

# Bilirubin reductase shows host-specific associations in animal large intestines

Keith Dufault-Thompson<sup>1</sup>, Sophia Levy<sup>2</sup>, Brantley Hall<sup>2,3</sup>, Xiaofang Jiang<sup>1,\*</sup>

<sup>1</sup>National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894, United States

<sup>2</sup>Department of Cell Biology and Molecular Genetics, University of Maryland, 4062 Campus Drive, College Park, Maryland 20742, United States

<sup>3</sup>Center for Bioinformatics and Computational Biology, University of Maryland, 8125 Paint Branch Drive, College Park, Maryland 20742, United States

\*Corresponding author: Xiaofang Jiang, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Building 38A, Room 6N607, 8600 Rockville Pike, Bethesda, MD 20894, United States. Email: [xiaofang.jiang@nih.gov](mailto:xiaofang.jiang@nih.gov)

## Abstract

Animal gastrointestinal tracts contain diverse metabolites, including various host-derived compounds that gut-associated microbes interact with. Here, we explore the diversity and evolution of bilirubin reductase, a bacterial enzyme that metabolizes the host-derived tetrapyrrole bilirubin, performing a key role in the animal heme degradation pathway. Through an analysis of the bilirubin reductase phylogeny and predicted structures, we found that the enzyme family can be divided into three distinct clades with different structural features. Using these clade definitions, we analyzed metagenomic sequencing data from multiple animal species, finding that bilirubin reductase is significantly enriched in the large intestines of animals and that the clades exhibit differences in distribution among animals. Combined with phylogenetic signal analysis, we find that the bilirubin reductase clades exhibit significant associations with specific animals and animal physiological traits like gastrointestinal anatomy and diet. These patterns demonstrate that bilirubin reductase is specifically adapted to the anoxic lower gut environment of animals and that its evolutionary history is complex, involving adaptation to a diverse collection of animals harboring bilirubin-reducing microbes. The findings suggest that bilirubin reductase evolution has been shaped by the host environment, providing a new perspective on heme metabolism in animals and highlighting the importance of the microbiome in animal physiology and evolution.

**Keywords:** heme degradation, bilirubin reductase, bilirubin, host-associated microbes, host adaptation

## Introduction

Animal-associated microbiomes play essential roles in health and disease, including digesting dietary metabolites, providing colonization resistance against pathogens, and stimulating host inflammatory responses [1]. Investigating these host-microbiome dynamics is essential to understanding nearly all facets of animal physiology and ecology, and our view of animals as holobionts, superorganisms consisting of a host and its associated microbes, is constantly evolving [1, 2]. As new microbial functions are characterized, it is essential to consider them within the context of their potential host-microbe interactions.

Heme and heme derivatives are essential cofactors in all domains of life, leading to a broad diversity in the metabolism of these tetrapyrrole metabolites across different organisms [3]. The evolution of closed circulatory systems resulted in the increased importance of heme degradation in animals, where heme needs to be converted into excretable waste products as part of the natural turnover of blood cells [4, 5]. The secretion of these oxidized heme derivatives like bilirubin into the gastrointestinal tract provides access to readily reducible electron acceptors for bacteria with the enzymes needed to utilize them. The reduction of bilirubin by gut microbes determines if the bilirubin will be reabsorbed by the animal or reduced by microbes and excreted as waste products [6].

In a previous study, we identified and characterized a microbial bilirubin reductase enzyme, BilR, filling a long-standing gap in our understanding of the heme degradation pathway [7]. Bilirubin reductase is an anaerobic reductase belonging to the Old Yellow Enzyme family [7]. The Old Yellow Enzyme family is broadly distributed across the microbial tree of life, catalyzing the reduction of various organic and inorganic compounds [8, 9]. This family of enzymes consists of a triose-phosphate isomerase (TIM barrel) domain, where the substrate binding site is located, and can have additional domains related to electron transfer [10]. The gene encoding bilirubin reductase, *bilR*, was identified primarily in anaerobic bacterial species from Bacilli and Clostridia classes [7]. These classes of bacteria are commonly found in the human gut, where they utilize a variety of alternative electron acceptors in the largely anoxic gut environment. The identification of bilirubin reductase has provided the foundation needed to understand how the bilirubin reductase enzyme has evolved in the context of animal physiology and ecology.

Host-associated microbes are constantly evolving in the context of the host environment, subject to host immune responses, changes in host physiology, and the production of host-derived metabolites [11]. Opportunistic pathogens like *Helicobacter pylori* have undergone significant diversification within different human populations, resulting in changes in virulence [12], and

Received: 6 September 2024. Revised: 2 December 2024. Accepted: 8 December 2024

© Published by Oxford University Press on behalf of the International Society for Microbial Ecology 2024.

This work is written by US Government employees and is in the public domain in the US.

commensal bacteria like *Bacteroides thetaiotaomicron* have evolved to degrade the complex carbohydrates present in their host's diet [13]. Recent studies have highlighted that microbes from the gut environment can interact with a wide range of host-derived metabolites, including bile acids [14], hormones [15, 16], and mucin [17], potentially giving them competitive advantages over other members of the microbiome that cannot utilize these compounds as carbon sources or electron acceptors. The excretion of bilirubin into the gut provides a potential metabolic niche for gut microbes to fill, and understanding this process can provide new insights into animal health and microbiome evolution.

In this study, we investigate the evolution and ecology of the bilirubin reductase enzyme family using phylogenetic and metagenomic approaches. We find that bilirubin reductase can be delineated into three clades with conserved structural differences. Through an analysis of gastrointestinal metagenomes, we found that bilirubin reductase is enriched in the anoxic niche of the large intestines of animals, highlighting its likely role in anaerobic respiration in the gut. We further investigate the ecological significance of bilirubin reductase in the animal gastrointestinal tract, finding species-specific distributions of bilirubin reductase clades and strong signals of clade-specific associations with different animal physiological traits. Overall, our results suggest bilirubin reduction is a metabolic trait adapted to the anoxic niche of the animal gut. These results have significant implications for understanding bilirubin metabolism in the gut and provide a foundation for future work with these enzymes.

## Materials and methods

### BilR and BilS sequence analysis

We identified putative bilirubin reductase genes using a previously generated HMM profile of bilirubin reductase profile sequences [7]. We performed a search against the representative genomes of the Genome Taxonomy Database (GTDB) (version r212) [36] using HMMSearch (version 3.4) [37]. The sequences were curated, keeping hits to the HMM profile with E-values less than  $1 \times 10^{-85}$  an E-value that was determined to be strict enough to differentiate putative BilR sequences from other closely related Old Yellow Enzymes [7], while also including 13 homologous genes from a neighboring gene family that represented the closest related sequences that were not in the BilR family that we had previously identified [7]. Lastly, the sequences dataset was curated through initial sequence alignments and phylogenetic tree construction, resulting in the identification of seven sequences that poorly aligned to the rest of the BilR proteins and lacked of the defining "HGDR" BilR active site motif [7] and were subsequently considered spurious hits and removed from the dataset.

The resulting putative BilR sequences and 13 outgroup sequences were aligned using ClustalO (version 1.2.4) [38], and the subsequent alignment was trimmed using TrimAl (version 1.2) [39] keeping only positions within the conserved TIM-barrel domain that were not gaps in greater than 10% of sequences. A phylogenetic tree was constructed based on the trimmed alignment using IQtree (version 2.2.0.5) [40] with 1000 ultrafast bootstraps using the best-fit model determined by the IQtree Model Finder [41] (LG amino acid exchange rate matrix, empirical among acid frequencies, with a FreeRate model of rate heterogeneity with eight categories). The phylogenetic tree was visualized using iTol [42].

Homologs of the putative BilS protein were identified by searching for hits to the COG0716 COG (Clusters of Orthologous Groups) using eggno-mapper (version 2.1.12) [43]. Based on our previous analyses, the BilS proteins is part of a broadly distributed group of flavodoxin-like proteins belonging to the COG0716 group [7]. Proteins from the GTDB representative genomes containing putative *bilR* genes were annotated using eggno-mapper to assign them to COG ortholog IDs. The COG annotations were filtered by an E-value of  $1 \times 10^{-20}$  to remove potential spurious hits. The set of proteins in this filtered hit dataset were then subset to keep proteins whose genes were adjacent to a putative *bilR* gene, giving us a set of putative *bilR*-associated *bilS* genes. These putative *bilS* genes were mapped onto the BilR phylogenetic tree using iTol [42].

Predicted structures for representative BilR sequences from the three clades (BilR-basal enzyme from *Ruminococcus gnavus* (GCF\_000507805.1), BilR-short enzyme from *Clostridium symbiosum* WAL-14163 (GCF\_000189595.1), BilR-insertion enzyme from *Faecalibacillus intestinalis* SNUG30099 (GCA\_003024685)) were generated using AlphaFold2 (version 2.3.2) [44] with default settings and visualized using PyMOL. Structures for the 118 BilR-Insertion proteins were generated using ESMFold (version 2.0.0) [45]. The sequences of the BilR-insertion proteins were aligned using ClustalO (version 1.2.4) [38], and the aligned secondary structures of the proteins were visualized using SSDraw [46].

### Metagenomic analysis

A collection of metagenomic samples from animal gastrointestinal tracts was curated and annotated with taxonomic, dietary, and anatomic information (Supplementary Table 1). Metagenomics samples were downloaded from the Sequence Read Archive and quality-trimmed using TrimGalore (version 0.6.7) [47]. Only samples with over 1 000 000 quality filtered reads were kept for subsequent analysis. The quality trimmed reads were mapped to a database containing all putative *BilR* gene sequences using Bowtie2 (version 2–2.5.3) [48], and the per-gene mapping statistics were summarized using Samtools (version 1.19) [49]. The reads were also mapped to the database of 120 marker gene sequences from all GTDB representative genomes.

Because of the variability in quality and availability of many of the host animal genomes, removing contamination by mapping reads to the host genome was not possible for many of the samples. Instead, the sum of the number of reads mapped to all 120 of the GTDB marker genes was used to account for the differences in bacterial sequence content in each sample. To normalize the amount of mapping to each *bilR* clade, the number of reads mapped to each gene was summed for all the genes in that clade for each sample. These values were then divided by the average length of the genes in that clade (BilR-Basal: 1899.9 nucleotides; BilR-short: 1096.5 nucleotides; BilR-Insertion: 1972.8 nucleotides), divided by the total mapping to the 120 marker genes, and finally multiplied by  $10^8$  (analogous to the per million scaling factor used in reads per kilobase million and transcripts per million normalization approaches) to give a normalized gene mapping for each BilR clade in each sample. The mapping data was analyzed in R and visualized using ggplot2 [50]. Diagrams of the gastrointestinal tracts of ruminants, chickens, and mice were made using Biorender.

Statistical comparisons of total BilR content in different gastrointestinal regions were made using the "wilcox.test" function from the R stats package (version 4.2.1) to perform a one-sided test if the normalized BilR abundance in the large intestine samples (cecum, colon, rectum) were higher than the samples from the stomachs (rumen, reticulum, omasum, abomasum) or small

intestines (duodenum, jejunum, ileum). The “adonis2” function from the vegan package (version 2.6-4) was used to perform PERMANOVA tests for significant differences in composition between groups of samples based on animal taxonomy, diet, or gastrointestinal anatomy. Significant enrichment of bilirubin reductase clades was tested for using Wilcoxon Rank Sum tests using the “wilcox.test” function in the R stats package (version 4.2.1). First, the normalized *bilR* read mapping in each sample was converted to relative abundance values, giving a relative abundance of each *bilR* clade out of the total mapping to *bilR* in that sample. The samples were then grouped based on host traits, and for each host trait, a one-sided Wilcoxon Rank Sum test was used to determine if the relative abundance of the *bilR* clade was higher in the samples associated with the trait (e.g. *ruminant samples*) compared to the background of the rest of the samples not associated with the trait (e.g. *non-ruminant samples*). Lastly, the Benjamin-Hochberg procedure was used to correct for multiple testing within each broader category (i.e. host class, order, species, diet, or gastrointestinal anatomy), and a corrected *P* value threshold of 0.05 was used to determine if a *bilR* clade was significantly enriched in a group or not.

### Phylogenetic signal analysis

Phylogenetic signal analysis was performed using the “delta\_statistic” R package [51]. First, the isolation information for each bacteria containing an annotated *bilR* gene was identified using the metadata available on GenBank and the literature describing the bacterial species. Based on this isolation information, each *bilR* gene and bacterial genome was associated with the host traits (taxonomy, diet, gastrointestinal anatomy) of the animals from which the bacteria were isolated. Bacteria that were not isolated from host-associated environments were assigned to a non-animal group and included in the subsequent phylogenetic signal analyses as part of the “non-trait” groups. For each host trait, the delta statistic was calculated by comparing the trait to all other species (e.g. carnivore vs non-carnivore) using the *BilR* sequence tree and the GTDB species tree subset to only include species with *BilR* genes [36]. One thousand random simulations were then performed for the gene and species trees where the host traits were shuffled throughout the trees, keeping the same proportion of positive and negative trait assignments, and the delta statistic was calculated for each simulation. A *P* value was then calculated based on a comparison of the delta statistics from the actual trait assignments compared to the distribution of the 1000 random simulation delta statistics.

## Results

### Bilirubin reductase diversification has led to structurally distinct clades

Bilirubin reductase was previously found to be part of the Old Yellow Enzyme family, a broadly distributed family of anaerobic reductases [7, 9]. Bilirubin reductase enzymes can be classified into three evolutionarily related clades distinguished by their protein domain composition and structural characteristics (Fig. 1). The *BilR*-basal clade proteins consist of a TIM-barrel domain, an FAD-binding domain, and an NADPH-binding domain and are structurally homologous to other characterized Old Yellow Enzymes, including 2,4 dienoil-CoA reductase from *Escherichia coli* [10] and daidzein reductase from *Eggerthella* sp. YY7918 [18] (Fig. 1). Based on the phylogenetic placement of the *BilR*-basal clade and its similarity to other Old Yellow Enzymes, we speculate that the ancestral form of the bilirubin reductase enzyme was similar to the *BilR*-basal proteins.

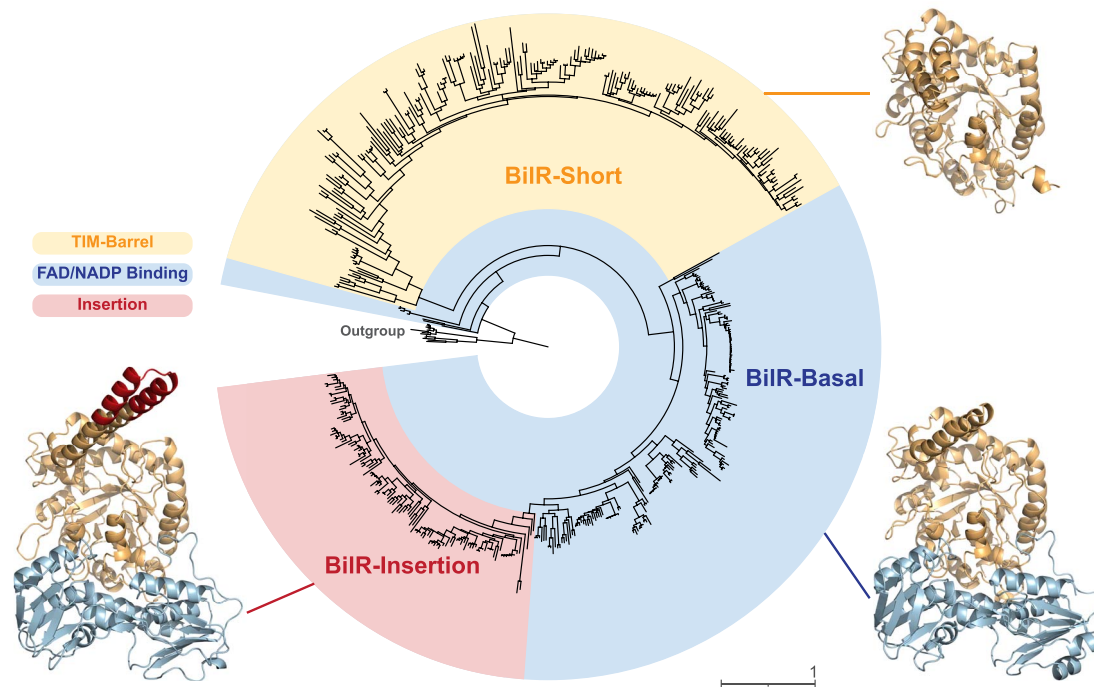
The *BilR*-short clade likely evolved from an ancestral bilirubin reductase enzyme that underwent a loss of the FAD-binding and NADPH-binding domains (Fig. 1). *BilR*-short proteins consist of only a TIM-barrel domain, similar to other well-studied old yellow enzymes from yeast [19]. The primary difference between these other Old Yellow Enzymes and *BilR*-short is the retention of a small C-terminal domain that contains an Iron-Sulfur center C-X2-C-X3-C binding motif like seen in homologous proteins like the *E. coli* 2,4-dienoil-CoA reductase protein [10] and the two other clades of bilirubin reductase enzymes (Supplementary Fig. 1). Despite the loss of these FAD and NAD binding domains, previous experiments have demonstrated that the *BilR*-short enzyme is still functional, possibly requiring an accessory flavodoxin-like protein, *BilS*, which is encoded by a gene that is frequently adjacent to *BilR* genes from the *BilR*-short clade [7] (Supplementary Fig. 2). Other flavodoxin-like proteins are involved in various reduction reactions [20], suggesting that the *BilS* protein may fulfill the electron transfer functions lost in the *BilR*-short clade. Further investigation is needed to fully characterize how the truncation of the protein impacts the function of *BilR*-Short enzymes and to understand its implications for bilirubin metabolism in the gut.

The *BilR*-insertion enzymes are similar to the *BilR*-basal enzymes, with the addition of a 23 amino acid insertion in the capping domain of the protein (Fig. 1). This capping domain region, located on the loop between the third beta-strand and third alpha helix of the TIM-barrel domain, sits near the predicted substrate-binding pocket. The sequence of the insertion region is not well conserved compared to the flanking regions of the TIM-barrel domain, but the predicted structures of this insertion are generally consistent, with only 12 of the 118 predicted structures being predicted to have different secondary structures (Supplementary Fig. 3). While the function of this region has not been extensively characterized, it has been found to vary in length between different Old Yellow Enzymes [21] and has been implicated in substrate binding in *Shewanella* yellow enzyme 1 from *Shewanella oneidensis* [22]. Overall, the bilirubin reductase phylogeny and predicted structural differences between clades suggest that bilirubin reductase evolution has involved at least two major structural mutations that have led to the three distinct clades we have identified in this protein family.

### Bilirubin reductase is a niche-specific gene adapted to the animal lower-gut

Bilirubin reductase’s function as an anaerobic reductase is reflected by its enrichment in the anoxic environment of the large intestines of multiple animal species. Animal gastrointestinal tracts exhibit variation in length, physiology, and microbial community composition, often reflecting the adaptation of the animals to feeding on different types of foods [23, 24]. Despite these differences, most animal gastrointestinal tracts can be broadly described as having relatively aerobic stomachs and small intestines followed by a primarily anoxic large intestine where most of the microbial degradation of food and host-derived metabolites occurs [23]. Bilirubin is secreted from the bile duct into the duodenum region of the small intestines of most animals (Fig. 2A–C). Bilirubin reductase is primarily found in the anaerobic bacterial species from the *Clostridia* and *Bacilli* classes of the *Bacillota\_A* and *Bacillota* phyla, respectively [7]. We hypothesized that if the bilirubin reductase enzyme has evolved in the context of animal heme metabolism, then it should be enriched within the anoxic regions of the gut that also contain bilirubin.

To examine if the bilirubin reductase gene is associated with a specific region of the gut, we analyzed shotgun metagenomic



**Figure 1.** Phylogenetic reconstruction and structures of the BilR enzyme family. (A) Phylogenetic reconstruction of BilR sequences based on 513 predicted bilirubin reductase proteins. Clades are colored and labeled to highlight major structural differences between the BilR sequences. AlphaFold2 predicted structures of a BilR-basal enzyme from *R. gnavus* (GCF\_000507805.1), BilR-short enzyme from *Clostridium symbiosum* WAL-14163 (GCF\_000189595.1), and BilR-insertion enzyme from *Faecalibacillus intestinalis* SNUG30099 (GCA\_003024685) are shown in the same orientation next to their respective clades. The protein structures are colored to highlight shared protein domains and significant features of interest in the proteins, with the TIM-barrel domain colored yellow, the Flavodoxin-like fold/NADP(H)-binding domain colored blue, and the capping domain insertion colored red.

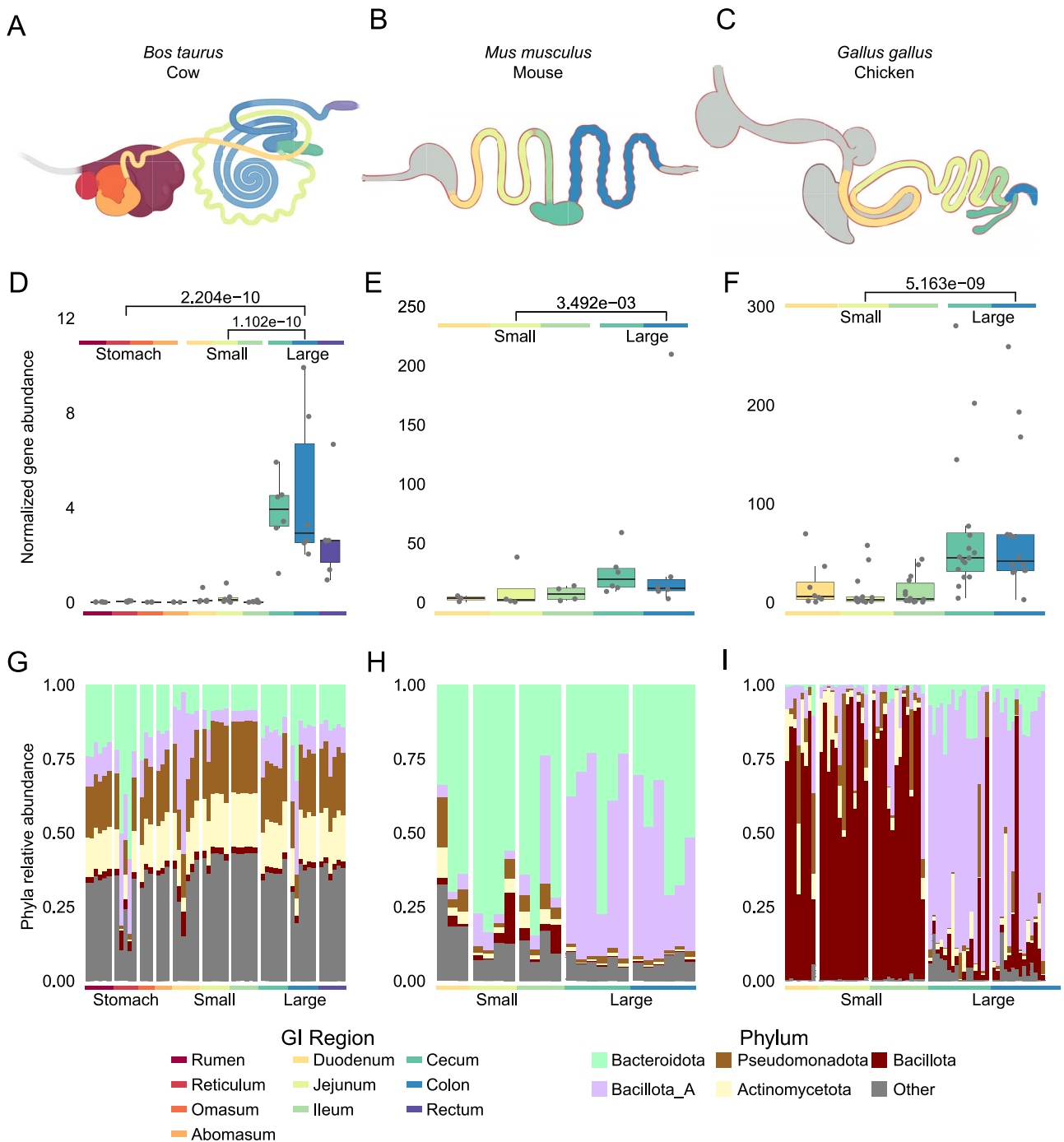
sequencing data from three studies that sampled the microbial communities along the length of the gastrointestinal tracts of different animals. These studies included samples from mice (*Mus musculus*) (PRJNA1055712), chickens (*Gallus gallus*) (PRJNA417359) [25], and seven different ruminant species (PRJNA657455) [26] (Supplementary Table 1). In all nine species, distinct changes in bilirubin reductase abundance were observed along the gastrointestinal tract, with the upper gastrointestinal tract (stomachs and small intestines) having very little detected bilirubin reductase and the large intestines showing significant enrichment of bilirubin reductase (Fig. 2D–F, Supplementary Fig. 5). Bacillota\_A, the predominant bacterial phylum harboring the BilR gene [7], is present in various non-lower gut environments, including the rumen of ruminants (Fig. 2G–I), the enrichment of BilR is seen specifically in the anoxic conditions of the large intestines.

The trends in bilirubin reductase presence in the anoxic environment of the lower gut were consistent across mice, chickens, and seven ruminant species despite the differences in diet and gastrointestinal anatomy between the animals (Fig. 2, Supplementary Fig. 4). Bilirubin reductase was significantly more abundant in the large intestines of cows (*Bos taurus*) than in the stomachs or small intestines (stomachs:  $P$  value  $2.204 \times 10^{-10}$ , Wilcoxon Rank Sum; small intestines:  $P$  value  $1.102 \times 10^{-10}$ , Wilcoxon Rank Sum). Similarly, the abundance of BilR in mice and chickens was higher in the largest intestines compared to the small intestines (mice:  $P$  value  $2.492 \times 10^{-3}$ , Wilcoxon Rank Sum; chickens:  $P$  value  $5.163 \times 10^{-9}$ , Wilcoxon Rank Sum). This trend was consistent in the other six ruminant species, except the water deer (*Hydropotes inermis*) samples, which had an enrichment of BilR-basal in the small intestines and then an enrichment of BilR-insertion in the large intestines (Supplementary Fig. 4). Despite the presence of bacteria from the

Bacillota and Bacillota\_A phyla in non-lower gut environments and the presence of bilirubin in the small intestines, the significant enrichment of bilirubin reductase in the anoxic environment of the large intestines underscores the gene's niche-specific adaptation. Significant functional heterogeneity has been observed in different gastrointestinal tract regions [27], and our results suggest that bilirubin reduction is a large intestine-specific function. Adaptation of bilirubin reductase enzyme to the anoxic niche of the lower gut appears to be reflected by the distinct enrichment of bilirubin reductase in the large intestines of mice, chickens, and multiple ruminant species, despite the differences in gastrointestinal anatomy and diet between these animals. This niche specificity highlights the importance of understanding host-associated environments' local biogeography and ecology, especially when considering highly variable environments like the animal gut.

### Bilirubin reductase clades are associated with different animal gastrointestinal tracts

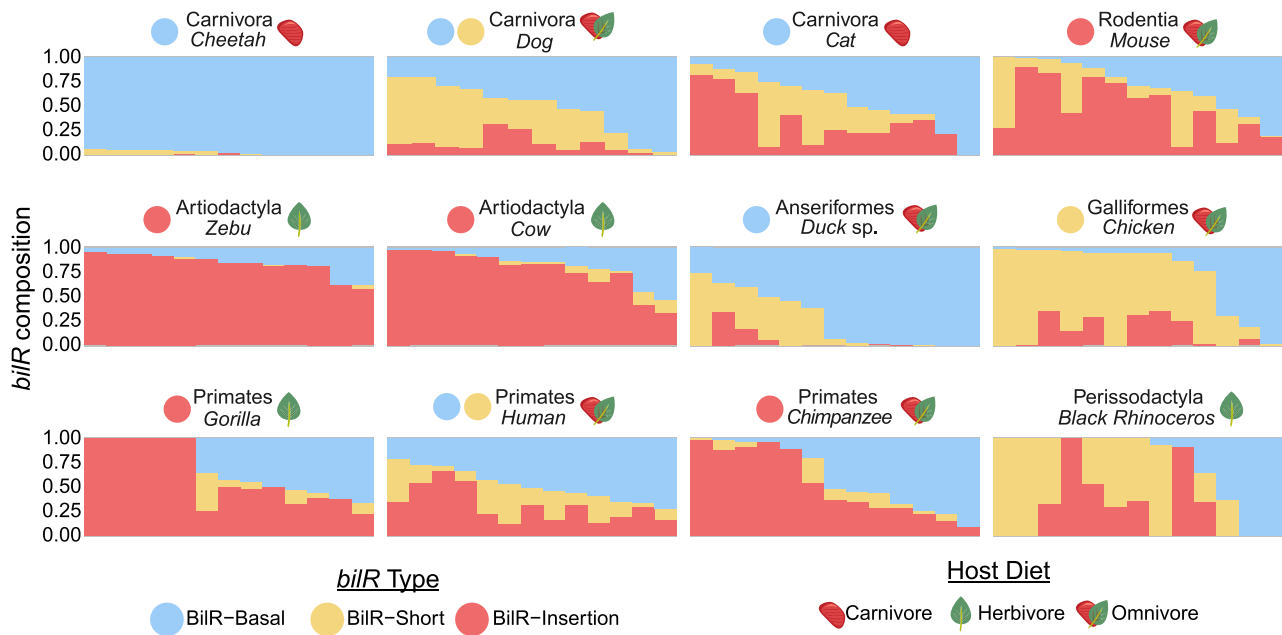
We investigated the distribution of BilR enzymes in different animals, finding species-specific compositional differences between animals. To identify possible differences in BilR content between animals, we curated a collection of 1025 animal gut metagenomes from 150 species and compared the relative abundance of each bilirubin reductase clade between the samples (Fig. 3, Supplementary Fig. 5). When considering only animals with at least 10 samples, we found that the variation in bilirubin reductase composition between animal species was significantly higher than the variation observed within species (PERMANOVA,  $P$  value 0.001), suggesting that there are species-specific patterns in bilirubin reductase composition.



**Figure 2.** Total bilirubin reductase presence in animal gastrointestinal tracts. Data is shown for an analysis of *Bos taurus* (A, D, G), *Mus musculus* (B, E, H), and *Gallus gallus* (C, F, I). Panels A–C show diagrams of the gastrointestinal tracts for each animal with the regions where samples are colored. Gray areas indicate that no samples were from those areas of the gastrointestinal tracts. Panels d–f show the normalized gene abundance of bilR within the metagenomes from each region of the gastrointestinal tracts in each animal. Boxplots show the median and middle quartiles and individual data points are shown as gray points. And are colored to indicate the gastrointestinal region they represent. Statistical comparisons between groups of samples are indicated above each plot, with the P values shown being from one-sided Wilcoxon rank sum tests testing if the large intestine samples had higher levels of bilR than the small intestine samples or stomachs. Panels G–I show relative abundance plots of bacterial taxa at the phylum level based on mapping marker genes to the GTDB reference gene set. The top five most abundant taxa across all samples are colored, with the remaining taxa being grouped into the other category and colored gray.

While most samples had a mixture of bilirubin reductase types, we observed consistent and significant trends across samples from multiple animal species (Fig. 3, Supplementary Fig. 5, Supplementary Table 2). Cheetahs (*Acinonyx jubatus*), domesticated dogs (*Canis lupus familiaris*), and domesticated cats (*Felis catus*), all had significant enrichment of BilR-basal, suggesting a

possible association between this bilirubin reductase clade and animals from the *Carnivora* order. BilR-insertion was enriched in cows (*B. taurus*) and zebu (*Bos indicus*), animals characterized by having the distinct ruminant gastrointestinal tracts found in the animals of the *Artiodactyla* order. The *Aves* and *Primates* order species showed species-specific trends, with individual species



**Figure 3.** Presence of BilR in different animal metagenomes. Plots show the composition of bilirubin reductase clades in each sample. Samples are grouped by animal species, and the diet of each animal is shown in the graphics to the right of each animal's name. The significantly enriched bilirubin reductase types, based on Wilcoxon rank sum tests, are shown by the colored circles to the left of each animal's name.

being enriched for different bilirubin reductase types. In the Aves order, chickens (*G. gallus*) were enriched for BilR-short, whereas the duck (*Anatidae*) samples were enriched for BilR-basal. In the primate order, humans (*Homo sapiens*) were enriched for BilR-basal, gorillas (*Gorilla gorilla*) were enriched for BilR-insertion, and chimpanzees (*Pan troglodytes*) were enriched for both BilR-basal and BilR-insertion. The samples analyzed for multiple animals were taken from different studies or different sampling locations when possible, including four different studies for domesticated cats (*F. catus*) and sampling from multiple geographic locations for chickens (*G. gallus*), providing a broader survey of the diversity in these animals and mitigating some of the effects of sample location and study. Although the limited sample sizes make it difficult to draw conclusions about the enrichment of bilirubin reductase clades within the other animals, the differences observed from animal to animal highlight the effects of bilirubin reductase specialization and niche differentiation-driven host-specific selective pressures.

Grouping the animal samples into broader categories highlighted the potential associations of animal taxonomy and physiology with the bilirubin reductase clades (Supplementary Fig. 6, Supplementary Table 2). In relation to animal taxonomy, significant differences between taxa were observed at both the class (Supplementary Fig. 6A) and order levels (Supplementary Fig. 6C) (PERMANOVA; Class:  $P$  value 0.001, Order:  $P$  value 0.001). Animals from the class Mammalia had an overall enrichment for BilR-insertion (Wilcoxon Rank Sum;  $P$  value 1.21E-36) (Supplementary Fig. 6A), whereas the Aves class showed an enrichment for BilR-short (Wilcoxon Rank Sum;  $P$  value 2.96E-35). When considering animals at the order taxonomic level, there was variation between different taxa (Supplementary Fig. 6C). Some Mammalia orders were enriched for BilR-basal (Wilcoxon Rank Sum; Carnivora:  $P$  value 7.61E-15, Primates;  $P$  value 1.17E-07) and some for BilR-insertion (Wilcoxon Rank Sum; Artiodactyla;  $P$  value 1.22E-28, Rodentia;  $P$  value 4.57E-11) (Supplementary Fig. 6C). Similarly, the Aves order Anseriformes was enriched for BilR-basal (Wilcoxon Rank Sum;  $P$  value 2.26E-07) whereas the Galliformes

order was enriched for BilR-short (Wilcoxon Rank Sum;  $P$  value 1.06E-55) (Supplementary Fig. 6C). While these results highlight some species-specific and taxonomic trends, there are various other factors including dietary differences, geographic differences and even study-specific factors that may play roles driving the inter- and intra-group differences we observe in bilirubin reductase types.

These bilirubin reductase trends may also reflect the differences in gastrointestinal anatomy or diet between the animals, which both showed significant differences in BilR composition (PERMANOVA; Diet  $P$  value 0.001, GI Tract  $P$  value 0.001). In terms of diet, we found that carnivores tended to be enriched for BilR-basal genes, whereas herbivores tended to be enriched for BilR-insertion genes, largely mirroring the trends seen for the Carnivora and Artiodactyla orders, which consist of carnivores and herbivores respectively (Supplementary Fig. 6B). Among the gastrointestinal tract categories, ruminants had a clear and significant enrichment for BilR-insertion genes consistent with the trends seen for samples from cows and zebu (Wilcoxon Rank Sum;  $P$  value 1.58E-32) (Supplementary Fig. 6D). These trends highlight the diversity of the bilirubin-reducing microbiota in different animals, suggesting potential differences in this pathway between different types of animals. Various other factors, including specific diet composition, antibiotic administration, and geography, may contribute to these different trends, highlighting the need for more cross-sectional studies and sampling of diverse animal populations to fully understand the drivers of these differences.

### Host-specific adaptation of bilirubin reductase is reflected in the gene level

Given the observed host-specific association of bilirubin reductase clades, we aimed to investigate whether this pattern is driven by selection pressure acting on species with the BilR gene or directly on the gene itself. We collected information on the environment where each of the bacteria with a bilR gene was collected (Materials and Methods, Supplementary Table 3).

Corresponding to what we had previously observed in the association of BilR with the lower gut environment, nearly all of the bacteria found to have *bilR* genes were from animal-associated environments (429/513 species), with only a few species being from environmental (27/513 species) or artificial ecosystems (4/513 species). We performed a phylogenetic signal analysis to assess the associations between different bacterial ecological information and phylogenies at the gene and species levels for bilirubin reductase and the bilirubin reductase encoding bacterial species (Fig. 4, Supplementary Table 4). The phylogenetic signal for the host animal diet was consistently stronger in the gene phylogeny compared to the species phylogeny of bilirubin reductase-containing species (Fig. 4). The strongest signal was associated with the carnivore diet (delta 43.1, *P* value 0.01), with no significant signal seen in the species tree (*P* value 0.9). The omnivore diet and herbivore diet associations also had stronger associations in the gene tree (omnivore: 6.7, *P* value 0; herbivore: 10.0, *P* value 0) compared to the species tree (omnivore: 2.7, *P* value 0; herbivore: 6.7, *P* value 0). These patterns are based on limited data, with some categories, like the carnivores or birds, being represented by isolates from only a few animal species (Supplementary Table 3). Despite this limited coverage and diversity of the bacterial genomes, these trends, in combination with the metagenomic analyses, suggest that the evolution of bilirubin reduction has been influenced by the gut environment that the bilirubin-reducing bacteria are found in.

The phylogenetic reconstruction of the bilirubin reductase genes also highlighted cases of likely horizontal gene transfer events that introduced the *bilR* gene into new taxa and new environments. We identified multiple clades within the bilirubin reductase phylogeny that contained genes from distinct taxonomic lineages, suggesting that these clades had acquired their genes through horizontal gene transfer (Supplementary Fig. 7). Two clades of Actinomycetota were seen in the phylogeny, one consisting primarily of genes from Bifidobacteriaceae strains from chicken and monkey-associated environments and another consisting of genes from Coriobacteriaceae and Atopobiaceae primarily from chicken-associated environments. We also observed a clade of genes from the species in the *Psychrobacter* genus of the Pseudomonadota phylum that were primarily from aquatic environments. These instances provide examples of how the evolution of bilirubin reductase can be influenced by horizontal gene transfer, resulting in the introduction of bilirubin reductase into new environments and animal microbiomes.

## Discussion

Here, we present an analysis of the bilirubin reductase enzyme family, placing this critical enzyme in an evolutionary and ecological context. We find that the bilirubin reductase enzyme is uniquely associated with the anoxic niche of animal large intestine and has differentiated into three structurally distinct clades with different animal host-association patterns. While there are still open questions about the functional differences between the three BilR clades, our results suggest that bilirubin reductase evolution has been shaped by the niche of the anoxic lower gut.

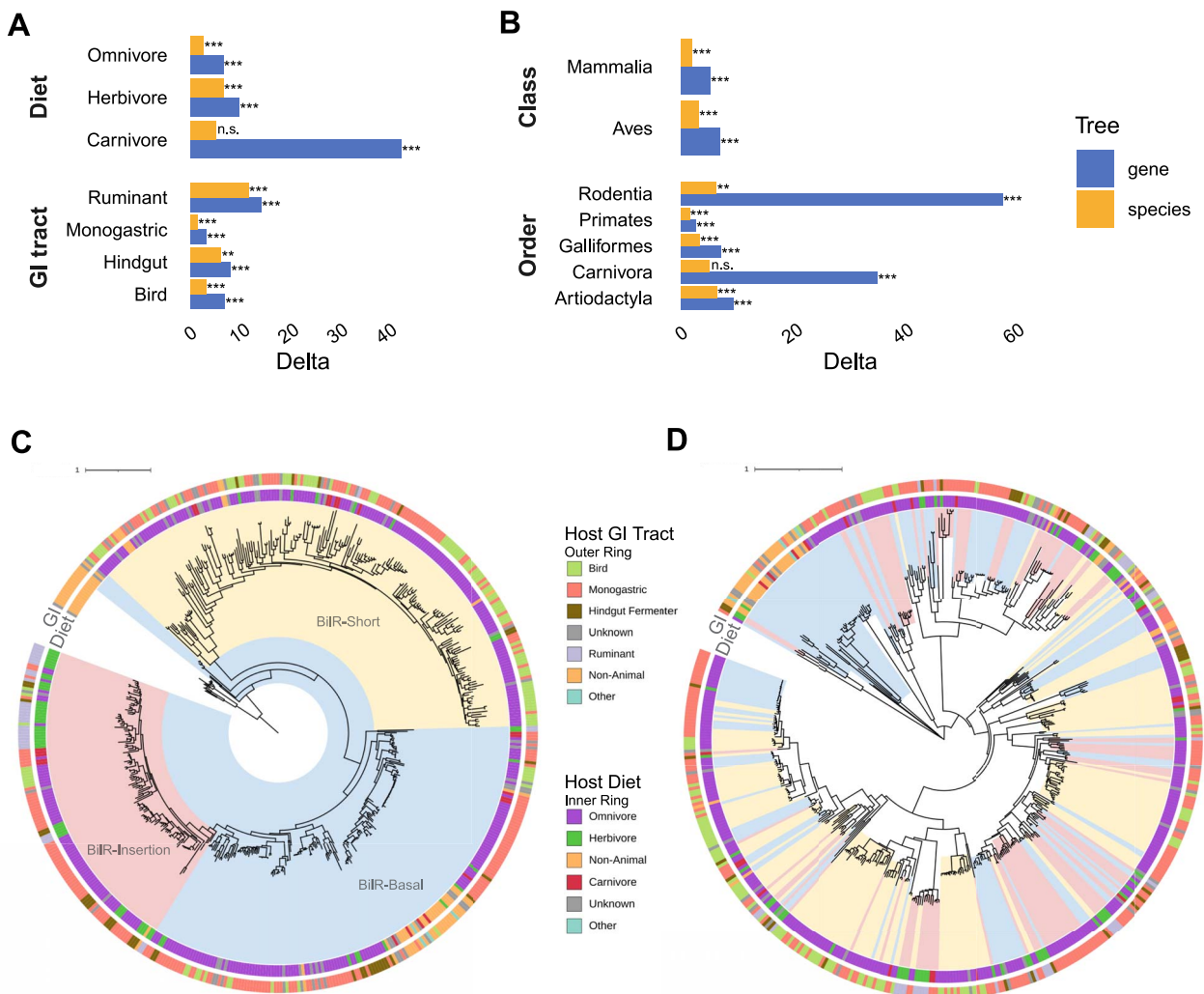
Bilirubin reductase's activity as an anaerobic reductase is likely involved in anaerobic respiration in the gut, providing a mechanism for bilirubin reducing microbes to the host-produced bilirubin molecules. This function suggests a possible metabolic niche that could provide bilirubin reducing microbes with a competitive advantage through the use of an electron acceptor that is not widely utilized by the other microbes in the gut. The description

of other bacterial enzymes in gut microbes, including the bile acid reductases [14] and a cholesterol reductase [28], suggests that there are a broad set of metabolic niches in the gut environment that involve the utilization of host derived metabolites. These metabolic activities often convert the metabolites into compounds with different physical properties (e.g. differences in solubility between bilirubin and urobilinogen) [7] and can directly impact the reabsorption of the metabolite by the host (e.g. cholesterol reduction) [28]. As more of these functions are characterized, we can start to build a more complete picture of the metabolic diversity that exists in the gut microbiome and better understand the complex set of interactions that are involved in host-microbiome dynamics.

Bilirubin reductase appears to be specifically associated with the lower gut in animals and was previously found to be nearly universally present in healthy adult humans [7]. This unique association with the gut environment suggests that bilirubin reductase could be used as a biomarker for gut health or for fecal contamination in the environment. Some commonly used biomarkers like fecal coliform bacteria [29] and crassphages [30] are to detect fecal contamination, and some microbial genes are used as disease biomarkers, including the *tcdB* toxin gene from *Clostridium difficile* [31]. Although taxonomic redundancy across different environments can often make bacterial taxa unreliable as biomarkers, genes can often be extremely sensitive to evolutionary pressures in different environments [32] and may serve as more precise markers. Bilirubin reductase was also observed to be missing in a significant fraction of young infants during the period when they are most susceptible to developing jaundice and in adult inflammatory bowel disease patients [7]. This suggests a possible role for bilirubin reductase, or absence of bilirubin reductase, as a biomarker for susceptibility to bilirubin-related disorders like hyperbilirubinemia and neonatal jaundice, where the ability of the microbiome to degrade bilirubin may play a significant role in disease etiology.

The differential presence of bilirubin reductase types in different animals likely reflect the differences in animal host physiology and diet. Bilirubin production varies greatly between animals as does the relative amount of biliverdin (a precursor of bilirubin) and bilirubin that are secreted into the gut [33], leading to differences in the availability and concentration of bilirubin in the animal gut environment. These differences in bilirubin availability may create environments that are more favorable for different types of bilirubin reductase, potentially explaining their differences in relative abundance between animals. Similarly, meat and plant based animal diets have significantly different abundances of sugars, complex polysaccharides, and amino acids, potentially creating environments that are more or less favorable for anaerobic respiration in the gastrointestinal tract. Carbohydrate-rich diets have the potential to act as rich energy sources, but the gastrointestinal tract is often lacking in available electron acceptors leading to a preference for fermentation over anaerobic respiration [34, 35]. The prevalence of fermentation in environments that are potentially rich energy sources may provide a metabolic niche for bacteria that are respiring using alternative electron acceptors like bilirubin.

We attempt to provide an overview of bilirubin reductase distribution among a diverse collection of animals, but our analysis is limited by the availability of samples from different types of animals. Animals that are model organisms (e.g. mice) or domesticated (e.g. chickens, cows, dogs, or cats) are overrepresented in our dataset, reflecting the availability of data from these animals that are easier to work with. In contrast, collecting samples from



**Figure 4.** Host and environment association with BilR at the gene and species level. (A) Phylogenetic signal analysis results for the species and gene trees based on animal diet and GI tract anatomy. (B) Phylogenetic signal analysis results for the species and gene trees based on animal class and order. (C) Tree based on protein sequences of identified BilR proteins. (D) Species tree of BilR-containing bacteria. Trees are colored to show the BilR clades. The inner color strip on the trees shows the diet of the animal from which the bacterial species was isolated. The outer color strip shows the GI tract anatomy of the animal from which the bacterial species was isolated.

wild animals is often difficult, and many studies focusing on these animals have limited sample sizes. The limited number of samples from many animals also makes it difficult to draw conclusions about the relationship between bilirubin reductase and traits that are often conflated with each other, such as diet and gastrointestinal tract anatomy. More samples from a diverse collection of animals are needed to better understand the trends seen in our analysis. In addition to these host-related limitations, it is important to recognize that the presence and abundance of the bilirubin reductase gene in a microbiome does not necessarily reflect the activity of the microbes. Bilirubin-reducing microbes may have significant differences in their activity and efficiency due to the type of bilirubin reductase, differential regulation of the gene, and differences in cellular transport. More biochemical analysis of a diverse set of bilirubin reducing microbes is needed to fully understand the functional and health implications of these differences in bilirubin reductase presence.

In conclusion, bilirubin reductase has diverged into three distinct clades, which have varied distributions across various animal species. The enrichment of bilirubin reductase in the anoxic niche of the large intestines, combined with the species-specific

distribution of bilirubin reductase clades, highlights the specific adaptation of bilirubin reductase to host-specific gastrointestinal niches. This highlights the importance of considering microbial metabolism within the context of the microbes' environment as we better understand the complex interactions that occur in host-associated environments like the gut. Further exploration into bilirubin reductase and other similar enzymes promises to expand our understanding of animal and bacterial evolution, providing valuable insights into both animal and bacterial physiology.

## Acknowledgements

This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>).

## Author contributions

K.D. and X.J. conceptualized the project. B.H. and X.J. supervised the project. K.D. and S.L. performed the experiments. K.D. performed data curation and formal analysis. K.D. wrote the original draft. K.D., S.L., B.H., and X.J. reviewed and edited.



## Supplementary material

Supplementary material is available at *The ISME Journal* online.

## Conflicts of interest

B.H. is the inventor on a provisional patent filed by the University of Maryland for the use of the enzyme bilirubin reductase. The other authors declare no competing interests.

## Funding information

K.D. and X.J. are supported by the Division of Intramural Research of the NIH, National Library of Medicine. B.H. is supported by startup funding from the University of Maryland.

## Data and materials availability

All genomic data analyzed in this study are available through the GTDB (<https://gtdb.ecogenomic.org/>). All metagenomic datasets analyzed in the study are publicly available, and the project and run information is detailed in [Supplementary Table 1](#).

## References

- Hou K, Wu Z-X, Chen X-Y et al. Microbiota in health and diseases. *Signal Transduct Target Ther* 2022;**7**:1–28. <https://doi.org/10.1038/s41392-022-00974-4>
- Egan S, Fukatsu T, Francino MP. Opportunities and challenges to microbial symbiosis research in the microbiome era. *Front Microbiol* 2020;**11**:1150. <https://doi.org/10.3389/fmicb.2020.01150>
- Bryant DA, Hunter CN, Warren MJ. Biosynthesis of the modified tetrapyrroles—the pigments of life. *J Biol Chem* 2020;**295**:6888–925. <https://doi.org/10.1074/jbc.REV120.006194>
- Snyder GK, Sheafor BA. Red blood cells: centerpiece in the evolution of the vertebrate circulatory system. *Am Zool* 1999;**39**:189–98. <https://doi.org/10.1093/icb/39.2.189>
- Takemoto JY, Chang C-WT, Chen D et al. Heme-derived bilins. *Isr J Chem* 2019;**59**:378–86. <https://doi.org/10.1002/ijch.201800167>
- Vitek L, Zelenka J, Zadinová M et al. The impact of intestinal microflora on serum bilirubin levels. *J Hepatol* 2005;**42**:238–43. <https://doi.org/10.1016/j.jhep.2004.10.012>
- Hall B, Levy S, Dufault-Thompson K et al. BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen. *Nat Microbiol* 2024;**9**:173–84. <https://doi.org/10.1038/s41564-023-01549-x>
- Shi Q, Wang H, Liu J et al. Old yellow enzymes: structures and structure-guided engineering for stereocomplementary bioreduction. *Appl Microbiol Biotechnol* 2020;**104**:8155–70. <https://doi.org/10.1007/s00253-020-10845-z>
- Peters C, Frasson D, Sievers M et al. Novel old yellow enzyme subclasses. *ChemBiochem* 2019;**20**:1569–77. <https://doi.org/10.1002/cbic.201800770>
- Hubbard PA, Liang X, Schulz H et al. The crystal structure and reaction mechanism of *Escherichia coli* 2,4-dienoyl-CoA reductase. *J Biol Chem* 2003;**278**:37553–60. <https://doi.org/10.1074/jbc.M304642200>
- Toft C, Andersson SGE. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 2010;**11**:465–75. <https://doi.org/10.1038/nrg2798>
- Muñoz-Ramírez ZY, Pascoe B, Mendez-Tenorio A et al. A 500-year tale of co-evolution, adaptation, and virulence: *helicobacter pylori* in the Americas. *ISME J* 2021;**15**:78–92. <https://doi.org/10.1038/s41396-020-00758-0>
- Groussin M, Mazel F, Alm EJ. Co-evolution and co-speciation of host-gut bacteria systems. *Cell Host Microbe* 2020;**28**:12–22. <https://doi.org/10.1016/j.chom.2020.06.013>
- Kang D-J, Ridlon JM, Moore DR 2nd et al. *Clostridium scindens* baiCD and baiH genes encode stereo-specific 7 $\alpha$ /7 $\beta$ -hydroxy-3-oxo- $\Delta^4$ -cholenoic acid oxidoreductases. *Biochim Biophys Acta* 2008;**1781**:16–25. <https://doi.org/10.1016/j.bbali.2007.10.008>
- Jacoby C, Scorza K, Ecker L et al. Gut bacteria metabolize natural and synthetic steroid hormones via the reductive OsrABC pathway. *bioRxiv* 2024.
- Arp G, Jiang A, Dufault-Thompson K et al. Gut bacteria encode reductases that biotransform steroid hormones. *bioRxiv* 2024.
- Glover JS, Ticer TD, Engevik MA. Characterizing the mucin-degrading capacity of the human gut microbiota. *Sci Rep* 2022;**12**:8456. <https://doi.org/10.1038/s41598-022-11819-z>
- Kawada Y, Goshima T, Sawamura R et al. Daidzein reductase of *Eggerthella* sp. YY7918, its octameric subunit structure containing FMN/FAD/4Fe-4S, and its enantioselective production of R-dihydroisoflavones. *J Biosci Bioeng* 2018;**126**:301–9. <https://doi.org/10.1016/j.jbiosc.2018.03.018>
- Karplus PA, Fox KM, Massey V. Flavoprotein structure and mechanism. 8. Structure-function relations for old yellow enzyme. *FASEB J* 1995;**9**:1518–26. <https://doi.org/10.1096/fasebj.9.15.8529830>
- Sancho J. Flavodoxins: sequence, folding, binding, function and beyond. *Cell Mol Life Sci* 2006;**63**:855–64. <https://doi.org/10.1007/s00018-005-5514-4>
- Opperman DJ, Piater LA, van Heerden E. A novel chromate reductase from *Thermus scotoductus* SA-01 related to old yellow enzyme. *J Bacteriol* 2008;**190**:3076–82. <https://doi.org/10.1128/JB.01766-07>
- van den Hemel D, Brigé A, Savvides SN et al. Ligand-induced conformational changes in the capping subdomain of a bacterial old yellow enzyme homologue and conserved sequence fingerprints provide new insights into substrate binding. *J Biol Chem* 2006;**281**:28152–61. <https://doi.org/10.1074/jbc.M603946200>
- McCallum G, Tropini C. The gut microbiota and its biogeography. *Nat Rev Microbiol* 2024;**22**:105–18. <https://doi.org/10.1038/s41579-023-00969-0>
- Youngblut ND, Reischer GH, Walters W et al. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun* 2019;**10**:1–15. <https://doi.org/10.1038/s41467-019-10191-3>
- Huang P, Zhang Y, Xiao K et al. The chicken gut metagenome and the modulatory effects of plant-derived benzylisoquinoline alkaloids. *Microbiome* 2018;**6**:211. <https://doi.org/10.1186/s40168-018-0590-5>
- Xie F, Jin W, Si H et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome* 2021;**9**:137. <https://doi.org/10.1186/s40168-021-01078-x>
- Regional diversity of the gastrointestinal microbiome. *Cell Host Microbe* 2019;**26**:314–24. <https://doi.org/10.1016/j.chom.2019.08.011>
- Kenny DJ, Plichta DR, Shungin D et al. Cholesterol metabolism by uncultured human gut bacteria influences host cholesterol level. *Cell Host Microbe* 2020;**28**:245–257.e6. <https://doi.org/10.1016/j.chom.2020.05.013>
- Cohen J, Shuval HI. Coliforms, fecal coliforms, and fecal streptococci as indicators of water pollution. *Water Air Soil Pollut* 1973;**2**:85–95. <https://doi.org/10.1007/BF00572392>

30. Sabar MA, Honda R, Haramoto E. CrAssphage as an indicator of human-fecal contamination in water environment and virus reduction in wastewater treatment. *Water Res* 2022;**221**:118827. <https://doi.org/10.1016/j.watres.2022.118827>
31. Madden GR, Rigo I, Boone R et al. Novel biomarkers, including PCR cycle threshold, for predicting recurrent *Clostridioides difficile* infection. *Infect Immun* 2023;**91**:e0009223. <https://doi.org/10.1128/iai.00092-23>
32. Hoffmann AA, Willi Y. Detecting genetic responses to environmental change. *Nat Rev Genet* 2008;**9**:421–32. <https://doi.org/10.1038/nrg2339>
33. Ostrow JD. *Bile Pigments and Jaundice: Molecular, Metabolic, and Medical Aspects*. Marcel Dekker Inc, New York, USA. 1986, 4.
34. Ungerfeld EM. Metabolic hydrogen flows in rumen fermentation: principles and possibilities of interventions. *Front Microbiol* 2020;**11**:589. <https://doi.org/10.3389/fmicb.2020.00589>
35. Campbell A, Gdanetz K, Schmidt AW et al. H generated by fermentation in the human gut microbiome influences metabolism and competitive fitness of gut butyrate producers. *Microbiome* 2023;**11**:133. <https://doi.org/10.1186/s40168-023-01565-3>
36. Parks DH, Chuvochina M, Rinke C et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;**50**:D785–94. <https://doi.org/10.1093/nar/gkab776>
37. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform* 2009;**23**:205–11.
38. Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 2014;**1079**:105–16. [https://doi.org/10.1007/978-1-62703-646-7\\_6](https://doi.org/10.1007/978-1-62703-646-7_6)
39. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>
40. Nguyen L-T, Schmidt HA, von Haeseler A et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74. <https://doi.org/10.1093/molbev/msu300>
41. Kalyaanamoorthy S, Minh BQ, Wong TKF et al. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–9. <https://doi.org/10.1038/nmeth.4285>
42. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;**23**:127–8. <https://doi.org/10.1093/bioinformatics/btl529>
43. Cantalapiedra CP, Hernández-Plaza A, Letunic I et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;**38**:5825–9. <https://doi.org/10.1093/molbev/msab293>
44. Jumper J, Evans R, Pritzel A et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. <https://doi.org/10.1038/s41586-021-03819-2>
45. Lin Z, Akin H, Rao R et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123–30. <https://doi.org/10.1126/science.ade2574>
46. Chen EA, Porter LL. SSDraw: software for generating comparative protein secondary structure diagrams. *Protein Sci* 2023;**32**:e4836. <https://doi.org/10.1002/pro.4836>
47. Krueger F, James F, Ewels P et al. FelixKrueger/TrimGalore: v0.6.10. 2023. [Zenodo. https://zenodo.org/records/7598955](https://zenodo.org/records/7598955)
48. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**:357–9. <https://doi.org/10.1038/nmeth.1923>
49. Danecek P, Bonfield JK, Liddle J et al. Twelve years of SAM-tools and BCFtools. *Gigascience* 2021;**10**:giab008. <https://doi.org/10.1093/gigascience/giab008>
50. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 2009. Springer New York. **1**, <https://doi.org/10.1007/978-0-387-98141-3>.
51. Borges R, Machado JP, Gomes C et al. Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics* 2019;**35**:1862–9. <https://doi.org/10.1093/bioinformatics/bty800>