

CATALYTIC MECHANISMS AND PHYSIOLOGICAL CONSEQUENCES OF  
MICROBIAL BILE ACID CONJUGATION

By

Douglas Van Allen Guzior

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Microbiology and Molecular Genetics – Doctor of Philosophy

2024

## ABSTRACT

Human bile has been studied for over 170 years and yet we are routinely reminded of how little we know. Early medicine considered bile an essential component of the four 'humors' governing health. This gradually transitioned to determining the structures of cholic acid and chenodeoxycholic acid, the two primary bile acids (BAs) present in humans, followed by investigating the nuances behind further modifications to these BA by bacteria in our intestines. Yet, prior to 2019, scientific dogma was that BA conjugation with glycine and taurine was solely performed by the host in the liver. Then, Quinn and colleagues described how bacteria in our gut are capable of ligating amino acids leucine, phenylalanine, and tyrosine to cholic acid. This was the first description of microbially conjugated bile acids (MCBAs). Given the recency of their discovery, the mechanisms behind MCBA production and their physiological relevance remained unknown prior to the work presented here.

In Chapter 2, I describe the *in vitro* acyl transfer of amino acids to BA by the enzyme bile salt hydrolase/transferase (BSH/T). I show that purified BSH/T from *Clostridium perfringens* is capable of transferring amino acids to taurocholic acid, glycocholic acid, and free cholic acid. I identify the pH optimum for this transfer and show that all 20 proteinaceous amino acids are not used. Finally, I examine the reaction kinetics of phenylalanine transfer to taurocholic acid.

In Chapter 3, I begin teasing apart the taxonomic diversity of bacteria capable of BA conjugation within the gastrointestinal tract. Because *Enterocloster bolteae* was the first bacterium implicated in MCBA production, culture-based screening for other producers focused on members of the *Lachnospiraceae* family. 19 of 29 species screened

demonstrated the ability to produce MCBAs and clustered based on amino acid use and total abundance of MCBAs produced. However, these groups did not correlate with taxonomy. Further analysis revealed instead that MCBA profiles correlated based on BSH/T amino acid sequence, leading to three distinct classes based on MCBA profile. I then compared MCBA production between wild type and variants containing active site substitutions to further understand how active site structure impacts MCBA production.

Chapter 4 begins to describe the physiological relevance of MCBAs from the level of individual bacteria to microbial communities, to human health and development. Given the antimicrobial effects of free BAs, I show how the hydrophobicity of the ligated amino acid impacts overall MCBA antimicrobial efficacy. I then show how high oral MCBA dosing correlates with shifts in the gut microbiome and that, at a lower dose, MCBAs are capable of entering enterohepatic circulation and infiltrating several tissues. Transitioning to direct human relevance, I show that MCBAs are enriched in a patient cohort undergoing sleeve gastrectomy surgery and shift dramatically following the operation. Finally, I shift from analyzing the BA pool in the context of gut dysbiosis to gastrointestinal development and describe fecal microbiome and metabolome changes through the first 12 months of life in an infant cohort. Certain classes, such as those resulting from host BA detoxification, show marked changes with time. These include MCBAs, where prevalence decreases as the infant matures.

*Science is iterative, building a body of tentative knowledge over time. No one does science alone, but is, in a sense, always in dialogue with their progenitors going back to Aristotle and Thales. It's really quite crowded in the lab when you're working alone.*  
– Zachary D. Blount

## ACKNOWLEDGEMENTS

It truly takes a village to train a graduate student and I would not be here without the army of people who have supported me through this journey. First, I want to thank Dr. Robert Quinn for allowing me to join his lab as his inaugural graduate student. From traveling across the pond to traveling across the continent, I am immensely grateful for the opportunities I had throughout my time in your lab. The diversity in training I received and the support in traveling for conferences are things I will value for the rest of my career. I can only hope to have had a fraction of the impact on you that you had on me, and I'm excited to watch you continue to develop and grow the lab.

Thank you to the members of my committee – Dr. Robert Hausinger, Dr. Linda Mansfield, Dr. Laura McCabe, and Dr. Gemma Reguera. Whether during a committee meeting or simply in the hall, having such a supportive group of scientific juggernauts to bounce ideas off was invaluable towards my development not only as a scientist, but as a member of the greater scientific community. I would also like to thank Dr. Victor DiRita. Your eagerness to find time to mentor a graduate student, without any hard requirement to do so, continues to be an inspiration and I truly cherish our chats over a simple cup of coffee.

I am immensely grateful for the brilliant collaborators I had the chance to work with during my time at Michigan State. Thank you to Dr. Jenna Wurster and Dr. Peter Belenky from Brown University; Dr. Julie Lumeng from the University of Michigan; Dr. Hilary Browne from University College Cork; Dr. Yan Shou, Dr. Claudio Durán, Dr. Bastian Haak, Dr. Andre Mu, Mr. Nicholas Dawson, and Dr. Trevor Lawley from the Wellcome Sanger Institute; Dr. Stewart Graham from Corewell Health; and Mr. Hao Wu and Dr. Gustavo de

los Campos from the Department of Epidemiology and Biostatistics at MSU. It has been both an honor and a pleasure to work with and learn from each of you. Additionally, Dr. Anthony Schillmiller and Dr. Casey Johnny from the MSU Research Technology Support Facility Mass Spectrometry Core deserve particular recognition. Whether it was to get the Q Exactive up and running again after a mistake I made or simply to talk through a half-baked metabolomics idea, thank you.

To the members of the Quinn lab, both past and present, you made my time in the lab exciting and it was my pleasure to work alongside you. To Dr. Christian Martin, for being there from the beginning of my graduate journey and for helping process infant samples analyzed in Chapter 4. You have been an incredible lab mate and brilliant resource as I learned how to navigate the world of metabolomics. Thank you to Yousi Fu for expanding the size of the bile acid team when you joined and for helping with many of the mouse experiments discussed in Chapter 4. To Dr. Kerri Neugebauer and Dr. Lydia-Ann Ghuneim, thank you for being both wonderful lab mates and wonderful mentors. To Mx. Lo Sosinski, thank you for not only helping me learn how to write R code but for being a wonderful friend. To Chris Bridges, thank you for allowing me the honor of being your mentor. I continue to admire your persistence, rigor, and scientific curiosity and wish you the best as you transition to your own graduate work. To Dr. Cely Gonzales, Dr. Hansani Karunarathne, and Dr. Nina Rosset, thank you for joining the lab and for your continued persistence in pushing the boundaries of science. I count myself fortunate to have had the opportunity to work alongside you.

I am incredibly grateful for the friends I have made throughout my time at MSU. Thank you to Carson Broeker for being a wonderful roommate and an even better friend.

Your Midwestern niceness is only matched by your perseverance and hard work. Thank you to my accountabilibuddy, Jasper Gomez. You have one of the sharpest scientific minds I know and a commitment to push the bounds of knowledge. There is nothing you can't do. To Kati Ford and Beth Ottosen, thank you both for years of friendship and for helping an old dog learn new tricks. I am immensely thankful for the time you took to show me the ins and outs of molecular biology, and I owe every ounce of cloning success to you. I would also be remiss if I didn't thank the Misfits, both those already mentioned in addition to Chris Speicher, Jenn Speicher, and May Napora. Playing dodgeball or kickball with our ragtag group of free agents-turned-close friends gave me something to look forward to every week (and an outlet to help work frustrations out).

Some friendships waver out between chapters in life; others are those you forge for a lifetime. Thank you to the nerdiest people I know, to the members of our self-proclaimed "nerd friends" – Logan Briggs, Jimmy Carl, Johan Harris, Jake Hilliker, Zakary Kadish, Ian Rice, and Spencer Schmitt. It's mind-boggling to think about how we've known each other for more than half of our lives; even more to think about each of our separate paths only to continue to keep in touch. Ryan Seltzer Houk deserves an honorable mention here; thank you for your constant willingness to hop online, hang out, and listen. To have such a supportive group of guys is something I will continue to appreciate for decades to come.

To the Sunday D&D squad – Logan Briggs, Jacob Griest, Jake Payne, Jon Shea, Natalie Lucas-Youngblood, and Nicholas Youngblood. Your creativity, week in and week out, has been a source of inspiration and your friendship means the world to me. Natalie deserves particular recognition here; thank you for your eagerness and willingness to

generate art for my research presentations. To have such a talented, scientifically-minded artist is a major boon, and your work has been included in every single talk I gave throughout my tenure as a graduate student.

It cannot be overstated how important my family is to me. First and foremost, I need to thank the two people who were my first role models, my parents, Pete and Sue. Your confidence and support remained unwavering throughout my time at MSU. You have shown me how to be kind while fair, confident while respectful, humble while proud, and patient while persistent. I would not be the man I am today without you, and I hope to continue to make you proud. To David, my younger brother, for keeping me humble and always managing to bring a smile to my face. Finally, to my step-parents, Dan and Beth, and step-brother, Finn, for always making space and being there for me. I am beyond lucky to have such a supportive and loving family. I also need to thank Peter Morgan here. You have been there for every big moment of my life, and I am eternally grateful to have you be part of it.

Finally, I owe so much of my success to the woman who has been here the whole time. Thank you, Halle. I would not have made it through this journey if not for your unwavering love, support, and care. Thank you for always having my back, you mean the world to me.

## TABLE OF CONTENTS

LIST OF TABLES .....	xi
LIST OF FIGURES .....	xiii
LIST OF ABBREVIATIONS .....	xv
CHAPTER 1: INTRODUCTION .....	1
1.1 - Preface .....	2
1.2 - Abstract .....	3
1.3 - Introduction .....	4
1.4 - Deconjugation .....	10
1.5 - Dehydroxylation .....	13
1.6 - Oxidation and Epimerization .....	15
1.7 - Isomerization .....	20
1.8 - Reconjugation: microbially conjugated bile acids .....	22
1.9 - Molecular diversity of microbially conjugated bile acids .....	24
1.10 - Microbial bile acid products and host health .....	27
1.11 - Conclusions.....	29
REFERENCES.....	31
CHAPTER 2: IDENTIFICATION AND CHARACTERIZATION OF ACYLTRANSFERASE ACTIVITY BY THE ENZYME BILE SALT HYDROLASE .....	42
2.1 - Preface .....	43
2.2 - Abstract .....	44
2.3 - Introduction .....	45
2.4 - Results .....	46
2.5 - Discussion.....	52
2.6 - Methods .....	53
2.7 - Data availability.....	55
REFERENCES.....	57
APPENDIX A: SUPPLEMENTARY TABLES .....	60
APPENDIX B: SUPPLEMENTARY FIGURES .....	61
CHAPTER 3: DIVERSITY OF BACTERIA CAPABLE OF MCBA PRODUCTION AND THEIR ASSOCIATED CONJUGATED BILE ACID PRODUCTS .....	63
3.1 - Preface .....	64
3.2 - Abstract .....	65
3.3 - Introduction .....	66
3.4 - Results .....	68
3.5 - Discussion.....	80
3.6 - Methods .....	82
3.7 - Data availability.....	85
REFERENCES.....	87
APPENDIX A: SUPPLEMENTARY TABLES .....	91

APPENDIX B: SUPPLEMENTARY FIGURES .....	102
CHAPTER 4: INTERPLAY BETWEEN MICROBIALLY CONJUGATED BILE ACIDS, THE MICROBIOME, AND THE METABOLOME .....	103
4.1 - Preface .....	104
4.2 - Abstract .....	106
4.3 - Introduction .....	108
4.4 - Results .....	109
4.5 - Discussion.....	133
4.6 - Methods .....	138
4.7 - Data availability.....	150
REFERENCES.....	152
APPENDIX A: SUPPLEMENTARY TABLES .....	161
APPENDIX B: SUPPLEMENTARY FIGURES .....	178
CHAPTER 5: CLOSING REMARKS .....	183
5.1 - Conclusions and significance .....	184
5.2 - Future directions .....	188
5.3 - Concluding remarks .....	190
REFERENCES.....	191

## LIST OF TABLES

Table 2.1: Abundance of amino acids used in acyl transfer when provided different BA substrates .....	48
Table 2.2: Goodness of fit for curves fit to determine pH optimum for amino acid acyl transfer by <i>C. perfringens</i> BSH.....	60
Table 3.1: Strains used in this work.....	91
Table 3.2: Individual amino acid use in conjugation for strains within MCBA profile cluster 1 .....	94
Table 3.3: Individual amino acid use in conjugation for strains within MCBA profile cluster 2 .....	95
Table 3.4: Individual amino acid use in conjugation for strains within MCBA profile cluster 3 .....	96
Table 3.5: Individual amino acid use in conjugation for strains within MCBA profile cluster 4 .....	97
Table 3.6: Individual amino acid use in conjugation for strains within MCBA profile cluster 5 .....	98
Table 3.7: Publicly available genome sequences for <i>Lachoclostridium scindens</i> used in phylogenetic analysis and BSH/T prediction .....	99
Table 3.8: Annotated MCBAs used for peak integration in mutagenesis studies .....	100
Table 3.9: Primers used in for <i>C. perfringens</i> <i>bsh/t</i> cloning and mutagenesis experiments .....	101
Table 4.1: Top 30 ASVs contributing to random forest classification of cecal samples by 100 mg kg <sup>-1</sup> MCBA gavage group.....	161
Table 4.2: Top 30 ASVs contributing to random forest classification of fecal samples by 100 mg kg <sup>-1</sup> MCBA gavage group.....	163
Table 4.3: Summary of previously reported MCBA concentrations in murine and human samples .....	165
Table 4.4: BA concentrations in murine tissue and feces following 10 mg kg <sup>-1</sup> MCBA dosing via PBFM.....	166
Table 4.5: List of BAs present in mass spectrometry standards.....	168
Table 4.6: Individual BA concentrations in human sleeve gastrectomy patient cohort.....	169

Table 4.7: BA concentrations, based on class, in human sleeve gastrectomy patient cohort.....	171
Table 4.8: Results from PERMANOVA testing of infant metabolome Bray-Curtis dissimilarity .....	172
Table 4.9: Results from PERMANOVA testing of infant microbiome Bray-Curtis dissimilarity .....	173
Table 4.10: EnvFit results based on Bray-Curtis dissimilarity for infant fecal metabolome data.....	174
Table 4.11: EnvFit results based on Bray-Curtis dissimilarity for infant fecal microbiome data.....	176

## LIST OF FIGURES

Figure 1.1: Diversity of known human bile acids .....	6
Figure 1.2: BA deconjugation reactions and enzyme homology present between gut bacteria .....	12
Figure 1.3: Pathway of bacterial dehydroxylation of primary BAs .....	14
Figure 1.4: Pathways of CA and CDCA epimerization .....	18
Figure 1.5: Pathways of allo-BA formation from 7 $\alpha$ -dehydroxylation intermediates .....	21
Figure 1.6: Potential increased diversity of the host BA pool as a result of MCBA production .....	25
Figure 2.1: <i>C. perfringens</i> BSH/T produces a broad range of MCBA at acidic pH .....	47
Figure 2.2: CpBSH/T deconjugation and acyl transfer kinetic characterization.....	51
Figure 2.3: TCA deconjugation by commercially available <i>C. perfringens</i> BSH/T at pH 3-10.....	61
Figure 2.4: pH-dependency of MCBA production by <i>C. perfringens</i> BSH/T .....	62
Figure 3.1: MS <sup>2</sup> -based molecular networking illustrates the unknown conjugated BA diversity.....	69
Figure 3.2: Dissimilarity between MCBA-producing strains based on amino acid use in conjugation .....	71
Figure 3.3: MCBA product identities correlate with BSH/T amino acid sequences .....	72
Figure 3.4: <i>Lachnoclostridium scindens</i> genome analysis for putative <i>bsh/t</i> annotation .....	74
Figure 3.5: BSH/T partial sequence alignment of strains screened for MCBA production .....	76
Figure 3.6: Nonessential active site residues drive amino acid selectivity in MCBA production .....	78
Figure 3.7: GCA and TCA extracted ion chromatograms following 24 h induction of <i>C. perfringens</i> BSH/T variants in <i>E. coli</i> .....	79
Figure 3.8: Crystal structure of <i>C. perfringens</i> BSH/T with co-crystalized taurine and DCA .....	102

Figure 4.1: MCBA show varied antimicrobial properties.....	111
Figure 4.2: Amino acid-dependency of MCBA antimicrobial efficacy .....	113
Figure 4.3: Broad microbiome community shifts following 100 mg kg <sup>-1</sup> MCBA gavage .....	114
Figure 4.4: Random Forest classification of murine microbiome community structure following 100 mg kg <sup>-1</sup> MCBA gavage.....	116
Figure 4.5: SerCA concentrations following 100 mg kg <sup>-1</sup> feeding.....	118
Figure 4.6: MCBA concentrations in fecal and tissue samples following mixed MCBA dosing via PBFM.....	120
Figure 4.7: BA concentrations in mouse tissue samples following MCBA feeding and in human feces of patients undergoing sleeve gastrectomy .....	122
Figure 4.8: MCBA-containing sample proportions across the first 12 months of life ....	124
Figure 4.9: Temporal shifts in alpha-diversity within infant fecal metabolomes and microbiomes driven by richness .....	126
Figure 4.10: Temporal beta-diversity shifts within infant fecal metabolome and microbiome communities .....	127
Figure 4.11: Univariate effects of 33 covariables on multi-omic sample dissimilarity ...	128
Figure 4.12: Longitudinal changes in presence/absence are highly correlated for certain pairs of metabolites and ASVs.....	131
Figure 4.13: Microbial and metabolite features with significant temporal shifts in zero- proportions.....	132
Figure 4.14: Microbiome community shifts following 10 mg kg <sup>-1</sup> MCBA dosing via PBFM.....	178
Figure 4.15: Extracted ion chromatograms of PheCA and SerCA exposed to pancreatic carboxypeptidases.....	179
Figure 4.16: Expected change in probability of zero for metabolites and ASVs by feature and individuals .....	180
Figure 4.17: Microbial and metabolite features with significant correlations in temporal shifts of zero-proportions .....	181
Figure 4.18: MS <sup>2</sup> comparison between putative cholestane glucuronide and annotated cholestane .....	182

## LIST OF ABBREVIATIONS

12-ECA	12-epicholic acid
AGC	Automatic gain control
AlaCA	Alanochoolic acid
AlaDCA	Alanodeoxycholic acid
Allo-BA	Allo-bile acid
Amp	Ampicillin
Ara	Arabinose
ArgCA	Arginochoolic acid
AsnCA	Asparagochoolic acid
AspCA	Aspartochoolic acid
ASV	Amplicon sequence variant
ATCC	American Type Culture Collection
auc	Area under the curve
BA	Bile acid
BAAT	Bile acid-CoA:amino acid N-acyltransferase
BAL	Bile acid-coenzyme A ligase
BEH	Bridged ethyl hybrid
BSH/T	Bile salt hydrolase/transferase
CA	Cholic acid
CCUG	Culture Collection, University of Gothenburg
CDCA	Chenodeoxycholic acid
CitCA	Citrullochoolic acid

CitDCA	Citrullodeoxycholic acid
Cm	Chloramphenicol
CpBSH/T	<i>Clostridium perfringens</i> bile salt hydrolase/transferase
CysCA	Cystocholic acid
DCA	Deoxycholic acid
DMSO	Dimethyl sulfoxide
EHC	Enterohepatic circulation
FC	Fold change
FC-AUC	Fold change in area under curve
FXR	Farnesoid x receptor
GABA	$\gamma$ -aminobutyric acid
GCA	Glycocholic acid
GCDCA	Glycochenodeoxycholic acid
GDCA	Glycodeoxycholic acid
GlnCA	Glutamocholic acid
GluCA	Glutamatocholic acid
GluCDCA	Glutamatochenodeoxycholic acid
GluDCA	Glutamatodeoxycholic acid
GNPS	Global Natural Products Social Molecular Networking
HCA	Hyocholic acid
HDCA	Hyodeoxycholic acid
HisCA	Histidocholic acid
HisDCA	Histidodeoxycholic acid

HPLC	High performance liquid chromatography
HSDH	Hydroxysteroid dehydrogenase
IBD	Inflammatory bowel disease
IleCA	Isoleucocholic acid
IQR	Inter-quartile range
isoCA	Isocholic acid
isoCDCA	Isochenodeoxycholic acid
LB	Lysogeny broth
LC-MS/MS	Liquid chromatograph-tandem mass spectrometry
LCA	Lithocholic acid
LeuCA	Leucocholic acid
LysCA	Lysocholic acid
LysUDCA	Lysoursocholic acid
<i>m/z</i>	Mass-to-charge ratio
MCA	Muricholic acid
MCBA	Microbially conjugated bile acid
MetCA	Methionocholic acid
MetDCA	Methionodeoxycholic acid
MSU	Michigan State University
NMDS	Non-metric data scaling
OD <sub>600</sub>	Optical density at 600 nm
OrnCA	Ornithocholic acid
PBFM	Peanut butter feeding method

PC	Principal component
PCA	Principal component analysis
PCo	Principal coordinate
PcoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational analysis of variance
PheCA	Phenylalanochoic acid
PheDCA	Phenylalanodeoxychoic acid
PheHDCA	Phenylalanoxydeoxychoic acid
ProCA	Prolochoic acid
RCM	Reinforced clostridial medium
RT	Retention time
S1PR2	Sphingosine-1-phosphate receptor 2
SDM	Site-directed mutagenesis
SerCA	Serochoic acid
SG	Sleeve gastrectomy
SPF	Specific pathogen-free
TCA	Taurochoic acid
TCDCA	Taurochenodeoxychoic acid
TDCA	Taurodeoxychoic acid
ThrCA	Threonochoic acid
ThrCDCA	Threonochenodeoxychoic acid
TrpCA	Tryptophanochoic acid

TrpDCA	Tryptophanodeoxycholic acid
TyrCA	Tyrosocholic acid
TyrUDCA	Tyrosoursodeoxycholic acid
UCA	Ursocholic acid
UDCA	Ursodeoxycholic acid
UPLC	Ultra performance liquid chromatography
ValCA	Valocholic acid
WT	Wild type

**CHAPTER 1:**  
**INTRODUCTION**

## 1.1 - Preface

The contents of this chapter were originally published in *Microbiome* in 2021 (Material from: Guzior, D.V., Quinn, R.A. Review: microbial transformations of human bile acids. *Microbiome* **9**, 140 (2021). <https://doi.org/10.1186/s40168-021-01101-1>). Per the publisher BioMed Central, “This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.” The full description of the Creative Commons license governing the original publication can be found at [creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/). This text has been revised slightly to reflect scientific advances in the field of bile acid metabolism since the time of its original publication.

## 1.2 - Abstract

Bile acids play key roles in gut metabolism, cell signaling, and microbiome composition. While the liver is responsible for the production of primary bile acids, microbes in the gut modify these compounds into myriad forms that greatly increase their diversity and biological function. Since the early 1960s, microbes have been known to transform human bile acids in four distinct ways: deconjugation of amino acids glycine or taurine, and dehydroxylation, dehydrogenation, and epimerization of the cholesterol core. Alterations in the chemistry of these secondary bile acids have been linked to several diseases, such as cirrhosis, inflammatory bowel disease, and cancer. In addition to the previously known transformations, a recent study has shown that members of our gut microbiota are also able to conjugate amino acids to bile acids, representing a new set of “microbially conjugated bile acids.” This new finding greatly influences the diversity of bile acids in the mammalian gut, but the effects on host physiology and microbial dynamics are mostly unknown. This review focuses on recent discoveries investigating microbial mechanisms of human bile acids and explores the chemical diversity that may exist in bile acid structures in light of the new discovery of microbial conjugations.

## 1.3 - Introduction

### 1.3.1 - *The history of bile*

Bile has been implicated as important in human health for millennia. Hippocrates developed the idea of humourism in the third century BC, which describes the body as being composed of four ‘humors,’ two of which involve bile. When these humors are balanced the body is healthy, but illness occurs when any become unbalanced (1). Even today, we are still trying to understand how the delicate balance between different bile acid (BA) concentrations throughout the body is associated with states of health or disease. Our gut microbiome, the consortium of microorganisms living in our gastrointestinal system, is a major mediator of BA chemistry and, consequently, the development of healthy or diseased states. For example, abnormally high levels of the microbially modified secondary BA deoxycholic acid ( $3\alpha, 12\alpha$ -dihydroxy- $5\beta$ -cholan-24-oic acid, DCA) is associated with gut dysbiosis and disease (2, 3). There has been increased research in recent years on the connection between our gut microbiome, BA pool composition, and human health, all of which build on our knowledge from the previous two millennia of BA chemistry.

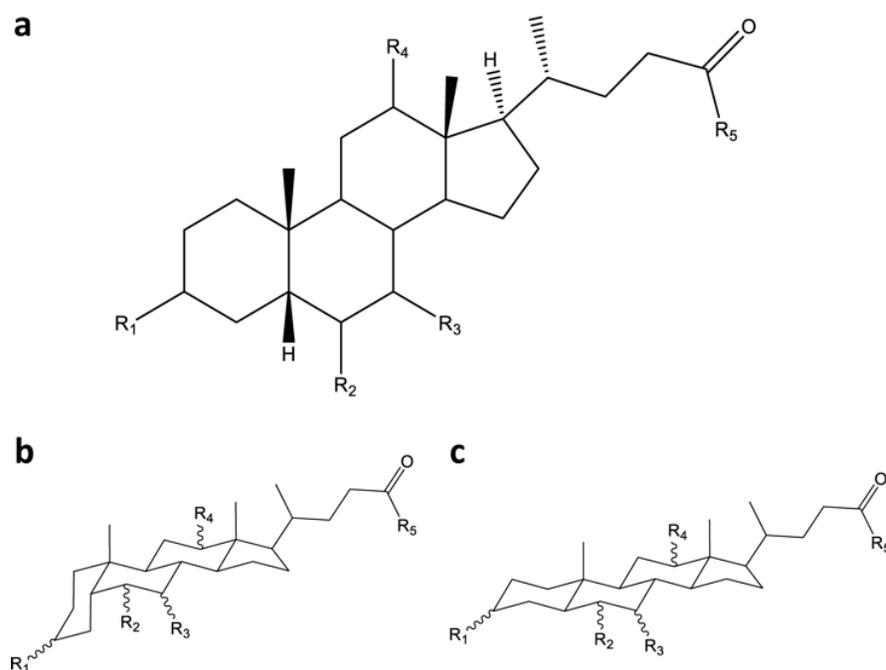
### 1.3.2 - *Bile acid biochemistry and physiology*

Primary BAs are those synthesized in the liver from cholesterol (4). The primary BA pool in humans consists of cholic acid ( $3\alpha, 7\alpha, 12\alpha$ -trihydroxy- $5\beta$ -cholan-24-oic acid, CA), chenodeoxycholic acid ( $3\alpha, 7\alpha$ -dihydroxy- $5\beta$ -cholan-24-oic acid, CDCA), and subsequent C24 taurine- or glycine-bound derivatives (Figure 1.1). Glycine and taurine bound BAs are also referred to as bile salts due to their low  $pK_a$  and complete ionization resulting in these compounds being present as anions *in vivo* (5–7). For the purposes of

this review, all compounds will be referenced in their protonated form, being named conjugated bile acids *in lieu* of conjugated bile salts. Primary BAs are heavily modified in the lower gastrointestinal tract to produce a broad range of secondary BAs (Figure 1.1). This microbial metabolism is so extensive that instead of primary BAs having the highest prevalence in human cecal contents, microbially modified BAs DCA (a CA derivative) and lithocholic acid (3 $\alpha$ -hydroxy-5 $\beta$ -cholan-24-oic acid, LCA, a CDCA derivative) are the most prevalent, together accounting for over 50% of the bile acid pool (8). On average, secondary BAs reached concentrations of 200  $\mu$ M for DCA and 160  $\mu$ M for LCA (9).

Relevant BAs within humans are not limited to hydroxylation at C3, C7, and C12, but are also found to be hydroxylated at C6 as is the case for  $\alpha$ -muricholic acid (3 $\alpha$ , 6 $\beta$ , 7 $\alpha$ -trihydroxy-5 $\beta$ -cholan-24-oic acid,  $\alpha$ MCA) and  $\beta$ -muricholic acid (3 $\alpha$ , 6 $\beta$ , 7 $\beta$ -trihydroxy-5 $\beta$ -cholan-24-oic acid,  $\beta$ MCA). Muricholic acids are predominant in mice and scarce in humans, though not absent. MCA forms of bile acids are present in infant urine and feces, though they decrease in concentration to below detectable levels in adults (10, 11). Due to their predominance in mice and rats, MCAs are important in gastrointestinal research using animal models (12).

BAs have traditionally been thought to undergo amino acid conjugation solely in the liver. There is a single human enzyme, bile acid-CoA:amino acid N-acyltransferase (hBAAT), that is responsible for acyl-conjugation. These conjugated primary BAs are secreted via the bile canaliculi into the gallbladder where they are stored until consumption of a meal. They are then secreted into the duodenum and travel through the small intestine, only to be subsequently reabsorbed in the terminal ileum and transported to the liver for re-conjugation, if necessary, followed by secretion into the gallbladder and



Bile Acid	Abbreviation	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>
Cholic acid	CA	α	H	α	α	OH
Chenodeoxycholic acid	CDCA	α	H	α	H	OH
Glycocholic acid	GCA	α	H	α	α	Glycine
Taurocholic acid	TCA	α	H	α	α	Taurine
Glychenodeoxycholic acid	GCDCA	α	H	α	H	Glycine
Taurochenodeoxycholic acid	TCDCA	α	H	α	H	Taurine
α-muricholic acid	αMCA	α	β	α	H	OH
β-muricholic acid	βMCA	α	β	β	H	OH
Deoxycholic acid	DCA	α	H	H	α	OH
Lithocholic acid	LCA	α	H	H	H	OH
Allodeoxycholic acid	aDCA	α	H	H	α	OH
Allolithocholic acid	aLCA	α	H	H	H	OH
7-epicholic acid	7-ECA	α	H	β	α	OH
Ursodeoxycholic acid	UDCA	α	H	β	H	OH
Isocholic acid	iCA	β	H	α	α	OH
Isochenodeoxycholic acid	iCDCA	β	H	α	H	OH
Isodeoxycholic acid	iDCA	β	H	H	α	OH
Isolithocholic acid	iLCA	β	H	H	H	OH
12-oxochenodeoxycholic acid	12-oxoCDCA	α	H	α	oxo	OH
7-oxodeoxycholic acid	7-oxoDCA	α	H	oxo	α	OH
3-oxocholic acid	3-oxoCA	oxo	H	α	α	OH
7-oxolithocholic acid	7-oxoLCA	α	H	oxo	H	OH
3-oxochenodeoxycholic acid	3-oxoCDCA	oxo	H	α	H	OH
Phenylalanochoic acid	PheCA	α	H	α	α	Phenylalanine
Tyroschoic acid	TyrCA	α	H	α	α	Tyrosine
Leuchoic acid	LeuCA	α	H	α	α	Leucine

### Figure 1.1: Diversity of known human bile acids

**a**, All BAs are built off the same sterol backbone with variations in hydroxylated positions, hydroxyl orientation, and the presence of ketones. CA and CDCA, along with GCA, GCDCA, TCA, and TCDCA, make up the primary BA pool. Remaining BAs in the list make up secondary and tertiary BA pools as a result of modifications from gut microbes (13–15). Allo-bile acids, although matching in hydroxyl positions to their standard bile acid counterparts, differ in ring orientation. Standard bile acids have the first ring in the **b**, *trans*-orientation, yielding 5β-BAs, while allo-bile acids have this ring in the **c**, *cis*-orientation, yielding 5α-BAs.

Recirculation(16). Enterohepatic circulation is very efficient, recirculating approximately 95% of secreted bile acids, including some of those modified by the microbiota (17). The remaining 5% undergoes a myriad of transformations throughout the gastrointestinal tract (13, 18). Although the specific chemistry of BA reabsorption is not completely elucidated, it is generally understood that conjugated BAs are actively transported by ileal transporters and some passive diffusion across the gut epithelium can occur for both conjugated and non-conjugated BAs, specifically those conjugated to glycine (17, 19). GCA and other glycine conjugates may be able to undergo passive diffusion due to the relatively small change in BA biochemistry caused by glycine conjugation.

BAs play an important role in regulating various physiological systems, such as fat digestion, cholesterol metabolism, vitamin absorption, liver function, and enterohepatic circulation through their combined signaling, detergent, and antimicrobial mechanisms (20). BAs are agonists of the Farnesoid X receptor (FXR), with varying degrees of activity depending on the structure of the compound (21). CDCA is the most potent FXR agonist, followed by DCA, LCA, and lastly, CA. Though their effects on FXR are less clear and more research is needed, conjugated BAs have also been observed to play a role as FXR agonists, notably within the small intestine where concentrations can reach as high as 10 mM (22, 23). FXR is responsible for regulating several steps in the synthesis of primary BAs CA and CDCA. The loss of FXR activity in mice results in metabolic perturbations and loss of host BA regulation (24). FXR plays a major role in protecting the small intestine from bacterial overgrowth, regulating key antimicrobial pathways including inducible nitric oxide synthase, IL-18, angiogenin production, and production of several antimicrobial peptides, such as those within the Defa gene family (22, 25). Tauro-BAs, specifically

T $\beta$ MCA, have also been shown to act as FXR antagonists, inhibiting BA synthesis via negative regulation (26). Additionally, BAs are agonists of g-protein coupled receptors such as TGR5 (Takeda G protein-coupled receptor 5) and S1PR2 (sphingosine-1-phosphate receptor 2). S1PR2 is expressed ubiquitously within the liver while TGR5 is expressed primarily in non-parenchymal cells (27). Expression of both S1PR2 and TGR5 is a balancing act within the liver between homeostasis and damage. S1PR2 is activated by conjugated BAs and results in pro-inflammatory effects that can increase liver damage while TGR5 is activated by all BAs along with several other steroids and results in anti-inflammatory effects in addition to anti-cholestatic and anti-fibrotic effects (27). These characteristics make S1PR2 inhibitors and TGR5 agonists attractive candidates for drug development.

### 1.3.3 - Microbial bile acid interactions

Bile acids are potent antimicrobials. As such, they play an important role in the innate immune defense within the intestine. Consequently, modifications of BAs are an essential microbial defense mechanism (28). BAs have been known to impact susceptible bacteria in both a bacteriostatic and bactericidal fashion since the late 1940s, impacting such genera as *Staphylococcus*, *Balantidium*, *Pneumococcus*, and *Enterococcus* in addition to members of the phylum Spirochaetes (29). BAs act as detergents in the gut and support the absorption of fats through the intestinal membrane. These same properties allow for the disruption of bacterial membranes. Primary BAs disrupt membranes in a dose-dependent fashion and non-conjugated BAs exact a greater reduction in viability than their conjugated counterparts when tested against *Staphylococcus aureus*, several *Lactobacillus* species, and several *Bifidobacterium*

species (28, 30). As a result of the conjugation to glycine or taurine, primary BAs are fully ionized at physiological pH. While this is important in the movement of BAs from the liver, complete ionization prevents significant interaction and passive diffusion across bacterial membranes whereas non-conjugated CA and CDCA are able to disrupt membranes, cross them, and cause intracellular damage (31). Conjugated BAs can have more indirect action on the gut microbiota, however, because at high concentrations in the small intestine they modulate FXR and other ileal receptors which control bile synthesis.

#### *1.3.4 - Microbial bile acid transformation pathways*

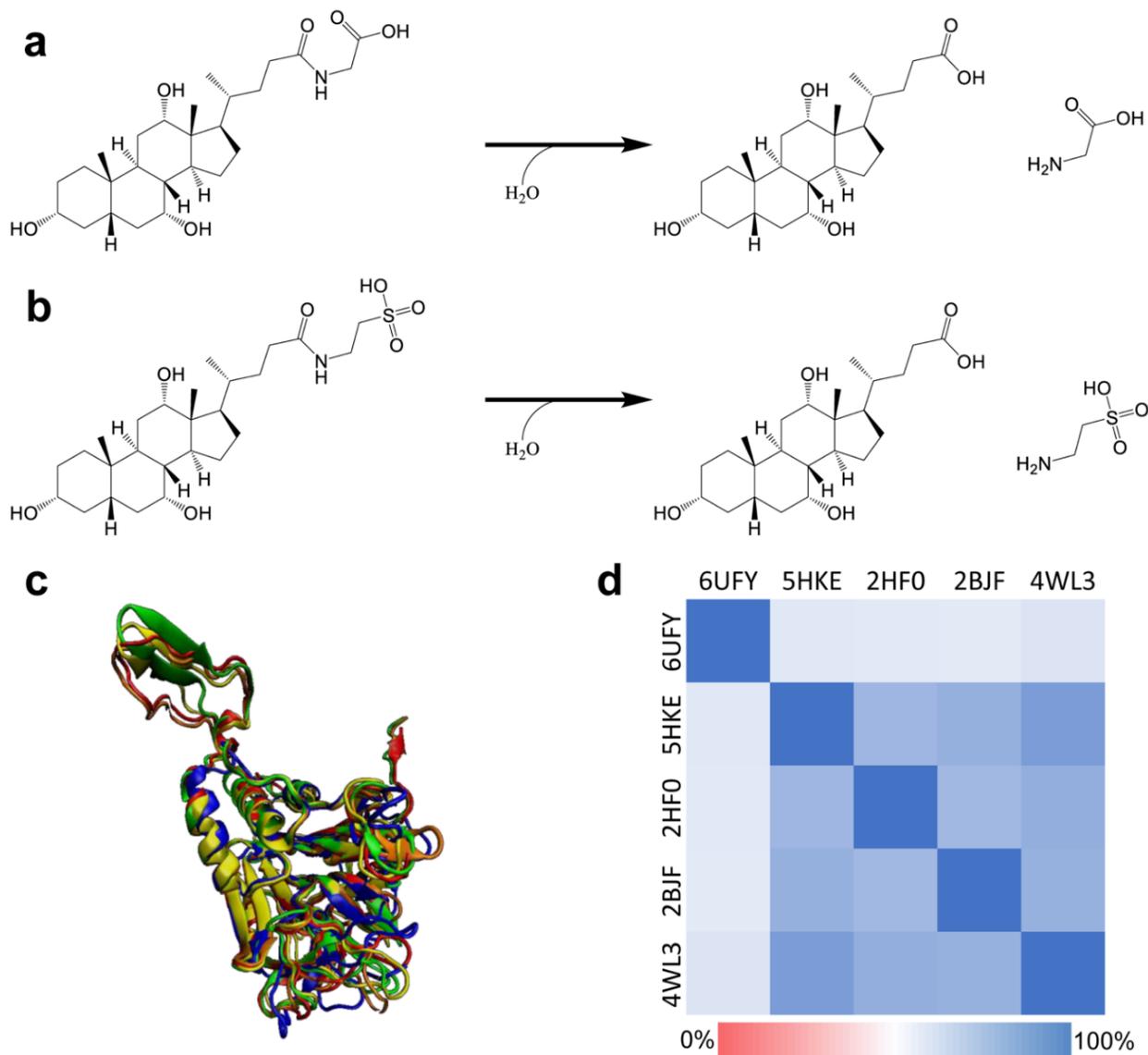
Traditionally, there have been four distinct pathways related to microbial transformations of BAs: deconjugation, dehydroxylation, oxidation, and epimerization. The latter two methods of BA transformations work hand in hand, as formation of oxoBAs is a key step prior to epimerization. Research into microbial bile salt hydrolases (BSHs) has been the latest boom in health-related BA research since their discovery in the 1970s with over 260 publications listed on PubMed from within the last 10 years (search term 'bile salt hydrolase', search performed in 2021). Additionally, several reviews have been written specifically about the biochemistry, diversity, and implications of microbially transformed BAs on host health (18, 32, 33). The diversity of BAs has recently been shown to be higher than originally thought as members of the gut microbiota demonstrated the ability to conjugate amino acids to cholic acid independent of the host liver (13).

## 1.4 - Deconjugation

Deconjugation of BAs is considered the “gateway reaction” to further modification (34). There are several hypotheses that could explain the importance of deconjugation. As previously discussed, deconjugated primary BAs can act as signaling molecules which modify the total bile acid pool, and therefore, the microbiota may have evolved the deconjugation mechanism to manipulate bile production further. Deconjugation also results in increased concentrations of antimicrobial BAs, CA and CDCA, that may drive shifts in microbiome composition and act as a possible form of microbial chemical warfare. BSHs (classified as EC 3.5.1.24) are able to deconjugate both glycine- and taurine-bound primary BAs, though differences in activity may indicate BSH substrate specificity (18). Members of the gut microbiota may also use the liberated glycine and taurine residues as nutrient sources. Regardless, deconjugation is an essential function of the gut microbiome. Enzymes capable of catalyzing the deconjugation reaction are found across all major bacterial phyla and within major archaeal species, suggesting that the genes encoding them are horizontally transferable (35, 36). *Bacteroides* spp. are among the phyla suggested to play a major role in deconjugating primary BAs (37). The diversity of bacteria capable of amino acid hydrolysis includes Gram-positive genera such as *Bifidobacterium* (38), *Lactobacillus* (39, 40), *Clostridium* (41), *Enterococcus* (42), and *Listeria* (43). However, BSH activity is not limited to Gram-positive bacteria. Gram-negatives such as *Stenotrophomonas* (44), *Bacteroides* (45), and *Brucella* (46) also contribute to amino acid hydrolysis within the gut. In the cases of *Brucella abortus* and *Listeria monocytogenes*, BSH genes are important for virulence and establishing infection within mouse models. A metagenomic study by Jones *et al.* found BSH-encoding genes

are conserved among all major bacterial and archaeal species within the gut (34). Bacteria capable of BSH activity comprise 26.03% of identified strains of gut bacteria present in humans, although some of these strains may be in low abundance as only 26.40% of BSH-capable strains are present in human guts throughout the globe (47). The mere ubiquity of BSHs in the gut exemplifies their importance to our microbiota.

All BSH reactions rely on amide bond hydrolysis in order to free taurine or glycine (Figure 1.2a-b). Optimal BSH activity occurs at neutral or slightly acidic pH (5–7) with reported optima around pH 6 (41, 48, 49). Interestingly, among *Bifidobacterium* spp. arose three separate classes of BSH within which two of the classes showed high activity and differed in substrate specificity (38). Both classes exhibited a preference for glycine-conjugated BAs but varied in activity for taurine-conjugated BAs. Although BSHs may utilize both taurine and glycine conjugates, encoding many BSHs may allow for slight changes in substrate specificity and more specific manipulation of the bile acid pool. BSH enzymes from *Ligilactobacillus salivarius* (PDB ID: 5HKE) (50, 51), *Bifidobacterium longum* (PDB ID: 2HF0) (52, 53), *Bacteroides thetaiotaomicron* (PDB ID: 6UFY) (54, 55), *Clostridium perfringens* (PDB ID: 2BJF) (56, 57), and *Enterococcus faecalis* (PDB ID: 4WL3) (58) have been crystalized (Figure 1.2c). Comparing structural homology (Figure 1.2d), *E. faecalis*, *L. salivarius*, *B. longum*, and *B. thetaiotaomicron* each maintained the  $\alpha\beta\alpha$  motif indicating that it is essential for activity (47). The BSH from *B. thetaiotaomicron* (Figure 1.2c, blue) is missing a turn which may be one of the driving factors for the decreased structural homology between the other crystalized BSHs. Analysis of key residues from *L. salivarius*, *B. longum*, *E. faecalis*, and *C. perfringens* amino acid



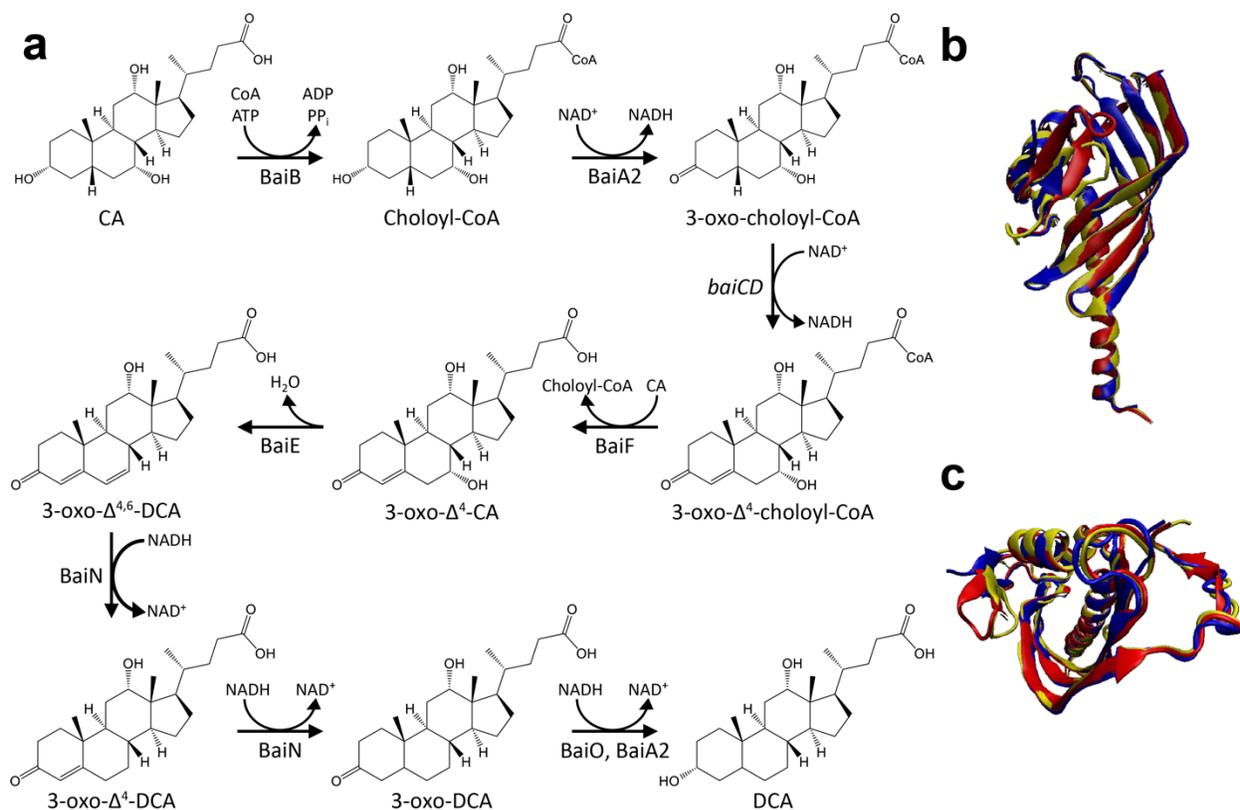
**Figure 1.2: BA deconjugation reactions and enzyme homology present between gut bacteria**

Regardless of hydroxylation positions, substitution of water for either **a**, glycine or **b**, taurine yields the same products. **c**, Structural homology between subunits from *B. thetaiotaomicron* (6UFY, blue), *L. salivarius* (5HKE, red), *B. longum* (2HF0, yellow), *C. perfringens* (2BJF, green), and *E. faecalis* (4WL3, orange) using Visual Molecular Dynamics (VMD) software(59). **d**, Structural homology ( $Q_H$ ) was measured utilizing VMD with a minimum of 0.5804 and a maximum of 0.8533. *E. faecalis* and *L. salivarius* BSHs had the greatest similarity while *B. thetaiotaomicron* was the most dissimilar to all other organisms.

sequences demonstrated highly conserved residues throughout the BSH active site across each genus (47).

### 1.5 - Dehydroxylation

One of the key transformations by gut microbes is BA dehydroxylation at C7. Within *Lachnoclostridium scindens*, formerly *Clostridium scindens*, the *bai* operon encodes several proteins needed for the sequential oxidation of CA (60). The *baiG* gene encodes a bile acid transporter, allowing for CA uptake. BaiG is also capable of transporting CDCA and DCA (61). This is followed by CoA ligation in an ATP-dependent manner by BaiB to form choloyl-CoA. Choloyl-CoA is then oxidized twice, first by BaiA and followed by BaiCD, to yield 3-oxo- $\Delta$ 4-choloyl-CoA. BaiF is then hypothesized to transfer CoA from 3-oxo- $\Delta$ 4-choloyl-CoA to CA, yielding 3-oxo- $\Delta$ 4-CA and choloyl-CoA (15). BaiF CoA transferase activity has already been observed with DCA-CoA, LCA-CoA, and allo-DCA-CoA acting as donors and CA acting as an acceptor (60). The rate limiting step occurs during the dehydroxylation of C7 via BaiE, a 7 $\alpha$ -dehydratase. The genes involved in CA 7 $\alpha$ -dehydroxylation are capable of recognizing intermediates in the CDCA dehydroxylation pathway as well. Interestingly, CoA-conjugation at C24 was not necessary for dehydratase activity to occur with CA as the substrate, and in some cases enabled for greater  $k_{cat}$  and lower  $K_M$  (62). Crystal structures of BaiE have been generated in the ligand-absent conformation from *L. scindens* (PDB ID: 4LEH) (63), *Lachnoclostridium hylemonae* (formerly *Clostridium hylemonae*, PDB ID: 4L8O) (64), and *Peptacetobacter hiranonis* (formerly *Clostridium hiranonis*, PDB ID: 4L8P) (65) (Figure 1.3). Each unit displayed structural similarity (QH) greater than 85%, as calculated in Visual Molecular Dynamics (VMD) (59). The enzymes responsible



### Figure 1.3: Pathway of bacterial dehydroxylation of primary BAs

CA (R: -OH) and CDCA (R: -H). **a**, The pathway to complete  $7\alpha$ -dehydroxylation is a multi-stage process that involves progressive substrate oxidation, likely for molecule stability, prior to dehydroxylation, followed by reduction at each previously oxidized position along the sterol backbone(60). The enzyme capable of dehydroxylation, BaiE, is highly conserved structurally between *L. scindens* (red), *C. hylemonae* (blue), and *P. hiranonis* (yellow), evident in both **b**, side and **c**, top-down views of BaiE.

for the reductive arm of BA 7 $\alpha$ -dehydroxylation within *L. scindens* are encoded by *baiN*, which is responsible for the sequential reduction of C6-C7 and C4-C5 after dehydroxylation, and by *baiA2*, which catalyzes the NADH-dependent 3-oxoreduction of both 3-oxodeoxycholic acid (3-oxoDCA) and 3-oxolithocholic acid(3-oxoLCA) (66, 67). BaiO is proposed to carry out a similar function to BaiA2 in the reductive arm of 7 $\alpha$ -dehydroxylation, though this has not yet been verified experimentally (15). This form of bile acid metabolism appears to be limited to members of the class Clostridia. *Dorea*, *Flavonifractor*, *Pseudoflavonifractor*, *Proteocatocola*, and *Ruminococcus* genera have been reported to harbor genes required for 7 $\alpha$ -dehydroxylation, in addition to the well-studied *Lachnoclostridium* and *Peptacetobacter* (68, 69).

7 $\beta$ -dehydroxylation occurs in a similar fashion, the key difference being that BaiH is used in the place of BaiCD for C4 oxidation (70, 71). 7 $\beta$ -dehydratase activity is likely the rate limiting step in 7 $\beta$ -dehydroxylation similar to BaiE above, though the exact gene has not yet been identified. This indicates that further research is needed to elucidate the impact and prevalence of organisms capable of 7 $\beta$ -dehydroxylation, especially given the relative absence of 7 $\beta$  BAs.

## 1.6 - Oxidation and Epimerization

Epimerization of BAs is carried out by gut microbes and further diversifies the chemistry of secondary BAs. This occurs in two distinct steps: oxidation of the hydroxyl group by a position-specific hydroxysteroid dehydrogenase, such as a 7 $\alpha$ -HSDH, followed by the reduction of another position-specific hydroxysteroid dehydrogenase, 7 $\beta$ -HSDH. Both reactions do not need to be carried out by the same organism and co-cultures of microbes are known to possess epimerization capabilities (72). CA can be

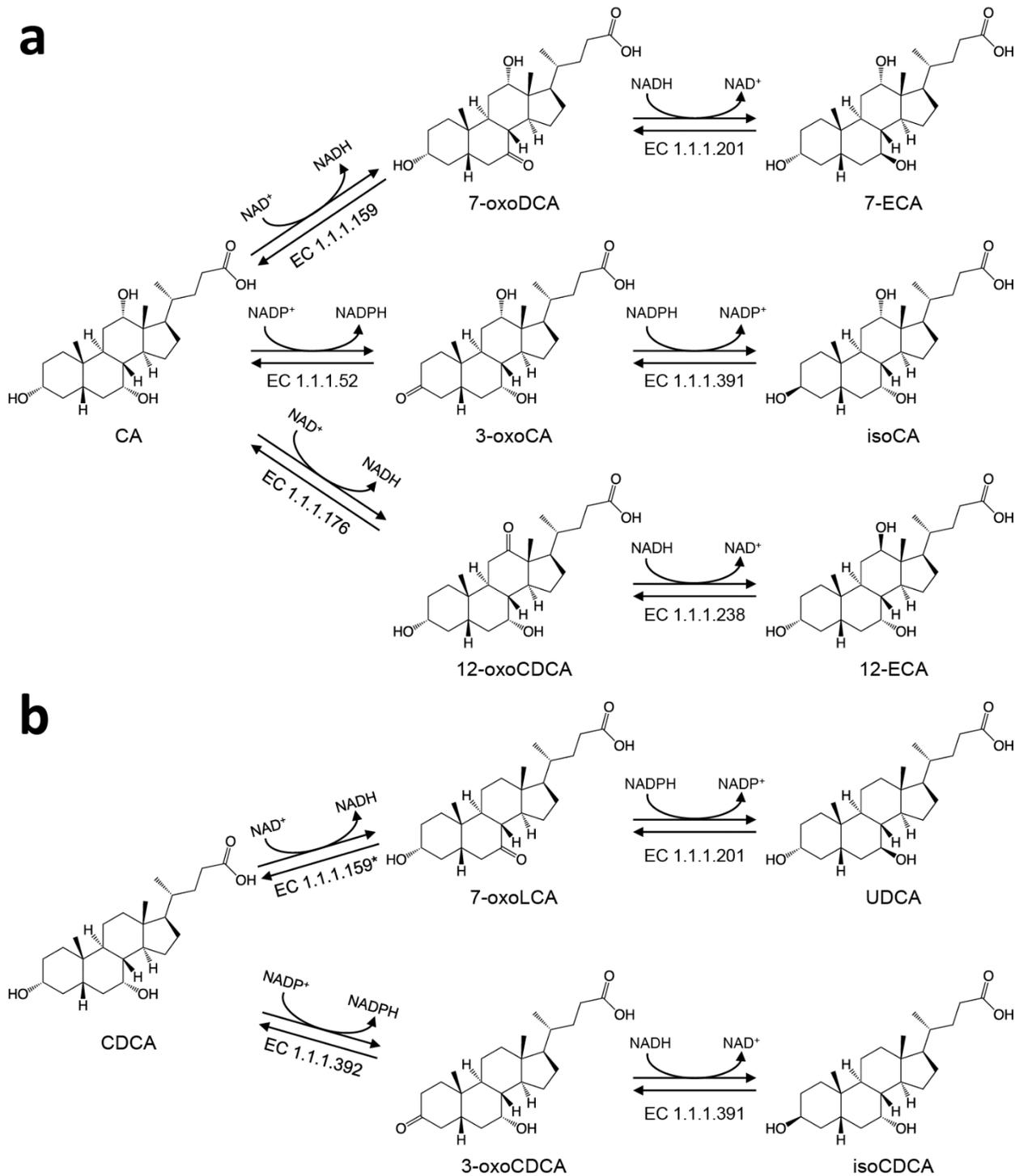
epimerized to form derivatives such as ursolic acid ( $3\alpha$ ,  $7\beta$ ,  $12\alpha$ -trihydroxy- $5\beta$ -cholan-24-oic acid, UCA), 12-epicholic acid ( $3\alpha$ ,  $7\alpha$ ,  $12\beta$ -trihydroxy- $5\beta$ -cholan-24-oic acid, 12-ECA), or isocholic acid ( $3\beta$ ,  $7\alpha$ ,  $12\alpha$ -trihydroxy- $5\beta$ -cholan-24-oic acid, isoCA) (Figure 1.4a), while CDCA can be epimerized to form either UDCA or isochenodeoxycholic acid ( $3\beta$ ,  $7\alpha$ -Dihydroxy- $5\beta$ -cholan-24-oic acid, isoCDCA) (Figure 1.4b). Both oxidation and subsequent epimerization have been observed at all three CA hydroxyl positions as well as both CDCA hydroxyl positions and are responsible for much of the diversity found in non-conjugated BAs.

Recently, *L. scindens*, *L. hylemonae*, *C. perfringens*, and *P. hiranonis* have all been observed to produce enzymes capable of hydroxysteroid  $3\alpha$ -dehydrogenation, an important step in the pathway toward  $7\alpha$ -dehydroxylation (36). However, unlike *L. scindens*, *L. hylemonae*, and *P. hiranonis*, *C. perfringens* has not been reported to produce LCA or DCA and its growth is inhibited by both secondary BAs (73).  $3\alpha$ -dehydrogenation also occurs outside of the genus *Clostridium* and includes other intestinal organisms such as *Blautia producta* and *Eggerthella lenta* (formerly *Eubacterium lentum*) in addition to environmental species such as *Acinetobacter lwoffii* (66, 74, 75). 3-oxoLCA production has also been observed within the genera *Adlercreutzia*, *Collinsella*, *Gordonibacter*, *Monoglobus*, *Peptoniphilus*, *Phocea*, and *Raoultibacter* (76). Of those, *Raoultibacter* was not observed to fully convert LCA to isoLCA but successfully converted 44% of provided LCA to 3-oxoLCA. Surprisingly, *E. lenta*  $3\alpha$ -HSDH is capable of utilizing both tauro-BAs and glyco-BAs as substrates and in the case of CDCA oxidation,  $3\alpha$ -HSDH activity increased when conjugated forms of CDCA were used as substrates (77). This goes against the notion that bile BA

deconjugation is the essential 'gateway' reaction and further investigation is required to elucidate if glycine and taurine residues impact molecular mechanisms of catalysis in addition to if conjugated BA oxidation impacts subsequent transformations.

Epimerization of CDCA, independent of conjugation, is important for producing the protective BA, UDCA. 7 $\alpha$ -epimerization to UDCA occurs in the gut by members such as *Clostridium baratii* among other isolates not yet identified (78, 79). *C. baratii* has been shown to epimerize CDCA to UDCA but was not capable of epimerizing glyco- and tauro-BAs and instead deconjugated TCDCA prior to epimerization (78).

*Ruminococcus gnavus*, *Clostridium absonum*, *Stenotrophomonas maltophilia*, and *Collinsella aerofaciens* all contribute to the UDCA pool via conversion of 7-oxoLCA in an NADH or NADPH-dependent fashion (75, 80–82). Optimum pH varied between species; *C. absonum* 7 $\beta$ -HSDH functioned optimally at pH 8.5 while *R. gnavus* and *C. aerofaciens* functioned optimally at pH 6. 12 $\beta$ -HSDH activity can occur in both acidic and alkaline conditions. *R. gnavus*, in contrast to *C. absonum* and *C. aerofaciens*, displayed a clear preference in catalyzing the conversion of 7-oxoLCA to UDCA with a specificity constant 55-fold higher than that of the conversion of UDCA to 7-oxoLCA (80). The specificity constant of an enzyme for a specific substrate, determined by the ratio of  $k_{cat}/K_m$ , where larger values correspond with greater catalytic efficiency compared to other substrates. This directionality of activity paired with the protective properties of UDCA support *R. gnavus* as a potential probiotic, and this role should be further investigated.



**Figure 1.4: Pathways of CA and CDCA epimerization**

**a**, CA undergoes three different epimerization pathways leading to the production of isoCA (via 3 $\alpha$ / $\beta$ -HSDH), UCA (via 7 $\alpha$ / $\beta$ -HSDH), or 12-ECA (via 12 $\alpha$ / $\beta$ -HSDH) while **b**, CDCA undergoes two distinct epimerization pathways leading to the production of UDCA (via 7 $\alpha$ / $\beta$ -HSDH) or isoCDCA (via 3 $\alpha$ / $\beta$ -HSDH). \**S. maltophilia* transforms CDCA to 7-oxoCDCA but the enzyme is categorized under EC 1.1.1.159, where the official reaction involves CA 7 $\alpha$ -oxidation (85).

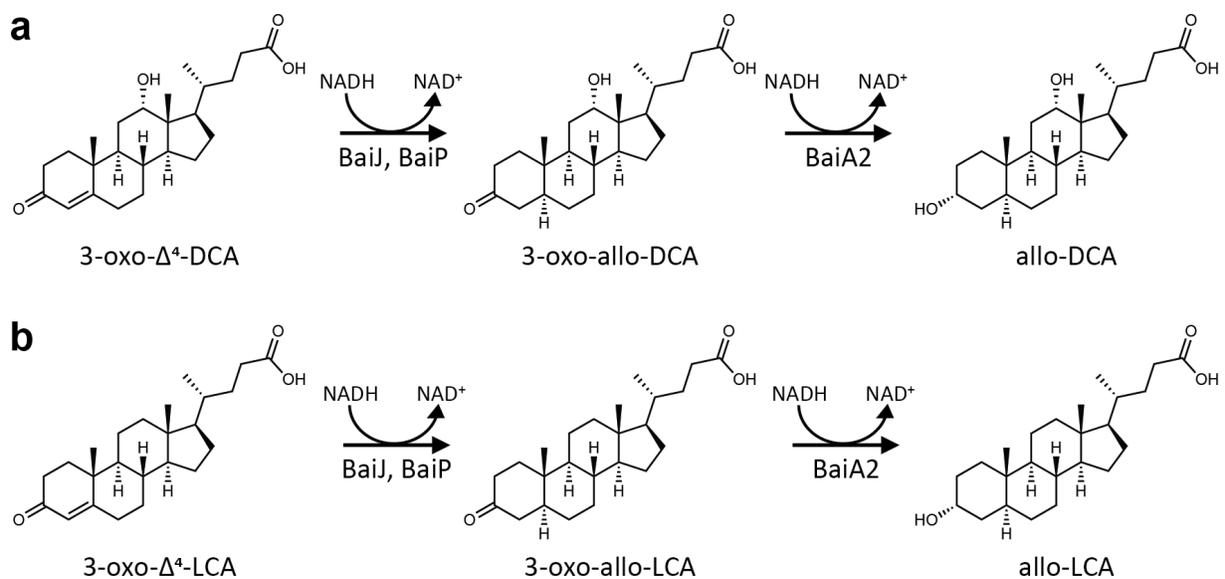
Several gut bacteria have recently been identified to produce 12 $\alpha$ -hydroxysteroid dehydrogenases (12 $\alpha$ -HSDH). *E. lenta* demonstrates 12 $\alpha$ -HSDH capabilities in addition to 3 $\alpha$ -HSDH. *E. lenta* 12 $\alpha$ -HSDH has an estimated molecular weight of 125 kDa and has a broad pH optimum, between pH 8 and 10.5 (83, 84). Catalysis requires NAD<sup>+</sup> or NADP<sup>+</sup> as a cofactor, though there is a preference for NAD<sup>+</sup> (66, 83). *E. lenta* 12 $\alpha$ -HSDH reaction velocity increased when tested with methylated BAs, suggesting a preference for hydrophobic BAs (84). Similar to its 3 $\alpha$ -HSDH, *E. lenta* 12 $\alpha$ -HSDH is capable of utilizing both glycine- and taurine-bound BAs (77). *Enterorhabdus mucosicola* is also capable of both 3 $\alpha$  and 12 $\alpha$  oxidation, although 12 $\alpha$ -HSDH activity is limited to when the C7 position has already been oxidized (86, 87). *L. scindens*, *P. hiranonis*, and *L. hylemonae* have since been reported to produce 12 $\alpha$ -HSDHs and it is hypothesized that *Bacteroides* species also encode 12 $\alpha$ -HSDHs (36, 66). Across all three clostridial species, there was a robust preference for 12-oxoLCA over 12-oxoCDCA suggesting the C7 hydroxyl group, or lack thereof, plays a large role in determining enzyme activity. Oxidation at C12 occurs for 12 $\beta$  BAs as well and has been observed in strains of *Clostridium paraputrificum*, *Clostridium tertium*, and *Clostridioides difficile* (88). These 12 $\beta$ -HSDHs are relatively stable at physiological conditions, maintaining activity at 37 °C for approximately 45 minutes at pH 8.5 (89). Based on these findings by Edenharder and Pfutzner, *C. paraputrificum* 12 $\beta$ -HSDH behaves in a similar manner to established 12 $\alpha$ -HSDHs, as shown by its pH optimum and molecular weight. The gene encoding the 12 $\beta$ -HSDH in *C. paraputrificum* was recently identified, allowing for investigation into the diversity of potential 12 $\beta$ -HSDH producers (90). Putative 12 $\beta$ -HSDH genes were found across Firmicutes, Actinobacteria, and Alphaproteobacteria. However, there may be

several forms of 12 $\beta$ -HSDH as the authors did not find homologs to the *C. paraputrificum* 12 $\beta$ -HSDH in *C. difficile* and *C. tertium* even though both species are capable of 12 $\beta$ -HSDH activity.

Members of the gut microbiota are not only capable of reducing BAs with a single position oxidized, but some also reduce BAs oxidized at two or three positions. Similar trends regarding non-target hydroxyl oxidation have been observed by other *Coriobacteriaceae*, such as *C. aerofaciens*, *E. lenta*, and *Lancefieldella parvula* (formerly *Atopobium parvulum*) (86). Not all members oxidized DCA at both C3 and C12 independent of the other position, but all of the strains observed to modify DCA were shown to oxidize at both positions (86). *L. scindens* and *P. hiranonis* were among the only bacteria capable of completely hydrogenating 3,7,12-trioxolithocholic acid, a fully oxidized derivative of CA, to CA (36). Oxidation may be a way for microbes to detoxify BAs. By decreasing their amphipathicity, oxidized BAs progressively lose the ability to act as detergents, preventing DNA and membrane damage.

### **1.7 - Isomerization**

Allo-bile acids (allo-BA), those with a 5 $\alpha$  ring resulting in a more planar structure, have received less attention in the field of microbial bile acid metabolism. *In vitro* production of allo-BAs, specifically allo-deoxycholic acid (3 $\alpha$ , 12 $\alpha$ -dihydroxy-5 $\alpha$ -cholan-24-oic acid, allo-DCA), were first associated with *L. scindens* VPI 12708 with production shown to be inducible when growing the bacterium in the presence of CA (91). After uncovering their microbial origins, the mechanism behind microbial BA isomerization then remained unknown for over 30 years. Recently, however, Lee *et al.* described BaiP, a BA-



**Figure 1.5: Pathways of allo-BA formation from  $7\alpha$ -dehydroxylation intermediates**  
 Formation of allo-bile acids **a**, allo-DCA from 3-oxo- $\Delta^4$ -DCA and **b**, allo-LCA from 3-oxo- $\Delta^4$ -LCA. 3-oxo- $\Delta^4$ -DCA and 3-oxo- $\Delta^4$ -LCA are intermediate products of  $7\alpha$ -dehydroxylation, and BaiA2 is shared across allo- and standard secondary BAs.

inducible 3-oxo- $\Delta^4$ -5 $\alpha$ -reductase produced by *L. scindens* ATCC 35704, and BaiJ, an isoform produced by *L. scindens* VPI 12708 and *L. hylemonae* DSM 15053 (68). BaiP and BaiJ convert 3-oxo- $\Delta^4$ -DCA and 3-oxo- $\Delta^4$ -LCA, intermediates in the 7 $\alpha$ -dehydroxylation pathway described above, to 3-oxo-allo-DCA and 3-oxo-allo-LCA respectively. These are then reduced to allo-DCA and allo-LCA by BaiA1 (Figure 1.5) (68).

The diversity of microbes capable of allo-BA production is an active area of research, but much remains to be determined. Screening bacterial isolates from allo-BA-enriched human centenarians for the BA isomerization showed that members of *Parabacteroides*, *Bacteroides*, *Alistipes*, and *Odoribacter* genera converted 3-oxo- $\Delta^4$ -LCA to 3-oxo-allo-LCA or iso-allo-LCA(92). Production of allo-LCA was minimal, potentially reflecting selective pressure due to high antimicrobial activity. Genomic analysis of *Proteocatella sphenisci* (93), first isolated from penguin guano, and *Peptacetobacter hiranonis*, a BA metabolizing bacterium and marker for canine health (94), suggest that these organisms encode the genes required for this transformation (68), but this needs to be experimentally validated.

### **1.8 - Reconjugation: microbially conjugated bile acids**

A novel set of recently discovered BAs were conjugated at the C24 acyl site similarly to the host conjugation mechanism (13). Instead of the traditional amino acids taurine and glycine, these compounds were conjugated with the amino acids phenylalanine, leucine, and tyrosine on a cholic acid backbone. The initial work associated these molecules with the gut microbiota and follow-up experiments identified the bacterium *Enterocloster bolteae*, formerly *Clostridium bolteae*, as a species responsible for their production. In light of their microbial origin and the mechanism

mirroring that of host-conjugation, we hereby refer to these compounds as “microbially conjugated bile acids” (MCBAs).

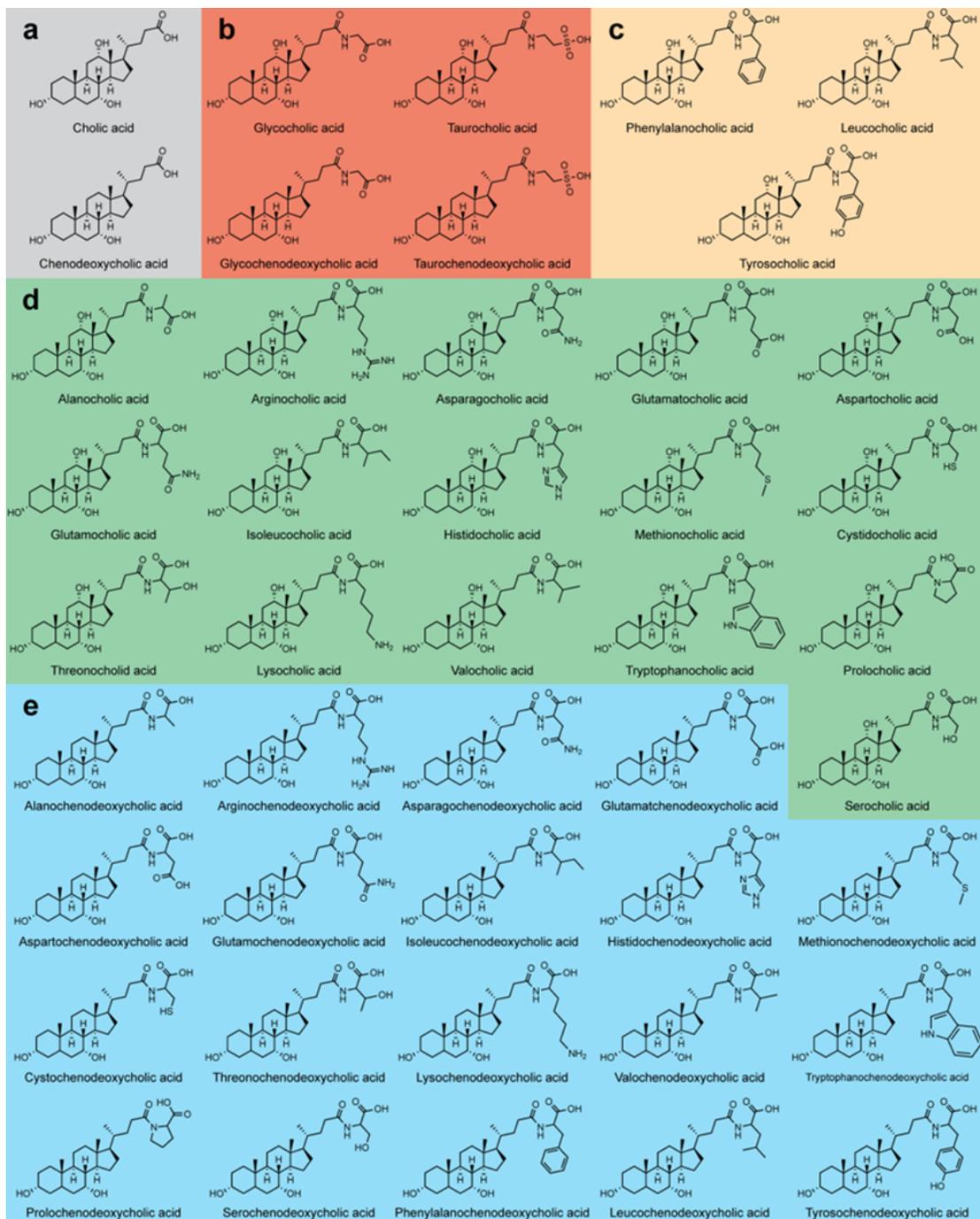
The exact mechanism of this microbially mediated conjugation had not been elucidated as of 2020, though it may rely on a similar mechanism to hBAAT within the liver involving a Cys-Asp-His triad, with cysteine functioning as the catalytic residue for nucleophilic attack (95). Regardless of their mechanism of production, the addition of unique amino acid chemistry on the BA acyl-site inevitably modifies its chemical properties. Phenylalanine, a large hydrophobic amino acid, will greatly increase the hydrophobicity of the BA itself and possibly induce steric hindrance to any binding mechanisms with ileal receptors or BA transporters. Leucine, too, is a relatively large hydrophobic residue, which may create similar chemical properties to that of phenylalanine. The additional hydroxyl group on the aromatic ring of tyrosine may create some unique properties as this will increase the compound’s hydrophilicity and create slightly a more polar, hydrophilic BA, similar to the increase in polarity provided by taurine conjugation to cholic acid though not as pronounced. The presence of any of these amino acids at the conjugation site will also alter the BA’s emulsifying properties, as a primary function of these compounds is to solubilize fat from our diet. Since their original discovery, the diversity of known MCBAs has increased dramatically. However, until the mechanism of their synthesis is enzymatically elucidated and exhaustive searches into MCBA diversity are performed, our knowledge of the limits on amino acid conjugation of BAs by the human microbiota remains incomplete.

The functions of phenylalanine, leucine, and tyrosine CA conjugates remain mostly unknown, though gavage of mice with these compounds has been shown to result in

agonism of FXR. Further investigation into the roles of known and unknown BA conjugates may yield novel drug targets or therapeutic agents for the treatment of numerous enteric diseases. Evidence already points toward BA hydrophilicity playing a major role in activity of several BA modifying enzymes; the three novel conjugates currently reported represent three of the four most lipophilic amino acids based on partition coefficient (96). Thus, identifying organisms responsible for conjugation of other amino acids to other BAs and amino acid-specific mechanisms are the necessary first steps to determining how microbes are utilizing these compounds to impact the host or competing members of the microbiota.

### **1.9 - Molecular diversity of microbially conjugated bile acids**

Over 140 amino acids are known to occur in natural proteins (97). The human BA pool consists of a sterol backbone capable of hydroxylation at four different positions (including C6, observed in MCA), which can be  $\alpha$ - or  $\beta$ -hydroxylated, oxidized to form a ketone, or absent. This backbone can also be present as one of two stereoisomers: 5 $\alpha$ -sterol or 5 $\beta$ -sterol, significantly broadening potential BA diversity. Limiting the bile acid backbone to only those known to be conjugated by the host (CA and CDCA) in addition to limiting the amino acid conjugated to those naturally occurring in humans, the potential diversity of the human conjugated BA pool increases over 5-fold from what is currently known (Figure 1.6). This estimate does not consider non-amino acid conjugates, such as ciliatocholic acid or choloyl-CoA, nor does it include the diversity of potential host hydroxyl modifications, such as sulfation (98, 99). Overall, the human bile acid pool is dominated by CA, CDCA, and DCA (100). Subsequent taurine and glycine conjugation increases



**Figure 1.6: Potential increased diversity of the host BA pool as a result of MCBA production**

With the current understanding of BA metabolism, **a**, primary BAs CA and CDCA are known to be conjugated in the liver to taurine and glycine to form **b**, GCA, TCA, GCDCA, and TCDCA, completing the pool of primary human BAs. In light of recent research, CA is also known to be conjugated by gut microbes to form **c**, PheCA, LeuCA, and TyrCA (13). Expanding the potential library of microbially conjugated BAs by including the remaining amino acids conjugates for **d**, CA and **e**, CDCA increases the diversity of human BAs over 5-fold for these backbones alone.

this pool to 9 BAs. Limiting the estimate of possible BA-amino acid conjugates to standard amino acids and the three BAs listed above increases the potential human BA pool to 66 unique conjugates. Finally, including all potential oxidized, epimerized, and dehydroxylated states of each hydroxyl group present on CA (C3, C7, C12) in addition to ring orientation expands the number of potential human BA conjugates to over 2800. Although it is unlikely that the number of physiologically relevant MCBAs is this high, one can imagine the potential diversity of MCBAs and the potential for their impact on the gut microbiota and the host.

Through 2020, only relatively hydrophobic amino acids have been reported to be conjugated to CA by microbes, lowering the overall partition coefficient of each molecule. The partition coefficient is the log-ratio of concentrations of a compound in a hydrophobic solvent, such as 1-octanol, compared to a hydrophilic solvent, such as water. This is to say that a higher value indicates that the compound is more present in the hydrophobic phase rather than the hydrophilic phase. As expected, hydrophobicity increases with the reduction of BAs. CA has a partition coefficient of 2.02, which increases to 3.28 when CA is reduced to CDCA and further increases to 3.5 when reduced to DCA (101). The conjugation of both glycine and taurine to any sterol significantly increases the hydrophilicity of the compound, thus decreasing the partition coefficient for each BA. Therefore, the acyl-conjugation of BAs undoubtedly affects their function. Similarly, microbial conjugation with hydrophobic amino acids would also affect their detergent, signaling, and antimicrobial properties, as well as BA transport. One may wonder then, why do gut microbiota conjugate our bile acids? There are a multitude of possible explanations ranging from enzyme promiscuity to antimicrobial metabolite production, to

targeted manipulation of the host BA signaling and regulatory system. Only further research on the genetic, biochemical, and microbiological characterization of the conjugation mechanism and its microbial and host effects will provide the answers. Nevertheless, the MCBA chemical diversity already detected in the mammalian gut, and the potential described above, will invariably diversify the chemical properties of the bile acid pool.

### **1.10 - Microbial bile acid products and host health**

Though BAs themselves function as important antimicrobial agents, microbial modification of BAs is equally important in disease prevention and maintenance of a healthy gut microbiome. Though *C. difficile* infections are devastating, fecal microbiota transplant can be a successful treatment in some cases. Successful transplants correlate with an increase in *bsh* copy number compared to levels prior to transplant, suggesting that microbial modifications of primary BAs play a role in protecting the host against microbial infection (102). The host microbiota plays an important role in protection against colonization by pathogenic organisms, and the involvement of BA modification in this protective effect is only beginning to be understood (103). Decreased bile acid deconjugation correlates with several irritable bowel diseases such as ulcerative colitis, Crohn's disease, and irritable bowel syndrome (35). Thus, supplementing diets with microbially transformed BAs may have profound and beneficial effects on host pathology.

LCA production, a result of CDCA dehydroxylation, is one of the more interesting transformations by gut microbes with a known impact on host health. LCA has been observed to act as an anti-inflammatory agent and protect against colitis in a mouse model (104). However, LCA and DCA, another the secondary BA, are known carcinogens.

While primary human bile acids are known to induce DNA damage within bacteria, LCA and DCA have been observed to damage DNA within mammalian cells (105). Recently, isoLCA and 3-oxoLCA were found to suppress TH<sub>17</sub> cell differentiation and 3 $\alpha$ / $\beta$ -HSDH gene abundance, required for their synthesis, varied between patients with IBD and non-IBD controls (76). DCA exposure has also been correlated with increased apoptosis and increased production of reactive oxygen and nitrogen species (106). These changes in eukaryotic cell death are a consequence of DCA intercalating into mitochondrial membranes at low concentrations (107, 108), while at concentrations above 250  $\mu$ M it causes cell death through necrosis (109). Notably, LCA and DCA are the most prevalent bile acids in human colorectal cancer (106).

Epimerized BAs also influence host health. UDCA, the 7 $\beta$  epimer of CDCA, exhibits protective effects in the gut, specifically through inhibition of TNF $\alpha$ , IL-1 $\beta$ , and IL-6 release (104). UDCA use has been shown to counteract the apoptotic effects of DCA (110). UDCA has also been approved for use in gallstone dissolution and in treating primary biliary cholangitis, the later indication as a result of the ability of UDCA to increase bile acid biosynthesis (104, 110). One caveat of UDCA use is that, at high doses (28–30 mg/kg/day), long-term use leads to increased risk of colorectal cancer in patients with ulcerative colitis and primary sclerosing cholangitis (111).

Products of BA isomerization are rarely identified in healthy human adults yet show significant enrichment in dysbiotic adults. Allo-BAs are primarily associated with pregnant women, fetuses, and infants, reaching trace concentrations in healthy adults (112–114). Interestingly, allo-LCA and iso-allo-LCA were enriched in a cohort of Japanese centenarians (92). However, allo-BAs are notably enriched in hepatocellular carcinoma

(115) and CRC patient cohorts (68), the latter reinforced by increased *baiP* and *baiJ* abundance across five CRC metagenomic datasets (68, 116–121). Traditionally underappreciated, the role of allo-BAs in carcinogenesis are only beginning to be uncovered.

It is possible that MCBAs may also play a role in disease mechanisms, as phenylalanochoholic acid (PheCA), tyrosochoholic acid (TyrCA), and leucochoholic acid (LeuCA) were more prevalent in patients with inflammatory bowel disease and cystic fibrosis (and though not disease related, were also found in infants) (13). However, one cannot know, simply by detection in a diseased population, whether MCBAs, or any BA for that matter, are cause or consequence of a particular diseased state; a conundrum that is well known in the microbiome field. There is evidence that at least one microbe that produces MCBAs, *E. bolteae* (referred to as *C. bolteae* in the referenced manuscript), may be involved in severe IBD and Crohn's disease, as it was identified as one of the most transcriptionally active microbes in the dysbiotic and diseased gut and MCBAs were elevated in these same samples (13, 122). This association indicates that MCBAs may be involved in severe IBD, but future research is required. Regardless, BAs can serve as markers for various disease states (105, 106) and can themselves be used as therapeutics, such as in the case of UDCA, making them an important group of compounds for identification and treatment of human disease.

### **1.11 - Conclusions**

Although BAs have been studied for centuries, recent discoveries show that we still have much to learn. The host BA pool controls microbial diversity, but so too does microbial metabolism of these BAs drive host physiology. In this sense, BAs act as the

language of an intricate molecular crosstalk between humans and their gut microbiota. Mechanisms of microbial modification of host BAs continue to be elucidated as do the roles that BA metabolism plays in host health. The presence of MCBAs in the human BA pool demonstrates the need for further study of microbial BA modification and further expands the chemical language our gut microbiota uses to communicate with its host.

## REFERENCES

1. Goodacre CJ, Naylor WP. 2020. Evolution of the temperament theory and mental attitude in complete denture prosthodontics: from Hippocrates to M.M. House. *J Prosthodont* 29:594–598.
2. Liu L, Dong W, Wang S, Zhang Y, Liu T, Xie R, Wang B, Cao H. 2018. Deoxycholic acid disrupts the intestinal mucosal barrier and promotes intestinal tumorigenesis. *Food Funct* 9:5588–5597.
3. Cao H, Xu M, Dong W, Deng B, Wang S, Zhang Y, Wang S, Luo S, Wang W, Qi Y, Gao J, Cao X, Yan F, Wang B. 2017. Secondary bile acid-induced dysbiosis promotes intestinal carcinogenesis. *Int J Cancer* 140:2545–2556.
4. Hofmann AF. 1999. The continuing importance of bile acids in liver and intestinal disease. *Arch Intern Med* 159:2647–2658.
5. Bortolini O, Medici A, Poli S. 1997. Biotransformations on steroid nucleus of bile acids. *Steroids* 62:564–577.
6. Russell DW. 2003. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* 72:137–174.
7. Moini J. 2019. Epidemiology of diet and diabetes mellitus. *Epidemiol Diabetes* 57–73.
8. Kakiyama G, Muto A, Takei H, Nittono H, Murai T, Kurosawa T, Hofmann AF, Pandak WM, Bajaj JS. 2014. A simple and accurate HPLC method for fecal bile acid profile in healthy and cirrhotic subjects: Validation by GC-MS and LC-MS. *J Lipid Res* 55:978–990.
9. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, Chatfield DA, Hagey LR, Steinbach JH, Hofmann AF. 2007. Human cecal bile acids: concentration and spectrum. *Am J Physiol-Gastrointest Liver Physiol* 293:G256–G263.
10. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. 2012. Targeted profiling of circulating and hepatic bile acids in human, mouse, and rat using a UPLC-MRM-MS-validated method. *J Lipid Res* 53:2231–2241.
11. Goto J, Hasegawa K, Nambara T, Iida T. 1992. Studies on steroids. CCLIV. Gas chromatographic-mass spectrometric determination of 4- and 6-hydroxylated bile acids in human urine with negative ion chemical ionization detection. *J Chromatogr B Biomed Sci App* 574:1–7.
12. Li J, Dawson PA. 2019. Animal models to study bile acid metabolism. *Biochim Biophys Acta - Mol Basis Dis* 1865:895–911.

13. Quinn RA, Melnik AV, Vrbanac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, Downes M, Welch RD, Quinn M, Humphrey G, Panitchpakdi M, Weldon KC, Aksenov A, da Silva R, Avila-Pacheco J, Clish C, Bae S, Mallick H, Franzosa EA, Lloyd-Price J, Bussell R, Thron T, Nelson AT, Wang M, Leszczynski E, Vargas F, Gauglitz JM, Meehan MJ, Gentry E, Arthur TD, Komor AC, Poulsen O, Boland BS, Chang JT, Sandborn WJ, Lim M, Garg N, Lumeng JC, Xavier RJ, Kazmierczak BI, Jain R, Egan M, Rhee KE, Ferguson D, Raffatellu M, Vlamakis H, Haddad GG, Siegel D, Huttenhower C, Mazmanian SK, Evans RM, Nizet V, Knight R, Dorrestein PC. 2020. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* 579:123–129.
14. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, Daniélsdóttir AD, Krecke M, Merten D, Haraldsdóttir HS, Heinken A, Heirendt L, Magnúsdóttir S, Ravcheev DA, Sahoo S, Gawron P, Friscioni L, Garcia B, Prendergast M, Puente A, Rodrigues M, Roy A, Rouquaya M, Wiltgen L, Žagare A, John E, Krueger M, Kuperstein I, Zinovyev A, Schneider R, Fleming RMT, Thiele I. 2019. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res* 47:D614–D624.
15. Heinken A, Ravcheev DA, Baldini F, Heirendt L, Fleming RMT, Thiele I. 2019. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. *Microbiome* 7:75.
16. Hofmann AF. 2009. The enterohepatic circulation of bile acids in mammals: form and functions. *Front Biosci* 14:2584–2598.
17. Dawson PA, Karpen SJ. 2015. Intestinal transport and metabolism of bile acids. *J Lipid Res* 56:1085–1099.
18. Ridlon JM, Kang DJ, Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* 47:241–259.
19. Aldini R, Roda A, Lenzi PL, Ussia G, Vaccari MC, Mazzella G, Festi D, Bazzoli F, Galletti G, Casanova S, Montagnani M, Roda E. 1992. Bile acid active and passive ileal transport in the rabbit: effect of luminal stirring. *Eur J Clin Invest* 22:744–750.
20. de Aguiar Vallim TQ, Tarling EJ, Edwards PA. 2013. Pleiotropic roles of bile acids in metabolism. *Cell Metab* 17:657–669.
21. Shin DJ, Wang L. 2019. Bile acid-activated receptors: a review on FXR and other nuclear receptors, p. 51–72. *In* *Handb. Exp. Pharmacol.* Springer New York LLC.
22. Inagaki T, Moschetta A, Lee YK, Peng L, Zhao G, Downes M, Yu RT, Shelton JM, Richardson JA, Repa JJ, Mangelsdorf DJ, Kliewer SA. 2006. Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proc Natl Acad Sci U S A* 103:3920–3925.

23. Hofmann AF, Eckmann L. 2006. How bile acids confer gut mucosal protection against bacteria. *Proc Natl Acad Sci U S A* 103:4333–4334.
24. Sinal CJ, Tohkin M, Miyata M, Ward JM, Lambert G, Gonzalez FJ. 2000. Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* 102:731–744.
25. Tremblay S, Romain G, Roux M, Chen X-L, Brown K, Gibson DL, Ramanathan S, Menendez A. 2017. Bile acid administration elicits an intestinal antimicrobial program and reduces the bacterial burden in two mouse models of enteric infection. *Infect Immun* 85:e00942-16.
26. Sayin SI, Wahlström A, Felin J, Jäntti S, Marschall HU, Bamberg K, Angelin B, Hyötyläinen T, Orešič M, Bäckhed F. 2013. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab* 17:225–235.
27. Keitel V, Stindt J, Häussinger D. 2019. Bile Acid-Activated Receptors: GPBAR1 (TGR5) and Other G Protein-Coupled Receptors, p. 19–49. *In Handbook of Experimental Pharmacology*. Springer New York LLC.
28. Sannasiddappa TH, Lund PA, Clarke SR. 2017. *In vitro* antibacterial activity of unconjugated and conjugated bile salts on *Staphylococcus aureus*. *Front Microbiol* 8:1581.
29. Stacey M, Webb M. 1947. Studies on the antibacterial properties of the bile acids and some compounds derived from cholanic acid. *Proc R Soc Med* 134:523–537.
30. Kurdi P, Kawanishi K, Mizutani K, Yokota A. 2006. Mechanism of growth inhibition by free bile acids in Lactobacilli and Bifidobacteria. *J Bacteriol* 188:1979–1986.
31. Urdaneta V, Casadesús J. 2017. Interactions between bacteria and bile salts in the gastrointestinal and hepatobiliary tracts. *Front Med* 4:163.
32. Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB. 2016. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* 7:22–39.
33. Gustafsson BE, Midtvedt T, Norman A. 1966. Isolated fecal microorganisms capable of 7-alpha-dehydroxylating bile acids. *J Exp Med* 123:413–432.
34. Jones BV, Begley M, Hill C, Gahan CGM, Marchesi JR. 2008. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A* 105:13580–13585.
35. Joyce SA, Gahan CGM. 2017. Disease-associated changes in bile acid profiles and links to altered gut microbiota. *Dig Dis* 35:169–177.

36. Doden H, Sallam LA, Devendran S, Ly L, Doden G, Daniel SL, Alves JMP, Ridlon JM. 2018. Metabolism of oxo-bile acids and characterization of recombinant 12 $\alpha$ -hydroxysteroid dehydrogenases from bile acid 7 $\alpha$ -dehydroxylating human gut bacteria. *Appl Environ Microbiol* 84:235–253.
37. Ovadia C, Perdones-Montero A, Spagou K, Smith A, Sarafian MH, Gomez-Romero M, Bellafante E, Clarke LCD, Sadiq F, Nikolova V, Mitchell A, Dixon PH, Santa-Pinter N, Wahlström A, Abu-Hayyeh S, Walters JRF, Marschall HU, Holmes E, Marchesi JR, Williamson C. 2019. Enhanced microbial bile acid deconjugation and impaired ileal uptake in pregnancy repress intestinal regulation of bile acid synthesis. *Hepatology* 70:276–293.
38. Kim GB, Yi SH, Lee BH. 2004. Purification and characterization of three different types of bile salt hydrolases from *Bifidobacterium* strains. *J Dairy Sci* 87:258–266.
39. Elkins CA, Moser SA, Savage DC. 2001. Genes encoding bile salt hydrolases and conjugated bile salt transporters in *Lactobacillus johnsonii* 100-100 and other *Lactobacillus* species. *Microbiology* 147:3403–3412.
40. Corzo G, Gilliland SE. 1999. Bile salt hydrolase activity of three strains of *Lactobacillus acidophilus*. *J Dairy Sci* 82:472–480.
41. Coleman JP, Hudson LL. 1995. Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens*. *Appl Environ Microbiol* 61:2514–2520.
42. Wijaya A, Hermann A, Abriouel H, Specht I, Yousif NMK, Holzapfel WH, Franz CMAP. 2004. Cloning of the bile salt hydrolase (*bsh*) gene from *Enterococcus faecium* FAIR-E 345 and chromosomal location of *bsh* genes in food Enterococci. *J Food Prot* 67:2772–2778.
43. Dussurget O, Cabanes D, Dehoux P, Lecuit M, Buchrieser C, Glaser P, Cossart P. 2002. *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol Microbiol* 45:1095–1106.
44. Dean M, Cervellati C, Casanova E, Squerzanti M, Lanzara V, Medici A, De Laureto PP, Bergamini CM. 2002. Characterization of cholyglycine hydrolase from a bile-adapted strain of *Xanthomonas maltophilia* and its application for quantitative hydrolysis of conjugated bile salts. *Appl Environ Microbiol* 68:3126–3128.
45. Kawamoto K, Horibe I, Uchida K. 1989. Purification and characterization of a new hydrolase for conjugated bile acids, chenodeoxycholytaurine hydrolase, from *Bacteroides vulgatus*. *J Biochem (Tokyo)* 106:1049–1053.
46. Delpino MV, Marchesini MI, Estein SM, Comerchi DJ, Cassataro J, Fossati CA, Baldi PC. 2007. A bile salt hydrolase of *Brucella abortus* contributes to the establishment of a successful infection through the oral route in mice. *Infect Immun* 75:299–305.

47. Song Z, Cai Y, Lao X, Wang X, Lin X, Cui Y, Kalavagunta PK, Liao J, Jin L, Shang J, Li J. 2019. Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* 7:9.
48. Percy-Robb IW. 1972. Bile acids: a pH dependent antibacterial system in the gut? *Br Med J* 3:813–815.
49. Stellwag EJ, Hylemon PB. 1976. Purification and characterization of bile salt hydrolase from *Bacteroides fragilis* subsp. *fragilis*. *BBA - Enzymol* 452:165–176.
50. Xu F, Guo F, Hu XJ, Lin J. 2016. Crystal structure of bile salt hydrolase from *Lactobacillus salivarius*. *Acta Crystallogr Sect F Struct Biol Commun* 72:376–381.
51. Hu X-J. 2016. Bile salt hydrolase from *Lactobacillus salivarius* (1.1). 5hke. RCSB PDB.
52. Kumar RS, Brannigan JA, Prabhune AA, Pundle AV, Dodson GG, Dodson EJ, Suresh CG. 2006. Structural and functional analysis of a conjugated bile salt hydrolase from *Bifidobacterium longum* reveals an evolutionary relationship with penicillin V acylase. *J Biol Chem* 281:32516–32525.
53. Suresh CG, Kumar RS, Brannigan JA. 2011. *Bifidobacterium longum* bile salt hydrolase (1.2). 2hf0. pdb. RCSB PDB.
54. Seegar TCM. 2020. *B. theta* bile salt hydrolase (1.2). 6ufy. pdb. RCSB PDB.
55. Adhikari AA, Seegar TCM, Ficarro SB, McCurry MD, Ramachandran D, Yao L, Chaudhari SN, Ndousse-Fetter S, Banks AS, Marto JA, Blacklow SC, Devlin AS. 2020. Development of a covalent inhibitor of gut bacterial bile salt hydrolases. *Nat Chem Biol* 16:318–326.
56. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2005. Conjugated bile acid hydrolase is a tetrameric N-terminal thiol hydrolase with specific recognition of its cholyl but not of its tauryl product. *Biochemistry* 44:5739–5748.
57. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2011. Crystal structure of conjugated bile acid hydrolase from *Clostridium perfringens* in complex with reaction products taurine and deoxycholate (1.2). 2bjf. pdb. RCSB PDB.
58. Ramasamy S, Chand D, Suresh C. 2016. Crystal structure determination of Bile Salt Hydrolase from *Enterococcus faecalis* (1.1). 4wl3. pdb. RCSB PDB.
59. Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual Molecular Dynamics. *J Mol Graph* [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).

60. Ridlon JM, Hylemon PB. 2012. Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens*, a bile acid 7 $\alpha$ -dehydroxylating intestinal bacterium. *J Lipid Res* 53:66–76.
61. Mallonee DH, Hylemon PB. 1996. Sequencing and expression of a gene encoding a bile acid transporter from *Eubacterium* sp. strain VPI 12708. *J Bacteriol* 178:7053–7058.
62. Bhowmik S, Chiu H-P, Jones DH, Chiu H-J, Miller MD, Xu Q, Farr CL, Ridlon JM, Wells JE, Elsliger M-A, Wilson IA, Hylemon PB, Lesley SA. 2016. Structure and functional characterization of a bile acid 7 $\alpha$  dehydratase BaiE in secondary bile acid synthesis. *Proteins Struct Funct Bioinforma* 84:316–331.
63. Joint Center for Structural Genomics. 2017. Crystal structure of a bile acid 7 $\alpha$ -dehydratase (CLOSCI\_03134) from *Clostridium scindens* ATCC 35704 at 2.90 Å resolution (1.1). 4l8h. pdb. RCSB PDB.
64. Joint Center for Structural Genomics. 2017. Crystal structure of a bile acid 7 $\alpha$ -dehydratase (CLOHYLEM\_06634) from *Clostridium hylemonae* DSM 15053 at 2.20 Å resolution (1.1). 4l8o. pdb. RCSB PDB.
65. Joint Center for Structural Genomics. 2017. Crystal structure of a bile acid 7 $\alpha$ -dehydratase (CLOHIR\_00079) from *Clostridium hiranonis* DSM 13275 at 1.60 Å resolution (1.1). 4l8p. pdb. RCSB PDB.
66. Harris SC, Devendran S, Méndez- García C, Mythen SM, Wright CL, Fields CJ, Hernandez AG, Cann I, Hylemon PB, Ridlon JM. 2018. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243 T. *Gut Microbes* 9:523–539.
67. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. 2020. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582:566–570.
68. Lee JW, Cowley ES, Wolf PG, Doden HL, Murai T, Caicedo KYO, Ly LK, Sun F, Takei H, Nittono H, Daniel SL, Cann I, Gaskins HR, Anantharaman K, Alves JMP, Ridlon JM. 2022. Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes. *Gut Microbes* 14:2132903.
69. Kim KH, Park D, Jia B, Baek JH, Hahn Y, Jeon CO. 2022. Identification and characterization of major bile acid 7 $\alpha$ -dehydroxylating bacteria in the human gut. *mSystems* 7:e00455-22.
70. Bhowmik S, Jones DH, Chiu HP, Park IH, Chiu HJ, Axelrod HL, Farr CL, Tien HJ, Agarwalla S, Lesley SA. 2014. Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. *Proteins Struct Funct Bioinforma* 82:216–229.

71. Kang DJ, Ridlon JM, Moore DR, Barnes S, Hylemon PB. 2008. *Clostridium scindens* *baiCD* and *baiH* genes encode stereo-specific 7 $\alpha$ /7 $\beta$ -hydroxy-3-oxo- $\Delta^4$ -cholenoic acid oxidoreductases. *Biochim Biophys Acta - Mol Cell Biol Lipids* 1781:16–25.
72. Hirano S, Masuda N. 1981. Epimerization of the 7-hydroxy group of bile acids by the combination of two kinds of microorganisms with 7  $\alpha$ - and 7  $\beta$ -hydroxysteroid dehydrogenase activity, respectively. *J Lipid Res* 22:1060–1068.
73. Wang S, Martins R, Sullivan MC, Friedman ES, Misic AM, El-Fahmawi A, De Martinis ECP, O'Brien K, Chen Y, Bradley C, Zhang G, Berry ASF, Hunter CA, Baldassano RN, Rondeau MP, Beiting DP. 2019. Diet-induced remission in chronic enteropathy is associated with altered microbial community structure and synthesis of secondary bile acids. *Microbiome* 7:1–20.
74. Eggert T, Bakonyi D, Hummel W. 2014. Enzymatic routes for the synthesis of ursodeoxycholic acid. *J Biotechnol* 191:11–21.
75. Giovannini PP, Grandini A, Perrone D, Pedrini P, Fantin G, Fogagnolo M. 2008. 7 $\alpha$ - and 12 $\alpha$ -Hydroxysteroid dehydrogenases from *Acinetobacter calcoaceticus* *lwoffii*: a new integrated chemo-enzymatic route to ursodeoxycholic acid. *Steroids* 73:1385–1390.
76. Paik D, Yao L, Zhang Y, Bae S, D'Agostino GD, Zhang M, Kim E, Franzosa EA, Avila-Pacheco J, Bisanz JE, Rakowski CK, Vlamakis H, Xavier RJ, Turnbaugh PJ, Longman RS, Krout MR, Clish CB, Rastinejad F, Huttenhower C, Huh JR, Devlin AS. 2022. Human gut bacteria produce T<sub>H</sub>17-modulating bile acid metabolites. *Nature* 603:907–912.
77. Mythen SM, Devendran S, Méndez-García C, Cann I, Ridlon JM. 2018. Targeted synthesis and characterization of a gene cluster encoding NAD(P)H-dependent 3 $\alpha$ -, 3 $\beta$ -, and 12 $\alpha$ -hydroxysteroid dehydrogenases from *Eggerthella* CAG:298, a gut metagenomic sequence. *Appl Environ Microbiol* 84.
78. Lepercq P, Gérard P, Béguet F, Raibaud P, Grill J-P, Relano P, Cayuela C, Juste C. 2004. Epimerization of chenodeoxycholic acid to ursodeoxycholic acid by *Clostridium baratii* isolated from human feces. *FEMS Microbiol Lett* 235:65–72.
79. Edenharder R, Knaflic T. 1981. Epimerization of chenodeoxycholic acid to ursodeoxycholic acid by human intestinal lecithinase-lipase-negative Clostridia. *J Lipid Res* 22:652–658.
80. Lee JY, Arai H, Nakamura Y, Fukiya S, Wada M, Yokota A. 2013. Contribution of the 7 $\beta$ -hydroxysteroid dehydrogenase from *Ruminococcus gnavus* N53 to ursodeoxycholic acid formation in the human colon. *J Lipid Res* 54:3062–3069.
81. Ferrandi EE, Bertolesi GM, Polentini F, Negri A, Riva S, Monti D. 2012. In search of sustainable chemical processes: Cloning, recombinant expression, and functional

- characterization of the 7 $\alpha$ - and 7 $\beta$ -hydroxysteroid dehydrogenases from *Clostridium absonum*. *Appl Microbiol Biotechnol* 95:1221–1233.
82. Liu L, Aigner A, Schmid RD. 2011. Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7 $\beta$ -hydroxysteroid dehydrogenase from *Collinsella aerofaciens*. *Appl Microbiol Biotechnol* 90:127–135.
  83. MacDonald IA, Mahony DE, Jellett JF, Meier CE. 1977. NAD-dependent 3 $\alpha$ - and 12 $\alpha$ -hydroxysteroid dehydrogenase activities from *Eubacterium lentum* ATCC no. 25559. *Biochim Biophys Acta BBALipids Lipid Metab* 489:466–476.
  84. MacDonald IA, Jellett JF, Mahony DE, Holdeman LV. 1979. Bile salt 3 $\alpha$ - and 12 $\alpha$ -hydroxysteroid dehydrogenases from *Eubacterium lentum* and related organisms. *Appl Environ Microbiol* 37:992–1000.
  85. Pedrini P, Andreotti E, Guerrini A, Dean M, Fantin G, Giovannini PP. 2006. *Xanthomonas maltophilia* CBS 897.97 as a source of new 7 $\beta$ - and 7 $\alpha$ -hydroxysteroid dehydrogenases and cholyglycine hydrolase: Improved biotransformations of bile acids. *Steroids* 71:189–198.
  86. Wegner K, Just S, Gau L, Mueller H, Gérard P, Lepage P, Clavel T, Rohn S. 2017. Rapid analysis of bile acids in different biological matrices using LC-ESI-MS/MS for the investigation of bile acid transformation by mammalian gut bacteria. *Anal Bioanal Chem* 409:1231–1245.
  87. Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk HP, Goodfellow M, Göker M. 2018. Genome-based taxonomic classification of the phylum Actinobacteria. *Front Microbiol* 9:2007.
  88. Edenharder R, Schneider J. 1985. 12 $\beta$ -dehydrogenation of bile acids by *Clostridium paraputrificum*, *C. tertium*, and *C. difficile* and epimerization at carbon-12 of deoxycholic acid by cocultivation with 12 $\alpha$ -dehydrogenating *Eubacterium lentum*. *Appl Environ Microbiol* 49:964–968.
  89. Edenharder R, Pfützner A. 1988. Characterization of NADP-dependent 12 $\beta$ -hydroxysteroid dehydrogenase from *Clostridium paraputrificum*. *Biochim Biophys Acta BBALipids Lipid Metab* 962:362–370.
  90. Doden HL, Wolf PG, Gaskins HR, Anantharaman K, Alves JMP, Ridlon JM. 2021. Completion of the gut microbial epi-bile acid pathway. *Gut Microbes* 13:1–20.
  91. Hylemon PB, Melone PD, Franklund CV, Lund E, Björkhem I. 1991. Mechanism of intestinal 7 $\alpha$ -dehydroxylation of cholic acid: evidence that allo-deoxycholic acid is an inducible side-product. *J Lipid Res* 32:89–96.
  92. Sato Y, Atarashi K, Plichta DR, Arai Y, Sasajima S, Kearney SM, Suda W, Takeshita K, Sasaki T, Okamoto S, Skelly AN, Okamura Y, Vlamakis H, Li Y, Tanoue T, Takei H, Nittono H, Narushima S, Irie J, Itoh H, Moriya K, Sugiura Y, Suematsu M, Moritoki

- N, Shibata S, Littman DR, Fischbach MA, Uwamino Y, Inoue T, Honda A, Hattori M, Murai T, Xavier RJ, Hirose N, Honda K. 2021. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* 599:458–464.
93. Pikuta EV, Hoover RB, Marsic D, Whitman WB, Lupa B, Tang J, Krader P. 2009. *Proteocatella sphenisci* gen. nov., sp. nov., a psychrotolerant, spore-forming anaerobe isolated from penguin guano. *Int J Syst Evol Microbiol* 59:2302–2307.
  94. Félix AP, Souza CMM, De Oliveira SG. 2022. Biomarkers of gastrointestinal functionality in dogs: a systematic review and meta-analysis. *Anim Feed Sci Technol* 283:115183.
  95. Sfakianos MK, Wilson L, Sakalian M, Falany CN, Barnes S. 2002. Conserved residues in the putative catalytic triad of human bile acid coenzyme A:amino acid N-acyltransferase. *J Biol Chem* 277:47270–47275.
  96. van de Waterbeemd H, Karajiannis H, El Tayar N. 1994. Lipophilicity of amino acids. *Amino Acids* 7:129–145.
  97. Ambrogelly A, Palioura S, Söll D. 2007. Natural expansion of the genetic code. *Nat Chem Biol* 3:29–35.
  98. Tamari M, Ogawa M, Kametaka M. 1976. A new bile acid conjugate, ciliatocholic acid, from bovine gall bladder bile. *J Biochem (Tokyo)* 80:371–377.
  99. Chiang JYL. 2009. Bile acids: regulation of synthesis. *J Lipid Res* 50:1955–1966.
  100. Chiang JYL. 2017. Recent advances in understanding bile acid homeostasis. *F1000Research* 6:2029.
  101. Roda A, Minutello A, Angellotti MA, Fini A. 1990. Bile acid structure-activity relationship: Evaluation of bile acid lipophilicity using 1-octanol/water partition coefficient and reverse phase HPLC. *J Lipid Res* 31:1433–1443.
  102. Mullish BH, McDonald JAK, Pechlivanis A, Allegretti JR, Kao D, Barker GF, Kapila D, Petrof EO, Joyce SA, Gahan CGM, Glegola-Madejska I, Williams HRT, Holmes E, Clarke TB, Thursz MR, Marchesi JR. 2019. Microbial bile salt hydrolases mediate the efficacy of fecal microbiota transplant in the treatment of recurrent *Clostridioides difficile* infection. *Gut* 68:1791–1800.
  103. Pickard JM, Zeng MY, Caruso R, Núñez G. 2017. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev* 279:70–89.
  104. Ward JBJ, Lajczak NK, Kelly OB, O'Dwyer AM, Giddam AK, Ní Gabhann J, Franco P, Tambuwala MM, Jefferies CA, Keely S, Roda A, Keely SJ. 2017. Ursodeoxycholic acid and lithocholic acid exert anti-inflammatory actions in the colon. *Am J Physiol - Gastrointest Liver Physiol* 312:G550–G558.

105. Bernstein H, Bernstein C, Payne CM, Dvorakova K, Garewal H. 2005. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res Mutat Res* 589:47–65.
106. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. 2011. Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol* 85:863–871.
107. Castro RE, Amaral JD, Solá S, Kren BT, Steer CJ, Rodrigues CMP. 2007. Differential regulation of cyclin D1 and cell death by bile acids in primary rat hepatocytes. *Am J Physiol Gastrointest Liver Physiol* 293:G327–334.
108. Sousa T, Castro RE, Pinto SN, Coutinho A, Lucas SD, Moreira R, Rodrigues CMP, Prieto M, Fernandes F. 2015. Deoxycholic acid modulates cell death signaling through changes in mitochondrial membrane properties. *J Lipid Res* 56:2158–2171.
109. Gumpricht E, Devereaux MW, Dahl RH, Sokol RJ. 2000. Glutathione status of isolated rat hepatocytes affects bile acid-induced cellular necrosis but not apoptosis. *Toxicol Appl Pharmacol* 164:102–111.
110. Goossens JF, Bailly C. 2019. Ursodeoxycholic acid and cancer: from chemoprevention to chemotherapy. *Pharmacol Ther* 203:107396.
111. Eaton JE, Silveira MG, Pardi DS, Sinakos E, Kowdley KV, Luketic VAC, Harrison ME, McCashland T, Befeler AS, Harnois D, Jorgensen R, Petz J, Lindor KD. 2011. High-dose ursodeoxycholic acid is associated with the development of colorectal neoplasia in patients with ulcerative colitis and primary sclerosing cholangitis. *Am J Gastroenterol* 106:1638–1645.
112. Kimura A, Mahara R, Inoue T, Nomura Y, Murai T, Kurosawa T, Tohma M, Noguchi K, Hoshiyama A, Fujisawa T, Kato H. 1999. Profile of urinary bile acids in infants and children: developmental pattern of excretion of unsaturated ketonic bile acids and 7 $\beta$ -hydroxylated bile acids. *Pediatr Res* 45:603–609.
113. Kimura A, Suzuki M, Murai T, Inoue T, Kato H, Hori D, Nomura Y, Yoshimura T, Kurosawa T, Tohma M. 1997. Perinatal bile acid metabolism: analysis of urinary bile acids in pregnant women and newborns. *J Lipid Res* 38:1954–1962.
114. Suzuki M, Murai T, Yoshimura T, Kimura A, Kurosawa T, Tohma M. 1997. Determination of 3-oxo- $\Delta^4$ - and 3-oxo- $\Delta^{4,6}$ -bile acids and related compounds in biological fluids of infants with cholestasis by gas chromatography-mass spectrometry. *J Chromatogr B Biomed Sci App* 693:11–21.
115. El-Mir MY, Badia MD, Luengo N, Monte MJ, Marin JJ. 2001. Increased levels of typically fetal bile acid species in patients with hepatocellular carcinoma. *Clin Sci Lond Engl* 1979 100:499–508.
116. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. *Nature* 449:804–810.

117. Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD. 2018. Diagnostic potential and interactive dynamics of the colorectal cancer virome. *mBio* 9:e02248-18.
118. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, Hercog R, Koch M, Luciani A, Mende DR, Schneider MA, Schrotz-King P, Tournigand C, Tran Van Nhieu J, Yamada T, Zimmermann J, Benes V, Kloor M, Ulrich CM, Von Knebel Doeberitz M, Sobhani I, Bork P. 2014. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 10:766.
119. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, Su L, Li X, Li X, Li J, Xiao L, Huber-Schönauer U, Niederseer D, Xu X, Al-Aama JY, Yang H, Wang J, Kristiansen K, Arumugam M, Tilg H, Datz C, Wang J. 2015. Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat Commun* 6:6528.
120. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, Wang X, Xu X, Chen N, Wu WKK, Al-Aama J, Nielsen HJ, Kiilerich P, Jensen BAH, Yau TO, Lan Z, Jia H, Li J, Xiao L, Lam TYT, Ng SC, Cheng AS-L, Wong VW-S, Chan FKL, Xu X, Yang H, Madsen L, Datz C, Tilg H, Wang J, Brünner N, Kristiansen K, Arumugam M, Sung JJ-Y, Wang J. 2017. Metagenomic analysis of fecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66:70–78.
121. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R. 2016. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLOS ONE* 11:e0155362.
122. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vázquez-Baeza Y, White RA, Bishai J, Bullock K, Deik A, Dennis C, Kaplan JL, Khalili H, McIver LJ, Moran CJ, Nguyen L, Pierce KA, Schwager R, Sirota-Madi A, Stevens BW, Tan W, ten Hoeve JJ, Weingart G, Wilson RG, Yajnik V, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569:655–662.

**CHAPTER 2:**  
**IDENTIFICATION AND CHARACTERIZATION OF ACYLTRANSFERASE ACTIVITY**  
**BY THE ENZYME BILE SALT HYDROLASE**

## 2.1 - Preface

Portions of this chapter were published in the journal *Nature* in 2024 (Material from: Guzior, D.V., Okros, M., Shivel, M. *et al.* Bile salt hydrolase acyltransferase activity expands bile acid diversity. *Nature* **626**, 852–858 (2024). <https://doi.org/10.1038/s41586-024-07017-8>). Per the publisher, Springer Nature, “Authors have the right to reuse their article’s Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.”

## 2.2 - Abstract

Bacterial metabolism of host bile acids (BA) has long been implicated in the development and propagation of gastrointestinal disease. Given the correlations between microbially conjugated bile acid (MCBA) production profiles and bile salt hydrolase amino acid sequences, I sought to determine if acyl transfer by this enzyme would be observed *in vitro*, making it a bile salt hydrolase/transferase, BSH/T. Incubating *Clostridium perfringens* BSH/T (CpBSH/T) with taurocholic acid (TCA) and an equimolar amino acid mix resulted in robust MCBA production. Under these conditions, 16 of 20 amino acids were transferred to cholic acid (CA) at an optimum pH of 5.3. When provided glycocholic acid (GCA), 11 of 19 amino acids were transferred (when excluding glycine). Surprisingly, 12 of 20 amino acids were transferred to free CA, although total MCBA abundance was lower compared to GCA and TCA. Proline and aspartate use was not observed regardless of the base BA provided. Formation of phenylalanochoic acid (PheCA) from phenylalanine transfer to TCA showed linear kinetics, suggesting that phenylalanine competes with water for nucleophilic activity against an enzyme-BA bound intermediate. This intermediate was validated by converting the catalytic cysteine at residue 2 to an alanine, thus inactivating the activity. Furthermore, I validated the importance of asparagine at position 82 in shaping active site structure, subsequently impacting MCBA profile when replaced by tyrosine (N82Y). *E. coli* expressing N82Y variants of BSH/T demonstrated decreased MCBA production, with an enrichment of small amino acid ligands. This work is the first to characterize acyl transfer by BSH/T in addition to validating important residues within the active site that contribute to MCBA production.

### 2.3 - Introduction

Scientific dogma has been that catalysis of bile salt deconjugation by the enzyme bile salt hydrolase (BSH) is the “gateway reaction” to further bile acid modification by bacteria in our intestines. This description was assigned originally due to the finding that *Lachnoclostridium scindens* (formerly *Clostridium scindens*) was unable to perform 7 $\alpha$ -dehydroxylation on glycine- and taurine-bound BAs, suggesting the C24 carboxyl is required (1). Decades of research were required to fully understand the enzymatic pathway required for the production of the 7 $\alpha$ -dehydroxylated BAs deoxycholic acid (DCA) and lithocholic acid (LCA). *L. scindens* was first described to perform 7 $\alpha$ -dehydroxylation in 1980, yet the full metabolic pathway was only recently elucidated by Funabashi and colleagues in 2020 (2, 3). This discovery was of critical importance, as both DCA and LCA exhibit potent antimicrobial activity, DNA damaging activity, and membrane disrupting effects (4). Understanding the biochemical changes resulting from microbial metabolism is key to describing relevant, physiological roles and impacts.

Modification at the C24 carboxyl group, specifically involving conjugation, is known to have drastic impacts on BA biochemistry. As BAs are dehydroxylated, either by the host or resident microbiota, they become more hydrophobic. This increase in hydrophobicity allows for increased intercalation into lipid membranes (5, 6). However, conjugation with taurine or glycine in the liver aids in their storage prior to meals; conjugated primary BAs are more hydrophilic and can thus be transported via active transport at the ileum and by the liver (7–12). However, taurine and glycine are small and polar, drastically altering chemical properties of these otherwise hydrophobic, amphipathic compounds. Conjugation with biochemically diverse ligands, ranging from larger or less polar amino

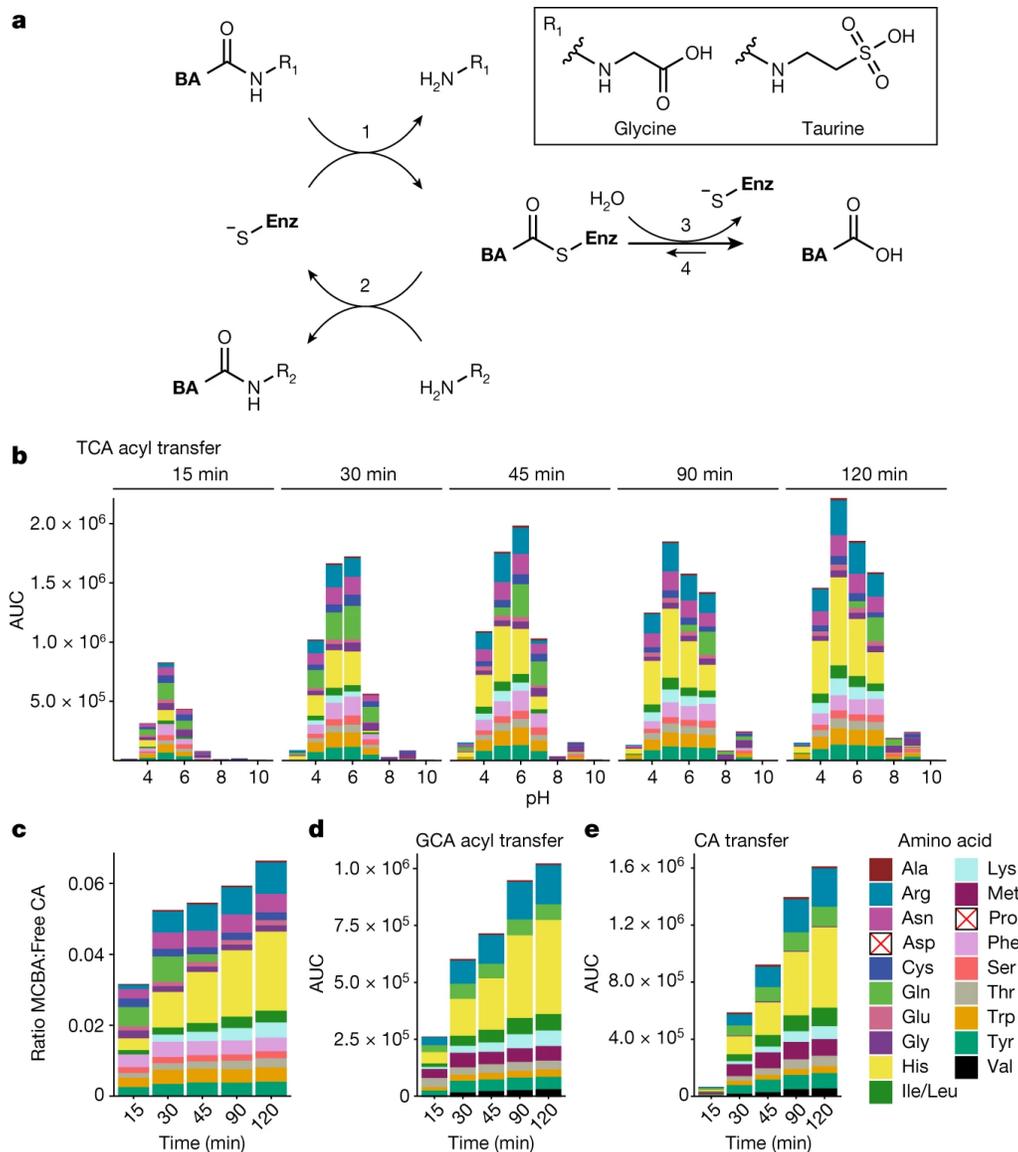
acids undoubtedly changes essential biochemistry of these compounds. Within this, properties of the ligand itself may impact selectivity for use in conjugation.

In this work, I first identify the ability of purified *C. perfringens* BSH to produce MCBA. I then characterize optimal conditions for acyl transfer of amino acids to BAs, both those that are specifically conjugated and to free CA. Following this analysis, I investigate kinetic characteristics of phenylalanine transfer to TCA and CA. Finally, I investigated how residues within the active site impact the capacity and substrate diversity for BA conjugation after aligning BSH/T amino acid sequence from bacterial strains and subsequently screening them for MCBA production.

## **2.4 - Results**

### *2.4.1 - BSH/T acyl transfer characterization*

The first BSH/T to be purified and have its hydrolase activity characterized was from *C. perfringens* (13). Due to its established interaction with conjugated BAs, also known as bile salts, we investigated the capacity of CpBSH/T to exchange the conjugated amino acid. Enzyme-catalyzed hydrolysis of bile salts occurs via a covalently bound cysteinyl intermediate (Figure 2.1a) (14) and CpBSH/T was active for hydrolysis over a broad pH range (pH 3-7, Table 2.3). When incubated with TCA and an equimolar mix of 20 essential amino acids, CpBSH/T rapidly hydrolyzed TCA to cholic acid (CA, Table 2.3), as expected, in addition to catalyzing acyl-conjugation of CA with a variety of amino acids (Figure 2.1b). Indeed, 16 of 20 amino acids became linked to CA, with aspartocholic acid (AspCA), methionocholic acid (MetCA), prolocholic acid (ProCA), and valocholeic acid (ValCA) not being produced under these conditions (Figure 2.1b, Table 2.1). CpBSH/T may catalyze acyl transfer through the reaction of amino acids with a covalently bound



**Figure 2.1: *C. perfringens* BSH/T produces a broad range of MCBAs at acidic pH**

**a**, Chemical reaction steps catalyzed by BSH/T. The enzyme is capable of (1) reacting with conjugated primary BAs through nucleophile attack using Cys2 to form a covalently bound enzyme-CA intermediate followed by (2) hydrolytic release of the BA or (3) reaction with other amino acids by an acyl transfer reaction, resulting in formation of MCBAs. In addition, MCBAs can be generated by (4) direct formation of the enzyme-CA intermediate from CA with subsequent acyl transfer. Enz, enzyme. **b**, Stacked area-under-the-curve (AUC) profiles of MCBAs products following CpBSH/T incubation with TCA and an equimolar amino acid mixture over a broad pH range (3.0–10.0), across time. **c**, Ratio of mean summed MCBAs abundance to CA abundance, derived from acyltransferase activity and hydrolase activity of CpBSH/T incubated with TCA and an equimolar amino acid mixture at pH 5.0. **d,e**, MCBAs profiles at pH 5.0 following CpBSH/T incubation with an equimolar amino acid mixture and 2.5 mM GCA (**d**) or 2.5 mM CA (**e**). Prolocholcic acid (ProCA) and aspartocholcic acid (AspCA) were not present in any samples.  $n = 3$  independent reactions per pH, per BA.

**Table 2.1: Abundance of amino acids used in acyl transfer when provided different BA substrates**

Data are presented as mean AUC (s.e.m.).  $n = 3$  independent reactions.

Amino Acid	TCA	GCA	CA
Ala	1.88E+04 (1.22E+02)	7.59E+03 (3.77E+02)	1.54E+04 (5.90E+02)
Arg	2.95E+05 (7.18E+03)	1.72E+05 (3.18E+03)	2.68E+05 (9.36E+03)
Asn	1.75E+05 (5.76E+03)	0.00E+00 (0.00E+00)	0.00E+00 (0.00E+00)
Cys	7.22E+04 (9.33E+01)	0.00E+00 (0.00E+00)	0.00E+00 (0.00E+00)
Gln	0.00E+00 (0.00E+00)	6.90E+04 (3.22E+03)	1.38E+05 (1.44E+03)
Glu	5.04E+04 (1.59E+02)	0.00E+00 (0.00E+00)	0.00E+00 (0.00E+00)
Gly	5.81E+04 (2.70E+03)	N/A	6.90E+03 (2.74E+03)
His	7.42E+05 (3.08E+04)	4.14E+05 (5.33E+03)	5.63E+05 (1.88E+04)
Ile/Leu	1.12E+05 (5.52E+03)	7.17E+04 (1.61E+03)	1.29E+05 (1.04E+03)
Lys	1.41E+05 (4.59E+03)	6.85E+04 (1.82E+03)	9.13E+04 (3.59E+03)
Met	0.00E+00 (0.00E+00)	6.40E+04 (2.59E+03)	1.16E+05 (2.76E+02)
Phe	1.29E+05 (4.83E+03)	0.00E+00 (0.00E+00)	0.00E+00 (0.00E+00)
Ser	6.60E+04 (1.08E+03)	0.00E+00 (0.00E+00)	0.00E+00 (0.00E+00)
Thr	8.54E+04 (2.66E+03)	3.95E+04 (1.42E+03)	7.32E+04 (3.22E+03)
Trp	1.36E+05 (4.85E+03)	3.20E+04 (1.27E+03)	4.84E+04 (6.61E+03)
Tyr	1.34E+05 (4.02E+03)	5.44E+04 (9.38E+02)	1.07E+05 (2.01E+03)
Val	0.00E+00 (0.00E+00)	3.33E+04 (3.73E+02)	5.48E+04 (1.41E+03)

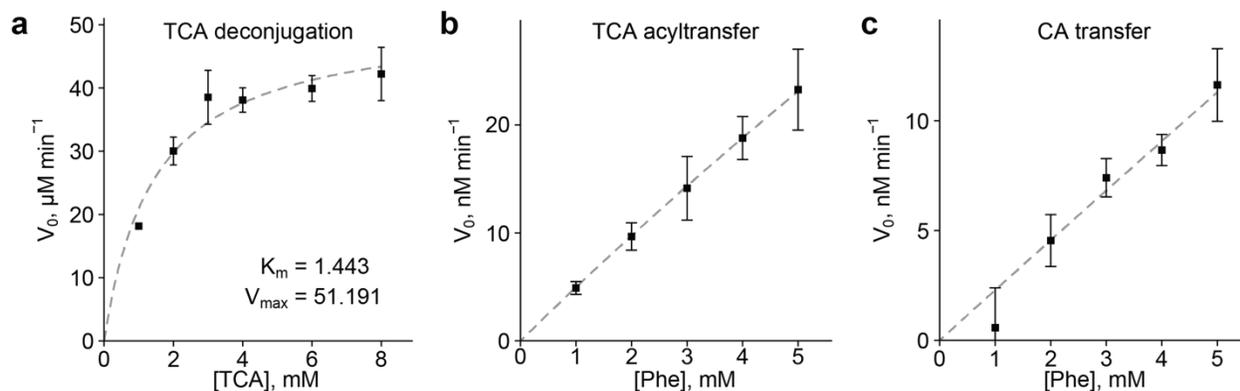
intermediate, where an amino acid acts as a nucleophile *in lieu* of water (Figure 2.1a, steps 1 and 3). I observed acyltransferase activity across a broad pH range (Figure 2.1b), with an optimum at pH 5.3 (, Table 2.1) based on the summed abundance of MCBAs following 120-min incubation at 37 °C. This value is slightly higher than the previously reported pH 4.5-4.9 optimum for TCA hydrolysis (15, 16). At peak activity, acyl transfer activity reaches 7.0% of hydrolysis activity. That is, one amino acid was incorporated for every 15 TCA molecules hydrolyzed to CA, showing acyl transfer by BSH/T is significant.

I then sought to determine if a similar panel of amino acids would be conjugated when provided GCA instead of TCA. CpBSH/T incubated with GCA at pH 5.0 transferred 11 of 19 supplied amino acids, excluding glycine due to its availability after hydrolysis (Figure 2.1d, Table 2.1). Surprisingly, CpBSH/T produced MetCA and ValCA, otherwise absent when provided TCA. The reduced number of amino acids transferred (11/19 compared to 16/20 when provided TCA) may be a consequence of competition by high glycine concentrations following hydrolysis (Table 2.1).

I also demonstrated that CpBSH/T can ligate amino acids directly to CA (Figure 2.1e), likely occurring through a covalent intermediate (Figure 2.1a, steps 4 and 3). CpBSH/T successfully ligated 12 of 20 amino acids (, including valine and methionine which were not observed with TCA transfer. However, consistent with TCA transfer, ProCA and AspCA were not observed. The absence of proline conjugation may be due to its unique secondary amine preventing proper nucleophilic attack; previous reports also have not observed proline conjugation (17).

#### *2.4.2 - Kinetic characterization of phenylalanochoic acid production by BSH/T*

To investigate the acyl transfer kinetics of CpBSH/T, I first determined a saturation concentration for TCA hydrolysis, as performed previously (13), and found 8 mM TCA more than sufficient to saturate the hydrolysis reaction (Figure 2.2a). Investigating constants for phenylalanine transfer resulted in linear kinetics for PheCA production with increasing phenylalanine concentration (up to 5 mM, Figure 2.2b). For phenylalanine transfer to TCA, doubling the phenylalanine concentration resulted in approximately double the reaction rate. There is also no clear saturating concentration of phenylalanine for this transfer, unlike when characterizing TCA deconjugation. These linear kinetics observed for PheCA formation are consistent with the formation of the enzyme-CA adduct, followed by a rate determining nucleophilic attack to achieve hydrolysis or amino acid acyl transfer (18). Kinetics of phenylalanine ligation to CA were also linear and exhibited rates that were nearly 10% of those for acyl transfer to TCA (Figure 2.2c), again supporting the rate-determining reaction after formation of an enzyme-CA intermediate.



**Figure 2.2: CpBSH/T deconjugation and acyl transfer kinetic characterization**

**a**, Deconjugation kinetics for commercial CpBSH/T when incubated with TCA.  $K_m$ , Michaelis constant;  $V_{max}$ , maximum rate of reaction. **b,c**, Reaction kinetics for the formation of PheCA when incubating 8 mM TCA (**b**) or 8 mM CA (**c**) with CpBSH/T and 1–5 mM phenylalanine. Data are presented as mean  $\pm$  s.e.m.;  $n = 3$  independent reactions per concentration.

## 2.5 - Discussion

Exploring the mechanisms behind microbial BA conjugation by the enzyme BSH, this work begins to shed light on the previously unknown catalytic capacity of one of the most intensely studied enzymes in the history of gut microbiome research. My characterization of phenylalanine acyl transfer by purified BSH/T at optimum pH showed linear kinetics. Given the evidence of a covalently bound BA-enzyme being a requirement for acyl transfer, phenylalanine is in direct competition with water for nucleophilic attack; increasing phenylalanine concentrations will increase the incidence of it outcompeting water to act upon the BA-enzyme intermediate. However, modification of active site structure at the asparagine present at position 79/82, with more detailed description in the following chapter, resulted in diminished MCBA production and reduced ligand diversity. Conversion of Asn82 to Tyr82 resulted in an enrichment of small amino acid ligands with TCA as the BA backbone, notably alanine and glycine. This result may be due to Tyr82 constricting the active site to the point where entry by large or more hydrophobic amino acids is more difficult.

I also observed that CpBSH/T substrate preference impacts amino acid preference in conjugation. *C. perfringens* BSH/T exhibits increased deconjugation of taurine-bound BAs compared to glycine-bound BAs. However, preference for glycine-bound BAs over taurine-bound has been observed for other BSH/T forms, as is the case with BSH/T from *L. plantarum* (formerly *Lactobacillus plantarum*) (19). This work shows that consequences of active site structure modification were more pronounced when CpBSH/T was provided GCA or free CA. Future work involving active site modification of BSH/T from *L. plantarum* could elucidate the impacts of deconjugation substrate preference on MCBA production.

## 2.6 - Methods

### 2.6.1 - Reaction conditions for enzyme characterization and acyl transfer kinetic determination

Lyophilized *C. perfringens* BSH (Creative Enzymes) was resuspended in 0.1 M phosphate buffer at pH 7.0 to a concentration of 2 units  $\mu\text{L}^{-1}$ . 100 mM GCA and TCA stocks were prepared in water, and CA stocks were produced in DMSO. Each enzyme reaction was run in triplicate in bicarbonate, Tris, or citrate-phosphate buffer at the indicated pH using 0.1 units  $\mu\text{L}^{-1}$  of enzyme, 2.5 mM BA (TCA, GCA, or CA), and 125  $\mu\text{M}$  complete amino acid mixture (Promega) for 6.25  $\mu\text{M}$  individual amino acid concentrations. Reactions were incubated at 37 °C, and 30  $\mu\text{L}$  aliquots of the reaction were quenched by the addition of 45  $\mu\text{L}$  methanol at each timepoint for a final concentration of 60% methanol (v:v). Extracts were stored at -80 °C prior to mass spectrometry analysis.

I examined the kinetics of the acyl transfer reaction using phenylalanine, CA, and TCA. Phenylalanine stocks were prepared in sterile water followed by 0.2  $\mu\text{m}$  filtering. Reactions were prepared in 0.1 M citrate buffer at pH 5.3 with a final concentration of 0.05 units  $\mu\text{L}^{-1}$  of enzyme and 8% DMSO (v:v). Reactions were sampled at 1, 5, 10, 15, and 20 min and quenched with cold methanol. Extracts were brought to a final concentration of 50% methanol (v:v). TCA deconjugation kinetics were determined by the addition of 1-8 mM, as previously described (13). Reaction velocities for phenylalanine transfer were determined by the addition of 1-5 mM phenylalanine and 8 mM CA or TCA. Extracts were stored at -80 °C prior to mass spectrometry analysis.

### 2.6.2 - Untargeted metabolomics for BA analysis

Extracts for measuring enzyme activity were not diluted before LC–MS/MS analysis. LC was performed using a Vanquish™ Autosampler (Thermo Scientific) and an Acquity ultra-performance liquid chromatograph (UPLC) bridged ethyl hybrid (BEH) C-18 column, 2.1 mm x 100 mm (Waters). MS was performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) running in positive ion mode. All analyses used a 10 µl injection volume, 0.4 ml min<sup>-1</sup> flow rate and 60 °C column temperature. Samples were eluted using a linear solvent gradient of water (A) and acetonitrile (B), each containing 0.1% formic acid, across a 12 min chromatographic run as follows: 0–1 min, 2% B; 1–8 min, 2–100% B; 8–12 min, 100% B; 10–12 min, 2% B. Data were collected using electrospray ionization in positive mode. MS<sup>1</sup> data were collected using a 35,000 resolution, automatic gain control (AGC) target of 1 × 10<sup>6</sup>, maximum injection time of 100 ms and a scan range set from 100 to 1,500 *m/z* (during min 1–10). Data-dependent MS<sup>2</sup> spectra were collected for the top five most abundant peaks identified in MS<sup>1</sup> survey scans. Files were converted to mzXML format through GNPS Vendor Conversion and submitted to the Global Natural Products Social Molecular Networking Database (GNPS, [gnps.ucsd.edu](http://gnps.ucsd.edu)) for molecular networking and spectral identification (20).

### 2.6.3 - Targeted mass spectrometry for phenylalanochoic acid quantification

Extracts for measuring enzyme activity were not diluted prior to LC-MS/MS analysis. LC was performed using a Vanquish™ Autosampler (Thermo Scientific) and an Acquity ultra-performance liquid chromatograph (UPLC) bridged ethyl hybrid (BEH) C-18 column, 2.1 mm x 100 mm (Waters). MS was performed using a Q Exactive™ Hybrid

Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) running in positive ion mode. All analyses used a 10  $\mu$ L injection volume, 0.4 mL min<sup>-1</sup> flow rate, and 60 °C column temperature. Samples were eluted using a linear solvent gradient of water (A) and acetonitrile (B), each containing 0.1% formic acid, across a 12-min chromatographic run as follows: 0–1 min, 2% B; 1–8 min, 2–100% B; 8–12 min, 100% B; 10–12 min, 2% B. Data were collected using electrospray ionization in positive mode. MS<sup>1</sup> data were collected using a 70,000 resolution, AGC target of  $1 \times 10^6$ , maximum injection time of 100 ms, and a scan range set from 100 to 1500  $m/z$  (during min 1–10). PheCA concentrations were calculated using XCaliber™ software (Thermo Scientific) and an 8-point standard curve containing labeled standards.

## 2.7 - Data availability

Protein structures are available on the Protein Data Bank. *C. perfringens* BSH/T in complex with DCA and taurine, from refs. (16, 21), is available under PDB ID [2BJG](https://doi.org/10.2210/pdb2BJG/pdb) (<https://doi.org/10.2210/pdb2BJG/pdb>). *L. salivarius* BSH/T in complex with TCA, from refs. (22, 23), is available under PDB ID [8BLT](https://doi.org/10.2210/pdb8blt/pdb) (<https://doi.org/10.2210/pdb8blt/pdb>). Raw mass spectrometry data are publicly available in the MassIVE database ([massive.ucsd.edu](https://massive.ucsd.edu)) for CpBSH/T variant analysis under [MSV000092138](https://doi.org/10.25345/C55D8NQ9V) (<https://doi.org/10.25345/C55D8NQ9V>).

GNPS molecular networks are available for CpBSH/T incubation with 1 mM BA and equimolar amino acid mix at [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=3dec8f7ab26d47098406a7e597825154](https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=3dec8f7ab26d47098406a7e597825154) and [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=33da5da024ed44848770a4a02b119d9e](https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=33da5da024ed44848770a4a02b119d9e),

for the CpBSH/T mutagenesis experiment at

[gnps.ucsd.edu/ProteoSAFe/status.jsp?task=30c88ca297a44f84be5fa32b376e5cb9](https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=30c88ca297a44f84be5fa32b376e5cb9).

## REFERENCES

1. Batta AK, Salen G, Arora R, Shefer S, Batta M, Person A. 1990. Side chain conjugation prevents bacterial 7-dehydroxylation of bile acids. *J Biol Chem* 265:10925–10928.
2. White BA, Lipsky RL, Fricke RJ, Hylemon PB. 1980. Bile acid induction specificity of 7 $\alpha$ -dehydroxylase activity in an intestinal Eubacterium species. *Steroids* 35:103–109.
3. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. 2020. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582:566–570.
4. Guzior DV, Quinn RA. 2021. Review: microbial transformations of human bile acids. *Microbiome* 9.
5. Jean-Louis S, Akare S, Ali MA, Mash EA, Meuillet E, Martinez JD. 2006. Deoxycholic acid induces intracellular signaling through membrane perturbations. *J Biol Chem* 281:14948–14960.
6. Zhou Y, Maxwell KN, Sezgin E, Lu M, Liang H, Hancock JF, Dial EJ, Lichtenberger LM, Levental I. 2013. Bile acids modulate signaling by functional perturbation of plasma membrane domains. *J Biol Chem* 288:35660–35670.
7. Dietschy JM. 1968. Mechanisms for the intestinal absorption of bile acids. *J Lipid Res* 9:297–309.
8. Dawson PA, Karpen SJ. 2015. Intestinal transport and metabolism of bile acids. *J Lipid Res* 56:1085–1099.
9. Aldini R, Roda A, Lenzi PL, Ussia G, Vaccari MC, Mazzella G, Festi D, Bazzoli F, Galletti G, Casanova S, Montagnani M, Roda E. 1992. Bile acid active and passive ileal transport in the rabbit: effect of luminal stirring. *Eur J Clin Invest* 22:744–750.
10. Aldini R, Montagnani M, Roda A, Hrelia S, Biagi PL, Roda E. 1996. Intestinal absorption of bile acids in the rabbit: different transport rates in jejunum and ileum. *Gastroenterology* 110:459–468.
11. Hofmann AF, Poley JR. 1972. Role of bile acid malabsorption in pathogenesis of diarrhea and steatorrhea in patients with ileal resection. I. Response to cholestyramine or replacement of dietary long chain triglyceride by medium chain triglyceride. *Gastroenterology* 62:918–934.
12. Dawson PA. 2011. Role of the intestinal bile acid transporters in bile acid and drug disposition. *Handb Exp Pharmacol* 169–203.

13. Gopal-Srivastava R, Hylemon PB. 1988. Purification and characterization of bile salt hydrolase from *Clostridium perfringens*. J Lipid Res 29:1079–1085.
14. Lodola A, Branduardi D, De Vivo M, Capoferri L, Mor M, Piomelli D, Cavalli A. 2012. A catalytic mechanism for cysteine N-terminal nucleophile hydrolases, as revealed by free energy simulations. PLoS ONE 7:e32397.
15. Coleman JP, Hudson LL. 1995. Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens*. Appl Environ Microbiol 61:2514–2520.
16. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2005. Crystal structure of conjugated bile acid hydrolase from *Clostridium perfringens* in complex with reaction products taurine and deoxycholate.
17. Lucas LN, Barrett K, Kerby RL, Zhang Q, Cattaneo LE, Stevenson D, Rey FE, Amador-Noguez D. 2021. Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. mSystems 6:e00805-21.
18. Hinberg I, Laidler KJ. 1972. The kinetics of reactions catalyzed by alkaline phosphatase: the effects of added nucleophiles. Can J Biochem 50:1360–1368.
19. Wang G, Yu H, Feng X, Tang H, Xiong Z, Xia Y, Ai L, Song X. 2021. Specific bile salt hydrolase genes in *Lactobacillus plantarum* AR113 and relationship with bile salt resistance. LWT 145:111208.
20. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kaponi CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya CAP, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Lington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828–837.

21. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2005. Conjugated bile acid hydrolase is a tetrameric N-terminal thiol hydrolase with specific recognition of its cholyl but not of its tauryl product. *Biochemistry* 44:5739–5748.
22. Karlov DS, Long SL, Zeng X, Xu F, Lal K, Cao L, Hayoun K, Lin J, Joyce SA, Tikhonova IG. 2023. Structure of *Lactobacillus salivarius* (Ls) bile salt hydrolase(BSH) in complex with taurocholate (TCA). <https://doi.org/10.2210/pdb8BLT/pdb>. Retrieved 2 December 2023.
23. Karlov DS, Long SL, Zeng X, Xu F, Lal K, Cao L, Hayoun K, Lin J, Joyce SA, Tikhonova IG. 2023. Characterization of the mechanism of bile salt hydrolase substrate specificity by experimental and computational analyses. *Structure* 31:629-638.e5.

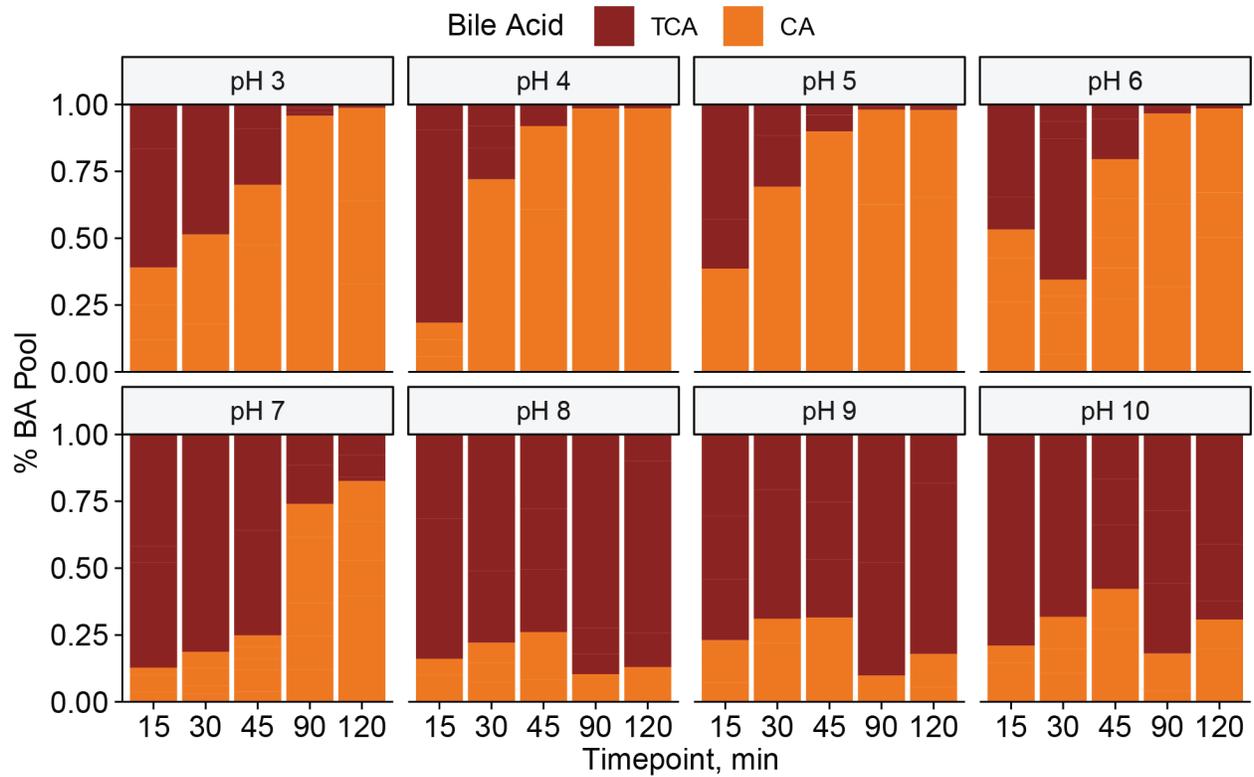
## APPENDIX A: SUPPLEMENTARY TABLES

**Table 2.2: Goodness of fit for curves fit to determine pH optimum for amino acid acyl transfer by *C. perfringens* BSH**

Values in parentheses are the s.e.m. The equation used to calculate the pH optimum (5-factor, bolded) was based on an adjusted R<sup>2</sup>. The coefficient significance was determined by a one-sided t test and the model significance was determined by one-way ANOVA without *P* value adjustment. \**P* < 0.1, \*\**P* < 0.05 \*\*\**P* < 0.01.

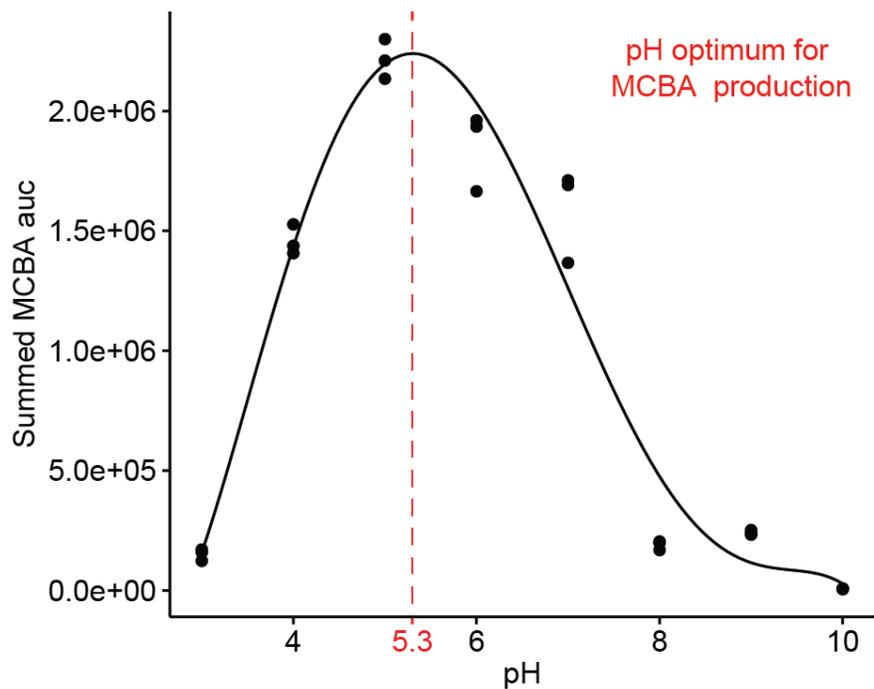
Dependent variable:					
	-1	poly(2)	y poly(3)	poly(4)	poly(5)
Constant	2.002e+06*** (-4.90E+05)	9.630e+05*** (-1.06E+05)	9.630e+05*** (-4.75E+04)	9.630e+05*** (-4.79E+04)	9.630e+05*** (-4.48E+04)
x	-1.599e+05** (-7.11E+04)	-1.795e+06*** (-5.19E+05)	-1.795e+06*** (-2.33E+05)	-1.795e+06*** (-2.35E+05)	-1.795e+06*** (-2.20E+05)
x2		-2.891e+06*** (-5.19E+05)	-2.891e+06*** (-2.33E+05)	-2.891e+06*** (-2.35E+05)	-2.891e+06*** (-2.20E+05)
x3			2.136e+06*** (-2.33E+05)	2.136e+06*** (-2.35E+05)	2.136e+06*** (-2.20E+05)
x4				1.96E+05 (-2.35E+05)	1.96E+05 (-2.20E+05)
x5					-4.231e+05* (-2.20E+05)
Observations	24	24	24	24	24
R <sup>2</sup>	0.187	0.672	0.937	0.939	0.95
Adjusted R <sup>2</sup>	0.15	0.641	0.928	0.926	0.936
Residual	7.98E+05	5.19E+05	2.33E+05	2.35E+05	2.20E+05
Std. Error	(df = 22)	(df = 21)	(df = 20)	(df = 19)	(df = 18)
F Statistic	5.061** (df = 1; 22)	21.526*** (df = 2; 21)	99.244*** (df = 3; 20)	73.479*** (df = 4; 19)	67.932*** (df = 5; 18)

## APPENDIX B: SUPPLEMENTARY FIGURES



**Figure 2.3: TCA deconjugation by commercially available *C. perfringens* BSH/T at pH 3-10**

The proportion of TCA and CA in the BA pool when *C. perfringens* BSH/T was incubated with 8 mM TCA at different pH values across time.  $n = 3$  separate reactions.



**Figure 2.4: pH-dependency of MCBA production by *C. perfringens* BSH/T**

Summed MCBA AUC after 120 min incubation of CpBSH/T, 2.5 mM TCA and 125  $\mu$ M equimolar amino acid mix at various pH values revealing the pH-dependence of BA conjugation,  $n = 3$  independent reactions. Red dashed line indicates the pH 5.3 optimum following derivation as determined by fitting a 5-factor polynomial equation (detailed curve fitting outputs in Table 2.2).

**CHAPTER 3:**  
**DIVERSITY OF BACTERIA CAPABLE OF MCBA PRODUCTION AND THEIR  
ASSOCIATED CONJUGATED BILE ACID PRODUCTS**

### 3.1 - Preface

Portions of this chapter were published in the journal Nature in 2024 (Material from: Guzior, D.V., Okros, M., Shivel, M. *et al.* Bile salt hydrolase acyltransferase activity expands bile acid diversity. *Nature* **626**, 852–858 (2024). <https://doi.org/10.1038/s41586-024-07017-8>). Per the publisher, Springer Nature, “Authors have the right to reuse their article’s Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.” Dr. Yousi Fu performed *Lachnoclostridium scindens* genome mining and provided the amino acid sequence used to generate Figure 3.4b. Dr. Robert A. Quinn performed the BSH/T amino acid sequence alignment and generated Figure 3.5 and 3.8). Dr. Robert P. Hausinger generated Figure 3.6a.

### 3.2 - Abstract

Microbial metabolism of human bile acids is one of the dominant drivers of gut microbiome structure and diversity. Microbial transformations including deconjugation, dehydroxylation, oxidation, and epimerization all increase the diversity of the human bile acid pool resulting in hundreds of potential forms (1). Given how recently microbially conjugated bile acids (MCBAs) were described, little is known about the breadth and specificity of amino acid use. Here, I investigate the diversity of conjugated bile acids produced by 29 bacterial strains when grown in medium supplemented with cholic acid (CA) and taurine. 16 of the 20 proteinogenic amino acids were used in conjugation in addition to the supplemented taurine and non-essential amino acid citrulline. Valine, methionine, proline, and arginine conjugation was not observed. 19 of 29 strains produced at least one conjugated BA. The most robust producer, *Lactiplantibacillus plantarum*, utilized all 16 proteinaceous amino acids. I then investigated connections between genome phylogeny and MCBA production, or lack thereof. MCBA production did not correlate with evolutionary relatedness, instead correlating with amino acid sequences of the enzyme bile salt hydrolase, henceforth referred to as bile salt hydrolase/transferase (BSH/T). Mapping MCBA profiles to BSH/T phylogenetic trees revealed three distinct clades enriched with BSH/T sequences from high-production strains, a mix of high and low producers, and low-production strains.

### 3.3 - Introduction

Microbial metabolism of human bile acids (BAs) is a dominant driver of structure and diversity of our gut microbiome (1). Taurocholic acid (TCA), a primary BA, is a known germinant for *Clostridioides difficile* within the small intestine (2, 3). However, the secondary BA deoxycholic acid (DCA) inhibits *C. difficile* infection by both inhibiting germination and decreasing the efficiency of sporulation (4). Conversion of a cholic acid (CA) to DCA through 7 $\alpha$ -dehydroxylation results in increased BA toxicity (1). Consequences are easily observed when measuring bacterial growth. Reported effective doses of DCA against *Staphylococcus aureus* were 5% of CA (5). Additionally, vegetative *C. difficile* grown in medium containing 50  $\mu$ M DCA showed marked inhibition compared to 50  $\mu$ M CA (6). Thus, both host and microbial BA metabolism are important drivers of microbiome dynamics and community structure.

Our understanding of BA metabolism has been taxonomically limited, namely involving members of the genus *Clostridium*. This comes as no surprise given the extensive history of research into microbial BA metabolism in the family Clostridiaceae. In 1966, Gustaffson *et al.* were the first to show that bacterial monocultures were able to convert the primary BAs CA and chenodeoxycholic acid (CDCA) to DCA and lithocholic acid (LCA) respectively (7). *Lachnoclostridium scindens* (formerly *Clostridium scindens*) is perhaps the most well-studied bile acid metabolizing microorganism; *L. scindens* was the first bacterium identified to dehydroxylate bile acids, with the full enzymatic pathway having been elucidated in the context of this species (8). When looking at the “gateway” reaction to bile acid metabolism, bile acid deconjugation, investigation into the taxonomic diversity of bacterial bile salt hydrolase genes within the gut has been an ongoing and

vigorous area of research. Work from Song *et al.* investigating sequence diversity found that *bsh* sequences were identified across bacterial taxa, being present in 591 strains across 117 genera (9).

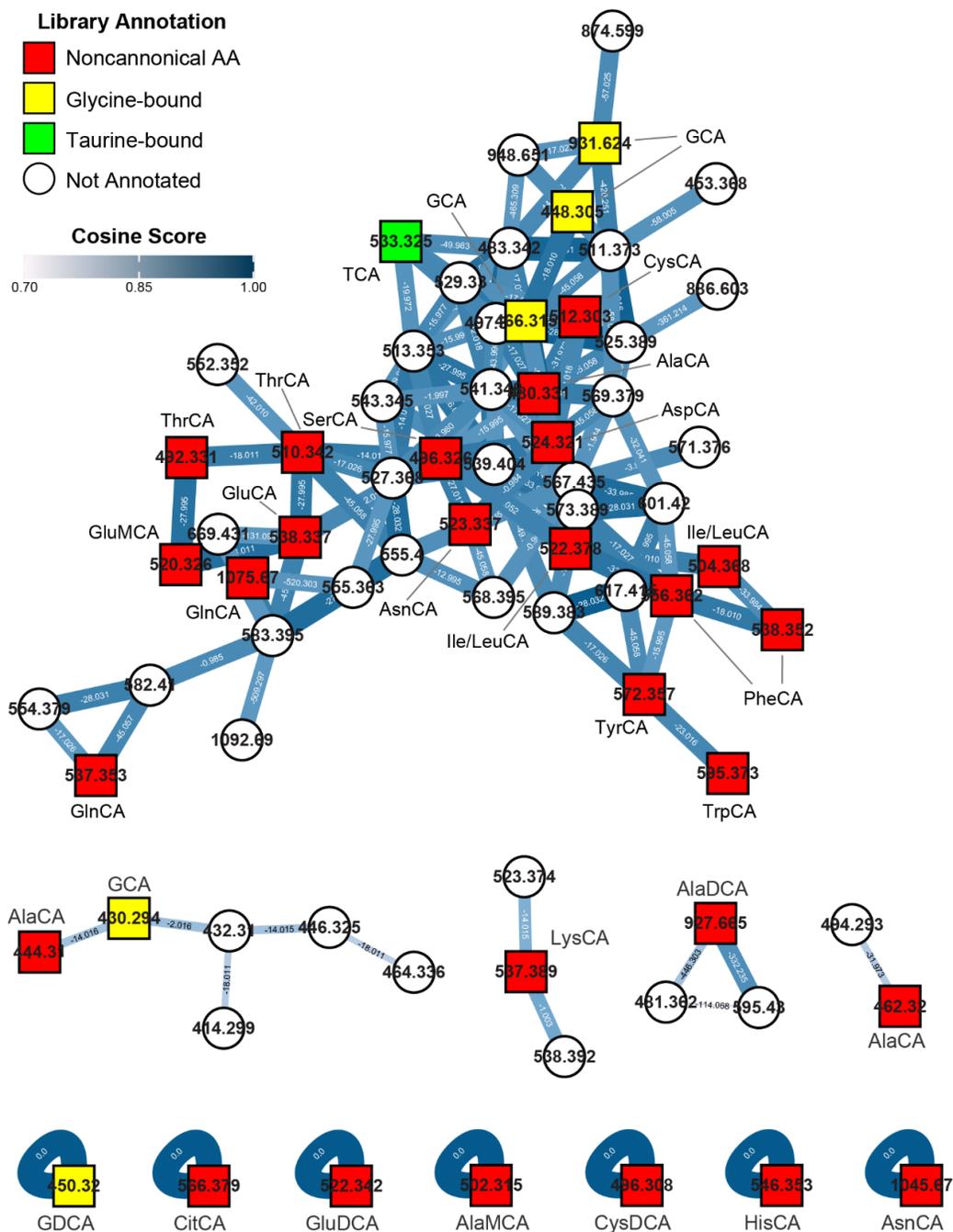
Given both the sequence diversity and widespread occurrence of *bsh in vivo*, the characterization of both bacterial strains capable of MCBA production in addition to the diversity of MCBA products is an important first step in understanding the biological implications of these recently identified molecules. It is essential to understand not only the purpose of MCBA production from the perspective of bacterial fitness and competition, but of teasing apart the context of the relationship between MCBA-producing bacteria and their host. Here, I screened 29 bacterial strains for MCBA production with a focus on the family *Lachnospiraceae*, of which the original producer *Enterocloster bolteae* is a member, in addition to other relevant gut bacteria. I leveraged molecular networking approaches to investigate the diversity of MCBAs produced by those strains in order to look at both annotated MCBAs and those without current annotation. This process was followed by phylogenetic analysis to investigate the drivers of MCBA production and subsequent product profiles. I show that evolutionary relatedness is not a key marker for MCBA production, whereas BSH/T amino acid sequences correlate with the MCBA profile. Further investigation highlighted the amino acid at position 82 as a determinant of MCBA production, whether that residue is coded as asparagine or tyrosine. Mutagenic studies showed that this residue is not essential for deconjugation, but it is a significant driver of total MCBA production and overall diversity of the MCBAs produced.

## 3.4 - Results

### 3.4.1 - Microbial bile acid conjugation results in a highly diverse suite of MCBAs

The original publication by Quinn *et al.* describing the identification and production of microbially conjugated bile acids verified production of phenylalanochoic acid (PheCA), leucochoic acid (LeuCA), and tyrosochoic acid (TyrCA) (10). I sought to investigate if additional amino acids could be used in this novel transformation. As taurine is absent in all standard media formulations, it was therefore supplemented to determine if both host-conjugated forms of CA can be produced by bacteria. I observed 16 of 20 proteinogenic amino acids being used in BA conjugation (Figure 3.1). Proline, methionine, valine, and arginine conjugation were not observed. Otherwise, the ability to utilize different amino acids was not driven by the size, charge, or polarity of the respective R group. The capacity of each amino acid for use in conjugation was also primarily uninfluenced by the amino acid structure, with proline as a potential exception. Among these conjugates were MCBAs built from DCA, notably a secondary BA and derivative of CA, *in lieu* of CA. Additional ligands outside of the 20 proteinogenic amino acids were used in this transformation. In addition to amino acid conjugates, citrulline-conjugated cholic acid (citrullochoic acid, CitCA) and TCA production were both observed.

Analysis of annotated compounds allowed for high-throughput characterization of microbial BA conjugation. However, utilizing an untargeted method allows for network construction of both known and unknown compounds based on MS<sup>2</sup> spectral similarity. Leveraging this analysis, the number of measured MCBA species underwent a stark increase (Figure 3.1). Features annotated as taurine and glycine-conjugated BAs form complex networks with MCBAs bound to noncanonical amino acids in addition to amino



**Figure 3.1: MS<sup>2</sup>-based molecular networking illustrates the unknown conjugated BA diversity**

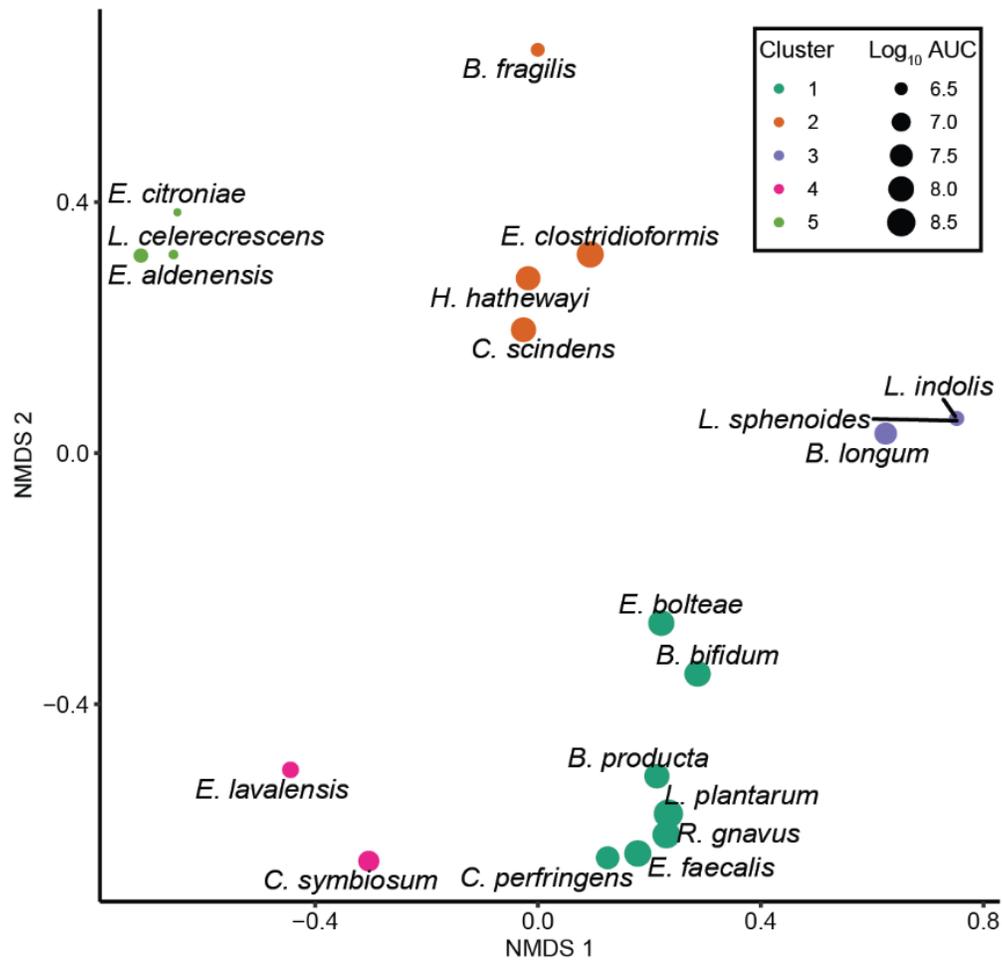
Molecular networks containing MCBA. Bacterial strains were screened for MCBA production after being solely provided with CA and taurine, thus any conjugated BAs are of microbial origin. The node shape corresponds to library annotation with the node color corresponding to the conjugated BA class. 18 different substrates were utilized for BA conjugation, including glycine and taurine in addition to previously unreported citrulline. Edge color and thickness represent the cosine score, a measure of spectral similarity between two metabolites. The mass difference between two nodes is labeled across each edge.

acid derivatives. These networks included dihydroxy-BA backbones, specifically CDCA and DCA, which were not provided initially but must have instead been produced *de novo* within CA-treated monocultures. Production of dihydroxy-MCBAs was primarily associated with *L. scindens*, an organism frequently associated with BA 7 $\alpha$ -dehydroxylation (1, 11, 12).

#### 3.4.2 - MCBA production is not linked to evolutionary relatedness

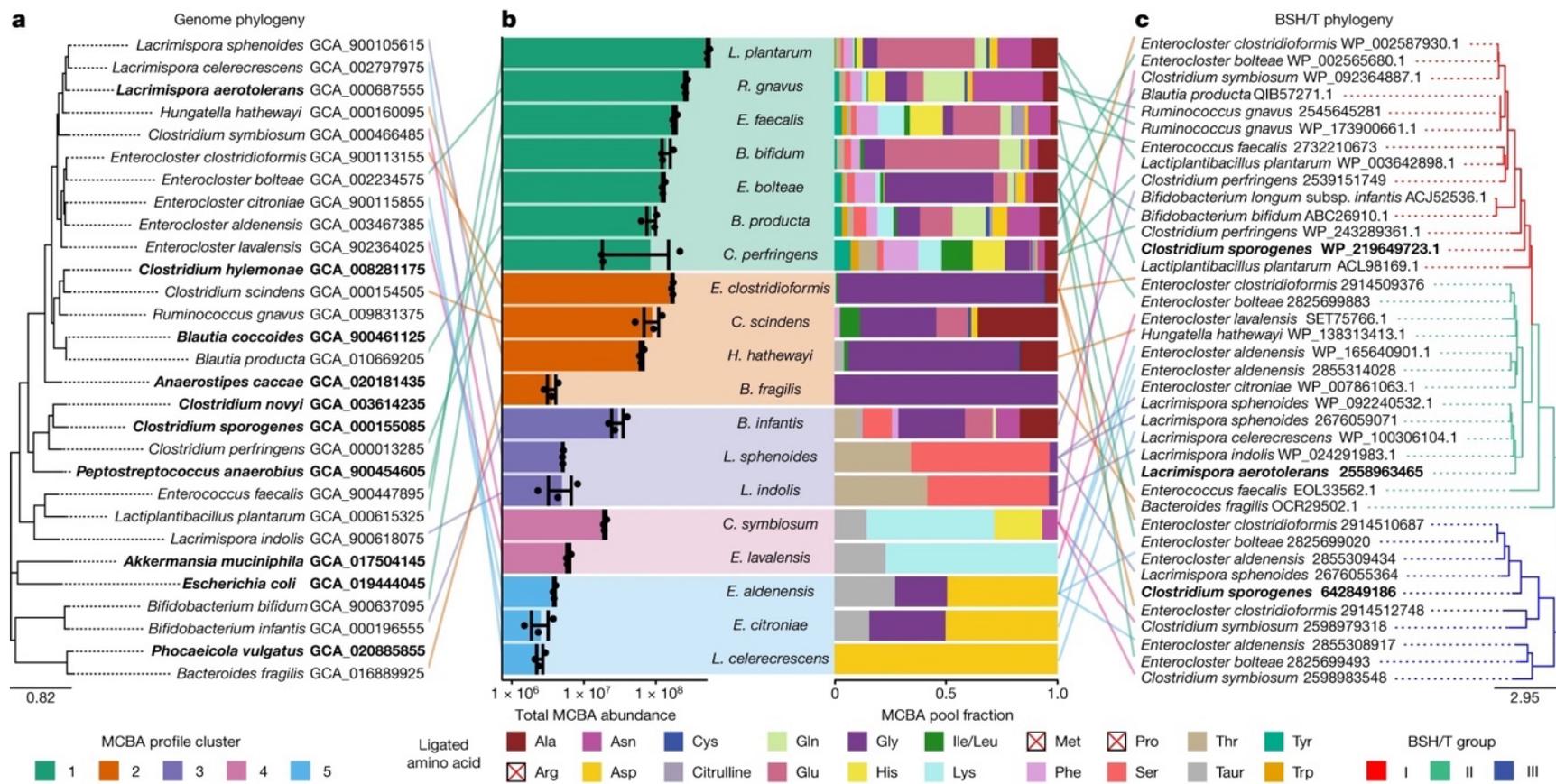
After confirming MCBA production by BSH/T *in vitro*, we analyzed the genomes of 29 strains that were subsequently screened for MCBA production to determine if phylogenetic relatedness correlated with conjugation capability (Figure 3.3a, Table 3.1). These included Actinomycetia, Verrucomicrobiae, Gammaproteobacteria, Bacilli, and Clostridia, with a focus on the *Lachnospiraceae* family in Clostridia. Of these strains, 19 produced at least one MCBA (Figure 3.3b) and production was particularly prevalent among the *Lachnospiraceae*, with only Verrucomicrobiae (*Akkermansia muciniphila*) being unable to produce MCBAs. The most robust MCBA producers, *Lactiplantibacillus plantarum*, *Ruminococcus gnavus*, *Enterococcus faecalis*, and *Bifidobacterium bifidum* subsp. *infantis*, were phylogenetically disparate, indicating little association between evolutionary relatedness and MCBA production (Figure 3.3a).

Hierarchical clustering was employed to investigate additional commonalities between robust and inefficient MCBAs producers. Calculating Bray-Curtis dissimilarity based on the amino acid used in conjugation revealed distinct groupings among the strains screened. Strains clustered based on total MCBA abundance in addition to the diversity of amino acids used (Figure 3.2 and Figure 3.3, Table 3.2–



**Figure 3.2: Dissimilarity between MCBA-producing strains based on amino acid use in conjugation**

Nonmetric data scaling using Bray-Curtis dissimilarity of amino acids used in BA conjugation, using average amino acid auc per strain. Color represents cluster assigned based on cluster analysis and dot size represents the average total MCBA abundance.  $n = 3$  independent cultures.



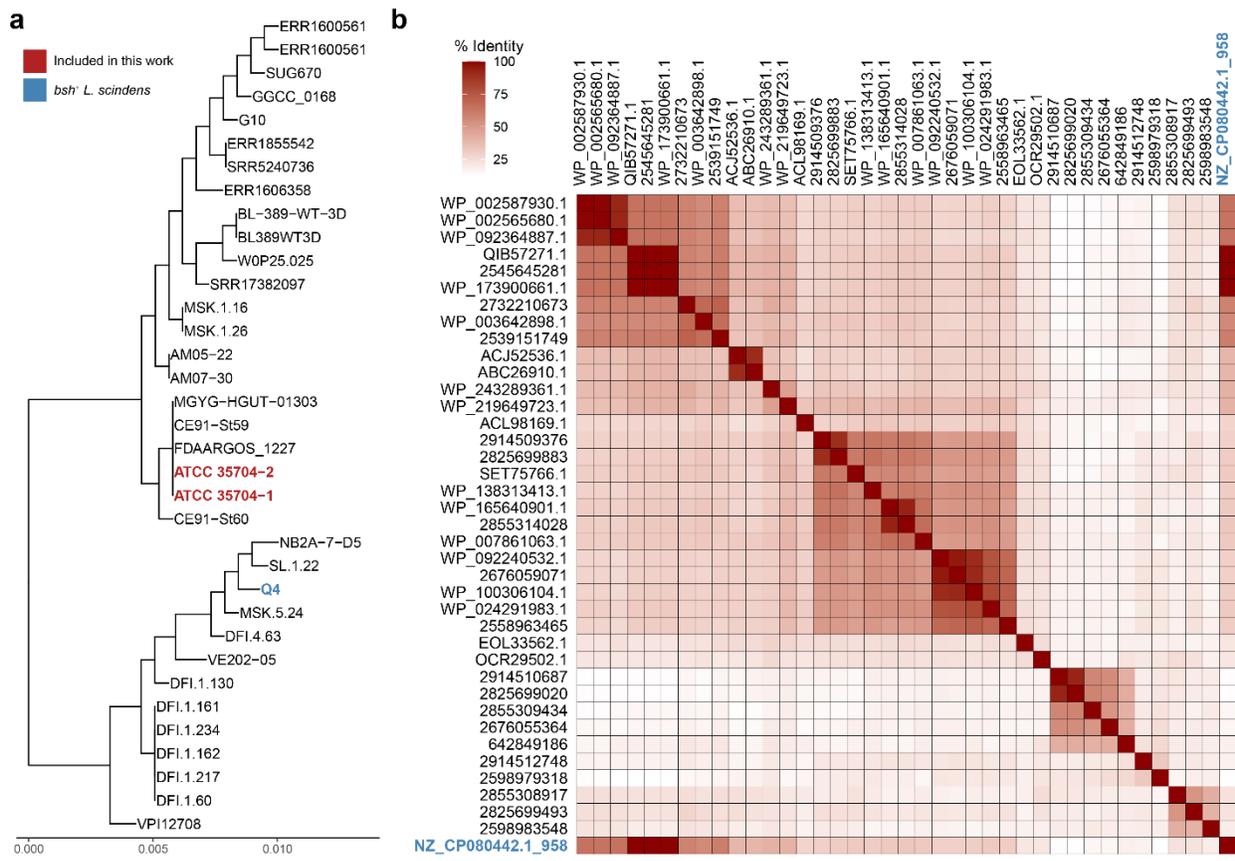
**Figure 3.3: MCBA product identities correlate with BSH/T amino acid sequences**

**a**, Genome phylogenetic tree for all strains screened, with strains in bold denoting those that did not produce MCBA. **b**, Summed MCBA abundances and profiles from gut bacteria that were grown in the presence of 1 mM CA, colored by MCBA profile cluster. Data are presented as mean  $\pm$  s.e.m.;  $n = 3$  independent cultures. **c**, Phylogenetic relatedness of BSH/T amino acid sequences showing three clusters of related sequences. Lines connect the genome and BSH/T sequences to the product profiles for each strain. Line color corresponds to MCBA profile cluster.

Figure 3.6). Cluster 1 strains showed robust conjugation of a wide variety of amino acids (Figure 3.2). Strains in cluster 2 favored glycine and alanine conjugation (Figure 3.3), whereas cluster 3 preferentially conjugated small, hydrophilic amino acids (Table 3.4). Cluster 4 showed extensive lysine conjugation (Table 3.5) and cluster 5 profiles were dominated by aspartate conjugation (Table 3.6). The most robust MCBA producer, *L. plantarum*, lies in cluster 1 and produced 16 of 18 observed MCBAs (Figure 3.3). Although clustering showed little phylogenetic correlation, clusters 3 and 5 were primarily associated with members of the *Lachnospiraceae* (Figure 3.3).

#### 3.4.3 - BSH/T sequence shapes the associated conjugation profile

Genomes of all 29 species were mined to search for *bsh/t* presence in order to investigate relationships between the translated protein sequences and the MCBA profiles (Figure 3.5, Table 3.1). Two species, *Clostridium sporogenes* and *Lacrimispora aerotolerans*, possess annotated *bsh/t* but did not produce MCBAs. Further analysis revealed valine in place of a traditional start codon in *C. sporogenes* BSH/T, which may explain why MCBA production was absent in this bacterium. In contrast, *Lachnoclostridium scindens* ATCC 35704 produced MCBAs while lacking an annotated *bsh/t*, matching previous reports (13). Analysis of 35 publicly available *L. scindens* genomes for *bsh/t* presence showed only strain, Q4, contained a predicted *bsh/t* (Figure 3.4a,b, Table 3.7). *L. scindens* Q4 BSH/T displayed high amino acid sequence similarity to BSH/T from *R. gnavus* and other sequences with BSH/T cluster 1 (Figure 3.4b). The absence of *bsh/t* in MCBA-producing *L. scindens* ATCC 35704 suggests that other enzymes capable of BA conjugation remain to be discovered.



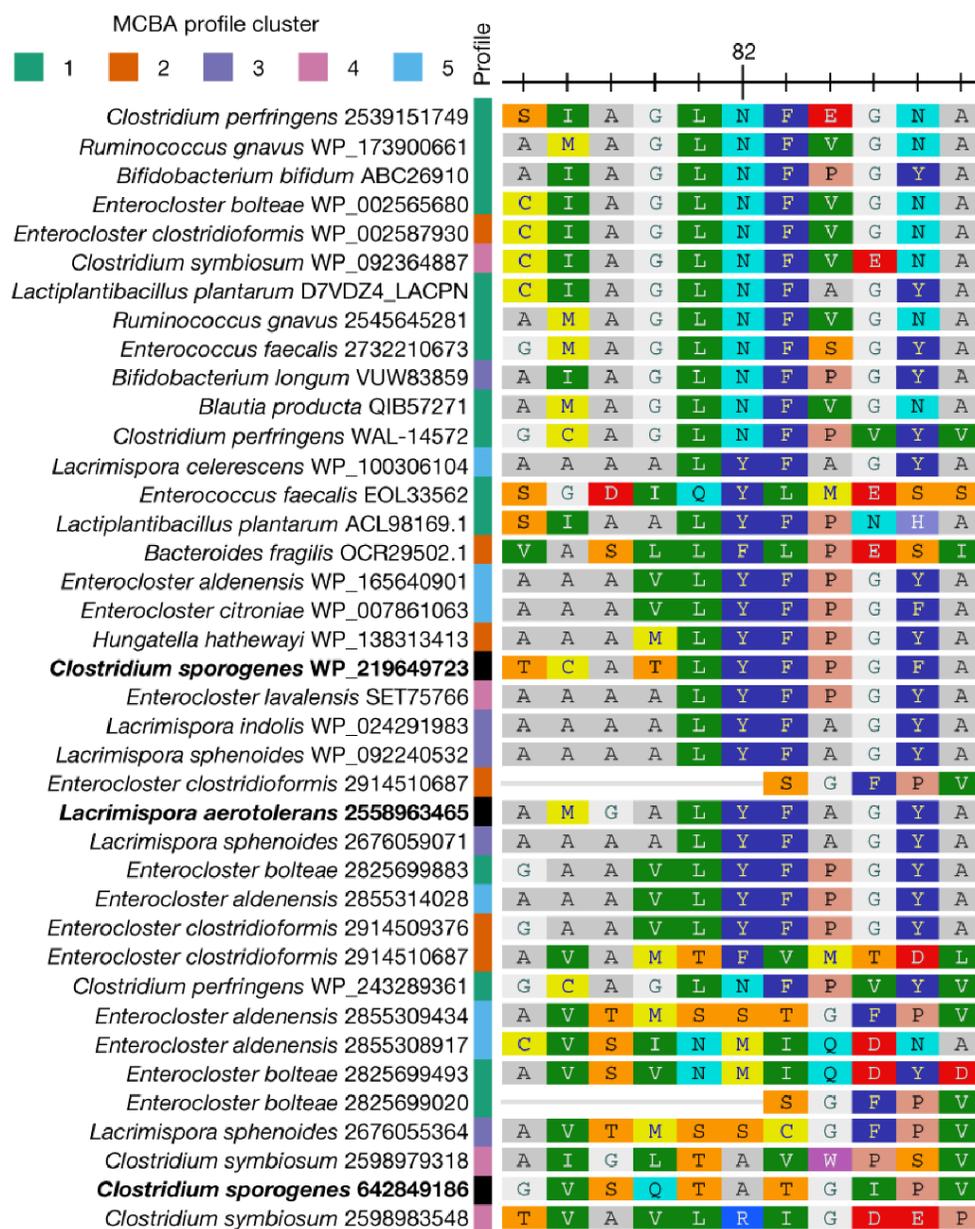
**Figure 3.4: *Lachnospiridium scindens* genome analysis for putative *bsh/t* annotation**

**a**, Phylogenetic analysis of 35 publicly available genomes for *L. scindens*. The ATCC type strain, used in this work, has two deposited genomes and is highlighted in red. The only strain with a predicted *bsh/t* was *L. scindens* strain Q4, highlighted in blue. **b**, Pairwise BSH/T amino acid sequence similarity of all strains included in this work (matching Fig. 2c), now including the predicted BSH/T present in *L. scindens* strain Q4 (NZ\_CP080442.1\_958, based on Prokka analysis).

The remaining 18 MCBA producers had at least one annotated or predicted *bsh/t* in their genome with some, such as *E. bolteae*, containing at least three. BSH/T phylogenetic tree topology (Figure 3.3c) showed limited correlation to the five MCBA profile clusters (Figure 3.3b). However, there were three main BSH/T lineages: group I, containing a set of diverse and robust MCBA producers; group II, primarily associating with MCBA clusters 3 and 5; and group III, showing significant sequence divergence from the other groups and little association with MCBA profiles. The last group may represent sequences with a high degree of similarity to other enzymes in the Ntn-hydrolase superfamily, indicating that these BSH/T homologues may have other functions. *E. bolteae* and *E. clostridioformis* contain BSH/T sequences from all three groups, yet *E. bolteae* produced a diverse MCBA profile whereas glycocholic acid (GCA) dominated MCBA produced by *E. clostridioformis*.

#### 3.4.4 - Active site residues impact BSH/T conjugated amino acid selectivity

Analysis of BSH/T amino acid sequence alignment showed an amino acid substitution that was potentially responsible for divergence in the conjugation profiles observed. Asn82 (Figure 3.5; *Clostridium perfringens* BSH/T as reference) (14, 15) was reported as being highly conserved in BSH/T sequences in previous studies (16). However, I show that this position is instead a tyrosine in BSH/T from several *Lachnospiraceae* species, most residing in BSH/T group II. This residue lies in the active site of the BSH/T crystal structure from *C. perfringens* (PDB ID 2bjg) (14, 15), adjacent to the carboxylate of co-crystallized DCA (Figure 3.8) and directly at the location of the amide bond of TCA in *Lactobacillus salivarius* BSH/T co-crystallized with TCA

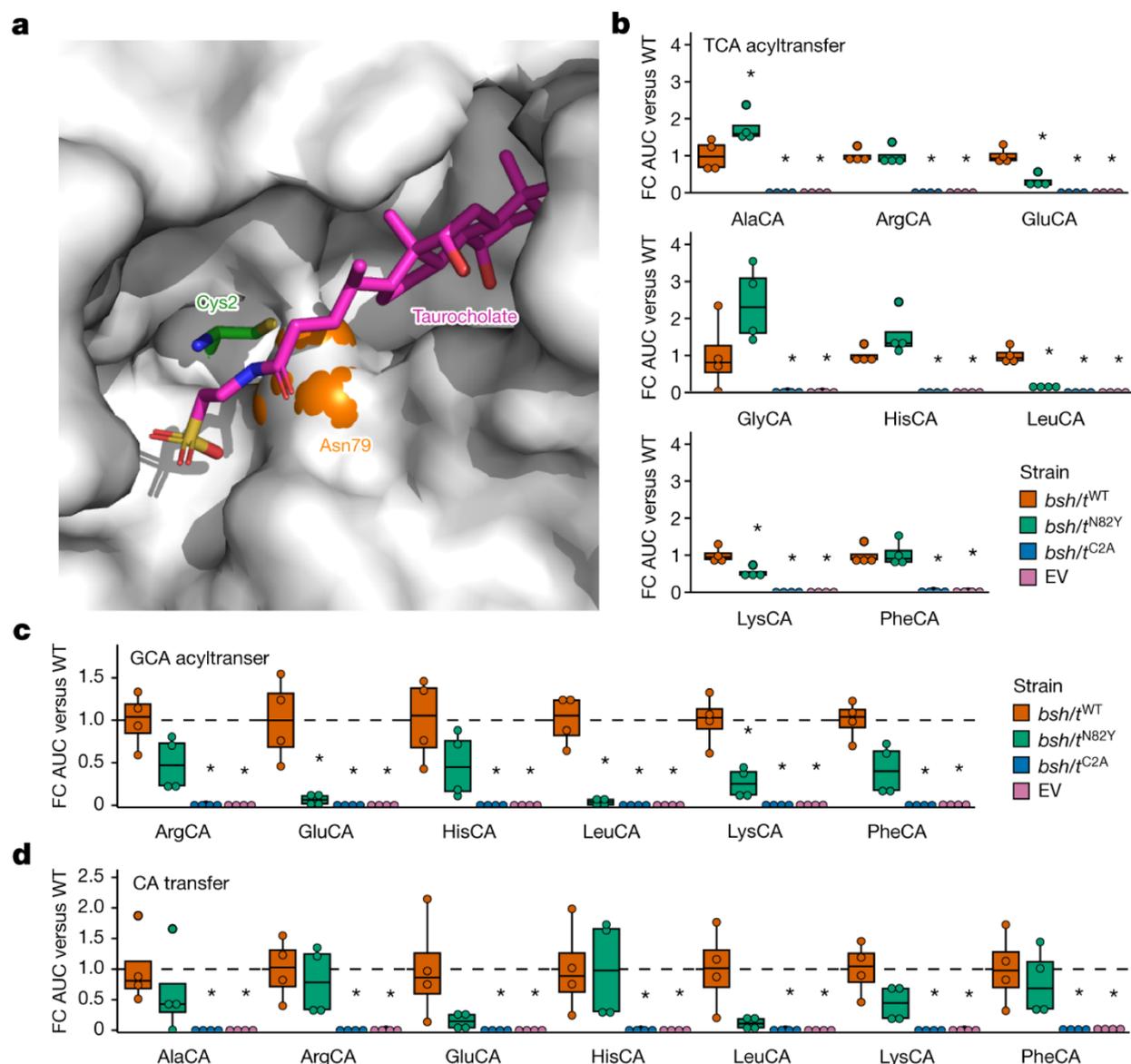


**Figure 3.5: BSH/T partial sequence alignment of strains screened for MCBA production**

BSH/T amino acid sequence alignment highlighting conserved Asn82 or Tyr82 for Clostridia-like and *Lachnospiraceae*-like BSH/T sequences, respectively, with MCBA profile cluster identified next to the strain and BSH/T accession number.

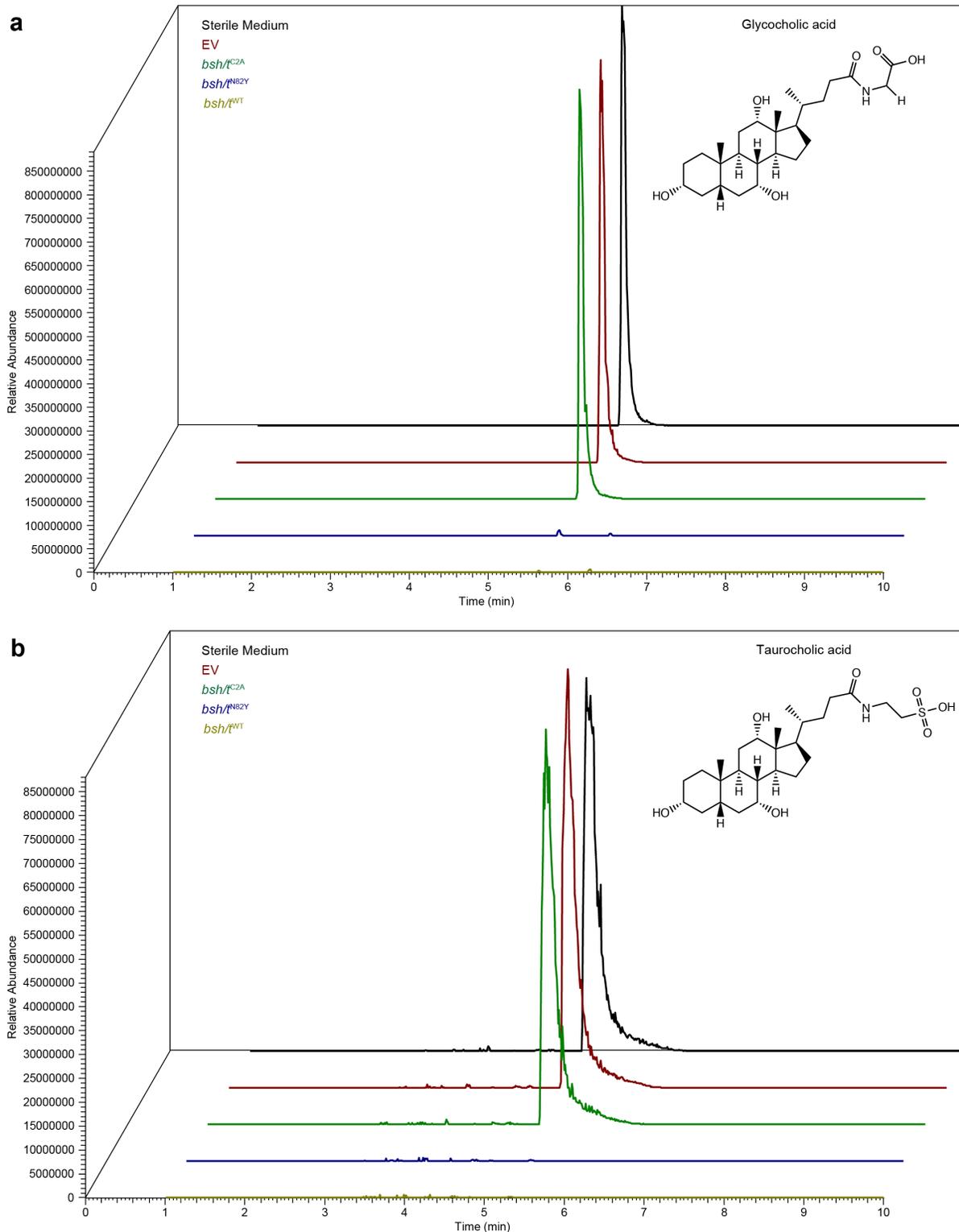
(PDB ID 8blt, Figure 3.6a) (17). We therefore proposed that BSH/T sequence variation at the active site determines its capacity for BA conjugation and I substituted both Asn82 and Cys2 of CpBSH/T in order to test this hypothesis.

Alteration of Asn82 to Tyr82 (N82Y) in CpBSH/T shaped the amino acid conjugation pool in a similar fashion observed for organisms encoding either of these variants. *Escherichia coli* expressing N82Y variants demonstrated significant deficits in BA conjugation, with decreased abundance of glutamatocholic acid (GluCA), lysocholic acid (LysCA), and LeuCA when grown in medium containing 1 mM GCA or 1 mM TCA (Figure 3.6). These trends were also seen when grown in medium containing 1 mM CA. However, alanocholic acid (AlaCA) was significantly enriched in the N82Y variant compared to the wild type (WT) protein when provided TCA. By contrast, Cys2 substitution (C2A) resulted in complete ablation of BA conjugation, regardless of substrate. C2A variants were also unable to hydrolyze TCA, whereas WT and N82Y variants completely hydrolyzed TCA (Figure 3.7a) and most GCA (Figure 3.7b) to CA.



**Figure 3.6: Nonessential active site residues drive amino acid selectivity in MCBA production**

**a**, Structure of *L. salivarius* BSH/T (PDB ID 8BLT)(17, 18) in complex with TCA (molecular surface representation) showing the proximity of Asn79 (Asn82 in CpBSH/T) and catalytic Cys2 with the amide bond of TCA. **b–d**, Fold-change (FC) in abundance of MCBA produced by *C. perfringens* BSH/T with substitutions in Asn82 (*bsh/t*<sup>N82Y</sup>) or Cys2 (*bsh/t*<sup>C2A</sup>) compared to WT when expressed by *E. coli* DH5α incubated with 1 mM TCA (**b**), 1 mM GCA (**c**) or 1 mM CA (**d**), using endogenous amino acids for BA conjugation. EV denotes pBAD18-Cm without insert; *n*=4 independent cultures. Data in **b–d** are presented as boxplots where the middle lines are the median, while lower and upper hinges represent the first and third quartiles, upper whiskers extend to maxima and lower whiskers extend to minima. Statistical significance in **b–d** was determined by Wilcoxon rank-sum test against the WT enzyme results with Benjamini–Hochberg *P* value correction, \**P* < 0.05.



**Figure 3.7: GCA and TCA extracted ion chromatograms following 24 h induction of *C. perfringens* BSH/T variants in *E. coli***

Representative **a**, GCA and **b**, TCA extracted ion chromatograms showing significantly diminished in WT and N82Y variant strains with minimal change in the C2A variant and EV control.

### 3.5 - Discussion

Ligand use in microbial BA conjugation was notably diverse in our experiments. We observed 16 of 20 proteinogenic amino acids ligated to BAs; only valine, proline, methionine, and arginine conjugation was not observed. The secondary amine structure of proline may contribute to its lack of use in BA conjugation, something that has been reported in other work (19). Observing citrulline and taurine use in conjugation demonstrates that MCBA diversity is not limited to essential amino acids alone. Given the propensity of TCA production by the human liver, it comes as no surprise that the carboxylate group present on common amino acids is not required for use in conjugation. The recent description of DCA conjugation with  $\gamma$ -aminobutyric acid (GABA) and tyramine further expand the known ligands involved in microbial BA conjugation (13). Similar to taurine, GABA and tyramine are decarboxylated forms of amino acid (glutamate and tyrosine, respectively) still containing an amine (20, 21). Independent of amino acid metabolism, teasing apart the true diversity of the BA pool becomes limited by the methods available to search for novel, otherwise undescribed ligands. As MS/MS spectral libraries continue to fill with references and more powerful *in silico* tools are developed, so too will our understanding of the truly rich diversity present in our bile.

The lack of observable valine, methionine, and arginine conjugation is not explained by steric hinderances. It may instead be the case that these three amino acids are used in conjugation, but at such low levels to be undetectable using our current methods. Other considerations include differences in composition and concentration of intracellular amino acids between strains. Similar to differences in GC content within bacterial genomes, biases for amino acid use in proteomes exist between species (22–

24). This may explain, in part, differences seen within MCBA profile clusters when total MCBA concentration is otherwise similar, particularly in the high production, high diversity MCBA profile cluster 1. *L. plantarum* and *R. gnavus* show similar diversity in amino acid use with neither producing TCA and only *R. gnavus* producing CitCA, yet there is clear enrichment of glutamatocholic acid (GluCA) production by *L. plantarum* compared to enrichment of glutamocholic acid (GlnCA) and asparagocholic acid (AsnCA), with slight enrichment of histidocholic acid (HisCA) in the case of *R. gnavus*.

Finally, our investigation into impacts of active site structure were limited to one catalytically nonessential residue. However, we show here that the overall structure of the active site is an important driver in amino acid use in BA conjugation. Several groups have reported additional active site residues important for bile acid deconjugation in addition to Asn79/82 analyzed here (9, 25). Further investigation into the contributions of these other residues in substrate selectivity in the context of BA conjugation not only builds the repertoire of nonessential residues known, but allows for potential engineering of BSH/T.

## 3.6 - Methods

### 3.6.1 - Bacterial strains, media, and growth conditions

All media used in anaerobic experiments were pre-reduced for at least 24 h in an anaerobic chamber prior to inoculation and cultures were grown in an atmosphere containing 98% nitrogen and 2% hydrogen. Bacterial cultures were grown from glycerol freezer stocks in Reinforced Clostridial Medium (RCM, Merck), Brain Heart Infusion medium (BHI, Merck), and BHI supplemented with 5  $\mu\text{g mL}^{-1}$  hemin, 1  $\mu\text{g mL}^{-1}$  vitamin K, 10  $\text{g L}^{-1}$  yeast extract, and 0.5  $\text{g L}^{-1}$  L-cysteine (BHIS). A full list of strains used in this work can be found in Table 3.1.

### 3.6.2 - Phylogenetic analysis and BSH/T visualization

Genomic sequences were acquired from GenBank and BSH/T amino acid sequences were obtained from the Joint Genome Institute (JGI) and the National Center for Biotechnology Information (NCBI) protein databases (Table 3.1). Phylogenetic trees were constructed using FastTree (26) and visualized in R (version 4.2.2) (27) using the 'ggtree' package (v.3.6.2) (28). BSH/T sequences were aligned using the NCBI constraint-based aligner tool (COBALT) (29). Parameters of the alignment were set to defaults, including an E-value of 0.003, word size of 4, and maximum cluster distance of 0.8. The 3D structure of *C. perfringens* BSH/T (PDB ID: 2bjg) (15, 30) was visualized using the Protein Data Bank online structure viewer, Mol\*Viewer (31). The 3D structure of *L. salivarius* BSH/T (PDB ID: 8blt) (17, 18) was visualized using PyMOL (v.2.5.4, Schrödinger Inc.).

For *L. scindens* predictive *bsh/t* analysis, genomic sequences were obtained from NCBI (Table 3.7,  $n = 35$  strains). Prodigal (version 2.6.3) (32) was used to predict protein-

encoding genes present within each genome followed by searching for predicted BSH/T sequences using DIAMOND (version 0.9.36.137) (33).

### 3.6.3 - *In vitro* screen for MCBA production

Overnight cultures were grown from freezer stocks, anaerobically at 37 °C. Once the optical density measured at 600 nm ( $OD_{600}$ ) reached at least 0.10, cultures were diluted to a final  $OD_{600}$  of 0.01 in medium with or without 1 mM CA and 100  $\mu$ M taurine in 96-deep-well plates (Thermo Fisher Scientific). Plates were sealed with a rubber mat (Thermo Fisher Scientific) and incubated for 24 h at 37 °C under anaerobic conditions (98%  $CO_2$ , 2%  $H_2$ ).  $OD_{600}$  was measured and metabolite extraction was performed by diluting whole cultures 2:3 (v:v) in 100% ice cold methanol in 1.7 mL microcentrifuge tubes (Axygen) followed by overnight incubation at 4 °C. Extracts were then centrifuged at 10,000  $g$  for 5 min to pellet cell debris followed by storage at -80 °C prior to liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS) analysis.

### 3.6.4 - *Untargeted metabolomics for BA analysis*

Bacterial culture extracts were diluted 1:1 (v:v) in 50% methanol containing 2.5  $\mu$ g  $mL^{-1}$  phenol red internal standard prior to LC-MS/MS analysis. LC was performed using a Vanquish™ Autosampler (Thermo Scientific) and an Acquity ultra-performance liquid chromatography (UPLC) bridged ethyl hybrid (BEH) C-18 column, 2.1 mm x 100 mm (Waters). MS was performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) running in positive ion mode. All analyses used a 10  $\mu$ L injection volume, 0.4  $mL\ min^{-1}$  flow rate, and 60 °C column temperature. Samples were eluted using a linear solvent gradient of water (A) and acetonitrile (B), each containing 0.1% formic acid, across a 12-min chromatographic run as follows: 0–1 min, 2% B; 1–8

min, 2–100% B; 8-12 min, 100% B; 10-12 min, 2% B. Data were collected using electrospray ionization in positive mode. MS<sup>1</sup> data were collected using a 35,000 resolution, automatic gain control (AGC) target of  $1 \times 10^6$ , maximum injection time of 100 ms, and a scan range set from 100 to 1500 *m/z* (during min 1–10). Data-dependent MS<sup>2</sup> spectra were collected for the top 5 most abundant peaks identified in MS<sup>1</sup> survey scans.

### 3.6.5 - Metabolite annotation and molecular network visualization

Thermo RAW files were converted to mzXML format via GNPS Vendor Conversion and submitted the Global Natural Products Social Molecular Networking database (GNPS, [gnps.ucsd.edu](http://gnps.ucsd.edu)) for molecular networking and spectral annotation (34, 35). First, MS/MS data were filtered by removing fragment ions within 17 Da of the precursor *m/z*. MS/MS spectra were window filtered by choosing only the top 6 fragment ions within the 50 Da window throughout the spectrum. Precursor ion and MS/MS fragment ion mass tolerance values were set to 0.02 Da. A molecular network was then created where edges between two nodes were filtered to have a cosine score above 0.7 and more than 4 matched peaks. Edges were only kept in the network if each of the nodes appeared in each other's top 10 most similar nodes. Spectra were then searched against GNPS' spectral libraries for molecular annotation. Library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. The resulting networks were visualized using Cytoscape (36).

For *bsh/t* expression and mutagenic studies, the converted files were submitted to GNPS for classic molecular networking to identify MCBAs present in each sample. Peak area-under-curve (auc) abundances were calculated using XCaliber™ software (Thermo

Scientific) based on  $m/z$  and the retention time of each MCBA annotated by GNPS (Table 3.8).

#### 3.6.6 - Site-directed mutagenesis of *C. perfringens* BSH/T active site

*C. perfringens* *bsh/t* was amplified and inserted into pBAD18-Cm via Gibson assembly (New England Biolabs). All primers can be found in Table 3.9. Resulting products were cloned into chemically competent *Escherichia coli* DH5 $\alpha$ . Plasmid purification was performed via Mini-Prep (Qiagen) and inserts were verified via PCR. Site-directed mutagenesis was performed for codons of residues Asn82 and Cys2 using a Q5 Site-Directed Mutagenesis kit (New England BioLabs) with mutations confirmed via Sanger sequencing. To compare MCBA production profiles, cultures of each strain were first grown overnight in LB and then diluted to OD<sub>600</sub> = 0.01 in LB with a final concentration of 1 mM CA/TCA/GCA or 1% DMSO, 1 mg mL<sup>-1</sup> arabinose, 100  $\mu$ M taurine, and 20  $\mu$ g mL<sup>-1</sup> chloramphenicol. All cultures were incubated aerobically for 24 h at 37 °C with 220 rpm shaking. The OD<sub>600</sub> was measured and metabolite extraction was performed by diluting whole cultures 2:3 (v:v) in 100% ice-cold methanol in 1.7 mL microcentrifuge tubes (Axygen) followed by overnight incubation at 4 °C. Extracts were then centrifuged at 10,000  $g$  for 5 min to pellet cell debris followed by storage at -80 °C prior to LC-MS/MS analysis.

### 3.7 - Data availability

Raw mass spectrometry data are publicly available in the MassIVE database ([massive.ucsd.edu](http://massive.ucsd.edu)) for the *in vitro* screen for MCBA production under [MSV000090234](https://doi.org/10.25345/C5S756Q1B) (<https://doi.org/10.25345/C5S756Q1B>) and for CpBSH/T variant analysis under [MSV000092138](https://doi.org/10.25345/C55D8NQ9V) (<https://doi.org/10.25345/C55D8NQ9V>).

GNPS molecular networks are available for the MCBA production screen at [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=565151309a874d5f97caa3f383c95382](https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=565151309a874d5f97caa3f383c95382) and for the CpBSH/T mutagenesis experiment at [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=30c88ca297a44f84be5fa32b376e5cb9](https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=30c88ca297a44f84be5fa32b376e5cb9).

## REFERENCES

1. Guzior DV, Quinn RA. 2021. Review: microbial transformations of human bile acids. *Microbiome* 9:140.
2. Sorg JA, Sonenshein AL. 2008. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J Bacteriol* 190:2505–2512.
3. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, Chatfield DA, Hagey LR, Steinbach JH, Hofmann AF. 2007. Human cecal bile acids: concentration and spectrum. *Am J Physiol-Gastrointest Liver Physiol* 293:G256–G263.
4. Usui Y, Ayibieke A, Kamiichi Y, Okugawa S, Moriya K, Tohda S, Saito R. 2020. Impact of deoxycholate on *Clostridioides difficile* growth, toxin production, and sporulation. *Heliyon* 6:e03717.
5. Sannasiddappa TH, Lund PA, Clarke SR. 2017. *In vitro* antibacterial activity of unconjugated and conjugated bile salts on *Staphylococcus aureus*. *Front Microbiol* 8:1581.
6. Kang JD, Myers CJ, Harris SC, Kakiyama G, Lee I-K, Yun B-S, Matsuzaki K, Furukawa M, Min H-K, Bajaj JS, Zhou H, Hylemon PB. 2019. Bile acid 7 $\alpha$ -dehydroxylating gut bacteria secrete antibiotics that inhibit *Clostridium difficile*: role of secondary bile acids. *Cell Chem Biol* 26:27-34.e4.
7. Gustafsson BE, Midtvedt T, Norman A. 1966. Isolated fecal microorganisms capable of 7-alpha-dehydroxylating bile acids. *J Exp Med* 123:413–432.
8. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. 2020. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582:566–570.
9. Song Z, Cai Y, Lao X, Wang X, Lin X, Cui Y, Kalavagunta PK, Liao J, Jin L, Shang J, Li J. 2019. Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* 7:9.
10. Quinn RA, Melnik AV, Vrbanac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, Downes M, Welch RD, Quinn M, Humphrey G, Panitchpakdi M, Weldon KC, Aksenov A, da Silva R, Avila-Pacheco J, Clish C, Bae S, Mallick H, Franzosa EA, Lloyd-Price J, Bussell R, Thron T, Nelson AT, Wang M, Leszczynski E, Vargas F, Gauglitz JM, Meehan MJ, Gentry E, Arthur TD, Komor AC, Poulsen O, Boland BS, Chang JT, Sandborn WJ, Lim M, Garg N, Lumeng JC, Xavier RJ, Kazmierczak BI, Jain R, Egan M, Rhee KE, Ferguson D, Raffatellu M, Vlamakis H, Haddad GG, Siegel D, Huttenhower C, Mazmanian SK, Evans RM, Nizet V, Knight R, Dorrestein PC. 2020. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* 579:123–129.

11. Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB. 2016. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* 7:22–39.
12. Ridlon JM, Kang DJ, Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* 47:241–259.
13. Mullaney MW, Fiebig A, Schnizlein MK, McMillin M, Rose AR, Koval J, Rubin D, Dalal S, Sogin ML, Chang EB, Sidebottom AM, Crosson S. 2024. Microbially catalyzed conjugation of GABA and tyramine to bile acids. *J Bacteriol* 206:e00426-23.
14. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2005. Conjugated bile acid hydrolase is a tetrameric N-terminal thiol hydrolase with specific recognition of its cholyl but not of its tauryl product. *Biochemistry* 44:5739–5748.
15. Rossocha M, Schultz-Heienbrok R, Von Moeller H, Coleman JP, Saenger W. 2011. Crystal structure of conjugated bile acid hydrolase from *Clostridium perfringens* in complex with reaction products taurine and deoxycholate (1.2). 2bjf. pdb. RCSB PDB.
16. Foley MH, Allen G, Rivera AJ, Stewart AK, Barrangou R, Theriot CM. 2021. *Lactobacillus* bile salt hydrolase substrate specificity governs bacterial fitness and host colonization <https://doi.org/10.1073/pnas.2017709118/-/DCSupplemental>.
17. Karlov DS, Long SL, Zeng X, Xu F, Lal K, Cao L, Hayoun K, Lin J, Joyce SA, Tikhonova IG. 2023. Structure of *Lactobacillus salivarius* (Ls) bile salt hydrolase(BSH) in complex with taurocholate (TCA) (1.2). 8blt. pdb. RCSB PDB.
18. Karlov DS, Long SL, Zeng X, Xu F, Lal K, Cao L, Hayoun K, Lin J, Joyce SA, Tikhonova IG. 2023. Characterization of the mechanism of bile salt hydrolase substrate specificity by experimental and computational analyses. *Structure* 31:629-638.e5.
19. Lucas LN, Barrett K, Kerby RL, Zhang Q, Cattaneo LE, Stevenson D, Rey FE, Amador-Noguez D. 2021. Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. *mSystems* 6:e00805-21.
20. Petroff OAC. 2002. Book Review: GABA and Glutamate in the Human Brain. *The Neuroscientist* 8:562–573.
21. Marcobal A, De las Rivas B, Landete JM, Tabera L, Muñoz R. 2012. Tyramine and phenylethylamine biosynthesis by food bacteria. *Crit Rev Food Sci Nutr* 52:448–467.
22. Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci* 99:3695–3700.

23. Tekaiia F, Yeramian E, Dujon B. 2002. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. *Gene* 297:51–60.
24. Okayasu T, Ikeda M, Akimoto K, Sorimachi K. 1997. The amino acid composition of mammalian and bacterial cells. *Amino Acids* 13:379–391.
25. Dong Z, Lee BH. 2018. Bile salt hydrolases: Structure and function, substrate preference, and inhibitor development. *Protein Sci Publ Protein Soc* 27:1742–1754.
26. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650.
27. R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria.
28. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36.
29. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23:1073–1079.
30. Lodola A, Branduardi D, De Vivo M, Capoferri L, Mor M, Piomelli D, Cavalli A. 2012. A catalytic mechanism for cysteine N-terminal nucleophile hydrolases, as revealed by free energy simulations. *PLoS ONE* 7:e32397.
31. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. 2021. Mol\*Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* 49:W431–W437.
32. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
33. Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 18:366–368.
34. Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard P-M, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin H. C, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M,

- Dorrestein PC. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* 17:905–908.
35. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kaponov CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya CAP, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Lington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837.
  36. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.
  37. Lensmire JM, Wischer MR, Kraemer-Zimpel C, Kies PJ, Sosinski L, Ensink E, Dodson JP, Shook JC, Delekta PC, Cooper CC, Havlichek DH, Mulks MH, Lunt SY, Ravi J, Hammer ND. 2023. The glutathione import system satisfies the *Staphylococcus aureus* nutrient sulfur requirement and promotes interspecies competition. *PLOS Genet* 19:e1010834.

## APPENDIX A: SUPPLEMENTARY TABLES

**Table 3.1: Strains used in this work**

Included are taxonomic family, genome sequence accession number for species used in phylogenetic tree construction, BSH/T amino acid sequence accession number for *bsh+* species, in addition to MCBA production group and BSH/T sequence group. All genome sequences were obtained from the National Institutes of Health. BSH/T sequences were obtained from Joint Genome Institute Integrated Microbial Genomes and Microbiomes system (JGI IMG/M) or National Center for Biotechnology Information (NCBI) databases. Sequences from JGI IMG/M start with 'IMG#' followed by the unique gene ID. American Type Culture Collection, ATCC; Culture Collection, University of Gothenburg, CCUG; NP, no production.

Species	Strain	Source	Taxonomic Family	MCBA Profile Cluster	Genome Accession #	BSH/T Protein Accession#	BSH/T Group
<i>Phocaeicola vulgatus</i>	8482	ATCC	<i>Bacteroidaceae</i>	NP	GCA_020885855		
<i>Blautia coccooides</i>	29236	ATCC	<i>Lachnospiraceae</i>	NP	GCA_900461125		
<i>Clostridium perfringens</i>	13124	ATCC	<i>Clostridiaceae</i>	1	GCA_000013285	WP_243289361.1 IMG#2539151749	I I
<i>Clostridium sporogenes</i>	15579	ATCC	<i>Clostridiaceae</i>	NP	GCA_000155085	WP_219649723.1 IMG#642849186	I III
<i>Lachnoclostridium scindens</i>	35704	ATCC	<i>Lachnospiraceae</i>	3	GCA_000154505		
<i>Clostridium symbiosum</i>	14940	ATCC	<i>Lachnospiraceae</i>	4	GCA_000466485	WP_092364887.1 IMG#2598983548 IMG#2598979318	I III III
<i>Enterocloster aldenensis</i>	BAA-1318	ATCC	<i>Lachnospiraceae</i>	2	GCA_003467385	WP_165640901.1 IMG#2855314028 IMG#2855308917 IMG#2855309434	II II III III
<i>Enterocloster bolteae</i>	BAA-613	ATCC	<i>Lachnospiraceae</i>	3	GCA_002234575	WP_002565680.1 IMG#2825699883 IMG#2825699493 IMG#2825699020	I II III III

Table 3.1 (cont'd)

<i>Enterocloster clostridioformis</i>	25537	ATCC	<i>Lachnospiraceae</i>	<b>3</b>	GCA_900113155	WP_002587930.1 IMG#2914509376 IMG#2914512748 IMG#2914510687	I II III III
<i>Lacrimispora aerotolerans</i>	43524	ATCC	<i>Lachnospiraceae</i>	<b>NP</b>	GCA_000687555	IMG#2558963465	II
<i>Lacrimispora sphenoides</i>	19403	ATCC	<i>Lachnospiraceae</i>	<b>5</b>	GCA_900105615	IMG#2676059071 WP_092240532.1 IMG#2676055364	II II III
<i>Ruminococcus gnavus</i>	29149	ATCC	<i>Lachnospiraceae</i>	<b>1</b>	GCA_009831375	WP_173900661.1 IMG#2545645281	I I
<i>Anaerostipes caccae</i>	47493T	CCUG	<i>Lachnospiraceae</i>	<b>NP</b>	GCA_020181435		
<i>Akkermansia muciniphila</i>	64013T	CCUG	<i>Akkermansiaceae</i>	<b>NP</b>	GCA_017504145		
<i>Bacteroides fragilis</i>	4856T	CCUG	<i>Bacteroidaceae</i>	<b>2</b>	GCA_016889925	OCR29502.1	III
<i>Bifidobacterium bifidum</i>	45217T	CCUG	<i>Bifidobacteriaceae</i>	<b>1</b>	GCA_900637095	ABC26910.1	I
<i>Bifidobacterium longum</i> subsp. <i>infantis</i>	30512BT	CCUG	<i>Bifidobacteriaceae</i>	<b>1</b>	GCA_000196555	ACJ52536.1	I
<i>Blautia producta</i>	9990T	CCUG	<i>Lachnospiraceae</i>	<b>1</b>	GCA_010669205	QIB57271.1	I
<i>Clostridium hylemonae</i>	45367T	CCUG	<i>Lachnospiraceae</i>	<b>NP</b>	GCA_008281175		
<i>Clostridium novyi</i>	57219T	CCUG	<i>Clostridiaceae</i>	<b>NP</b>	GCA_003614235		
<i>Enterocloster citroniae</i>	52203T	CCUG	<i>Lachnospiraceae</i>	<b>2</b>	GCA_900115855	WP_007861063.1	II
<i>Enterocloster lavalensis</i>	54291T	CCUG	<i>Lachnospiraceae</i>	<b>4</b>	GCA_902364025	SET75766.1	II
<i>Hungatella hathewayi</i>	43506T	CCUG	<i>Clostridiaceae</i>	<b>3</b>	GCA_000160095	WP_138313413.1	II
<i>Lacrimispora celerecrescens</i>	68430T	CCUG	<i>Lachnospiraceae</i>	<b>2</b>	GCA_002797975	WP_100306104.1	II
<i>Lacrimispora indolis</i>	55582T	CCUG	<i>Lachnospiraceae</i>	<b>5</b>	GCA_900618075	WP_024291983.1	II
<i>Lactiplantibacillus plantarum</i>	30503T	CCUG	<i>Lactobacillaceae</i>	<b>1</b>	GCA_000615325	ACL98169.1 D7VDZ4_LACPN	III I

Table 3.1 (cont'd)

<i>Peptostreptococcus anaerobius</i>	7835T	CCUG	<i>Clostridiaceae</i>	<b>NP</b>	GCA_900454605		
<i>Enterococcus faecalis</i>	DF4	Oral Isolate	<i>Enterococcaceae</i>	<b>1</b>	GCA_900447895	IMG#2732210673	<b>I</b>
						EOL33562.1	<b>III</b>
<i>Escherichia coli</i>	DH5 $\alpha$	Dr. Robert Hausinger	<i>Enterobacteriaceae</i>	<b>NP</b>	GCA_019444045		
<i>Pseudomonas aeruginosa</i>	PA01	Dr. Christopher Waters	<i>Pseudomonadaceae</i>				
<i>Pseudomonas aeruginosa</i>	PA14	Dr. Christopher Waters	<i>Pseudomonadaceae</i>				
<i>Salmonella enterica</i> serovar <i>Typimurium</i>	14028S	Dr. Kristen Parent	<i>Enterobacteriaceae</i>				
<i>Staphylococcus aureus</i> (37)		Dr. Neal Hammer	<i>Staphylococcaceae</i>				
<i>Staphylococcus epidermidis</i> (37)		Dr. Neal Hammer	<i>Staphylococcaceae</i>				

**Table 3.2: Individual amino acid use in conjugation for strains within MCBA profile cluster 1**

Values represent mean auc  $\pm$  s.e.m.  $n = 3$  independent cultures; NO, not observed.

Amino Acid	<i>L. plantarum</i>	<i>R. gnavus</i>	<i>E. faecalis</i>	<i>B. bifidum</i>	<i>E. bolteae</i>	<i>B. producta</i>	<i>C. perfringens</i>
<b>Ala</b>	1.24E7 $\pm 2.04E5$	3.36E6 $\pm 9.71E4$	1.21E6 $\pm 1.11E5$	2.45E6 $\pm 3.86E5$	2.71E6 $\pm 2.06E5$	1.41E6 $\pm 1.47E5$	9.53E5 $\pm 6.2E5$
<b>Asn</b>	3.98E7 $\pm 2.17E6$	4.1E7 $\pm 1.06E6$	8.92E6 $\pm 5.04E5$	2.89E6 $\pm 3.52E5$	2.26E6 $\pm 2.45E5$	6.27E6 $\pm 8.59E5$	1.41E6 $\pm 1.15E6$
<b>Asp</b>	1.93E7 $\pm 3.34E5$	5.69E6 $\pm 2.59E5$	2.41E6 $\pm 3.89E4$	2.24E6 $\pm 3.45E5$	5.35E6 $\pm 3.51E5$	5.85E6 $\pm 6.66E5$	3.77E5 $\pm 3.08E5$
<b>Cit</b>	NO	7.82E5 $\pm 1.04E5$	1.05E7 $\pm 4.79E5$	2.15E6 $\pm 2.35E5$	9.7E5 $\pm 6.25E4$	1.12E6 $\pm 9.44E4$	1.62E6 $\pm 1.32E6$
<b>Cys</b>	4.1E6 $\pm 3.04E5$	1.56E6 $\pm 1.71E5$	2.75E5 $\pm 1.13E5$	2.67E5 $\pm 1.27E5$	1.09E5 $\pm 8.93E4$	7.13E5 $\pm 1.36E5$	1.63E5 $\pm 1.33E5$
<b>Gln</b>	2.77E7 $\pm 1.58E6$	4.71E7 $\pm 1.45E6$	9.25E6 $\pm 5.72E5$	1.33E7 $\pm 1.84E6$	3.61E6 $\pm 2.34E5$	1.28E7 $\pm 1.36E6$	NO
<b>Glu</b>	5.74E7 $\pm 3.55E6$	4.8E6 $\pm 6.97E5$	9.65E6 $\pm 4.42E5$	1.79E7 $\pm 2.54E6$	2.04E6 $\pm 1.6E5$	3.15E6 $\pm 1.08E6$	1.72E5 $\pm 1.4E4$
<b>Gly</b>	7.02E6 $\pm 3.28E5$	4.92E6 $\pm 2.24E5$	1.66E6 $\pm 1.52E5$	2.58E6 $\pm 2.62E5$	1.24E7 $\pm 6.18E5$	1.8E6 $\pm 6.43E4$	1.85E6 $\pm 2.73E5$
<b>His</b>	1.42E7 $\pm 1.43E6$	2.04E7 $\pm 2.2E6$	2.69E7 $\pm 1.44E6$	NO	5.18E5 $\pm 2.25E5$	NO	1.21E7 $\pm 9.85E6$
<b>Ile/Leu</b>	2.69E6 $\pm 1.48E6$	7.54E5 $\pm 2.23E4$	2.25E6 $\pm 8.1E5$	1.15E6 $\pm 1.37E5$	9.38E5 $\pm 2.69E4$	6.06E5 $\pm 9.15E4$	5.79E6 $\pm 4.1E6$
<b>Lys</b>	4.32E6 $\pm 2.48E5$	5.54E6 $\pm 1.24E5$	2.14E7 $\pm 2.57E6$	NO	2.76E6 $\pm 8.72E5$	6.31E6 $\pm 1.2E6$	8.79E6 $\pm 6.66E6$
<b>Phe</b>	1.03E7 $\pm 2.95E6$	6.91E6 $\pm 8.7E5$	8.83E6 $\pm 5.5E5$	2.8E6 $\pm 8.56E5$	5.9E6 $\pm 5.26E5$	2.09E6 $\pm 2.06E5$	6.52E6 $\pm 4.1E6$
<b>Ser</b>	6.67E6 $\pm 2.23E5$	5.12E6 $\pm 1.91E5$	3.96E6 $\pm 2.99E5$	4.22E6 $\pm 3.85E5$	4.E6 $\pm 2.14E5$	5.27E6 $\pm 3.28E5$	2.15E6 $\pm 6.71E5$
<b>Taur</b>	NO	NO	1.89E5 $\pm 1.54E5$	1.15E6 $\pm 1.27E5$	1.05E6 $\pm 5.79E5$	3.19E5 $\pm 2.6E5$	9.34E5 $\pm 1.18E5$
<b>Thr</b>	5.91E6 $\pm 2.8E6$	3.13E6 $\pm 2.36E5$	2.07E6 $\pm 6.45E5$	1.87E6 $\pm 5.31E5$	6.47E5 $\pm 1.47E5$	8.92E5 $\pm 2.31E5$	3.02E6 $\pm 2.22E6$
<b>Trp</b>	2.39E6 $\pm 5.43E5$	8.74E5 $\pm 2.06E4$	3.41E6 $\pm 9.2E4$	NO	1.09E6 $\pm 2.98E4$	2.1E6 $\pm 1.67E5$	2.94E6 $\pm 2.2E6$
<b>Tyr</b>	1.48E6 $\pm 1.05E5$	5.72E6 $\pm 1.21E6$	5.91E6 $\pm 9.45E5$	1.42E6 $\pm 1.67E5$	3.86E6 $\pm 1.7E6$	2.86E6 $\pm 6.35E5$	6.00E6 $\pm 4.7E6$

**Table 3.3: Individual amino acid use in conjugation for strains within MCBA profile cluster 2**

Values represent mean auc  $\pm$  s.e.m.  $n = 3$  independent cultures; NO, not observed.

<b>Amino Acid</b>	<i>B. fragilis</i>	<i>L. scindens</i>	<i>E. clostridioformis</i>	<i>H. hathewayi</i>
<b>Ala</b>	NO	6.36E6 $\pm$ 1.84E6	1.97E6 $\pm$ 1.11E5	2.16E6 $\pm$ 1.4E5
<b>Asn</b>	NO	NO	NO	NO
<b>Asp</b>	NO	2.5E6 $\pm$ 5.11E5	NO	NO
<b>Cit</b>	NO	NO	NO	NO
<b>Cys</b>	NO	6.56E5 $\pm$ 1.3E5	5.18E5 $\pm$ 9.51E4	3.35E5 $\pm$ 1.37E5
<b>Gln</b>	NO	NO	NO	NO
<b>Glu</b>	NO	3.15E6 $\pm$ 5.84E5	NO	NO
<b>Gly</b>	7.22E5 $\pm$ 7.64E4	6.09E6 $\pm$ 9.69E5	3.12E7 $\pm$ 4.23E5	9.62E6 $\pm$ 6.42E5
<b>His</b>	NO	NO	NO	NO
<b>Ile/Leu</b>	NO	3.93E6 $\pm$ 1.15E6	7.81E5 $\pm$ 1.45E4	5.14E5 $\pm$ 3.52E4
<b>Lys</b>	NO	NO	8.06E5 $\pm$ 4.12E4	NO
<b>Phe</b>	NO	1.02E6 $\pm$ 1.64E5	NO	NO
<b>Ser</b>	NO	NO	NO	NO
<b>Taur</b>	NO	NO	NO	2.68E6 $\pm$ 3.47E5
<b>Thr</b>	NO	NO	NO	NO
<b>Trp</b>	NO	NO	NO	NO
<b>Tyr</b>	NO	NO	NO	NO

**Table 3.4: Individual amino acid use in conjugation for strains within MCBA profile cluster 3**

Values represent mean auc  $\pm$  s.e.m.  $n = 3$  independent cultures; NO, not observed.

Amino Acid	<i>B. infantis</i>	<i>L. indolis</i>	<i>L. sphenoides</i>
Ala	1.01E6 $\pm$ 7.66E4	NO	NO
Asn	1.56E6 $\pm$ 3.39E5	NO	NO
Asp	NO	NO	NO
Cit	NO	NO	NO
Cys	NO	NO	NO
Gln	4.27E5 $\pm$ 3.48E5	NO	NO
Glu	9.22E5 $\pm$ 2.99E5	NO	NO
Gly	1.78E6 $\pm$ 1.09E5	3.81E4 $\pm$ 3.11E4	3.75E4 $\pm$ 3.06E4
His	NO	NO	NO
Ile/Leu	NO	NO	NO
Lys	NO	NO	NO
Phe	4.22E5 $\pm$ 7.93E4	NO	NO
Ser	3.9E6 $\pm$ 3.73E5	2.71E6 $\pm$ 7.25E5	3.18E6 $\pm$ 2.08E5
Taur	8.85E5 $\pm$ 4.65E4	NO	NO
Thr	1.43E6 $\pm$ 6.43E5	1.03E6 $\pm$ 3.97E5	8.82E5 $\pm$ 1.83E5
Trp	NO	NO	NO
Tyr	NO	NO	NO

**Table 3.5: Individual amino acid use in conjugation for strains within MCBA profile cluster 4**

Values represent mean auc  $\pm$  s.e.m.  $n = 3$  independent cultures; NO, not observed.

<b>Amino Acid</b>	<i>C. symbiosum</i>	<i>E. lavalensis</i>
<b>Ala</b>	NO	NO
<b>Asn</b>	6.86E5 $\pm$ 1.96E4	NO
<b>Asp</b>	NO	NO
<b>Cit</b>	NO	NO
<b>Cys</b>	NO	NO
<b>Gln</b>	NO	NO
<b>Glu</b>	NO	NO
<b>Gly</b>	NO	NO
<b>His</b>	4.18E6 $\pm$ 2.83E5	NO
<b>Ile/Leu</b>	NO	NO
<b>Lys</b>	1.12E7 $\pm$ 5.82E5	4.72E6 $\pm$ 3.53E5
<b>Phe</b>	NO	NO
<b>Ser</b>	NO	NO
<b>Taur</b>	2.82E6 $\pm$ 8.9E4	1.4E6 $\pm$ 3.95E5
<b>Thr</b>	NO	NO
<b>Trp</b>	NO	NO
<b>Tyr</b>	NO	NO

**Table 3.6: Individual amino acid use in conjugation for strains within MCBA profile cluster 5**

Values represent mean auc  $\pm$  s.e.m.  $n = 3$  independent cultures; NO, not observed.

<b>Amino Acid</b>	<i>E. aldenensis</i>	<i>E. citroniae</i>	<i>L. celerecrescens</i>
<b>Ala</b>	NO	NO	NO
<b>Asn</b>	NO	NO	NO
<b>Asp</b>	1.93E6 $\pm$ 8.89E4	1.27E6 $\pm$ 1.36E5	2.46E6 $\pm$ 1.99E5
<b>Cit</b>	NO	NO	NO
<b>Cys</b>	NO	NO	NO
<b>Gln</b>	NO	NO	NO
<b>Glu</b>	NO	NO	NO
<b>Gly</b>	1.83E5 $\pm$ 3.83E3	1.74E5 $\pm$ 2.73E4	NO
<b>His</b>	NO	NO	NO
<b>Ile/Leu</b>	NO	NO	NO
<b>Lys</b>	NO	NO	NO
<b>Phe</b>	NO	NO	NO
<b>Ser</b>	NO	NO	NO
<b>Taur</b>	1.06E6 $\pm$ 7.53E4	3.95E5 $\pm$ 3.23E5	NO
<b>Thr</b>	NO	NO	NO
<b>Trp</b>	NO	NO	NO
<b>Tyr</b>	NO	NO	NO

**Table 3.7: Publicly available genome sequences for *Lachoclostridium scindens* used in phylogenetic analysis and BSH/T prediction**

Strains ATCC 35704 and Q4 are bolded. ATCC 35704 produced MCBA's when grown in 1 mM CA and strain Q4 was the only strain predicted to encode a BSH/T based on Prokka analysis.

Strain	Assembly Accession #	Size (Mb)
<b>ATCC 35704-1</b>	<b>GCA_004295125.1</b>	<b>3.65804</b>
VPI12708	GCA_027941655.1	3.98305
<b>Q4</b>	<b>GCA_019597925.1</b>	<b>3.94184</b>
G10	GCA_020892115.1	3.31559
BL389WT3D	GCA_009684695.1	3.78553
FDAARGOS_1227	GCA_016889005.1	3.6191
CE91-St60	GCA_022845835.1	3.60809
CE91-St59	GCA_022845815.1	3.60808
<b>ATCC 35704-2</b>	<b>GCA_000154505.1</b>	<b>3.62261</b>
MSK.5.24	GCA_013304085.1	4.07271
NB2A-7-D5	GCA_024125195.1	4.1826
SL.1.22	GCA_020555615.1	3.97009
DFI.1.234	GCA_022137935.1	4.31663
MGYG-HGUT-01303	GCA_902373645.1	3.62261
AM05-22	GCA_027662895.1	3.33015
GGCC_0168	GCA_017565985.1	3.41709
DFI.1.217	GCA_020562885.1	4.38987
DFI.1.162	GCA_020563365.1	4.39631
DFI.1.161	GCA_024463895.1	4.32554
MSK.1.26	GCA_013304105.1	3.22797
DFI.1.60	GCA_020561885.1	4.30944
AM07-30	GCA_027662765.1	3.33167
MSK.1.16	GCA_013304115.1	3.2301
DFI.1.130	GCA_020563525.1	4.56586
BL-389-WT-3D	GCA_009696415.1	3.61438
DFI.4.63	GCA_020560435.1	4.16737
SUG670	GCA_022777065.1	2.82745
ERR1600561_bin.107_CONCOCT_v1.1_MAG	GCA_938001855.1	2.88328
ERR1606358_bin.2_metaWRAP_v1.3_MAG	GCA_945908315.1	2.946
SRR5240736_bin.6_metaWRAP_v1.3_MAG	GCA_945830785.1	3.09947
ERR1855542_bin.22_metaWRAP_v1.3_MAG	GCA_945875235.1	3.1274
ERR1600561-bin.52	GCA_905206435.1	2.98552
SRR17382097_bin.48_metaWRAP_v1.3_MAG	GCA_945871535.1	2.87824
W0P25.025	GCA_004558675.1	2.88522
VE202-05	GCA_000471845.1	3.91239

**Table 3.8: Annotated MCBAs used for peak integration in mutagenesis studies**

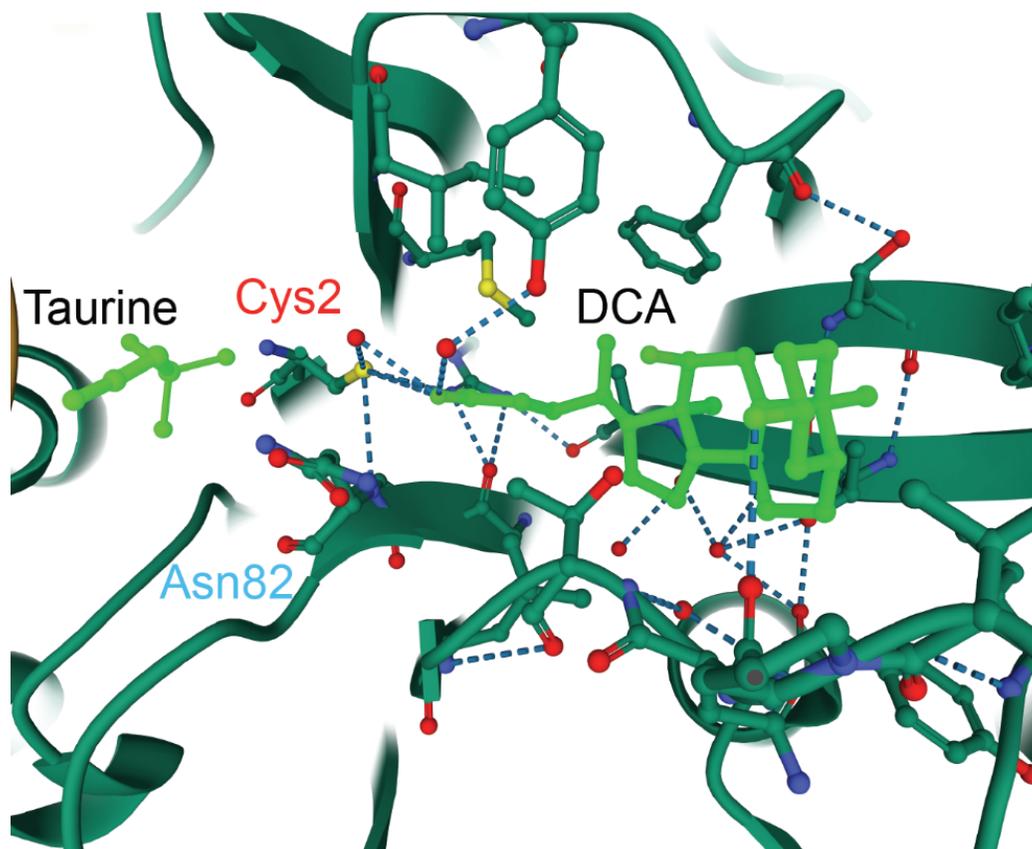
<b>Bile Acid</b>	<b>Abbr.</b>	<b><i>m/z</i></b>	<b>RT, min</b>
Alanochoic acid	AlaCA	480.3316	5.87
Arginochoic acid	ArgCA	565.3957	5.11
Glutamatochoic acid	GluCA	538.3378	5.53
Glycochoic acid	GCA	448.3054	5.65
Histidochoic acid	HisCA	546.3538	5.04
Iso/Leucochoic acid	Ile/LeuCA	522.3787	6.36
Lysochoic acid	LysCA	537.3898	5.15
Phenylalanochoic acid	PheCA	556.3628	6.40
Taurochoic acid	TCA	516.2987	5.33
Cholic acid	CA	426.3212	6.12

**Table 3.9: Primers used in for *C. perfringens bsh/t* cloning and mutagenesis experiments**

Bases matching codons changed for site-directed mutagenesis experiments are underlined. For amplification of *C. perfringens bsh/t*, base overhangs homologous with linearized pBAD18-Cm are highlighted in bold.

Primer	Sequence (5'-3')
pBAD-FWD	TCTAGAGTCGACCTGCAG
pBAD-REV	CGAGCTCGAATTCGCTAG
pBAD-Screen-FWD	GGCGTCACACTTTGCTATGCCATAGC
pBAD-Screen-REV	CTACGGCGTTTCACTTCTGAGTTCGGC
CpBST-N82Y-FWD	TGCTGGCTTAT <u>ATTT</u> CCCTGTTTATG
CpBST-N82Y-REV	CATCCTAATCCCTTTTCATTC
CpBST-C2A-FWD	AGTTTTTATGGCTACAGGATTAGCCTTAGAAACAAAAG
CpBST-C2A-REV	CACTCCTCGAGCTCGAAT
CpBST-BAD-FWD	CGTTTTTTTGGGCTAGCGAATTCGAGCTCGAGGAGTGAGTTTTTATGTGTA CAGG
CpBST-BAD-REV	AAGCTTGCATGCCTGCAGGTCGACTCTAGACCCATGCAACAACTAATTTA CATG

## APPENDIX B: SUPPLEMENTARY FIGURES



**Figure 3.8: Crystal structure of *C. perfringens* BSH/T with co-crystallized taurine and DCA**

Publicly available structure of *C. perfringens* BSH/T (PDB ID: 2bjg) (14, 15) co-crystallized with taurine and DCA, products of incubation with TDCA. Residues important for BA deconjugation are highlighted in addition to Asn82, the residue playing a key role in specificity of microbial BA conjugation.

**CHAPTER 4:**  
**INTERPLAY BETWEEN MICROBIALLY CONJUGATED BILE ACIDS, THE  
MICROBIOME, AND THE METABOLOME**

## 4.1 - Preface

Some contents of this chapter were published in the journal *Nature* in 2024 (Material from: Guzior, D.V., Okros, M., Shivel, M. *et al.* Bile salt hydrolase acyltransferase activity expands bile acid diversity. *Nature* **626**, 852–858 (2024). <https://doi.org/10.1038/s41586-024-07017-8>). Per the publisher, Springer Nature, “Authors have the right to reuse their article’s Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.” Maxwell Okros, Madison Shivel, and Bruin Armwald conducted work with mice . Dr. Wendy M. Miller, Dr. Kathryn M. Ziegler, Dr. Matthew D. Sims, Dr. Michael E. Maddens, and Dr. Stewart F. Graham coordinated sample collection, treated patients, performed bariatric surgeries, completed all clinical follow-up for work presented related to the sleeve gastrectomy patient cohort.

Additional contents of this chapter were in review when submitting this dissertation in a manuscript titled “A novel multi-omics analysis approach for population and subject-specific microbiome-metabolome trajectories” by authors Guzior, D.V., Wu H., Martin, C., Neugebauer, K.A., Rzepka, M.M., Lumeng, J.C., Quinn, R.A., and de los Campos, G. where Hao Wu and I contributed to this work equally as co-first authors. Hao Wu from the de los Campos lab developed and validated the computational modeling methods using random regression included here (method sections 4.6.11-15) and contributed Figure 4.12a in addition to Figures 4.3 and 4.4. Dr. Christian Martin aided in sample metabolite extraction and preparation for mass spectrometry analysis. Dr. Kerri A. Neugebauer and Madison R. Rzepka aided in DNA extraction and quality control prior to submission for

16S amplicon sequencing. Dr. Julie M. Lumeng secured funding for the project from the National Institutes of Health (Grant R01HD084163) and provided access to collected fecal swabs.

## 4.2 - Abstract

While extensive research has examined the impacts of secondary bile acids (BAs) on the host and their microbiome, little is known about the roles played by microbially conjugated bile acids. It has been well established that BAs are key drivers of intestinal microbiome structure, given their potent antimicrobial and signaling effects. Following up on previously performed screens for microbial BA conjugation, 12 anaerobic and 6 aerobic bacteria were grown in the presence of eight different microbially conjugated bile acids (MCBAs) with varying biochemical properties based on the amino acid ligand. Hydrophilic conjugates were less antimicrobial compared to free cholic acid. The hydrophobic conjugates phenylalanochoic acid, leucochoic acid, and tyrosochoic acid exhibited more potent antimicrobial effects, particularly against *Lactiplantibacillus plantarum* and *Peptostreptococcus anaerobius*. Investigating these effects using an *in vivo* mouse model revealed that high doses of MCBAs resulted in shifted microbiome structures. Reducing the concentration of MCBAs resulted in a lower magnitude shift but revealed the capability for MCBAs to enter enterohepatic circulation. We then investigated how these MCBAs related to important changes in the human gut microbiome, including those that occur during a surgical intervention for an obese disease state and the dynamic changes during the first year of human life. MCBA concentrations significantly decreased before and after sleeve gastrectomy surgery, whereas concentrations of primary and secondary BAs did not. Initial investigation into overall microbiome shifts matched existing dogma; as infants mature, bacterial diversity within the gut increases. However, we observed that increased microbial diversity correlates with decreased metabolite diversity, with both results being driven by sample richness. Further analysis revealed maternal

health and infant diet as key factors shaping both the metabolome and microbiome community structure. This decrease in metabolite richness extends to the BA pool where we observe that MCBAs become less prevalent as infants mature. Similarly, glucuronidated BAs decreased in abundance as infants matured, a potential indicator of proper gastrointestinal maturation through decreases in toxic BA concentration. Together, these results are a key first description of the role of MCBAs across a broad scale, from single-celled organisms to murine models to human cohorts.

### 4.3 - Introduction

Our intestines host a dense, diverse, and dynamic microbial community that plays an essential role in human physiology. One way we shape and drive the structure of this community is through antimicrobial and signaling properties of bile acids (BAs). Products of host and microbial BA metabolism are known to have highly variable effects on the resident gut microbiota, dependent on both bacterial species and BA biochemical properties. Conjugation with glycine or taurine, both small and polar molecules, increases BA hydrophilicity and is then reflected in greater microbial BA tolerance *in vitro* compared to free BAs (1, 2).

Dynamic changes in the gastrointestinal microbiome are known to occur in concert with shifts within the human BA pool. Use of antibiotics has been shown to promote *Clostridioides difficile* colonization by reducing concentrations of otherwise inhibitory secondary BAs, a consequence of the loss of secondary BA-producing bacteria (3). Diet and lifestyle are also known to impact the gastrointestinal microbiome and metabolome (4–7). High-fat diets, often used as a proxy for diets of the Western world, have been shown to enrich for BA deconjugating and dehydroxylating bacteria resulting in higher serum concentrations of deoxycholic acid (DCA) and other secondary BAs (8). Diet is also known to play a role in infant gastrointestinal microbiome development in concert with other early-life exposures, such as antibiotics use, household pets, and delivery mode (9–12).

Both primary and secondary BAs function as key signaling molecules, regulating nutrient uptake and gut homeostasis. Ursodeoxycholic acid (UDCA), an epimer of CDCA, and LCA exhibit anti-inflammatory effects by inhibiting the release of pro-inflammatory

cytokines TNF $\alpha$ , IL-1 $\beta$ , and IL-6 (13). Secondary BAs are also known to be taken up by the host, entering enterohepatic circulation and being subsequently conjugated in the liver with glycine or taurine (14). Understanding compositional shifts in the human BA pool is an active area of biomarker research for predicting onset and severity of multiple diseases. For decades, BAs have been used as markers of gastrointestinal disease severity, notably for irritable bowel syndrome, gallstones (15), liver damage & diseases (16–20), and gut-related cancers (21). Implications of BAs in carcinogenesis, specifically DCA, was first reported by Cook, Kennaway, and Kennaway in 1940 (22). Recent work has begun to illustrate their roles as biomarkers of Alzheimer’s disease (23–25), heart diseases (26), diabetes (27, 28), and lung inflammation (29) among other disorders with primary manifestations outside the gastrointestinal system.

While both primary and secondary BAs continue to be investigated at length for their roles in host pathology, the impacts of microbially conjugated bile acids (MCBAs) remain largely unknown. Here, we describe MCBA dynamics within the gut across several scales. At the microbial level, we investigated how individual MCBAs impact bacterial growth *in vitro*. We then utilized murine models to determine if these effects translated *in vivo*, followed by using two human cohorts known to be experiencing significant changes in their gut microbiome to investigate changes within humans.

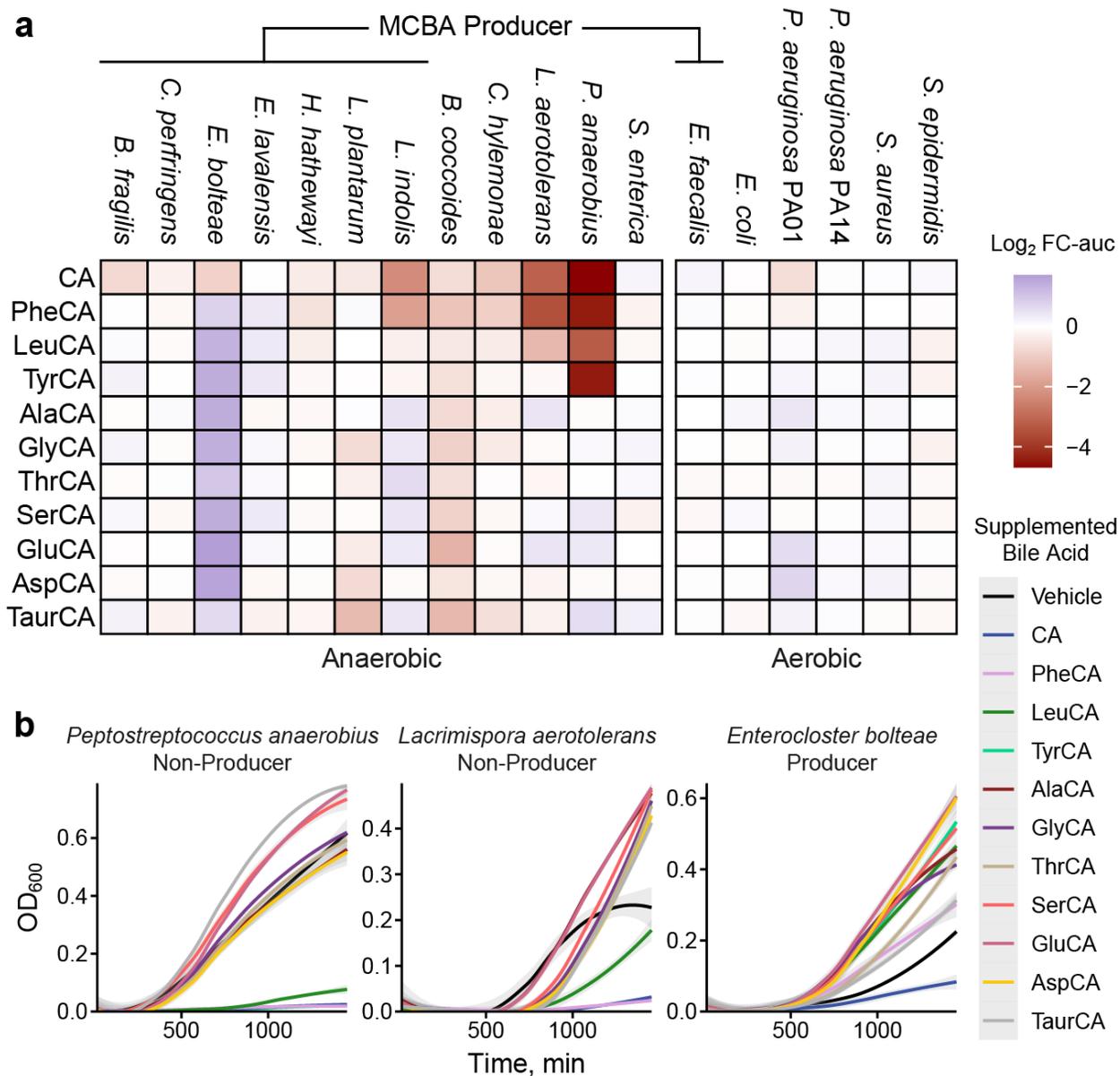
## **4.4 - Results**

### *4.4.1 - Antimicrobial efficacy of MCBAs*

Free BAs are known to exert antimicrobial activity by damaging cell membranes and chromosomal DNA (2), a mechanism not limited to bacterial cells. This antimicrobial activity is a well-known property of secondary BAs (21, 30–32), whereas conjugated

primary BAs are less antimicrobial (2, 33). We therefore suggested that microbial BA conjugation may be a means for bacteria to modulate BA toxicity. To test this hypothesis, we first determined the effects of medium supplementation with 1 mM CA or individual MCBA on *Enterocloster bolteae*, the first organism identified to produce MCBA. *E. bolteae* showed increased growth in the presence of any MCBA but growth with CA showed a slight detriment (Figure 4.1a). We chose to use 1 mM CA as it is known to be inhibitory against most BA-susceptible bacteria and represents the higher range of native BA concentrations in the human gastrointestinal (GI) tract (34, 35). Impacts of MCBA administration on further species showed variable antimicrobial efficacy. The most marked reductions in growth were observed for *Clostridium hylemonae*, *Blautia coccoides*, *Peptostreptococcus anaerobius*, *Lacrimispora aerotolerans*, and *Lacrimispora indolis*, of which only *L. indolis* produced MCBA (Figure 4.1a). *P. anaerobius* and *L. aerotolerans* showed the most marked deficit as growth was significantly reduced if not completely inhibited (Figure 4.1a,b).

Antimicrobial efficacy depended on the amino acid conjugated, where hydrophobic conjugates showed the strongest effects, particularly PheCA and LeuCA (Figure 4.1). Importantly, these effects were not observed for host conjugates GCA or TCA, indicating that microbial conjugation with these non-canonical amino acids can increase BA toxicity. *L. aerotolerans* showed growth defects when grown in LeuCA, with an effective dose nearly half that of CA (236 versus 425  $\mu\text{M}$ ; Figure 4.2a,d) but showed slightly increased resistance to PheCA ( $\text{ED}_{50} = 460 \mu\text{M}$ ; Figure 4.2) compared to CA. However, PheCA effectively inhibited *P. anaerobius* growth at nearly two-thirds the concentration of CA



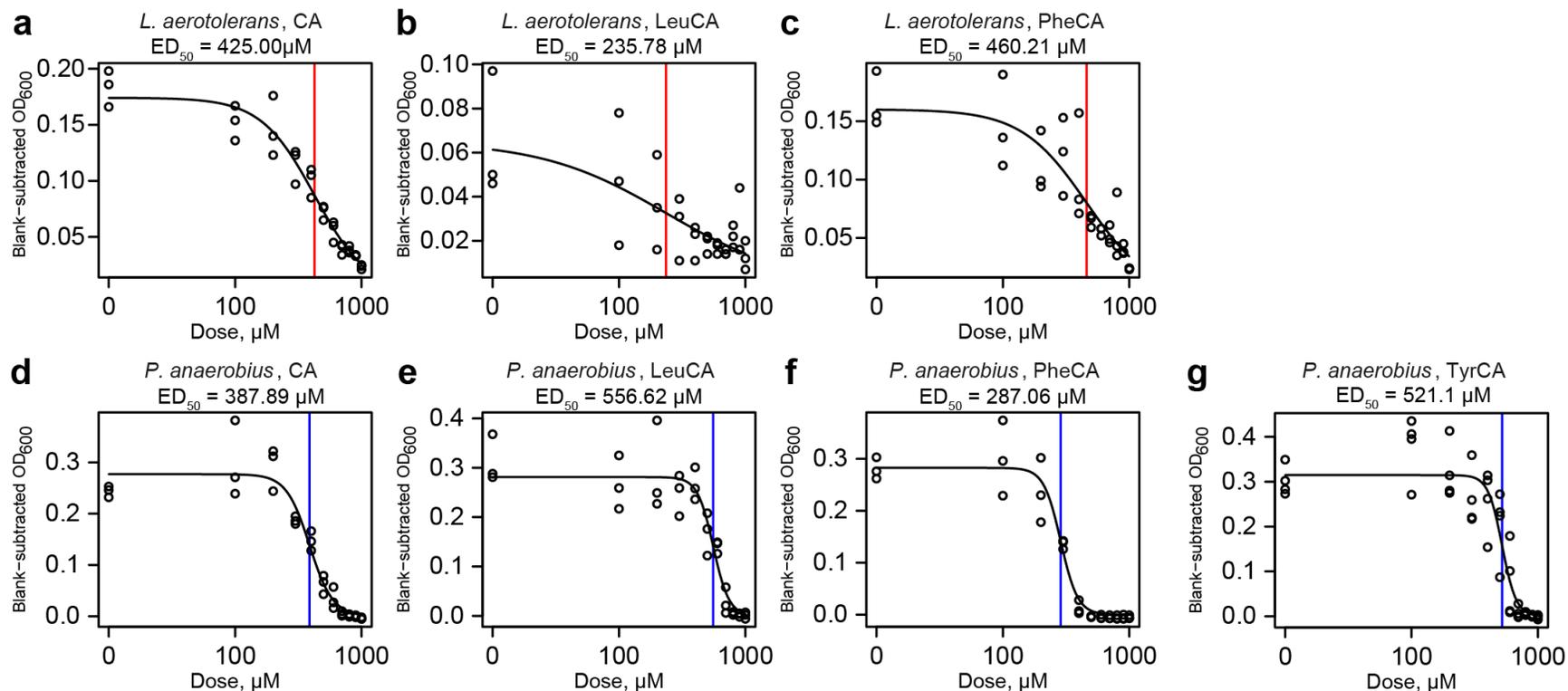
**Figure 4.1: MCBAs show varied antimicrobial properties**

**a**, Average  $\log_2$ -fold-change in area under the growth curve (FC-AUC) between BA-treated cultures and control. **b**, Representative growth curves for *P. anaerobius* and *L. aerotolerans*, species showing growth detriments in the presence of 1 mM CA and CA conjugated with hydrophobic amino acids, in addition to *E. bolteae*, demonstrating slight increases in FC-AUC for all MCBAs administered. Growth curve data are presented as smoothed mean  $OD_{600}$  with the 95% confidence interval shaded behind the line.  $n = 3$  independent cultures for anaerobic growth and 4 independent cultures for aerobic growth.

(ED<sub>50</sub> = 287 versus 388 μM; Figure 4.2) whereas LeuCA and TyrCA were half as effective (ED<sub>50</sub> = 557 and 521 μM, respectively; Figure 4.2).

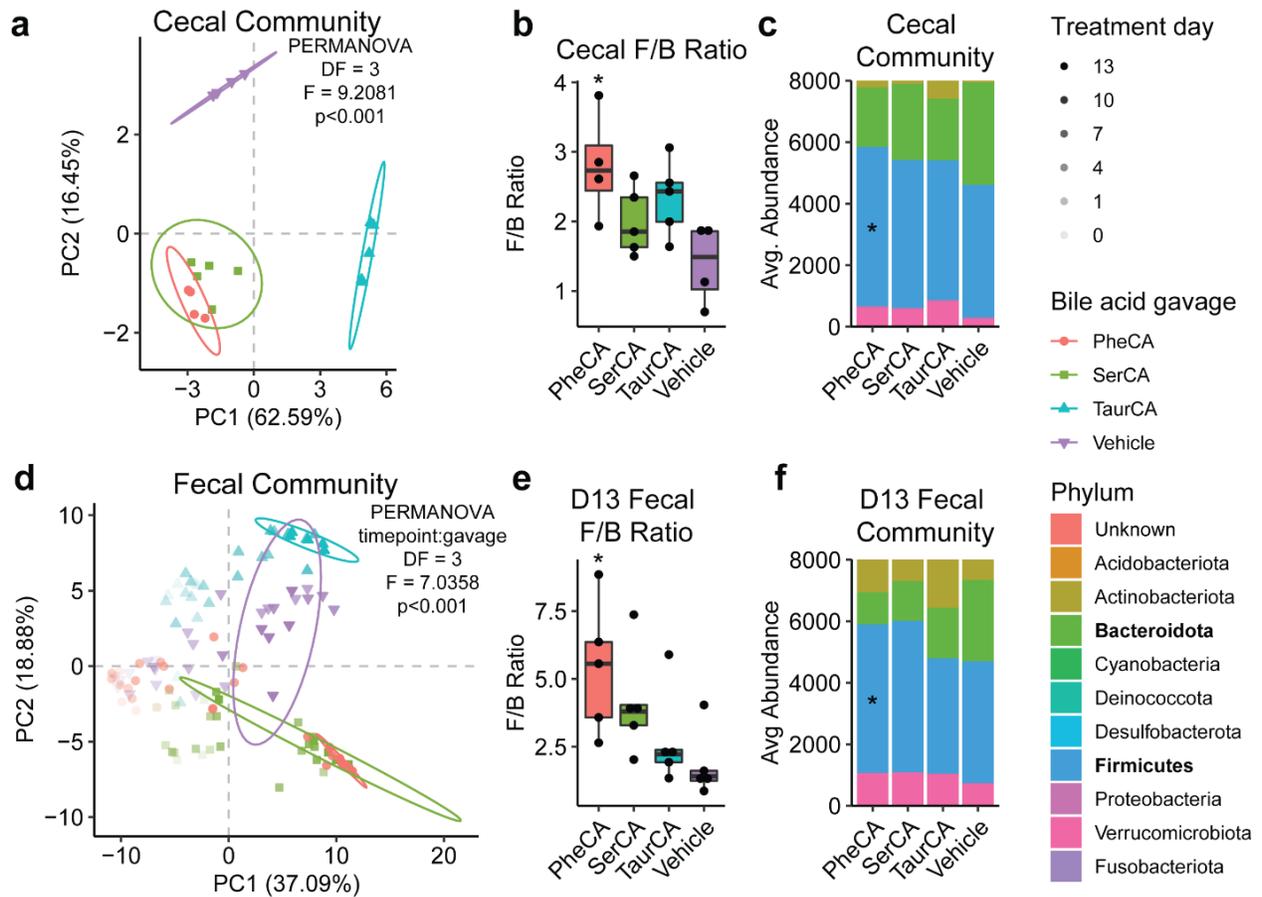
#### 4.4.2 - Murine microbiome shifts following high-dose MCBA administration

After the discovery of the antimicrobial properties of MCBAs and their functional dependence on the amino acid conjugated, we investigated these effects *in vivo*. C57BL/6 mice were administered MCBAs by oral gavage or feeding to monitor changes in the gut microbiome. Wild-type C57BL/6 mice were gavaged with 100 mg kg<sup>-1</sup> PheCA, SerCA, TCA, or mock control for 13 days and then sacrificed for sampling and 16S rRNA gene amplicon microbiome analysis on day 14. These were chosen to represent CA bound to a large hydrophobic amino acid (phenylalanine, PheCA), a small hydrophilic amino acid (serine, SerCA), with the host conjugate TCA used for comparison. Significant differences in cecal microbiome communities of female mice were observed between the groups (Figure 4.3a, PERMANOVA;  $F = 9.2081$ ,  $P < 0.001$ ), though differences in PheCA and SerCA gavage alone were less significant (PERMANOVA;  $F = 1.8692$ ,  $P = 0.033$ ). Microbiome shifts were also seen in the fecal samples, notably at day 13, where changes in community structure significantly differed between gavage groups over time (Figure 4.3d, PERMANOVA;  $F = 7.0358$ ,  $P < 0.001$ ). The ratio of Firmicutes to Bacteroidota (F/B ratio), formerly Bacteroidetes, has been of recent interest for its use as a marker for gut health. Previous reports have shown that a higher abundance of Firmicutes in feces has been associated with obesity (36, 37) while, conversely, an increase in the abundance of Bacteroidota is associated with inflammatory bowel disease (37). Female mice gavaged with PheCA had a significant increase in F/B ratio compared to vehicle controls for both



**Figure 4.2: Amino acid-dependency of MCBA antimicrobial efficacy**

Dose-response curves for *L. aerotolerans* when grown for 24 h in a, CA, b, LeuCA, or c, PheCA with calculated ED<sub>50</sub> shown in red. Dose-response curves for *P. anaerobius* when grown for 24 h in d, CA, e, LeuCA, f, PheCA, or g, TyrCA with ED<sub>50</sub> shown in blue. *n* = 4 independent cultures per strain.



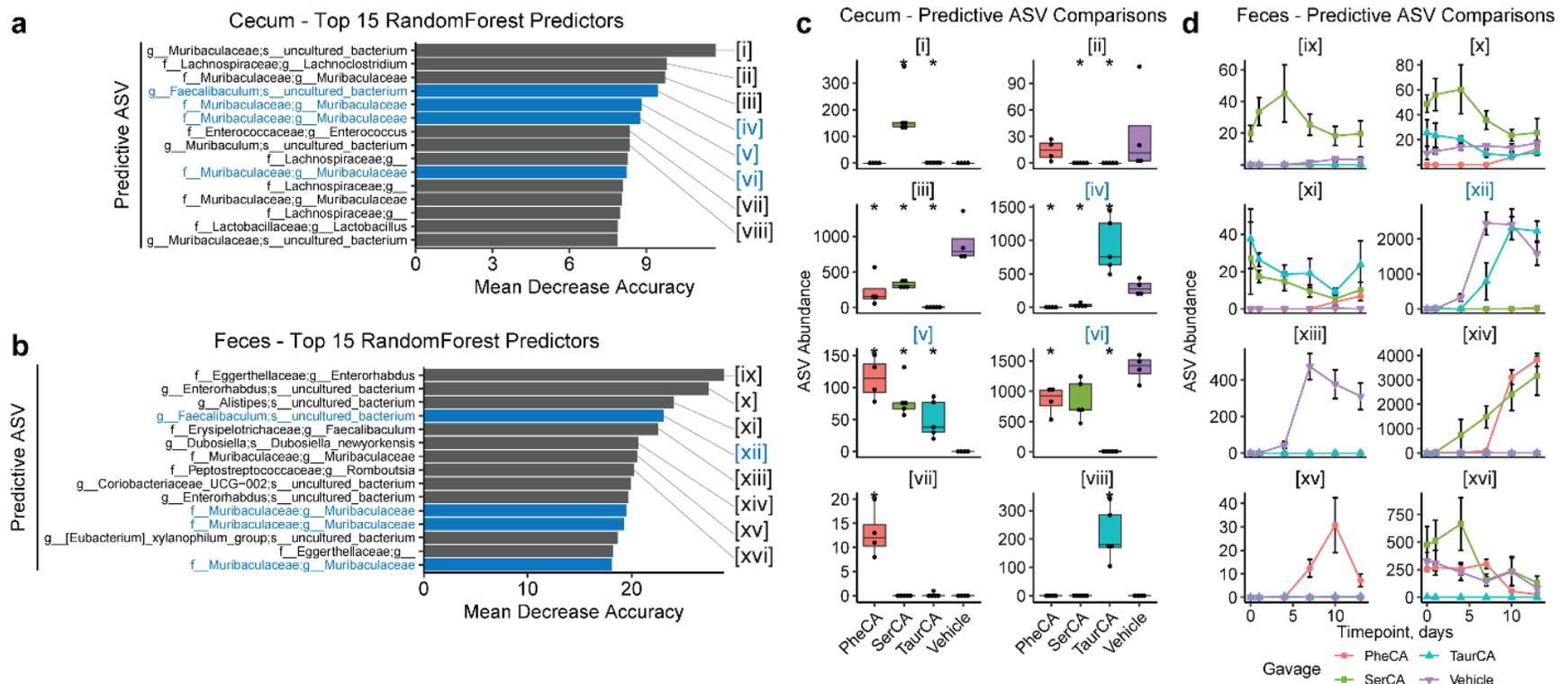
**Figure 4.3: Broad microbiome community shifts following 100 mg kg<sup>-1</sup> MCBA gavage**

Principal coordinate analysis (PCoA) of microbiome community structure via Bray-Curtis dissimilarity of **a**, cecal and **d**, fecal samples after oral gavage of different MCBAs in C57BL/6 mice. The ratio of Firmicutes/Bacteroidota (F/B ratio) between gavage groups for **b**, cecum and **e** fecal samples at day 13, with corresponding phylum-level community profiles for both **c**, cecum and **f**, fecal samples. Ellipses were drawn at 95% confidence for cecum and day 13 fecal samples. Cecal 16S analysis,  $n = 4-5$  per group; fecal 16S analysis,  $n = 5$  per group, per timepoint. Statistical significance determined by Wilcoxon rank sum tests, using vehicle gavage as a reference group. \* $P < 0.05$ .

cecum (Figure 4.3b,c) and fecal (Figure 4.3e,f) samples following gavage. This increase was not significant in fecal samples of mice gavaged with TCA ( $P = 0.222$ ) or SerCA ( $P = 0.056$ ), though a trend is apparent. Collectively, these data indicate that MCBAs produced by BSH/T can alter the gut microbiome differently than host-conjugated TCA or a mock control.

Random forest classification was used to determine the effects of BA gavage on cecal and fecal bacterial communities (Figure 4.4a,b). Of the 30 most important amplicon sequence variants (ASVs) for model accuracy, 18 were present in both cecal (Table 4.1) and fecal (Table 4.2) classifications, with four being present in the top 15 ASVs in both sample types (Figure 4.4a,b). PheCA gavage resulted in an increased abundance of cecal *Enterococcus*, members of which have been shown to produce MCBAs, in addition to a member of the genus *Muribaculaceae* (Figure 4.4). SerCA gavage resulted in an increased abundance of an uncultured *Muribaculaceae* species. TCA and vehicle gavage resulted in enrichment of a *Faecalibacterium* species that was absent in mice gavaged with either SerCA or PheCA (Figure 4.4) for both cecal and fecal samples (Figure 4.4). Both SerCA and PheCA gavage resulted in an increased fecal abundance of *Dubosiella newyorkensis* (Figure 4.4), a species first isolated in 2017 with little currently known about its role in the murine gut microbiome (38).

After this initial gavage experiment, we sought to quantify concentrations of administered MCBAs in the guts of these animals to determine the observed effects on the microbiome at physiologically relevant concentrations compared to those published in the literature (Table 4.3) (39, 40). Using a peanut butter feeding method (PBFM) where a  $100 \text{ mg kg}^{-1}$  SerCA dose was administered via peanut butter pellet resulted in SerCA



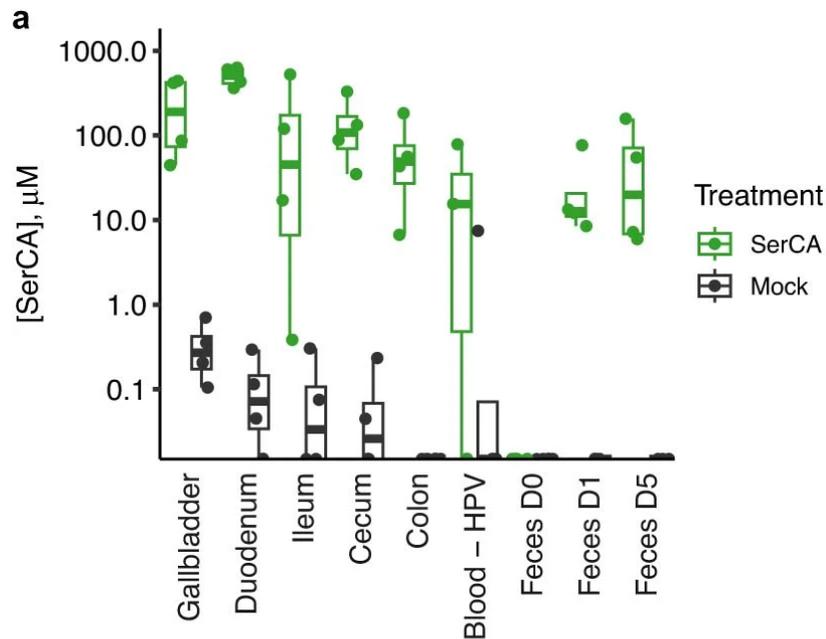
**Figure 4.4: Random Forest classification of murine microbiome community structure following 100 mg kg<sup>-1</sup> MCBA gavage**

Bar charts displaying the top 15 bacterial groups impacting the mean accuracy of Random Forest classification based on MCBA gavage group of **a**, cecal and **b**, fecal samples from days 1 to 13. ASVs highlighted in blue represent those that were matched between cecal and fecal classifications. Comparisons between gavage groups for the top predictive ASVs in **c**, cecal samples and **d**, fecal samples over time, with blue graph titles indicating shared features between the top 15 predictors in both analyses. Boxes represent the interquartile range (IQR), the center line represents the median, and whiskers represent 1.5 x IQR. Line plots show mean  $\pm$  s.e.m. Cecal 16S analysis,  $n = 4-5$  per group; fecal 16 analysis,  $n = 5$  per group, per timepoint. Statistical significance determined by Wilcoxon rank sum tests, using vehicle gavage as a reference group. \* $P < 0.05$ .

concentrations of  $506\pm65$   $\mu\text{M}$  in the duodenum,  $165\pm123$   $\mu\text{M}$  in the ileum,  $146\pm64$   $\mu\text{M}$  in the cecum,  $72\pm38$   $\mu\text{M}$  in the colon, and  $56\pm35$   $\mu\text{M}$  in feces after five days of feeding (Figure 4.5). These MCBA concentrations were 10 to 50-fold higher than average MCBA concentrations reported in the human gut, though reports vary considerably with some human samples reaching levels detected in these mice (Table 4.3). Therefore, we reduced the dose 10-fold and repeated the same treatments via PBFM. Significant microbiome shifts were not observed compared to TCA and amino acid + BA controls (Figure 4.14), likely due to concentrations several fold lower than the  $\text{ED}_{50}$  values reported. These *in vitro* and *in vivo* experiments demonstrate that MCBAs do have antimicrobial properties, but this depends on the conjugated amino acid and effects *in vivo* occurred only at the higher levels of their physiological concentration in humans.

#### 4.4.3 - MCBAs enter enterohepatic circulation (EHC)

An important question about MCBAs is whether or not they can enter EHC, a tightly regulated process for recycling BAs starting at the terminal ileum that returns BAs to the liver through the hepatic portal vein (HPV) (41). To investigate the propensity for MCBAs to enter EHC, C57BL/6J mice were fed  $100\text{ mg kg}^{-1}$  of SerCA through the PBFM (42). SerCA was detected in all GI tissues analyzed, blood from the HPV, and appeared in fecal pellets after 24 h (Figure 4.5). We then sought to determine the ability of MCBAs with various conjugated amino acids to enter EHC. Mice were fed a mixture of eight MCBAs for 5 days ( $10\text{ mg kg}^{-1}$  of AlaCA, AspCA, GluCA, LeuCA, PheCA, SerCA, ThrCA and TyrCA) through PBFM. These MCBAs were observed throughout the GI tract and in the liver, kidney, serum, and gallbladder with particularly high abundances of PheCA and SerCA across all samples (Figure 4.6 and Table 4.4). However, these were also detected



**b**

Sample	SerCA-treated	Mock
Blood - HPV	31.3 ± 24 μM	1.86 ± 1.86 μM
Gallbladder	247 ± 105 μM	0.328 ± 0.131 μM
Duodenum	506 ± 65.1 μM	0.103 ± 0.0628 μM
Ileum	166 ± 123 μM	0.0875 ± 0.069 μM
Colon	72 ± 38.2 μM	0 ± 0 μM
Cecum	146 ± 64.1 μM	0.0625 ± 0.053 μM
Feces - D0	0 ± 0 μM	0 ± 0 μM
Feces - D1	27.6 ± 16.3 μM	0 ± 0 μM
Feces - D5	56.5 ± 35.8 μM	0 ± 0 μM

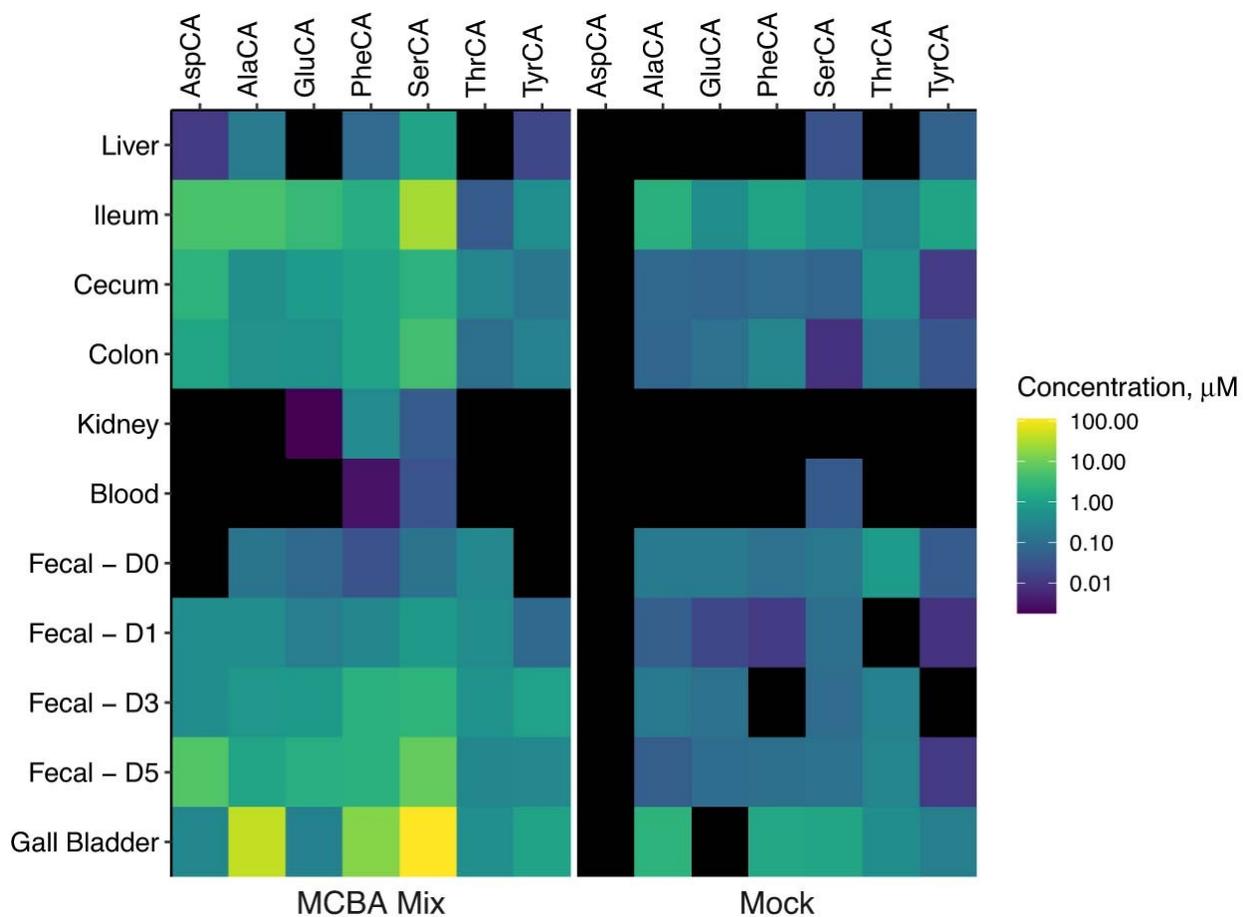
**Figure 4.5: SerCA concentrations following 100 mg kg<sup>-1</sup> feeding**

**a**, SerCA concentrations in murine tissue and fecal samples following 100 mg kg<sup>-1</sup> SerCA dosing via PBFM. Data are presented as boxplots where the middle lines are the median, lower and upper hinges represent the first and third quartiles, upper whiskers extend to maxima and lower whiskers extend to minima. **b**, Table showing SerCA concentration by sample type, presented as mean ± s.e.m., *n* = 4 mice per group.

at low concentrations in mock-fed controls, probably due to basal concentrations of production *in vivo*. Therefore, we further verified the ability of both SerCA and PheCA to enter EHC in a follow-up experiment including more equimolar amino acid + BA controls (matching 10 mg kg<sup>-1</sup> of individual BA). Interestingly, SerCA concentrations were eight-fold higher in the gallbladder of SerCA-treated animals than PheCA in PheCA-treated animals (359 ± 80 μM versus 42 ± 12 μM) and SerCA was detected consistently in the liver (0.95 ± 0.20 μM) when PheCA was not (Figure 4.6 and Table 4.4). Gallbladder and liver samples from amino acid + BA controls contained low concentrations of these compounds, supporting *de novo* conjugation and subsequent circulation, but these concentrations were significantly less than MCBA-fed animals. It is possible that the limited EHC observed with PheCA was due to specific hydrolysis by pancreatic carboxypeptidases, as reported for TyrCA and other conjugated BAs (43, 44). However, neither carboxypeptidase showed activity when incubated with PheCA or SerCA while each catalyzed near-complete hydrolysis of positive controls (Figure 4.15). It is also possible that preferential microbial hydrolysis of PheCA in the gut, which has been recently described, may have occurred (45). These experiments support the finding that MCBAs can enter EHC intact when fed to mice at physiologically relevant concentrations, potentially affecting liver and BA metabolism and that the degree of EHC depends on the amino acid conjugated.

#### 4.4.4 - Human bariatric surgery affects fecal MCBAs

To investigate whether MCBA concentrations change in the context of human gastrointestinal health, we analyzed fecal samples from patients who underwent sleeve gastrectomy as a treatment modality for obesity. The concentrations of MCBAs and other



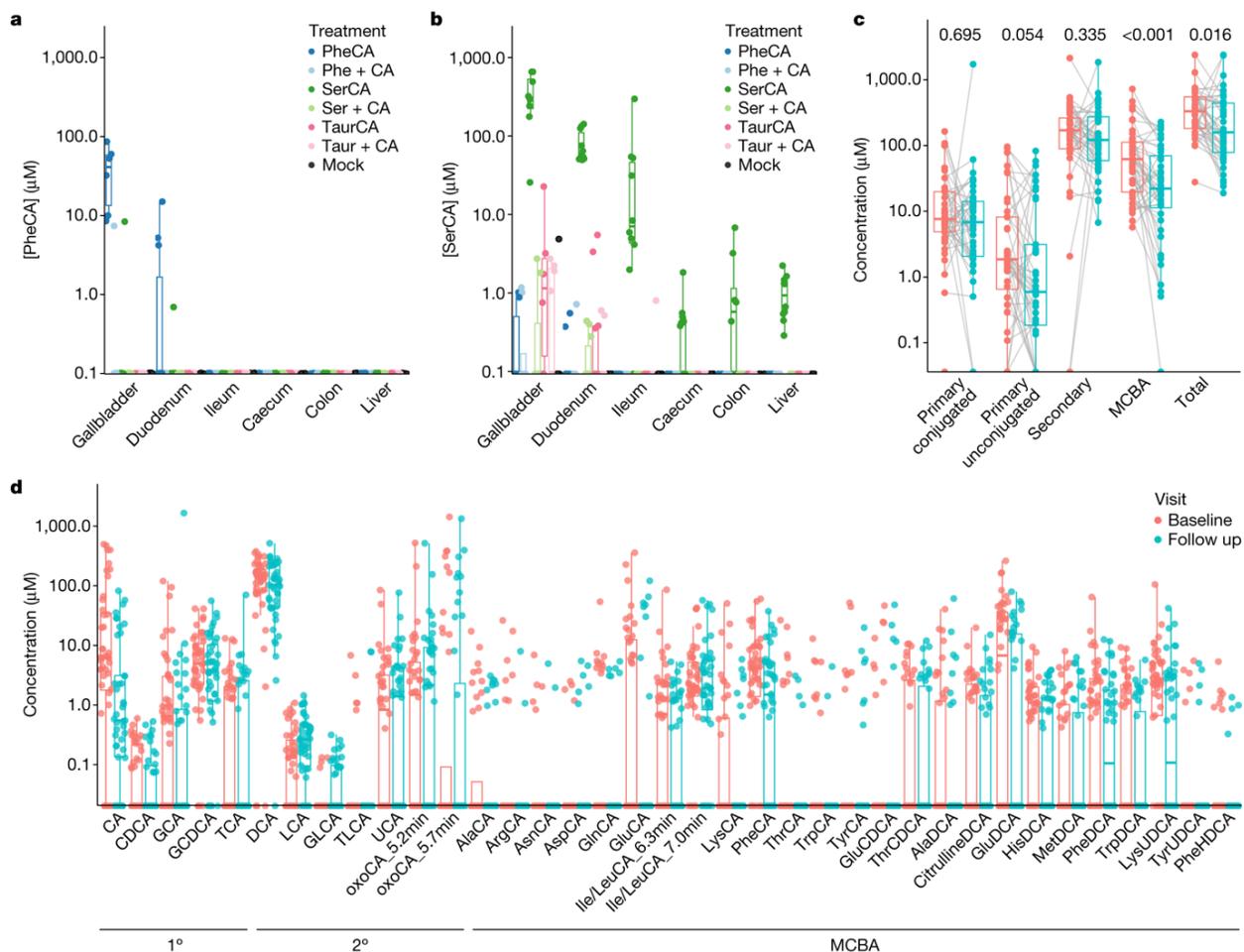
**Figure 4.6: MCBA concentrations in fecal and tissue samples following mixed MCBA dosing via PBFM**

Data are presented as the average concentration of each MCBA included in the MCBA mix ( $80 \text{ mg kg}^{-1}$  total,  $10 \text{ mg kg}^{-1}$  per individual MCBA).  $n = 3$  treatment, 2 control.

BAs were quantified from samples collected before surgery and 3 months post-operation. A diverse complement of MCBAAs were detected, including at least 25 unique compounds (Figure 4.7c, Table 4.6) with an average total MCBA concentration of  $78 \pm 12 \mu\text{M}$ . In comparison, primary conjugated BAs were measured at  $34 \pm 20 \mu\text{M}$ , free primary BAs at  $10.8 \pm 2.4 \mu\text{M}$  and secondary BAs at  $223 \pm 33 \mu\text{M}$ . This is evidence that fecal MCBAAs can reach concentrations at or above primary BAs in feces and approximately one-third the concentration of secondary BAs (Table 4.7). Furthermore, collective analysis of BA chemistry showed significant reductions in fecal concentrations of MCBAAs ( $P = 8.9 \times 10^{-4}$ ) and total BAs ( $P = 0.016$ ) after sleeve gastrectomy but conjugated primary BAs, free BAs and secondary BAs were not (Figure 4.7d). This supports the hypothesis that MCBAAs are a substantial component of the human BA pool and are affected by surgical treatment for obesity.

#### *4.4.5 - Contrasting diversity trajectories within infant fecal metabolomes and microbiomes*

We explored changes in the Shannon index of diversity in the microbiome and metabolome through the first year of life and found that infant metabolomes became slightly less diverse (Figure 4.9a,  $\rho = -0.11$ ,  $P = 0.0143$ ) over time while their microbiomes became more diverse (Figure 4.9b,  $\rho = 0.38$ ,  $P = 4.02 \times 10^{-14}$ ). Further inspection of the underlying drivers of this relationship showed it was a consequence of reductions in metabolite richness and an increase in microbial richness. Across samples, the overall count of unique molecular features decreased significantly as infants matured through their first year (Figure 4.9c, Spearman's  $\rho = -0.28$ ,  $P = 9.76 \times 10^{-11}$ ). However, the opposite was observed for the microbiome (Figure 4.9d,  $\rho = 0.46$ ,  $P = 6.99 \times 10^{-21}$ ). This finding highlighted zero inflation as a marked aspect of the structure of both datasets, which led



**Figure 4.7: BA concentrations in mouse tissue samples following MCBA feeding and in human feces of patients undergoing sleeve gastrectomy**

**a,b**, Concentrations of PheCA (**a**) and SerCA (**b**) in mouse tissue samples following 10-day dosing with the indicated treatments through PBFM. PheCA is highly abundant in the gallbladder and present in the duodenum of PheCA-fed mice, while SerCA is enriched in all tissues sampled, including the colon and liver.  $n = 5$  male and 5 female mice per treatment. **c,d**, BA class shifts in a patient population undergoing sleeve gastrectomy before (baseline) and 3 months after (follow up) surgery, with significant decreases in total MCBA concentration and total BA (**c**) and changes in individual BA concentrations in that cohort (**d**).  $n = 44$  patients, with paired samples at each timepoint. Data in **a–d** are presented as boxplots where the middle lines are the median, lower and upper hinges represent the first and third quartiles, upper whiskers extend to maxima and lower whiskers extend to minima. The significance between timepoints was determined by Wilcoxon signed-rank tests with Benjamini–Hochberg  $P$  value correction.

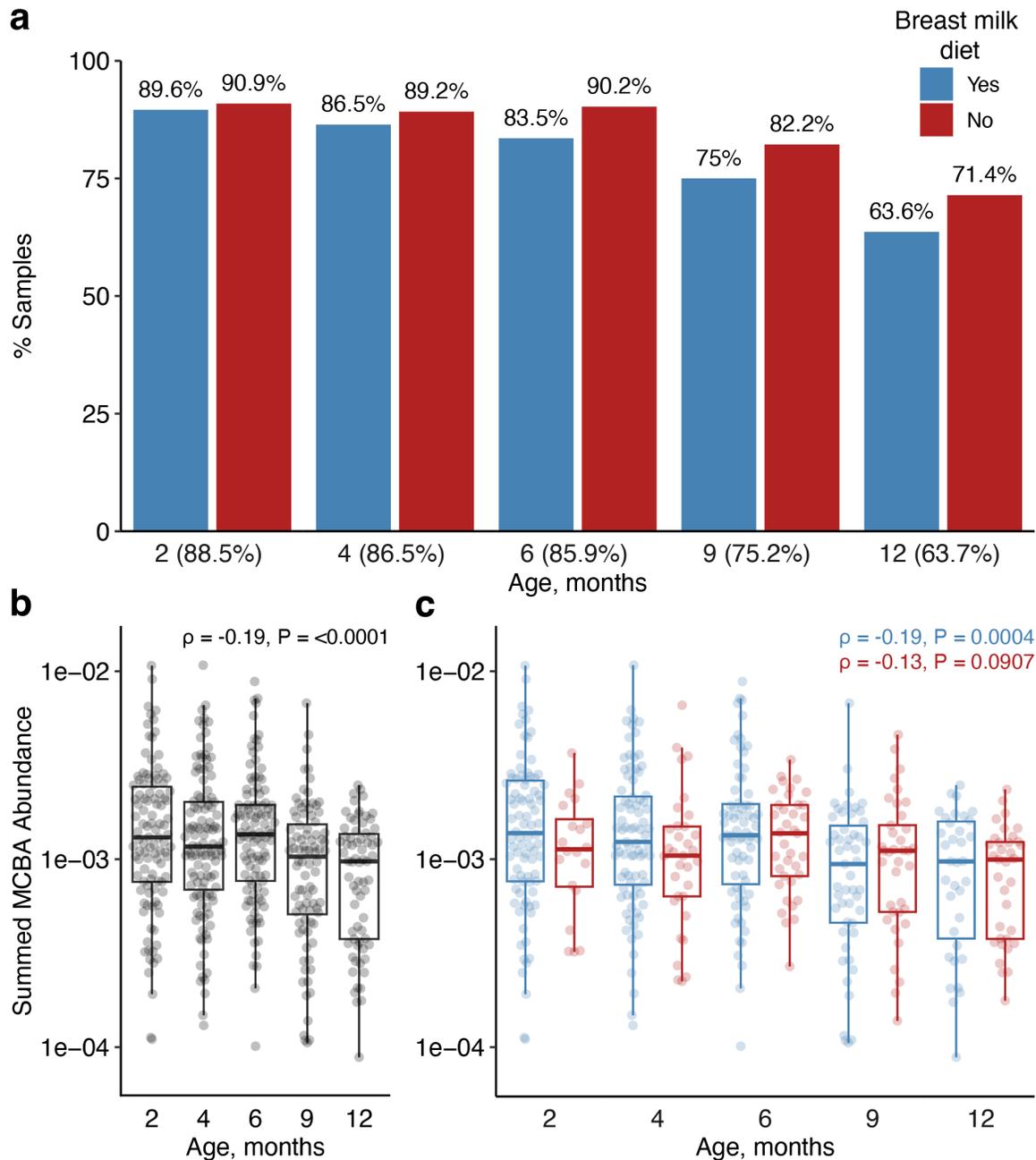
to further analysis of changes in the proportion of zeros for each feature and the development of the statistical methods below to characterize the trajectories of presence/absence of these features through infant development.

#### *4.4.6 - MCBA shifts across the first 12 months of life*

After investigating how external variables correlate with shifts in the infant metabolome and microbiome and uncovering BA shifts indicative of robust detoxification, we then sought to investigate temporal changes associated with MCBA. Investigating the proportion of samples at each age with at least one MCBA present revealed that the fraction of MCBA-containing samples decreased with time (Figure 4.8a). Given the extensive literature surrounding diet early in life, namely breast-fed or formula-fed, we investigated whether or not early diet had impacts on these proportions and found decreases in MCBA-positive samples were similar regardless of diet. With zero-inflation prevalent within these datasets, we then investigated shifts in total MCBA abundance for samples containing at least one annotated MCBA and observed that the MCBA abundance significantly decreased with time across the dataset (Figure 4.8b, Spearman's  $\rho = -0.19$ ,  $P < 0.0001$ ). Stratifying by diet revealed that samples from infants still fed a breast milk diet exhibited more significant correlations between age and MCBA abundance (Figure 4.8c,  $\rho = -0.19$ ,  $P < 0.0004$ ) compared to those on solid food or formula (Figure 4.8c,  $\rho = -0.13$ ,  $P = 0.0907$ ).

#### *4.4.7 - Maternal health significantly impacts metabolome and microbiome development*

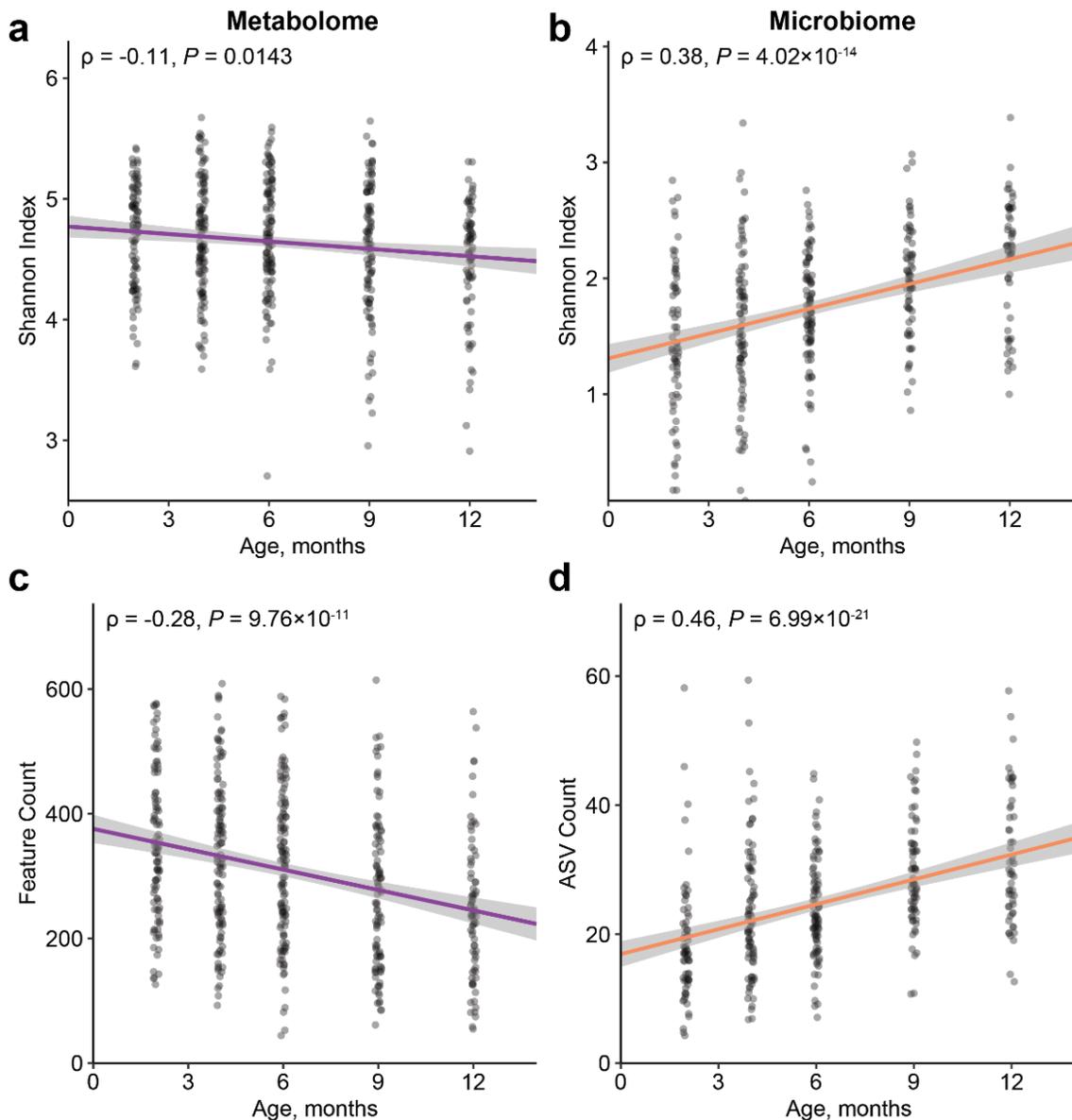
In order to further understand the multi-omic shifts occurring within the developing infant, we applied ordination analysis to both microbiome and metabolome datasets as has been performed previously (12, 46). We observed that as infants aged, shifts in



**Figure 4.8: MCBA-containing sample proportions across the first 12 months of life**

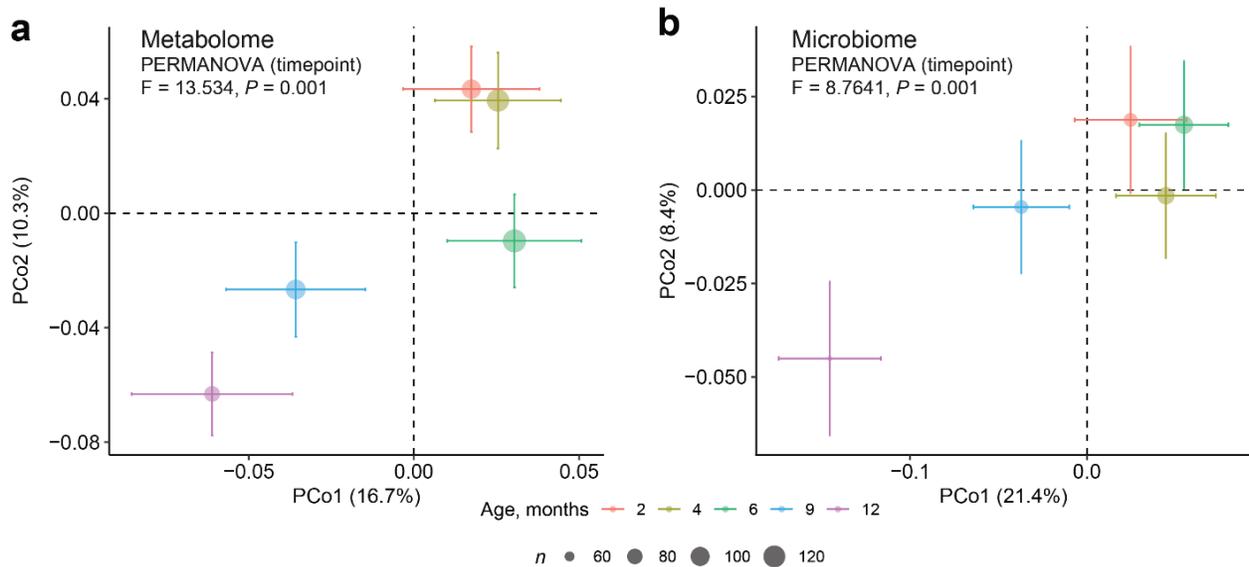
**a**, Proportion of all samples within each timepoint that contained at least one annotated MCBA, stratified by whether the child was still fed primarily a breast milk diet. Percentages in parentheses represent the combined sample proportion irrespective of diet. Changes in MCBA abundance for MCBA-positive samples **b**, across the entire data set and **c**, separated by breast milk diet. The correlation was determined by Spearman's rank correlation ( $\rho$ ) and associated  $P$  value.

metabolome structure occur in similar magnitude to the microbiome; timepoints closer to birth are more similar to each other but shift dramatically across the first month of life (Figure 4.10, Table 4.8 and Table 4.9). We then sought to investigate other covariates that may be driving differences in metabolome and microbiomes. Univariate analysis was used to determine the association of 33 covariables with microbiome or metabolome profile and because of the obvious association with age across both datasets, we stratified this analysis by infant age (Figure 4.11). This procedure revealed 8 significant associations with the metabolome (Table 4.10) and 5 significant associations with the microbiome (Table 4.11). Maternal health was strongly associated with the metabolome. Maternal body mass index (BMI) before and after pregnancy was significantly associated, with pre-pregnancy BMI having the greatest explained variance for differences in the metabolome by timepoint. Conversely, maternal BMI pre-pregnancy was not significantly associated with changes in the microbiome and BMI post-pregnancy was not significant following  $p$ -value adjustment ( $P_{adj} = 0.0845$ ). Dietary factors related to breastfeeding (recently breastfed, primarily breastmilk diet, and formula feeding) were significantly associated with both metabolome and microbiome. However, the frequency of an infant finishing a pumped meal was significantly associated with the infant fecal metabolome but not the microbiome. Finally, the self-reported race of the mother was significantly associated with both the metabolome and microbiome while the reported race of the infant was only significantly associated with the metabolome.



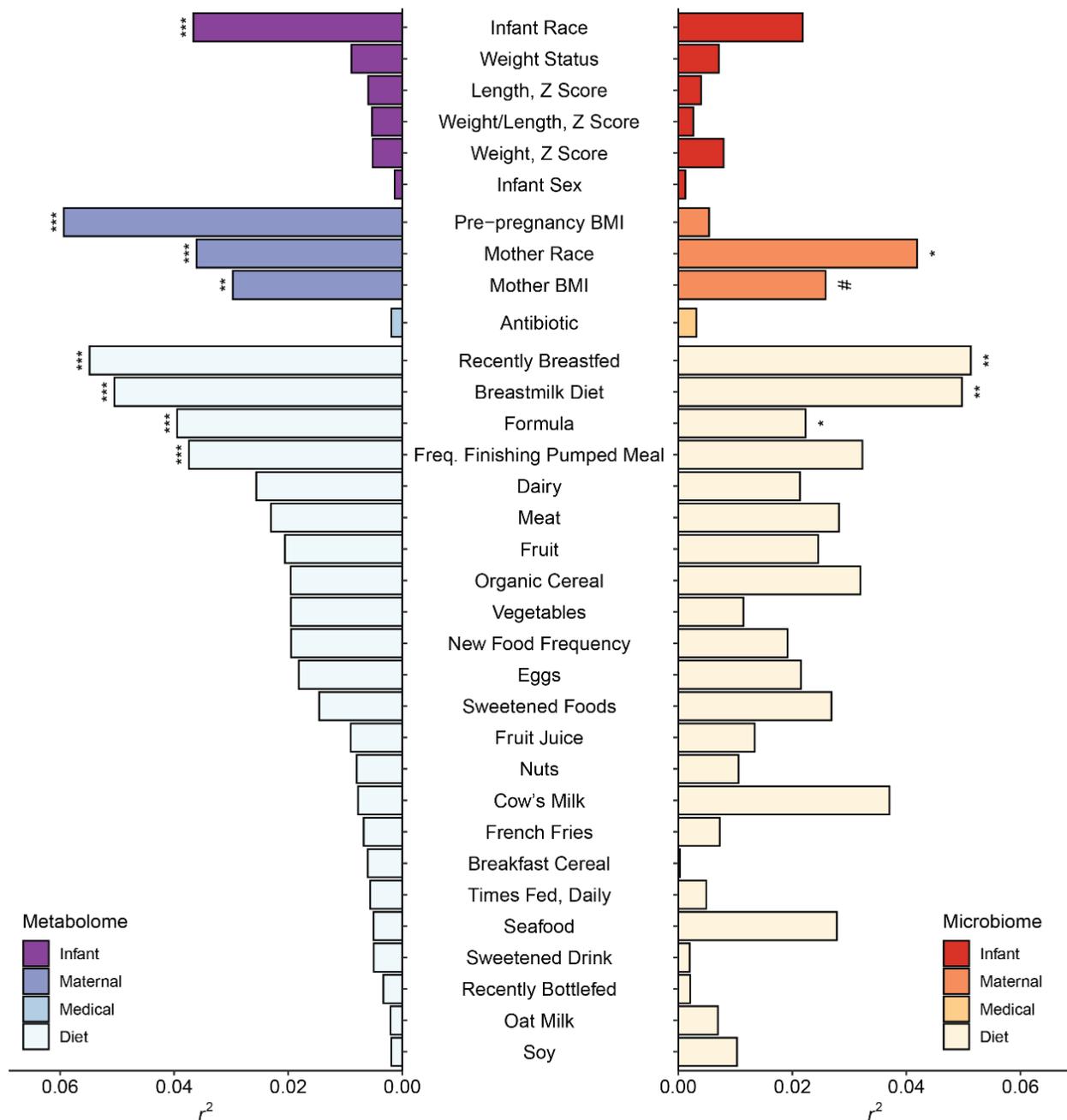
**Figure 4.9: Temporal shifts in alpha-diversity within infant fecal metabolomes and microbiomes driven by richness**

Multi-omic changes in Shannon index, a measure of beta-diversity, over time within infant fecal samples. **a**, Metabolome diversity shows significant decreases with time while **b**, microbiome diversity exhibits significant positive correlations with time. Significant changes in **c**, metabolome feature count and **d**, microbiome ASV count support their role in driving changes in overall beta-diversity. The significance was determined by Spearman correlation, with Spearman's rho ( $\rho$ ) and associated  $P$  values provided.



**Figure 4.10: Temporal beta-diversity shifts within infant fecal metabolome and microbiome communities**

Bray-Curtis dissimilarity ordinations showing the mean centroid of each timepoint for the **a**, metabolome and **b**, microbiome. The centroid size was based on the number of samples at a given timepoint, error bars represent s.e.m. Sampling timepoint was significantly associated with microbiome and metabolome profiles across the first year of life (detailed results in Table 4.8 and Table 4.9).



**Figure 4.11: Univariate effects of 33 covariables on multi-omic sample dissimilarity**

Significance and correlation coefficients by each of 33 covariables against the metabolome (left, shaded blue) and microbiome (right, shaded red), as modelled by envfit with stratification by timepoint. Significance was determined by permutation testing with Benjamini-Hochberg false discovery rate (FDR)  $P$  value adjustment. # $P < 0.1$ , \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

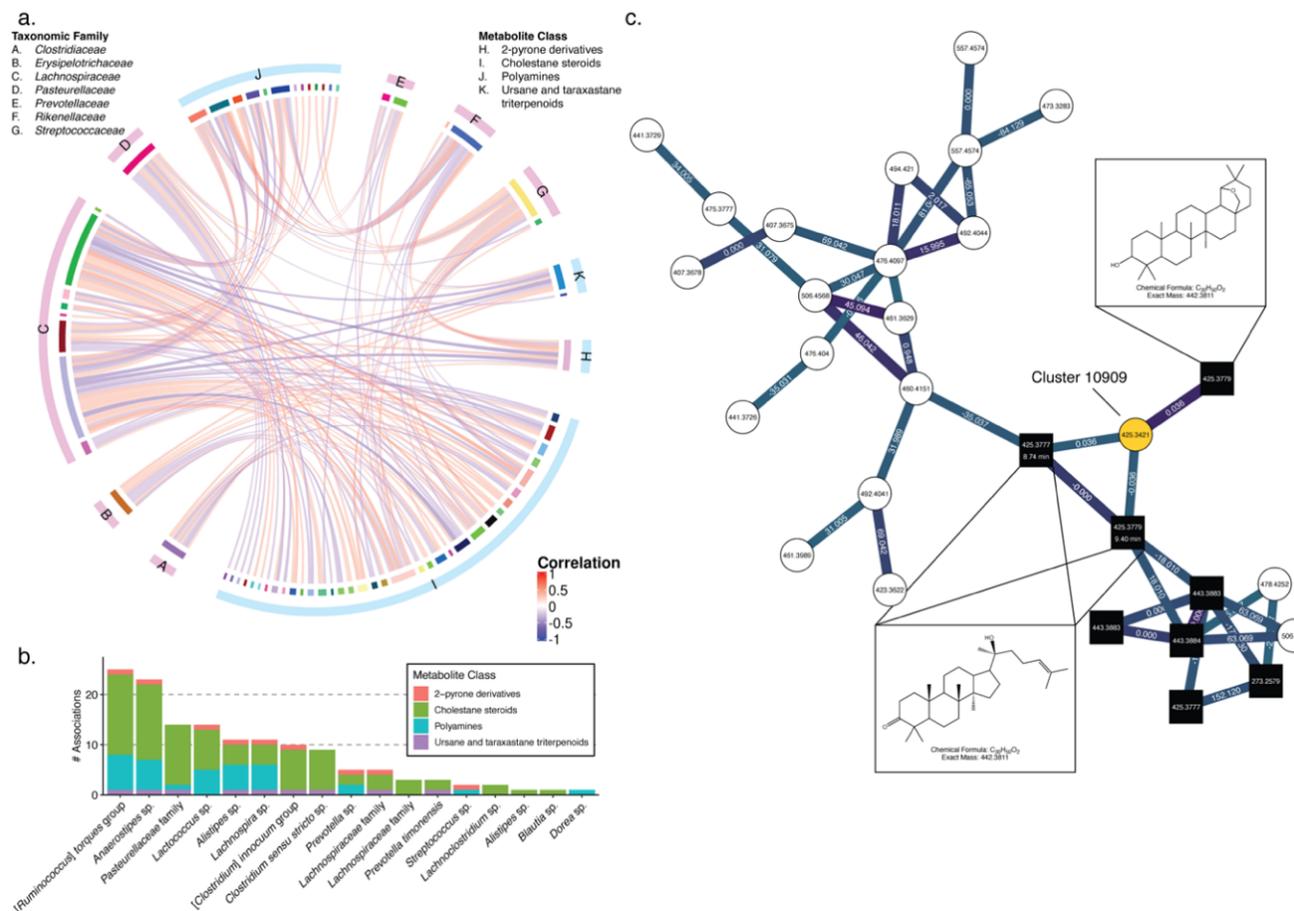
#### 4.4.8 - Microbiome maturation correlates with changes in the metabolome

We used the estimated subject-level expected change for the proportion of zeroes (gray dots in Figure 4.17) to identify metabolite-ASV pairs that appear to change in a coordinated manner. To do this, for each metabolite-ASV pair, we computed Pearson's correlation (across subjects) between the predicted change in the probability of not being detected and tested if the correlation significantly differed from zero ( $P_{\text{adj}} < 0.05$ ). Here, a positive correlation indicates that the metabolite and the ASV pair show concordant trajectories (e.g., both increasing or decreasing in the first-year change of not-being-detected probability, Figure 4.17a-c) and a negative correlation indicates the opposite (Figure 4.17d-f). Among the 93,660 metabolite-ASV pairs, 940 pairs had correlations significantly different than zero suggesting that these pairs may be changing, within subject, in a coordinated manner. These pairs involved 62 ASVs and 410 metabolites. Of these correlations, 473 (50.3%) were positive and the remaining were negative.

To gain insight into these results we conducted an enrichment analysis to associate groups of metabolites and ASVs that have seemingly coordinated changes in the proportion of zeros over time. The enrichment analysis pointed to seven microbial families (17 ASVs) and four metabolite groups (51 metabolites) which appear to be significantly enriched for coordinated longitudinal changes (Figure 4.12a). For instance, two ASVs within the *Streptococcaceae* family often changed in a coordinated fashion with metabolites classified as polyamines and cholestane steroids (Figure 4.12a, red ties between groups G-J and G-I). Significant associations were dominated by two members of the *Lachnospiraceae*, a member of the genus *Ruminococcus* and a member of the genus *Anaerostipes*, and the bulk of associations were with compounds related to

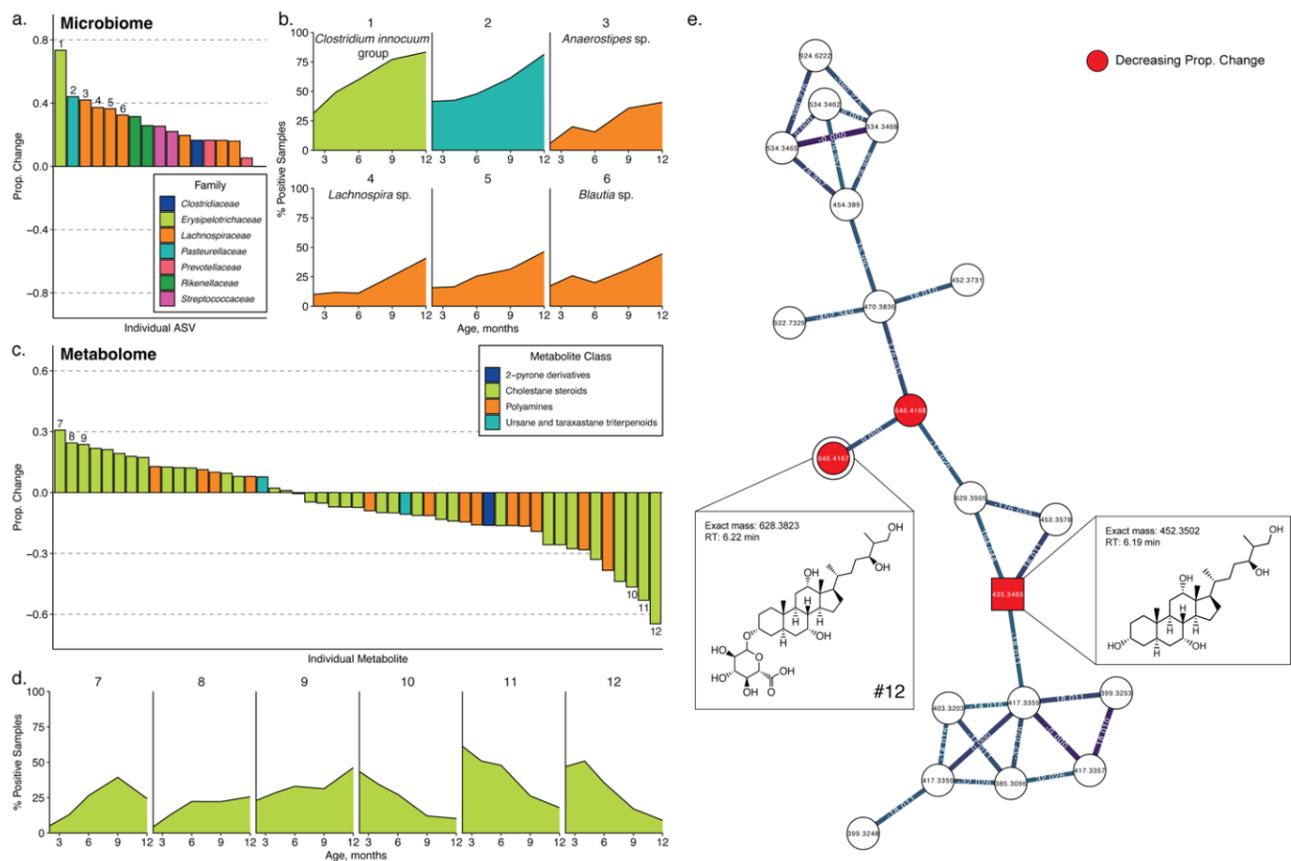
cholestane steroids, a class of C27 bile acids (Figure 4.12b). The most significant correlation was between Cluster 10909, an unknown metabolite, and the ASV related to *Ruminococcus torques* mentioned above ( $P = 7.8e-11$ ). However, molecular networking reveals Cluster 10909 is spectrally similar to annotated metabolites containing a steroid core but differing primarily in the number of ring structures present (Figure 4.12c).

We further investigated the 17 ASVs and 51 metabolites from the enriched groups. For the microbiome, most ASVs appear to have an increase in the probability of being detected over time, further supporting the model that the gut microbiome becomes more complex through early life (Figure 4.13a). These increases are driven by the acquisition of members of the *Lachnospiraceae*, common bile acid metabolizing bacteria with several implications in host health (47). The top ASV significantly increasing in population prevalence based on sample trajectories is a member of the *Clostridium innocuum* group, with 31.4% of 2-month samples increasing to 83.3% of 12-month samples containing the ASV (Figure 4.13b). Of the top 10 metabolites, 8 were structurally related to cholestane steroids as determined by MS<sup>2</sup> spectra (Figure 4.13c). The top metabolite that significantly decreased in probability of being detected, Cluster 5583 (646.4167 *m/z*), did not match any annotated compounds; however, molecular networking revealed spectral similarity of this unknown to an annotated cholestane steroid (Cluster 6185, Fig. 5d-e), which also decreased in sample prevalence as infants matured. MS<sup>2</sup> spectral alignment supported the relatedness of these metabolites, but 5583 contained a set of peaks exhibiting a  $194.0417 \pm 0.0024$  Da increase from those matching Cluster 6185 (Figure 4.18). This shift matches what one would expect following glucuronidation, a known method of molecular detoxification for bile acids (48).



**Figure 4.12: Longitudinal changes in presence/absence are highly correlated for certain pairs of metabolites and ASVs**

**a**, Co-correlation network between metabolite and ASVs where each node in the inner circle represents one metabolite or ASV and tracks in the outer circle indicate the taxonomy groups of inner molecules (groups A-G are ASV families, and groups H-K are metabolite groups). The colors of the strings indicate the correlation of longitudinal trends between two nodes. **b**, The number of significant metabolite associations for each ASV, colored by NPC class, with the most specific taxonomic assignment displayed. **c**, Molecular network with the most significant microbe-metabolite correlation, demonstrating close relatedness between the unknown metabolite (Cluster 10909) and several annotated cholestane steroids. White, circular nodes are unknown metabolites and black boxes are metabolites with library hits from GNPS.



**Figure 4.13: Microbial and metabolite features with significant temporal shifts in zero-proportions**

The change in probability of being detected for **a**, the 17 ASVs with significant changes ( $P < 0.05$ ), where ASVs are colored by assigned taxonomic family, and **b**) the top 6 ASVs with increases, colored by taxonomic family and labeled with the most specific taxonomic assignment, where applicable. Similarly, the change in probability of being detected for **c**, 51 metabolites with significant changes ( $P < 0.05$ ), where metabolites are colored by molecular class as determined by the Natural Products Classifier (49), and **d**) the top three metabolites with increases and the top three metabolites with declines over time, as shown by the change in positive sample proportion. **e**, The molecular network for the metabolite with the greatest, negative proportional change (#12), putatively described as the glucuronidated form of the annotated cholestane steroid within the same network. White, circular nodes are unknown metabolites and boxes are metabolites with library hits from GNPS. Nodes highlighted in red significantly decrease in sample proportion over time.

## 4.5 - Discussion

We investigated correlations between MCBAs and dynamic microbial and biochemical systems. We first characterized the *in vitro* impacts of MCBAs on bacterial monocultures of relevant gut bacteria, including known MCBA producers and pathogens. We show that MCBAs exhibit variable impacts on bacterial growth, with hydrophobic amino acid bound MCBAs being more antimicrobial than polar, hydrophilic amino acids. However, this generality varied by bacterial strain with some showing slight increases in growth compared to the vehicle control, such as *E. bolteae*. For *L. aerotolerans* and *P. anaerobius*, more hydrophobic MCBAs showed greater growth inhibition compared to controls. ED<sub>50</sub> concentrations further support the view that MCBA impacts are strain specific; *L. aerotolerans* was more susceptible to LeuCA than CA alone, yet *P. anaerobius* showed decreased susceptibility to LeuCA than CA. The inverse was then seen for PheCA.

We then hypothesized that MCBAs would impact gut microbiome structure. The effects we observed *in vitro* translated when investigated *in vivo*, where mice fed MCBAs displayed changes in the structure of their gut microbiomes. These shifts were only observed at the highest concentrations of BAs found in the murine gut detected in humans. Although not tested in this study, other base BAs with stronger antimicrobial properties, such as DCA or LCA, may become more potent when conjugated. Further study of conjugates with the strongest antimicrobial effects will help elucidate their potential as agents of microbial warfare in the human gut or, in contrast, as a means of detoxification.

The fate of MCBA *in vivo* remains a significant question. Recent work has shown that BSH/T can hydrolyze these compounds *in vitro* (45) as can pancreatic carboxypeptidases A and B (44). However, we did not observe PheCA and SerCA deconjugation by pancreatic carboxypeptidases A and B *in vitro*. Utilizing murine models, we showed that MCBA are capable of entering EHC. Liver and gallbladder concentrations of SerCA were higher than PheCA, indicating some selection for entry into these organs. Whether this behavior was due to specific selectivity by BA transporters in the ileum, favored microbial hydrolysis or other forms of metabolism remains unknown. Follow-up studies are needed to better understand the fate of MCBA and their effects on the host and its microbiome.

Bariatric surgeries are known to result in compositional changes in both host BA and their microbiome (50–52). We therefore sought to characterize fecal BA shifts within a human patient cohort following sleeve gastrectomy. We observed significant decreases in total BA concentration 3-months post-operation, contrary to established reports (50–52). It is important to note that we quantified fecal BA concentrations whereas serum BA concentrations are more frequently measured clinically. This difference may also be due in part to our ability to identify and quantify MCBA shifts whereas many earlier reports focused on primary BAs (conjugated and free) in addition to notable secondary BAs. Significant shifts in the concentration of primary and secondary BAs were not observed in our cohort. However, MCBA concentrations significantly decreased following surgery and may be driving the shifts observed in total BA concentrations.

We also observed MCBA shifts within the developing fecal metabolome. As infants matured, MCBA prevalence within each sampling timepoint decreased. What remains to

be understood are additional causes of decreased MCBA prevalence, independent of breast milk diets. Given the evidence shown here that MCBAs vary in antimicrobial activity, the decrease in prevalence may be due to infant bile becoming less toxic not only to the host but to the host microbiota as well. Decreased MCBA presence matched overall decreases in metabolome diversity as infants grew older through the first 12 months. Investigating broad changes in fecal metabolome and microbiome diversity showed contrasting trajectories; where microbiome diversity increased with age, metabolome diversity decreased with age. This may be a consequence of increased metabolic capacity of the gastrointestinal microbiota. That is, as species and strain diversity increase, so too does genetic variation within the community resulting in metabolism of more complex metabolites into a smaller, core metabolome.

Our approach to characterizing microbiome and metabolome changes in the first year of life, while accounting for the prevalence of zeros, also allowed prioritization of features with the most dynamic changes. Although annotation is always a significant challenge in untargeted metabolomics, and we encountered a large proportion of the metabolites that were prioritized but remain structurally unknown, detailed spectral analysis of our changing metabolites enabled annotation of some features. In our longitudinal trajectory analysis, the top feature decreasing in prevalence as infants aged was a glucuronidated form of a cholestane steroid. Glucuronidation and sulfation are similar methods of bile acid modification by liver enzymes, both of which facilitate the excretion and detoxification of these detergent compounds (48, 53, 54). Bile acid detoxification is well-studied in infants due to the prevalence of diseases such as cholestasis and jaundice which affect 1 in ~2500 infants globally (55). These conditions

result in extensive detoxification and excretion of bile acids to avoid liver toxicity, often comprised of glucuronidation and sulfation. Though we do not have a record of cholestasis in this cohort, our finding may represent some degree of bile acid detoxification as a natural feature of the developing infant gut-liver axis and warrants further investigation of the role these biochemical transformations play in the incidence of infant cholestasis.

It is difficult to fully unravel whether some of the shifts we identified using our longitudinal modeling are due to metabolome-microbiome interplay or simply due to tandem maturation of both the host and their microbiota. Longitudinal modeling identified related cholestane steroids as decreasing in prevalence over time, with certain ASVs increasing or decreasing in correlation with those cholestane-related metabolites. Cholestanic acids (C<sub>27</sub> bile acids) are structurally similar to cholanoic acids (C<sub>24</sub> bile acids) as both contain the same sterane core due to their synthesis from cholesterol, with cholestanic acids having a three-carbon longer arm branching off the core (56). Thus, decreases in the putative cholestane glucuronide may be indicative of proper infant development and less of increased microbial metabolism. As pressures from bile lessen, so too does the need for members of the host microbiota to detoxify these compounds through hydrophilic amino acid ligation.

It is well established that the infant gastrointestinal tract undergoes significant physiological shifts within the first 12 months of life; neonatal diets primarily composed of mother's milk transitions to solid foods as the infant matures and exposures throughout this period drive changes in both host physiology and associated microbiota (9–12, 57, 58). Given this knowledge, we then sought to identify additional variables that best explain

microbiome and metabolome variation with time. We found similar environmental and behavioral impacts on the data as reported in other studies (10, 12, 54, 59), with breastfeeding and formula feeding driving the microbiome variation, but we also identified impacts of the mother's BMI, race and feeding behavior on the metabolome. The effect of mother's BMI on fecal biochemistry is of particular interest, as this may be a manifestation of feeding habits of the mother that translate to the child, either through breastmilk or directly. Further exploration of the impacts of metabolome dynamics in early life is warranted and may reveal molecular features important for shaping the gut chemical and microbial environment.

## 4.6 - Methods

### 4.6.1 - Assaying MCBA impacts on bacterial growth

Pure cultures were grown in reduced RCM and incubated overnight under anaerobic conditions at 37 °C without shaking. The OD<sub>600</sub> was measured followed by dilution to a final OD<sub>600</sub> of 0.01 in clear bottom 96-well plates containing RCM supplemented with 1 mM BA or vehicle (DMSO). Growth curves were generated via a BioTek Synergy HTX plate reader equipped with Gen5 imaging software (version 3.10, Agilent). Plates were incubated at 37 °C under aerobic or anaerobic conditions, as described above, with 205 cycles per min (cpm) orbital shaking. OD<sub>600</sub> measured every 15 min for 24 h. Measurements were blank-corrected and subsequent growth curve analyses were performed in R (60) and R studio (version 2023.06.2+561, Posit). Growth curves were analyzed using the 'growthcurver' R package (version 0.3.1) (61) and comparisons were drawn between fold-change differences in the logarithmic auc. ED<sub>50</sub> values in were determined using terminal OD<sub>600</sub> after growing *L. aerotolerans* or *P. anaerobius* in RCM supplemented with 0-1000 µM PheCA, TyrCA, LeuCA, or CA for 24 h.

### 4.6.2 - Ethics statement

All mouse experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at Michigan State University. Animal health was routinely assessed by laboratory technicians as well as the Michigan State University veterinary staff.

### 4.6.3 - Animals, housing, bile acid dosing, and sample collection

C57BL/6J mice were purchased from Jackson Laboratories (Bar Harbor, ME) and acclimated in the new facility for 1 week prior to bile acid administration via oral gavage.

Cage changes were performed weekly in a laminar flow hood by core facility staff. Mice were housed under a 12-h cycle of light and darkness. Male and female 6-week-old C57BL/6J mice ( $n = 5$  per sex, per group) were administered  $100 \text{ mg kg}^{-1}$  TCA, PheCA, or SerCA dissolved in corn oil via daily oral gavage for 14 days. A control group was administered corn oil alone (vehicle). Treatments were randomized upon receiving the mice. Longitudinal fecal samples and weights were collected from individual mice daily throughout the duration of treatment. On day 13, mice were fasted for approximately 12 h to clear fecal material from the gut prior to necropsy and tissue collection. Animals used for necropsy and tissue collection were euthanized humanely via anesthesia using isoflurane followed by cervical dislocation. Prior to analysis, phosphate buffered saline (Sigma) was added 3:1 (v:w) to fecal samples while 200  $\mu\text{L}$  was added to cecum samples followed by homogenization via bead bashing at maximum speed for 10 min using a Bead Ruptor 96 (Omni International Inc., Kennesaw, GA).

Male and female C57BL/6Crl mice were bred in-house and pups were weaned at 3 weeks of age. At P38, mice were singly housed in cages lined with paper towels, and weights were recorded. Mice were allowed to acclimate to singly housed environments for 3 days. After this acclimation period, paper towel lining was replaced, and mice were fasted for 12 h prior to training (P41). The next day, mice were given a plain peanut butter pellet and were observed to determine whether they would consume the pellet. Normal chow was returned to the cages and all mice that consumed the pellet were advanced to the next stage of training in which they were given a plain peanut butter pellet at approximately the same time of day for 3 consecutive days. On the third day of non-fasted training (P45), the mice were again weighed. All mice that successfully completed training

(consumed the plain peanut butter training pellet within 1 h) were included in subsequent studies. The pilot mixed MCBA dosing experiment lasted for 5 days (starting at P46, day 0). Each day, mice were weighed and treated with either mock peanut butter pellets or pellets containing 10 mg kg<sup>-1</sup> of each of the following for a total dose of 80 mg kg<sup>-1</sup>: AlaCA, AspCA, GluCA, LeuCA, PheCA, SerCA, ThrCA, and TyrCA. Treatments were randomized for mice that successfully completed PBFM training. Fecal samples were collected on days 0, 1, 3, and 5. Paper towel linings were replaced to refresh cages on days 0 and 3. On day 5 (P51), mice were euthanized humanely via anesthesia using isoflurane followed by cervical dislocation. Tissue samples were collected, flash frozen in liquid nitrogen immediately after harvest, and stored at -80 °C prior to further analysis.

The experimental phase of the individual MCBA study including equimolar amino acid-BA controls lasted for 10 days (starting at P46, day 0). Mice were administered MCBA using the Peanut Butter Feeding Method (first developed by Zapata *et al*<sup>29</sup>, see information), a means of administering hydrophobic compounds to mice in a palatable and controlled manner. Each day, mice were weighed and fed peanut butter pellets containing 10 mg kg<sup>-1</sup> of the following: TCA, equimolar taurine and CA (Taur+CA), SerCA, equimolar serine and CA (Ser+CA), PheCA, equimolar phenylalanine and CA (Phe+CA), or a vehicle containing just peanut butter (5 mice per sex per group). Fecal samples were collected on days 0, 1, 4, 7 and 10. Paper towel linings were replaced to refresh cages on days 0, 3, 6 and 9. On day 10 (P56), mice were euthanized humanely via anesthesia using isoflurane followed by cervical dislocation. Tissue samples were collected and flash frozen in liquid nitrogen immediately after harvest.

All mice were housed under a 12-h cycle of light and darkness and treatments, weights, fecal collections, and paper towel changes were conducted in a laminar flow hood. Calculations for dosage for treatments were done using weights at P45. Prior to analysis, phosphate buffered saline was added 3:1 (v:w) to fecal samples and 5:1 (v:w) to tissue samples followed by homogenization via bead bashing at 20 s<sup>-1</sup> for 30 s with 1 min of rest 3 times using a Bead Ruptor 96 (Omni International, Inc., Kennesaw, GA).

#### 4.6.4 - *Untargeted metabolomics for bile acid analysis*

Metabolite extracts were diluted 1:1 (v:v) in 50% methanol (v:v, water) prior to LC-MS/MS analysis. Ultra-high performance liquid chromatography (UPLC) was performed using a Vanquish Autosampler (Thermo) and separation was achieved using an Acquity ultra-performance liquid chromatography (UPLC) bridged ethyl hybrid (BEH) C18 column, 2.1 mm × 100 mm (Waters). All analysis used a 10 µL injection volume, 0.4 mL min<sup>-1</sup> flow rate, and 60 °C column temperature. Samples were eluted using a linear solvent gradient of water (A) and acetonitrile (B), each containing 0.1% formic acid, across a 12-min chromatographic run as follows: 0-1 min, 2% B; 1-8 min, 2-100% B; 8-12 min, 100% B; 10-12 min, 2% B. Mass spectrometry was performed using a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo). Data were collected using electrospray ionization in positive mode. MS<sup>1</sup> data were collected using a 35,000 resolution, automatic gain control (AGC) target of 1×10<sup>6</sup>, maximum injection time of 100 ms, and a scan range set from 100 to 1500 *m/z* during min 1-10. Data-dependent MS<sup>2</sup> spectra were collected for the top 5 most abundant peaks identified in MS<sup>1</sup> survey scans. Resulting raw data files were converted to mzXML format via GNPS Vendor Conversion prior to data mining using MZmine3 (version 3.2.8) (62, 63). Outputs were submitted the

Global Natural Products Social Molecular Networking Database (GNPS, [gnps.ucsd.edu](http://gnps.ucsd.edu)) for spectral annotation and molecular networking (64, 65). Due to the large number of unknown features present, SIRIUS structural prediction and molecular classification was employed for metabolites with  $m/z$  less than 850 Da (66).

#### *4.6.5 - Sample collection from patients undergoing sleeve gastrectomy, processing and analysis*

This prospective, single-arm study enrolled 44 obese patients participating in our health system's (Corewell Health (formerly Beaumont Health), Royal Oak, MI) bariatric surgery program and planning on sleeve gastrectomy (SG). This study was approved by the Beaumont Institutional Review Board (IRB no. 2017-201) and reviewed by the Michigan State University IRB (STUDY00003064). All participants provided informed consent before participating in the study. Participants were recruited from the Royal Oak Weight Control Center, an affiliate of the Corewell Health William Beaumont University Hospital, during the pre-operative bariatric surgery process. This involves a medical work-up, surgical risk stratification and multidisciplinary education before moving on to surgery. Information about the study was presented at the free informational bariatric surgery seminar which prospective patients attend before starting the bariatric surgery program. Fliers about the trial were posted and distributed at the Beaumont Weight Control Center and mailed to patients with the bariatric surgery approval letter. The approval letter is written by a Weight Control Center physician and indicates that a patient is approved from a medical and multidisciplinary team perspective to move forward with bariatric surgery. Inclusion criteria followed the National Institutes of Health criteria for bariatric surgery: BMI at or above  $40 \text{ kg m}^{-2}$  or a BMI of  $35\text{--}40 \text{ kg m}^{-2}$  with an obesity comorbidity such

as type 2 diabetes, heart disease or obstructive sleep apnoea<sup>58</sup>. A further inclusion requirement was being between 18 and 70 years old. Patients were excluded if they had poorly controlled medical or psychiatric conditions which, in the opinion of the investigator, made the patient unlikely to be able to properly participate in the study. Biases include self-selection bias and that only patients planning the SG bariatric surgical procedure were included. It is unlikely that these biases had a significant impact on the results of this study.

Fecal specimens were provided by the William Beaumont Research Institute biorepository. Demographic data were collected, weight and height were measured and BMI was calculated on enrolment into the bariatric surgery program. Weight was measured again on the morning of SG surgery and at 3 months post-SG. Fecal samples were collected pre-operatively and 3 months following SG.

Fecal samples were extracted 1:5 (w:v) in 70% HPLC-grade ice-cold methanol and BA concentrations were calculated on the basis of targeted analysis described above. One fecal sample was lost during extraction, therefore resulting in a cohort of 44 subjects with paired fecal samples before and after SG. The extracts were spun in a microcentrifuge at 12,000g to pellet protein and the methanol supernatant was diluted 1:1 (v:v) in 50% methanol before mass spectrometry analysis. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) protocols were the same as described above for all untargeted metabolomics analysis of microbial and mouse samples. Data were processed with MZmine and GNPS feature-based molecular networking as previously described.

#### *4.6.6 - Targeted metabolomics for bile acid quantification*

For mouse and human samples, BAs were quantified by running a standard mix of known concentrations of various BAs (Table 4.5) as an eight-point standard curve using the same LC–MS/MS as described above. The standard curve samples were added to the end of the run for all samples for which quantification was applied. Data were processed with MZmine3 (v.3.3.0) (62) to obtain AUC abundance for BAs present in both standards and analyzed samples. BA concentrations were determined on the basis of the equation of the curve that fit either a linear or power function, depending on the ionization behavior of individual BAs. Because many MCBAs were detected for which there were no available standards, we used the standard curve from AlaCA to calculate pseudo-concentrations of these compounds.

#### *4.6.7 - DNA isolation and bacterial 16S rRNA amplicon gene sequencing*

DNA from mouse feces and tissue was extracted using the Quick-DNA Faecal/Soil Microbe Miniprep kit (Zymo) according to the manufacturer's instructions. To test extraction efficacy, full-length 16S ribosomal RNA genes were amplified using primers 27f and 1492r and analyzed through gel electrophoresis. Subsequent microbiome sequencing was performed using Illumina compatible primers 515f and 806r to amplify the V4 hypervariable region of the 16S rRNA gene. Sequencing was performed at the Michigan State University RTSF Genomics Core following the protocol previously described (67). PCR products were batch normalized through SequalPrep DNA Normalization plate (Invitrogen) and product recovered from the plates was pooled. This pool was concentrated and cleaned up using a QIAquick Spin column (Qiagen) and AMPure XP magnetic beads (Beckman Coulter). Quality was checked and quantified

using a combination of Qubit dsDNA HS (Thermo Fisher Scientific), 4200 TapeStation HS DNA1000 (Agilent) and Colibri Illumina Library Quantification qPCR assays (Invitrogen). This pool was loaded onto a MiSeq v.2 standard flow cell and sequencing was carried out in a 2 × 250 base pair paired end format using a MiSeq v.2 500 cycle reagent cartridge. Custom sequencing and index primers complementary to the 515f/806r oligomers were added to appropriate wells of the reagent cartridge. Base calling was done by Real Time Analysis v.1.18.54 (RTA, Illumina) and output of RTA was demultiplexed and converted to FastQ format with Bcl2fastq v.2.20.0 (Illumina).

Raw sequences were analyzed using Qiita, a web-based QIIME 2 analysis platform (68, 69). Sequences were filtered on the basis of quality to generate amplicon sequence variants through the Deblur method (70). Taxonomy was assigned using the q2-feature-classifier against the 99% SILVA 16S rRNA gene sequence database (release 138) (71, 72). Sample data were rarefied to 8,000 reads per sample and core diversity metrics, such as Bray–Curtis dissimilarity, were calculated. We performed statistical analysis in R and random forest classification was performed using the ‘randomforest’ package (v.4.7-1.1) (60, 73, 74).

#### *4.6.8 - Sample collection within the ABC baby cohort*

Dual-headed fecal swabs, collected from infant-mother pairs, were separated for metabolite and DNA extraction. Metabolite extractions were performed in 96 deep-well plates (Thermo). One swab head was inserted into 600 µL cold methanol, the plate then sealed with a rubber mat, and incubated at 4 °C overnight. Swab heads were then removed. Plates containing the resulting metabolite extracts were centrifuged for 10 min

at 4,100 g to pellet cell debris followed by storage at -80 °C prior to liquid chromatograph-tandem mass spectrometry analysis (LC-MS/MS).

#### *4.6.9 - ABC baby data quality control and pre-processing*

Data processing and statistical analysis were performed in R. Metabolome samples were filtered to only include samples collected using swabs with wooden handles due to overwhelming signal from plastic-handled swabs severely impacting subsequent analysis (data not shown). Additionally, samples were collected at two weeks following birth, but were not included in this analysis due to a lack of available metadata. Microbiome data were initially rarefied to 5000 counts without replacement and samples unable to meet that threshold were excluded from further analysis. Resulting metabolome and microbiome data are featured with zero-inflation. Although our model could handle zero-inflated data, the molecules with excessive zeros provide little valuable information in the study. Therefore, we removed molecules with proportions of non-zeros less than 10% among samples. Then, to make different omics data more “integrable”, we normalized the data by sum. To ensure enough sample sizes to compute the expected changes for each subject, we removed the subjects with less than three samples (subjects with samples collected at less than three different timepoints).

#### *4.6.10 - Statistics*

Initial bacterial 16S amplicon data diversity metrics were calculated using Qiita (68). Unless otherwise mentioned, statistical analyses were performed in R (v.4.2.2) (60) using RStudio (v. 2023.12.0+369, Posit). Data normalization, alpha and beta diversity calculation, and permutational analysis of variance (PERMANOVA) tests were performed using the `adonis2` function from the `vegan` R package (v.2.6-4) (75), using 999

permutations. Additional statistics tests were performed using the `rstatix` package (v0.7.2) (76). Results were visualized using `ggplot2` (v.3.4.4) (77, 78) and `ggpubr` (v.0.6.0) (79) packages.

Univariate analyses of infant fecal metabolome and microbiome structure was performed using the `envfit` function from the `vegan` package, using 10,000 permutations, with stratification based on infant age (timepoint). Metadata variables were fit onto PCoA ordinations calculated based on Bray-Curtis dissimilarity(80). Resulting P values were then adjusted using the Benjamini-Hochberg false discovery rate (FDR) method(81). Calculating PERMANOVA statistics for infant metabolome and microbiome datasets was performed using the following formula:

$$features \sim (timepoint * bmilkstop) * antibiotic$$

#### 4.6.11 - Mixed-effects longitudinal logistic regressions

Models were fitted using the function `glmer` from package `lme4` (version 1.1-34) (82). For each metabolite or ASV, we have its probability of presence  $\pi_{ij}$  and corresponding collecting age  $t_{ij}$  for subject  $i = 1, \dots, n$  at time point  $j = 1, \dots, n_i$ . We fitted the mixed-effects longitudinal logistic regression models:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = (\mu + \mu_i) + t_{ij}(\beta_T + \beta_{Ti}) + F_i\beta_F + B_i\beta_B + H_i\beta_H + O_i\beta_O,$$

$$\mu_i \sim N(0, \sigma_\mu^2), \beta_{Ti} \sim N(0, \sigma_{\beta_T}^2)$$

where  $\mu$  and  $\beta_T$  are the fixed intercept and time slope,  $\mu_i$  and  $\beta_{Ti}$  are the random intercepts and time slopes. We also include a dummy variable for female ( $F_i$ ) and dummy variables for black ( $B_i$ ), Hispanic ( $H_i$ ), and others ( $O_i$ ) to account for sex and ancestry differences.

#### 4.6.12 - Evaluation of model prediction with testing data

We further examined models' prediction performance in cross-validations and compared it with standard logistic regression models (without subject-specific random effects on the intercepts,  $\mu_i$ , and slopes,  $\beta_{Ti}$ ). To estimate prediction accuracy, for each metabolite or ASV, we randomly selected one sample from each subject to form the testing set. The remaining samples were used as the training set to fit the models (mixed-effects and standard logistic regression models) and predict the left-out testing samples. We then computed the area under the ROC curve (AUC, using the function `auc` from package `pROC` (version 1.18.0) (83) between the predicted probabilities and the observed presence/absence outcome in the testing data. We repeated the above validation process 20 times and computed the average of the AUC for each metabolite and ASV. Using these validation results, we filtered out metabolites and ASVs that had an AUC that was not significantly greater than 0.5 ( $P < 0.05$ , for our sample size, that corresponds to an  $AUC \geq 0.65$ ). Therefore, all the longitudinal trajectories that we report, as well as the cross-subject correlation and enrichment analysis that followed are based on ASVs and metabolites with testing  $AUC \geq 0.65$ .

#### 4.6.13 - Prediction of longitudinal trajectories

We used the fitted mixed-effects longitudinal logistic regression models to predict the expected change in the probability of detection for each metabolite and ASV. At the population level, the expected changes in the proportion of zeros between two different time points ( $t_2 > t_1$ ) are (here we predicted for sex and ancestry indices - male, white):

$$\Delta P = \left[ 1 - \frac{1}{1 + e^{-(\mu + t_2 \beta_T)}} \right] - \left[ 1 - \frac{1}{1 + e^{-(\mu + t_1 \beta_T)}} \right],$$

And the expected changes in the proportion of zeros between two different time points at the individual level are:

$$\Delta P_i = \left\{ 1 - \frac{1}{1 + e^{-[(\mu + \mu_i) + t_2(\beta_T + \beta_{T_i}) + \tau_i]}} \right\} - \left\{ 1 - \frac{1}{1 + e^{-[(\mu + \mu_i) + t_1(\beta_T + \beta_{T_i}) + \tau_i]}} \right\},$$

$$\tau_i = F_i\beta_F + B_i\beta_B + H_i\beta_H + O_i\beta_O.$$

where  $i = 1, \dots, n$  is an index for the subject. The curves and predicted changes we report is for 12 months during the first year of life (since our first samples were collected at month 2, we used  $t_1 = 2$  and  $t_2 = 14$ ).

#### 4.6.14 - Bootstrap analysis of Pearson correlation

To identify metabolite-ASV pairs that were changing in a seemingly coordinated fashion, we correlated the vectors containing the  $\Delta P'_i$ 's for all subjects for a metabolite with that of a vector containing the  $\Delta P'_i$ 's for an ASV. We did this for all metabolite-ASV pairs. For the sample size that we have, Pearson's correlation can be highly affected by outliers; therefore, to smooth-out the influence of outliers we reported an (approximately unbiased) Bootstrap estimate for the correlation coefficients (84)

$$\hat{r} = \frac{1}{B} \sum_{b=1}^B r_b \left[ 1 + \frac{1 - r_b^2}{2(n-3)} \right]$$

where  $r_b$  is the traditional Pearson correlation coefficient for Bootstrap sample  $b$ . A standard error for these estimates was computed using the Bootstrap sample and t-tests were used to determine if the correlation was significantly different from 0.

#### 4.6.15 - Enrichment analysis

With the taxonomy grouping information for each metabolite and ASV, as well as the above results indicating whether each of these molecules had a significant longitudinal change, we used hypergeometric tests to identify groups of metabolites or

ASVs that change significantly over time. These taxonomic groups with expected relative abundance change coordinatively over time were used to form a network.

#### 4.7 - Data availability

Raw mass spectrometry data are publicly available in the MassIVE database ([massive.ucsd.edu](http://massive.ucsd.edu)) for MCBA gavage samples under MSV000093173 ([doi.org/10.25345/C57S7J35N](https://doi.org/10.25345/C57S7J35N)), for mixed MCBA PBFM dosing under MSV000093171 ([doi.org/10.25345/C5H98ZQ3R](https://doi.org/10.25345/C5H98ZQ3R)), for 100 mg<sup>-1</sup> kg of SerCA PBFM dosing at MSV000093169 ([doi.org/10.25345/C5RV0DB2C](https://doi.org/10.25345/C5RV0DB2C)), 10 mg<sup>-1</sup> kg of MCBA PBFM dosing under MSV000093172 ([doi.org/10.25345/C5CJ87W9C](https://doi.org/10.25345/C5CJ87W9C)) and for SG fecal samples under MSV000093167 ([doi.org/10.25345/C51834C9N](https://doi.org/10.25345/C51834C9N)). Raw mass spectrometry data for the ABC baby cohort are available under MassIVE ID MSV000092782 ([doi.org/10.25345/C5DJ58S9M](https://doi.org/10.25345/C5DJ58S9M)).

GNPS molecular networks are available for SG fecal samples at [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f11eaab1cf1d43b1a5f754575d171e87](http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f11eaab1cf1d43b1a5f754575d171e87) and for the ABC Baby cohort at [gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7454748a6baa406b909540b1c90a4e7e](http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7454748a6baa406b909540b1c90a4e7e).

16S rRNA gene amplicon data were deposited in the EMBL-EBI European Nucleotide Archive. Data from the 100 mg kg<sup>-1</sup> gavage experiment can be found under project PRJEB68000, study accession ERP153011. Tissue data from the 10 mg kg<sup>-1</sup> PBFM experiment can be found under project PRJEB68146, study accession ERP153132. Fecal data available from the 10 mg kg<sup>-1</sup> PBFM experiment are available under project PRJEB68149, study accession ERP153135. Fecal data from the ABC Baby cohort can be found under project PRJEB72674, study accession ERP157451. Analyses

can be found on Qiita under analysis ID 53128 for the 100 mg kg<sup>-1</sup> gavage and IDs 57407 and 57481 for tissue and fecal samples, respectively, from the 10 mg kg<sup>-1</sup> PBFM experiment. Analysis for the ABC Baby cohort can be found under analysis ID 48437.

## REFERENCES

1. Sung JY, Shaffer EA, Costerton JW. 1993. Antibacterial activity of bile salts against common biliary pathogens. *Dig Dis Sci* 38:2104–2112.
2. Sannasiddappa TH, Lund PA, Clarke SR. 2017. *In vitro* antibacterial activity of unconjugated and conjugated bile salts on *Staphylococcus aureus*. *Front Microbiol* 8:1581.
3. Theriot CM, Bowman AA, Young VB. 2016. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. *mSphere* 1:e00045-15.
4. Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A, Antonopoulos DA, Jabri B, Chang EB. 2012. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10<sup>-/-</sup>* mice. *Nature* 487:104–108.
5. Von Schwartzberg RJ, Bisanz JE, Lyalina S, Spanogiannopoulos P, Ang QY, Cai J, Dickmann S, Friedrich M, Liu S-Y, Collins SL, Ingebrigtsen D, Miller S, Turnbaugh JA, Patterson AD, Pollard KS, Mai K, Spranger J, Turnbaugh PJ. 2021. Caloric restriction disrupts the microbiota and colonization resistance. *Nature* 595:272–277.
6. Li Y, Yang X, Zhang J, Jiang T, Zhang Z, Wang Z, Gong M, Zhao L, Zhang C. 2021. Ketogenic diets induced glucose intolerance and lipid accumulation in mice with alterations in gut microbiota and metabolites. *mBio* 12:e03601-20.
7. Liu Y, Zhong W, Li X, Shen F, Ma X, Yang Q, Hong S, Sun Y. 2023. Diets, gut microbiota and metabolites. *Phenomics* 3:268–284.
8. Wan Y, Yuan J, Li J, Li H, Zhang J, Tang J, Ni Y, Huang T, Wang F, Zhao F, Li D. 2020. Unconjugated and secondary bile acid profiles in response to higher-fat, lower-carbohydrate diet and associated with related gut microbiota: a 6-month randomized controlled-feeding trial. *Clin Nutr* 39:395–404.
9. Mueller NT, Whyatt R, Hoepner L, Oberfield S, Dominguez-Bello MG, Widen EM, Hassoun A, Perera F, Rundle A. 2015. Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *Int J Obes* 39:665–670.
10. Moore RE, Townsend SD. 2019. Temporal development of the infant gut microbiome. *Open Biol* 9:190128.
11. Yang I, Corwin EJ, Brennan PA, Jordan S, Murphy JR, Dunlop A. 2016. The infant microbiome: implications for infant health and neurocognitive development. *Nurs Res* 65:76–88.
12. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni HV, Metcalf GA, Muzny D, Gibbs RA, Vatanen T, Huttenhower C, Xavier RJ, Rewers M, Hagopian W, Toppari J, Ziegler AG, She JX,

- Akolkar B, Lernmark A, Hyoty H, Vehik K, Krischer JP, Petrosino JF. 2018. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562:583–588.
13. Ward JBJ, Lajczak NK, Kelly OB, O'Dwyer AM, Giddam AK, Ní Gabhann J, Franco P, Tambuwala MM, Jefferies CA, Keely S, Roda A, Keely SJ. 2017. Ursodeoxycholic acid and lithocholic acid exert anti-inflammatory actions in the colon. *Am J Physiol - Gastrointest Liver Physiol* 312:G550–G558.
  14. Hofmann AF. 1999. The continuing importance of bile acids in liver and intestinal disease. *Arch Intern Med* 159:2647–2658.
  15. Han X, Wang J, Wu Y, Gu H, Zhao N, Liao X, Jiang M. 2023. Predictive value of bile acids as metabolite biomarkers for gallstones: A protocol of systematic review and meta-analysis. *PLOS ONE* 18:e0284138.
  16. Puri P, Daita K, Joyce A, Mirshahi F, Santhekadur PK, Cazanave S, Luketic VA, Siddiqui MS, Boyett S, Min H, Kumar DP, Kohli R, Zhou H, Hylemon PB, Contos MJ, Idowu M, Sanyal AJ. 2018. The presence and severity of nonalcoholic steatohepatitis is associated with specific changes in circulating bile acids. *Hepatology* 67:534–548.
  17. Jiao N, Baker SS, Chapa-Rodriguez A, Liu W, Nugent CA, Tsompana M, Mastrandrea L, Buck MJ, Baker RD, Genco RJ, Zhu R, Zhu L. 2018. Suppressed hepatic bile acid signalling despite elevated production of primary and secondary bile acids in NAFLD. *Gut* 67:1881–1891.
  18. Luo L, Aubrecht J, Li D, Warner RL, Johnson KJ, Kenny J, Colangelo JL. 2018. Assessment of serum bile acid profiles as biomarkers of liver injury and liver disease in humans. *PLOS ONE* 13:e0193824.
  19. Dasarathy S, Yang Y, McCullough AJ, Marczewski S, Bennett C, Kalhan SC. 2011. Elevated hepatic fatty acid oxidation, high plasma fibroblast growth factor 21, and fasting bile acids in nonalcoholic steatohepatitis. *Eur J Gastroenterol Hepatol* 23:382–388.
  20. Aranha MM, Cortez-Pinto H, Costa A, Da Silva IBM, Camilo ME, De Moura MC, Rodrigues CMP. 2008. Bile acid levels are increased in the liver of patients with steatohepatitis. *Eur J Gastroenterol Hepatol* 20:519–525.
  21. Cao H, Xu M, Dong W, Deng B, Wang S, Zhang Y, Wang S, Luo S, Wang W, Qi Y, Gao J, Cao X, Yan F, Wang B. 2017. Secondary bile acid-induced dysbiosis promotes intestinal carcinogenesis. *Int J Cancer* 140:2545–2556.
  22. Cook JW, Kennaway EL, Kennaway NM. 1940. Production of tumours in mice by deoxycholic acid. *Nature* 145:627–627.

23. Mahmoudian Dehkordi S, Arnold M, Nho K, Ahmad S, Jia W, Xie G, Louie G, Kueider-Paisley A, Moseley MA, Thompson JW, St John Williams L, Tenenbaum JD, Blach C, Baillie R, Han X, Bhattacharyya S, Toledo JB, Schafferer S, Klein S, Koal T, Risacher SL, Allan Kling M, Motsinger-Reif A, Rotroff DM, Jack J, Hankemeier T, Bennett DA, De Jager PL, Trojanowski JQ, Shaw LM, Weiner MW, Doraiswamy PM, Van Duijn CM, Saykin AJ, Kastenmüller G, Kaddurah-Daouk R, for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium. 2019. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease—An emerging role for gut microbiome. *Alzheimers Dement* 15:76–92.
24. Nho K, Kueider-Paisley A, Mahmoudian Dehkordi S, Arnold M, Risacher SL, Louie G, Blach C, Baillie R, Han X, Kastenmüller G, Jia W, Xie G, Ahmad S, Hankemeier T, Van Duijn CM, Trojanowski JQ, Shaw LM, Weiner MW, Doraiswamy PM, Saykin AJ, Kaddurah-Daouk R, for the Alzheimer's Disease Neuroimaging Initiative and the Alzheimer Disease Metabolomics Consortium. 2019. Altered bile acid profile in mild cognitive impairment and Alzheimer's disease: Relationship to neuroimaging and CSF biomarkers. *Alzheimers Dement* 15:232–244.
25. Dilmore AH, Martino C, Neth BJ, West KA, Zemlin J, Rahman G, Panitchpakdi M, Meehan MJ, Weldon KC, Blach C, Schimmel L, Kaddurah-Daouk R, Dorrestein PC, Knight R, Craft S, Alzheimer's Gut Microbiome Project Consortium. 2023. Effects of a ketogenic and low-fat diet on the human metabolome, microbiome, and foodome in adults at risk for Alzheimer's disease. *Alzheimers Dement* 19:4805–4816.
26. Chong Nguyen C, Duboc D, Rainteau D, Sokol H, Humbert L, Seksik P, Bellino A, Abdoul H, Bouazza N, Treluyer J-M, Saadi M, Wahbi K, Soliman H, Coffin B, Bado A, Le Gall M, Varenne O, Duboc H. 2021. Circulating bile acids concentration is predictive of coronary artery disease in human. *Sci Rep* 11:22661.
27. Qi L, Chen Y. 2023. Circulating bile acids as biomarkers for disease diagnosis and prevention. *J Clin Endocrinol Metab* 108:251–270.
28. Zheng X, Chen T, Zhao A, Ning Z, Kuang J, Wang S, You Y, Bao Y, Ma X, Yu H, Zhou J, Jiang M, Li M, Wang J, Ma X, Zhou S, Li Y, Ge K, Rajani C, Xie G, Hu C, Guo Y, Lu A, Jia W, Jia W. 2021. Hyocholic acid species as novel biomarkers for metabolic disorders. *Nat Commun* 12:1487.
29. Ahmed M, Levy L, Hunter SE, Zhang KC, Huszti E, Boonstra KM, Sage AT, Azad S, Zamel R, Ghany R, Yeung JC, Crespín OM, Frankel C, Budev M, Shah P, Snyder LD, Belperio J, Singer LG, Weigt SS, Todd JL, Palmer SM, Keshavjee S, Martinu T. 2019. Lung bile acid as biomarker of microaspiration and its relationship to lung inflammation. *J Heart Lung Transplant* 38:S255–S256.
30. Bernstein H, Bernstein C, Payne CM, Dvorakova K, Garewal H. 2005. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res Mutat Res* 589:47–65.

31. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. 2011. Carcinogenicity of deoxycholate, a secondary bile acid. *Arch Toxicol* 85:863–871.
32. Jia W, Xie G, Jia W. 2018. Bile acid–microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nat Rev Gastroenterol Hepatol* 15:111–128.
33. Kurdi P, Kawanishi K, Mizutani K, Yokota A. 2006. Mechanism of growth inhibition by free bile acids in *Lactobacilli* and *Bifidobacteria*. *J Bacteriol* 188:1979–1986.
34. Hamilton JP, Xie G, Raufman J-P, Hogan S, Griffin TL, Packard CA, Chatfield DA, Hagey LR, Steinbach JH, Hofmann AF. 2007. Human cecal bile acids: concentration and spectrum. *Am J Physiol-Gastrointest Liver Physiol* 293:G256–G263.
35. Northfield TC, McColl I. 1973. Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine. *Gut* 14:513–518.
36. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Human gut microbes associated with obesity. *Nature* 444:1022–1023.
37. Stojanov S, Berlec A, Štrukelj B. 2020. The influence of probiotics on the *Firmicutes/Bacteroidetes* ratio in the treatment of obesity and inflammatory bowel disease. *Microorganisms* 8:1715.
38. Cox LM, Sohn J, Tyrrell KL, Citron DM, Lawson PA, Patel NB, Izumi T, Perez-Perez GI, Goldstein EJC, Blaser MJ. 2017. Description of two novel members of the family *Erysipelotrichaceae*: *Ileibacterium valens* gen. nov., sp. nov. and *Dubosiella newyorkensis*, gen. nov., sp. nov., from the murine intestine, and emendation to the description of *Faecalibacterium rodentium*. *Int J Syst Evol* 67:1247–1254.
39. Gentry EC, Collins SL, Panitchpakdi M, Belda-Ferre P, Stewart AK, Carrillo Terrazas M, Lu H, Zuffa S, Yan T, Avila-Pacheco J, Plichta DR, Aron AT, Wang M, Jarmusch AK, Hao F, Syrkin-Nikolau M, Vlamakis H, Ananthakrishnan AN, Boland BS, Hemperly A, Vande Castele N, Gonzalez FJ, Clish CB, Xavier RJ, Chu H, Baker ES, Patterson AD, Knight R, Siegel D, Dorrestein PC. 2024. Reverse metabolomics for the discovery of chemical structures from humans. *Nature* 626:419–426.
40. Shalon D, Culver RN, Grembi JA, Folz J, Treit PV, Shi H, Rosenberger FA, Dethlefsen L, Meng X, Yaffe E, Aranda-Díaz A, Geyer PE, Mueller-Reif JB, Spencer S, Patterson AD, Triadafilopoulos G, Holmes SP, Mann M, Fiehn O, Relman DA, Huang KC. 2023. Profiling the human intestinal environment under physiological conditions. *Nature* 617:581–591.
41. Hofmann AF. 2009. The enterohepatic circulation of bile acids in mammals: form and functions. *Front Biosci* 14:2584–2598.
42. Zapata RC, Zhang D, Chaudry B, Osborn O. 2021. Self-Administration of drugs in mouse models of feeding and obesity. *J Vis Exp* 62775.

43. Huijghebaert SM, Hofmann AF. 1988. Influence of the amino acid moiety on deconjugation of bile acid amidates by cholylglycine hydrolase or human fecal cultures. *J Lipid Res* 27:742–752.
44. Huijghebaert SM, Hofmann AF. 1986. Pancreatic carboxypeptidase hydrolysis of bile acid-amino acid conjugates: Selective resistance of glycine and taurine amidates. *Gastroenterology* 90:306–315.
45. Foley MH, Walker ME, Stewart AK, O’Flaherty S, Gentry EC, Patel S, Beaty VV, Allen G, Pan M, Simpson JB, Perkins C, Vanhoy ME, Dougherty MK, McGill SK, Gulati AS, Dorrestein PC, Baker ES, Redinbo MR, Barrangou R, Theriot CM. 2023. Bile salt hydrolases shape the bile acid landscape and restrict *Clostridioides difficile* growth in the murine gut. *Nat Microbiol* 8:611–628.
46. Beller L, Deboutte W, Falony G, Vieira-Silva S, Tito RY, Valles-Colomer M, Rymenans L, Jansen D, Van Espen L, Papadaki MI, Shi C, Yinda CK, Zeller M, Faust K, Van Ranst M, Raes J, Matthijssens J. 2021. Successional stages in infant gut microbiota maturation. *mBio* 12:e01857-21.
47. Guzior DV, Quinn RA. 2021. Review: microbial transformations of human bile acids. *Microbiome* 9:140.
48. Trottier J, Perreault M, Rudkowska I, Levy C, Dallaire-Theroux A, Verreault M, Caron P, Staels B, Vohl M-C, Straka RJ, Barbier O. 2013. Profiling serum bile acid glucuronides in humans: gender divergences, genetic determinants, and response to fenofibrate. *Clin Pharmacol Ther* 94:533–543.
49. Kim HW, Wang M, Leber CA, Nothias L-F, Reher R, Kang KB, van der Hooff JJJ, Dorrestein PC, Gerwick WH, Cottrell GW. 2021. NPClassifier: a deep neural network-based structural classification tool for natural products. *J Nat Prod* 84:2795–2807.
50. Ferrannini E, Camastra S, Astiarraga B, Nannipieri M, Castro-Perez J, Xie D, Wang L, Chakravarthy M, Haeusler RA. 2015. Increased bile acid synthesis and deconjugation after biliopancreatic diversion. *Diabetes* 64:3377–3385.
51. Albaugh VL, Flynn CR, Cai S, Xiao Y, Tamboli RA, Abumrad NN. 2015. Early increases in bile acids post Roux-en-Y gastric bypass are driven by insulin-sensitizing, secondary bile acids. *J Clin Endocrinol Metab* 100:E1225-1233.
52. Han H, Wang L, Du H, Jiang J, Hu C, Zhang G, Liu S, Zhang X, Liu T, Hu S. 2015. Expedited biliopancreatic juice flow to the distal gut benefits the diabetes control after duodenal-jejunal bypass. *Obes Surg* 25:1802–1809.
53. Perreault M, Bialek A, Trottier J, Verreault M, Caron P, Milkiewicz P, Barbier O. 2013. Role of glucuronidation for hepatic detoxification and urinary elimination of toxic bile acids during biliary obstruction. *PLoS One* 8:e80994.

54. Alnouti Y. 2009. Bile acid sulfation: a pathway of bile acid elimination and detoxification. *Toxicol Sci Off J Soc Toxicol* 108:225–246.
55. Feldman AG, Sokol RJ. 2019. Neonatal cholestasis: emerging molecular diagnostics and potential novel therapeutics. *Nat Rev Gastroenterol Hepatol* 16:346–360.
56. Crick PJ, Yutuc E, Abdel-Khalik J, Saeed A, Betsholtz C, Genove G, Björkhem I, Wang Y, Griffiths WJ. 2019. Formation and metabolism of oxysterols and cholestenic acids found in the mouse circulation: Lessons learnt from deuterium-enrichment experiments and the *CYP46A1* transgenic mouse. *J Steroid Biochem Mol Biol* 195:105475.
57. Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG. 2015. The infant microbiome development: Mom matters. *Trends Mol Med* 21:109–117.
58. The CHILD Study Investigators, Tun HM, Konya T, Takaro TK, Brook JR, Chari R, Field CJ, Guttman DS, Becker AB, Mandhane PJ, Turvey SE, Subbarao P, Sears MR, Scott JA, Kozyrskyj AL. 2017. Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios. *Microbiome* 5:40.
59. van Best N, Rolle-Kampczyk U, Schaap FG, Basic M, Olde Damink SWM, Bleich A, Savelkoul PHM, von Bergen M, Penders J, Hornef MW. 2020. Bile acids drive the newborn's gut microbiota maturation. *Nat Commun* 11.
60. R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria.
61. Sprouffske K, Wagner A. 2016. Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* 17.
62. Pluskal T, Castillo S, Villar-Briones A, Orešič M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform* 11.
63. Myers OD, Sumner SJ, Li S, Barnes S, Du X. 2017. One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: New algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks. *Anal Chem* 89:8696–8703.
64. Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard P-M, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouvellec A, Ludwig M, Martin H. C, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweiger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H,

- Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* 17:905–908.
65. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kaponov CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrew K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya CAP, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Lington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837.
  66. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* 16:299–302.
  67. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120.
  68. Gonzalez A, Navas-Molina JA, Kosciulek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods* 15:796–798.
  69. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvall C, Edwards CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C,

- Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolk T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Lottfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Priesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotech* 37:852–857.
70. Amnon A, Daniel M, A N-MJ, Evguenia K, T MJ, Zhenjiang ZX, P KE, R TL, R HE, Antonio G, Rob K, A GJ. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2:e00191-16.
  71. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 41.
  72. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Priesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “all-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42.
  73. Breiman L, Cutler A, Liaw A, Wiener M. 2022. randomForest: Breiman and Cutler’s Random Forests for Classification and Regression.
  74. Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* 2:18–22.
  75. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista HBA, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill MO, Lahti L, McGlenn D, Ouellette M-H, Ribeiro Cunha E, Smith T, Stier A, Ter Braak CJF, Weedon J. 2022. vegan: Community Ecology Package.
  76. Kassambara A. 2021. rstatix: pipe-friendly framework for basic statistical tests.
  77. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D. 2022. ggplot2: create elegant data visualizations using the grammar of graphics.
  78. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
  79. Kassambara A. 2020. ggpubr: ggplot2 based publication ready plots.

80. Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349.
81. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300.
82. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48.
83. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77.
84. Olkin I, Pratt JW. 1958. Unbiased estimation of certain correlation coefficients. *Ann Math Stat* 29:201–211.

## APPENDIX A: SUPPLEMENTARY TABLES

**Table 4.1: Top 30 ASVs contributing to random forest classification of cecal samples by 100 mg kg<sup>-1</sup> MCBA gavage group**

Higher percent mean decrease in accuracy (%MDA) corresponds the ASVs that are more important in the development of the random forest model.

%MDA	Phylum	Class	Order	Family	Genus	Species
11.70	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
9.79	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnoclostridium	
9.74	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
9.44	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Faecalibaculum	uncultured bacterium
8.82	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
8.76	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
8.36	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	
8.35	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculum	uncultured bacterium
8.27	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae		
8.23	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
8.08	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae		
8.04	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
7.98	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae		
7.89	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	
7.86	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
7.83	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
7.68	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium] xylanophilum group	uncultured bacterium
7.53	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
7.48	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	uncultured bacterium
7.38	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae		
7.38	Firmicutes	Clostridia	Peptostreptococcales- Tissierellales	Peptostreptococcaceae	Romboutsia	
7.31	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospiraceae UCG-001	uncultured bacterium
7.27	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	uncultured bacterium
7.26	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae		
7.14	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
7.12	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Dubosiella	Dubosiella newyorkensis

Table 4.1 (cont'd)

7.07	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Faecalibaculum
6.93	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	Oscillibacter
6.91	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia-Shigella
6.88	Firmicutes	Clostridia	Oscillospirales	Oscillospiraceae	

**Table 4.2: Top 30 ASVs contributing to random forest classification of fecal samples by 100 mg kg<sup>-1</sup> MCBA gavage group**

Higher percent mean decrease in accuracy (%MDA) corresponds the ASVs that are more important in the development of the random forest model.

%MDA	Phylum	Class	Order	Family	Genus	Species
28.87	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	
27.44	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	uncultured bacterium
24.07	Bacteroidota	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	uncultured bacterium
23.09	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Faecalibaculum	uncultured bacterium
22.56	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Faecalibaculum	
20.64	Firmicutes	Bacilli	Erysipelotrichales	Erysipelotrichaceae	Dubosiella	Dubosiella newyorkensis
20.55	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
20.25	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	
19.94	Actinobacteriota	Coriobacteriia	Coriobacteriales	Atopobiaceae	Coriobacteriaceae UCG-002	uncultured bacterium
19.68	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	uncultured bacterium
19.53	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
19.31	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
18.69	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	[Eubacterium] xylanophilum group	uncultured bacterium
18.22	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae		
18.09	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
18.03	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
17.49	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnoclostridium	
17.09	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	uncultured bacterium
16.35	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	
15.73	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Enterorhabdus	mouse gut
15.51	Firmicutes	Clostridia	Monoglobales	Monoglobaceae	Monoglobus	
15.11	Firmicutes	Bacilli	RF39	RF39	RF39	
14.68	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculum	uncultured bacterium
14.67	Firmicutes	Bacilli	RF39	RF39	RF39	human gut
14.59	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
14.39	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium
14.24	Bacteroidota	Bacteroidia	Bacteroidales	Muribaculaceae	Muribaculaceae	uncultured bacterium

Table 4.2 (cont'd)

13.21	Firmicutes	Clostridia	Oscillospirales	Ruminococcaceae	Incertae Sedis
13.01	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Dorea
12.43	Firmicutes	Clostridia	Peptostreptococcales- Tissierellales	Peptostreptococcaceae	Romboutsia

**Table 4.3: Summary of previously reported MCBA concentrations in murine and human samples**

‡Total MCBA molar concentrations were estimated using the mass of glutamate conjugated cholic acid (glutamatocholic acid, GluCA; 537 Da).

Study	Bile Acid	Concentration
Gentry <i>et al.</i> (preprint) Population: Healthy adults doi: 10.21203/rs.3.rs-820302/v1	PheCA	1 $\mu\text{g kg}^{-1}$ feces (~1.8 $\mu\text{M}$ , maximum 11 $\mu\text{M}$ )
Shalon <i>et al.</i> (2023) Population: Healthy adults doi: 10.1038/s41586-023-05989-7	Total MCBAs	3,000-10,000 ng/mL (5.55-18.5 $\mu\text{M}^\ddagger$ ) Average 4,000 ng/mL (7.4 $\mu\text{M}^\ddagger$ )
	GlnCA	1,000 ng/mL (1.86 $\mu\text{M}$ , highest individual concentration)
This work Population: Patients undergoing sleeve gastrectomy	Total MCBAs	77.7 $\mu\text{M}$ (range: 0-724 $\mu\text{M}$ )
	Individual MCBA	Average 9.77 $\mu\text{M}$ (range: 0.307-359 $\mu\text{M}$ ) across MCBAs present
	GluCA	Average 15.1 $\mu\text{M}$ (range: 0-359 $\mu\text{M}$ )
	GluDCA	Average 20.1 $\mu\text{M}$ (range: 0-263 $\mu\text{M}$ )

**Table 4.4: BA concentrations in murine tissue and feces following 10 mg kg<sup>-1</sup> MCBA dosing via PBFM**

All concentrations are presented as micromolar concentrations. Data are presented as mean (s.e.m.), *n* = 5 male and 5 female per treatment group

	Group	CA	GCDCA	GCA	Ile/LeuCA	PheCA	SerCA	TMCA	TCA	ThrCA	TyrCA
Cecum	Mock	6.6 (1.9)	0 (0)	1.2 (0.5)	0 (0)	0 (0)	0 (0)	39 (13)	26 (8)	0 (0)	0 (0)
	Phe+CA	19 (7)	0 (0)	1.7 (0.5)	0 (0)	0 (0)	0 (0)	77 (42)	28 (14)	0 (0)	0 (0)
	PheCA	17 (8)	0 (0)	1.8 (0.5)	0 (0)	0 (0)	0 (0)	93 (60)	64 (39)	0 (0)	0 (0)
	Ser+CA	5.7 (1.3)	0 (0)	1.8 (0.4)	0 (0)	0 (0)	0 (0)	30 (14)	14 (7)	0.25 (0.25)	0 (0)
	SerCA	15 (7)	0 (0)	1.6 (0.6)	0 (0)	0 (0)	0.36 (0.16)	50 (12)	25 (10)	0.32 (0.32)	0 (0)
	Taur+CA	61 (40)	0 (0)	2.1 (0.5)	0 (0)	0 (0)	0 (0)	120 (50)	120 (71)	0 (0)	0 (0)
	TCA	11 (5)	0 (0)	2.0 (0.4)	0 (0)	0 (0)	0 (0)	30 (14)	13 (10)	0.27 (0.27)	0 (0)
Colon	Mock	19 (6)	0 (0)	0.07 (0.07)	0 (0)	0 (0)	0 (0)	160 (60)	40 (19)	0 (0)	0 (0)
	Phe+CA	38 (13)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	140 (80)	45 (20)	0 (0)	0 (0)
	PheCA	32 (22)	0 (0)	0.34 (0.34)	0 (0)	0 (0)	0 (0)	160 (93)	19 (19)	0 (0)	0 (0)
	Ser+CA	53 (23)	0 (0)	2.0 (2.0)	0 (0)	0 (0)	0 (0)	2100 (1900)	19 (12)	0 (0)	0 (0)
	SerCA	68 (37)	0 (0)	0.082 (0.054)	0 (0)	0 (0)	1.4 (0.8)	180 (70)	15 (13)	0 (0)	0 (0)
	Taur+CA	68 (42)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	190 (90)	52 (31)	0.031 (0.031)	0 (0)
	TCA	14 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	81 (13)	20 (17)	0 (0)	0 (0)
Duodenum	Mock	41 (15)	0 (0)	5.8 (0.9)	0 (0)	0 (0)	0 (0)	6400 (700)	16000 (2000)	0 (0)	0 (0)
	Phe+CA	61 (23)	0 (0)	5.6 (2.6)	0 (0)	0 (0)	0.063 (0.063)	5300 (800)	16000 (3000)	0 (0)	0 (0)
	PheCA	110 (40)	0 (0)	9.8 (3.0)	0 (0)	2.4 (1.5)	0.074 (0.051)	7900 (1400)	16000 (2000)	0 (0)	0 (0)
	Ser+CA	68 (25)	0 (0)	4.2 (1.2)	0 (0)	0 (0)	0.084 (0.045)	5500 (700)	15000 (2000)	0 (0)	0.10 (0.10)
	SerCA	118 (46)	0 (0)	8.0 (1.7)	0 (0)	0.059 (0.059)	80 (12)	6500 (900)	17000 (2000)	0 (0)	0 (0)
	Taur+CA	69 (15)	0 (0)	5.6 (2.4)	0 (0)	0 (0)	0.093 (0.063)	4600 (900)	12000 (2000)	0 (0)	0 (0)

Table 4.4 (cont'd)

	TCA	39 (10)	0 (0)	5.5 (0.8)	0 (0)	0 (0)	0.92 (0.59)	5700 (700)	17000 (2000)	0 (0)	0 (0)
Gall bladder	Mock	80 (46)	0 (0)	21 (11)	0 (0)	0 (0)	0.60 (0.60)	9100 (2600)	5300 (2200)	0.97 (0.97)	0 (0)
	Phe+CA	320 (10)	0 (0)	57 (11)	0 (0)	0.91 (0.91)	0.25 (0.17)	15000 (3000)	600 (400)	3.3 (1.0)	0 (0)
	PheCA	280 (180)	0 (0)	37 (8.4)	0 (0)	42 (13)	0.29 (0.18)	16000 (3000)	8700 (5600)	3.4 (1.7)	0 (0)
	Ser+CA	350 (90)	0 (0)	59 (17)	0 (0)	0 (0)	0.62 (0.41)	20000 (3000)	440 (300)	0 (0)	0 (0)
	SerCA	710 (490)	0.77 (0.77)	81 (29)	0 (0)	1.0 (1.0)	360 (80)	13000 (3000)	1600 (1300)	0.91 (0.60)	0 (0)
	Taur+CA	320 (140)	0.38 (0.38)	48 (14)	0 (0)	0 (0)	1.1 (0.4)	10000 (4000)	1500 (1100)	1.0 (1.0)	0 (0)
	TCA	94 (45)	0 (0)	31 (13)	0 (0)	0 (0)	4.7 (3.6)	12000 (4000)	9900 (5900)	0.83 (0.83)	0 (0)
Ileum	Mock	59 (43)	0 (0)	3.0 (0.9)	0.13 (0.06)	0 (0)	0 (0)	810 (340)	1400 (600)	0 (0)	0 (0)
	Phe+CA	130 (107)	0 (0)	3.7 (1.8)	0.17 (0.07)	0 (0)	0 (0)	990 (520)	2400 (1300)	0.094 (0.094)	0 (0)
	PheCA	280 (130)	0 (0)	4.0 (1.1)	0.13 (0.06)	0 (0)	0 (0)	860 (340)	1800 (900)	0.047 (0.047)	0 (0)
	Ser+CA	290 (110)	0 (0)	5.3 (1.8)	0.61 (0.41)	0 (0)	0 (0)	1700 (800)	1900 (900)	0.042 (0.042)	0 (0)
	SerCA	565 (300)	0 (0)	13 (9)	0.20 (0.08)	0 (0)	46 (29)	2700 (1400)	3700 (1700)	0.15 (0.12)	0.10 (0.10)
	Taur+CA	140 (60)	0 (0)	3.3 (1.1)	0.10 (0.06)	0 (0)	0.071 (0.071)	950 (450)	1700 (800)	0.023 (0.023)	0 (0)
	TCA	270 (110)	0 (0)	9.6 (3.2)	0.16 (0.06)	0 (0)	0 (0)	4900 (2000)	8600 (3500)	0.089 (0.089)	0.12 (0.12)
Liver	Mock	1.5 (0.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	220 (60)	690 (150)	0 (0)	0.090 (0.090)
	Phe+CA	2.8 (0.8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	130 (20)	420 (50)	0 (0)	0 (0)
	PheCA	2.3 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	170 (40)	630 (150)	0 (0)	0 (0)
	Ser+CA	2.3 (0.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	150 (30)	480 (110)	0 (0)	0 (0)
	SerCA	2.0 (0.7)	0 (0)	0 (0)	0 (0)	0 (0)	0.95 (0.20)	140 (30)	470 (80)	0 (0)	0 (0)
	Taur+CA	1.4 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	150 (20)	380 (30)	0 (0)	0 (0)
	TCA	1.9 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	170 (30)	610 (110)	0 (0)	0.083 (0.083)

**Table 4.5: List of BAs present in mass spectrometry standards**

<b>Bile Acid</b>	<b>Abbr.</b>
Glutamatocholic acid	GluCA
Glycocholic acid	GCA
Threonochoholic acid	ThrCA
Tyrosocholic acid	TyrCA
Taurolithocholic acid	TLCA
Taurochenodeoxycholic acid	TCDCA
Glycochenodeoxycholic acid	GCDCA
Taurodeoxycholic acid	TDCA
Alanochoholic acid	AlaCA
Aspartocholic acid	AspCA
Taurocholic acid	TCA
Deoxycholic acid	DCA
Ursodeoxycholic acid	UDCA
Chenodeoxycholic acid	CDCA
3-oxocholic acid	3-oxoCA
Cholic acid	CA
Serocholic acid	SerCA
Phenylalanochoholic acid	PheCA

**Table 4.6: Individual BA concentrations in human sleeve gastrectomy patient cohort**

Concentrations shown are for baseline samples collected and for samples collected 3 months after surgery. Data are presented as mean  $\pm$  s.e.m.,  $n = 44$  per visit, 88 combined.

Bile Acid	Baseline	Follow up	Combined
AlaCA	1.1 $\pm$ 0.5 $\mu$ M	0.4 $\pm$ 0.1 $\mu$ M	0.8 $\pm$ 0.2 $\mu$ M
AlaDCA	2.5 $\pm$ 1.1 $\mu$ M	2.2 $\pm$ 1.4 $\mu$ M	2.4 $\pm$ 0.9 $\mu$ M
ArgCA	1.3 $\pm$ 0.7 $\mu$ M	0.2 $\pm$ 0.2 $\mu$ M	0.8 $\pm$ 0.4 $\mu$ M
AsnCA	0.3 $\pm$ 0.2 $\mu$ M	0.2 $\pm$ 0.1 $\mu$ M	0.2 $\pm$ 0.1 $\mu$ M
AspCA	0.2 $\pm$ 0.1 $\mu$ M	0.4 $\pm$ 0.2 $\mu$ M	0.3 $\pm$ 0.1 $\mu$ M
CA	66 $\pm$ 20 $\mu$ M	8.7 $\pm$ 2.7 $\mu$ M	37 $\pm$ 11 $\mu$ M
CDCA	0.1 $\pm$ 0.0 $\mu$ M	0.1 $\pm$ 0.0 $\mu$ M	0.1 $\pm$ 0.0 $\mu$ M
CitDCA	1.7 $\pm$ 0.5 $\mu$ M	1.3 $\pm$ 0.4 $\mu$ M	1.5 $\pm$ 0.3 $\mu$ M
DCA	150 $\pm$ 20 $\mu$ M	130 $\pm$ 20 $\mu$ M	140 $\pm$ 10 $\mu$ M
GCA	8.6 $\pm$ 3.7 $\mu$ M	39 $\pm$ 38 $\mu$ M	24 $\pm$ 19 $\mu$ M
GCDCA	8.6 $\pm$ 1.5 $\mu$ M	8.0 $\pm$ 1.7 $\mu$ M	8.3 $\pm$ 1.1 $\mu$ M
GLCA	0.0 $\pm$ 0.0 $\mu$ M	0.0 $\pm$ 0.0 $\mu$ M	0.0 $\pm$ 0.0 $\mu$ M
GlnCA	2.3 $\pm$ 1.3 $\mu$ M	0.5 $\pm$ 0.2 $\mu$ M	1.4 $\pm$ 0.6 $\mu$ M
GluCA	22 $\pm$ 10 $\mu$ M	8.4 $\pm$ 3.6 $\mu$ M	15 $\pm$ 5 $\mu$ M
GluCDCA	2.8 $\pm$ 1.3 $\mu$ M	2.3 $\pm$ 1.2 $\mu$ M	2.5 $\pm$ 0.9 $\mu$ M
GluDCA	30 $\pm$ 8 $\mu$ M	11 $\pm$ 3 $\mu$ M	20 $\pm$ 4 $\mu$ M
HisDCA	1.7 $\pm$ 0.3 $\mu$ M	0.8 $\pm$ 0.2 $\mu$ M	1.2 $\pm$ 0.2 $\mu$ M
Ile/LeuCA_6.3min	3.7 $\pm$ 2.0 $\mu$ M	0.8 $\pm$ 0.2 $\mu$ M	2.3 $\pm$ 1 $\mu$ M
Ile/LeuCA_7.0min	5.9 $\pm$ 1.6 $\mu$ M	7 $\pm$ 2.1 $\mu$ M	6.5 $\pm$ 1.3 $\mu$ M
LCA	0.2 $\pm$ 0.0 $\mu$ M	0.3 $\pm$ 0.1 $\mu$ M	0.3 $\pm$ 0.0 $\mu$ M
LysCA	2.6 $\pm$ 1.3 $\mu$ M	0.4 $\pm$ 0.2 $\mu$ M	1.5 $\pm$ 0.7 $\mu$ M
LysUDCA	6.0 $\pm$ 2.4 $\mu$ M	3.5 $\pm$ 1.2 $\mu$ M	4.8 $\pm$ 1.3 $\mu$ M
MetDCA	0.9 $\pm$ 0.3 $\mu$ M	0.8 $\pm$ 0.2 $\mu$ M	0.8 $\pm$ 0.2 $\mu$ M
PheCA	8.4 $\pm$ 2 $\mu$ M	4.5 $\pm$ 1.1 $\mu$ M	6.5 $\pm$ 1.1 $\mu$ M
PheDCA	3.7 $\pm$ 1.5 $\mu$ M	1.7 $\pm$ 0.5 $\mu$ M	2.7 $\pm$ 0.8 $\mu$ M
PheHDCA	0.3 $\pm$ 0.1 $\mu$ M	0.1 $\pm$ 0 $\mu$ M	0.2 $\pm$ 0.1 $\mu$ M
TCA	1.8 $\pm$ 0.5 $\mu$ M	2.7 $\pm$ 1.6 $\mu$ M	2.2 $\pm$ 0.8 $\mu$ M
TLCA	0.3 $\pm$ 0.2 $\mu$ M	0.4 $\pm$ 0.2 $\mu$ M	0.3 $\pm$ 0.2 $\mu$ M
ThrCA	1.6 $\pm$ 0.8 $\mu$ M	0.2 $\pm$ 0.1 $\mu$ M	0.9 $\pm$ 0.4 $\mu$ M
ThrCDCA	1.7 $\pm$ 0.5 $\mu$ M	1.4 $\pm$ 0.4 $\mu$ M	1.5 $\pm$ 0.3 $\mu$ M
TrpCA	0.8 $\pm$ 0.4 $\mu$ M	0.1 $\pm$ 0.1 $\mu$ M	0.5 $\pm$ 0.2 $\mu$ M
TrpDCA	1.5 $\pm$ 0.4 $\mu$ M	0.6 $\pm$ 0.2 $\mu$ M	1.0 $\pm$ 0.2 $\mu$ M
TyrCA	2.6 $\pm$ 1.5 $\mu$ M	1.0 $\pm$ 0.5 $\mu$ M	1.8 $\pm$ 0.8 $\mu$ M

Table 4.6 (cont'd)

TyrUDCA	$0.4 \pm 0.2 \mu\text{M}$	$0.3 \pm 0.2 \mu\text{M}$	$0.3 \pm 0.1 \mu\text{M}$
UCA	$4.9 \pm 2.2 \mu\text{M}$	$4.7 \pm 1.9 \mu\text{M}$	$4.8 \pm 1.4 \mu\text{M}$
oxoCA_5.2min	$20 \pm 13 \mu\text{M}$	$21 \pm 12 \mu\text{M}$	$20 \pm 9 \mu\text{M}$
oxoCA_5.7min	$68 \pm 35 \mu\text{M}$	$61 \pm 32 \mu\text{M}$	$64 \pm 24 \mu\text{M}$

**Table 4.7: BA concentrations, based on class, in human sleeve gastrectomy patient cohort**

Concentrations are shown before surgery, at a 3-month post-operation visit (follow up), or across all samples. Total includes BAs included in other classes, such as sulfated or acetylated BAs.  $n = 44$  per visit, 88 combined. Data are presented as mean  $\pm$  s.e.m.

<b>Visit</b>	<b>Baseline</b>	<b>Follow up</b>	<b>Combined</b>
Primary Conjugated	19 $\pm$ 5 $\mu$ M	49 $\pm$ 39 $\mu$ M	34 $\pm$ 19 $\mu$ M
Primary Unconjugated	12 $\pm$ 4 $\mu$ M	8.7 $\pm$ 2.7 $\mu$ M	11 $\pm$ 2 $\mu$ M
Secondary	230 $\pm$ 40 $\mu$ M	210 $\pm$ 40 $\mu$ M	220 $\pm$ 30 $\mu$ M
MCBA	110 $\pm$ 20 $\mu$ M	50 $\pm$ 9 $\mu$ M	78 $\pm$ 12 $\mu$ M
Total	430 $\pm$ 60 $\mu$ M	360 $\pm$ 80 $\mu$ M	390 $\pm$ 50 $\mu$ M

**Table 4.8: Results from PERMANOVA testing of infant metabolome Bray-Curtis dissimilarity**

Infant age at time of sampling (timepoint), if the infant ceased breastfeeding (bmilkstop), and antibiotic use within 14 days of sampling (antibiotic) were assessed individually in addition to assessing interactions between variables. \*\*\* $P < 0.001$ .

Variable	DF	Sum of Squares	R <sup>2</sup>	F	Pr(>F)	
timepoint	1	3.535	0.02548	13.5336	0.001	***
bmilkstop	1	2.593	0.01869	9.9263	0.001	***
antibiotic	1	0.366	0.00264	1.4020	0.110	
timepoint:bmilkstop	1	0.363	0.00261	1.3888	0.121	
timepoint:antibiotic	1	0.237	0.00170	0.9055	0.531	
bmilkstop:antibiotic	1	0.270	0.00194	1.0324	0.373	
timepoint:bmilkstop:antibiotic	1	0.258	0.00186	0.9873	0.416	
Residual	502	131.131	0.94507			
Total	509	138.752	1.00000			

**Table 4.9: Results from PERMANOVA testing of infant microbiome Bray-Curtis dissimilarity**

Infant age at time of sampling (timepoint), if the infant ceased breastfeeding (bmilkstop), and antibiotic use within 14 days of sampling (antibiotic) were assessed individually in addition to assessing interactions between variables. \* $P < 0.05$ , \*\*\* $P < 0.001$ .

Variable	DF	Sum of Squares	R <sup>2</sup>	F	Pr(>F)	
timepoint	1	2.477	0.02287	8.7641	0.001	***
bmilkstop	1	1.829	0.01688	6.4695	0.001	***
antibiotic	1	0.381	0.00352	1.3484	0.123	
timepoint:bmilkstop	1	0.532	0.00491	1.8824	0.042	*
timepoint:antibiotic	1	0.670	0.00618	2.3697	0.011	*
bmilkstop:antibiotic	1	0.226	0.00208	0.7979	0.681	
timepoint:bmilkstop:antibiotic	1	0.157	0.00145	0.5560	0.916	
Residual	361	102.044	0.94210			
Total	368	108.316	1.00000			

**Table 4.10: EnvFit results based on Bray-Curtis dissimilarity for infant fecal metabolome data**

Variable	Grouping	P value
Infant Race	White (Non-Hispanic), Black (Non-Hispanic), Hispanic, Other (American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Multiracial, Non-Hispanic)	0.0006
Weight Status	Underweight, Normal, Overweight, Obese	0.4226
Length, Z Score	Numeric	0.4859
Weight/Length, Z Score	Numeric	0.4859
Weight, Z Score	Numeric	0.4859
Infant Sex	Female, Male	0.6243
Pre-pregnancy BMI	Numeric	0.0006
Mother Race	White (Non-Hispanic), Black (Non-Hispanic), Hispanic, Other (American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Multiracial, Non-Hispanic)	0.0006
Mother BMI	Numeric	0.0029
Antibiotic	Yes, No	0.6243
Recently Breastfed	Yes, No	0.0006
Breastmilk Diet	Yes, No	0.0006
Formula	Yes, No	0.0006
Freq. Finishing Pumped Meal	Never, Rarely, Sometimes, Most of the time, Always	0.0009
Dairy	Yes, No	0.2255
Meat	Yes, No	0.4226
Fruit	Yes, No	0.4226
Organic Cereal	Yes, No	0.4859
Vegetables	Yes, No	0.4226
New Food Frequency	None, 1 per week, Every 2 days, Every 3 days, Every 4 or 5 days, Every day, More than 1 per day	0.4859
Eggs	Yes, No	0.4226
Sweetened Foods	Yes, No	0.4706
Fruit Juice	Yes, No	0.4226
Nuts	Yes, No	0.7644
Cow's Milk	Yes, No	0.8626
French Fries	Yes, No	0.5467
Breakfast Cereal	Yes, No	0.4859
Times Fed, Daily	Numeric	0.4706
Seafood	Yes, No	0.8626
Sweetened Drink	Yes, No	0.6243
Recently Bottle-fed	Yes, No	0.4589

Table 4.10 (cont'd)

Oat Milk	Yes, No	<b>0.8626</b>
Soy	Yes, No	0.9221

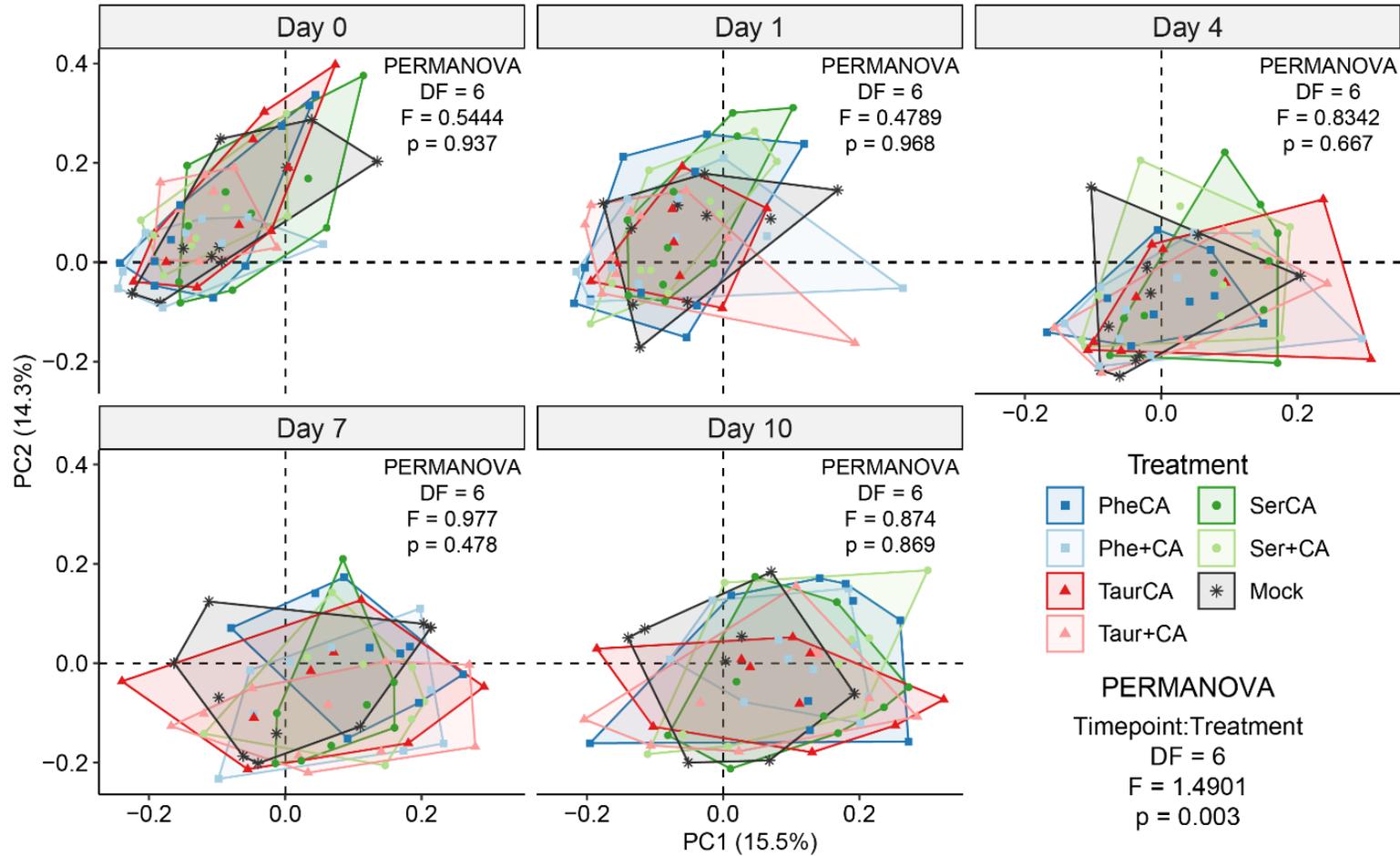
**Table 4.11: EnvFit results based on Bray-Curtis dissimilarity for infant fecal microbiome data**

Variable	Grouping	P value
Infant Race	White (Non-Hispanic), Black (Non-Hispanic), Hispanic, Other (American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Multiracial, Non-Hispanic)	0.1442
Weight Status	Underweight, Normal, Overweight, Obese	0.7238
Length, Z Score	Numeric	0.7215
Weight/Length, Z Score	Numeric	0.7238
Weight, Z Score	Numeric	0.6017
Infant Sex	Female, Male	0.7238
Pre-pregnancy BMI	Numeric	0.7215
Mother Race	White (Non-Hispanic), Black (Non-Hispanic), Hispanic, Other (American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Multiracial, Non-Hispanic)	0.0132
Mother BMI	Numeric	0.0845
Antibiotic	Yes, No	0.7215
Recently Breastfed	Yes, No	0.0017
Breastmilk Diet	Yes, No	0.0017
Formula	Yes, No	0.0363
Freq. Finishing Pumped Meal	Never, Rarely, Sometimes, Most of the time, Always	0.1094
Dairy	Yes, No	0.7932
Meat	Yes, No	0.7238
Fruit	Yes, No	0.3171
Organic Cereal	Yes, No	0.5797
Vegetables	Yes, No	0.7238
New Food Frequency	None, 1 per week, Every 2 days, Every 3 days, Every 4 or 5 days, Every day, More than 1 per day	0.7215
Eggs	Yes, No	0.7215
Sweetened Foods	Yes, No	0.5797
Fruit Juice	Yes, No	0.3019
Nuts	Yes, No	0.7932
Cow's Milk	Yes, No	0.4809
French Fries	Yes, No	0.7238
Breakfast Cereal	Yes, No	0.9930
Times Fed, Daily	Numeric	0.7215
Seafood	Yes, No	0.4809
Sweetened Drink	Yes, No	0.9447
Recently Bottle-fed	Yes, No	0.7215

Table 4.10 (cont'd)

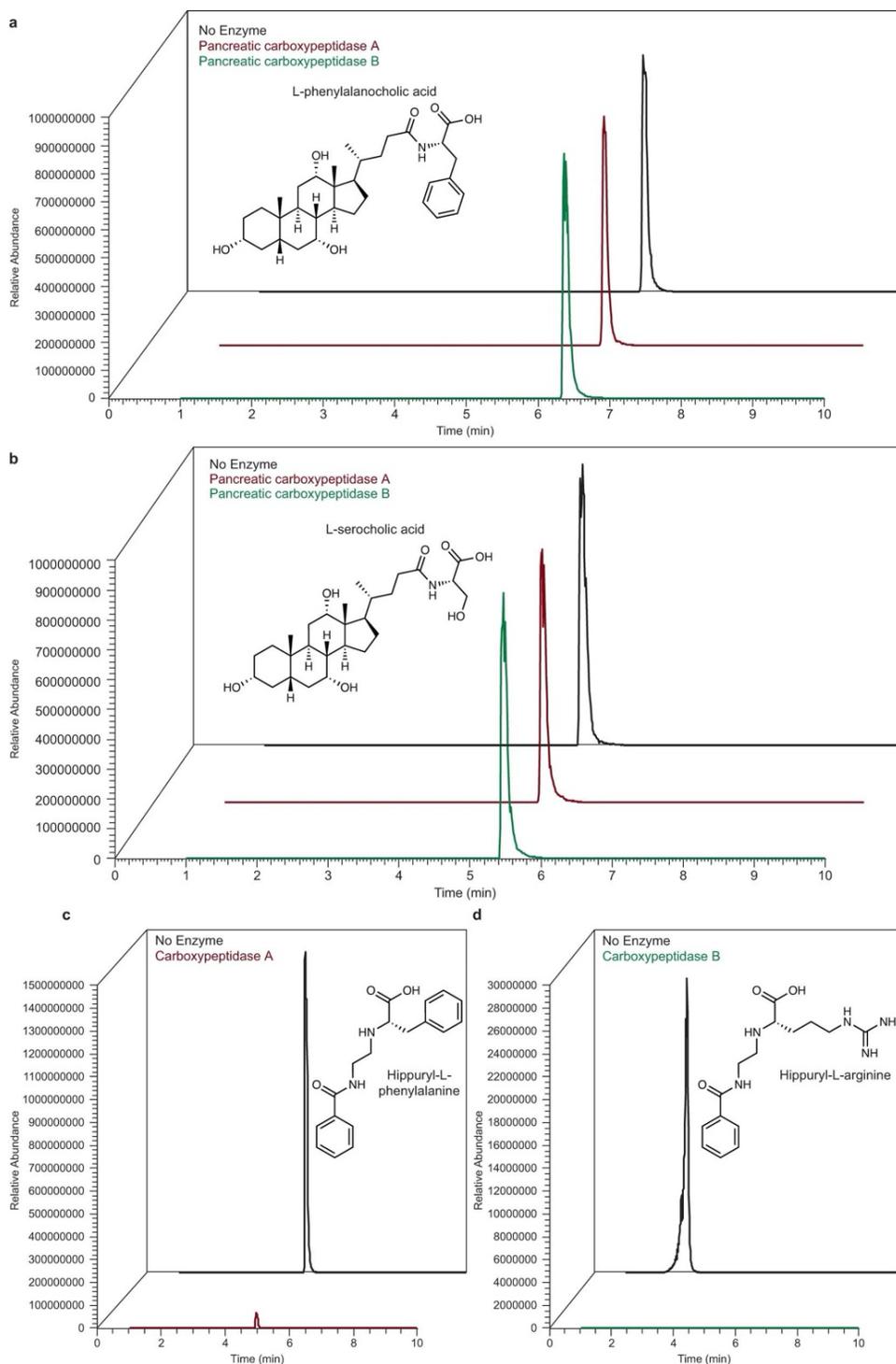
Oat Milk	Yes, No	0.7215
Soy	Yes, No	0.7215

## APPENDIX B: SUPPLEMENTARY FIGURES



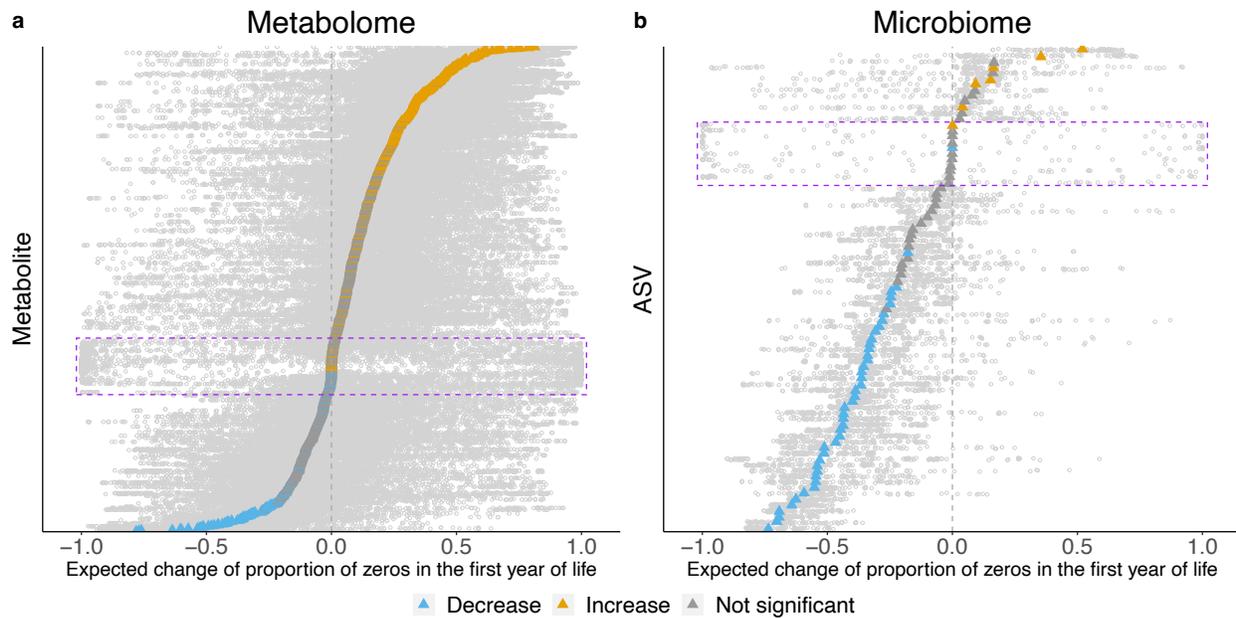
**Figure 4.14: Microbiome community shifts following 10 mg kg<sup>-1</sup> MCBA dosing via PBFM**

Timepoint-nested PERMANOVA reveals significant shifts by treatment, though significance is lost when tested within individual timepoints.  $n = 5$  male, 5 female per treatment.



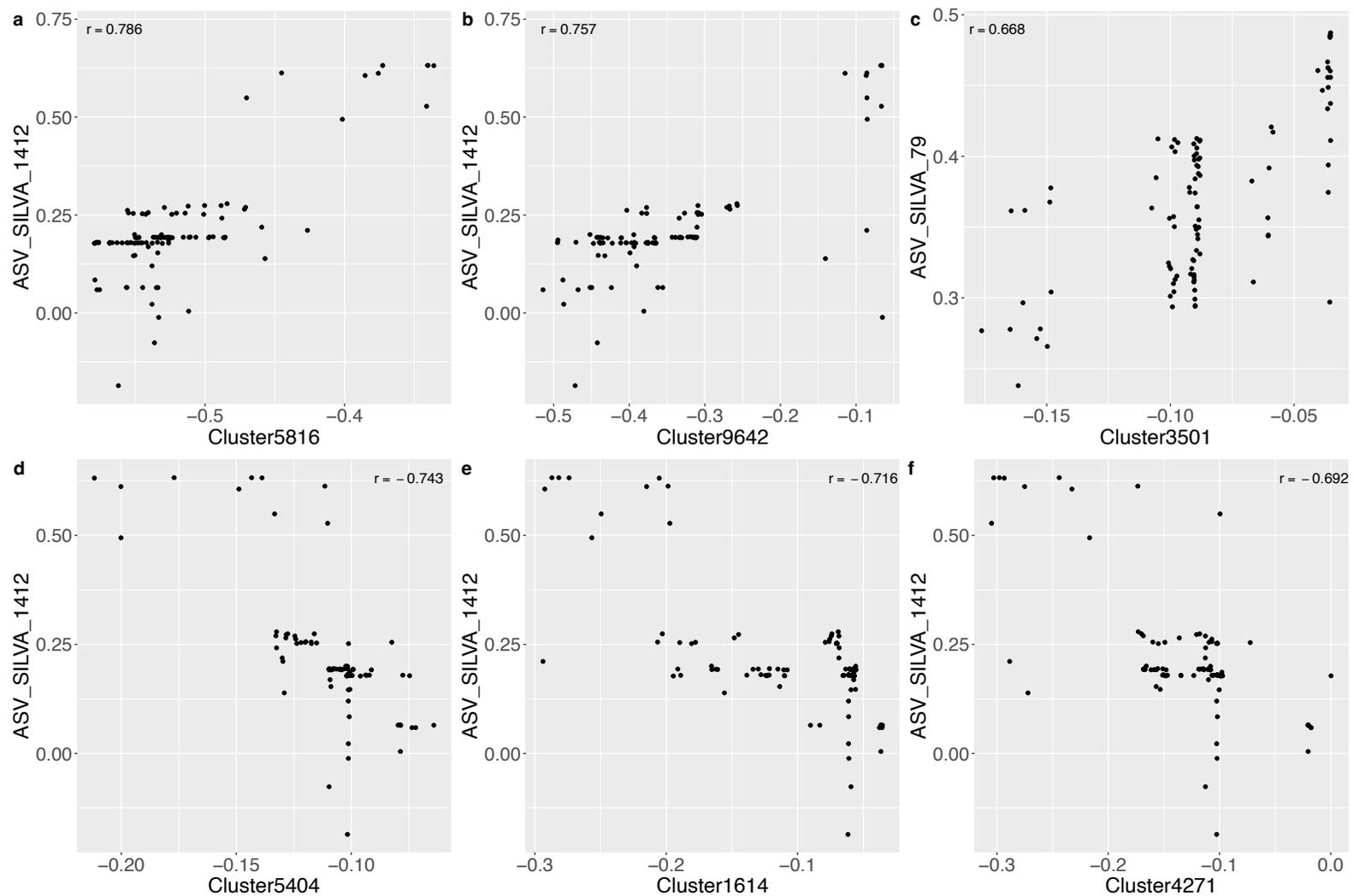
**Figure 4.15: Extracted ion chromatograms of PheCA and SerCA exposed to pancreatic carboxypeptidases**

When incubated with **a**, 1 mM PheCA or **b**, 1 mM SerCA, neither pancreatic carboxypeptidase A nor pancreatic carboxypeptidase B were able to deconjugate the supplemented MCBA while still showing near-complete elimination of native substrates **c**, hippuryl-L-phenylalanine and **d**, hippuryl-L-arginine for carboxypeptidase A and carboxypeptidase B, respectively. Reactions were performed in triplicate.

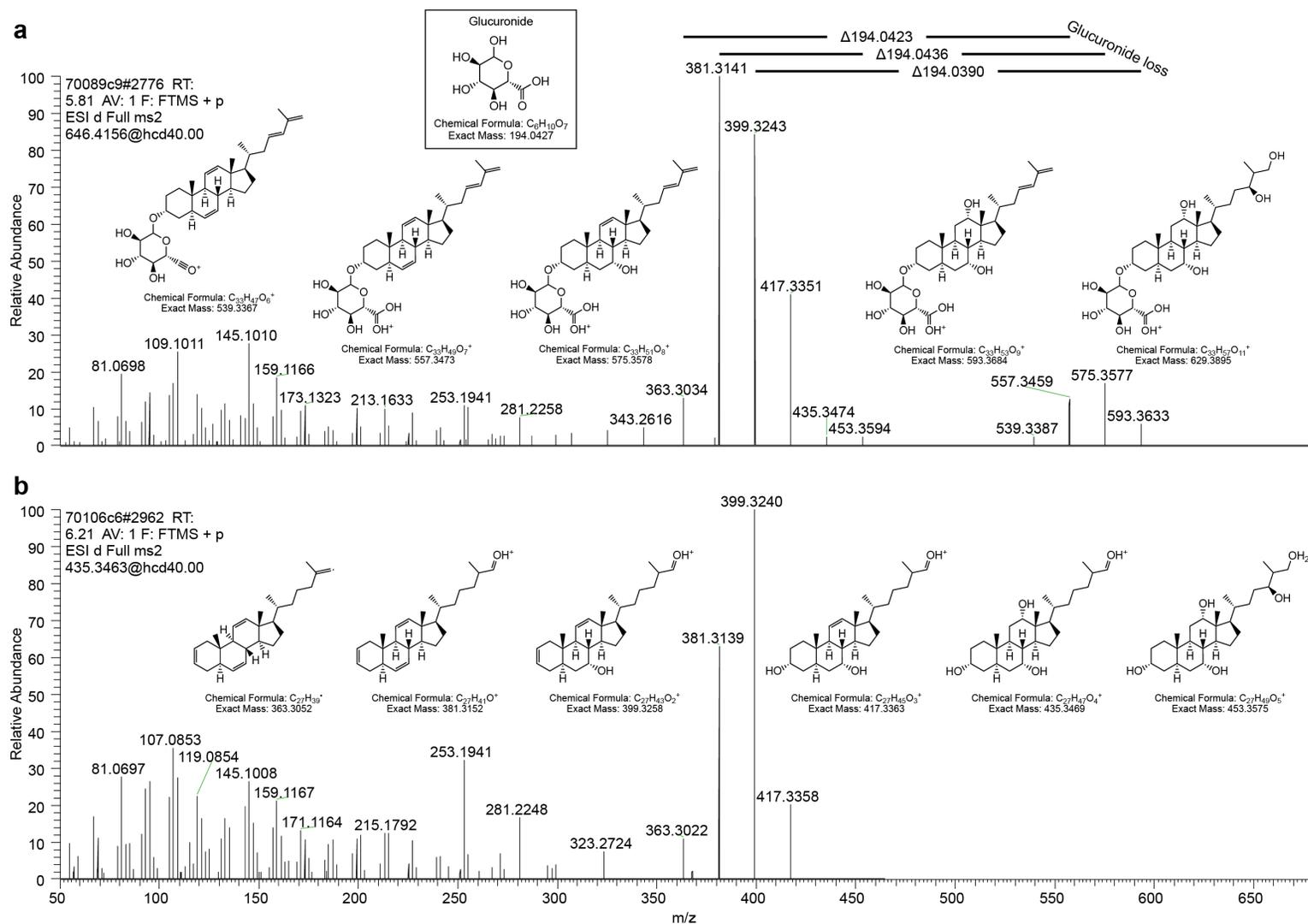


**Figure 4.16: Expected change in probability of zero for metabolites and ASVs by feature and individuals**

Each gray dot represents the expected change in the probability of not being detected (probability of zero) for a metabolite (left) or an ASV (right) in a subject. Dots within the same row correspond to the same feature (metabolite or ASV). The colored triangle is the average change in probability of zero for the corresponding feature (the color used for the average expected changes matches the colors of the curves in Fig. 2: blue indicates a significant decline and yellow an increase in the probability of not being detected, gray represents features with no significant change in prevalence over time). The horizontal dispersion of dots within a row represents how heterogeneous a metabolite or ASV trajectory was across subjects. The dots within the purple rectangle include metabolites and ASVs with overall non-significant changes in the proportion of zeros.



**Figure 4.17: Microbial and metabolite features with significant correlations in temporal shifts of zero-proportions**  
 Each dot represents a subject, the x- and y-coordinates of the points are the predicted change in the probability of zeros for a metabolite (x-axis) ASV (y-axis) pair. **a-c**, Panels showing metabolite-ASV pairs with positive correlations and **d-f** pairs with negative correlations.



**Figure 4.18: MS<sup>2</sup> comparison between putative cholestane glucuronide and annotated cholestane**  
 MS<sup>2</sup> spectra for **a**, the putative glucuronidated cholestane decreasing in abundance over time as infants mature compared to **b**, the annotated cholestane metabolite present within the same molecular network. Mass shifts corresponding to loss of the glucuronide moiety are shown in addition to putative structural annotations for prevalent MS<sup>2</sup> fragment

**CHAPTER 5:**  
**CLOSING REMARKS**

## 5.1 - Conclusions and significance

Building upon the nearly two century-long history of bile research to understand how products of microbial metabolism interact with both the host and their microbiome is essential for responding to and reducing the burden of gastrointestinal diseases. An estimated 24% of adults have been diagnosed with a digestive disease with annual medical costs ranging from approximately \$10,000 to over \$100,000 for each case (1). This work contributes to that goal by identifying one enzyme responsible for producing MCBAs (Chapter 2), characterizing the diversity of bacteria capable of producing these compounds (Chapter 3), and describing interactions between MCBAs, the host, and their resident microbiome (Chapter 4).

### 5.1.1 - Structural nuances of MCBA production

Prior to my investigation into the capacity for amino acid ligation to bile acids (BAs) by the enzyme bile salt hydrolase/transferase (BSH/T), bacterial BA deconjugation was thought to be a unidirectional transformation (2). The catalytic mechanism first involves forming a covalent bond between glycine- or taurine-conjugated BA and BSH/T, liberating the amino acid in the process. As a result, water can easily act as a nucleophile, freeing this deconjugated BA and regenerating the catalytic cysteine present in BSH/T (3). I show that unidirectionality of this mechanism is simply untrue by incubating purified BSH/T from *Clostridium perfringens* (CpBSH/T) with a mix of all 20 proteinaceous amino acids and taurocholic acid (TCA), glycocholic acid (GCA), or cholic acid (CA). Across all three primary BA forms, CpBSH/T was able to conjugate noncanonical amino acids to a CA backbone. I observed differences in conjugation efficiency under these conditions by BA substrate. Overall MCBA production was lower for GCA and CA compared to TCA,

matching substrate preferences for CpBSH/T in the context of catalyzing deconjugation (4, 5).

In Chapter 2, I show that the structure of the BSH/T active site is an important driver of amino acid specificity during MCBA production based on both sequence analysis and mutagenesis experiments. After comparing BSH/T amino acid sequences across strains screened for MCBA production, I identified Asn82 as an important driver of amino acid selection in BA conjugation. By comparing MCBA production between WT and N82Y variants, I show that this residue plays a role in active site structure. When including comparisons between N82Y and C2A variants, I conclude that Asn82 is important in MCBA product diversity but it is not essential for the catalysis of acyl transfer or deconjugation. These findings have significant implications for the over 50 year history of research on BSH/T, which has long been confused about substrate specificity between taurine and glycine BAs. This confusion in the literature may be due to evolutionary pressure acting on the previously unappreciated transferase activity of the enzyme shaping the active site in a manner that dictates both acyl-transfer and hydrolysis.

### *5.1.2 - Diverse products of microbial bile acid metabolism*

Not only are the bacteria capable of this transformation diverse, but so too is the resulting MCBA pool. While production of the secondary BAs deoxycholic acid (DCA) and lithocholic acid (LCA) has traditionally been associated with a subset of clostridial species, the diversity of bacterial taxa containing *bsh* is significantly larger, shown in over 1 in 4 strains in the Human Microbiome Project microbiota reference genome containing at least one (2, 6–8). However, unlike the multi-step pathway for 7 $\alpha$ -dehydroxylation, I show that simply having a *bsh/t* gene encoding a functional enzyme is enough to produce MCBAs.

I show that evolutionary relatedness is not a strong predictor of the capacity for MCBA production. Instead, BSH/T production is associated with the enzyme's amino acid sequence. I also show that BSH/T itself is not the sole enzyme capable of MCBA production as *L. scindens* was observed to produce MCBA, despite its lack of an annotated *bsh/t* gene. Other groups have also reported MCBA production by *L. scindens* (9). Interestingly, but perhaps unsurprisingly, *L. scindens* was observed to produce MCBA using DCA as the BA backbone. I also show that amino acid use in production extends beyond essential amino acids, including citrulline and ornithine.

### 5.1.3 - MCBA capacity for microbiome remodeling

In this work, I showed that MCBA have highly variable effects on bacterial growth. These effects match previous reports that increases in BA hydrophilicity via conjugation with glycine and taurine reduce antimicrobial efficacy (10). Across all strains grown in the presence of an MCBA, only the more hydrophobic forms showed notable impacts on bacterial growth. Interestingly, I observed that *E. bolteae* grew better when exposed to any MCBA than when provided CA alone. Whether this is due to resistance mechanisms resulting from its robust ability to produce MCBA, the biochemical changes resulting from conjugation, or a combination of both remains to be determined. However, given the response of *L. plantarum* and *P. anaerobius*, *E. bolteae* may only need to control for hydrophobic conjugates, as MCBA containing more hydrophilic amino acids demonstrated little to no antimicrobial activity independent of established BA resistance mechanisms.

These effects translated to the murine gastrointestinal tract when provided at a high dose. However, the magnitude of shifts in the microbiome decreased upon reducing

the MCBA dose 10-fold to more physiologically relevant concentrations. Corn oil gavage alone is sufficient to cause microbiome community shifts within mice (11). Therefore, adding a detergent likely impacting host lipid uptake may only exacerbate changes already occurring. Observing community shifts at the highest concentration may be indicative that these compounds are more impactful within gastrointestinal microenvironments, such as the space between intestinal villi or within small intestinal crypts. In addition to uncertainties involving the potential roles played within these microenvironments, it is difficult to fully quantify the concentration of MCBAs within the gut, particularly within the mouse due to the small size of the collected sample. Folz *et al.* showed that MCBAs are enriched in the human small intestine compared to feces utilizing an ingestible sampling device, contrary to what one would expect with the significantly smaller microbial population present (12, 13). Measuring these compounds accurately is challenging due to their diversity. Limiting the acyl-conjugate repertoire to only the 20 essential amino acids and two primary BA backbones, would result in 40 separate compounds, making it difficult to measure collectively as standards in an LC-MS run. Expanding to include products of other BA transformations increases the number to the thousands while still excluding C27 bile acids, C24 bile alcohols, and other related compounds (2, 14).

#### *5.1.4 - MCBA implications in human health*

With the recency of the discovery of MCBAs, understanding their roles in host health remains an important and ongoing pursuit. Previous work has shown that these compounds are enriched in patients with irritable bowel diseases, specifically Crohn's disease (15), and multiple groups have shown that MCBAs are capable of modulating

host signaling pathways depending on the ligated amino acid (16, 17). In the work presented here, I build upon the other evidence showing these compounds are enriched within a dysbiotic gut. The concentration of MCBA significantly decreases following sleeve gastrectomy as a treatment modality for obesity, whereas primary and secondary BA concentrations did not show this significant change. I also show that infant maturation correlates with decreasing MCBA prevalence, matching reductions in other forms of detoxified BAs, namely glucuronidated BAs.

## **5.2 - Future directions**

### *5.2.1 - Identifying alternate routes of microbial BA conjugation and the true diversity of the BA pool*

One of the surprising observations in this work was the capacity for *L. scindens* ATCC 35704 to conjugate bile acids, though it does not contain an annotated or predicted *bsh/t* allele. Human BAAT has previously been shown to transfer glycine to various fatty acids (18). Therefore, microbial fatty acid transferases are a prime target for investigation into their capacity for amino acid transfer to bile acids, perhaps as a detoxification mechanism for bile acids and fatty acids alike.

Untangling the capacity for individual bacterial strains to conjugate bile acids merely scratches the surface of truly understanding the biochemical importance of these molecules. The work here focused on *C. perfringens* BSH/T, given its availability commercially and its prevalence within BA research throughout previous literature. However, recent work by Song *et al.* revealed BSH/T presence across 117 genera within 12 phyla from the Human Microbiome Project database (8). Using these sequences as references for analyzing global metagenomic data, they found that global BSH/T

sequences grouped into 8 distinct BSH/T phylotypes within which the capacity for deconjugation varied significantly. Marrying their *in silico* analysis with experimental validation in the context of MCBA production could bolster our understanding of not only how the amino acid sequence impacts conjugation capacity but the relevance of these compounds to global health.

### 5.2.2 - Interrogating microbe-MCBA-host interactions

The work here focused primarily on MCBA production from CA, GCA, and TCA, but these core molecules are not the only conjugated BAs to which gastrointestinal bacteria are exposed. Primary BA hydrolysis by BSH/T enzymes from different organisms shows BSH/T sequences can be grouped by preference for either glycine- or taurine-bound BAs. Certain BSH/T are capable of hydrolyzing MCBAs, exhibiting preferences for certain amino acids (19, 20). For example, BSH/T from *C. perfringens* is capable of deconjugating TyrCA but is practically unable to deconjugate PheCA or LeuCA, yet BSH/T from *L. plantarum* demonstrated the ability to deconjugate all three MCBAs with only slight reductions in its ability to deconjugate LeuCA (20). Just as substrate preferences for deconjugation by BSH/T have been routinely described, much remains unknown for both BA and amino acid preferences across bacterial taxa.

I show here that MCBAs have the propensity for species-specific inhibition of both commensal and pathogenic bacteria. For example, PheCA and TyrCA have been reported to inhibit *C. difficile* germination (21). Teasing out the structural nuances of this inhibition would provide useful information in the context of drug development for the treatment of *C. difficile* infections with the potential for broad application to additional gastrointestinal pathogens, such as *Vibrio cholerae*.

One of the most significant drawbacks of 16S-V4 amplicon microbiome and untargeted metabolome analysis is that findings are primarily correlative. We show here that supplementing the diet with a high individual MCBA dose is sufficient for causing observable microbiome shifts within the gastrointestinal microbiome, but a 10-fold reduction in dose results in a loss of these shifts. Future work involving more direct, causal analysis is essential to understand the true physiological role these compounds play within the gut and throughout the body.

### **5.3 - Concluding remarks**

Findings from the work presented in this thesis provide novel insights into the production and physiological impacts of MCBAs. To the best of our knowledge, we reported the first describe a microbial enzyme capable of conjugating BAs, a discovery that not only moved the entire field of BA research forward, but also opens up an entirely new area of research to understand how the gut microbiome broadly modifies host lipids through acyl-transfer. Beyond this, the discoveries presented here act as the foundation for future research into microbial, murine, and human consequences of MCBA production. It is my hope that this thesis lays the groundwork for investigation and characterization of microbial BA metabolism that we have only just discovered.

## REFERENCES

1. Mathews SC, Izmailyan S, Brito FA, Yamal J-M, Mikhail O, Revere FL. 2022. Prevalence and financial burden of digestive diseases in a commercially insured population. *Clin Gastroenterol Hepatol* 20:1480-1487.e7.
2. Guzior DV, Quinn RA. 2021. Review: microbial transformations of human bile acids. *Microbiome* 9:140.
3. Lodola A, Branduardi D, De Vivo M, Capoferri L, Mor M, Piomelli D, Cavalli A. 2012. A catalytic mechanism for cysteine N-terminal nucleophile hydrolases, as revealed by free energy simulations. *PLoS ONE* 7:e32397.
4. Coleman JP, Hudson LL. 1995. Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens*. *Appl Environ Microbiol* 61:2514–2520.
5. Gopal-Srivastava R, Hylemon PB. 1988. Purification and characterization of bile salt hydrolase from *Clostridium perfringens*. *J Lipid Res* 29:1079–1085.
6. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. 2020. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582:566–570.
7. Kim KH, Park D, Jia B, Baek JH, Hahn Y, Jeon CO. 2022. Identification and characterization of major bile acid 7 $\alpha$ -dehydroxylating bacteria in the human gut. *mSystems* 7:e00455-22.
8. Song Z, Cai Y, Lao X, Wang X, Lin X, Cui Y, Kalavagunta PK, Liao J, Jin L, Shang J, Li J. 2019. Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* 7:9.
9. Lucas LN, Barrett K, Kerby RL, Zhang Q, Cattaneo LE, Stevenson D, Rey FE, Amador-Noguez D. 2021. Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. *mSystems* 6:e00805-21.
10. Sannasiddappa TH, Lund PA, Clarke SR. 2017. *In vitro* antibacterial activity of unconjugated and conjugated bile salts on *Staphylococcus aureus*. *Front Microbiol* 8:1581.
11. Gokulan K, Kumar A, Lahiani MH, Sutherland VL, Cerniglia CE, Khare S. 2021. Differential toxicological outcome of corn oil exposure in rats and mice as assessed by microbial composition, epithelial permeability, and ileal mucosa-associated immune status. *Toxicol Sci* 180:89–102.

12. Folz J, Culver RN, Morales JM, Grembi J, Triadafilopoulos G, Relman DA, Huang KC, Shalon D, Fiehn O. 2023. Human metabolome variation along the upper intestinal tract. *Nat Metab* 5:777–788.
13. Shalon D, Culver RN, Grembi JA, Folz J, Treit PV, Shi H, Rosenberger FA, Dethlefsen L, Meng X, Yaffe E, Aranda-Díaz A, Geyer PE, Mueller-Reif JB, Spencer S, Patterson AD, Triadafilopoulos G, Holmes SP, Mann M, Fiehn O, Relman DA, Huang KC. 2023. Profiling the human intestinal environment under physiological conditions. *Nature* 617:581–591.
14. Mohanty I, Allaband C, Mannocho-Russo H, El Abiead Y, Hagey LR, Knight R, Dorrestein PC. 2024. The changing metabolic landscape of bile acids – keys to metabolism and immune regulation. *Nat Rev Gastroenterol Hepatol* <https://doi.org/10.1038/s41575-024-00914-3>.
15. Quinn RA, Melnik AV, Vrbanac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, Downes M, Welch RD, Quinn M, Humphrey G, Panitchpakdi M, Weldon KC, Aksenov A, da Silva R, Avila-Pacheco J, Clish C, Bae S, Mallick H, Franzosa EA, Lloyd-Price J, Bussell R, Thron T, Nelson AT, Wang M, Leszczynski E, Vargas F, Gauglitz JM, Meehan MJ, Gentry E, Arthur TD, Komor AC, Poulsen O, Boland BS, Chang JT, Sandborn WJ, Lim M, Garg N, Lumeng JC, Xavier RJ, Kazmierczak BI, Jain R, Egan M, Rhee KE, Ferguson D, Raffatellu M, Vlamakis H, Haddad GG, Siegel D, Huttenhower C, Mazmanian SK, Evans RM, Nizet V, Knight R, Dorrestein PC. 2020. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature* 579:123–129.
16. Rimal B, Collins SL, Tanes CE, Rocha ER, Granda MA, Solanki S, Hoque NJ, Gentry EC, Koo I, Reilly ER, Hao F, Paudel D, Singh V, Yan T, Kim MS, Bittinger K, Zackular JP, Krausz KW, Desai D, Amin S, Coleman JP, Shah YM, Bisanz JE, Gonzalez FJ, Vanden Heuvel JP, Wu GD, Zemel BS, Dorrestein PC, Weinert EE, Patterson AD. 2024. Bile salt hydrolase catalyses formation of amine-conjugated bile acids. *Nature* 626:859–863.
17. Fu T, Huan T, Rahman G, Zhi H, Xu Z, Oh TG, Guo J, Coulter S, Tripathi A, Martino C, McCarville JL, Zhu Q, Cayabyab F, Low B, He M, Xing S, Vargas F, Yu RT, Atkins A, Liddle C, Ayres J, Raffatellu M, Dorrestein PC, Downes M, Knight R, Evans RM. 2023. Paired microbiome and metabolome analyses associate bile acid changes with colorectal cancer progression. *Cell Rep* 42:112997.
18. O'Byrne J, Hunt MC, Rai DK, Saeki M, Alexson SEH. 2003. The human bile acid-CoA:amino acid N-acyltransferase functions in the conjugation of fatty acids to glycine. *J Biol Chem* 278:34237–34244.
19. Foley MH, Allen G, Rivera AJ, Stewart AK, Barrangou R, Theriot CM. 2021. *Lactobacillus* bile salt hydrolase substrate specificity governs bacterial fitness and host colonization <https://doi.org/10.1073/pnas.2017709118/-/DCSupplemental>.

20. Malarney KP, Chang PV. 2023. Electrostatic interactions dictate bile salt hydrolase substrate preference. *Biochemistry* 62:3076–3084.
21. Foley MH, Walker ME, Stewart AK, O’Flaherty S, Gentry EC, Patel S, Beaty VV, Allen G, Pan M, Simpson JB, Perkins C, Vanhoy ME, Dougherty MK, McGill SK, Gulati AS, Dorrestein PC, Baker ES, Redinbo MR, Barrangou R, Theriot CM. 2023. Bile salt hydrolases shape the bile acid landscape and restrict *Clostridioides difficile* growth in the murine gut. *Nat Microbiol* 8:611–628.