-value significance level set at 0.01 and the permutation number set to 999 was run 3 times to test for the host-*Bifidobacterium* co-phylogenetic signal. P-values obtained from each run were adjusted using the Benjamini-Hochberg correction. Tanglegrams were created with 'phytools'[83] package implemented in R.

Genomic and functional annotation

For consistency, all genomes were annotated using Prokka v.1.14.6.[84] Functional traits were profiled using eggNOG-mapper v.2.1.11[85] with eggNOG database v.5.0.[101] Standalone version of dbCAN3 with HMMdb v.12 and "–hmm_cov 0.50" option was used to annotate carbohydrate-active enzymes (CAZYmes).[86] R packages 'dplyr'[87] and 'ggplot2' were used to summarise and visualise basic genomic and functional abundance features.

Assessment of between-group differences in abundance of functional features

Between-group analysis of differences in abundance of functional features (KEGG Orthology (KO) and CAZyme) in bifidobacteria associated with particular host groups was performed in R for a subset of the genomic data. We only included host order categories for which we had more than 20 *Bifidobacterium* genomes available (Hymenoptera (bees and bumblebees) (n = 35), Artiodactyla (even-toed ungulates) (n = 58), Carnivora (n = 22), Rodentia (n = 47), Primates (n = 172)). KO and CAZyme abundance matrices generated for each of the *Bifidobacterium* strains used in this analysis were filtered at mean prevalence of >20% and subsequently log-transformed (log+1) to achieve a less skewed distribution. Differences in abundance of functional features were assessed following the approach described in Ruehlemann et al.[111] Briefly, linear regression analysis was employed using abundances as dependent variable and Hymenoptera/Mammals and Primate/non-primate dichotomies as explanatory variables in a single model for each function defined as lm(abundance ~ Mammals + Primates). P-values were calculated from the t-values of the resulting models using the summary.lm() function. Log-fold differences were calculated using group mean abundances and a pseudo count of 0.01. P-values were adjusted using Bonferroni correction. Features with significant ($Q < 0.05$) positive association were grouped into "Hymenoptera", "Mammals", "Primates" and "non-primates" categories. Features without abundance differences were grouped as "other".

Additionally, CAZyme abundance data for host order groups was summarised and visualised with R packages 'dplyr', 'ggplot2', 'cowplot'[88] and 'UpSetR'[89]. Principal component analysis of the CAZyme abundance data was performed using 'factoextra' v.1.0.7.[90]

Prediction of protein structure and ligand binding of selected glycoside hydrolases

AlphaFold 3 available through the Google DeepMind platform was used to predict protein structures of selected glycoside hydrolases based on their amino acid sequences.[91] Reference protein models coupled with ligands were downloaded from the Protein Data

Bank.[102] We used Coot v.0.9.8.93 to align models produced in this study to reference models.[92] PyMOL v.3.0.0 was used to visualise superimposed structures.[93] PeSTo-Carbs was used to predict carbohydrate binding sites in protein structures, to which we were not able to align available reference ligands.[94]

4303 amino acid sequences predicted with dbCAN3 as belonging to the GH13 glycoside hydrolase family were used to construct the maximum likelihood phylogenetic tree. The multiple sequence alignment was generated with MAFFT v7[95] and trimmed using trimAL v.1.5.rev0 with options "-gt 0.5 -cons 50"[96]. Model testing and the subsequent tree generation according to the best-fitting model (WAG+F+R10) was performed in IQ-Tree v.2.0.5.[80]

Data visualisation

Inkscape v.1.4[112] was used to produce final versions of figures.

QUANTIFICATION AND STATISTICAL ANALYSIS

PERMANOVA was conducted using the 'adonis2()' function and PERMDISP was performed using the 'betadisper()' function implemented in the R package 'vegan'[97]. Pairwise comparisons of Euclidean distances between host taxonomic classes were performed using the 'adonis()' function implemented in the R package 'vegan'. ANOSIM statistics was implemented in the R package 'vegan'. Kruskal-Wallis test and Dunn's post-hoc test were performed using the kruskal_test() and dunn_test() functions, respectively, implemented in the 'rstatix' package in R.[98] 'parafit()' function was implemented in the R package 'ape'[99]. Linear regression models were built using the 'lm()' function and p-values were calculated from the t-values of the linear regression models using the 'summary.lm()' function, both implemented in R. For all presented statistics, the hypothesis test used, the values and representation of n, and measures of precision can be found in the corresponding figure legends, main text, and STAR Methods.