Abstract

Title of Dissertation: HEALTH-RELEVANT FUNCTIONS OF THE GUT MICROBIOME

Sophia Columbus Sery Levy
Doctor of Philosophy, 2025

Dissertation Directed by: Brantley Hall, PhD
Department of Cell Biology and Molecular Genetics
Center for Bioinformatics and Computational Biology

The gut microbiome serves vital functions in host nutrition, immunity, and metabolism. Perturbations in the gut microbiome, commonly referred to as dysbiosis, have been correlated to a wide range of diseases. However, the mechanistic basis of these relationships are poorly understood. Here, I elucidated the enzymatic mechanisms underlying three important, health-relevant functions of the gut microbiome using our lab's integrated approach, combining experimental techniques with bioinformatic analysis.

In Chapter 2, I identify the novel bilirubin reductase BilR using a fluorescence assay to measure bilirubin reduction to urobilinogen. We delineate BilR from similar reductases via my identification and mutation of key residues to find that BilR is primarily encoded by Bacillota species. Our analysis of human gut metagenomes revealed that BilR is ubiquitous in healthy adults, but shows decreased prevalence in neonates and individuals with Inflammatory Bowel Disease (IBD). Our discovery sheds light on the role of the gut microbiome in bilirubin metabolism and highlights the significance of the gut-liver axis in maintaining healthy serum bilirubin homeostasis.

In Chapter 3, I demonstrate that disparate taxa can recapitulate the metabolism of oxidized sugars utilizing enzymatically divergent, yet functionally equivalent *gud/gar* pathways. We identify novel enzymes in the divergent commensal pathway, including putative 5-KDG aldolase GudL and an uncharacterized ABC transporter GarABC that can recapitulate the function of their nonhomologous pathogen counterparts. Our findings reveal a phenomenon parallel to the pathogen growth promotion from inflammation-derived nitric oxide: oxidized sugars, which are also produced during inflammation, serve as alternative carbon sources for

commensal microbes and may contribute to their increased relative abundance during gut inflammation.

In Chapter 4, I discover SpiR, a microbial steroid $\Delta^{5\text{-}4}$ isomerase/3-keto reductase responsible for cholesterol conversion to coprostanone in the gut. We verify that SpiR catalyzes the stereospecific oxidation of both cholesterol and pregnenolone, but that it preferentially binds to cholesterol over other related steroids. Phylogenetic analysis revealed SpiR is restricted to an uncultured Actualibacteraceae clade, and multi-omics analysis shows SpiR is not only enriched in cholesterol-converting individuals, but is a predictor of cholesterol conversion. Our findings refine the enzymatic model of gut cholesterol metabolism and establish *spiR* as a driver for microbial cholesterol reduction in the gut.

HEALTH-RELEVANT FUNCTIONS OF THE GUT MICROBIOME

by

Sophia Columbus Sery Levy

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of Doctor of Philosophy
2025

Advisory Committee:
Professor Brantley Hall, Chair
Professor Amy Karlsson, Dean's Representative
Professor Steven Jay
Professor Najib El-Sayed
Professor Brian Pierce

Preface

These discoveries, methods, and results have either been published in peer reviewed journals, or are currently under review. At the time of this writing, Chapter 2 has been published in *Nature Microbiology,* Chapter 3 has been published in *Nature Communications*, and Chapter 4 is under review at *Nature Microbiology.* Each chapter has been reformatted here. I am deeply grateful to each of my co-authors. Their expertise and commitment were instrumental in the success of my research.

- **Chapter 2**

  Brantley Hall, **Sophia Levy**, Keith Dufault-Thompson, Gabriela Arp, Aoshu Zhong, Glory Minabou Ndjite, Ashley Weiss, Domenick Braccia, Conor Jenkins, Maggie R. Grant, Stephanie Abeysinghe, Yiyan Yang, Madison D. Jermain, Chih Hao Wu, Bing Ma, and Xiaofang Jiang. "BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen." *Nature Microbiology* 9 (2024). My contributions include (1) method development, (2) experimental design, (3) performing experiments, and (4) writing and editing the manuscript.

- **Chapter 3**

  Sophia Levy, Angela K. Jiang, Maggie R. Grant, Gabriela Arp, Glory Minabou Ndjite, Xiaofang Jiang, and Brantley Hall. "Convergent evolution of oxidized sugar metabolism in commensal and pathogenic microbes in the inflamed gut." *Nature Communications* 16 (2025). My contributions include (1) conceptualization of the project, (2) method development, (3) experimental design, (4) performing experiments, (5) data analysis, and (6) writing and editing the manuscript.

- **Chapter 4**

  Sophia Levy, Gabriela Arp, Angela K. Jiang, Keith Dufault-Thompson, Aoshu Zhong, Maggie Grant, Yue Li, Xiaofang Jiang, and Brantley Hall. "SpiR is a microbial

enzyme that drives cholesterol conversion." *Manuscript under review.* (2025). My contributions include (1) method development, (2) experimental design, (3) performing experiments, and (4) writing and editing the manuscript.

Below are other publications I have contributed to during my PhD:

- Gabriela Arp, Angela Jiang, Keith Dufault-Thompson, **Sophia Levy**, Aodhu Zhong, Jyotsna Talreja Wassan, Maggie Grant, Yue Li, Brantley Hall, and Xiaofang Jiang. "Gut bacteria encode reductases that biotransform steroid hormones." *Manuscript under review* (2025).
- Gabriela Arp, **Sophia Levy**, and Brantley Hall. "Bacterial androgen production and prostate cancer." *Nature Microbiology* 10 (2025).
- Keith Dufault-Thompson, **Sophia Levy**, Brantley Hall, and Xiaofang Jiang. "Bilirubin reductase shows host-specific associations in animal large intestines." *The ISMA Journal* 18 (2024).
- Domenick J. Braccia, Glory Minabou Ndjite, Ashley Weiss, **Sophia Levy**, Stephanie Abeysinghe, Xiaofang Jiang, Mihai Pop, and Brantley Hall. "Microbiome-wide search for bacterial azoreductases reveals potentially uncharacterized azoreductases encoded in the human gut microbiome." *Drug Metabolism and Disposition* 51 (2023).

Acknowledgements

I would like to express my gratitude to my advisor, Dr. Brantley Hall, for his unwavering support and mentorship throughout my research journey. His guidance and advice was invaluable in the completion of this thesis and my growth as a scientist and individual.

I am also deeply grateful to my committee - Dr. Steven Jay, Dr. Najib El-Sayed, and Dr. Brian Pierce - for their time and insightful feedback on my work.

A special thank you to Dr. Amy Karlsson, whose support during my undergraduate years made a lasting impact and who, years later, continued to encourage my growth by serving as the Dean's Representative on my committee.

To my family, thank you for your love, encouragement, and confidence in me, even when I was disheartened by a lack of progress or negative results. Josh, you listened to me talk through hours of scientific work that you didn't understand, but you paid attention anyway. Thank you.

To Peter, I could not have done this without you. You provided me with so much support throughout my doctoral journey, from hot tea and dinner when I got home from the lab to pep talks when I threw my hands up and said I couldn't do this.

A final thank you to my beloved Mia.

Table of Contents

List of Figures

List of Abbreviations

IBD	**I**nflammatory **B**owel **D**isease
EC	**E**nzyme **C**ommission
RMSD	**R**oot **M**ean **S**quare **D**eviation
LC-MS/MS	**L**iquid **C**hromatography with tandem **M**ass **S**pectrometry
GTDB	**G**enome **T**axonomy **D**ata**b**ase
NOS	**N**itric **O**xide **S**ynthase
IPTG	**I**so**p**ropyl β-D-1-**T**hio**g**alactopyranoside
HMP2	**H**uman **M**icrobiome **P**roject **2**
UC	**U**lcerative **C**olitis
CD	**C**rohn's **D**isease
MGX	**M**eta**g**enomics
MTX	**M**eta**t**ranscriptomics

Chapter 1: Introduction

The gut microbiome serves vital functions in human health, and is often referred to as the 'metabolic organ'.[1] Sequencing technology allowed us to advance our understanding of the gut microbiome from a single collection of commensal microbes to a dynamic community.[2,3] Population studies have shown strong variation in microbiome composition between individuals - regardless of age, sex, and lifestyle - making it difficult to draw strong conclusions based solely on the overall composition of the microbiota.[4] Recently, the microbiome field has shifted from cataloging microbial taxa via sequencing to investigating their function, with an emphasis on the impact of microbial metabolism on host physiology. [5,6] Instead of relying solely on the composition of the microbiome, we now strive to resolve microbial activity at the functional level, mapping those functions to the responsible enzymes and genes, and categorize the microbial community based on potential function rather than species.

1.1 Functions of the gut microbiome

Gut microbes metabolize nutrition in the diet, train, regulate, and mature the immune system, prevent pathogen growth, transform the epithelial barrier, and impact the nervous system. Previous work has highlighted the role of the gut microbiota in maintaining health across virtually every system in the body.[7–9]

The gut microbiome is responsible for the catabolism of many host-inaccessible components of our diet. Indigestible polysaccharides are fermented into short chain fatty acids such propionate, butyrate, and acetate, which feed colonocytes and modulate inflammation.

Our microbiota also contribute to immune development and regulation. *Bacteroides fragilis* educates the early immune system through production of polysaccharide A, which modulates the ratio of Th1:Th2, promoting self-tolerance and repressing inflammatory cytokines.[10] Aberrant immune responses to microbial perturbations have been implicated in the development of various autoimmune diseases. Through its very existence, the microbiota contribute to pathogen colonization resistance, utilizing resources necessary for pathogens to colonize and releasing antibacterials.

The microbiome supports gut barrier integrity, playing a key role in maintaining tight junctions. Probiotics containing *Lactobacillus farciminis* and *Bifidobacterium longum* have been

shown to promote epithelial integrity and suppress stress signaling via the hypothalamus-pituitary-adrenal axis.[11]

Finally, the microbiome is thought to be a vital component of the poorly understood gut-brain axis. This bidirectional communication network is implicated in disorders ranging from inflammatory bowel syndrome to neurological diseases, potentially via serotonin signaling and modulation of gut motility.[12]

The gut microbiome serves as a dynamic, metabolically active 'organ' that is essential to human health. A mechanistic understanding of these processes will be critically important in uncovering microbiome-based strategies for disease prevention and therapeutic intervention.

1.2 Microbial transformation of host-derived compounds

The classic influence of the gut microbiome on human health is via transformation of host-derived compounds.[13] Whether through metabolism of indigestible polysaccharides to short chain fatty acids, deconjugation of steroids and other substrates, or dehydroxylation of primary bile acids to secondary bile acids, the microbiota are responsible for a number of unique metabolic pathways.[14–16] These biotransformations can have profound impacts on the availability and impact of host metabolites.

One such example is the impact of the microbiome on bilirubin homeostasis. Bilirubin is a byproduct of heme degradation, and is metabolized to bilirubin diglucuronide by the host before being secreted into the gut. There, both host and microbial beta-glucuronidases deconjugate the bilirubin. Without further bacterial intervention, the bilirubin is reabsorbed, contributing to high serum bilirubin levels and hyperbilirubinemia. However, bilirubin-reducing bacteria can further metabolize deconjugated bilirubin to urobilinogen or stercobilinogen, which is not resabsorbable and can be safely excreted as urobilin or stercobilin.[17]

Through a similar mechanism, the microbiome metabolizes cholesterol that enters the gut, which can be reabsorbed into serum and contribute to hypercholesterolemia, to coprostanol, which is non-absorbable and excreted in waste.

In both of these instances, the gut microbiome converts a bioactive molecule that contributes to disease to a less bioactive molecule that can be easily removed from the system. Beyond bilirubin and cholesterol, microbial enzymes modify a wide range of host derived metabolites, impacting host signaling, immune activation, and overall health.

1.3 Inflammation-related perturbations of the microbiome

Though the gut microbiome is quite resilient, various host and environmental factors can lead to disruption. Perturbations in the gut microbiome, commonly referred to as dysbiosis, have been correlated to a wide range of diseases, including Inflammatory Bowel Disease (IBD).[18] Previous studies have primarily relied on species-level resolution to associate dysbiosis with disease. However, we suggest that changes in community functional potential may better explain these disease associations, as multiple taxa may occupy the same functional niche.

Inflammation profoundly impacts the gut environment and microbiota.[19] Interestingly, it remains unclear whether inflammation changes the composition of the microbiome, or if the changes to the microbiome cause host inflammation. Current dogma includes a circular feedback mechanism, where either the host environment or the microbiome is impacted first, and then causes the subsequent changes to the latter, as exemplified by β-oxidation of butyrate in colonocytes. Butyrate is the preferred energy source for colonocytes, and the majority of butyrate in the gut is produced by obligate anaerobic commensal bacteria.[20] When the obligate anaerobes producing butyrate are plentiful, butyrate activates PPAR-γ in colonocytes, promoting mitochondrial β-oxidation of the butyrate. β-oxidation consumes significant amounts of oxygen, therefore contributing to the anoxic environment in the gut lumen. However, if butyrate is low, the lack of PPAR-γ signaling shifts the colonocyte metabolism towards anaerobic glycolysis, which consumes very little oxygen. Since the oxygen diffusing from the blood vessels is no longer being consumed by the colonocytes, this results in a loss of hypoxia in the gut lumen.[21] This feedback loop can be triggered in either direction - the gut microbiome changes composition so there are fewer butyrate-producing obligate anaerobes, resulting in an oxidizing, inflammatory environment, or the host immune response creates an oxidizing environment, killing the butyrate-producing obligate anaerobes. Thus, either inflammation can cause changes to the microbiome, or changes to the microbiome can cause inflammation.

IBD is often used as a model for gut inflammation and microbiome related disease.[22] IBD consists of two subsets of chronic inflammatory gut disease: ulcerative colitis (UC), which is characterized by continuous inflammation spreading from the rectum, and Crohn's Disease (CD), which presents as discontinuous inflammation, often in the terminal ileum.[23] Though studies on microbial perturbations during IBD often offer contradictory results, Schirmer et al. compiled a meta-analysis of four studies of microbiota changes during IBD inflammation to

examine which species are increased in relative abundance across multiple studies.[22] Shifts in taxonomic composition included increases in the relative abundance of *Escherichia, Veillonella, Klebsiella, Blautia,* and *Lachnoclostridium* spp.[24] Some of these increases can be attributed to environmental changes caused by the host inflammatory response.

As part of the NF-κB host inflammatory response, inducible nitric oxide synthase (iNOS) and dual oxidase (DUOX2) are induced and highly expressed.[25] This leads to high concentrations of nitric oxide, superoxides, and related radicals in the gut lumen, resulting in an oxidizing gut environment. The impact of the newly oxidizing gut environment on facultative anaerobes such as Enterobacteriaceae has been well studied.[25,26] These species can utilize the nitrates and other N-oxides produced as terminal electron acceptors, gaining a competitive advantage over the obligate anaerobes unable to take advantage of the inflamed environment and contributing to the characteristic dysbiosis of inflammation.

Alongside the alternative terminal electron acceptors provided by the oxidizing environment, the high concentration of nitric oxide and superoxides also contribute to host-mediated sugar oxidation. As the nitric oxide diffuses through the colonic epithelial barrier to the lumen, it spontaneously oxidizes simple sugars found in the cell membranes, turning glucose to glucaric acid (glucarate or saccharic acid) and galactose to galactaric acid (galactarate or mucic acid).[27] This creates an abundance of oxidized sugars in the inflamed gut, providing a new carbon source for microbes able to metabolize them. Previous work demonstrated that *Salmonella enterica* serovar Typhimurium was able to utilize these oxidized sugars to gain a competitive advantage in the inflamed gut, leading to bacterial blooms.[27]

Though most previous work on the competitive advantages conferred by oxidized sugar metabolism was done on pathogens, it would follow that there are commensal species that have also adapted to be able to utilize this alternative carbon source available during inflammation to gain a similar competitive advantage.

Chapter 2: BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen

This chapter contains material previously published in Nature Microbiology in BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen,[17] *which was a joint work with Brantley Hall, Keith Dufault-Thompson, Gabriela Arp, Aoshu Zhong, Glory Minabou Ndjite, Ashley Weiss, Domenick Braccia, Conor Jenkins, Maggie R. Grant, Stephanie Abeysinghe, Yiyan Yang,*

Madison D. Jermaine, Chih Hao Wu, Bing Ma, and Xiaofang Jiang. S.L. assisted in optimizing the fluorescence assay for bacterial media, performed bacterial screening, created the gene constructs and mutants, and contributed to writing the manuscript.

B.H. and X.J. conceptualized and supervised the project. S.L., B.H., X.J., K.D.-T., G.M.N., A.Z., G.A., A.W., D.B., C.J., S.A.,Y.Y., M.D.J., C.H.W., M.R.G. and B.M. performed the experiments and/or analyses. S.L., B.H., X.J., and K.D.-T. wrote the paper.

2.1 Introduction

Bilirubin, an intermediate of the heme degradation pathway, plays a critical role in human physiology through the gut-liver axis[28]. Alongside other organic molecules in bile such as cholesterol and bile acids, bilirubin diglucuronide (conjugated bilirubin) is secreted into the gut where it is either excreted or reabsorbed. When bilirubin diglucuronide is deconjugated by human or bacterial beta-glucuronidases into unconjugated bilirubin, it can be readily reabsorbed into enterohepatic circulation or further metabolized via reduction reactions by gut microbes into the more excretable metabolites urobilinogen and stercobilinogen[29]. Bilirubin reabsorption elevates serum bilirubin levels, while excretion as urobilinogen and stercobilinogen in stool and urine facilitates its clearance, completing the heme degradation pathway[30–32] (Figure 2.1a).

Dysregulation of gut microbial bilirubin reduction affects serum bilirubin levels, which can have significant health implications. In moderate concentrations, bilirubin serves as an important antioxidant with potential health benefits[33,34]. However, elevated serum bilirubin concentrations can become toxic, leading to jaundice and in extreme cases, kernicterus, a type of bilirubin-induced neurological damage[35]. Similarly, urobilinogen can be reabsorbed and has been associated with multiple diseases, highlighting the key role of bilirubin reduction in the homeostasis of multiple metabolites[30,31,36–38]. While the importance of bilirubin enterohepatic

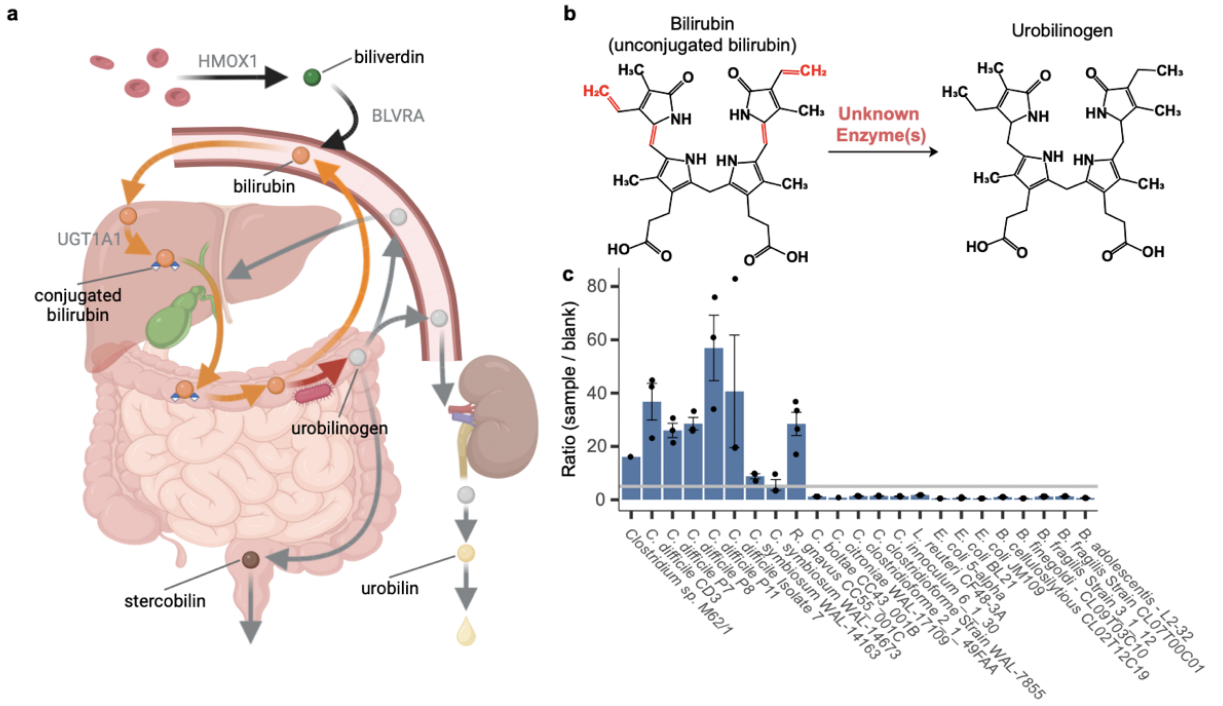

Figure 2.1 Identification of bilirubin reducing bacterial strains. **a)** Illustrated representation of the heme degradation pathway. Key human enzymes are labeled with gray text. **b)** Diagram of the structures of bilirubin and urobilinogen. The bonds reduced during bilirubin reduction are shown in red. **c)** Results of fluorescence assay screening of bacterial strains. Measurements from n=3 independent biological replicates are shown as black points. Bars show the ratios of the samples' fluorescence to a corresponding abiotic media sample with bilirubin added. Error bars indicate one standard error above and below the mean values. The gray line marks a ratio of 5, above which the sample was considered to be positive for bilirubin reduction. *Clostridium sp.* M62/1 and *C. citroniae* WAL-17108 are represented by single data points.

circulation to serum bilirubin levels was suggested in the 1960s and 1970s, it was largely undervalued until the work of Vitek *et al.* in the early 2000s, which showed the direct impact of microbial bilirubin metabolism on serum bilirubin levels in rats[39–42].

Gut microbes are solely responsible for the reduction of bilirubin to urobilinogen, potentially utilizing the molecule as a terminal electron acceptor in anaerobic respiration. Despite the recognized role of gut microbes in bilirubin reduction, the gut microbial enzyme that reduces bilirubin to urobilinogen, hereafter called bilirubin reductase, has remained undiscovered[43]. Multiple bilirubin reducing bacteria have been identified, including strains of *Clostridioides difficile*[40], *Clostridium ramosum*[44], *Clostridium perfringens*[40], and *Bacteroides fragilis*[45], but much of this work was performed before genome sequencing was widely available, making it difficult to know which strains were studied. Without knowledge of the gene encoding bilirubin reductase, it is difficult to draw conclusions about how gut microbial metabolism affects serum

bilirubin homeostasis, leaving a significant gap in our understanding of the heme degradation pathway.

In this study, we aimed to identify the enzyme(s) responsible for bilirubin reduction to urobilinogen and apply this knowledge to understand the relationship between microbial bilirubin reduction and human health. Combining experimental screening of common gut bacteria, comparative genomics and biochemical inference, we were able to identify a candidate bilirubin reductase gene we named *bilR* and confirm its bilirubin reductase activity through fluorescence assays and metabolomics. When analyzing human gut metagenomes, we found that bilirubin reductase was nearly universally present in healthy adults, while the prevalence of the gene was much lower in patients with Inflammatory Bowel Disease (IBD) and in infants, especially during the first few months of life when infants are most susceptible to developing jaundice. This study fills a long-standing knowledge gap in the heme degradation pathway and provides a building block to understand disorders of serum bilirubin homeostasis like jaundice.

2.2 Results

2.2.1 Identification of a putative bilirubin reductase

Our goal was to pinpoint the gene encoding bilirubin reductase through experimental screening and comparative genomics. The bilirubin reductase enzyme is likely an oxidoreductase that acts on carbon-carbon double bonds (EC:1.3.-,-)(Figure 2.1b). Additionally, this gene should be present in the genomes of species that can reduce bilirubin and absent in those that cannot. By examining closely related species with different abilities to reduce bilirubin, we aimed to identify a specific gene responsible for this activity.

To identify gut microbial species that were putatively capable of bilirubin reduction, we employed a fluorescence assay on extracts from bacteria grown in media with bilirubin. The premise of the assay is that the unstable potential products of bilirubin reduction, urobilinogen and stercobilinogen, can be readily oxidized to the more stable urobilin and stercobilin with the addition of iodine and measured via fluorescence, while non-protein bound bilirubin is not fluorescent[46]. As a positive control, we demonstrated that the known bilirubin reducer, *C. difficile* CD3, returned a positive result in the fluorescence assay and that no fluorescence was produced when no bilirubin was added to the culture (Figure 2.1c).[47]

To search for other potential bilirubin reducing species, we grew various bacteria from the major phyla of the human gut microbiome, prioritizing species previously reported to reduce bilirubin and their taxonomic relatives, in media supplemented with bilirubin and assayed for urobilin with fluorescence. Using our fluorescence assay we identified putative bilirubin reduction activity in nine strains, including three species that were previously not known to be capable of bilirubin reduction: *Clostridium symbiosum* (strains WAL-14163 and WAL-14673), *Clostridium sp.* M62/1, and *Ruminococcus gnavus* CC55_001C (Figure 2.1) Additionally, 13 gut species did not return a positive fluorescence result (Figure 2.1c). Of the tested species, all bilirubin reducers were found to be in the class Clostridia of the phylum Firmicutes. However, close relatives of these reducers, such as *Clostridium clostridioforme* 2_1_49FAA and *Clostridium bolteae* CC43_001B, were not found to be reducers.

We leveraged the variability in bilirubin reduction between closely related strains to identify candidate bilirubin reductase genes through a comparative genomics analysis. Our analysis focused on the genomes of five putative bilirubin reducing strains and five closely related strains that did not show any bilirubin reduction activity in our fluorescence assay. Among the ten genomes analyzed, a total of 6,256 orthogroups were identified, of which 389 were predicted to be putative oxidoreductases. Only two of these orthogroups fit the same pattern of presence and absence in the strains as the bilirubin reduction phenotype. One of them was predicted to be a 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase (EC: 1.17.1.2) which is an enzyme that acts on CH or CH2 groups and is involved in isoprenoid biosynthesis which did not meet the proposed requirements for a bilirubin reductase enzyme[48]. The other reductase orthogroup was an unannotated enzyme found to be homologous to 2,4-Dienoyl CoA reductase (EC:1.3.1.34), an oxidoreductase that reduces carbon-carbon double bonds, similar to the expected bilirubin reduction reaction.

We then analyzed the operons that contained the putative bilirubin reductases. Three versions of a putative bilirubin reductase operon were identified consisting of different combinations of three genes hereafter referred to as *bilQ*, *bilR*, and *bilS* (Figure 2.2). *bilR* is the putative bilirubin reductase, *bilS* is a flavodoxin-like protein, and *bilQ* is a MarR family transcriptional regulator. *C. difficile* CD3 and both strains of *C. symbiosum* contained versions of

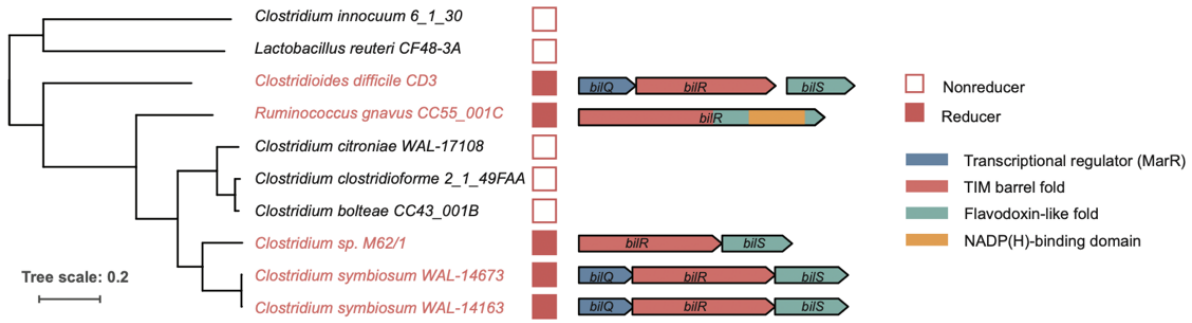

Figure 2.2 Putative bilirubin reductase operons. The phylogenetic tree shows the relationship between five bilirubin reducers and five non-reducers. Genes are represented as arrows. Genes are colored to show predicted domains.

the operon that included all three genes. *Clostridium sp.* M62/1 had only *bilR* and *bilS* and *R. gnavus* CC55_001C had just *bilR*. The *R. gnavus* CC55_001C *bilR* differs in that it encodes two extra C-terminal domains in comparison to the other four *bilR* genes. The N-terminal domain of the *R. gnavus* CC55_001C BilR protein is predicted to be a TIM barrel fold, and it is homologous to the full protein sequences of other *bilR* genes. The two extra domains of the *R. gnavus* CC55_001C BilR are a flavodoxin-like domain and an NADP(H)-binding domain. There was no sequence homology detected between the flavodoxin-like domain encoded by the *R. gnavus* CC55_001C *bilR* and the flavodoxin-like protein encoded by *bilS*, although they are predicted to have the same fold and are likely involved in electron transfer. The BilR of *R. gnavus* CC55_001C was predicted to belong to the Old Yellow Enzyme family (COG1902) and exhibits clear homology to *E. coli* 2,4-Dienoyl CoA reductase (PDB:1PS9)(protein identity=29.82%, RMSD of predicted BilR and 2,4-Dienoyl CoA reductase =2.79, normalized TM-score 0.90), an enzyme that catalyzes the reduction of both 2-trans,4-cis and 2-trans,4-trans-dienoyl-CoA thioesters[49]. This homology supports the identification of BilR as a putative bilirubin reductase, as bilirubin reductase is hypothesized to be an oxidoreductase that can act on the multiple CH-CH groups of bilirubin.

2.2.2 BilR confers bilirubin reductase activity

Once BilR was identified as a candidate bilirubin reductase, we sought to experimentally validate that it is sufficient to confer bilirubin reductase activity through heterologous expression in *E. coli*. The *bilRS* genes from *C. symbiosum* and *C. difficile* (short *bilRS*), as well as *bilR* from

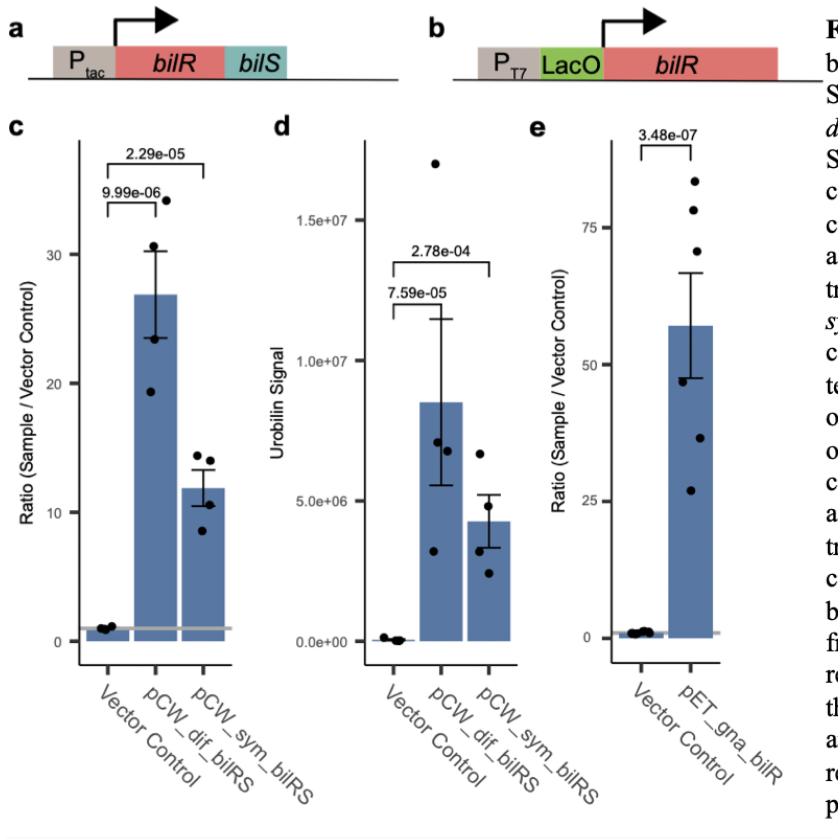

Figure 2.3 Confirmation of *bilR(S)* bilirubin reductase activity. **a)** Schematic of *C. symbiosum* and *C. difficile bilRS* construct. **b)** Schematic of *R. gnavus bilR* construct. **c)** Fluorescence assay comparing bilirubin reduction activities of *E. coli* 10-beta transformed with empty vector, *C. symbiosum* construct, and *C. difficile* construct (two-sample one-sided t-test). **d)** Metabolomics confirmation of urobilin production (two-sample one-sided t-test). **e)** Fluorescence confirmation of bilirubin reduction activity of *E. coli* 10-beta transformed with the *R. gnavus bilR* construct (two-sample t-test). The bar heights show the mean values from n=4 independent biological replicates that were used to generate the plots and statistics in panels c and d. N=3 independent biological replicates were used to generate the plots and statistics for panel e.

R. gnavus (long *bilR*), were separately cloned into a pCW-lic vector and then transformed into *E. coli* 10-beta or into a pET-28a(+) vector and transformed into *E. coli* T7 express lysY/Iq (Figure 2.3a,b). To validate that the fluorescence assay could be used a proxy for bilirubin reduction to urobilinogen we assayed for the reduction of bilirubin to urobilinogen via fluorescence assay and verified the results with LC MS/MS (Figure 2.3c,d). The strains transformed with the *C. difficile* and *C. symbiosum* BilRS both demonstrated significant fluorescence after being incubated with bilirubin and the presence of urobilin was observed in the corresponding LC-MS samples (fluorescence assay: *C. difficile* BilRS *P*=9.99e-06, *C. symbiosum* BilRS P=2.29e-05, two sample t-test on log2 transformed data; metabolomics: *C. difficile* P=7.59e-05, *C. symbiosum* P=2.78e-04, two sample t-test on log2 transformed data (urobilin m/z 591.4→m/z 343.2)). The samples did not have any peaks indicating that stercobilin was present, suggesting that the assay could be used as a reliable indication of bilirubin reduction to urobilinogen. The long BilR construct was also capable of reducing bilirubin to urobilinogen (fluorescence assay: *R. gnavus* BilR *P*=3.48e-07, two sample t-test on log2 transformed data) (Figure 2.3e), and *E. coli*

with the empty vector backbone pCW-lic did not reduce bilirubin (P=0.167, one-sided t-test on log2 transformed data against abiotic media control). Neither mesobilirubin (m/z 589.3→m/z 301.2) nor stercobilin (m/z 595.4→m/z 345.2) were consistently detected. Native bilirubin reducers and *E. coli* expressing *bilRS* were also able to reduce mesobilirubin to urobilinogen suggesting that the enzyme can act on all four double bonds in bilirubin. These results support that BilR is sufficient to reduce bilirubin or mesobilirubin to urobilinogen.

2.2.3 Delineation of the BilR clade

To delineate the BilR sequences from other members of the Old Yellow Enzyme family, a combination of structural prediction and sequence conservation was used to identify key

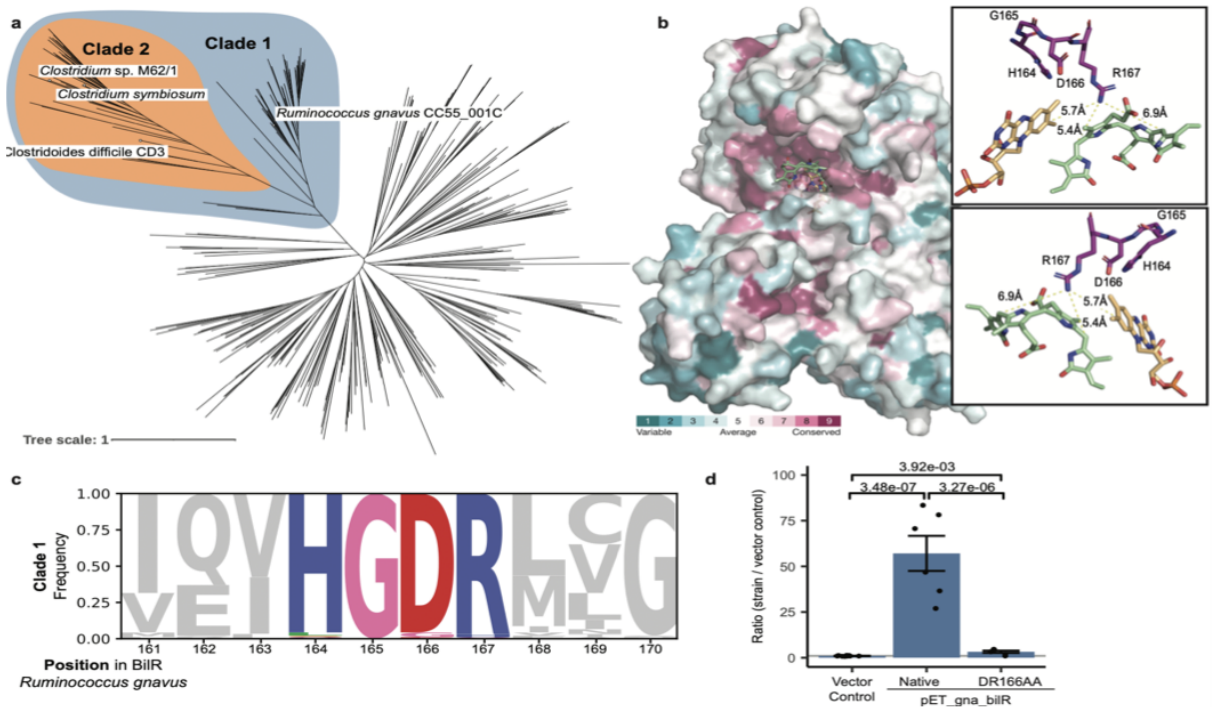

Figure 2.4 Identification of a *bilR* clade. **a)** Gene tree constructed from putative *bilR* sequences and related reductases from the Old Yellow Enzyme family. Experimentally confirmed bilirubin reducers are labeled in the tree. Clade 1 indicates the overall *bilR* clade, while Clade 2 indicates the short *bilR* clade. The same tree with all bootstrap values and species labels is included in Extended Data Fig.8. **b)** AlphaFold predicted structure of the *R. gnavus* BilR with sequences colored to show their degree of conservation within Clade1 based on a ConSurf conservation analysis. A docked bilirubin (green) and FMN molecule (orange), are shown on the structure and the insets show the potential site of interaction between the R167 residue and the putative enzyme substrates. The HGDR residues are colored in purple in the insets. **c)** Diagram showing the conservation of positions within the Clade 1 BilR sequences. The conserved HGDR motif positions are colored. **d)** Fluorescence assay results comparing bilirubin reductase active in *E. coli* 10-beta transformed with a vector control, *R. gnavus bilR*, and *R. gnavus bilR* with the DR residues at position 166-167 mutated to AA.

residues in the putative BilR sequences and differentiate BilR from the other sequences (Figure 2.4a). Our analysis has revealed that one particular clade, referred to as Clade 1, is likely to be the bilirubin reductase clade (Figure 2.4a). The predicted structure of BilR was compared to the structurally homologous *E. coli* 2,4-dienoyl-CoA reductase enzyme, providing an indication of the putative active site in the BilR enzyme. The Y166 residue in the *E. coli* 2,4-dienoyl-CoA reductase is thought to protonate one of the carbon atoms that forms the carbon double bond in the enzyme's substrate completing the reduction of that bond[49,50]. The similar positioning of the positively charged arginine residue in the predicted BilR structure suggests that this R167 residue or the neighboring D166 residue may also be acting as a proton donor, completing the reduction reaction after a hydride transfer to the carbon-carbon double bond from a flavin cofactor of NAD(P)H. The D166 and R167 residues, along with the neighboring H164 and G165 residues were found to be conserved among the BilR genes from the five confirmed bilirubin reducing species. A search for this HGDR motif within the old yellow enzyme family revealed that it was nearly universally conserved within a larger clade of proteins that includes all five bilirubin reductases within it (Figure 2.4b,c). The conservation of these residues within the Clade 1 sequences (Figure 2.4a) combined with the similar positioning of these residues to the active site residues in the *E. coli* 2,4-dienoyl-CoA reductase enzyme, suggests that Clade 1 contains the bilirubin reductase enzymes.

To confirm that Clade 1 is composed of bilirubin reductase enzymes, we performed experiments to investigate the functional impact of the D166 and R167 residues on the enzymes activity. This included mutating the residues in question to D166A and R167A and heterologously expressing the mutated protein to measure the impact on the enzyme's activity. The mutated protein showed similar structural properties as the wild type protein, confirming that the mutation did not lead to a major disruption in the overall folding of the enzyme. When expressed in *E. coli* the mutated protein demonstrated a significantly lower bilirubin reductase activity than the wild type protein (P=3.07e-06, two sample t-test) (Figure 2.4d). This experimental approach provides direct evidence that the D166 and R167 residues in the active site are critical for bilirubin reduction, a characteristic that is specific to bilirubin reductase enzymes and distinct from other members of the Old Yellow Enzyme family, supporting the assignment of Clade 1 as bilirubin reductase (Figure 2.4a).

We then examined the taxonomic distribution and diversity of bilirubin reducers, identifying putative bilirubin reductases in 658 of the representative genomes in the Genome

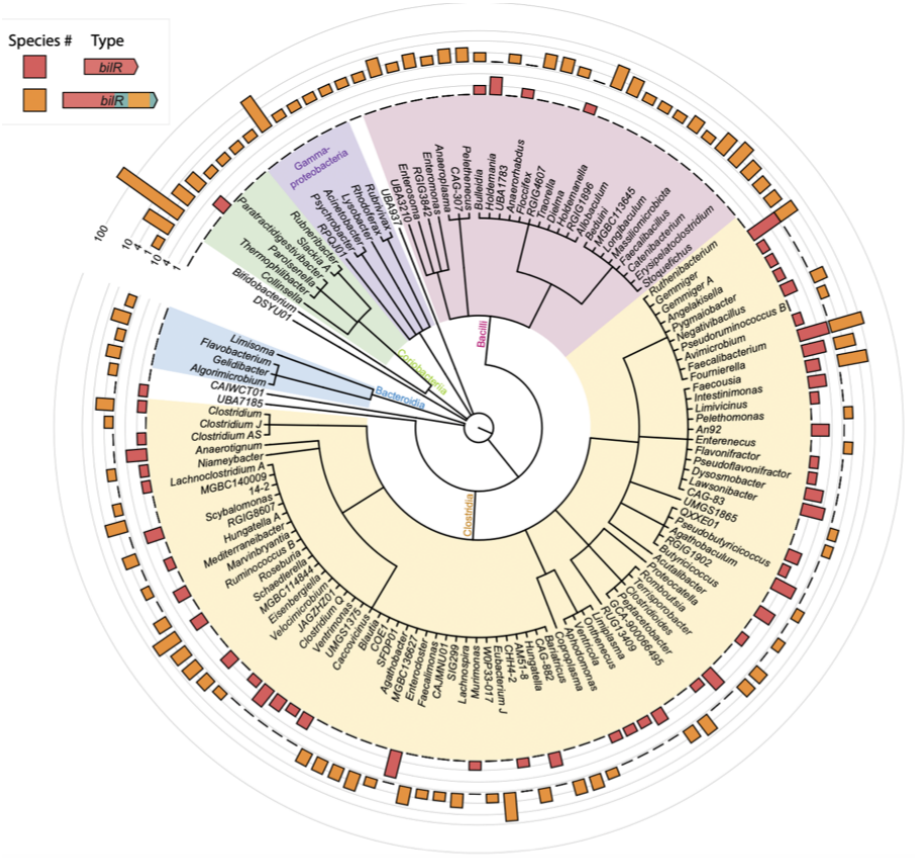

Figure 2.5 Taxonomic distribution of bilirubin reductase. The cladogram shows the relationships between different taxa with detected bilirubin reductase genes. The outer rings show the presence of the short *bilR* gene (red) and long *bilR* gene (orange).

Taxonomy Database (GTDB) (Figure 2.5). The distribution of *bilR* genes is largely in agreement with previously reported bilirubin reducing species with species like *C. ramosum* and *C. difficile* having predicted *bilR* genes. While the representative genome of *C. perfringens* does not have bilirubin reductase, 137 out of 315 genomes assigned to the *C. perfringens* species have it, suggesting that bilirubin reduction is a strain-specific feature. Notably, no species within the *Bacteroidaceae* family were predicted to be bilirubin reducers, conflicting with a previous report from 1972, but this could be due to this being a strain specific feature or incorrect taxonomic assignment[45].

The long variant of the *bilR* gene is widely distributed across multiple phyla, while the short variant seems to be more exclusive to Firmicutes, only being seen in the *Parolsenella* genus outside of Firmicutes. The majority of *bilR* genes were identified in Firmicutes species, including common human gut species such as *Roseburia intestinalis*, *Roseburia inulinivorans*, and *Faecalibacterium prausnitzii*, with a mixture of both *bilR* versions being seen. The short variant of the *bilR* gene is monophyletic (Clade 2 in Figure 2.4a), suggesting that there has been

extensive horizontal gene transfer during the evolution of *bilR*, particularly in Firmicutes. Some *bilR* genes were found in bacteria from the Flavobacteriales, which are typically found in aquatic and soil environments, suggesting a possible role for bilirubin reductase in breaking down bilirubin or similar metabolites in other environments. Nine species in the genus *Bifidobacterium* were predicted to be bilirubin reducers and were isolated from the feces of non-human animals such as chickens, lemurs, and monkeys. This implies that the bilirubin reducing capacity of the human gut microbiota is largely influenced by the abundance of Firmicutes-associated bilirubin reducers.

2.2.4 Bilirubin reductase is often absent in the neonate gut

To examine the relationship of bilirubin reductase to age and health we performed a large-scale analysis of human gut metagenomes. We examined the presence and absence of bilirubin reductase across 4296 infant gut metagenomes from the first year of life. We saw a

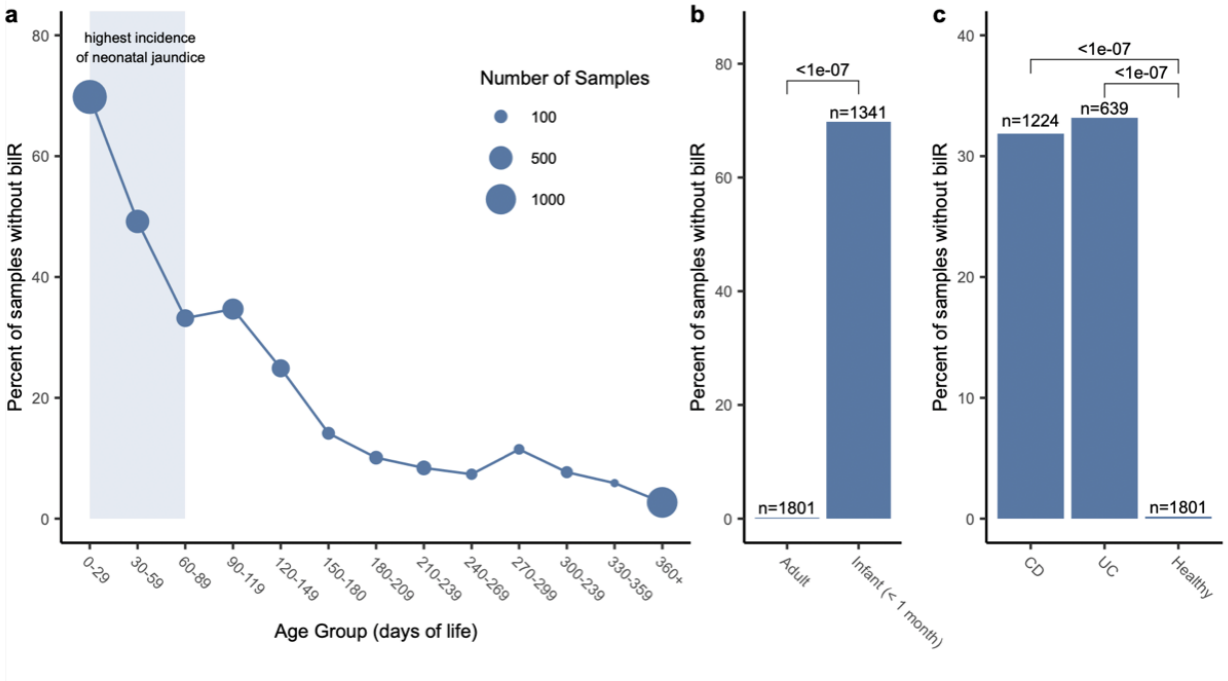

Figure 2.6 Presence of *bilR* in the human gut during development and disease. **a)** Percentage of infant gut metagenomes missing *bilR* during their first year of life. The period of highest jaundice susceptibility is indicated by a shaded blue area on the plot. **b)** Comparison of *bilR* absence in samples from healthy adults and infants in their first month of life. **c)** Comparison of the percentage of samples with no *bilR* detected from healthy adults and adults with IBD. The number of metagenomic samples included in each dataset is indicated above each bar. The P-values for each comparison show the results of a test of equal proportions testing if the fraction of samples with no *bilR* detected was different between groups, without adjusting for multiple testing.

stark trend where bilirubin reductase was often absent in infants during the first few months, when neonatal jaundice risk is highest, but was mostly present by the end of the first year of life only[51] (Figure 2.6a). While various factors are likely to influence the serum bilirubin levels in infants, the correspondence between the period of life where we observe the most samples missing bilirubin reductase and the risk of neonatal jaundice suggests a strong connection between the microbiome composition and the development of jaundice in infants.

2.2.5 Bilirubin reduction is a core function of the human gut

Serum bilirubin levels are significantly altered in IBD patients, suggesting that the bilirubin metabolizing microbiome is significantly disrupted in this disease[52]. When examining metagenomes from healthy adults we found that bilirubin reductase was absent in only 0.1% of the 1801 samples analyzed, a significantly lower fraction of samples ($P<2.2e-16$, test of equal proportions) than was seen in the young infants (Figure 2.6b). Compared to healthy adults IBD patients have been observed to have significantly lower levels of urobilin, and have been observed to form pigmented gallstones containing high levels of bilirubin[53,54]. Bilirubin reductase was absent in a significantly higher fraction of samples from patients with Crohn's disease ($P<2.2e-16$, test of equal proportions) or ulcerative colitis ($P<2.2e-16$, test of equal proportions) compared with healthy samples (Figure 2.6c). This provides evidence that the microbial taxonomic alterations associated with IBD and other inflammatory diseases may be having a direct impact on the bilirubin reduction activity of the gut microbiome.

2.3 Discussion

Despite the identification of urobilin as the yellow pigment in urine more than 125 years ago, the enzyme responsible for its production has remained a mystery[55]. Though it was previously thought that multiple enzymes were involved in the reduction of bilirubin, our results support that a single enzyme performs the reduction of bilirubin to urobilinogen[47]. The identification of bilirubin reductase allowed us to profile its abundance in 7,960 metagenomes showing that bilirubin reduction is a core feature of a healthy adult human microbiome and that neonates are often missing bilirubin reductase during the period of the highest incidence of neonatal jaundice[56,57]. Furthermore, we found that the prevalence of bilirubin reductase is decreased in IBD patients.

Our experiments have shown that bilirubin reductase can perform the complete reduction of bilirubin to urobilinogen and has the potential to act on multiple substrates, placing this single enzyme in a central role in bilirubin homeostasis. The importance of this process is further highlighted by the limited capacity of the UGT1A1 enzyme in the liver to conjugate serum bilirubin, a rate limiting step in heme degradation[58]. Additionally, conjugated bilirubin is thought to be readily deconjugated by gut microbes, and the activity of microbes on conjugated bilirubin has been shown to be limited, suggesting that bilirubin reduction is the key step that determines the degree to which heme degradation byproducts are reabsorbed or excreted[39,47,59,60].

While the discovery of bilirubin reductase illuminates a key gap in our understanding of the health relevant steps of heme excretion, there are additional aspects of the pathway that require elucidation. First, while it is likely that bilirubin is deconjugated by beta-glucuronidases before it is reduced by BilR, additional experiments with bilirubin diglucuronide are necessary to confirm the order of the pathway[47]. Second, there is an additional step in the pathway downstream of bilirubin reduction performed by the gut microbiome where the enzyme is still unknown: the reduction of urobilinogen to stercobilinogen. Similar to bilirubin, urobilinogen may serve as a terminal electron acceptor for microbes encoding these reductases, providing them a competitive advantage in the anaerobic environment of the gut.

With the knowledge of the species, genes, and enzymes involved in bilirubin reduction, future research can now focus on the extent to which gut microbial bilirubin metabolism affects serum bilirubin homeostasis and the role of bilirubin reduction in health and disease. Our finding that bilirubin reductase is often absent in young infants supports the hypothesis that neonatal jaundice is exacerbated by the absence or low abundance of bilirubin reducing microbes in the gut[39,40]. Multiple factors related to both the host and microbiome undoubtedly contribute to the development of jaundice. To address these factors, new cohort studies that simultaneously measure serum bilirubin, fecal urobilinoids, and the absolute abundance of bilirubin reducing bacteria in infants will be needed to disentangle their roles in this disease[61,62]. In addition to neonatal jaundice, bilirubin reduction by gut microbes is important to overall health because serum bilirubin concentrations have been observed to correlate inversely with adiposity and cardiovascular disease[63,64]. Urobilin concentrations also have an impact on human health being positively correlated with insulin resistance and heart failure[37,65]. This further highlights the complexity and importance of the heme degradation pathway, confirming the key role that bilirubin reductase plays in determining the balance of multiple health-relevant metabolites.

The differences observed between the metagenomes of healthy patients and IBD patients highlight the potential disruption of bilirubin metabolism in IBD. The lowered prevalence of bilirubin reducing bacteria in IBD patients leads us to hypothesize that the significant disruption of bilirubin metabolism combined with the previously established increase in unconjugated primary bile acids in IBD patients could contribute to the increased incidence of calcium bilirubinate gallstones that has been observed in these patients[53,66]. While these results suggest a role for bilirubin metabolizing microbes in these diseases, more work is needed before any conclusions can be made. Our work fills a long-standing knowledge gap, providing foundational knowledge that will serve as the basis of future investigations of the importance of bilirubin metabolism in human health.

2.4 Materials and Methods

2.4.1 Culture methods

Anaerobic bacteria: Bacterial strains were obtained from the NIH Biodefense and Emerging Infections Research Resources Repository (BEI). Strains were inoculated from a glycerol stock and grown under anaerobic conditions (90% N_2, 5% CO_2, 5% H_2) at 37℃ in an anaerobic chamber (Coy Laboratory Products). Strains were grown in 200 mL of liquid brain heart infusion broth (BHI, Research Products International (RPI), B11000) or yeast casitone fatty acid broth (YCFA) with 4.4 mg/100 mL sterile filtered bilirubin dissolved in dimethylsulfoxide (DMSO, Sigma-Aldrich). The bacteria were grown anaerobically at 37°C for 24 hours before starting the chloroform extraction.

Transformed E. coli: Transformed *E. coli* strains were inoculated from an agar colony plate into 300 mL BHI media with 100 ug/mL carbenicillin (GoldBio, 00901C103) or 50 ug/mL kanamycin and 100 uM isopropyl β-d-1-thiogalactopyranoside (IPTG, GoldBio, 221105I2481) and shaken aerobically at 37°C for 16 hours. The bacteria were then pelleted by centrifugation at 3260 x g for 6 minutes and moved to the anaerobic chamber. The pellet was resuspended in 200 mL of BHI with 4.4 mg/100 mL sterile filtered bilirubin dissolved in DMSO, 100 ug/mL carbenicillin or 50 ug/mL kanamycin, and 100 uM IPTG to induce expression of the cloned *bilR* gene(s). The strains were incubated anaerobically at 37°C for 24 hours before starting the chloroform extraction.

2.4.2 Measuring bilirubin reduction

Fluorescence Assay: Bacterial cultures were removed from the anaerobic chamber and pelleted by centrifugation at 3260 x g for six minutes. The supernatant was filtered through a non-sterile 0.22 um filter. Five mL of chloroform (Sigma-Aldrich) was added to the filtered supernatant and separated via organic extraction in a separatory funnel. The resultant chloroform-urobilinogen solution was air dried until the chloroform was completely evaporated, producing the extract used in the subsequent fluorescence assays and LC MS/MS.

Because the urobilinogen product of bilirubin reduction is unstable in oxygen, measuring urobilinogen itself was infeasible. Therefore, to measure bilirubin reduction, urobilinogen was oxidized to urobilin via the addition of iodine. The urobilin was quantified by adding zinc acetate solution to form a fluorescent zinc-urobilin complex[67–69]. However, this assay also detects stercobilin, which fluoresces at the same wavelength as urobilin.[70,71] To differentiate the two potential products, the fluorescence assay was verified by mass spectrometry.

The instability of urobilinogen requires this assay be performed immediately after the chloroform from the chloroform extraction had evaporated. The extracts were rehydrated using 600 µL of deionized water. A 400 µL aliquot of the rehydrated solution was transferred to a separate tube to continue the assay. 10 µL of 10% Povidone-iodine solution (CVS, A47194) was added to oxidize the urobilinogen to urobilin, followed by 10 µL of a 100 mM cysteine solution to reduce the rest of the iodine to prevent bilirubin oxidation after addition of Schlesinger's reagent (545 mM zinc acetate in methanol). 400 µL of Schlesinger's reagent was added to increase the fluorescence of urobilin. The resulting solution was portioned into 100 µL triplicates or quadruplets onto a 96 well acrylic plate. Each well's fluorescence was measured by SpectraMax M5 plate reader using a 495/525 nm wavelength at medium gain. Samples that produced a fluorescence signal greater than five times the fluorescence of an abiotic media control were identified as reducers. Samples below the five times threshold were considered to be non-reducers.

LC with tandem mass spectrometry (LC-MS/MS): 1 mL of the chloroform solution from the chloroform extraction was aliquot and allowed to dry completely. The samples were sent with overnight shipping in dry ice to the Duke metabolomics core. 60 uL of 10% Povidone-iodine solution was added to the dried samples. Samples were vortexed, then centrifuged at 15,000 x g

for 5 minutes at 4°C. The supernatant was transferred into the 1.7 mL Macherey-Nagal glass vial for the LC-MS/MS injection.

Samples were analyzed with the 6500+ QTRAP LC–MS/MS system (Sciex, Framingham, MA). The Sciex ExionLC UPLC system includes a degasser, an AD autosampler, an AD column oven, a controller, and an AD pump. The liquid chromatography separation was performed on an Agilent Eclipse Plus C18 RRHD column (2.1 x 50mm, 1.8 μm) with mobile phase A (0.2% formic acid in water) and mobile phase B (0.2% formic acid in acetonitrile). The flow rate was 0.35 mL/min. The linear gradient was as follows: 0–0.5min, 100% A; 4.0–5.5min, 0%A; 5.6–7.5min, 100%A. The autosampler was set at 10°C and the column was kept at 35°C. The injection volume was 5 μL. Mass spectra were acquired under positive electrospray ionization with the ion spray voltage of 5500 V. The source temperature was 450 °C. The curtain gas, ion source gas 1, and ion source gas 2 were 33, 55, and 60 psi, respectively. Multiple reaction monitoring (MRM) was used to detect mesobilirubin (m/z 589.3 →m/z 301.2), urobilin (m/z 591.4→ m/z 343.2) and stercobilin (m/z 595.4→m/z 345.2). A co-injection of *C. symbiosum* incubated with bilirubin and the urobilinogen standard was performed to ensure that the retention times seen in the samples and standards were consistent. All data was analyzed in Software Analyst 1.7.1.

2.4.3 Construct development

pCW-sym-bilRS: To express the *bilRS* genes from *C. symbiosum* in *E. coli,* the gene was amplified and inserted into an expression vector backbone pCW-lic (Addgene, 26908). A polymerase chain reaction (PCR) amplification using OneTaq Master Mix (NEB, M0482) was performed on *C. symbiosum* genomic DNA with forward primer TAAGCACATATGCTGGGAAAGACGAAGGAGGCTC and reverse primer TAAGCAGGTACCGGCAGCGTCCCTGAGT to amplify *C. symbiosum bilRS*. The product was purified with a Monarch PCR and DNA cleanup kit (NEB, T1030) and restricted with enzymes Nde1 (NEB, R0111) and Kpn1-HF (NEB, R3142) using NEBcloner® protocols to prepare for ligation into the pCW-lic backbone. The pCW-lic vector backbone was restricted with the same enzymes. pCW-lic was dephosphorylated with antarctic phosphatase (NEB,

M0289) to prevent self-ligation. The restricted insert and dephosphorylated backbone were ligated with T4 DNA ligase (NEB, M0202) to form the construct.

pCW-dif-bilRS: The same protocol was used as for pCW-sym-bilRS, with two exceptions. PCR amplification was performed on *C. difficile* genomic DNA with forward primer TAAGCAGGATCCCAGCTGTGGAGGAAGAATAGGATG and reverse primer TGCTTAAAGCTTCTACTACACAATACTAGCTTTAATCATCATA to amplify *C. difficile bilRS*. The product was restricted with enzymes BamH1-HF (NEB, R3136) and HindIII-HF (NEB, R3104) before ligating.

pCW-gna-bilR: The same protocol was used as for pCW-sym-bilRS, with two exceptions. PCR amplification was performed on *R. gnavus* genomic DNA with forward primer TAAGCAGGATCCGCAGAAAGAAAATGTTAAGGAGAGGCTG and reverse primer TAAGCAGGTACCGAAGGATGTTTCATCCACCTGTACG to amplify *R. gnavus bilR*. The product was restricted with enzymes BamH1-HF and Kpn1-HF before ligating.

Transformation with pCW constructs: The constructs were individually transformed into NEB 10β competent *E. coli* cells using the manufacturer's protocol (NEB, C3019). The cells were plated on Luria-Bertani (LB) plates with 100 ug/mL carbenicillin to select for successfully transformed colonies. The transformed colonies were verified through Sanger sequencing by Azenta.

pET28-bilR: In an effort to optimize the expression of bilR, a second construct was developed using a pET28a(+) vector backbone. To clone the *bilR* gene from *R. gnavus* CC55_001C into pET28a(+) for protein expression, the insert was amplified from a glycerol stock of *R. gnavus* using a reported protocol[72]. The forward primer used was CGGCAGCCATATGAGATTATTAGAACCAATTAAAG, and the reverse primer was GGCCGCAAGCTTATAAAACGTTTGCTGCC. The vector backbone was amplified from pET28a(+) with the forward primer CGTTTTATAAGCTTGCGGCCGCACTCGAG and the reverse primer ATAATCTCATATGGCTGCCGCGCGGCAC. The resulting insert and vector backbone PCR products were assembled using the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, E2621S) to yield pET28-BilR. The Q5® High-Fidelity 2X Master

Mix (New England Biolabs, M0492L) was used in PCR, and NEB® Turbo Competent *E. coli* (New England Biolabs, C2984H) was used as competent cells in cloning.

pET28-bilR-166AA: Site directed mutagenesis was performed using protocols from the Agilent QuikChange II Site-Directed Mutagenesis kit, with the exception that KOD Hot Start DNA polymerase (Sigma-Aldrich, 71086-3) and NEB® Turbo Competent *E. coli* were used. Briefly, the PCR product was amplified with pET28-BilR as the template, using the forward primer GAAGTACACGGAGCCGCTCTGATCGGATCATTC and the reverse primer GAATGATCCGATCAGAGCGGCTCCGTGTACTTC. The rest of the steps were performed as described in the site directed mutagenesis protocols to yield pET28-BilR-D166A+R167A. Mutations were confirmed by Sanger sequencing.

Transformation with pET construct: The construct was transformed into *E. coli* T7 express lysY/Iq competent cells using the manufacturer's protocol (NEB, C3019). The cells were plated on Luria-Bertani (LB) plates with 50 ug/mL kanamycin to select for successfully transformed colonies. The transformed colonies were verified through Sanger sequencing by Azenta and plasmids were verified through Plasmidsaurus.

2.4.4 Protein purification and circular dichromism

For protein purification the *E. coli* T7 express lysY/Iq transformed with pET28-bilR(DR166AA) or pET28-bilR was growth overnight in LB broth with 50 ug/mL kanamycin. 5 mL of inoculum was transferred from the overnight cultures into fresh LB broth with 50 ug/mL kanamycin and was grown to an OD of approximately 0.5. Expression of the proteins was induced with 0.4 millimolar isopropyl β-d-1-thiogalactopyranoside and the cultures were incubated at 25°C for 18 hours. Cells were centrifuged at 3000 x g and were frozen for two hours before being resuspended in lysis buffer (20 mM Tris, 0.2 M NaCl, 1 mM phenylmethylsulfonyl fluoride, 16 ug/mL RNase, 0.5 mg/mL lysozyme). Samples were then sonicated and centrifuged at 21,130 x g. SDS-Page was used to verify that the protein was present in the supernatant of the lysate. The supernatant was then filtered using a sterile 0.2 um filter and purified using a Nickel column (G Biosciences Nickel Chelating Resin) and 5 mL gravity filtration column (G Biosciences Columns, 5 mL). The purified proteins were then eluted using filter sterilized elution buffer (20 mM Tris, 0.2 M NaCl, 200 mM Imidazole, 6.5 pH). Circular dichroism measurements were taken for the purified wild type BilR protein and mutant BilR(DR166AA) in a buffer of 20 mM tris buffer with

50 mM sodium chloride at a pH of 7. Circular dichroism measurements were done using the Jasco J-810 Spectropolarimeter with a path length of 1 millimeter and then the measurements were normalized and visualized using R.

2.4.5 Bioinformatic analysis

Identification of candidate bilR genes: Five genomes from the experimentally confirmed bilirubin reducers and five genomes from closely related confirmed non-reducers were downloaded from NCBI. The assembled genomes for *C. bolteae* CC43_001B and *Clostridium innocuum* 6_1_30 were not available so the raw sequencing data was downloaded from their corresponding bioprojects. Fastq files for the two species were download from SRA using SRA-Tools (version 2.11.0)(https://github.com/ncbi/sra-tools), trim-galore (version 0.6.7) (https://github.com/ncbi/sra-tools) was used to trim and quality filter the fastq files, and the genomes were assembled using spades (version 3.15.5)[73]. The genomes were annotated with Prokka (version 1.14.6)[74]. Orthologer (version 2.7.1) was used to group the predicted protein sequences from each genome into orthogroups using default settings[75]. ECPred (version 1.1) was then used to assign the putative E.C. numbers to each protein sequence using default setting[76]. The orthogroups were then subset to only keep groups that contained more than 2 proteins assigned to the oxidoreductase E.C. number (EC: 1.-,-,-) leaving a total of 389 orthogroups. The taxonomic distribution of the proteins within these oxidoreductase orthogroups was profiled to identify orthogroups which were present in the bilirubin reducing bacteria and absent in the non-reducing species. Only two orthogroups fit the taxonomic distribution. Of them, one was assigned to EC:1.17.7.4 (4-hydroxy-3-methylbut-2-enyl diphosphate reductase) and unlikely to be bilirubin reductase; the other one was assigned to EC:1.-,-,- and was further investigated as a putative bilirubin reductase.

Structural prediction and molecular docking: The structures for the putative BilR protein from *R. gnavus* CC55_001C and the BilR and BilS proteins from *C. symbiosum* WAL-14163 were predicted using AlphaFold (version 2.2.0)[77]. Putative substrate binding pockets were predicted using Fpocket (version 4.0.2) with default settings[78]. The predicted pockets were visualized using PyMOL (version 2.5.0) and were compared to the substrate binding region of the homologous *E. coli* 2,4-Dienoyl CoA reductase (PDB:1PS9) to identify three putative substrate

binding regions on the BilR structure (http://www.pymol.org/). Structures for bilirubin (PubChem CID 5280352) and FMN (PubChem CID 643976) were docked on the *R. gnavus* BilR structure using AutoDock Vina (version 1.2.0)[79,80]. The docking simulations were performed within 20 by 20 by 20 angstrom cubes centered on the center points of the three chosen Fpocket substrate binding pocket predictions with the exhaustiveness set to 32. The docking results were visualized using PyMOL and the results using within figures were chosen based on what is known about what bonds are reduced on bilirubin and based on comparisons to the known binding conformations of the E. coli 1PS9 protein. Protein structural alignments were performed between the *Ruminococcus gnavus* CC55_001C predicted structure and each of the *Clostridium symbiosum* WAL-14163 BilR, *C. symbiosum* WAL-14163 BilS, and *Escherichia coli* 1PS9 structures using TM-Align[81]. Protein sequence conservation was visualized on the predicted *R. gnavus* BilR structure using ConSurf based on the aligned Clade 1 BilR sequences (https://consurf.tau.ac.il/consurf_index.php).

Search for BilR in GTDB database: The alignment of the amino acid sequences from species in Clade 1 and the positions in the alignment corresponding to the first 373 residues of the *R. gnavus* CC55_001C BilR protein were extracted to generate an HMM profile using the hmmbuild tool. This profile represents the TIM barrel domain of the BilR proteins. The HMM profile was then used to perform a search within genomes of the GTDB[82] to profile the taxonomic distribution of BilR. The search was performed using hmmsearch and only hits below an e-value of 1e-100 were considered. A protein domain prediction was then performed on the putative BilR hits using InterProScan (version 5.57-90.0)[83]. The putative BilR sequences were filtered requiring that they be at least 50% of the length of the BilR sequences from *C. symbiosum* WAL-14163, they have the highly conserved HGDR motif present in their sequence, and that they only have the expected domains present (PF00724 for the shorter bilR or PF07992 and PF00724 for longer BilR version). The presence or absence of BilR was then summarized across the different bacterial taxa and the results were visualized using iTOL[84].

Profiling of bilR presence in the human gut: A collection of metagenome datasets passing basic quality control was curated to provide a cross section of gut metagenomes from infants in the first year of life (n=4296), IBD patients (n=1863), and healthy adults (n=1801) (**Supplementary Table 2**). A *bilR* reference gene dataset was generated by searching the Unified Human

Gastrointestinal Genome collection using the previously generated HMM profile of the BilR TIM barrel domain. The resulting hits were filtered based on a 1e-100 e-value threshold and were subjected to the same quality control used in the search against the GTDB database, resulting in a total of 11158 *bilR* reference sequences. The metagenomes were all processed using a standard workflow that consisted of the following steps 1) The reads were downloaded from SRA, 2) the adaptors were trimmed using Trim-Galore with default settings, 3) potential human contaminant reads were identified by mapping the reads to a human genome reference (assembly T2T-CHM13v2.0) and were removed using Samtools (version 1.16.1)[85], 4) samples with less than one million reads were not kept, 5) reads were aligned the *bilR* reference gene set using Bowtie2[86]. The number of reads mapped to the *bilR* reference database was then summarized for each sample by normalizing by the total number of reads in the sample and multiplying by one million to give a *bilR* counts per million value (CPM).

Infant related metagenomes were binned into age groups of 30 days from 0 days to 1 year of life and samples without age metadata were not considered in this analysis. Samples from IBD patients were categorized based on if the patient had Ulcerative Colitis or Crohn's Disease. For healthy gut metagenomes samples were excluded if they were from patients under three years of age. *bilR* was considered to be present in a sample if the *bilR* CPM value was greater than five CPM. The prevalence of *bilR* between groups was compared using a test of equal proportions, using the *prop.test* function from the R *stats* library, to test if the given proportion of samples with *bilR* present was different between two groups.

Chapter 3: Convergent evolution of oxidized sugar metabolism in commensal and pathogenic microbes in the inflamed gut

This chapter contains material previously published in Nature Communications in Convergent evolution of oxidized sugar metabolism in commensal and pathogenic microbes in the inflamed gut,[87] *which was a joint work with Angela K. Jiang, Maggie R. Grant, Gabriela Arp, Glory Minabou Ndjite, Xiaofang Jiang, and Brantley Hall. S.L. contributed to the conceptualization of the project, performed all experimental validation, and wrote the manuscript with A.K.J..*

S.L., B.H., and X.J. conceptualized the project. S.L., A.K.J., M.G., G.A., and G.M.N., performed the experiments and/or analyses. S.L. and A.K.J. wrote the paper.

3.1 Introduction

Inflammation-driven perturbations of the gut microbiome play an important role in the positive feedback circuit underlying gastrointestinal inflammation, as exemplified in Inflammatory Bowel Disease (IBD), where microbial perturbations trigger an aberrant immunological response leading to the characteristic chronic inflammation and gastrointestinal symptoms.[88,89] This maladaptive interaction highlights the complex relationship between the commensal gut microbiome and the host immune system in the etiology of gut inflammation in both IBD and colitis.[90,91]

The taxonomic changes associated with inflammation are well documented, but the mechanisms underlying these changes remain understudied. Shifts in taxonomic composition often include increases in the relative abundance of *Escherichia, Veillonella, Klebsiella, Blautia,* and *Lachnoclostridium* spp.[24,92–94]

Recent studies have begun elucidating pathways responsible for these inflammation-induced alterations to the microbial community, including harmful effects of the oxidizing environment, increased abundance of terminal electron acceptors, and reduction of butyrate available to colonocytes.[21,26,95–98] Several key pathways trace back to the overexpression of *NOS2,* which encodes inducible nitric oxide synthase (iNOS), in the inflamed gut. *NOS2* expression is induced by transcription factor (NF)-kB during the inflammatory response.[99] Consequently, *NOS2* is highly upregulated during IBD. The Integrative Human Microbiome Project reported 4-12 fold higher *NOS2* expression in those with active flares of IBD compared to healthy individuals.[100–104]

In colonic epithelial cells, *NOS2* expression produces nitric oxide radicals that diffuse directly into the lumen and spontaneously oxidize glucose and galactose to glucarate (saccharic acid) and galactarate (mucic acid), respectively. PRISM metabolomics revealed concordantly high abundance oxidized sugars in the inflamed IBD gut.[27,105]

Metabolism of oxidized sugars during inflammation has been implicated as a pathogen colonization factor. Faber et al. demonstrated enhanced colonization of *Salmonella enterica* in the inflamed murine gut via oxidized sugar metabolism, while genetic ablation of the oxidized sugar pathway reduced *S. enterica* competition.[27]

The oxidized sugar metabolic pathway - hereafter referred to as the *gud/gar* pathway - enables bacteria to use glucarate and/or galactarate as a carbon source by catabolizing the oxidized sugars to pyruvate and 2-phosphoglycerate (2-PG) which can then be further metabolized through the remaining steps of canonical glycolysis.[106,107] During the penultimate stage of glycolysis, enolase converts 2-PG to phosphoenolpyruvate, which pyruvate kinase subsequently turns into pyruvate, producing ATP.[108,109]

The well-characterized *E. coli gud/gar* pathway contains five core genes: glucarate/galactarate permeases (*gudP/garP*) imports the oxidized sugars, glucarate/galactarate dehydratases (*gudD/garD*) convert the oxidized sugars into a common 5-keto-4-deoxy-D-glucarate (5-KDG) intermediate, 5-KDG aldolase (*garL*) splits 5-KDG into pyruvate and tartronate semialdehyde (TSA), TSA dehydrogenase (*garR*) forms glycerate, and, as a final step, glycerate kinase (*garK*) produces 2-PG (Figure 3.1a).[106,110,111] Despite the thorough characterization of the *gud/gar* pathway in *Enterobacteriaceae* spp., it is unknown whether other inflammation-adapted microbes can also utilize this pathway.

Here, we sought to systematically investigate the breadth of the diversity of commensal gut microbes that metabolize oxidized sugars. We found that several Bacillota (previously Firmicutes) species with increased relative abundance during IBD were able to use oxidized sugars as a carbon source. We searched for the *gud/gar* pathway in these species, and found that while some genes homologous to those in the Enterbacteriaceae pathway - *gud/garD* and *garR* - were present, other core pathway genes, like the permease *gud/garP* and aldolase *garL* were conspicuously missing. However, we hypothesized that unannotated genes within the same operons as the homologous genes might underlie the ability of these species to metabolize oxidized sugars. We individually tested the function of each gene in single-knockout *E. coli,* and observed

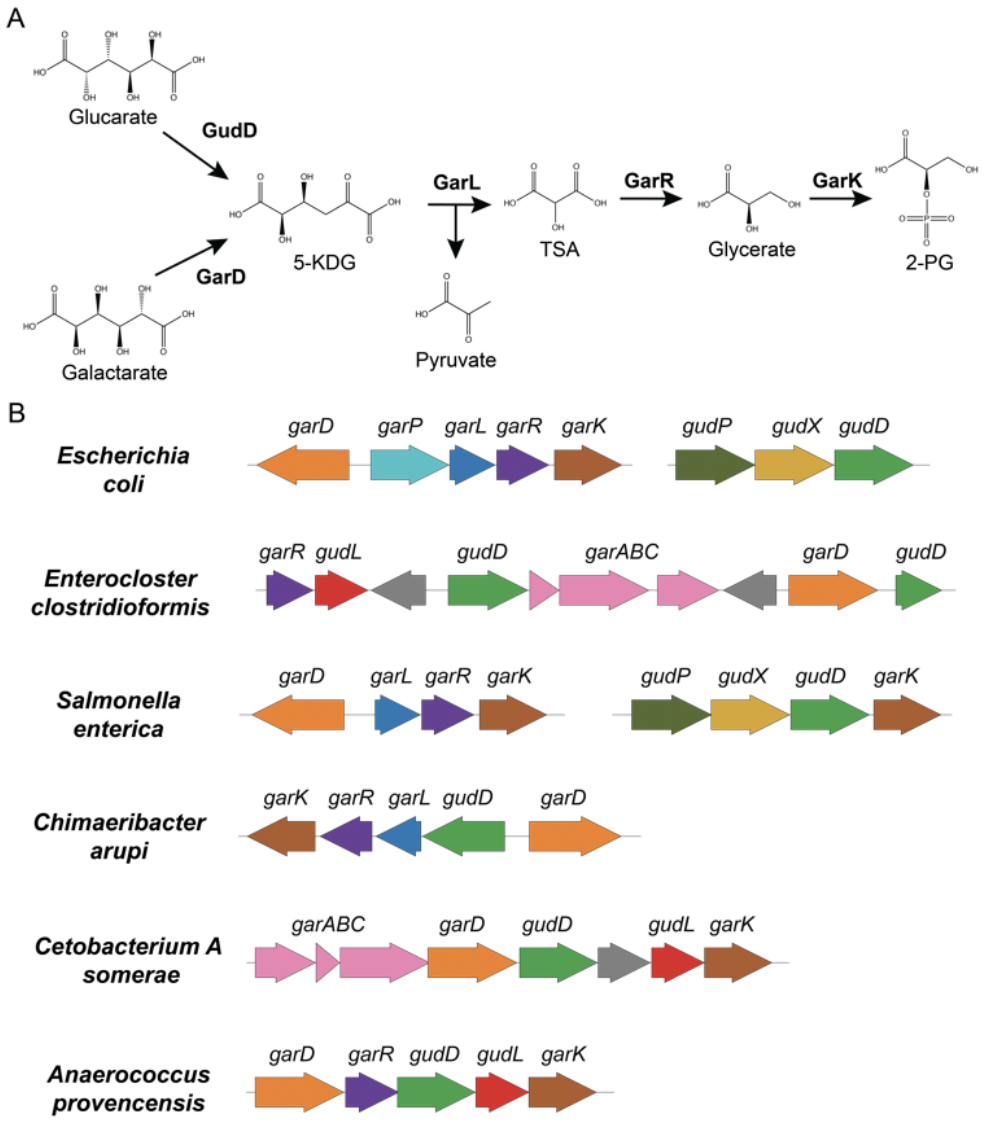

Figure 3.1 Oxidized sugar metabolism. **a)** Glucarate and galactarate catabolic pathway. **b)** *Gud/gar* gene cluster organization from *E. coli, E. clostridioformis, S. enterica, C. arupi, C. somerae,* and *A. provencensis.*

that genes from *Enterocloster clostridioformis* (previously known as *Clostridium clostridioforme)*, including an uncharacterized ABC transporter and aldolase, were able to rescue the corresponding knockout. Though *E. clostridioformis* contains no *garL* homolog, the uncharacterized aldolase was able to restore oxidized sugar metabolism of *E. coli garL* knockouts, revealing a novel putative 5-KDG aldolase we named *gudL*. Similarly, *E. clostridioformis* does not contain permeases *gudP* or *garP,* but instead utilizes an ABC transporter within the same operon as a homologous *gudD.* The previously uncharacterized *E. clostridioformis* oxidized sugar transporter *garABC* complements *gudP* and *garP E. coli* knockouts. Based on our discovery that *E. clostridioformis* metabolizes

oxidized sugars using a novel combination of homologous genes, convergently evolved enzymes, and alternative transporters, we systematically searched for gut microbial species that contained the *E. coli* or *E. clostridioformis gud/gar* pathway, identifying 887 species with homologous genes, including *Blautia producta* and *Fusobacterium nucleatum.* Thus, we can conclude that oxidized sugar metabolism is far more widespread than previously known, and exemplifies the convergent evolution of metabolic adaptation to the inflamed gut.

3.2 Results

3.2.1 Screening for gut microbial species capable of metabolizing oxidized sugars

We hypothesized the increased abundance of certain gut microbes in IBD may be attributed to their capacity to metabolize oxidized sugars produced during inflammation. To investigate this theory, we screened a selection of species that were both increased in relative abundance during IBD, and lacked a known mechanism behind their increased abundance for growth on oxidized sugars as a carbon source.[22]

E. coli 5α, *E. clostridioformis* WAL-7855, and *Enterocloster citroniae* WAL-17108 (*Clostridium citroniae)* grew on galactarate as a carbon source, while *Clostridium symbiosum* WAL-14163, *Mediterraneibacter lactaris* CC59-002D (*Ruminococcus lactaris)*, and *Bacteroides fragilis* CL03T12C07 did not, despite being associated with IBD (Figure 3.2).[22] The observed growth of these gut microbial species suggests the presence of genes essential for oxidized sugar metabolism. Consequently, we set out to identify these putative metabolic genes.

3.2.2 Identification of oxidized sugar metabolism gene cluster in *Enterocloster clostridioformis*

We theorized that the observed growth of select gut species on galactarate as a carbon source was facilitated by the *gud/gar* pathway previously characterized in *E. coli*. To identify these genes, we searched for homologs using *E. coli* proteins GudD*,* GarD*,* GarL*,* GarR, and GarK as reference for the minimal viable enzymatic pathway for oxidized sugar catabolism.

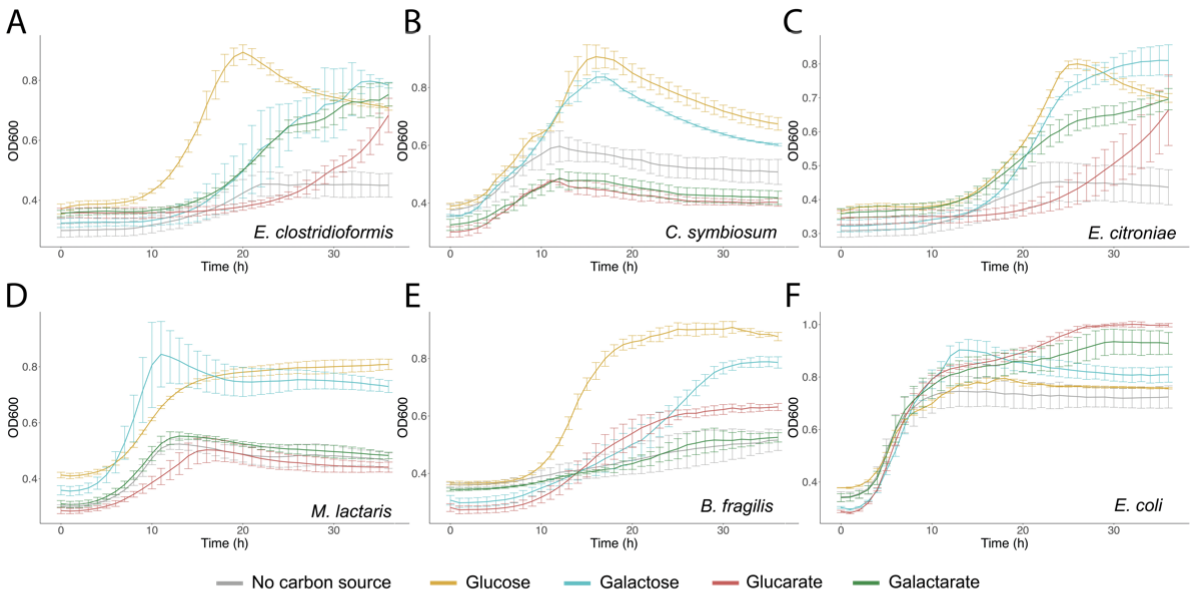

Figure 3.2 Identification of gut microbial species increased during inflammation capable of metabolizing glucarate or galactarate as a carbon source. OD600 of **a)** *Enterocloster clostridioformis* WAL7855 **b)** *Clostridium symbiosum* WAL14163, **c)** *Enterocloster citroniae* WAL-17108, **d)** *Mediterraneibacter lactaris* CC59-002D, **e)** *Bacteroides fragilis* CL03T12C07, and **f)** *E. coli* 5α grown in media with glucose, galactose, glucarate, or galactarate as an additional carbon source (n=3 biological replicates). Data are presented as the mean +/- standard deviation. Source data are provided as a Source Data file.

Notably, BLASTp only revealed proteins in *E. clostridioformis* homologous to GudD (percent identity 64.3%, e-value 7.37e-221, bitscore 611), GarD (percent identity 53.9%, e-value 5.11e-191, bitscore 541), GarR (percent identity 67.8%, e-value 2.27e-140, bitscore 395), and two copies of GarK (percent identity 43.5%, e-value 2.14e-98, bitscore 285; percent identity 42.3%, e-value 1.22e-100, bitscore 291). However, these proteins alone are unlikely to fully recapitulate the metabolic pathway to catabolize oxidized sugars, as two key enzymes are missing: Gud/GarP permeases necessary for oxidized sugar import and aldolase *garL,* responsible for 5-KDG catabolism to pyruvate and TSA. The missing enzymes raised questions as to how *E. clostridioformis* was able to metabolize galactarate without a complete oxidized sugar catabolic pathway.

Investigating the *E. clostridioformis* genes further, we manually examined the neighborhood of the homologous genes. Interestingly, we found an aldolase that was within the same operon as *garR,* but was not homologous to the *E. coli garL* aldolase. Additionally, while *E. clostridioformis* lacked the *gudP* and *garP* permeases, in the same operon as *gudD,* there was an uncharacterized ABC transporter (Figure 3.1b). Given the close proximity of these genes to the homologous *gud/gar* genes, we theorized that the novel aldolase *gudL* and uncharacterized ABC transporter may also be involved in oxidized sugar metabolism.

3.2.3 Both novel and homologous genes from the E. clostridioformis gud/gar gene cluster rescue *E. coli* knockouts

Due to the discovery of novel genes and divergent operon organization, we sought to validate the functionality of the putative *E. clostridioformis gud/gar* gene cluster through complementation of *E. coli gud/gar* knockouts. Since most *E. coli* K-12 derivatives natively contain the *gud/gar* metabolic pathway, we needed mutant strains of each *gud/gar* gene of interest. For this purpose, we employed the Keio collection, an *E. coli* BW25113 mutant library containing in-frame, single gene deletions of all nonessential genes, including *gudP, garP, garL,* and *garR*.[112]

To create our complementation library, the putative *gud/gar* genes from *E. clostridioformis* were cloned into a pCW-lic vector backbone under the control of an inducible *tac* promoter, creating plasmids to individually express the putative oxidized sugar transporter *garABC,* the putative oxidized sugar aldolase *gudL,* and the homologous reductase *garR*.[113] Then, we transformed each plasmid into the respective Keio *E. coli* knockout strain: pCW-*Clos*-*gudL* was transformed into Δ*garL E. coli,* pCW-*Clos*-*garR* into a Δ*garR* mutant, and pCW-*garABC* was transformed into both Δ*gudP* and Δ*garP* mutants, as we hypothesized that the *E. clostridioformis* ABC transporter would import both glucarate and galactarate.

In order to verify the *E. clostridioformis gud/gar* pathway functionality, we employed a sugar fermentation assay adapted from Faber et al. used to measure fermentation of carbon sources.[27] A minimal peptone fermentation media containing bromothymol blue pH indicator was supplemented with 10 g/L glucose, glucarate, or galactarate as a carbon source.[27] This media changes from a deep blue color to yellow based on the pH change that occurs as bacteria grow and acidify the media (Figure 3.3a). We inoculated *E. coli* into the fermentation media and induced expression with isopropyl β-D-1-thiogalactopyranisode (IPTG). After 48 hours, the color change of the media was measured as an indication of carbon source utilization.

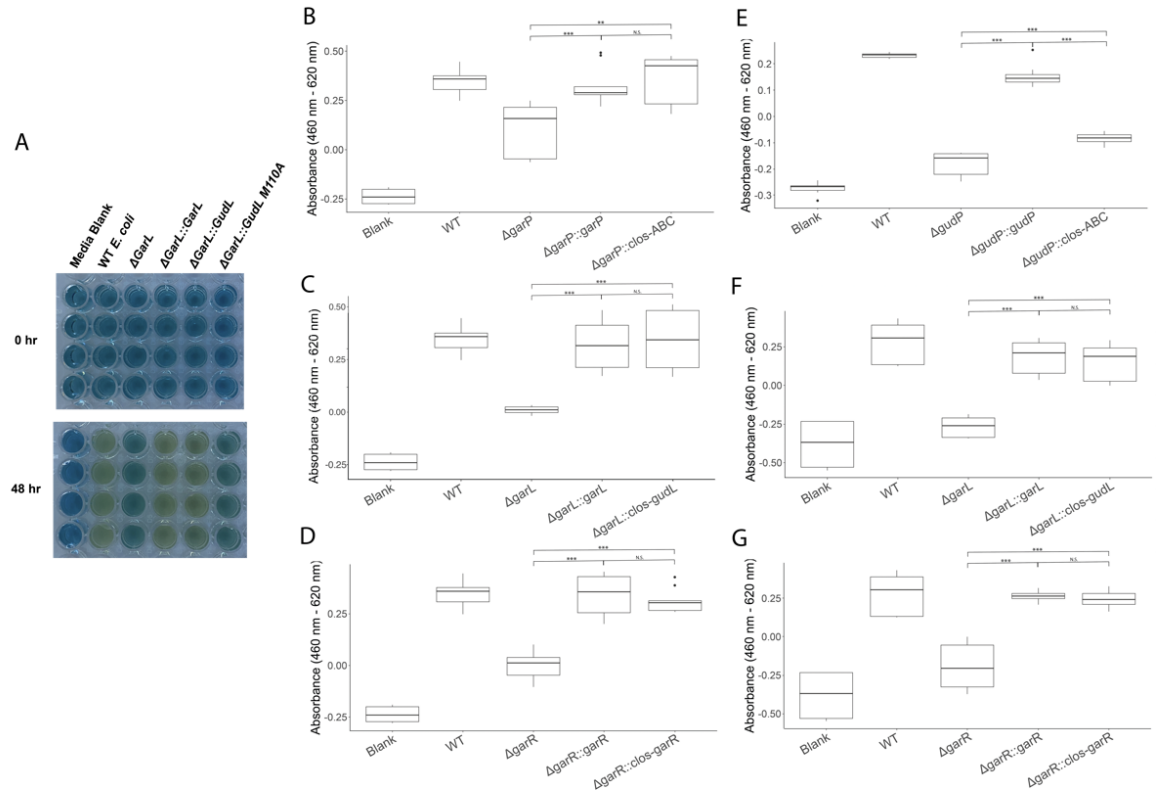

Figure 3.3 Complement metabolism of oxidized sugars. **a)** Representative images of fermentation media with 10 g/L galactarate inoculated with *E. coli* BL21, *ΔgarL, ΔgarL::garL, ΔgarL::Clos-gudL* or *ΔgarL::gudL-M110A* at 25°C for 48 hours. **b)** Fermentation assay after 48 hours of *ΔgarP*, *ΔgarP::garP*, and *ΔgarP::Clos-ABC E. coli* on galactarate as a carbon source. **c)** Fermentation assay after 48 hours of *ΔgarL*, *ΔgarL::garL*, and *ΔgarL::Clos-gudL E. coli* on galactarate as a carbon source. **d)** Fermentation assay after 48 hours of *ΔgarR*, *ΔgarR::garR*, and *ΔgarP::Clos-garR E. coli* on galactarate as a carbon source. **e)** Fermentation assay after 48 hours of *ΔgudP*, *ΔgudP::gudP*, and *ΔgudP::Clos-ABC E. coli* on glucarate as a carbon source **f)** Fermentation assay of *ΔgarL*, *ΔgarL::garL*, and *ΔgarL::Clos-garL E. coli* on glucarate as a carbon source **g)** Fermentation assay of *ΔgarR*, *ΔgarR::garR*, and *ΔgarR::Clos-garR E. coli* on glucarate as a carbon source. Wild type *E. coli* BL21 used as a positive control, and a media blank as a negative control.

Knockouts of *E. coli gudP* and *garL* abolished fermentation of glucarate and galactarate carbon sources, while deletion of *garP* and *garR* attenuated oxidized sugar fermentation (Figure 3A-G, Supplementary Figure 2). Complementation of the native *E. coli* gene back into the mutant strain restored utilization of glucarate and galactarate to that of the wild-type *E. coli* in all cases, validating the ectopic expression methodology (Figure 3.3a-g).

Complementation of the Δ*garP* mutant with *E. clostridioformis garABC* rescued the knockout, while the same genes in Δ*gudP* mutants only partially rescued the metabolic defects (Figure 3.3b,e). Complementation of Δ*garR* mutants with the homologous *E. clostridioformis garR* fully recovered the oxidized sugar metabolism phenotype on both glucarate and galactarate. Most notably, cross-species complementation of the *E. coli* Δ*garL* mutants with the putative *E. clostridioformis* aldolase *gudL* restored fermentation of glucarate and galactarate as strongly as isogenic complementation (Figure 3.3a,c,f).

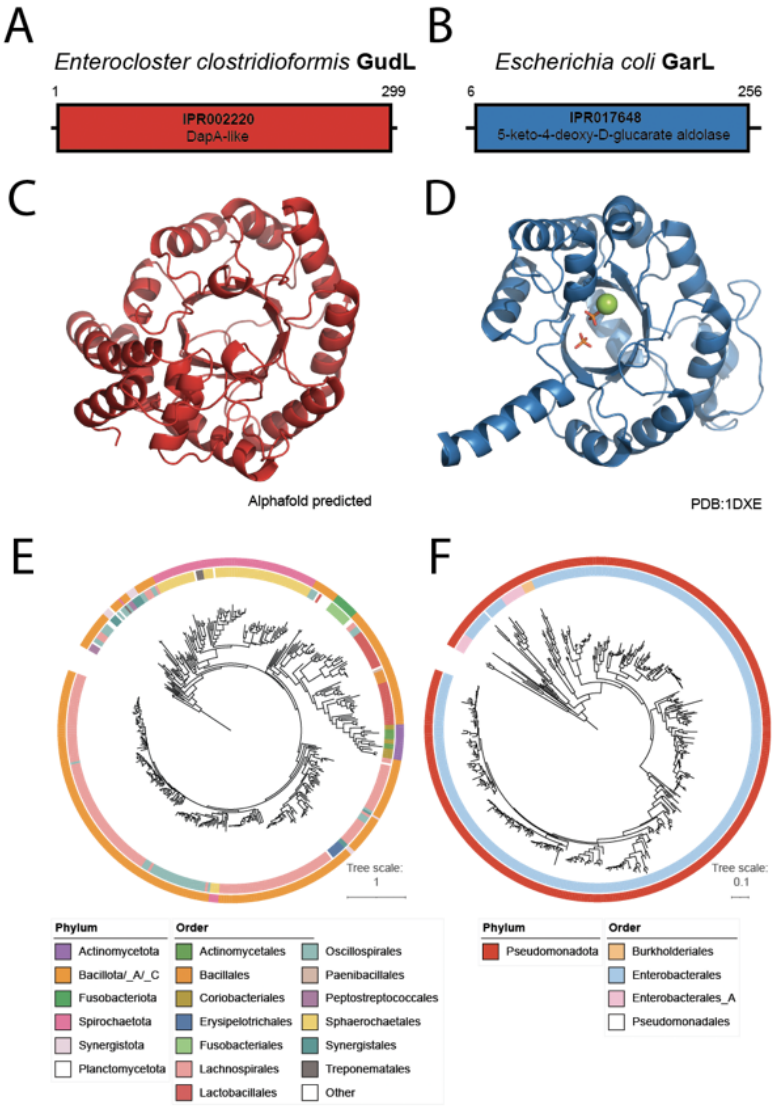

Figure 3.4 Comparison of GudL and GarL. Interproscan annotations of **a)** *E. clostridioformis* GudL and **b)** *E. coli* GarL. **c)** Alphafold-predicted structures of *E. clostridioformis* GudL and **d)** the monomer X-ray diffraction structure *E. coli* GarL obtained from the RCSB Protein Data Bank (ID: 1DXE). The RMSD of these two structures was 3.05, calculated using TM-align. Phylogenetic gene trees of **e)** *E. clostridioformis* GudL and **f)** *E. coli* GarL.

The restoration of *E. coli* knockout fermentation of oxidized sugars by heterologous expression of *E. clostridioformis* genes *garABC, gudL,* and *garR* demonstrates the functional equivalency of the encoded enzymes in bacterial metabolism of oxidized sugars.

3.2.4 The *dapA*-like gene *gudL* encodes a putative 5-KDG aldolase

The discovery of the novel GudL aldolase in *E. clostridioformis,* which catalyzes the same function as GarL in *E. coli,* prompted us to study its evolutionary history. Thus, we investigated the evolution of GudL to gain insights into its emergence and diversification across bacterial species. Despite the low amino acid pairwise identity of 7% between the *E. coli* 5-KDG aldolase

GarL and *E. clostridioformis* aldolase GudL, their shared functionality led to our hypothesis that this similarity is likely due to convergent evolution. An Interproscan search revealed that the *E. clostridioformis* GudL is annotated as "DapA-like" (IPR017648) and shares no common protein domain annotations with the *E. coli* 5-KDG aldolase GarL, substantiating the convergent evolution hypothesis (Figure 3.4a,b).[114]

We obtained further evidence supporting different common ancestors through structural comparisons of GarL and GudL. The Alphafold-predicted structures of GudL and the X-ray crystallography structure of *E. coli* GarL (PDB: 1DXE) exhibit a root mean square deviation (RMSD) of 3.05, indicating structural divergence (Figure 3.4c,d).[115] In contrast, as previously mentioned, the *E. clostridioformis* GarD, GudD, and GarR exhibit high shared pairwise amino acid identity to their *E. coli* counterparts (57%, 64%, and 68%, respectively), implying homology.

Together, these data suggest the *E. clostridioformis gud/gar* pathway consists of GudD, GarD and GarR enzymes homologous to their *E. coli* counterparts, but uses a newly discovered, convergently-evolved aldolase GudL predicted to catalyze the conversion of 5-KDG to TSA and

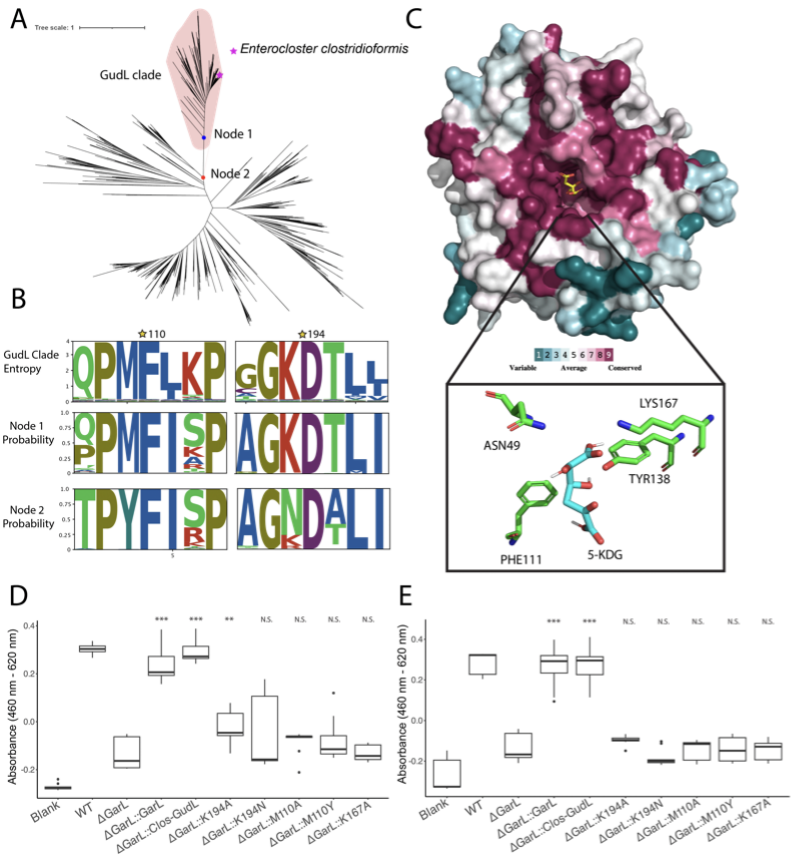

Figure 3.5 Delineation of GudL. **a)** Gene tree constructed from putative GudL sequences and related enzymes annotated as dapA-like (COG0329), showing a possible delineation of the GudL clade and the location of *E. clostridioformis* GudL. **b)** Diagram showing the conservation (entropy) of active site residues in clade 1, as well as the GRASP predicted ancestral states of Node 1 (GudL clade) and Node 2 (putative ancestral node). Active site residues with a predicted change from the ancestral state are labeled with a star. **c)** AlphaFold-predicted structure colored by degree of conservation of amino acid residues based on a ConSurf analysis. Red represents conserved regions, and green represents variable regions. A docked 5-KDG is shown on the structure (yellow and blue). The potential active site residue interactions with 5-KDG are shown, as predicted by Autodock-vina. **d)** Fermentation assay after 48 hours of wild type *E. coli BL21, ΔgarL*, *ΔgarL::garL*, *ΔgarL::Clos-GudL E. coli,* and *gudL* mutants K194A, K194N, M110A, M110Y, and K167A on galactarate and **e)** glucarate as a carbon source Significance is relative to *ΔgarL*. Experimental controls are shared with Figure 7D.

pyruvate, and an ABC transporter that functionally compensates for the absence of GudP or GarP to facilitate galactarate import (Figure 3.1b).

3.2.5 Delineation of the GudL clade

To delineate GudL from other DapA-like proteins, we used a combination of phylogenetic analyses, structural prediction, and sequence conservation, identifying key residues that differentiate GudL from DapA. Our phylogenetic analysis of DapA-like proteins revealed a putative GudL clade, based on the positions of the *E. clostridioformis* GudL and the branch length of the clade (Figure 3.5a).

Ancestral sequence reconstruction showed changes in the catalytic residues from the ancestral state (Node 2) (Figure 3.5b). Notably, there was a predicted change from Y110 to M110 and N194 to K194 from the ancestral state to the GudL clade, which are universally conserved among the putative GudL clade. These residues unique to the GudL clade may contribute to its function, and delineate it from other DapA-like enzymes. We mutated the *E. clostridioformis* GudL sequence to change position M110 and K194 to a neutral amino acid with no side chain (M110A, K194A), and to their ancestral states (M110Y, K194N). Mutating these residues to alanine abolished fermentation of both glucarate and galactarate, with the exception of K194A, which reduced but did not abolish oxidized sugar metabolism (Figure 3.5d,e). These results indicate that the GudL clade can be delineated from other DapA at Node 1.

We identified putative catalytic residues of GudL through structural alignments with the homologous 4-hydroxy-tetrahydrodipicolinate synthase (3PB2), revealing similar positioning of K167 to the DapA catalytic residue K161 (Figure 3.5c). In DapA, K161 catalyzes schiff base formation during the synthesis of 2,3-dihydrodipicolinate from pyruvate and L-aspartate-4-semialdehyde.[116] Consequently, we hypothesize that K167 may play an important role in the predicted catalytic reaction of 5-KDG by GudL. To verify the importance of K167, we mutated the GudL sequence to change position K167 into alanine. The mutation abolished utilization of both glucarate and galactarate (Figure 3.5d,e).

3.2.6 Oxidized sugar metabolism is a strain-specific capability widely distributed across the microbiome

Our original hypothesis was that commensal microbes may have adapted to utilize oxidized sugars in order to gain a competitive advantage in the inflamed gut. Once we demonstrated through genetic complementation that the divergent pathway for oxidized sugar metabolism in *E. clostridioformis* was functional, we expanded our search to other IBD-associated taxa and commensal microbes. Using ProkFunFind, a tool to annotate genes across gut microbial species, we searched the 85,202 prokaryotic genomes from the Genome Taxonomy Database (GTDB) for species that contained either homologs to the *E. coli* GudD/GarD, GarL, and GarR proteins, or homologs to *E. clostridioformis* GudD/GarD, GudL, and GarR.[117,118] Species that contained GudD or GarD, GarL or GudL, and GarR were considered to contain the minimum viable *gud/gar* pathway. GarK was excluded from the minimum viable pathway, as upon constructing the *garK* gene tree, we found that *garK* was polyphyletic, suggesting independent recruitment of kinases multiple times throughout evolution. This pattern contrasts sharply with the other *gud/gar* genes

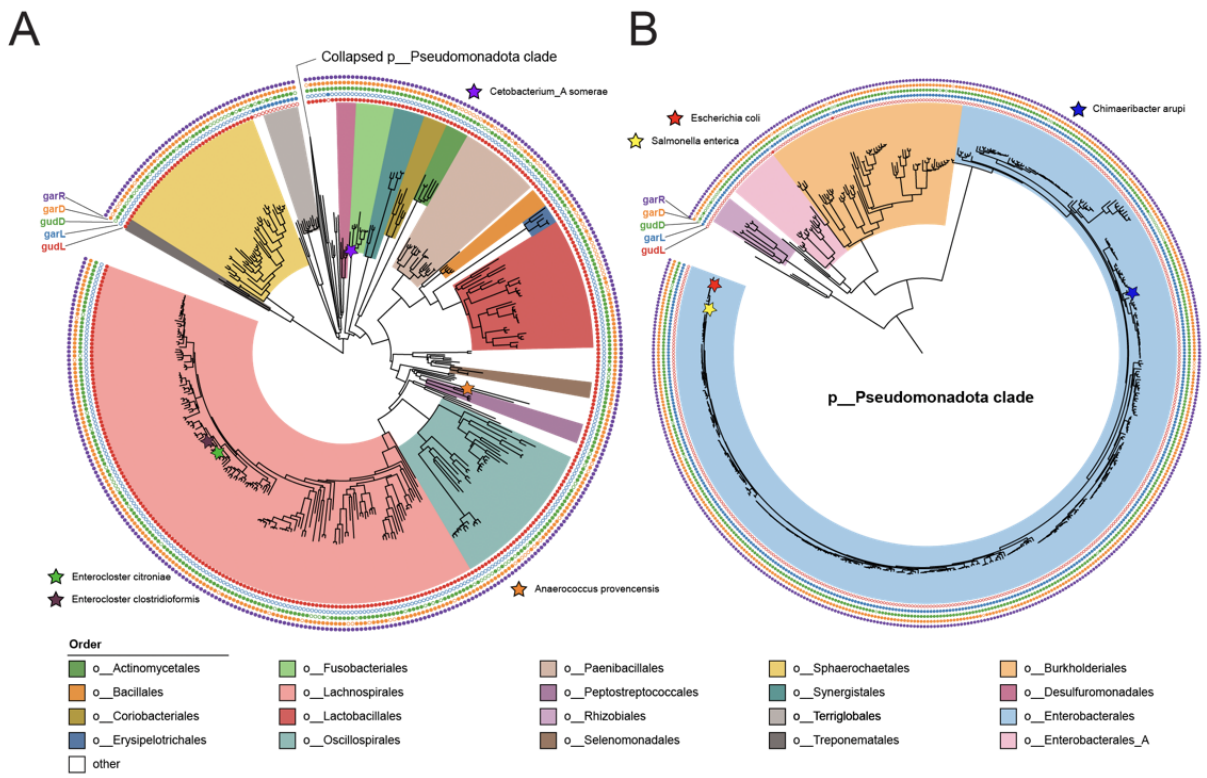

Figure 3.6 Species tree of GTDB taxa. The minimum viable gud/gar pathway is annotated with colored dots showing the presence or absence of GarR, GarD, GudD, and GudL in each species. The tree is divided into two segments, with one section collapsed in each part. **a)** Section of tree showing primarily species with GudL instead of GarL, including phyla such as Bacillota and Fusobacteriota. **b)** Subsection of the Pseudomonadota clade, showing primarily species with GarL instead of GudL. The species tree was generated by pruning the GTDB species tree using the gotree prune command.

(*gudD/garD*, *gudL/garL*, and *garR*), which show consistent co-evolution as a single functional unit. This method identified the putative minimum viable *gud/gar* pathway in 887 gut microbial species, including *E. citroniae,* and *Fusobacterium nucleatum E* (Figure 3.6). *Gud/gar* was primarily found in the phyla Bacillota (316 species), Pseudomonadota/Proteobacteria (505 species), and Fusobacteriota (10 species). Interestingly, GarL was primarily found in Pseudomonadota, while GudL is widely spread among a diverse range of phyla, including Bacillota and Fusobacteriota (Figure 3.4e,f).

Oxidized sugar metabolism appears to be a strain-specific function. For example, 2 out of 38 *Fusobacterium polymorphum* genomes, and 6 out of 18 *Fusobacterium mortiferum* genomes contain the *gud/gar* pathway.

3.2.7 Horizontal gene transfer of the *gud/gar* gene cluster from Bacillota to Fusobacteriota

Given that the gud/gar pathway was present in divergent phyla such as Bacillota, Fusobacteriota, and Pseudomonadota, we sought to understand the evolutionary history of the *gud/gar* pathway by reconstructing phylogenetic trees of the GudD*,* GarD*,* GudL*,* and GarR genes of the 887 species with the gud/gar pathway. The incongruence of the species cladogram and the GarD, GudD, and GarR gene trees suggest that the gene cluster may have spread via horizontal gene transfer, notably indicating a potential transfer from Bacillota to Fusobacteriota, possibly from the *Enterococcus* spp. clade. The GarD*,* GudD*,* and GarR gene trees show the Fusobacteriota clade nested within the Bacillota clade, despite being from a different phyla (Figure 3.7). The similarity between the *Enterococcus* spp. GarR to the *Fusobacterium* spp. GarR in both sequence identity and gene cluster organization suggests that a horizontal gene transfer event of the *garR* gene cluster occurred between Bacillota and Fusobacteriota (Figure 3.7). For example, the similarity of the GarR sequences between *Fusobacterium B.* sp. and *Enterococcus D. casseliflavus* is 82%. Furthermore, *Fusobacterium B.* sp. and *Enterococcus D. casseliflavus* share 5 genes in the *garR* gene clusters, these being *gudD*, *gudL*, *garK*, *garR*, and and *cDaR*, implicating horizontal gene transfer. This suggests that Fusobacteriota horizontally acquired a gene cluster that could be

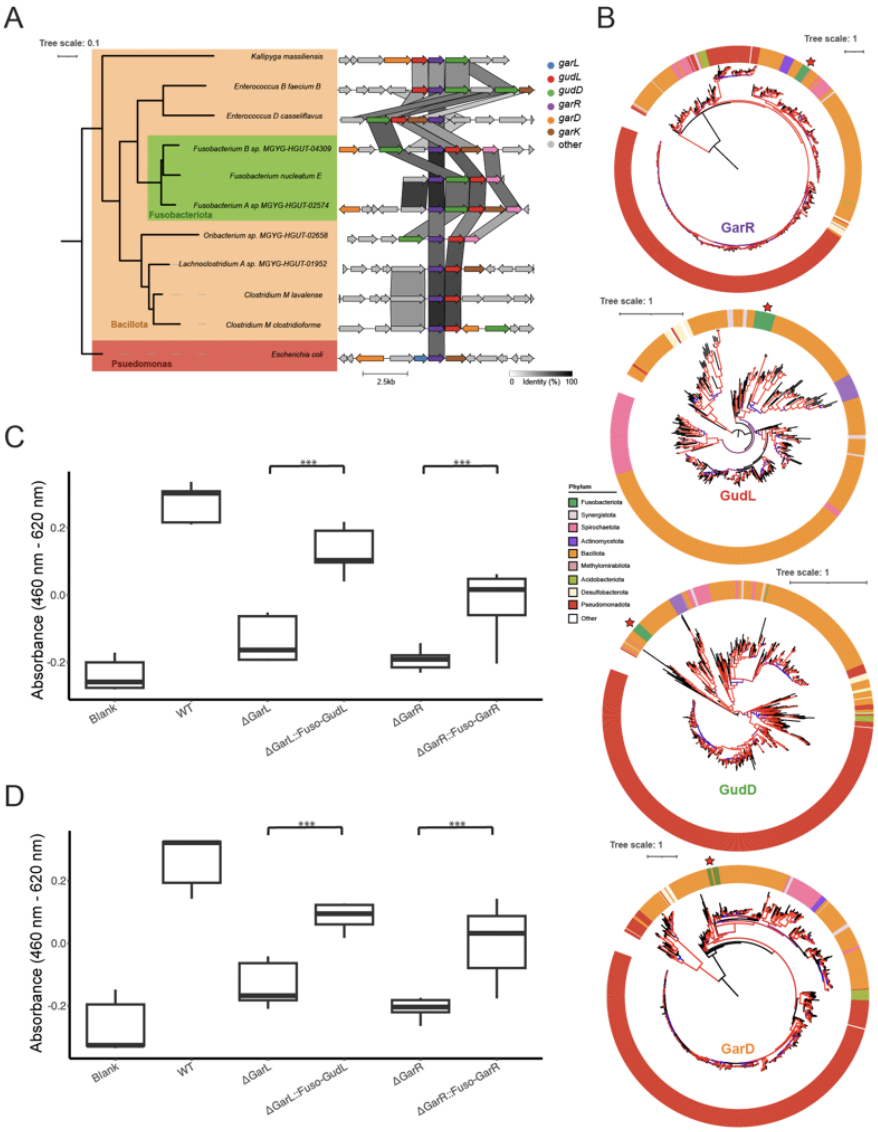

Figure 3.7 Oxidized sugar metabolism in *Fusobacterium* **a)** Phylogenetic tree and synteny analysis of the *garR* and *gudL* cluster in representative *Fusobacterium* spp. and Bacillota species. Homologous genes are represented by the same color, and darker links show higher similarity between homologous genes. Diagram produced by clinker v0.0.28. The phylogenetic tree was generated using IQ-TREE2 from the putative GarR proteins of the representative Bacillota and *Fusobacterium* spp., with *Escherichia coli* GarR as the outgroup. **b)** Gene trees of GarR, GudL, GarD, and GudD in gut microbial proteome. Genomes from GTDB. The trees are colored by phylum. Branches are colored to reflect bootstrap values, ranging from 75 (blue) to 100 (red). **c)** Fermentation assay after 48 hours of wild type *E. coli, ΔgarL*, *ΔgarL::Fuso-gudL*, *ΔgarR,* and *ΔgarR::Fuso-garR E. coli* on and **d)** glucarate as a carbon source

adaptive for inflammatory environments, potentially contributing a mechanism to the previously observed association of Fusobacteriota with IBD and colorectal cancer.[119]

To verify the functionality of the *Fusobacterium nucleatum gud/gar* genes, we complemented Keio *E. coli* knockouts with the *F. nucleatum* genes *gudL* and *garR*, which partially restored fermentation of both glucarate and galactarate (Figure 3.7c,d).

3.2.8 IBD patients exhibit increased relative abundance of the *gud/gar* pathway in fecal metagenomics and metatranscriptomics

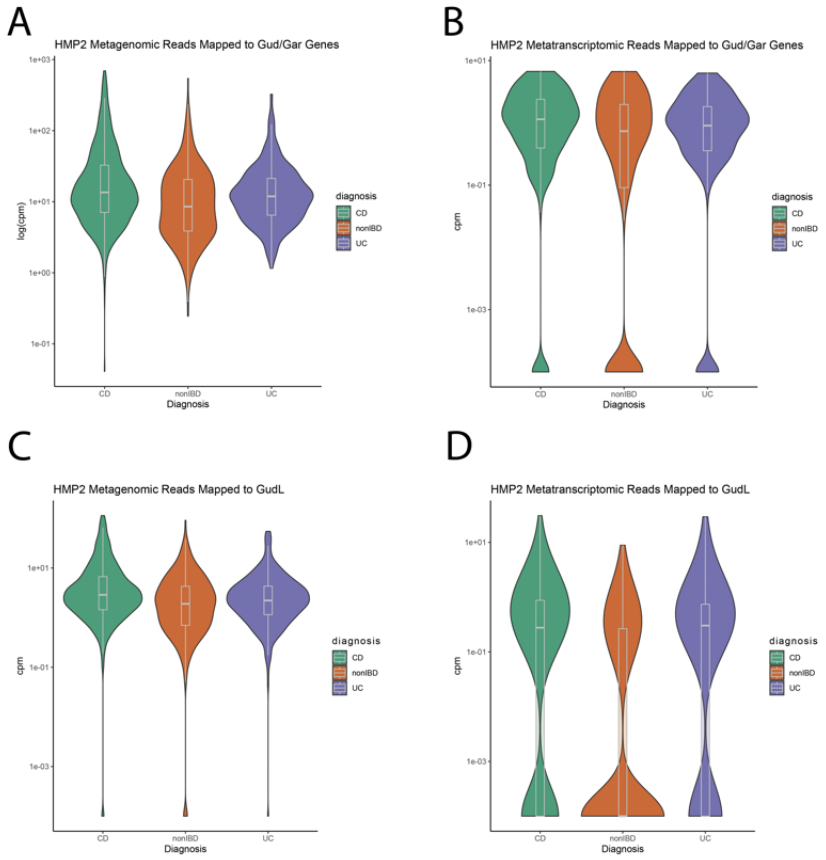

Figure 3.8 Metagenomic and metatranscriptomic analysis of *gud/gar* and *gudL* in IBD and non-IBD patients. **a)** Metagenomic analyses of *gudD*, *garD*, *garL*, and *garR* (Two-sided Mann Whitney U Test, CD *p=3.331e-10 W=126260,* UC *p=0.0002354 W=50742,* UC-CD *p=0.008988 W=105092*) and **b)** *gudL* only (Two-sided Mann Whitney U Test, CD *p=1.107e-09 W=125511,* UC *p=0.01401 W=53978,* UC-CD *p=0.0002738 W=108977*). **c)** Metatranscriptomic analysis of genes *gudD, garD, garL, gudL,* and *garR* in UC, CD, or non-IBD patients (Two-sided Mann Whitney U Test, CD *p=0.002261 W=103445,* UC *p=0.07443 W=52967,*UC-CD *p=0.3782 W=28152*). **d)** Metatranscriptomic analyses of *gudL* in UC, CD, or non-IBD patients (Two-sided Mann Whitney U Test, CD *p=1.548e-19 W=123662,* UC *p=1.614e-14 W=38031,*UC-CD *p=0.7831 W=27261*). Data from HMP2 and HPFS.

We hypothesized that the ability to utilize oxidized sugars was an adaptation that conferred a competitive advantage to microbes in the inflamed gut. Therefore, we sought to determine if the relative abundance of microbes encoding the gud/gar pathway were increased in IBD compared to non-inflamed controls. However, due to our previous discovery that *gud/gar* are accessory genes, we mapped metagenomic reads directly to a database of gud/gar operons rather than examine species-level changes. Using the homologous *gud/gar* genes *garD*, *gudD*, and *garR* as a reference database, we mapped the stool metagenomic reads from 1261 metagenomes of Human Microbiome Project 2 (HMP2) data (UC n=337, CD n=565, non-IBD n=359) and determined that the genes in the *gud/gar* pathway are found in higher relative abundance in IBD patients compared to non-IBD patients (Mann U Whitney test, CD *p=6.181e-08 W=122824,* UC *p=0.01763 W=54200,* UC-CD *p=0.008988 W=105092*) (Figure 3.8a).

We further investigated whether these genes were being expressed *in vivo*. Stool metatranscriptomic analysis of 1106 metatranscriptomes (UC n=184, CD n=292, non-IBD n=630) found increased relative abundance of *gud/gar* pathway transcripts in IBD patients (MTX: CD *p=0.002261 W=103445*, UC *p=0.07443 W=52967*, UC-CD *p=0.3782 W=28152)*, suggesting that

not only that the relative abundance of these species is increased during IBD, but they are also actively transcribing the *gud/gar* genes suggesting their functional utility (Figure 3.8c).

However, we wanted to confirm whether the alternative pathway identified in *E. clostridioformis* was contributing to the increase in *gud/gar* in IBD patients (UC n=166, CD n=270, non-IBD n=570). To do this, we used 380 putative *gudL* homologs as a reference database, and determined that *gudL* was increased in patients with IBD for both the metatranscriptomes and metagenomes (Figure 3.8b,d) (Mann Whitney U Test, MGX: CD *p=3.331e-10 W=126260*, UC *p=0.0002354 W=50742*, UC-CD *p=0.008988 W=105092*, MTX: CD *p<2.2e-16 W=123662,* UC *p=1.614e-14 W=38031,* UC-CD *p=0.7831 W=27261*). When we examined changes in *garL* between IBD patients and non-IBD individuals, we found that *garL* genes and transcripts were increased in CD patients, but not in UC patients compared to healthy controls (Mann Whitney U Test, MGX: CD *p=8.808e-05 W=115671*, UC *p=.07631 W=56227*, UC-CD *p=.03518 W=102641*, MTX: CD *p=1.752e-08 W=99977,* UC *p=0.2358 W=57046,* UC-CD *p=0.005956 W=28776).*

3.3 Discussion

In this study, we have shown that disparate taxa recapitulate the metabolism of glucarate and galactarate utilizing enzymatically divergent, yet functionally equivalent, *gud/gar* pathways. We identified this alternative enzymatic pathway in *E. clostridioformis* and demonstrated that the homologous reductase *garR,* uncharacterized transporter *garABC,* and putative 5-KDG aldolase *gudL* are able to complement the relevant *E. coli gud/gar* knockouts. Based on the discovery of this parallel pathway, we systematically searched for commensal species with the *gud/gar* pathway, and identified 887 species, spanning multiple phyla, with the minimum viable pathway. From this, we can conclude that the *gud/gar* pathway, which was previously considered a pathogen colonization factor, is present in many commensals, often via the alternative enzymatic pathway seen in *E. clostridioformis.*

Functional complementation of *E. coli* knockouts by the divergent *E. clostridioformis* genes reveals intriguing gaps in our understanding of oxidized sugar metabolism. *E. coli* Δ*garP* mutants show attenuated, rather than abolished, metabolism of galactarate, implying alternative methods of galactarate transport. GudP may be capable of importing galactarate given the structural similarity and the known promiscuity of GudD to dehydrate both glucarate and

galactarate.[110] Furthermore, the *E. clostridioformis garABC* rescues both Δ*gudP* and Δ*garP* mutants, suggesting a single ABC transporter mediates glucarate and galactarate uptake in *E. clostridioformis,* unlike the unique *E. coli* symporters.[111]

Of particular note is our discovery of the novel *E. clostridioformis* aldolase GudL*,* which shares a function with the *E. coli* aldolase GarL, even though they do not share a recent common ancestor. Complementation of an *E. coli* Δ*garL* knockout with *E. clostridioformis gudL* can fully rescue the defects in oxidized sugar metabolism, implicating *gudL* as a putative novel 5-KDG aldolase. However, the divergent structures of GudL and GarL, along with a lack of significant sequence similarity, suggests that they likely do not share a recent common ancestor. The presence of similar catalytic activities in two evolutionarily unrelated aldolases reflects a potential case of convergent evolution. Previous studies have documented cases of convergent evolution in aldolases, such as L- and D- threonine aldolases, across various bacterial species.[120]

Interestingly, although GudL and GarL are distinct, *E. coli* and *E. clostridioformis gud/gar* enzymes GudD, GarD, and GarR share high sequence similarity. We speculate that there could be horizontal gene transfer of these *gud/gar* genes between Bacillota and Pseudomonadota, although the directionality is unclear. However, our phylogenetic and gene cluster analyses do provide evidence for the direction of the gut adaptive horizontal gene transfer of the *gud/gar* pathway between Fusobacteriota and Bacillota. The *garR* and *garD* gene trees show the Fusobacteriota clade nested within the Bacillota clade. Combined with our synteny analyses, this suggests that Fusobacteriota acquired the *gud/gar* pathway from Bacillota. Past literature shows that horizontal gene transfer between *Fusobacterium nucleatum* and Bacillota is prevalent, and can be attributed to their close proximity in the oral cavity.[121,122] Inflammation and oxidative stress are reported to be involved in the initiation of colorectal cancer, and IBD patients are more likely to get colorectal cancer.[123] Given this, the potential involvement of *Fusobacterium* in sugar oxidation is an important finding, as the presence of *Fusobacterium* in the gut is correlated with colon cancer and increased in patients with Crohn's disease.[119,124,125]

Between the ABC transporter and novel *gudL* aldolase, we observe that divergent phyla have recapitulated the same function using alternative enzymes. Using the different operon organization and non-homologous genes identified in *E. clostridioformis* and *E. coli,* we have expanded the annotation of a function to include 887 additional gut microbial species. However, given the non-homologous aldolase we identified, it is possible that there are additional pathways for the metabolism of oxidized sugars that were not found during our search.

However, the strain-specificity of *gud/gar* introduces complexities when evaluating contributions to shifts in the microbial community. Taxonomic profiling of the microbiome has, at best, species-level resolution, precluding definitive conclusions about whether the enriched strains contain *gud/gar* and therefore metabolize oxidized sugars. Similarly, though *gud/gar* transcripts were increased in IBD, limited read coverage prevented evaluation of changes in transcriptional regulation.[126] The regulation of *gud/gar* transcription during inflammation, particularly regulatory changes between flares and periods of remission, remains an open avenue of investigation.

Though persisting knowledge gaps hinder our comprehensive understanding of how *gud/gar* contributes to inflammation-driven perturbations of the gut microbiome, utilization of alternative carbon sources via oxidized sugar metabolism offers additional insight into potential mechanisms underpinning the increased abundance of several obligate anaerobic taxa. Many obligate anaerobes are decreased in IBD due to a lack of mechanisms to detoxify the reactive oxygen and nitrogen species produced during inflammation.[127–129] Other genera benefit from the increased oxidative and nitrosative stress.[26,97,98] However, *Blautia* and *Lachnoclostridium* are obligate anaerobes enriched during inflammation that cannot utilize nitrates as electron acceptors, deviating from the pattern of decreased abundance seen in most obligate anaerobes and leaving the mechanism behind their increased abundance a mystery.[24]

These phenomena suggest additional mechanisms at work behind the inflammation-induced microbial perturbations. Here, we provide a potential mechanism through which *Blautia* and *Lachnoclostridium* spp., alongside many other commensal species, are able to metabolize the oxidized sugars produced in the inflamed gut as orthogonal carbon sources, and demonstrate that the *gud/gar* genes providing this functionality are increased in abundance and transcription in IBD patients compared to non-IBD controls. Based on this evidence, we propose that the metabolism of oxidized sugars, which was previously considered a pathogen colonization factor, may also be one of the many metabolic adaptations by commensal species to the inflamed gut. Our findings implicate oxidized sugar metabolism as a potential mechanism contributing to inflammation related gut dysbiosis, and shed new light on microbial adaptation to the inflamed gut.

3.4 Materials and methods

3.4.1 Growth of anaerobic microbes

Bacterial strains were obtained from the NIH Biodefense and Emerging Infections Research Resources Repository (BEI). Strains were inoculated from a glycerol stock and grown under anaerobic conditions (90% N2, 5% CO2, 5% H2) at 37 °C in an anaerobic chamber (Coy Laboratory Products). Strains were grown in minimal brain–heart infusion (BHI) broth (Research Products International, B11000) without glucose, and supplemented with 2 g/L glucose or galactarate.

3.4.2 Fermentation assay

Following the method provided in Faber et al., the fermentation media consisted of 10 g/L peptone (Criterion, C6641), 10 g/L carbon source (no carbon source, glucose, glucarate, or galactarate), and 24 mg/L bromothymol blue (Sigma-Aldrich, 114421).[27] The media was corrected to a pH of 7.6 and autoclaved to sterilize. Transformed *E. coli* were shaken in LB media supplemented with 100 ug/mL carbenicillin (GoldBio, 00901C103) overnight at 37℃. 20 uL of the overnight culture was inoculated into 1 mL of fermentation media supplemented with 100 uM isopropyl β-d-1-thiogalactopyranoside (IPTG, GoldBio, 221105I2481) and 100 ug/mL carbenicillin. 200 uL triplicates or quadruplicates were pipetted into a clear-bottom 96 well acrylic plate (Celltreat, 229592), sealed with a Breathe Easy membrane (Electron Microscopy Sciences, 70536), and incubated at 25℃ for 48 hours. The endpoint absorbance at 460 nm and 620 nm were read with a SpectraMax M5 plate reader and subtracted as a reference wavelength.

3.4.3 Construct development

As controls, empty pCW-lic vectors were transformed into each knockout strain. The wild-type genes *gudP, garP, garL,* and *garR* from *E. coli* were also cloned into pCW-lic backbones under a *tac* promoter, and transformed into their respective knockout strains.

pCW-Clos constructs: To ectopically express *garABC, garL,* and *garR* from *E. clostridioformis* in Keio *E. coli* mutants, the gene(s) of interest were amplified and inserted into a pCW-lic vector backbone (Addgene, 26908). A polymerase chain reaction was performed on *E. clostridioformis* genomic DNA via polymerase chain reaction (PCR) using Phusion High Fidelity DNA Polymerase (NEB, M0530S) and the primers listed in Supplementary Table 2. The product was purified using a Monarch PCR & DNA Cleanup Kit (NEB, T1030S). The pCW-lic backbone was cut with restriction enzymes NdeI (NEB, R0111S) and KpnI-HF (NEB, R3142S) and purified with a Monarch PCR & DNA Cleanup Kit. A Gibson assembly was performed with Gibson Assembly Master Mix (NEB, E2611S) according to manufacturer's protocols and stored at -20°C until use.

pCW-Keio constructs: The gene of interest (*gudP, garP, garL,* or *GarR)* was amplified off of a wild type Keio *E. coli* genome via and ligated into the pCW-lic vector using the same protocol as noted above, with primers listed in Supplementary Table 2.

pCW-Fuso constructs: The genes of interest (*gudL, garR)* from *Fusobacterium nucleatum* W1481 were ordered in a pCW-lic backbone from GenScript (https://www.genscript.com/).

Chemical competency: Keio collection *E. coli* were made chemically competent using Mix & Go! Transformation Kit and Buffer (Zymo, T3001) according to the manufacturer's protocol and maintained at -80°C until use.

Transformation: Each construct was individually transformed into chemically competent Keio collection *E. coli* according to the manufacturer's protocol (Zymo, T3001). The transformed cells were plated onto Luria-Bertani (LB) plates supplemented with 100 ug/mL carbenicillin to select for transformed colonies. Appropriate transformation was verified through Oxford Nanopore sequencing by Plasmidsaurus.

3.4.4 Bioinformatic analysis

Identification of gud/gar pathway in the GTDB genomes: All representative genomes from the Genome Taxonomy Database (GTDB) (release r207) (Parks et al., 2022) were downloaded, and protein sequences for each genome were predicted using Prokka (version 1.14.6) (Seemann, 2014). The HMM profiles for garD, garL, garR, gudD, and gudL were constructed. For garK, we used

TIGR00045 retrieved from TIGRFAM.[130] Corresponding e-values were employed to search for homologs in the GTDB proteins using ProkFunFind. Details on the HMM profiles and e-values can be found at the provided GitHub link (https://github.com/frikinzi/sugar_oxidation_bioinfo).

The hits were filtered based on a maximum 1e-100 identity threshold, resulting in putative 3,366 gud/gar sequences from 887 representative species. A genome was considered to have the minimum viable *gud/gar* pathway if it had GudD or GarD, GarR, and GudL or GarL.

Structural prediction and molecular docking: The structure for the putative *E. clostridioformis* GudL protein was predicted using AlphaFold2 (v2.3.0). Binding pockets were predicted using fpocket (v4.0) with default parameters. The pockets were compared to the homologous *Thermotoga maritima* dihydrodipicolinate synthase (3PB2) to identify putative substrate binding regions and catalytic residues.[131] The structure for 5-KDG (PubChem compound identifier: 5288442) was docked onto the predicted GudL structure using AutoDock Vina (v4.2).[132] The docking simulation was performed within 20 Å × 20 Å × 20 Å cubes centered on the centre points of the chosen fpocket substrate binding pocket with exhaustiveness set to 32. Docking results were visualized using PyMOL.[133] Catalytic residues were chosen based on previous studies on DapA. The predicted GudL protein structure was aligned with the *T. maritima* 3PB2 protein using TM-Align.[134] Protein sequence conservation of GudL was visualized using ConSurf based on the putative GudL clade (https://consurf.tau.ac.il/consurf_index.php).

Profiling gud/gar pathway abundance in human gut: To characterize the abundance of gud/gar transcripts in IBD compared to non-IBD, Human metagenomic and metatranscriptomic reads from HMP2 and HPFS were downloaded from the NCBI SRA and mapped to a reference database constructed with these 3,366 gud/gar genes using Bowtie2 v2.5.1. We filtered out samples with less than 5 million total reads. The number of reads mapped to the reference were normalized by the total number of reads per sample and then multiplied by 1 million to give counts per million (CPM). Patients with IBD were categorized based on whether they had Crohn's disease or ulcerative colitis. Using the Shapiro-Wilk normality test, we determined that both the MTX and MGX counts data did not follow a normal distribution (Shapiro-Wilk normality test: MGX W = W = 0.40175, p-value < 2.2e-16, MTX W = 0.32301, p-value < 2.2e-16). Therefore, we used a Mann–Whitney U test for each comparison among the 3 disease categories to see if counts of the *gud/gar* pathway genes significantly differed between the different groups. We did not correct the

p-value, and decided to not normalize based on species abundance because of the low number of reads mapped to the reference.

Phylogenetic analyses: Amino acid alignments for GudD/GarD/GarR/GarL/GudL found in the GTDB database were generated using ClustalW v2.1 using the BLASTp hits generated by ProkFunFind.[135] If the same genome had multiple hits for a gene, we selected the gene that had the lowest e-value to either the *E. clostridioformis* or *E. coli* gene. This resulted in 887 sequences in the GarR alignment, 809 sequences in the GarD alignment, 782 sequences in the GudD alignment, 507 GarL sequences, and 380 GudL sequences. The multiple sequence alignment was then trimmed using Goalign v0.3.7 by removing positions with over 97% gaps.[136] Gene trees for GudD, GarD and GarR were created from these trimmed alignments using IQ-Tree v2.2.0.3, with default parameters and 1000 bootstraps.[137] Ancestral sequence reconstruction was performed on the GudL tree using GRASP (http://grasp.scmb.uq.edu.au/), with default parameters.

Chapter 4: SpiR is a gut microbial enzyme that drives cholesterol conversion

This chapter contains material currently under review at Nature Microbiology in SpiR is a gut microbial enzyme that drives cholesterol conversion, *which was a joint work with Gabriela Arp, Angela K. Jiang, Keith Dufault-Thompson, Aoshu Zhong, Maggie Grant, Yue Li, Xiaofang Jiang, and Brantley Hall. S.L. developed the derivatization methods, performed the whole cell and bacterial lysate experiments, optimized the LC-MS methods, and contributed to writing and editing the manuscript.*

S.L., X.J., B.H., and G.A. contributed to the conceptualization of the study. S.L. and G.A. developed the methodology and performed validation. Formal analysis was conducted by A.K.J. and X.J., while investigation was carried out by S.L., G.A., and A.K.J. Data curation was performed by A.K.J. and X.J. The original draft of the manuscript was written by S.L., K.D.T., G.A., A.K.J., and X.J. with subsequent review and editing by S.L., K.D.T., A.Z., M.G., G.A., A.K.J., X.J., and B.H. Visualization efforts were led by G.A., X.J., and A.K.J. Supervision was provided by Y.L., B.H. and X.J.

4.1 Introduction

Hypercholesterolemia is a major risk factor for cardiovascular disease and the leading cause of global mortality. While diet and genetics are major contributors to elevated serum cholesterol levels, the gut microbiome is being increasingly recognized as a key aspect of cholesterol homeostasis. Gut bacteria affect cholesterol metabolism by converting cholesterol into coprostanol, a non-absorbable sterol that can be excreted as a waste product, thereby reducing serum cholesterol.[138] Approximately one gram of cholesterol enters the intestine daily from a combination of diet, bile, and epithelial turnover; 50–60% is reabsorbed, while the rest is excreted or metabolized by the gut microbiota.[139]

An inverse correlation between fecal coprostanol-to-cholesterol ratios and serum cholesterol levels has been reported, suggesting that individuals with efficient gut microbial cholesterol conversion tend to have lower circulating cholesterol.[140] Early evidence of a cholesterol-lowering effect associated with microbial cholesterol metabolism was found in the 1980s.[141] Further studies confirmed this relationship, showing that individuals classified as low

converters of cholesterol to coprostanol often have higher serum cholesterol levels than those classified as high converters.[142] The microbial origin of this process is reinforced by the finding that germ-free animals do not convert cholesterol to coprostanol, and exhibit higher circulating cholesterol levels.[143] Similarly, antibiotic treatment has been shown to reduce cholesterol-to-coprostanol conversion and correspondingly increase serum cholesterol levels.[144] To further support microbial involvement, a study found the abundance of coprostanoligenic bacteria in the gut is associated with greater cholesterol conversion efficiency and lower serum cholesterol levels.[145] These findings underscore the potential of microbial cholesterol metabolism to act as a regulatory mechanism influencing serum cholesterol levels, and highlight the need to elucidate the genetic and enzymatic basis of this process.

The first bacterium capable of converting cholesterol to coprostanol was identified in 1973 when the *Eubacterium* strain ATCC 21,408 was isolated from rat cecum.[146] Subsequent studies led to the isolation of cholesterol-reducing bacteria from humans,[147] baboons,[148] and hog waste lagoons.[149] These bacteria were identified as strictly anaerobic, non-spore-forming, gram-positive bacilli and were grouped under the genus *Eubacterium* based on their shared morphological and physiological characteristics.[148] It has been hypothesized that they form a monophyletic clade because of their conserved sterol-reducing metabolism, although genetic confirmation remains impossible, as most strains were isolated before sequencing was available and have since been lost.[146] Unlike most gut microbes, cholesterol-reducing *Eubacterium* species depend on Δ^5-3β-hydroxy steroids as essential metabolic substrates and do not proliferate in sterol-deficient environments.[150] Their metabolic activity suggests that cholesterol serves as a terminal electron acceptor, supported by their strict requirement for high cholesterol concentrations and selective biohydrogenation of the 5,6-double bond.[146] The transformation follows an indirect hydrogenation pathway, converting cholesterol to cholestenone, then to coprostanone, and finally to coprostanol[151]. This process is highly efficient, with some strains converting over 90% of the cholesterol to coprostanol[151]. Given the prevalence of *Eubacterium* species in the gut and their metabolic specialization, these bacteria are considered the primary contributors to cholesterol biohydrogenation in mammalian hosts[152].

Identifying its cholesterol-metabolizing enzymes is essential for understanding microbial contributions to host cholesterol metabolism. *Eubacterium coprostanoligenes* ATCC 51222, the only cholesterol-reducing *Eubacterium* species with a sequenced genome and available isolates, serves as a key model for elucidating the enzymatic basis of cholesterol biohydrogenation[149]. The

ismA gene, found in *E. coprostanoligenes* ATCC 51222, was proposed to catalyze cholesterol oxidation to cholestenone and reduction of coprostanone to coprostanol[153]. While metagenomic and biochemical evidence supports its involvement, some coprostanol-positive samples lack *ismA*, suggesting that additional enzymes may contribute to this process[153].

Here, we identified and characterized SpiR (Sterol processing isomerase and Reductase), a steroid $\Delta^{5\text{-}4}$ isomerase/3-keto reductase from *Eubacterium coprostanoligenes*, which catalyzes the first and obligate step in gut microbial cholesterol metabolism. Biochemical assays revealed that SpiR oxidizes cholesterol and related steroids with high specificity and stereoselectivity, functions as an NAD(H)-dependent homodimer, and exhibits stronger substrate affinity than the previously implicated enzyme IsmA. Phylogenetic analysis showed that SpiR clusters with steroid $\Delta^{5\text{-}4}$ isomerases from both prokaryotic and eukaryotic organisms. While *spiR* homologs are taxonomically restricted to a clade of uncultured *Acutalibacteraceae*, metagenomic surveys have revealed that they are broadly prevalent in human gut microbiomes. Across three independent cohorts integrating metagenomic and metabolomic data, the presence of *spiR* homologs was consistently tracked with coprostanone production and fecal cholesterol depletion, including samples lacking *ismA*. Classifiers based on *spiR* outperformed *ismA*-based models in predicting cholesterol conversion (AUC = 0.95 vs. 0.81), resolving prior inconsistencies and establishing SpiR as a more robust biomarker. These findings redefine the enzymatic basis of microbial cholesterol metabolism and highlight SpiR as a mechanistic driver and predictive marker of microbiome-mediated cholesterol homeostasis.

4.2 Results

4.2.1 SpiR is a putative steroid $\Delta^{5\text{-}4}$ isomerase/3-keto reductase involved in cholesterol metabolism in *Eubacterium coprostanoligenes*

To identify the key genetic components involved in cholesterol metabolism in *E. coprostanoligenes*, we investigated homologs of known steroid-transforming enzymes encoded in its genome. In our previous study, we identified a gene frequently fused with steroid 5β-reductase, which facilitated the conversion of pregnenolone to tetrahydropregnenolone. The conversion of pregnenolone to progesterone is catalyzed by 3β-hydroxysteroid

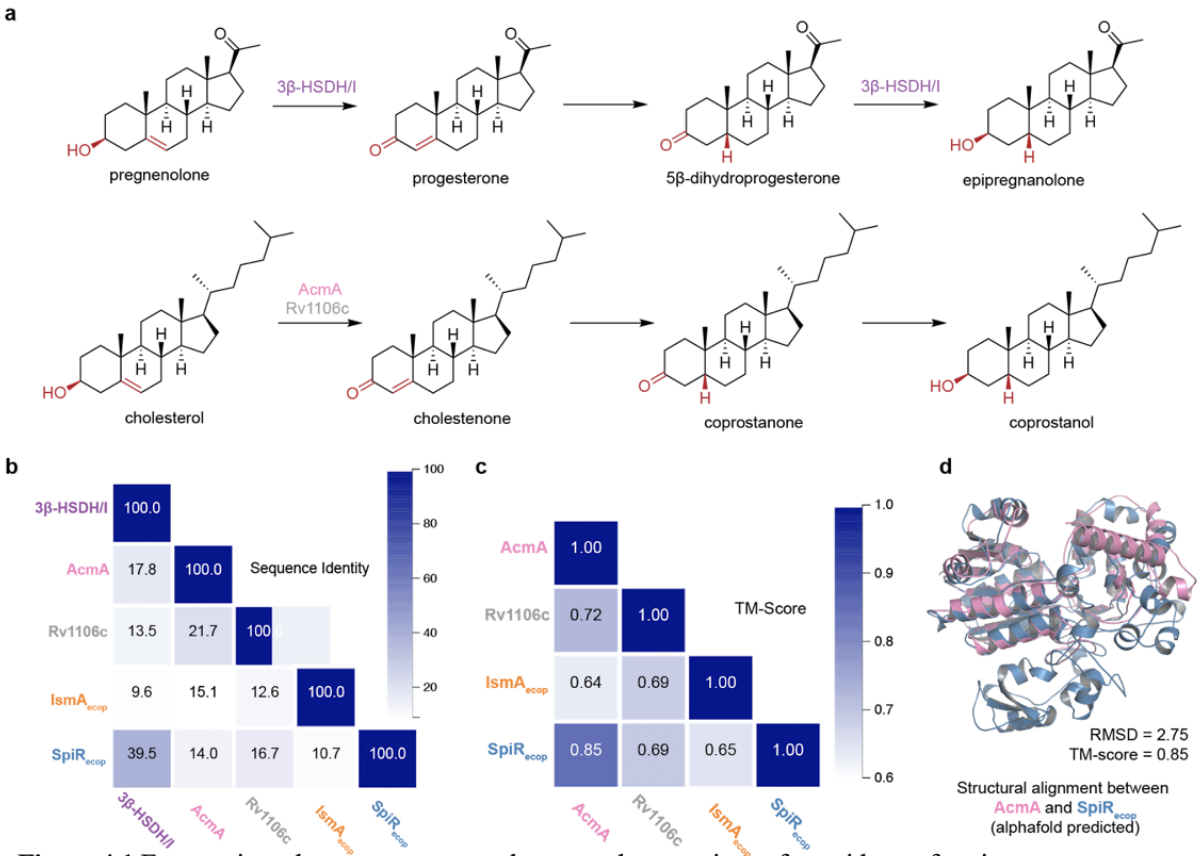

Figure 4.1 Enzymatic pathway, sequence, and structural comparison of steroid-transforming enzymes. **a)** Enzymatic pathways for the conversion of pregnenolone to epipregnanolone (top), and cholesterol to coprostanol (bottom). Known enzymes catalyzing the substeps of the reaction are indicated above. **b)** Pairwise sequence identity matrix of Rossmann fold-containing enzymes. Values indicate percent identity, with deeper blue coloration representing higher sequence identity. **c)** Predicted structural identity matrix of enzymes that may act on cholesterol. Values indicate the TM score, with deeper blue coloration representing higher structural similarity. **d)** Alphafold-predicted structural alignment of AcmA (pink) and SpiR (blue).

dehydrogenase/isomerase (3β-HSDH/I). Given the close resemblance between pregnenolone-to-progesterone and cholesterol-to-cholestenone conversions, we hypothesized that if *E. coprostanoligenes* contain an enzyme homologous to 3β-HSDH/I, they may catalyze a similar reaction in cholesterol metabolism (Figure 4.1a).

Through our search for homologs of 3β-HSDH/I, we identified SpiR (WP_242941742.1), which shares 39.5% protein sequence identity with the verified 3β-HSDH/I (dw0523) (Figure 4.1b). SpiR is distinct from IsmA, which has been previously linked to cholesterol dehydrogenation, with only 10.7% identity between the two proteins (Figure 4.1b). SpiR and IsmA belong to the SDR family, which comprises a functionally diverse group of oxidoreductases characterized by a single domain with a Rossmann fold. Other enzymes that possess the Rossmann fold and play a role in cholesterol metabolism include AcmA from

Sterolibacterium denitrificans and Rv1106c from *Mycobacterium tuberculosis*[154,155]. These enzymes catalyze the oxidation of cholesterol to cholestenone in an oxygen-independent manner[156,157]. Although neither AcmA nor Rv1106c shares high sequence similarity (>25%) with 3β-HSDH/I or SpiR, their predicted structural similarities suggest that SpiR plays a role in cholesterol metabolism (Figure 4.1c). Specifically, AcmA exhibits a high degree of structural similarity to SpiR, suggesting that SpiR performs an analogous function in cholesterol metabolism within *E. coprostanoligenes* (Figure 4.1d). Given its sequence similarity to 3β-HSDH/I, high structural homology to AcmA, and presence within *E. coprostanoligenes*, we hypothesized that SpiR functions as a steroid $\Delta^{5\text{-}4}$ isomerase/3-keto reductase.

4.2.2 SpiR catalyses the conversion of cholesterol to cholestenone

We investigated whether SpiR catalyses the conversion of cholesterol to cholestenone. To assess SpiR activity, we conducted assays using cell lysates from *E. coli* heterologously expressing SpiR incubated with cholesterol, followed by liquid chromatography-mass spectrometry (LC-MS) analysis. Prior to the activity assay, we confirmed the expression of recombinant SpiR and IsmA by SDS-PAGE. Distinct bands at the expected molecular weights of approximately 72 kDa for SpiR and 33 kDa for IsmA were observed in the induced samples but were absent in lysates from cells carrying an empty vector. These bands were also present in the post-lysis supernatant fractions, indicating that both proteins were expressed in soluble form.

We utilized electrospray ionization (ESI) mass spectrometry to assay cholestenone production by transformed E. coli strains expressing SpiR and IsmA. To improve the detection of cholestenone, we derivatized it with Girard's Reagent P, as has been done in previous studies[158,159](Figure 4.2a). *E. coli* lysates containing an empty vector served as a control and showed no cholesterol conversion, whereas those expressing SpiR efficiently produced cholestenone (Figure 4.2b). These results confirm that SpiR functions as a cholesterol oxidase. A comparison of cholestenone signals between *E. coli* lysates expressing IsmA and SpiR revealed

that SpiR generated a higher cholestenone signal, with a marginally significant difference (U = 9.0, p = 0.05, Mann-Whitney U test).

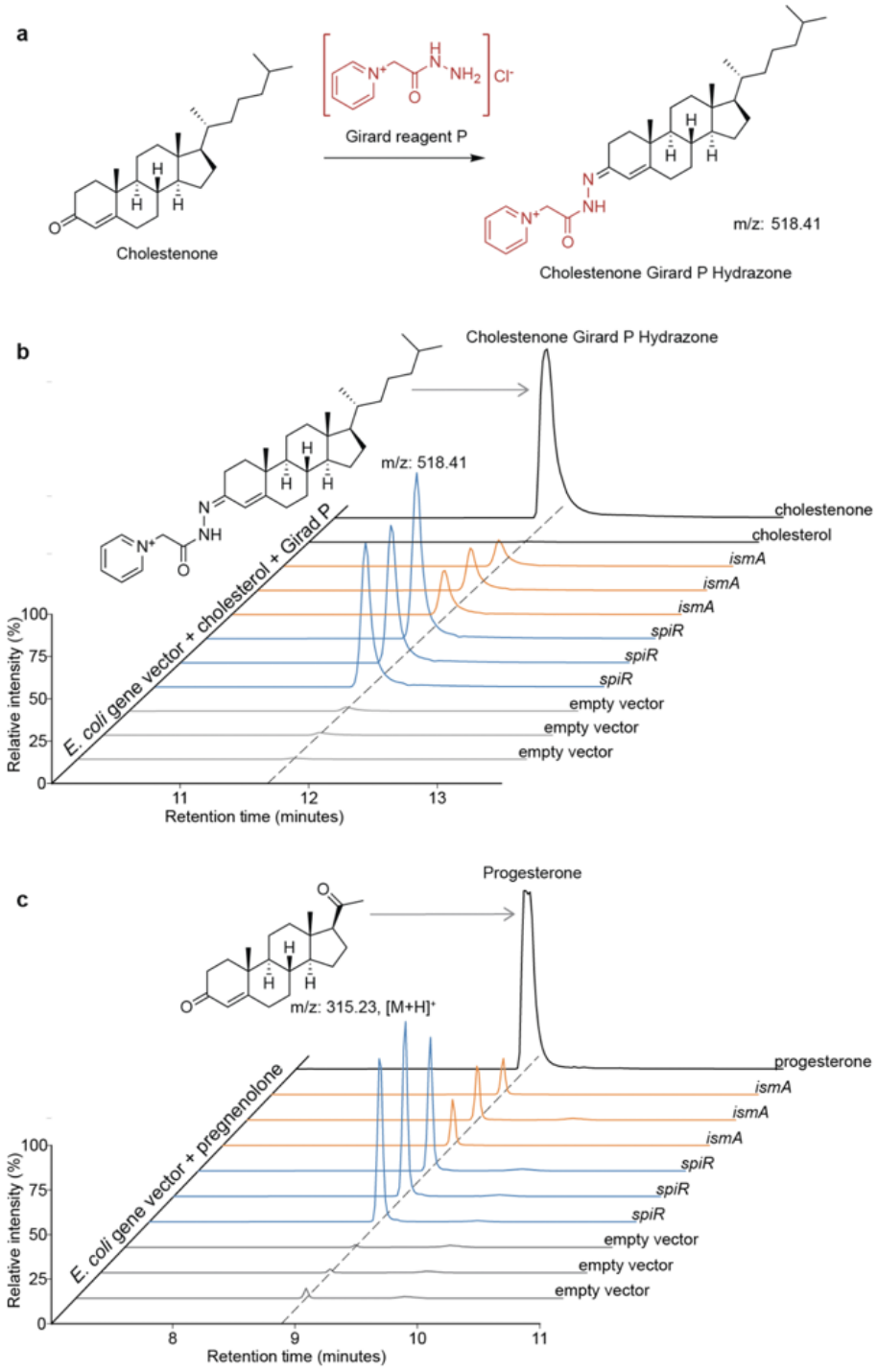

Figure 4.2 LC-MS detection of cholesterol- and pregnenolone-transforming activity. **a)** Derivatization of cholestenone to cholestenone hydrazone using Girard's Reagent P. **b)** Chromatogram showing the detection of cholestenone hydrazone in derivatized extracts of *E. coli* expressing a vector control (gray), *spiR* (blue), or *ismA* (orange) incubated with cholesterol. Standards of derivatized cholesterol and cholestenone are shown in black. **c)** Chromatogram showing the detection of progesterone in extracts of *E. coli* expressing a vector control (gray), *spiR* (blue), or *ismA* (orange) incubated with pregnenolone. Progesterone standard is shown in black.

4.2.3 SpiR catalyses pregnenolone-to-progesterone conversion, indicating broader steroid activity

The ability of *E. coprostanoligenes* to convert pregnenolone to progesterone has been previously reported[151]. We hypothesized that SpiR, which has been shown to catalyze cholesterol oxidation, may be responsible for this activity. To test this hypothesis, we evaluated whether SpiR catalyzes the conversion of pregnenolone to progesterone. Consistent with our observations for cholesterol, *E. coli* expressing SpiR converted pregnenolone to progesterone, whereas negative controls containing an empty vector showed no conversion (Figure 4.2c). These results confirm that SpiR catalyzes 3β-hydroxysteroid oxidation and $\Delta^{5\text{-}4}$ isomerization, which are pivotal reactions in the microbial metabolism of steroid hormones and support the conclusion that SpiR plays a broader role in microbial steroid metabolism.

4.2.4 SpiR reduces 3-keto steroids into stereospecific 3β-hydroxyl-steroids

To assess the enzymatic activity and stereospecificity of SpiR, we investigated its ability to catalyze the reduction of the 3-keto group in dihydroprogesterone to form tetrahydroprogesterone (Figure 4.3a). Because 3-keto reduction can yield either 3α- or 3β-hydroxysteroid products, we used LC-MS with a phase column capable of resolving the four major stereoisomers: allopregnanolone (3α,5α), isopregnanolone (3β,5α), pregnanolone (3α,5β), and epipregnanolone (3β,5β).

When *E. coli* heterologously expressing SpiR was incubated with either 5α-dihydroprogesterone or 5β-dihydroprogesterone, the predominant product detected matched the retention times of the corresponding 3β-hydroxy isomers (isopregnanolone and epipregnanolone, respectively)(Figure 4.3b,c). These results demonstrate that SpiR functions as a 3-keto reductase, reducing both 5α- and 5β-dihydroprogesterone to 3β-hydroxy tetrahydroprogesterone with consistent stereochemistry across divergent substrate conformations.

Together, these findings establish SpiR as a multifunctional enzyme capable of catalyzing three key transformations in microbial steroid metabolism: oxidation of the 3β-hydroxyl group in cholesterol and pregnenolone, $\Delta^{5\text{-}4}$ double-bond isomerization, and stereospecific reduction of 3-keto groups to 3β-hydroxy configurations.

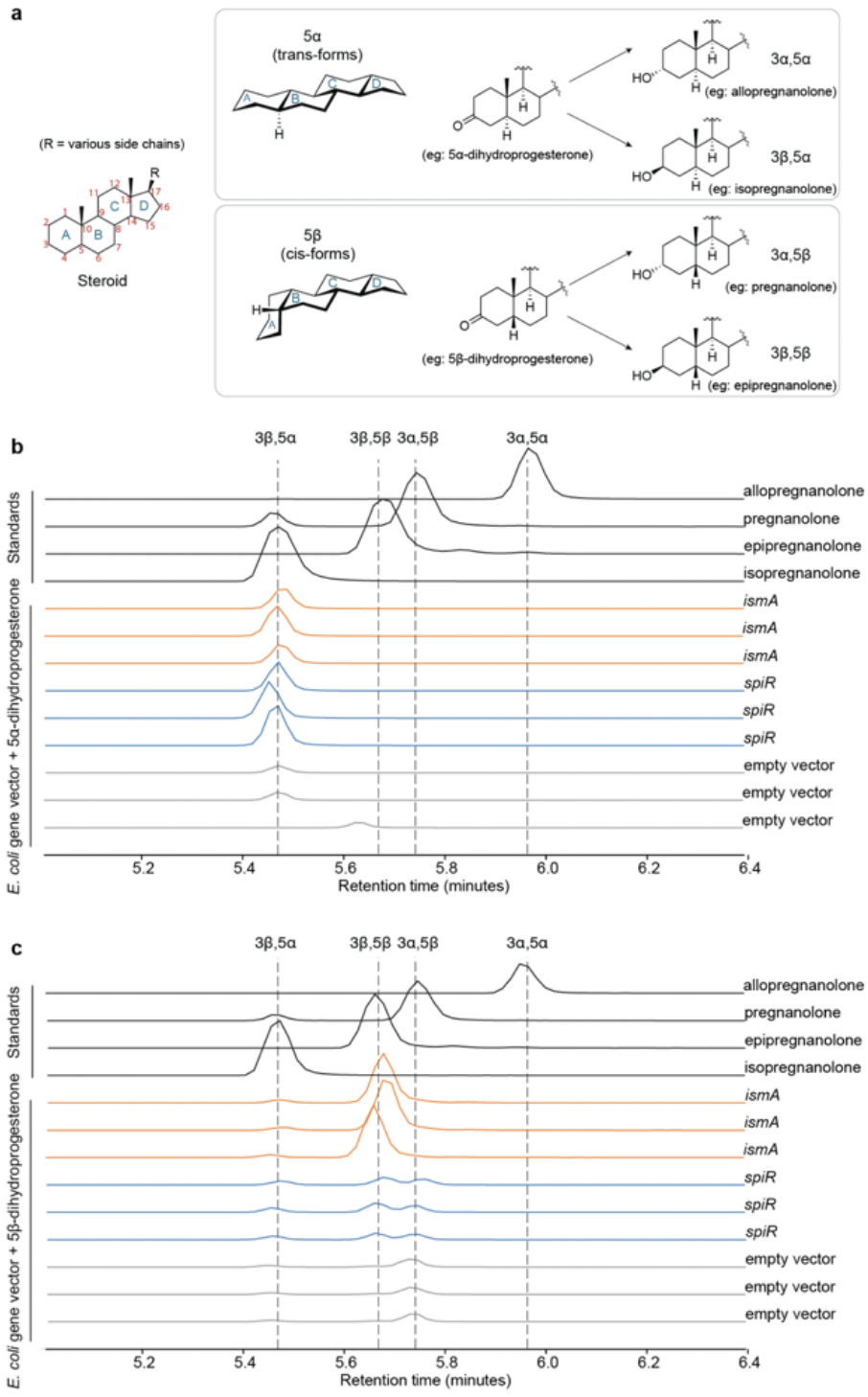

Figure 4.3 LC-MS detection of stereospecific 3β-reduction of steroids. **a)** Structure of the steroid core showing carbon numbering and ring designations (A–D). Right: stereoisomers of tetrahydroprogesterone derived from 5α- and 5β-dihydroprogesterone. **b)** LC–MS analysis of steroid products from *E. coli* expressing empty vector (grey), *ismA* (orange), or *spiR* (blue) incubated with 5α-dihydroprogesterone or **c)** 5β-dihydroprogesterone. Authentic standards (black) define the retention times for 3α,5α (allopregnanolone), 3β,5α (isopregnanolone), 3α,5β (pregnanolone), and 3β,5β (epipregnanolone) isomers.

4.2.5 SpiR preferentially binds to cholesterol

To characterize the substrate specificity of SpiR, we evaluated its binding affinity for the key substrates involved in these transformations: cholesterol, coprostanone, pregnenolone, and dihydroprogesterone. Isothermal titration calorimetry (ITC) experiments revealed a higher binding affinity for cholesterol (Kd = 539 nM) and coprostanone (Kd = 660 nM) (Figure 4.4b,c) than for homologous steroid hormones. Pregnenolone, which differs from cholesterol only in the absence of a hydrocarbon tail at C17, exhibited 6.12-fold weaker binding affinity (Kd = 3.3 µM). 5β-dihydroprogesterone, which lacks the same hydrocarbon tail present in coprostanone, bound SpiR with a 1.34-fold lower affinity (Kd = 887 µM). This suggests a substrate preference for

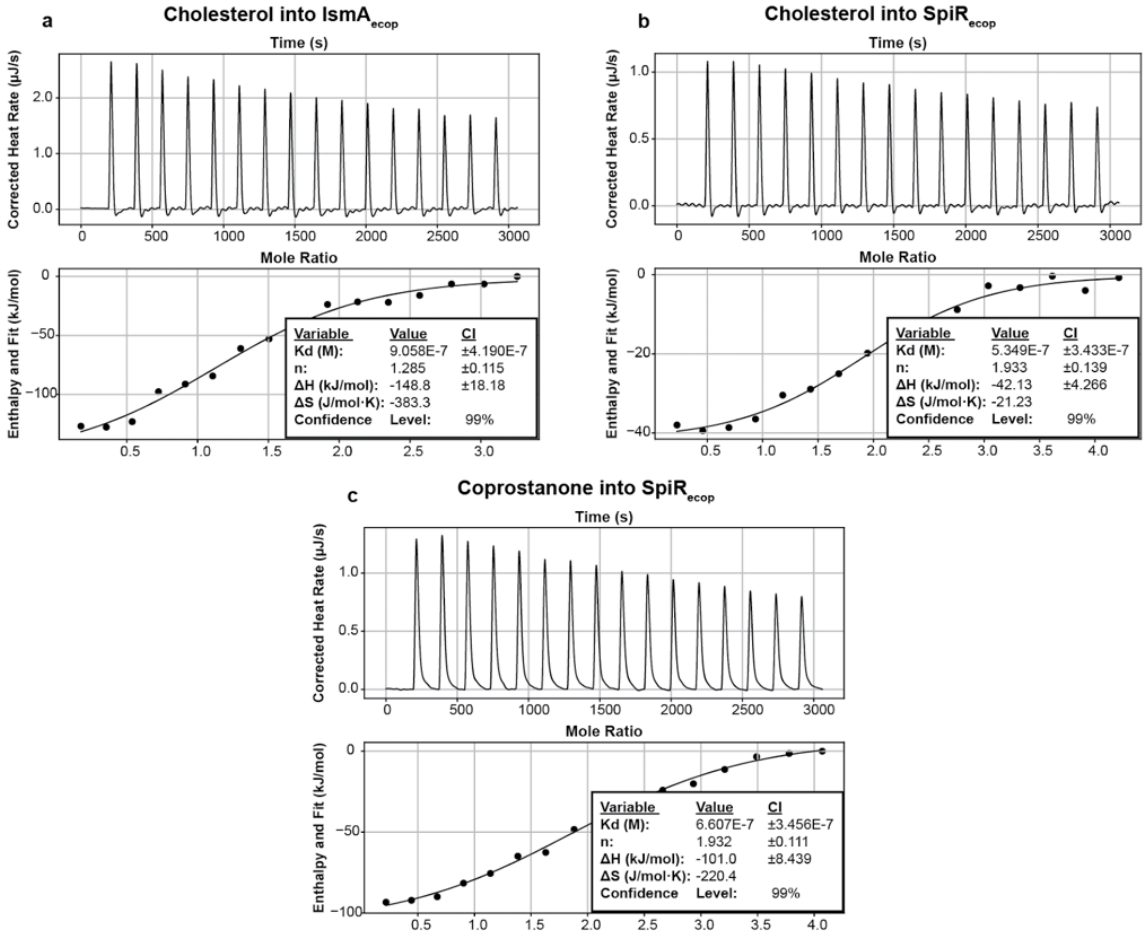

Figure 4.4 Steroid binding to IsmA and SpiR. Binding of IsmA to **a)** cholesterol and binding of SpiR to **b)** cholesterol and **c)** coprostanone as measured by ITC. Raw data (heats of binding) are shown in the upper graph of each panel, whereas integrated binding isotherms with their independent binding model fitting are shown in the lower panel. Recombinant IsmA and SpiR were placed in the calorimeter cell, and the experiment was performed at 25°C with successive 3.1µL injections of the ligand. Binding of cholesterol to IsmA, SpiR to cholesterol, and SpiR to coprostanone was performed using a 50 µm ligand and 5 µm protein. The thermodynamic parameters with standard errors were calculated based on each curve using the NanoAnalyze Software package.

coprostanone over 5β-dihydroprogesterone, although this preference is less marked than that between cholesterol and pregnenolone. These differences in binding affinity are likely the result of interactions between the enzyme and the functional groups decorating the steroid core, despite the similarities in the core structures. These results indicate that cholesterol and coprostanone are the preferred ligands of SpiR, underscoring their likely role in gut microbial cholesterol metabolism.

As IsmA was also found to metabolize cholesterol, we compared its cholesterol-binding affinity to SpiR. IsmA titrated with cholesterol showed 1.68-fold lower binding affinity than that observed between SpiR and cholesterol (Kd = 908 μM) (Figure 4.4a). This comparatively lower affinity suggests that even though IsmA may participate in cholesterol metabolism, SpiR is likely the more specialized enzyme for cholesterol reduction in the gut.

To further characterize the likely enzyme cofactors, we tested their binding to NADH, NADPH, and NAD^+. SpiR was able to bind to both NADH (Kd = 1.03 μM) with n = 1.813 ± 0.140 and NAD+ (Kd = 1.14 μM) with n = 1.974 ± 0.120, exhibiting a comparable binding affinity for NADH and NAD^+ as cofactors. However, no binding was observed between SpiR and NADPH. Both substrate and cofactor binding were further supported by NADH/NAD^+ oxidation-reduction assays, which confirmed that SpiR and IsmA are biologically active and reinforced ITC results demonstrating that SpiR binds NADH and NAD^+. These results align with the cofactors bound to similar reductase enzymes and validate our prediction of the cofactors necessary for the oxidation and reduction of enzyme substrates[160–162].

Cholesterol, coprostanone, pregnenolone, and 5β-dihydroprogesterone exhibited stoichiometric values of 1.933 ± 0.139, 1.932 ± 0.111, 1.970 ± 0.263, and 1.909 ± 0.149, respectively (Figure 4.4b,c). These values indicate that either two ligand molecules bind to a single enzyme monomer, or that the enzyme primarily exists as a dimer. Some enzymes can accommodate two substrates within the same binding site owing to π–π interactions within the aromatic core[163]. To investigate the oligomeric state of SpiR, a dynamic light scatterer (DLS) analysis was performed. The measured hydrodynamic radius was 3.76 nm, which is consistent with the theoretical radius of its dimeric state (3.68 nm) and larger than the expected radius of its monomeric form (2.94 nm)[164]. This indicates that SpiR is predominantly present as a dimer in the solution. Given that other hydroxysteroid dehydrogenase (HSDH) families have also been

observed to form homodimers[165,166], the stoichiometry and DLS results suggest that SpiR functions as a homodimer.

4.2.6 SpiR clusters phylogenetically with known steroid $\Delta^{5\text{-}4}$ isomerases within the SDR superfamily

The structural similarities between steroid substrates and their stoichiometric values suggest a conserved model of ligand-SpiR binding; therefore, we sought to identify its catalytic residues through structural modeling and ligand-docking simulations. Using the AlphaFold2-predicted structure of SpiR, we docked NAD^+ with cholesterol as a representative steroid substrate to define the putative active site. This analysis revealed that three residues, T126, Y155, and K159, are located in close proximity to the docked ligands (Figure 4.5a). These residues correspond to the canonical catalytic triad found in the SDR superfamily, in which tyrosine typically acts as a nucleophile, lysine as a general base, and threonine as a proton donor[167]. ConSurf analysis and multiple sequence alignment confirmed that these residues are

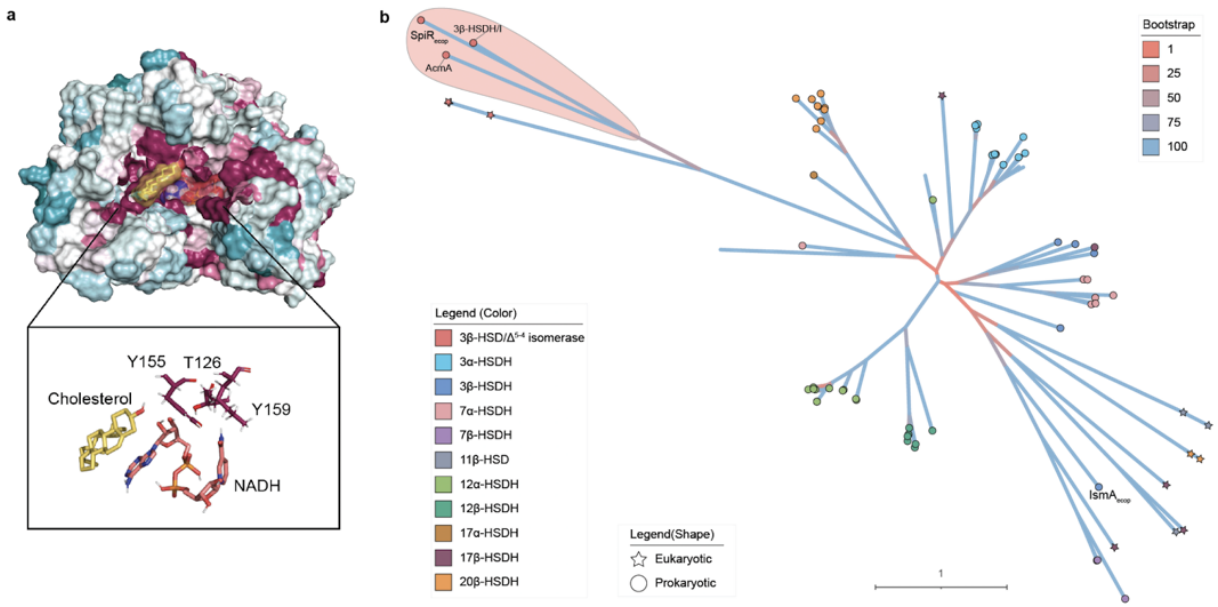

Figure 4.5 Structural conservation and phylogenetic reconstruction of SpiRs. **a)** The predicted AlphaFold2 structure of SpiR, colored according to the degree of conservation of amino acid residues based on ConSurf analysis. Red represents conserved regions and green represents variable regions. Docked cholesterol and NADH are shown on the structure (yellow and pink). Potential active-site residue interactions with cholesterol and NADH are shown, as predicted by Autodock vina. **b)** Phylogenetic tree representing evolutionary relationships between HSDH enzymes. Branch colors indicate bootstrap support, with cooler colors (e.g., blue) representing higher support and warmer colors (e.g., red) indicating lower support. Tip shapes differentiate taxonomy, where circles represent prokaryotic sequences and stars denote eukaryotic sequences. Tip labels are color-coded based on the hydroxyl functional group position on steroid substrates and stereochemistry.

highly conserved across SpiR homologs (Figure 4.5a). These findings suggest that SpiR uses a conserved SDR mechanism, with T126, Y155, and K159 forming the catalytic triad that is required for cholesterol oxidation.

To place SpiR in an evolutionary context, we compiled a curated dataset of biochemically characterized HSDHs. This included representative enzymes from previously published surveys of microbial HSDH diversity[166], as well as newly analyzed sequences of interest, including SpiR, AcmA, and related homologs. To broaden the functional and taxonomic scope of the phylogenetic analysis, we also incorporated additional enzymes not featured in the prior study, such as 17β-HSDH (DesG), 17α-HSDH (DesF), and 3β-HSDH/I[168–170]. By comparing SpiR with other HSDH enzymes, we elucidated the evolutionary context of the enzyme. Phylogenetic analysis revealed extensive diversity among the HSDH enzymes, indicating multiple evolutionary trajectories for steroid metabolism. SpiR, 3β-HSDH/I, and AcmA form a distinct clade that is a sister group to the eukaryotic 3β-HSD/$\Delta^{5\text{-}4}$ isomerase, suggesting a close evolutionary relationship between these prokaryotic and eukaryotic enzymes. AcmA catalyzes the conversion of cholesterol to cholestenone through $\Delta^{5\text{-}4}$ isomerization, a function that is also carried out by eukaryotic 3β-HSD/$\Delta^{5\text{-}4}$ isomerase. This clade is phylogenetically distant from the rest of the HSDH enzymes, including IsmA, which belongs to a more divergent and functionally distinct group. The separation of SpiR and its related enzymes from other HSDHs involved in bacterial sterol metabolism suggests distinct evolutionary pressures shaping their functions, with SpiR aligning closely with the known cholesterol isomerase, AcmA.

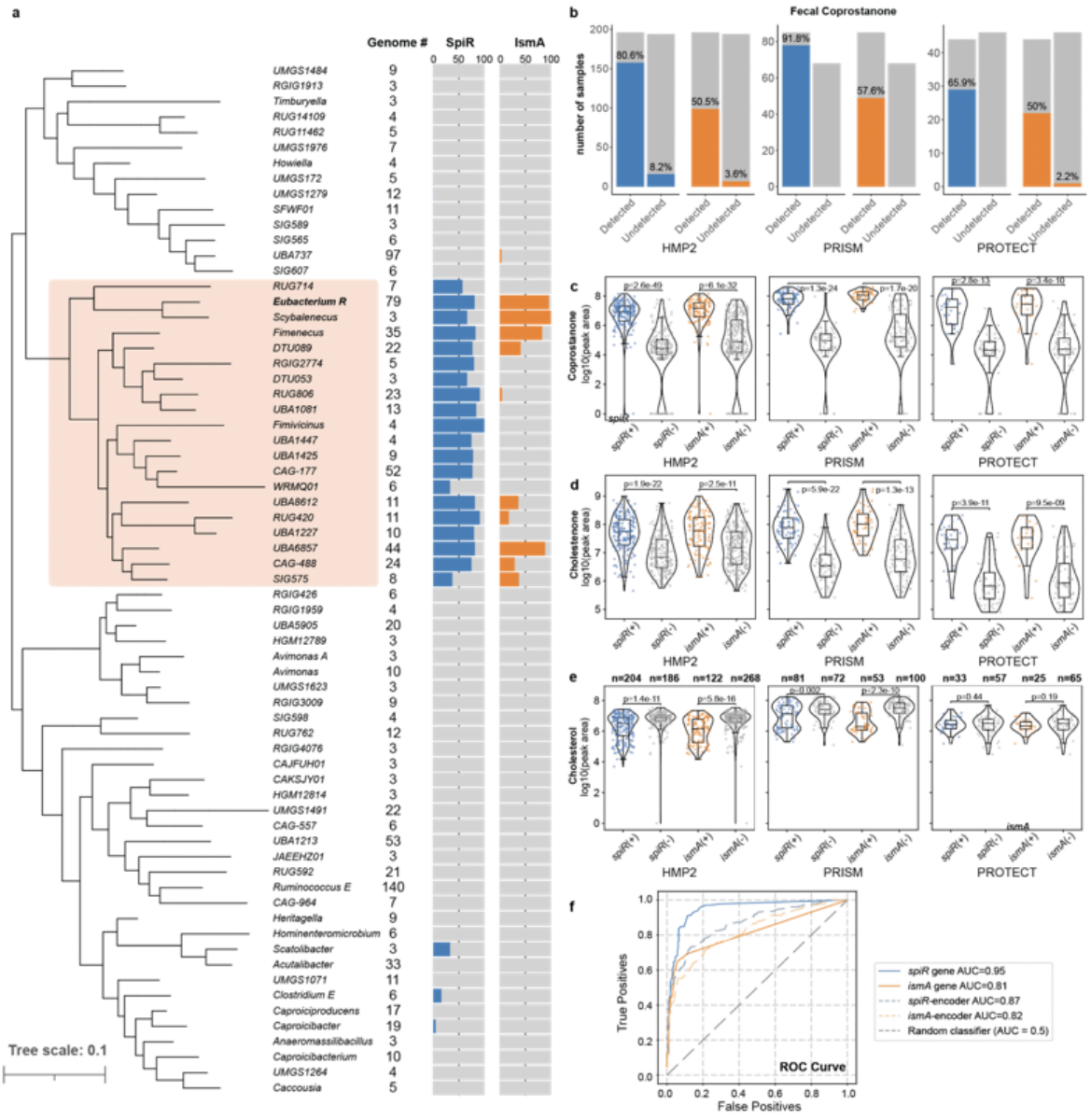

Fig 4.6 Genus distribution and presence of *spiR* in human gut metagenomes. **a)** Genus-level phylogeny of the family Acutalibacteraceae showing the number of genomes and percentage of species with *spiR* homologs and *ismA* homologs per genus. A clade containing species with both *spiR* homologs and *ismA* homologs features is highlighted in orange. The genus-level tree was generated by pruning the GTDB species tree using the gotree prune command and then collapsing the species by genus. Genera with fewer than three species were removed. **b)** Investigation of the association between *spiR* homologs, *ismA* homologs, and coprostanone formation in three paired human metagenomic and metabolomic datasets (HMP2, PRISM, and PROTECT). The presence of *spiR* was more strongly associated with coprostanone detection than *ismA*. **c)** The presence of *spiR* homologs or *ismA* homologs is associated with lower stool cholesterol levels. Each point represents a paired metabolomics–metagenomic sample. The y-axis represents the log10 area of the identified cholesterol peak area as a proxy for concentration. *P*-values represent the results of a one-sided Mann-Whitney U test (alternative=greater) comparing the cholesterol levels of samples with and without SpiR. **d)** and **e)** The presence of *spiR* homologs was more strongly associated with lower stool **d)** cholestenone and **e)** coprostanone levels than *ismA*. *P*-values represent the results of a one-sided Mann-Whitney U test (alternative=less) comparing the cholesterol levels of samples with and without SpiR. **f)** Receiver Operating Characteristic curve (ROC) for prediction of coprostanone detection in stool samples from all three cohorts (HMP2, PRISM, and PROTECT) based on the presence of *spiR* homologs, *ismA* homologs, *spiR*-containing species, or *ismA*-containing species with varying cpm thresholds (0–30 cpm for genes, 0–40,000 cpm for species). The presence of *spiR* homologs (AUC=0.95) showed the best predictive performance compared with the other curves.

4.2.7 SpiR homologs are encoded by uncultured Acutalibacteraceae

We performed a survey of the *spiR*-like gene family and identified them exclusively in bacteria from the Clostridia class and Oscillospirales order in the Acutalibacteraceae family (Figure 4.6a). Acutalibacteraceae are prevalent in anoxic gut environments and are linked to bile acid metabolism in the large intestine, where they play a role in deconjugation and transformation pathways[171]. The Acutalibacteraceae family is primarily represented by uncultured genera, consisting of nearly all genomes derived from metagenomes (96%, 1096 out of 1144) suggesting that much of the diversity within this family and the cholesterol-reducing microbial community has not been explored. Among the identified species, 308 contained *spiR* homologs and 180 possessed *ismA* homologs. Most *spiR* homologs (314/317) were encoded by species in one Acutalibacteraceae clade, and 148 species within this clade encoded both *ismA* and *spiR* homologs, indicating a significant overlap between the distribution of both gene types (Figure 4.5b). The restricted taxonomic distribution and frequent co-occurrence of *spiR* homologs and *ismA* homologs suggest that this clade represents a niche-adapted lineage specialized for cholesterol metabolism in the anaerobic gut, consistent with earlier reports describing non-spore-forming, strictly anaerobic *Eubacterium* species that depend on sterols as essential electron acceptors[146,148].

4.2.8 spiR homologs strongly correlate with cholesterol conversion across human gut microbiome cohorts

Many species encode both *spiR* homologs and *ismA* homologs, raising the possibility that previous associations between *ismA* and cholesterol conversion may have been confounded by *spiR*. Kenny et al. (2020)[153] found coprostanol in samples lacking *ismA*, implying that other enzymes may contribute. We hypothesized that *spiR* homologs may underlie this association and evaluated their link to cholesterol conversion. To evaluate this hypothesis, we analyzed paired metagenomic and metabolomic datasets, including the two cohorts used by Kenny et al. (2020)[153]: HMP2 (n=400)[172] and PRISM (n=151)[173], and included a new cohort, PROTECT (n=90)[174,175].

Our results suggest that *spiR* homologs are more strongly enriched in samples with detected cholesterol conversion than *ismA* homologs. Samples were grouped based on the presence ($\geq$ 1 CPM) or absence of *spiR* and *ismA* homologs in fecal metagenomes, as well as

the detection of cholesterol conversion in the fecal metabolome. Consistent with previous studies[176–179], we found that the samples had high variability in cholesterol conversion, with the overall count of converters (325 individuals) roughly equal to that of non-converters (308 individuals) (Figure 4.6b). There was a significantly higher prevalence of *spiR* homologs in converter samples than in non-converters across all three cohorts, as determined by one-sided Fisher's exact tests. In HMP2, the odds ratio was 43.57 (95% CI: 25.83–∞, $P = 1.80 \times 10^{-53}$); in PRISM, 408.18 (95% CI: 97.39–∞, $P = 1.36 \times 10^{-34}$); and in PROTECT, 55.11 (95% CI: 13.98–∞, $P = 1.54 \times 10^{-12}$). On average, 89.7% of converters harbored *spiR* homologs compared to only 6.18% of non-converters. Although *ismA* homologs were also significantly enriched in converter samples compared to non-converters, as determined by one-sided Fisher's exact tests (HMP2: OR = 22.16, 95% CI: 12.30–∞, $P = 6.29 \times 10^{-31}$; PRISM: OR = 109.80, 95% CI: 20.79–∞, $P = 7.74 \times 10^{-18}$; PROTECT: OR = 26.39, 95% CI: 6.85–∞, $P = 2.96 \times 10^{-8}$), the overall prevalence was lower than that of *spiR* homologs, with *ismA* homologs detected in only 52.2% of cholesterol converters on average. In addition, these trends were not confounded by disease status, with neither *spiR* homologs nor *ismA* homolog presence differing significantly across ulcerative colitis (UC), Crohn's disease (CD), and non-inflammatory bowel disease (non-IBD) samples**.** These patterns suggest that the presence of genes, not diseases, drives the observed differences in cholesterol metabolism.

Moreover, the presence of the *spiR* homologs was more strongly associated with increased levels of cholestenone and coprostanone in stool than *ismA* homologs across all three cohorts (HMP2, PRISM, and PROTECT). *spiR* homolog presence yielded robust strong associations with coprostanone (HMP2: $P = 2.6 \times 10^{-49}$; PRISM: $P = 1.3 \times 10^{-24}$; PROTECT: $P = 2.8 \times 10^{-13}$), exceeding those for *ismA* homologs (HMP2: $P = 6.1 \times 10^{-32}$; PRISM: $P = 1.7 \times 10^{-20}$; PROTECT: $P = 3.4 \times 10^{-10}$)(Figure 4.6c). Cholestenone followed the same trend, where those with *spiR* homologs tended to have lower cholestenone levels (one-sided Mann-Whitney U test, $P = 1.9 \times 10^{-22}$, 5.9×10^{-22}, and 3.9×10^{-11}, respectively) than those without the gene, whereas the corresponding *P*-values for *ismA* homologs were consistently higher (weaker associations: $P = 2.5 \times 10^{-11}$, 1.3×10^{-13}, and 9.5×10^{-9}, respectively)(Figure 4.6d). In addition, *spiR* homologs and *ismA* homologs were both associated with significantly reduced stool cholesterol

concentrations in HMP2 and PRISM, but not in PROTECT (*spiR* homologs: HMP2: $P = 1.4 \times 10^{-11}$; PRISM: $P = 0.002$; PROTECT: $P = 0.44$; *ismA* homologs: HMP2: $P = 5.8 \times 10^{-16}$; PRISM: $P = 2.3 \times 10^{-10}$; PROTECT: $P = 0.19$) (Figure 4.6e).

We extended this analysis to the species level by examining the presence of species encoding *spiR* homologs (*spiR*+ species) and species encoding ismA homologs (*ismA*+ species) in metagenomic samples. We applied an approach similar to that of Kenny et al. (2020) to assess species presence in relation to cholesterol conversion. We found similar trends compared to the gene-level analysis, where the *spiR*+ species were more strongly associated with cholesterol conversion than the *ismA*+ species and higher levels of coprostanone, cholestenone, and lower cholesterol levels, in all three studies. We then classified the species into four groups based on *the presence of ismA and spiR homologs*. Importantly, among the converters, there were very few species with only *ismA*, whereas most species had either *spiR* only or both *spiR* and *ismA* homologs. This suggests that *spiR*+ species account for most of the cholesterol conversion activity previously attributed to *ismA* and cholesterol converters, which lack *ismA* homologs.

We evaluated the predictive power of *ismA* and *spiR* homolog presence in relation to fecal coprostanone production and performed receiver operating characteristic (ROC) curve analysis at both the gene and species (encoder) levels (Figure 4.6f). The presence of the *spiR* homolog alone yielded the highest classification performance, with an area under the curve (AUC) of 0.95, outperforming the *ismA* homologs (AUC = 0.81). Similarly, classifiers based on $spiR^+$ species achieved higher performance (AUC = 0.87) than those based on $ismA^+$ species (AUC = 0.82). These results confirm that *spiR* homologs are a stronger and more consistent predictor of coprostanone production than *ismA* homologs, whether assessed directly at the gene level or inferred from the presence of the encoding species. This reinforces the conclusion that *spiR* homologs play a central role in cholesterol metabolism in the gut microbiome, including communities lacking an *ismA* homolog.

4.3 Discussion

This study identified SpiR as a bacterial enzyme that catalyzes the oxidation of cholesterol to cholestenone, the first step in the conversion of cholesterol to coprostanol. Previous work implicated *ismA* in this transformation. However, across the three independent

cohorts, *spiR* homologs were more predictive of the phenotype than *ismA* homologs, showing stronger associations with cholesterol conversion, coprostanol levels, and fecal sterol profiles. The vast majority of coprostanol-producing individuals encode *spiR* homologs. Individuals that lack *ismA* but still show robust cholesterol conversion often contain *spiR* homologs. Figure 4.6b highlights that 89.7% of coprostanol-producing individuals contained *spiR* homologs on average, while only 52.2% on average contained *ismA* homologs. This suggests that some cholesterol conversion previously attributed to *ismA* may have resulted from co-occurrence with *spiR* homologs. These findings challenge the current *ismA*-centric model of gut microbial cholesterol metabolism and establish *spiR* as a key enzymatic contributor to this pathway.

As the activating enzyme in the cholesterol reduction pathway, SpiR oxidizes cholesterol to cholestenone, forming the electron-withdrawing group necessary for subsequent ene reduction. This activating step suggests that *E. cop* may contain a dedicated cholestenone ene-reductase to convert cholestenone to coprostanone.

Compared to many other nutrients in the gut, cholesterol metabolism presents challenges owing to its hydrophobicity and inability to passively diffuse across cellular membranes[180–182]. Cholesterol is amphipathic, with a flat tetracyclic steroid ring system and poor solubility. Its structure promotes deep integration into lipid bilayers, but prevents free diffusion in the aqueous gastrointestinal environment, making it largely inaccessible[183]. In the gut, electron acceptors are limited; therefore, the utilization of cholesterol as a terminal electron acceptor could be highly advantageous[25]. However, our findings identified one phylogenetically restricted clade within the Acutalibacteraceae family capable of cholesterol metabolism, suggesting that it may be a rare trait driving niche adaptation in the gut.

Given the unique physiological properties of cholesterol, we hypothesized that there are likely unidentified transporters and downstream enzymes involved in its metabolism. The Acutalibacteraceae family described in this study is an ideal target for identifying other lineage-specific cholesterol metabolism genes. However, as cholesterol-metabolizing taxa are largely uncultured, direct experimental characterization remains challenging. Comparative genomics offers a powerful alternative for uncovering additional genes and a clearer understanding of how gut microbial cholesterol metabolism affects human health.

Our findings also provide a framework for leveraging microbial cholesterol metabolism in a translational context. *SpiR*-driven variation in coprostanol production between individuals likely contributes to the observed differences in systemic cholesterol levels[184]. The strong

predictive power of *spiR* for cholesterol-metabolizing activity supports its use as a functional biomarker to stratify individuals according to their microbiome's capacity for cholesterol conversion. Culturable members of the *spiR*-positive clade represent promising candidates for next-generation probiotics designed to enhance cholesterol excretion[153,185]. In parallel, dietary interventions that selectively enrich these taxa may offer a prebiotic route for modulating microbial cholesterol metabolism. For example, plant-derived sterols are metabolized by *Eubacterium* species and have been shown to influence both microbiota composition and host lipid profiles[186]. The precise delineation of the microbial genes and lineages responsible for cholesterol-to-coprostanol conversion in the gut thus opens new avenues for the therapeutic modulation of host lipid homeostasis via microbiome engineering.

4.4 Materials and methods

4.4.1 Cloning and transformation

Plasmids containing the genes of interest (*ismA* and *spiR*) were acquired from GenScript (Genscript.com). Plasmids were individually transformed into T7 Express *lysY/I^q* Competent *E. coli* (New England Biolabs, C3013I) using the manufacturer's protocol. The cells were plated on Luria-Bertani (LB, Sigma Aldrich, L3022) agar plates with 50 µg/mL kanamycin (BioBasic, 70560) to select for successfully transformed colonies. The transformed colonies were sequenced and verified using Plasmidsaurus (Plasmidsaurus.com).

4.4.2 Whole cell assays

Transformed *E. coli* strains were inoculated from a glycerol stock into 250 mL of LB media with 50 µg/mL kanamycin and shaken at 32℃ overnight. The bacteria were pelleted by centrifugation at 3,260 × *g* for 7 min and transferred to the anoxic chamber. The cell pellet was resuspended in 100 mL of anoxic LB medium with 10 µg/mL steroid, 50 µg/mL kanamycin, and 400 µM isopropyl-B-D-1-thiogalactopyranoside (IPTG, Goldbio, 22110512481) and incubated anoxically at 37 °C for 48 h.

After incubation in the anoxic chamber, the samples were pelleted by centrifugation at 3,260 x *g* for 5 min. Three milliliters of chloroform (Fisher Chemical, C298) were added to the supernatant, and the products were separated via organic extraction in a separatory funnel. The

resulting chloroform-steroid solution was air-dried until the chloroform was fully evaporated, and the extract was resuspended in 1 mL of methanol (Fisher Chemical, 232699).

4.4.3 Lysate assays

Ten milliliters of overnight *E. coli* culture were seeded in 500 mL of LB medium with 50 µg/mL kanamycin and shaken at 32 °C until an OD_{600} of 0.5-0.6 and 400 µM IPTG was added to the media. The bacteria were shaken overnight at 25 °C, pelleted by centrifugation at 3,260 × *g* for 10 min, and frozen at -20℃. Pellets were thawed on ice for 30 min before the addition of pH-adjusted cell lysis buffer (20 mM Tris, 0.2M sodium chloride), lysozyme (Research Products International, L38100), 100 µL beta-mercaptoethanol (BME, G Biosciences BC98), and protease inhibitor (Sigma Aldrich, 04693124001). Samples were sonicated at 4 °C using a Branson 550 sonicator set to 40% amplitude, with a cycle of 15 s on and 30 s off, for a total sonication time of two minutes. The lysate was centrifuged at 14,000 × *g* for 10 min, and the supernatant was incubated with 100 µL of 10 mM cholesterol in ethanol (IBI Scientific, IB15721), 10 µL of 100 mM NAD^+ (Sigma-Aldrich, 10127981001) in water, and 10 µL of 100 mM $NADP^+$ (Sigma-Aldrich, 10128031001) in water at 4 °C for 48 h. Prior to addition to the lysate, cholesterol was solubilized using methyl-β-cyclodextrin (Sigma Aldrich, C4555). The products were extracted using chloroform and resuspended in methanol.

4.4.4 Derivatization and LC-MS

After chloroform extraction, the samples were derivatized with Girard's Reagent P (Cayman Chemical Company, 601541). 40 µL of the sample was diluted 10-fold in distilled water, and three volumes of 2 mg/mL Girard's Reagent P in methanol containing 1% acetic acid (LabChem, LC101003) were added. The samples were incubated overnight in the dark at 37 °C.

Samples were analyzed with the Bruker Maxis-II QTOF, an ultra-high resolution Q-TOF mass spectrometer coupled with a Waters Acquity I-Class PLUS LC system. Liquid chromatography separation was performed on a Phenomenex Kinetex C18 100 Å LC Column (100 × 3 mm, 2.6 µm particle size) with mobile phase A (water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid) according to the Oxysterol Derivatization MaxSpec® Kit (Cayman Chemical, 601540) manufacturer's protocol. The gradient program began at 2% B for 0.5 minutes, increased to 15% B at 1 min, and was maintained for 9 min

before increasing to 80% B. From 10.2 minutes-13 minutes, B was increased to 95% and then decreased to 2% for an additional 3 min. The column was maintained at 25 °C. The injection volume was 5 µl. Mass spectra were acquired under positive electrospray ionization with an ion spray voltage of 4,500 V. The source temperature (dry gas) was set at 220°C at a flow rate of 5.0 L/min. All data were visualized using the Bruker Compass Analysis software.

4.4.5 Phase column LC-MS

To separate isomers (5α- and 5β-dihydroprogesterone), samples were analyzed with the Bruker Maxis-II QTOF, an ultra-high resolution Q-TOF mass spectrometer coupled with a Waters Acquity I-Class PLUS LC system. Liquid chromatography separation was performed on a Phenomenex Kinetex F5 100 Å LC Column (100 × 3 mm, 2.6 µm particle size) with mobile phase A (water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid). The gradient program began with 10% B for 2 min, increased to 55% B over the next minute, and further increased to 65% B for 11 min. B was increased to 90% over the next minute, maintained at 90% B for 2 min, decreased to 10% over the course of 1 min, and maintained for 3 min. The column was maintained at 45 °C. The injection volume was 5 µL. Mass spectra were acquired under positive electrospray ionization with an ion spray voltage of 4,500 V. The source temperature (dry gas) was set to 220°C with a flow rate of 5.0 L/min. All data was visualized using the Bruker Compass Analysis software.

4.4.6 Protein purification

Transformed E. coli strains were grown in LB media with 50 µg/mL kanamycin overnight at 37°C. Five milliliters of culture was seeded into 500 mL of LB and grown to an OD_{600} of ~ 0.5. For *spiR,* expression was induced with 400 µM IPTG, whereas IsmA expression was induced with 100 µM IPTG supplemented with 100 µM l-arginine at the time of induction. Both the proteins were incubated overnight at 25°C. The cells were pelleted for 20 min at 3,260 × g and the supernatant was discarded. For SpiR, cells sodium frozen, -20°C and subsequently thawed, ice, then resuspended in lysis buffer (0.1M sodium phosphate, 0.2M sodium chloride, 20 mM BME, pH = 6.5), whereas IsmA was resuspended in 20mM Tris, 0.2M sodium chloride, 20mM BME at pH 8.5. Both slurries were treated with lysozyme, benzonase (Sigma-Aldrich, E8263-5KU), and protease inhibitors.

The samples were sonicated at 4 °C using a Branson 550 sonicator at 40% amplitude with 15 s on and 15 s off for a total of 2 min. The cell debris was centrifuged at 14,000 × g for 10 min, and the supernatant was filtered through a 0.22 μm PES syringe filter (Celltreat, 229746). The resulting solution was passed through a Nickel HisTrap Excel column (Cytiva, 17371205) attached to a peristaltic pump. Nonspecific binding was eliminated using 50 mM imidazole (Fisher, BP305). The protein was eluted using 0.2M imidazole. The SpiR eluent was concentrated and buffer switched to 0.1M sodium phosphate, 0.2M sodium chloride, and 20 mM BME using Amicon ultra centrifugal filter 30 kDa MWCO (Millipore Sigma, UFC5030). The IsmA eluent was concentrated and the buffer was switched to 20mM Tris, 0.2M sodium chloride, 20mM BME at a pH of 8.5.

4.4.7 Binding characterization

The protein stocks were concentrated to the desired concentrations in the respective buffers. All the compounds tested with isothermal titration calorimetry (ITC) were prepared in the same buffer as the protein, SpiR, or IsmA tested. Most of the tested compounds were hydrophobic and required additional preparation to ensure proper solubility. Hydrophobic compounds were suspended in glass vials in phosphate buffer as 10mM stocks and shaken for 48 hours at RT. The compounds were diluted to the desired concentrations and all solutions were degassed prior to use.

Binding experiments were performed at 25 °C using a TA Instruments Nano ITC instrument. Inverse titrations were performed for pregnenolone (Selleck Chemicals, S1914) and 5β-dihydroprogesterone (Cayman Chemical, 34644). The ligand was present in the sample cell, whereas the protein was present in the syringe. Protein was delivered into the cell in 2.5 uL injections, with a spacing interval of 180 s and a syringe speed of 150 RPM. NADH (Cayman Chemical, 16078), NAD^+, coprostanone (Chemscence, CS-0643978), and cholesterol (AA Blocks, AA00EBSJ) were used as standard titrations, the and ligand was titrated into the protein. To improve the solubility of cholesterol and coprostanone in aqueous solutions, 5% (v/v) DMSO was added to the buffer used for the ligand and protein. The initial titrations were performed with 20 injections, which led to aggregation. To mitigate this problem, the reaction time was decreased by titrating 3.1 μL injections over a total of 16 injections. The same method was used for NADH and NAD^+.

Subtraction of the heats of dilution and analysis were performed using NanoAnalyze software (TA Instruments). The measured heat was integrated into a binding curve, wherein the enthalpy change (ΔH), entropic change (ΔS), stoichiometry (n), and binding (Kd) were calculated based on the fitting of the curve using the independent binding model for all ligand–protein interactions. The respective protein and ligand concentrations used are described in the figure captions.

4.4.8 Protein gel electrophoresis

Samples from lysate assays, along with a vector control, were collected at various time points for SDS-PAGE analysis to support the product conversion data. Samples were collected at three key stages: pre-induction, post-induction, and post-lysis supernatant, to monitor protein expression. The samples were prepared by mixing 20 μL of the experimental conditions, 8 μL of Lithium Dodecyl Sulfate (LDS) 4X (GenScript, M00676-10), and 4 μL of Sample Reducing Agent (SRA) 10X (Bioworld, 21440022-2). The samples were then incubated at 95°C for 5 min. BioRAD Mini-PROTEAN® TGX™ Precast Protein Gels, 4–15% (Bio-Rad, 4561084) were prepared, and 8 μL of samples were individually loaded into each well, along with a molecular weight marker (Cell Signaling Technology, 59329S) for protein molecular weight determination. The gels were run for 45 min at 120 volts and subsequently stained with 1% Coomassie Blue Stain (Research Products International, B43000-25.0) in a solution containing 65% deionized water, 35% methanol, and 5% acetic acid (LabChem, LC101003). The results for the samples are described in the figure captions.

4.4.9 Structural prediction and analysis

The structure of SpiR was predicted using AlphaFold2[187]. Binding pockets were predicted using fpocket (v4.0) with the default parameters[188]. The pockets were compared to the homologous DP-D-glycero-D-manno-heptose 4,6-Dehydratase (PDB: 7US5)[189] to identify the putative substrate-binding regions and catalytic residues. The structures for cholesterol (PubChem compound identifier: 5997) and NADH (PubChem compound identifier: 439153) were docked onto the predicted SpiR structure using AutoDock Vina (v4.2). The docking simulation was performed within 10 Å × 10 Å × 10 Å cubes centered on the center points of the

chosen fpocket substrate-binding pocket with exhaustiveness set to 32. The docking results were visualized using PyMOL[190].

4.4.10 Phylogenetic reconstruction and distribution of SpiR homologs

All representative genomes from the Genome Taxonomy Database (GTDB) (release r226)[191] were downloaded, and protein sequences for each genome were predicted using Prokka (version 1.14.6)[192]. All protein sequences were annotated using InterProScan v5.73[193]. We used DIAMOND[194] to search for homologous amino acid sequences of SpiR in all GTDB sequences in the superfamily SSF51735 (NAD(P)-binding Rossmann-fold domain) and filtered hits based on a cutoff of 500 bitscore.

We aligned SpiR with other previously biochemically characterized HSDHs[166] using MUSCLE v5[195] with default parameters and trimmed to remove columns with more than 97% gaps with goalign[196]. Trimmed alignment was used to construct a phylogenetic tree using RAxML-NG v1.2.2[197], with 1000 bootstrap replicates and the substitution model LG[198].

4.4.11 Cross-cohort identification of cholesterol derivatives via retention time and m/z harmonization

We identified the retention times of cholesterol, cholestenone, and coprostanone using LC-MS data from three untargeted metabolomic datasets generated from human cohorts: PRISM (PR000677), HMP2 (PR000639), and PROTECT (ST002471). These datasets were generated using similar protocols, which allowed us to align the retention time and m/z features across cohorts. Our approach follows a comparative annotation strategy similar to that described by Kenny et al. (2020)[153].

In the PRISM dataset, we identified cholesterol and cholestenone based on their retention times, as previously reported by Kenny et al. (2020)[153] (7.21 and 7.00 minutes, respectively) and expected m/z values. A third compound eluted at ~7.50 minutes was initially labeled as coprostanol based on the annotations reported by Cell Host & Microbe (2018). However, detailed inspection of its ionization profile, specifically m/z 385.3465 $[M+H]^+$, 409.3441 $[M+Na]^+$, and 369.3516 $[M+H–H_2O]^+$, strongly supports its identity as a coprostanone. Coprostanol is known to ionize poorly under ESI and has not been observed consistently in these datasets, further supporting this reassignment[199].

We then cross-referenced these features in the HMP2 and PROTECT. In HMP2, ME233441 was annotated as cholesterol with a retention time of 7.35 minutes, consistent with PRISM. In the PROTECT dataset, ME584140 and ME584114 corresponded to cholesterol and coprostanone, respectively, with retention times of 7.59 and 7.87 minutes, respectively. Based on the consistent retention behavior and m/z values, we subsequently assigned the retention times of cholestenone and coprostanone in HMP2 and cholestenone in PROTECT. These assignments were guided by exact mass values and the presence of expected adducts ($[M+H]^+$ and $[M+Na]^+$) co-eluted at the same retention time, enabling the confident harmonization of compound identities across all three metabolomics cohorts.

4.4.12 Metagenomic profiling of spiR homologs and ismA homologs in the human gut microbiome

We built a reference database using 323 *spiR* homolog genes as previously described. A reference dataset of 188 *ismA* homologs was also collected by aligning previously identified *ismA* homologs with MAFFT (v7.526) and building a profile HMM using HMMER (v3.4)[200]. Sequences from the superfamily SSF51735 were aligned to the HMM profile using hmmsearch, and hits were filtered based on a 348 bit score threshold.

We downloaded samples from 400 HMP2[201,] 151 PRISM[202] and 90 PROTECT, which had paired metagenomic and metabolomic data. The metagenomic reads were first processed by trimming the adapter sequences using TrimGalore (version 0.6.7)[203] with default settings. The reads were then mapped to a human reference (assembly T2T-CHM13v2.0) to identify potential contaminants and removed them using Samtools (v1.16)[204]. We removed samples with less than one million total reads after curation. We aligned the reads to the *spiR* homolog and *ismA* homolog reference databases using bowtie2 (v2.4.1)[205]. The number of reads mapped to each reference was summarized by normalizing the number of reads in the sample and multiplying by one million to obtain counts per million (cpm). A sample was considered to have *spiR* homologs or *ismA* homologs present if the cpm was greater than 1.

To determine species-level presence, we compiled all strains belonging to the species containing *spiR* homolog (n = 308) or *ismA* homolog (n = 180) genes and constructed a reference database using the complete genomic sequences of these strains. Metagenomic reads were mapped to this reference genome `using the alignment method described`

previously. A threshold of CPM ≥ 18,000 was applied to define species presence.

Chapter 5: Conclusion

In this dissertation, I have identified three key enzymes involved in health-relevant functions of the microbiome, elucidating their role in pathways critical for maintaining human health. Though there are innumerable ways the gut microbiome contributes to health, by focusing on the enzymatic basis for each function, we can take small, yet tangible, steps towards teasing apart this complex relationship. The research performed here provides the basis for future work to modulate the gut microbiome - via prebiotics, probiotics, or synbiotics - in order to impact human health.

5.1 Modulation of the microbiome

The gut microbiome represents an exciting frontier for modern therapeutics. Unfortunately, progress in this area has been hindered by several challenges. A combination of a poor understanding of causation behind observed associations between specific commensal microbes and health outcomes, insufficient regulation of easily available probiotic supplements, and misleading, widespread media hype that oversimplifies complex community interactions has led to an oversaturated probiotic market, claiming to address ailments from bloating to mental health disorders despite limited evidence. Though therapeutic modulation of the microbiome holds significant potential, better understanding of microbial functions and the enzymatic basis driving these associations are badly needed.

Though there are many ways of modulating the microbiome, including fecal microbiome transplants and antibiotic therapy, we will focus on a discussion of the common dietary interventions - namely, prebiotics and probiotics - as the commonly employed therapeutics for modulating the gut microbiome.

5.1.1 Prebiotics

Prebiotics consist of carbohydrates indigestible by and inaccessible to the host that confer health benefits by selectively promoting the growth of beneficial commensal bacteria. A key feature of prebiotics is their resistance to absorption in the upper gastrointestinal tract, ensuring that they reach the large intestine, where they can be fermented by the gut microbiota.

Common prebiotics include inulin from chicory root or Jerusalem artichoke, fructo-oligosaccharides, galacto-oligosaccharides, and xylo-oligosaccharides from hemicellulose. Prebiotics aim to promote the growth of specific beneficial commensal bacteria that are already present in the gut, and provide them with carbon sources to stimulate production of beneficial metabolites. Unlike probiotics, which introduce live bacteria, prebiotics support the growth of existing bacterial communities in the host. For example, *Clostridium butyricum* is a notable butyrate producing commensal. When provided with inulin, *C. butyricum* can metabolize it to butyrate, which is known to support the colonic epithelial barrier. Therefore, inulin supplements should increase butyrate levels by promoting *C. butyricum* growth and providing a carbon source for butyrate production.

In recent years, the definition of prebiotics has broadened to include new indigestible molecules, like plant-derived polyphenols. Prebiotics are generally defined as compounds oxidized or fermented by the gut bacteria. Although the mechanism of action of polyphenols is varied, some work has found that they promote the growth of beneficial microbes like *Akkermansia muciniphilia.* However, despite their modulatory effects on the gut microbiome, little work has been done on the metabolism of polyphenols in the gut. Preliminary studies show that polyphenols are likely not being oxidized by the microbiota, and their metabolic products remain unknown. This distinction expands the conventional definition of prebiotics, highlighting the need to better understand the metabolic products of polyphenols in the gut.

5.1.2 Probiotics

While prebiotics support the growth of existing commensal microbes, probiotics aim to introduce new, beneficial species or strains. Probiotic pills and powders often contain a mixture of species with the hope that at least one of them will be able to colonize the gut. They often include species like *Lactobacillus casei, Lactobacillus acidophilus, Streptococcus thermophilus,* and a wide variety of Bifidobacterium and Bacillus spp..

A number of considerations go into selecting strains for probiotic use. Ideal probiotics must be (1) non-pathogenic and categorized as generally regarded as safe (GRAS), (2) genetically stable to prevent unintended mutations, (3) able to survive in the GI tract, (4) likely to colonize the gut once consistently introduced, and (5) provide a demonstrated health benefit.[206] The health benefits of probiotics can include inhibition of pathogen growth, production of antimicrobial or bioactive molecules like bacitracin and short chain fatty acids, or synthesis of

vitamins, though probiotics often struggle to provide lasting changes to the composition of the gut microbiome.

Synbiotics combine prebiotics and probiotics, promoting the survival and colonization of the probiotic stain by providing carbon sources that selectively favor that strain. Synbiotics are often used to overcome difficulties in probiotic gut colonization.

However, today's prebiotic and probiotics tend to be untargeted, with generally beneficial effects. In order to specifically alter the gut microbiome to have specific effects, like lowering serum cholesterol or bilirubin, deliberate engineering and modulation of the microbiome is necessary.

5.2 Identifying the enzymatic basis for microbial functions

As more research is done on the enzymatic basis for microbial associations, the use of genetically modified bacteria as targeted probiotics becomes viable. Some of the microbes responsible for beneficial functions of the gut microbiome - like the reduction of bilirubin by *C. difficile* - are not viable candidates for probiotic therapy. *C. difficile* is known to cause infections with severe consequences in elderly or immunocompromised patients, and is therefore not categorized as GRAS. Given that *C. difficile* is an opportunistic pathogen, even a gene from *C. difficile* heterologously expressed in an approved strain of *E. coli* would be unlikely to be approved. However, by identifying the bacterial gene and enzyme responsible for bilirubin reduction, we were able to find *bilR* in other species, like *C. symbiosum,* which, while also unlikely to be approved as a probiotic, is a commensal species. It may be possible to express the *bilR* gene from *C. symbiosum* or another bilirubin reduction species in *E. coli* for use as a genetically engineered probiotic. A number of *E. coli* Nissle 1979 with heterologously expressed genes have already been proposed as probiotic therapeutics for cancer and IBD.[207]

5.2.1 BilR-based therapeutics

Narrowing the association between bilirubin reduction and *C. difficile* to the specific identification of long and short *bilR* broadens its applicability to disease prevention and treatment. Current diagnostics for jaundice and hyperbilirubinemia are limited to visual and light-based diagnosis followed by a serum bilirubin blood test. These are all metrics that only surface after bilirubin levels in the blood have risen to the diagnostic threshold. Utilizing the

knowledge that a lack of *bilR* in the gut microbiome is highly correlated with the time period of the highest incidence of jaundice, screening for *bilR*, or a lack thereof, in the infant gut could provide an early sign for infants more susceptible to jaundice, although a clinical study to correlate the presence of *bilR* specifically with jaundice in infants would be needed to provide stronger evidence.

If genetically modified *E. coli* containing *bilR* were approved as a probiotic, they could be given as preventative treatments to infants or other populations susceptible to hyperbilirubinemia to colonize the gut with commensal bilirubin reducing bacteria and preemptively reduce the reuptake of bilirubin in the gut.

5.2.2 SpiR-based therapeutics

Lowering the reuptake of cholesterol in the gut is already a validated therapeutic target, with cholesterol absorption inhibitors like ezetimibe becoming popular alongside statins. Ezetimibe works by inhibiting the NPC1L1/cholesterol complex from interfacing with clathrin coated vesicles for reuptake, rather than via the microbiota, and has been shown to be effective in reducing serum cholesterol. However, the core principle behind the mechanism of ezetimibe is decreasing the cholesterol reabsorbed in the gut.

The same core principle is at the heart of using probiotic therapy to decrease the amount of free cholesterol in the gut available for reuptake, rather than inhibiting the reabsorption process itself. With our identification of SpiR as an enzyme that oxidizes cholesterol to cholestenone, *spiR* becomes a possible therapeutic to be expressed in *E. coli* and used as a probiotic to reduce cholesterol reuptake in the gut.

More clinical data is necessary to associate the presence of *spiR* in the gut microbiome with decreased serum cholesterol, although our metagenomics work has already strongly correlated *spiR* with higher fecal coprostanol levels. However, if the presence of *spiR* is associated with lower serum cholesterol, or if the absence is correlated with hypercholesterolemia, *spiR* could make for an excellent candidate for preventative screening and treatment for high serum cholesterol. Other work on steroid-reducing enzymes may identify additional bacterial steroid-reduction genes that could be added to a screening panel, identifying which steroid reducing genes may be missing from the gut and leading to the development and prescription of personalized probiotics.

References

1. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262–1267 (2012).
2. Ha, C. W. Y., Lam, Y. Y. & Holmes, A. J. Mechanistic links between gut microbial community dynamics, microbial functions and metabolic health. *World J. Gastroenterol.* **20**, 16498–16517 (2014).
3. Robitaille, S. *et al.* Community composition and the environment modulate the population dynamics of type VI secretion in human gut bacteria. *Nat. Ecol. Evol.* **7**, 2092–2107 (2023).
4. Zhu, A., Sunagawa, S., Mende, D. R. & Bork, P. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol.* **16**, 82 (2015).
5. Martin, A. M., Sun, E. W., Rogers, G. B. & Keating, D. J. The influence of the gut microbiome on host metabolism through the regulation of gut hormone release. *Front. Physiol.* **10**, 428 (2019).
6. Joyce, S. A. & Clarke, D. J. Microbial metabolites as modulators of host physiology. *Adv. Microb. Physiol.* **84**, 83–133 (2024).
7. Kim, S., Seo, S.-U. & Kweon, M.-N. Gut microbiota-derived metabolites tune host homeostasis fate. *Semin. Immunopathol.* **46**, 2 (2024).
8. Lee, J.-Y., Bays, D. J., Savage, H. P. & Bäumler, A. J. The human gut microbiome in health and disease: time for a new chapter? *Infect. Immun.* **92**, e0030224 (2024).
9. Afzaal, M. *et al.* Human gut microbiota in health and disease: Unveiling the relationship. *Front. Microbiol.* **13**, 999001 (2022).
10. Troy, E. B. & Kasper, D. L. Beneficial effects of Bacteroides fragilis polysaccharides on the immune system. *Front. Biosci. (Landmark Ed.)* **15**, 25–34 (2010).
11. Kelly, J. R. *et al.* Breaking down the barriers: the gut microbiome, intestinal permeability and stress-related psychiatric disorders. *Front. Cell. Neurosci.* **9**, 392 (2015).
12. Martin, C. R., Osadchiy, V., Kalani, A. & Mayer, E. A. The brain-gut-microbiome axis. *Cell. Mol. Gastroenterol. Hepatol.* **6**, 133–148 (2018).
13. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components.

Eur. J. Nutr. **57**, 1–24 (2018).

14. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the intestinal Microbiota and its influence on human health and disease. *J. Mol. Biol.* **428**, 3230–3252 (2016).
15. Wilson, I. D. & Nicholson, J. K. Gut microbiome interactions with drug metabolism, efficacy, and toxicity. *Transl. Res.* **179**, 204–222 (2017).
16. Guzior, D. V. & Quinn, R. A. Review: microbial transformations of human bile acids. *Microbiome* **9**, 140 (2021).
17. Hall, B. *et al.* BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen. *Nat. Microbiol.* **9**, 173–184 (2024).
18. Hou, K. *et al.* Microbiota in health and diseases. *Signal Transduct. Target. Ther.* **7**, 135 (2022).
19. Al Bander, Z., Nitert, M. D., Mousa, A. & Naderpoor, N. The gut Microbiota and inflammation: An overview. *Int. J. Environ. Res. Public Health* **17**, 7618 (2020).
20. Park, B. *et al.* Crosstalk between butyrate oxidation in colonocyte and butyrate-producing bacteria. *iScience* **27**, 110853 (2024).
21. Litvak, Y., Byndloss, M. X. & Bäumler, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, (2018).
22. Schirmer, M., Garner, A., Vlamakis, H. & Xavier, R. J. Microbial genes and pathways in inflammatory bowel disease. *Nat. Rev. Microbiol.* **17**, 497–511 (2019).
23. Gohil, K. & Carramusa, B. Ulcerative colitis and Crohn's disease. *P T* **39**, 576–577 (2014).
24. Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med.* **9**, 103 (2017).
25. Winter, S. E. *et al.* Host-derived nitrate boosts growth of E. coli in the inflamed gut. *Science* **339**, 708–711 (2013).
26. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467**, 426–429 (2010).
27. Faber, F. *et al.* Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. *Nature* **534**, 697–699 (2016).

28. Koizumi, S. Human heme oxygenase-1 deficiency: a lesson on serendipity in the discovery of the novel disease. *Pediatr. Int.* **49**, 125–132 (2007).
29. Lester, R. & Schmid, R. Intestinal Absorption of Bile Pigments. *New England Journal of Medicine* vol. 269 178–182 Preprint at https://doi.org/10.1056/nejm196307252690402 (1963).
30. Lester, R. & Schmid, R. Intestinal Absorption of Bile Pigments. III. The Enterohepatic Circulation of Urobilinogen in the Rat. *J. Clin. Invest.* **44**, 722–730 (5 1965).
31. Saxerholt, H., Skar, V. & Midtvedt, T. HPLC separation and quantification of bilirubin and its glucuronide conjugates in faeces and intestinal contents of germ-free rats. *Scand. J. Clin. Lab. Invest.* **50**, 487–495 (1990).
32. With, T. K. *Bile Pigments : Chemical, Biological, and Clinical Aspects*. (1968).
33. Ching, S., Ingram, D., Hahnel, R., Beilby, J. & Rossi, E. Serum levels of micronutrients, antioxidants and total antioxidant status predict risk of breast cancer in a case control study. *J. Nutr.* **132**, 303–306 (2002).
34. Osiak, W., Wątroba, S., Kapka-Skrzypczak, L. & Kurzepa, J. Two Faces of Heme Catabolic Pathway in Newborns: A Potential Role of Bilirubin and Carbon Monoxide in Neonatal Inflammatory Diseases. *Oxid. Med. Cell. Longev.* **2020**, 7140496 (2020).
35. Kapitulnik, J. Bilirubin: an endogenous product of heme degradation with both cytotoxic and cytoprotective properties. *Mol. Pharmacol.* **66**, 773–779 (2004).
36. With, T. K. *Bile Pigments : Chemical, Biological, and Clinical Aspects*. (1968).
37. Stenemo, M. *et al.* The metabolites urobilin and sphingomyelin (30:1) are associated with incident heart failure in the general population. *ESC Heart Fail* **6**, 764–773 (2019).
38. Kipp, Z. A. *et al.* Bilirubin Levels Are Negatively Correlated with Adiposity in Obese Men and Women, and Its Catabolized Product, Urobilin, Is Positively Associated with Insulin Resistance. *Antioxidants (Basel)* **12**, (2023).
39. Vítek, L., Zelenka, J., Zadinová, M. & Malina, J. The impact of intestinal microflora on serum bilirubin levels. *J. Hepatol.* **42**, 238–243 (2005).
40. Vítek, L. *et al.* Intestinal colonization leading to fecal urobilinoid excretion may play a role in the

pathogenesis of neonatal jaundice. *J. Pediatr. Gastroenterol. Nutr.* **30**, 294–298 (2000).

41. Poland, R. L. & Odell, G. B. Physiologic jaundice: the enterohepatic circulation of bilirubin. *N. Engl. J. Med.* **284**, 1–6 (1971).
42. Brodersen, R. & Hermann, L. S. Intestinal reabsorption of unconjugated bilirubin. A possible contributing factor in neonatal jaundice. *Lancet* **1**, 1242 (1963).
43. Gustafsson, B. E. & Lanke, L. S. Bilirubin and urobilins in germfree, ex-germfree, and conventional rats. *J. Exp. Med.* **112**, 975–981 (1960).
44. Midtvedt, T. & Gustafsson, B. E. Microbial conversion of bilirubin to urobilins in vitro and in vivo. *Acta Pathol. Microbiol. Scand. B* **89**, 57–60 (1981).
45. Fahmy, K., Gray, C. H. & Nicholson, D. C. The reduction of bile pigments by faecal and intestinal bacteria. *Biochim. Biophys. Acta* **264**, 85–97 (1972).
46. Kumagai, A. *et al.* A bilirubin-inducible fluorescent protein from eel muscle. *Cell* **153**, 1602–1611 (2013).
47. Vítek, L. *et al.* Identification of bilirubin reduction products formed by Clostridium perfringens isolated from human neonatal fecal flora. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **833**, 149–157 (2006).
48. Rekittke, I. *et al.* Structure of (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate reductase, the terminal enzyme of the non-mevalonate pathway. *J. Am. Chem. Soc.* **130**, 17206–17207 (2008).
49. Hubbard, P. A., Liang, X., Schulz, H. & Kim, J.-J. P. The crystal structure and reaction mechanism of Escherichia coli 2,4-dienoyl-CoA reductase. *J. Biol. Chem.* **278**, 37553–37560 (2003).
50. Tu, X., Hubbard, P. A., Kim, J.-J. P. & Schulz, H. Two distinct proton donors at the active site of Escherichia coli 2,4-dienoyl-CoA reductase are responsible for the formation of different products. *Biochemistry* **47**, 1167–1175 (2008).
51. Dennery, P. A., Seidman, D. S. & Stevenson, D. K. Neonatal hyperbilirubinemia. *N. Engl. J. Med.* **344**, 581–590 (2001).
52. Zhao, X. *et al.* The Relationship between Serum Bilirubin and Inflammatory Bowel Disease. *Mediators Inflamm.* **2019**, 5256460 (2019).

53. Brink, M. A. *et al.* Enterohepatic cycling of bilirubin: a putative mechanism for pigment gallstone formation in ileal Crohn's disease. *Gastroenterology* **116**, 1420–1427 (1999).

54. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

55. Garrod, A. E. Note on the Origin of the Yellow Pigment of Urine. *J. Physiol.* **21**, 190–191 (1897).

56. Weller, S. D. V. Bile pigments in the stools of infants. *Arch. Dis. Child.* **26**, 86–88 (1951).

57. Maclagan, N. F. Faecal urobilinogen; clinical evaluation of a simplified method of estimation. *Br. J. Exp. Pathol.* **27**, 190–200 (1946).

58. Chen, S. & Tukey, R. H. Humanized UGT1 Mice, Regulation of UGT1A1, and the Role of the Intestinal Tract in Neonatal Hyperbilirubinemia and Breast Milk-Induced Jaundice. *Drug Metab. Dispos.* **46**, 1745–1755 (2018).

59. With, T. K. *Bile Pigments : Chemical, Biological, and Clinical Aspects*. (1968).

60. Koníčková, R. *et al.* Reduction of bilirubin ditaurate by the intestinal bacterium Clostridium perfringens. *Acta Biochim. Pol.* **59**, 289–292 (2012).

61. Vítek, L. & Ostrow, J. D. Bilirubin chemistry and metabolism; harmful and protective aspects. *Curr. Pharm. Des.* **15**, 2869–2883 (2009).

62. Zhou, S. *et al.* Association of serum bilirubin in newborns affected by jaundice with gut microbiota dysbiosis. *J. Nutr. Biochem.* **63**, 54–61 (2019).

63. Vítek, L., Novotný, L., Sperl, M., Holaj, R. & Spácil, J. The inverse association of elevated serum bilirubin levels with subclinical carotid atherosclerosis. *Cerebrovasc. Dis.* **21**, 408–414 (2006).

64. Kipp, Z. A. *et al.* Bilirubin Levels Are Negatively Correlated with Adiposity in Obese Men and Women, and Its Catabolized Product, Urobilin, Is Positively Associated with Insulin Resistance. *Antioxidants (Basel)* **12**, (2023).

65. Kipp, Z. A. *et al.* Bilirubin Levels Are Negatively Correlated with Adiposity in Obese Men and Women, and Its Catabolized Product, Urobilin, Is Positively Associated with Insulin Resistance. *Antioxidants (Basel)* **12**, (2023).

66. The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).

67. Naumann, H. N. Schlesinger's test for urobilin in the presence of riboflavin and other fluorescent compounds. *J. Lab. Clin. Med.* **32**, 1503–1507 (1947).

68. Elman, R. & McMaster, P. D. STUDIES ON UROBILIN PHYSIOLOGY AND PATHOLOGY : I. THE QUANTITATIVE DETERMINATION OF UROBILIN. *J. Exp. Med.* **41**, 503–512 (1925).

69. Kotal, P. & Fevery, J. Quantitation of urobilinogen in feces, urine, bile and serum by direct spectrophotometry of zinc complex. *Clin. Chim. Acta* **202**, 1–9 (1991).

70. Nordin, A. & Wagner, M. Application of 2D fluorescence spectroscopy on faecal pigments in water. https://stud.epsilon.slu.se/10251/1/daub_b_170622.pdf.

71. Miyabara, Y., Tabata, M., Suzuki, J. & Suzuki, S. Separation and sensitive determination of i-urobilin and l-stercobilin by high-performance liquid chromatography with fluorimetric detection. *J. Chromatogr.* **574**, 261–265 (1992).

72. Direct amplication of DNA from colonies of Bacillus subtilis and Escherichia coli by the polymerase chain reaction. *J. Microbiol. Methods* **11**, 121–126 (1990).

73. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020).

74. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

75. Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451 (2023).

76. Dalkiran, A. *et al.* ECPred: a tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature. *BMC Bioinformatics* **19**, 334 (2018).

77. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

78. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).

79. Eberhardt, J., Santos-Martins, D., Tillack, A. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Preprint at https://doi.org/10.26434/chemrxiv.14774223.

80. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
81. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
82. Parks, D. H. *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.* **50**, D785–D794 (2022).
83. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
84. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
85. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
87. Levy, S. *et al.* Convergent evolution of oxidized sugar metabolism in commensal and pathogenic microbes in the inflamed gut. *Nat. Commun.* **16**, 1121 (2025).
88. Shan, Y., Lee, M. & Chang, E. B. The Gut Microbiome and Inflammatory Bowel Diseases. *Annu. Rev. Med.* **73**, 455–468 (2022).
89. Glassner, K. L., Abraham, B. P. & Quigley, E. M. M. The microbiome and inflammatory bowel disease. *J. Allergy Clin. Immunol.* **145**, 16–27 (2020).
90. Shen, Z.-H. *et al.* Relationship between intestinal microbiota and ulcerative colitis: Mechanisms and clinical application of probiotics and fecal microbiota transplantation. *World J. Gastroenterol.* **24**, 5–14 (2018).
91. Świrkosz, G., Szczygieł, A., Logoń, K., Wrześniewska, M. & Gomułka, K. The Role of the Microbiome in the Pathogenesis and Treatment of Ulcerative Colitis-A Literature Review. *Biomedicines* **11**, (2023).
92. Morgan, X. C. *et al.* Dysfunction of the intestinal microbiome in inflammatory bowel disease and

treatment. *Genome Biol.* **13**, R79 (2012).

93. Lewis, J. D. *et al.* Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* **18**, 489–500 (2015).
94. Santana, P. T., Rosas, S. L. B., Ribeiro, B. E., Marinho, Y. & de Souza, H. S. P. Dysbiosis in Inflammatory Bowel Disease: Pathogenic Role and Potential Therapeutic Targets. *Int. J. Mol. Sci.* **23**, (2022).
95. Recharla, N., Geesala, R. & Shi, X.-Z. Gut Microbial Metabolite Butyrate and Its Therapeutic Role in Inflammatory Bowel Disease: A Literature Review. *Nutrients* **15**, (2023).
96. Gasaly, N., Hermoso, M. A. & Gotteland, M. Butyrate and the Fine-Tuning of Colonic Homeostasis: Implication for Inflammatory Bowel Diseases. *Int. J. Mol. Sci.* **22**, (2021).
97. Winter, S. E. *et al.* Host-derived nitrate boosts growth of E. coli in the inflamed gut. *Science* **339**, 708–711 (2013).
98. Hughes, E. R. *et al.* Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017).
99. Kelleher, Z. T., Matsumoto, A., Stamler, J. S. & Marshall, H. E. NOS2 regulation of NF-kappaB by S-nitrosylation of p65. *J. Biol. Chem.* **282**, 30667–30672 (2007).
100. The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).
101. Schirmer, M. *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3**, 337–346 (2018).
102. Avdagić, N. *et al.* Nitric oxide as a potential biomarker in inflammatory bowel disease. *Bosn. J. Basic Med. Sci.* **13**, 5–9 (2013).
103. Dhillon, S. S. *et al.* Higher activity of the inducible nitric oxide synthase contributes to very early onset inflammatory bowel disease. *Clin. Transl. Gastroenterol.* **5**, e46 (2014).
104. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* **4**, 293–305 (2019).
105. de Vera, M. E. *et al.* Transcriptional regulation of human inducible nitric oxide synthase (NOS2) gene by cytokines: initial analysis of the human NOS2 promoter. *Proceedings of the National*

Academy of Sciences **93**, 1054–1059 (1996).

106. Rosas-Lemus, M. *et al.* Structure of galactarate dehydratase, a new fold in an enolase involved in bacterial fitness after antibiotic treatment. *Protein Sci.* **29**, 711–722 (2020).
107. Monterrubio, R., Baldoma, L., Obradors, N., Aguilar, J. & Badia, J. A common regulator for the operons encoding the enzymes involved in D-galactarate, D-glucarate, and D-glycerate utilization in Escherichia coli. *J. Bacteriol.* **182**, 2672–2674 (2000).
108. Lenzen, S. A fresh view of glycolysis and glucokinase regulation: history and current status. *J. Biol. Chem.* **289**, 12189–12194 (2014).
109. Chaudhry, R. & Varacallo, M. *Biochemistry, Glycolysis*. (StatPearls Publishing, 2023).
110. Blumenthal, H. J. & Jepson, T. Asymmetric dehydration of galactarate by bacterial galactarate dehydratase. *Biochem. Biophys. Res. Commun.* **17**, 282–287 (1964).
111. Hubbard, B. K., Koch, M., Palmer, D. R., Babbitt, P. C. & Gerlt, J. A. Evolution of enzymatic activities in the enolase superfamily: characterization of the (D)-glucarate/galactarate catabolic pathway in Escherichia coli. *Biochemistry* **37**, 14369–14375 (1998).
112. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
113. Addgene: pCW-LIC. https://www.addgene.org/26098/.
114. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
115. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
116. Karsten, W. E., Nimmo, S. A., Liu, J. & Chooback, L. Identification of 2, 3-dihydrodipicolinate as the product of the dihydrodipicolinate synthase reaction from Escherichia coli. *Arch. Biochem. Biophys.* **653**, 50–62 (2018).
117. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
118. Dufault-Thompson, K. & Jiang, X. Annotating microbial functions with ProkFunFind. *mSystems*

e0003624 (2024).

119. Abed, J. *et al.* Colon Cancer-Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon Tumors via the Circulatory System. *Front. Cell. Infect. Microbiol.* **10**, 400 (2020).

120. Paiardini, A., Contestabile, R., D'Aguanno, S., Pascarella, S. & Bossa, F. Threonine aldolase and alanine racemase: novel examples of convergent evolution in the superfamily of vitamin B6-dependent enzymes. *Biochim. Biophys. Acta* **1647**, 214–219 (2003).

121. Mira, A., Pushker, R., Legault, B. A., Moreira, D. & Rodríguez-Valera, F. Evolutionary relationships of Fusobacterium nucleatum based on phylogenetic analysis and comparative genomics. *BMC Evol. Biol.* **4**, 50 (2004).

122. Karpathy, S. E. *et al.* Genome sequence of Fusobacterium nucleatum subspecies polymorphum - a genetically tractable fusobacterium. *PLoS One* **2**, e659 (2007).

123. Bardelčíková, A., Šoltys, J. & Mojžiš, J. Oxidative Stress, Inflammation and Colorectal Cancer: An Overview. *Antioxidants (Basel)* **12**, (2023).

124. Stidham, R. W. & Higgins, P. D. R. Colorectal Cancer in Inflammatory Bowel Disease. *Clin. Colon Rectal Surg.* **31**, 168–178 (2018).

125. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).

126. Shakya, M., Lo, C.-C. & Chain, P. S. G. Advances and Challenges in Metatranscriptomic Analysis. *Front. Genet.* **10**, 904 (2019).

127. Grasberger, H. *et al.* DUOX2 variants associate with preclinical disturbances in microbiota-immune homeostasis and increased inflammatory bowel disease risk. *J. Clin. Invest.* **131**, (2021).

128. McCord, J. M., Keele, B. B., Jr & Fridovich, I. An enzyme-based theory of obligate anaerobiosis: the physiological function of superoxide dismutase. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 1024–1027 (1971).

129. Henson, M. A. & Phalak, P. Microbiota dysbiosis in inflammatory bowel diseases: in silico investigation of the oxygen hypothesis. *BMC Syst. Biol.* **11**, 145 (2017).

130. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31**, 371–373 (2003).

131. Pearce, F. G., Dobson, R. C. J., Jameson, G. B., Perugini, M. A. & Gerrard, J. A. Characterization of monomeric dihydrodipicolinate synthase variant reveals the importance of substrate binding in optimizing oligomerization. *Biochim. Biophys. Acta* **1814**, 1900–1909 (2011).

132. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).

133. PyMOL. https://pymol.org/.

134. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).

135. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

136. Lemoine, F. & Gascuel, O. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genom Bioinform* **3**, lqab075 (2021).

137. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

138. Juste, C. & Gérard, P. Cholesterol-to-Coprostanol Conversion by the Gut Microbiota: What We Know, Suspect, and Ignore. *Microorganisms* **9**, (2021).

139. Deng, C., Pan, J., Zhu, H. & Chen, Z.-Y. Effect of Gut Microbiota on Blood Cholesterol: A Review on Mechanisms. *Foods* **12**, 4308 (2023).

140. Sekimoto, H., Shimada, O., Makanishi, M., Nakano, T. & Katayama, O. Interrelationship between Serum and Fecal Sterols. *Japanese Journal of Medicine* **22**, 14–20 (1983).

141. Sekimoto, H., Shimada, O., Makanishi, M., Nakano, T. & Katayama, O. Interrelationship between Serum and Fecal Sterols. *Japanese Journal of Medicine* **22**, 14–20 (1983).

142. Kriaa, A. *et al.* Microbial Reduction of Cholesterol to Coprostanol: An Old Concept and New Insights. *Catalysts* **9**, 167 (2019).

143. Kriaa, A. *et al.* Microbial Reduction of Cholesterol to Coprostanol: An Old Concept and New Insights. *Catalysts* **9**, 167 (2019).

144. Deng, C., Pan, J., Zhu, H. & Chen, Z.-Y. Effect of Gut Microbiota on Blood Cholesterol: A Review on Mechanisms. *Foods* **12**, 4308 (2023).

145. Veiga, P. *et al.* Correlation between faecal microbial community structure and cholesterol-to-coprostanol conversion in the human gut. *FEMS Microbiol Lett* **242**, 81–86 (2005).

146. Eyssen, H. J., Parmentier, G. G., Compernolle, F. C., De Pauw, G. & Piessens-Denef, M. Biohydrogenation of sterols by Eubacterium ATCC 21,408--Nova species. *Eur. J. Biochem.* **36**, 411–421 (1973).

147. Sadzikowski, M. R., Sperry, J. F. & Wilkins, T. D. Cholesterol-reducing bacterium from human feces. *Appl Environ Microbiol* **34**, 355–362 (1977).

148. Mott, G., Brinkley, A. W. & Mersinger, C. L. Biochemical characterization of cholesterol-reducing Eubacterium. *Appl. Environ. Microbiol.* **40**, 1017–1022 (1980).

149. Freier, T., Beitz, D. C., Li, L. & Hartman, P. A. Characterization of Eubacterium coprostanoligenes sp. nov., a cholesterol-reducing anaerobe. *Int. J. Syst. Bacteriol.* **44**, 137–142 (1994).

150. Sadzikowski, M. R., Sperry, J. F. & Wilkins, T. D. Cholesterol-reducing bacterium from human feces. *Appl Environ Microbiol* **34**, 355–362 (1977).

151. Ren, D., Li, L., Schwabacher, A., Young, J. & Beitz, D. Mechanism of cholesterol reduction to coprostanol by Eubacterium coprostanoligenes ATCC 51222. *Steroids* **61**, 33–40 (1996).

152. Parmentier, G. & Eyssen, H. Mechanism of biohydrogenation of cholesterol to coprostanol by Eubacterium ATCC 21408. *Biochim Biophys Acta* **348**, 279–284 (1974).

153. Kenny, D. J. *et al.* Cholesterol metabolism by uncultured human gut bacteria influences host cholesterol level. *Cell Host Microbe* **28**, 245–257.e6 (2020).

154. Yang, X., Dubnau, E., Smith, I. & Sampson, N. S. Rv1106c from Mycobacterium tuberculosis is a 3beta-hydroxysteroid dehydrogenase. *Biochemistry* **46**, 9058–9067 (2007).

155. Chiang, Y.-R. *et al.* Study of anoxic and oxic cholesterol metabolism by Sterolibacterium denitrificans. *J Bacteriol* **190**, 905–914 (2008).

156. Chiang, Y.-R. *et al.* Study of anoxic and oxic cholesterol metabolism by Sterolibacterium denitrificans. *J Bacteriol* **190**, 905–914 (2008).

157. Yang, X., Dubnau, E., Smith, I. & Sampson, N. S. Rv1106c from Mycobacterium tuberculosis is a 3beta-hydroxysteroid dehydrogenase. *Biochemistry* **46**, 9058–9067 (2007).

158. Lin, L. *et al.* Genome-centric investigation of bile acid metabolizing microbiota of dairy cows and associated diet-induced functional implications. *ISME J* **17**, 172–184 (2023).

159. Crick, P. J. *et al.* Evaluation of novel derivatisation reagents for the analysis of oxysterols. *Biochem Biophys Res Commun* **446**, 756–761 (2014).

160. Tanaka, N. *et al.* Crystal structures of the binary and ternary complexes of 7 alpha-hydroxysteroid dehydrogenase from Escherichia coli. *Biochemistry* **35**, 7715–7730 (1996).

161. Nakamura, S. *et al.* Apo- and holo-structures of 3alpha-hydroxysteroid dehydrogenase from Pseudomonas sp. B-0831. Loop-helix transition induced by coenzyme binding. *J. Biol. Chem.* **281**, 31876–31884 (2006).

162. Pakhomova, S. *et al.* Crystal structure of short chain dehydrogenase reductase 9 (short chain dehydrogenase reductase 9C4) in complex with NADH. Preprint at https://doi.org/10.2210/pdb9b2g/pdb (2025).

163. Kasbekar, M. *et al.* Selective small molecule inhibitor of the Mycobacterium tuberculosis fumarate hydratase reveals an allosteric regulatory site. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7503–7508 (2016).

164. Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431 (1999).

165. Grimm, C. *et al.* The crystal structure of 3alpha -hydroxysteroid dehydrogenase/carbonyl reductase from Comamonas testosteroni shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family. *J. Biol. Chem.* **275**, 41333–41339 (2000).

166. Doden, H. L. *et al.* Completion of the gut microbial epi-bile acid pathway. *Gut Microbes* **13**, 1–20 (2021).

167. Mullowney, M. W., McClure, R. A., Robey, M. T., Kelleher, N. L. & Thomson, R. J. Natural products from thioester reductase containing biosynthetic pathways. *Nat. Prod. Rep.* **35**, 847–878 (2018).

168. Jacoby, C. *et al.* Gut Bacteria Metabolize Natural and Synthetic Steroid Hormones via the Reductive

OsrABC Pathway. *bioRxiv* (2024) doi:10.1101/2024.10.08.617280.

169. Arp, G. *et al.* Gut Bacteria Encode Reductases that Biotransform Steroid Hormones. *bioRxiv* (2024) doi:10.1101/2024.10.04.616736.

170. Wang, T. *et al.* An expanded metabolic pathway for androgen production by commensal bacteria. *Nature Microbiology* 1–15 (2025).

171. Lin, L. *et al.* Genome-centric investigation of bile acid metabolizing microbiota of dairy cows and associated diet-induced functional implications. *ISME J* **17**, 172–184 (2023).

172. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

173. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* **4**, 293–305 (2019).

174. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* **4**, 293–305 (2019).

175. Schirmer, M. *et al.* Linking microbial genes to plasma and stool metabolites uncovers host-microbial interactions underlying ulcerative colitis disease course. *Cell Host Microbe* **32**, 209–226.e7 (2024).

176. Website. https://doi.org/10.1080/08910600500519854 doi:10.1080/08910600500519854.

177. Website. https://academic.oup.com/femsle/article/242/1/81/584220?login=true.

178. High-fat diet impact on intestinal cholesterol conversion by the microbiota and serum cholesterol levels. *iScience* **26**, 107697 (2023).

179. Juste, C. & Gérard, P. Cholesterol-to-Coprostanol Conversion by the Gut Microbiota: What We Know, Suspect, and Ignore. *Microorganisms* **9**, (2021).

180. Cohn, J. S., Kamili, A., Wat, E., Chung, R. W. S. & Tandy, S. Dietary phospholipids and intestinal cholesterol absorption. *Nutrients* **2**, 116–127 (2010).

181. Ikeda, I. & Kato, M. Mechanisms of intestinal cholesterol absorption. *Oleoscience* **12**, 107–114 (2012).

182. Wang, D. Q. H. New concepts of mechanisms of intestinal cholesterol absorption. *Ann. Hepatol.* **2**, 113–121 (2003).

183. Song, Y., Kenworthy, A. K. & Sanders, C. R. Cholesterol as a co-solvent and a ligand for membrane proteins. *Protein Sci.* **23**, 1–22 (2014).

184. Veiga, P. *et al.* Correlation between faecal microbial community structure and cholesterol-to-coprostanol conversion in the human gut. *FEMS Microbiol Lett* **242**, 81–86 (2005).

185. Li, C. *et al.* Gut microbiome and metabolome profiling in Framingham heart study reveals cholesterol-metabolizing bacteria. *Cell* **187**, 1834–1852.e19 (2024).

186. Plant sterols and human gut microbiota relationship: An in vitro colonic fermentation study. *Journal of Functional Foods* **44**, 322–329 (2018).

187. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

188. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).

189. Xiang, D. F., Thoden, J. B., Ghosh, M. K., Holden, H. M. & Raushel, F. M. Reaction mechanism and three-dimensional structure of GDP-d-glycero-α-d-manno-heptose 4,6-dehydratase from Campylobacter jejuni. *Biochemistry* **61**, 1313–1322 (2022).

190. PyMOL. https://www.pymol.org/.

191. Parks, D. H. *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.* **50**, D785–D794 (2022).

192. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

193. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

194. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**, 59–60 (2014).

195. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

196. Lemoine, F. & Gascuel, O. Gotree/Goalign: toolkit and Go API to facilitate the development of

phylogenetic workflows. *NAR Genom. Bioinform.* **3**, lqab075 (2021).

197. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

198. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

199. Honda, A. *et al.* Highly sensitive analysis of sterol profiles in human serum by LC-ESI-MS/MS. *J. Lipid Res.* **49**, 2063–2073 (2008).

200. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

201. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).

202. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* **4**, 293–305 (2019).

203. Krueger, F. *TrimGalore: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data*. (Github).

204. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

205. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

206. Pandey, K. R., Naik, S. R. & Vakil, B. V. Probiotics, prebiotics and synbiotics- a review. *J. Food Sci. Technol.* **52**, 7577–7587 (2015).

207. Ma, J. *et al.* Engineered probiotics. *Microb. Cell Fact.* **21**, 72 (2022).