

CHARACTERIZATION OF THE HUMAN GUT RESISTOME, MICROBIOME, AND  
METABOLOME DURING ENTERIC INFECTION

By

Zoe A. Hansen

A DISSERTATION

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

Microbiology and Molecular Genetics – Doctor of Philosophy  
Ecology, Evolutionary Biology and Behavior – Dual Major

2022

## ABSTRACT

### CHARACTERIZATION OF THE HUMAN GUT RESISTOME, MICROBIOME, AND METABOLOME DURING ENTERIC INFECTION

By

Zoe A. Hansen

The human gut environment is replete with host-microbe and microbe-microbe interactions that shape human health. This system is also a known reservoir for antimicrobial resistance (AMR). The ubiquity of AMR is alarming, as greater than 2.8 million antibiotic-resistant infections and 35,000 deaths occur annually in the United States. Multiple human pathogens have demonstrated reduced susceptibility to various antibiotics, including enteric pathogens such as *Campylobacter*, *Salmonella*, *Shigella*, and STEC, which cause millions of foodborne infections each year. The increasing incidence of antibiotic resistant enteric infections substantiates a need to further characterize these pathogens' role in the curation and dissemination of AMR across environments.

In this dissertation, a total of 223 human stools were assessed using shotgun metagenomics sequencing to investigate gut microbiome changes associated with enteric infection. Sixty-three stools were collected from patients suffering from enteric infection between 2011-2015 by the Michigan Department of Health and Human Services (MDHHS). Sixty-one of these patients submitted a follow-up sample between 1- and 29-weeks post-infection, and 99 healthy household members also submitted stools to serve as controls.

In Chapter 2, a subset of patients infected with *Campylobacter* spp. and their related controls were investigated to assess the gut resistome, or collection of all antimicrobial resistance genes (ARGs) and their genetic precursors, related to infection. This examination revealed significantly higher ARG diversity in infected patients compared to healthy controls.

Specifically, levels of multi-drug resistance (MDR) were greatly increased during infection. Three case clusters with distinct resistomes were identified; two of these clusters had unique ARG profiles that differed from those of healthy family members.

In Chapter 3, a larger subset of 120 paired samples (60 infected vs. 60 recovered) were investigated to further characterize resistome and microbiome fluctuations related to infection and recovery. Again, infected patients harbored greater resistome diversity; however, recovered individuals displayed higher diversity in their microbiota composition. Despite their lower overall microbial diversity, patients with acute infections showed an increase in the abundance of members of *Enterobacteriaceae*, with specific expansion of the genus *Escherichia*. Host-tracking analysis revealed that many *Enterobacteriaceae* carried ARGs related to MDR and biocide resistance, a finding with broad implications for the ecology of resistance during infection.

The fourth chapter explored metabolic capacity of gut microbial communities. In addition to metabolic pathway prediction, untargeted metabolomics was performed via LC/MS for 122 paired samples. Pathway annotation suggested that infected individuals contain greater microbial functional capacity, but metabolomics indicated greater overall metabolite diversity among recovered patients. Infection was associated with enhanced nitrogen and amino acid metabolism pathways. Although many metabolites remain uncharacterized, their presence or absence among individuals suggest their importance during and after infection.

Altogether, the findings of this dissertation further characterize ecological consequences related to enteric infection in the human gut. Specifically, this research illustrates the importance of enteric infection in the dissemination and persistence of resistance determinants. Moreover, the expansion of *Enterobacteriaceae* and the evident increase in nitrogen- and amino acid-related metabolism during infection represent potential targets for future intervention practices.

*For those who have taught me  
to love and be loved*



## ACKNOWLEDGEMENTS

*Whoever you are, no matter how lonely,  
the world offers itself to your imagination,  
calls to you like the wild geese, harsh and exciting –  
over and over announcing your place  
in the family of things.*

Mary Oliver  
Excerpt from **Wild Geese**

My heart is full of gratitude for the many, many people that have helped me define my place throughout graduate school. Without them, my pursuit and achievement of this doctoral degree would not have been possible. First, I would like to thank my mentor, Dr. Shannon Manning, for her continued guidance, expertise, and support throughout my graduate career. Upon joining her lab, I was unsure of whether I was “cut out” for the project she proposed I take on; it was primarily computational, and I had very little experience in any sort of programming language. But she saw right through my doubts. She urged me to apply myself, granting me the independence and grace to make mistakes, troubleshoot, and make mistakes again. Without her openness and belief in my ability, I may never have risen to the challenge of this project, nor would I have experienced the tremendous personal and professional growth that has accompanied it. I continue to be amazed by Shannon and her capacity to make space for patience, empathy, and poise, especially in the face of her ever-demanding schedule. She has been one of the most inspirational women in STEM that I have had the honor to work with, and I cannot thank her enough for serving as such a phenomenal mentor and role model.

I would also like to extend copious thanks to my graduate committee members, Dr. Lixin Zhang, Dr. Kim Scribner, Dr. Ashley Shade, and Dr. Rob Abramovitch, for their continued service and guidance throughout my research. Each of these individuals have offered invaluable advice, strengthening not only the methods and approaches of my projects, but my ability to

think more critically as a scientist. I would also like to extend my thanks to Dr. Jim Tiedje, who has graciously shared his insight and expertise with me and other members of Shannon's lab. I am also very grateful to the Microbiology and Molecular Genetics Department for fostering such a collaborative, supportive research community at MSU. Being a part of such a close-knit, friendly group of peers, faculty, and staff has been incredibly rewarding.

My completion of this degree would not have been possible without generous financial support. I am very thankful for the funding provided by the National Institutes of Health and the United States Department of Agriculture. Additionally, I am honored to have received the Thomas S. Whittam Travel Award and the Ronald and Sharon Rogowski Fellowship from the Microbiology and Molecular Genetics Department, the James. M. Tiedje Graduate Student Travel Award in Microbial Ecology from the Department of Plant, Soil, and Microbial Sciences, a Summer Fellowship in Ecology, Evolution, and Behavior from the Ecology, Evolution, and Behavior Program, and a Dissertation Continuation Fellowship and Dissertation Completion Fellowship from the College of Natural Science at MSU.

I am so very lucky to have been surrounded by many truly wondrous friends while at MSU, without whom I cannot fathom getting through graduate school. I will first thank my dear cohort friends – Leah Johnson, Reid Longley, Jake Bibik, Paul Fiesel, Garret Miller, Irving Salinas, and Davis Mathieu. From our first week together in the BMS program, to the many bonfires, Cinco de Mayo parties, basketball viewings, nature outings, tailgates, and board game nights, I have been so lucky to call them my dearest friends. I want to extend a special thanks to Leah for being a best friend and running buddy, and for all of our trips to Playmakers, Wharton Center shows, coffee dates, Chen's dinners, and Horrocks runs.

I am also deeply grateful for the friendships I have with my lab mates – Jose Rodrigues, Macy Pell, Cole McCutcheon, Sanjana Mukherjee, Karla Vasco, Claudia Sepulveda, and Bailey Bowcutt. It is not often that your lab mates become your family, but I am lucky enough to say this is the case. In addition to being amazing scientists and co-workers, they are also the goofiest, most lovely group one could ever hope to be a part of. I will particularly always cherish memories from our trip to Florida, tagging along with Macy and Cole for yet another coffee, pranking Jose (maybe a few too many times), visiting Sanjana in Washington D.C., spa days and massages with Macy, teaching and venting with Cole, and dog walks and bonding with Jose.

I also need to thank my volleyball friends, old and new, who have been helping to keep me sane over the past few years – Macy Pell, Mike Pajkos, Nick Valverde, Adam Kawash, Tara Watkins, Kurt Walcheske, and Meeshon Rogers. Playing volleyball with this group has been an incredible outlet for me to release pent up stress, exercise my competitiveness, and stay active. Not only are they the best teammates in the world, but they are also dear friends.

My pursuit of a PhD would not have been possible without the monumental support of my family. I wish to recognize the boundless love, time, support, and patience my parents, Zack and Nancy, have graced me with my entire life. In times of indecision, doubt, fear, or apprehension, they have been a steadfast source of security and love. My parents' love of nature and science drove me to pursue science myself; I can never thank them enough for instilling a lifelong sense of wonder and curiosity in me. I cannot wait to continue learning with and from them. I am also eternally grateful for the love and support of my sister, Anika. Even from across the country, she has always made me feel embraced, valued, and loved, reminding me of my strengths and dispelling my inner self-doubts. I cherish her wisdom and am fortunate to have a younger sister who sees the world through the most forgiving and kind eyes. I also wish to

recognize the unconditional love given to me by my dogs, Pearl and Rowan. They have shared with me the strongest of friendships and I adore them both for their ability to convey strength, hope, and love without words.

Finally, I would like to thank my partner, Jordan Dull. Despite having to deal with the business of his own PhD pursuit, Jordan has always made time for me and our relationship. He has shown endless amounts of patience, support and love during this time. He is perhaps the most creative, kind, genuine, and selfless person I know. I have learned much from Jordan's capacity for forgiveness, for his ability to show patience even in the most frustrating of scenarios, and I so appreciate the ways in which he challenges me to grow and evolve. Maintaining our closeness while being apart has not been easy, but Jordan's devotion to making me feel valued, heard, and seen, even over a video chat or phone call, helped me muster the strength to tackle difficult days. Among the trying times that accompany graduate school, he has always given me a home to return to. I cannot thank him enough for sharing this life with me.

When I first set out to earn my PhD, I never could have dreamed of how rewarding the process would be. In addition to developing new skills, mastering techniques, asking profound questions, and performing rigorous analyses, I was also able to forge lifelong friendships and strengthen existing ones. Being surrounded by these incredible people is what propelled me through my degree here. They have all helped me to further define my place in this family of things.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> .....	xiii
<b>LIST OF FIGURES</b> .....	xv
<b>KEY TO ABBREVIATIONS</b> .....	xxi
<b>CHAPTER 1</b> Literature Review: The Human Gut Microbiome, Antibiotic Resistance, and Enteric Infection.....	1
<b>MICROBIAL ECOLOGY OF THE HUMAN GUT</b> .....	2
Community assembly of the early human gut microbiome .....	2
Diversity and composition of the human gut microbiota.....	4
The gut microbiome during periods of ecological change.....	6
<b>SIGNIFICANCE OF ENTERIC INFECTION AND NOTABLE REPERCUSSIONS RELATED TO ENTERIC PATHOGENS</b> .....	10
<b>MECHANISMS OF COLONIZATION RESISTANCE PROTECT THE GUT FROM ENTERIC INFECTION</b> .....	12
<b>OTHER FACTORS INFLUENCING THE TRAJECTORY AND SEVERITY OF ENTERIC INFECTIONS</b> .....	15
<b>RELEVANCE OF ANTIBIOTIC RESISTANCE TO THE HUMAN GUT MICROBIOME AND ENTERIC PATHOGENS</b> .....	16
Antibiotics and Antibiotic Resistance.....	17
Mobility of antibiotic resistance .....	20
The ecology of antimicrobial resistance .....	21
Antibiotic resistance among enteric pathogens.....	25
<b>ROLES OF THE GUT MICROBIOME AND RELATED METABOLIC CAPACITY IN SHAPING HUMAN HEALTH</b> .....	28
Interplay between the gut microbiome and metabolism.....	28
The importance of short-chain fatty acids (SCFAs) in gut health .....	29
Other microbially-mediated metabolites of importance .....	32
<b>COMPUTATIONAL METHODS FOR STUDYING THE HUMAN GUT MICROBIOME, RESISTOME, AND METABOLOME</b> .....	34
16S rRNA and ITS sequencing.....	34
Shotgun metagenomics sequencing .....	36
Metabolomics and other ‘omics technologies.....	41
Downstream interpretation and statistical analysis.....	43
<b>SUMMARY</b> .....	46
<b>REFERENCES</b> .....	49
<b>CHAPTER 2</b> Comparing gut resistome composition among patients with acute <i>Campylobacter</i> infections and healthy family members .....	75
<b>ABSTRACT</b> .....	76
<b>INTRODUCTION</b> .....	77

<b>MATERIALS AND METHODS</b> .....	79
Study population .....	79
Sample preparation and sequencing analysis.....	80
Identification of antimicrobial resistance genes (ARGs).....	81
Identification of microbial taxa.....	82
Ecological analyses.....	83
Hierarchical clustering and epidemiological analysis.....	83
<b>RESULTS</b> .....	84
Characteristics of the study population.....	84
Number and diversity of ARGs vary depending on health status.....	85
Specific ARGs define case and control samples.....	85
Taxonomic diversity differs between cases and controls .....	90
Specific ARGs are not strongly associated with <i>Campylobacter</i> in the case samples.....	91
Clusters 1 and 3 have more diverse resistomes than Cluster 2.....	92
Case epidemiological data is linked to specific resistome profiles.....	93
Family relation is less influential than health status in shaping the gut resistome during enteric infection.....	95
<b>DISCUSSION</b> .....	96
<b>APPENDIX</b> .....	106
<b>REFERENCES</b> .....	122
<b>CHAPTER 3</b> Exploring recovery of the gut microbiome following enteric infection and the persistence of resistance genes in specific microbial hosts .....	129
<b>ABSTRACT</b> .....	130
<b>INTRODUCTION</b> .....	131
<b>METHODS</b> .....	134
Study population .....	134
Sample preparation and sequencing analysis.....	135
AmrPlusPlus – Read-based pipeline.....	135
Identification of microbial taxa.....	136
Ecological analyses.....	136
<i>Abundance and diversity analyses</i> .....	136
<i>Differential abundance of taxa and ARGs</i> .....	138
<i>Identification of continuous population structure</i> .....	138
Co-occurrence network construction .....	139
Anvi'o – Assembly-based pipeline.....	141
ARG-carrying contigs host-tracking analysis.....	141
<b>RESULTS</b> .....	142
Study population characteristics .....	142
Changes in the composition and diversity of the resistome and microbiome after recovery from enteric infection .....	144
<i>Resistome diversity</i> .....	144
<i>Microbiome diversity</i> .....	147

<i>Exploring potential for continuous structure of resistome and microbiome compositions using MMUPHin</i> .....	149
<i>Resistome composition</i> .....	151
<i>Microbiome composition</i> .....	156
<i>Covariate-controlled batch effect adjustment and differential abundance testing with MMUPHin</i> .....	158
Co-occurrence network analysis reveals connections between taxa and ARGs.....	160
<i>Global network construction</i> .....	160
<i>Co-occurrence networks relevant to beta-lactam ARGs</i> .....	162
<i>Investigating global co-occurrence networks related to infectious pathogen</i> .....	164
Host-tracking analysis.....	166
<i>ARG-harboring microbial hosts detected in cases vs. follow-ups</i> .....	166
<i>Comparing across enteric pathogens</i> .....	167
<i>Investigating the potential persistence of clinically relevant ESBLs post-recovery</i> .....	171
<b>DISCUSSION</b> .....	175
<b>APPENDIX</b> .....	188
<b>REFERENCES</b> .....	224
<b>CHAPTER 4</b> Recovery from enteric infection demonstrates a shift in functional capacity and metabolite composition .....	235
<b>ABSTRACT</b> .....	236
<b>INTRODUCTION</b> .....	237
<b>METHODS</b> .....	239
Study population .....	239
Sample preparation and metagenome sequencing analysis .....	240
Metagenome assembly and metabolic prediction profiling with Anvi'o.....	240
Metabolic prediction profiling with HUMAnN 3.0.....	241
Ecological analyses of metagenome and metabolome data .....	242
<i>Abundance and diversity analyses</i> .....	242
<i>Differential abundance of metabolic pathways</i> .....	243
<i>Identification of continuous population structure</i> .....	243
Metabolite extraction .....	244
Liquid Chromatography Mass Spectrometry (LC/MS) .....	245
Feature-based Molecular Networking (FBMN) of metabolites.....	246
Intensity normalization and Random Forest in R .....	247
Statistical analysis of metabolites using MetaboAnalyst 5.0.....	248
<b>RESULTS</b> .....	249
Characteristics of the study population.....	249
Variation in the metabolic potential of the gut during and after enteric infection .....	249
Functional differences in metabolic pathways during and after infection .....	252

Specific metabolic pathways differ between sample groups .....	254
Untargeted metabolomics of polar metabolites reveal crucial differences between samples .....	258
Nonpolar metabolites are distinct between infected and recovered metabolomes .....	262
<b>DISCUSSION</b> .....	269
<b>APPENDIX</b> .....	278
<b>REFERENCES</b> .....	335
 <b>CHAPTER 5</b> Conclusions and Future Directions .....	343
<b>REFERENCES</b> .....	350



## LIST OF TABLES

Table A.1. Characteristics of 26 patients with <i>Campylobacter</i> infections (cases) and 44 healthy individuals (controls).....	107
Table A.2. Differentially abundant antimicrobial resistance genes (ARGs) detected in stool samples from cases and controls.....	108
Table A.3. Correlation values between highly abundant antimicrobial resistant genes (ARGs) and specific taxa detected in <i>Campylobacter</i> cases.....	109
Table A.4. Differentially abundant antimicrobial resistance gene (ARG) classes detected in 25 cases with specific resistome profiles (clusters) determined by hierarchical clustering.....	110
Table A.5. Differentially abundant genes detected among cases living in urban versus rural settings. ....	111
Table B.1. Shapiro-Wilk Test results for case and follow-up samples in the microbiome and resistome datasets.....	189
Table B.2. Top-25 co-occurrence associations between <i>Escherichia</i> and ARG groups in cases. ....	189
Table B.3. Co-occurrence associations between <i>Escherichia</i> and other taxa and ARGs in follow-ups. ....	190
Table B.4A. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups. ....	191
Table B.4B. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups. ....	194
Table B.4C. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups. ....	197
Table C.1. Differentially abundant metabolic pathways in cases and follow-ups predicted by HUMAnN 3.0. ....	279
Table C.2. Top-30 polar clusters most important to health status classification via random forest. ....	281
Table C.3. Confusion matrix for classification of samples by health status generated by random forest on polar metabolites.....	282
Table C.4. Output generated by fold-change analysis exploring differentially abundant polar metabolites among cases and follow-ups.....	283

Table C.5. Top-30 nonpolar clusters most important to health status classification via random forest. ....	285
Table C.6. Confusion matrix for classification of samples by health status generated by random forest on nonpolar metabolites.....	286
Table C.7. Output generated by fold-change analysis exploring differentially abundant nonpolar metabolites among cases and follow-ups. ....	286

## LIST OF FIGURES

Figure 2.1. Resistomes in cases are more diverse than resistomes of controls.....	86
Figure 2.2. Resistomes of cases and controls are distinct.....	87
Figure 2.3. Relative abundance of ARGs differs among cases and controls.....	88
Figure 2.4. Hierarchical clustering illustrates group level ARG abundance differences between cases and controls. ....	89
Figure 2.5. Taxonomic relative abundance notably differs between cases and controls.....	91
Figure 2.6. Case Cluster 1 resistomes are more diverse than resistomes of Clusters 2 and 3 combined.....	94
Figure 2.7. Case resistomes cluster separately and case Cluster 2 is more similar to control samples.....	95
Figure 2.8. Diversity among different families is not significantly different. ....	98
Figure 2.9. Beta diversity analyses do not reveal clear similarities among families.....	99
Figure A.1. Sequencing run does not appear to impact resistome similarity among cases and controls.....	112
Figure A.2. Estimated sequencing coverage curves for cases and controls.....	113
Figure A.3. Comparing the average genome size and number of genome equivalents among case and control samples.....	114
Figure A.4. Linear discriminant analysis (LDA) scores showing differentially abundant antimicrobial resistance gene (ARG) classes by health status.....	115
Figure A.5. Linear discriminant analysis (LDA) scores for differentially abundant antimicrobial resistance genes (ARGs) at the group (gene) level by health status.....	116
Figure A.6. Controls display higher taxonomic diversity than cases. ....	117
Figure A.7. Actual abundances of bacterial taxa differ considerably between cases and controls.....	118
Figure A.8. Hierarchical clustering reveals three distinct resistome profiles among the cases..	119
Figure A.9. Linear discriminant analysis (LDA) showing differentially abundant antimicrobial resistance gene (ARG) classes between case clusters. ....	120

Figure A.10. Relative abundance of ARG classes varies across families but maintains the case versus control dichotomy in most circumstances. ....	121
Figure 3.1. Resistome diversity is greater during enteric infection than after recovery. ....	145
Figure 3.2. Resistomes during infection differ significantly from those of recovered samples. ....	146
Figure 3.3. Microbiome diversity is greater after recovering from enteric infection. ....	148
Figure 3.4. Compositional differences between case and follow-up microbiomes are nuanced. ....	149
Figure 3.5. Continuous structure analysis reveals taxonomic gradients driving distribution of samples across the population. ....	152
Figure 3.6. Continuous structure analysis highlights ARG abundance gradients driving differences among cases and follow-ups. ....	154
Figure 3.7. Relative abundance of the Top-10 resistance gene classes notably differs between case and follow-up samples. ....	155
Figure 3.8. Relative abundance of microbial phyla notably differs between cases and follow-ups. ....	157
Figure 3.9. Global network analysis highlighting ARG connections among cases and follow-ups. ....	163
Figure 3.10. Host-tracking via investigation of ARG-carrying contigs reveals genera responsible for harboring ARGs among cases. ....	168
Figure 3.11. Host-tracking via investigation of ARG-carrying contigs reveals genera responsible for harboring ARGs among follow-ups. ....	169
Figure B.1. Average Genome Size (AGS) and estimated number of Genome Equivalents (GE) for paired case and follow-up samples. ....	200
Figure B.2. Metagenomic sequencing coverage of short paired-end reads was determined using Nonpareil. ....	201
Figure B.3. Pairwise comparison of Bray-Curtis dissimilarity among cases and follow-ups. ...	202
Figure B.4. Exploring potential batch effects related to sequencing run using principal coordinate analysis of cases and follow-ups. ....	203
Figure B.5. Assembly coverage statistics quantified by the Quality Assessment Tool for Genome Assemblies (QUAST). ....	204
Figure B.6. Resistome and microbiome diversity analyses between cases, healthy household member (controls) and recovered cases (FollowUp). ....	205

Figure B.7. Various intrinsic factors influence case and follow-up resistomes.....	207
Figure B.8. Alpha and beta diversity of the resistome do not appear to differ across the four different enteric pathogens.....	208
Figure B.9. Multiple intrinsic factors influence case and follow-up microbiomes. ....	210
Figure B.10. Alpha and beta diversity of the microbiome do not appear to differ across the four different enteric pathogens. ....	211
Figure B.11. Continuous structure analysis reveals species gradients among cases and follow-ups. ....	213
Figure B.12. Relative abundance of resistance gene types and classes among cases and follow-ups. ....	215
Figure B.13. Relative abundance of microbial genera notably differ between cases and follow-ups. ....	216
Figure B.14. Differentially abundant ARG classes and groups among cases and follow-ups. ..	217
Figure B.15. Differential abundance of phyla, genera, and species among cases and follow-ups. ....	218
Figure B.16. Differential abundance of taxonomic and resistance gene features for cases and follow-ups with ANCOM-BC.....	219
Figure B.17. Global co-occurrence network analysis reveals interesting patterns among cases and follow-ups. ....	220
Figure B.18. Co-occurrence of Beta-lactam ARGs with other ARGs and taxa varies between cases and follow-ups. ....	221
Figure B.19. ARG-ARG and ARG-taxa connections are different for individuals infected with and recovering from <i>Salmonella</i> .....	222
Figure B.20. ARG-ARG and ARG-taxa connections are different for individuals infected with and recovering from <i>Campylobacter</i> . ....	223
Figure 4.1. Predicted MetaCyc pathways identified via HUMAnN 3.0 indicate significant differences in metabolic potential between cases and follow-ups. ....	251
Figure 4.2. Differentially abundant MetaCyc pathways among cases and follow-ups with UNMAPPED reads removed. ....	255
Figure 4.3. Relative abundances of PWY-5100: pyruvate fermentation to acetate and lactate II among cases and follow-ups.....	257

Figure 4.4. Volcano plot demonstrating fold-change of polar metabolites in cases and follow-ups. ....	260
Figure 4.5. Molecular network and MS <sup>2</sup> spectra for polar Cluster 221 and related clusters in case samples. ....	261
Figure 4.6. Heatmap displaying abundance of the top-50 polar metabolites based on significance determined by paired Wilcoxon tests in cases and follow-ups. ....	263
Figure 4.7. Volcano plot demonstrating fold-change of nonpolar metabolites in cases and follow-ups. ....	266
Figure 4.8. Molecular network and MS <sup>2</sup> spectra for Cluster 2756 and related cluster (2739), which were greatly increased in follow-ups. ....	267
Figure 4.9. Heatmap displaying abundance of the top-50 nonpolar metabolites based on significance determined by paired Wilcoxon tests in cases and follow-ups. ....	268
Figure C.1. Metabolic diversity of KEGG modules significantly differs among patients during infection and after recovery. ....	288
Figure C.2. Investigation of continuous structure within pathways identified by HUMANN 3.0 reveals metabolic tradeoffs. ....	290
Figure C.3. Investigation of continuous structure within module compositions reveals metabolic tradeoffs. ....	292
Figure C.4. Metabolic subcategories and modules demonstrate different frequencies among cases and follow-ups. ....	294
Figure C.5. Relative abundance of KEGG metabolic categories is consistent across cases and follow-ups. ....	295
Figure C.6. Relative abundance of KEGG metabolic subcategories also demonstrate consistency among cases and follow-ups. ....	296
Figure C.7. Investigating relative abundance of KEGG metabolic modules reveals slight discrepancies between cases and follow-ups. ....	297
Figure C.8. Relative abundances of PWY-7254: TCA cycle VII (acetate-producers). ....	298
Figure C.9. Relative abundances of P163-PWY: L-lysine fermentation to acetate and butanoate. ....	299
Figure C.10. Relative abundances of three butanoate production pathways. ....	300
Figure C.11. Relative abundances of P108-PWY: pyruvate fermentation to propanoate I. ....	301

Figure C.12. Relative abundances of PWY-5971: palmitate biosynthesis (type II fatty acid synthase). .....	302
Figure C.13. Relative abundances of lipopolysaccharide (LPS) biosynthesis among cases and follow-ups. ....	303
Figure C.14. Relative abundances of toluene degradation via p-cresol among cases and follow-ups. ....	304
Figure C.15. Richness and composition of polar metabolites significantly differs among patients during infection and after recovery. ....	305
Figure C.16. Richness and composition of polar metabolites does not appear to be influenced by infecting pathogen. ....	306
Figure C.17. Mean decrease in accuracy plot from random forest analysis of polar metabolites. ....	307
Figure C.18. Normalized abundance of Cluster 313 among cases and follow-ups separated by infecting pathogen. ....	308
Figure C.19. Normalized abundance of Cluster 2705 among cases and follow-ups separated by infecting pathogen. ....	309
Figure C.20. Molecular network and MS <sup>2</sup> spectra for three related clusters prevalent in follow-ups. ....	310
Figure C.21. Molecular network and MS <sup>2</sup> spectra for Cluster 2113, which was highly represented in follow-ups. ....	311
Figure C.22. Molecular network and MS <sup>2</sup> spectra for Cluster 2666, which was prevalent in follow-ups. ....	312
Figure C.23. Molecular network and MS <sup>2</sup> spectra for Cluster 970 and related Cluster 318 which were present primarily in cases. ....	313
Figure C.24. Molecular network and MS <sup>2</sup> spectra for Cluster 806, which was abundant in cases. ....	314
Figure C.25. Molecular network and MS <sup>2</sup> spectra for Cluster 313, which was abundant in cases. ....	315
Figure C.26. Richness and composition of nonpolar metabolites significantly differs among patients during infection and after recovery. ....	316
Figure C.27. Richness and composition of nonpolar metabolites does not appear to be influenced by infecting pathogen. ....	317

Figure C.28. Mean decrease in accuracy plot from random forest analysis of nonpolar metabolites. ....	318
Figure C.29. Normalized abundance of Cluster 2659 among cases and follow-ups separated by infecting pathogen. ....	319
Figure C.30. Normalized abundance of Cluster 321 among cases and follow-ups separated by infecting pathogen. ....	320
Figure C.31. Normalized abundance of Cluster 299 among cases and follow-ups separated by infecting pathogen. ....	321
Figure C.32. Mean decrease in accuracy plot from random forest analysis of nonpolar metabolites stratified by infecting pathogen. ....	322
Figure C.33. Normalized abundance of Cluster 2964 among cases and follow-ups separated by infecting pathogen. ....	323
Figure C.34. Normalized abundance of Clusters 6581 and 8369 among cases and follow-ups separated by infecting pathogen. ....	324
Figure C.35. MS <sup>2</sup> spectra for Cluster 321, which was a prevalent nonpolar metabolite in cases. ....	325
Figure C.36. Molecular network and MS <sup>2</sup> spectra for Cluster 1618, which was prevalent among cases. ....	325
Figure C.37. MS <sup>2</sup> spectra for Cluster 244, which was prevalent among cases. ....	326
Figure C.38. Molecular networks and MS <sup>2</sup> spectra for Clusters 4470 and 5193, metabolites found more consistently among follow-ups. ....	327
Figure C.39. The molecular network and MS <sup>2</sup> spectra for Cluster 2964, which was indicated to be more abundant among people with Salmonella infection. ....	328
Figure C.40. The molecular network and MS <sup>2</sup> spectra for Clusters 6581 and 8369, which were indicated to be more abundant among people with Campylobacter infection. ....	329
Figure C.41. Relative abundances of PWY-5675: nitrate reduction V among infected and recovered patients. ....	330
Figure C.42. Relative abundances of various arginine metabolism pathways among infected and recovered patients. ....	331
Figure C.43. Relative abundances of various ornithine metabolism pathways among infected and recovered patients. ....	333
Figure C.44. Relative abundances of PWY-6803: phosphatidylcholine acyl editing among infected and recovered patients. ....	334



## KEY TO ABBREVIATIONS

AAD	Antibiotic-Associated Diarrhea
ABC	ATP-Binding Cassette
ACC	ARG-carrying Contig
ACSSuT	Ampicillin, Chloramphenicol, Streptomycin, Sulfamethoxazole, Tetracycline
AGC	Automatic Gain Control
AGS	Average Genome Size
AMR	Antimicrobial Resistance
ANCOM-BC	Analysis of Compositions of Microbiomes with Bias Correction
ARG	Antimicrobial Resistance Gene
ATP	Adenosine Triphosphate
BA	Bile Acid
BEH	Ethylene Bridged Hybrid
BH	Benjamini-Hochberg
BHT	Butylated Hydroxytoluene
BLAST	Basic Local Alignment Search Tool
BSH	Bile Salt Hydrolase
BWA	Burrows-Wheeler Aligner
CA	Cholic Acid
CA	Correspondence Analysis
CAP	Cationic Antimicrobial Peptide
CARD	Comprehensive Antibiotic Resistance Database
CBA	Conjugated Bile Acid
CD	Crohn's Disease

CDA	Chenodeoxycholic Acid
CDC	Centers for Disease Control and Prevention
COG	Clusters of Orthologous Genes
DA	Deoxycholic Acid
DA	Differential Abundance
DNA	Deoxyribonucleic Acid
EHEC	Enterohemorrhagic <i>Escherichia coli</i>
ESBL	Extended-Spectrum Beta-Lactamase
FBMN	Feature-Based Molecular Networking
FC	Fold Change
FDR	False Discovery Rate
FMT	Fecal Microbiota Transplant
GC/MS	Gas Chromatography/Mass Spectrometry
GE	Genome Equivalent
GNPS	Global Natural Product Social Molecular Networking
HGT	Horizontal Gene Transfer
Hi-C	High-Throughput Chromosomal Confirmation Capture
HILIC	Hydrophilic Interaction Liquid Chromatography
HMM	Hidden Markov Model
HMP	Human Microbiome Project
HPLC	High-Performance Liquid Chromatography
HUMAnN	HMP Unified Metabolic Analysis Network
HUS	Hemolytic Uremic Syndrome
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome

ICE	Integrative and Conjugative Element
IEC	Intestinal Epithelial Cell
IL	Interleukin
iMP	Intestinal Mononuclear Phagocyte
IRB	Institutional Review Board
ITS	Internal Transcribed Spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC/MS	Liquid Chromatography/Mass Spectrometry
LCA	Lithocholic Acid
LDA	Linear Discriminant Analysis
LEfSe	Linear Discriminant Analysis Effect Size
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
MAG	Metagenome-Assembled Genome
MATE	Multidrug And Toxic Compound Extrusion
MDA	Mean Decrease in Accuracy
MDHHS	Michigan Department of Health and Human Services
MDR	Multidrug Resistance
MDRC	Multidrug Resistance Cluster
MDSS	Michigan Disease Surveillance System
MFS	Major Facilitator Superfamily
MGE	Mobile Genetic Element
MLS	Macrolide, Lincosamide, Streptogramin
MMUPHin	Meta-Analysis Methods with Uniform Pipeline for Heterogeneity in Microbiome Studies
MS <sup>2</sup>	Tandem Mass Spectrometry

MSU	Michigan State University
NCBI	National Center for Biotechnology Information
NCHS	National Center for Health Statistics
NK	Natural Killer
NLR	Nod-Like Receptor
NMDS	Nonmetric Multidimensional Scaling
NO	Nitric Oxide
NTS	Nontyphoidal <i>Salmonella</i>
OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
PERMDISP	Permutational Multivariate Analysis of Dispersions
PSD	Pairwise Alignment Sequence Dissimilarity
PWY	Pathway
QUAST	Quality Assessment Tool
RDP	Ribosomal Database Project
RF	Radio Frequency
RGI	Resistance Gene Identifier
RND	Resistance-Nodulation-Cell Division
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
RPK	Reads Per Kilobase
rRNA	Ribosomal Ribonucleic Acid
RTSF	Research Technology Support Facility

SASARI	<i>Salmonella</i> Streptomycin and Azithromycin Resistance Island
SCFA	Short-Chain Fatty Acid
SMR	Small Multidrug Resistance
SMRT	Single-Molecule Real-Time
SNP	Single Nucleotide Polymorphism
STEC	Shiga Toxin-Producing <i>Escherichia coli</i>
TCA	Tricarboxylic Acid
TCDC	Taurochenodeoxycholic Acid
TE	Total Extract
TLR	Toll-Like Receptor
TMA	Trimethylamine
TMAO	Trimethylamine N-oxide
TRACA	Transposon-Aided Capture
TSS	Total Sum Scaling
UC	Ulcerative Colitis
UHPLC	Ultra-High-Performance Liquid Chromatography
UMP	Uridine Monophosphate
UPLC	Ultra-Performance Liquid Chromatography
VFDB	Virulence Factor Database

## **CHAPTER 1**

Literature Review: The Human Gut Microbiome, Antibiotic Resistance, and Enteric Infection

## **MICROBIAL ECOLOGY OF THE HUMAN GUT**

Humans and microbes have been interacting for millennia. Microbes have come to colonize nearly every area of the human body and play key roles in shaping human health. Although previous work estimated the ratio of bacterial to human cells in the body to be 10:1, more refined recent research suggests that this ratio is closer to 1.3:1 (1, 2). These microbial symbionts compose the human microbiota; the comprehensive set of genomes belonging to the microbiota is termed the ‘microbiome’ and holds the key to various traits that broadly benefit human health (3). The human gut microbiome has been extensively explored in periods of health and disease (4). Yet, the complexities of this ornate relationship between the gut microbiota and their human host require ongoing exploration and elucidation.

### **Community assembly of the early human gut microbiome**

Establishment of the gut microbiome begins very early in life. However, the composition of gut microbiota at birth differs immensely from that of adults, and the nuances of community assembly continue to be investigated (5, 6). From the perspective of ecological theory, various frameworks have been considered for assembly of the human gut microbiome. Neutral theory is an important model which assumes the only driving factor shaping microbial communities is random chance; it negates ecological forces such as dispersal, diversification, ecological drift, and selection while also dismissing any species-level differences (6, 7). This theory assumes that most species share the same general niche (also referred to as “ecological equivalence”), suggesting that multiple members of the community may be capable of fulfilling certain community functions, a phenomenon called “functional redundancy” (3, 8). Although this theory serves as an important null model and successfully captures many aspects of the human gut microbiome, dismissal of other ecological processes such as dispersal and selection can be

limiting. Another ecological theory which considers the dynamic nature of microbial populations in addition to community-level traits is metacommunity theory. This theory considers the ecological world to be comprised of habitable patches that are spatially distinct; dispersal of species among these separate patches or communities results in a metacommunity (6, 9). This framework is especially useful for characterizing host-associated communities such as the gut microbiota, as it considers dispersal among patches (host-to-host or host-to-environment) as well as environmental selection (such as diet, antibiotic use, or disease state) when identifying forces driving community assembly and maintenance (6, 10). Indeed, metacommunity theory enables researchers to explore and predict community trajectories during succession, whether this be related to initial community establishment, recovery from an environmental disturbance (e.g., antibiotics), or in response to a microbial invasion (e.g., influx of pathogens) (6).

Empirically, many studies have been performed to explore the early stages of gut microbiome assembly during infancy. The initial composition of the infant gut microbiota has been found to depend on method of delivery, as vaginally delivered neonates were dominated by members of their mothers' vaginal microbiota (including *Lactobacillus* and *Prevotella*), while those delivered via Caesarean section were primarily colonized by skin-related bacteria such as *Staphylococcus*, *Corynebacterium*, and *Propionibacterium* spp. (11). Multiple studies have gone on to further document the trajectories of gut microbiome composition during the ensuing years of infancy as well. Ongoing community assembly of the infant gut was found to be nonrandom, as this process appears to be punctuated by discrete developmental events related to the expansion of specific community members (12, 13). Notably, this process was observed for both preterm and full term neonates, although the pace of progression observed in preterm babies was slower (14). Additionally, maturation of the gut microbiota in preterm infants was shown to be



influenced by gestational age at time of birth (14), whereas delivery mode (vaginal vs. C-section) and feeding patterns (breastfed, bottle-fed, or solid food) had significant effects on gut microbiota assembly in term babies (13). This study also demonstrated that cessation of breastfeeding, rather than introduction of solid foods, was associated with the development of adult-like microbiota, suggesting that breastfeeding acts as a primer for the gut microbiota, which can influence other aspects of metabolic and immune health (13).

### **Diversity and composition of the human gut microbiota**

Although the composition of gut microbial communities is dynamic and fluctuates over time, general patterns of diversity define a healthy human gut. Most taxa identified in the gut belong to two bacterial phyla: *Bacteroidetes* and *Firmicutes* (3, 15, 16). Other key members include *Actinobacteria*, *Proteobacteria*, *Fusobacteria* and *Verrucomicrobia*, though these taxa are present in significantly smaller numbers with some variation depending on the source (15). A study using 16S rRNA sequencing, for instance, identified over 60 OTUs in stool samples from 17 healthy adults that were prevalent in >50% of samples and included *Faecalibacterium*, *Ruminococcus*, *Eubacterium*, *Dorea*, *Bacteroides*, *Alistipes*, and *Bifidobacterium* (17). Indeed, another study investigating fecal samples from 124 individuals identified 75 taxa that were common to >50% of samples, and 57 that were present in >90% of samples (18). Turnbaugh et al. (2009) (19) also explored the microbiota of monozygotic and dizygotic twin pairs and demonstrated that the gut microbiome is shared among family members, though the specific bacterial lineages varied. Despite some of these similarities, the relative abundances of these microbial members vary significantly among different individuals and even within the same individual over time (16), trends that may be linked to diet and disease status (20).

Numerous factors can influence the composition and diversity of the gut microbiota. Age, host genetics, diet, environment, and disease have all been found to impact the composition and functionality of the gut microbiome (16). Disease status, in particular, is an extensively explored area of the microbiome field, though most studies have focused on chronic versus acute conditions. People with conditions such as inflammatory bowel disease (IBD) (including ulcerative colitis (UC) and Crohn's Disease (CD)), obesity, Type 2 diabetes, and others, for instance, were shown to have reduced diversity in their gut microbiome (4, 21). For example, obese individuals had markedly lower numbers of *Bacteroidetes* and overall higher levels of *Firmicutes* compared to lean counterparts (20). However, after practicing diet therapy, these individuals lost weight and developed microbiomes that more closely mirrored those of the lean cohort. Despite the finding that obese individuals display an increased *Firmicutes:Bacteroidetes* ratio (which has been corroborated in a handful of other studies), other groups have documented conflicting results. In fact, Walters et al. (2014) (22) deduced that signatures of obesity are not consistent between studies and demonstrated a lack of significance when performing a meta-analysis of compositional differences. Nonetheless, this meta-analysis identified notable microbial signatures defining communities belonging to individuals with and without IBD. Specifically, patients with IBD had decreased abundances of *Firmicutes* and *Bacteroidetes* and increased proportions of *Proteobacteria* and *Actinobacteria* (22), which was also observed in a prior study (23). For patients with Type 2 diabetes, an increased abundance of *Betaproteobacteria* and decrease in *Clostridia* abundance was observed and the ratio of *Bacteroidetes:Firmicutes* correlated with higher blood glucose levels (24). These findings illustrate that disease state is intricately related to gut microbiota composition, which has broad implications for metabolic health.

Although many studies have documented compositional variation across individuals, studies exploring functional diversity of the gut microbiome have demonstrated highly similar functional gene profiles (16, 19). Even among individuals with few genus- and phylum-level similarities, a majority of metabolic pathways are present in comparable proportions (18, 25). In a study of monozygotic and dizygotic twins, for instance, a ‘core microbiome’ was uncovered at the gene level rather than organismal level, suggesting that redundancy of community function may be more important than community composition (19). Importantly, shifts in the overall functional capacity of the gut microbiome may be indicative of alterations in physiological state (16). It is clear that disease status influences and can be influenced by the metabolic health of the gut microbiome, a concept that will be explored extensively in subsequent sections.

### **The gut microbiome during periods of ecological change**

Although a healthy adult gut microbiome maintains a stable state of composition and function, this homeostatic state is not always resistant to disturbance. Microbial communities experiencing disturbance often lose a proportion of members, creating opportunities for new or remaining community members to increase in abundance (6). Those communities that can withstand the effects of disturbance and remain relatively unchanged are considered highly resistant (26). Whether a community can successfully return to its original pre-disturbance stable state after the onslaught of disturbance is indicative of its resilience (26, 27). Communities that fail to resist changes brought on by disturbance may still display high levels of resilience.

One of the most well-studied disturbances relevant to the human gut microbiome is the use of antibiotics. Although antibiotic treatment is designed to eliminate pathogenic bacteria, the systemic effects of antibiotics can also detrimentally impact beneficial commensals (6). Multiple studies have documented a marked decrease in microbial diversity in the gut during antibiotic

therapy, regardless of drug type (28-31). Each of these studies also observed partial recovery of the gut microbiota and a return to a stable state following antibiotic cessation. Despite this, certain microbial taxa failed to recover after multiple weeks even in the absence of antibiotics. Palleja et al. (2018) (31), for instance, showed that antibiotic treatment greatly reduced the number of butyrate-producing bacteria including multiple *Faecalibacterium prausnitzii* strains as well as *Eubacterium* and *Coprococcus* spp. Even at 180 days post-antibiotics, six of the eight *F. prausnitzii* strains were no longer detectable in the gut community. This partial reduction or long-term elimination of these beneficial bacteria can have a detrimental impact on gut health. For example, Young and Schmidt (2004) (30) characterized the onset of antibiotic-associated diarrhea (AAD), a condition that is linked to antibiotic use, which results in severe diarrhea without clear cause. Relatedly, it is widely understood that a reduction in gut richness and diversity following antibiotic exposure creates prime conditions for the expansion of pathogens such as *Clostridium difficile*, a known contributor to AAD (32, 33). *C. difficile* is a common nosocomial pathogen whose survival and transmissibility can be attributed to its ability to form spores (34). Importantly, approximately 30% of patients with *C. difficile* infection experience recurrence or reinfection (35). The severity of these infections can, at times, only be addressed with fecal microbiota transplant (FMT) in which stool from a healthy donor is used to seed the suffering patient's gut community with beneficial microbes (36). Certainly, the negative consequences related to antibiotic use provide ample evidence that this treatment leads to widespread disturbance of the microbial gut environment.

Another ecological phenomenon relevant to the human gut is invasion. In microbial invasions, a foreign microbe, such as a pathogen, is introduced into a stable environment. If the pathogen is successful in establishing itself in this environment and overcoming colonization

resistance, it can uproot original community members (37). The ability of microbes to invade a given community depends upon that community's "niche opportunity," which refers to conditions that would promote microbial invasion (38). Examples include available resources, optimal abiotic conditions, interactions with resident microbiota, and relative flux within the community. In the human gut environment, the host, in conjunction with the resident commensal microbiota, dictate these niche opportunities (6). The role of the host and its associated gut microbiota in resistance to pathogen invasion is discussed in great detail in a later section.

Despite their relevance, microbial invasions are relatively understudied compared to macroorganisms. Even among studies that explore microbial invasion ecology, there are inconsistencies in how researchers study, execute, and interpret invasions (39). Kinnunen et al. (2016) (39), for instance, argues that microbial invasion, when viewed through the lens of community ecology, is biased towards selection with evident negation of other processes such as dispersal or drift. Vila et al. (2019) (40) sought to further characterize eco-evolutionary "rules" for microbial invasions that encapsulate some of these other ecological processes. Their simulation-based "nearly neutral" model identified five of these rules: 1) greater fitness increases the chance of persistence after invading, an attribute more often found in larger communities; 2) if invaders are relatively poor competitors, they are not likely to succeed in the community; 3) if invaders' competitive capacity is somewhat neutral, propagule pressure can determine the outcome of invasion (i.e., a greater density of invaders is less likely to result in stochastic extinction); 4) increased diversity of invaders results in similar outcomes as having a higher density of invaders; 5) more diverse resident communities show greater success in resisting invasions, an attribute likely related to resource partitioning and competition. Indeed, previous findings appear to corroborate these rules. One such study sought to characterize the importance

of resident community structure, invader diversity, and stage of succession and found that co-invasion by genotypically distinct bacteria resulted in greater persistence and alteration of resident community structure (41). These findings align well with Vila et al.'s Rule 4. Another study explored microbial invasions *in vitro* and determined that during the early stages of invasion, specifically, propagule pressure was the strongest explanatory variable for invasion success (42), a finding that has been corroborated in algal communities (43) and is in line with Rule 3. Importantly, later stages of invasion appeared to be more heavily dictated by community diversity and composition (42), suggesting that different ecological phenomena influence invasion success at different times during invasion.

Microbial invasion in the human gut is often explored in the context of dysbiosis or infection, as pathogens are non-resident microbes disrupting a resident community. Baumgartner et al. (2021) (44) used a gut microcosm system to elucidate which factors contributed to greater susceptibility and resistance to pathogen invasion. This study found that taxonomic composition played a crucial role in colonization resistance; the abundance of their invading strain was starkly reduced in microcosms containing human gut microbiota than their community-free control. These researchers also explored the role of abiotic factors. Notably, they found that changing abiotic conditions (e.g., nutrient type and availability) did not directly suppress the invading strain. However, altering nutrient status did impact the resident communities' ability to resist and suppress invasion (44). Examples of microbial invasions within human hosts include various pathogens. Just as *Clostridium difficile* exploits the dysbiotic state of the human gut microbiota after an antibiotic-induced disturbance, this microbe is also an unwelcome invader. Although it can exist as a member of the resident microbiota in some individuals, *C. difficile* transforms into an opportunistic pathogen in the vacuum created by antibiotic therapy and is capable of rapidly

taking over niche space (45). True pathogens have also been explored for their role in gut invasion. In a prior study from our lab, the impacts of enteric infection were explored using 16S rRNA sequencing for patients infected with one of four enteric pathogens including *Campylobacter*, *Salmonella*, *Shigella*, and Shiga toxin-producing *E. coli* (STEC). This study demonstrated that patients with infection demonstrated markedly lower gut microbiota diversity than healthy controls with distinct microbiota profiles (46). Notably, infected patients had drastically increased abundances of *Proteobacteria*. The influence of various ecological processes on the trajectory, health, and overall function of the human gut microbiota is undoubtedly extensive.

#### **SIGNIFICANCE OF ENTERIC INFECTION AND NOTABLE REPERCUSSIONS RELATED TO ENTERIC PATHOGENS**

Greater than 9.4 million foodborne infections caused by enteric pathogens occur each year in the United States (47). In 2020, the Centers for Disease Control and Prevention (CDC) reported increased incidence of infections caused by *Campylobacter* and Shiga toxin-producing *Escherichia coli* (STEC), among others; *Salmonella* and *Shigella* also maintained a high level of incidence (48). Each of these four pathogens is recognized as a leading cause of diarrheal disease among humans and has been implicated by their detection via culturing and nucleic acid amplification methods in the clinic (49). Importantly, the consequences of enteric infection go beyond experiencing symptoms; Singh et al. (2015) (46) found that gut microbiome composition changed drastically during infection compared to uninfected, healthy controls.

Although *Campylobacter*, *Salmonella*, *Shigella*, and STEC are all Gram-negative pathogens, each of these microbes behaves differently during infection. Nonetheless, all trigger inflammation in the gut during infection (50-53). While *Salmonella*, *Shigella*, and STEC are represented within the family *Enterobacteriaceae*, *Campylobacter* belongs to

*Campylobacteraceae*. *Campylobacter jejuni*, in particular, infects humans by penetrating the intestinal mucus barrier and invading epithelial cells, where it typically elicits a strong immune reaction (50). Most *Campylobacter* infections are self-limiting, however, some individuals can develop severe autoimmune disorders such as Guillain-Barré syndrome post-infection. This syndrome is classified as a neuropathological condition that results due to mimicry between the *Campylobacter* lipooligosaccharide (LOS) and human gangliosides (50).

Non-typhoidal *Salmonella* also invade intestinal epithelial cells, specifically specialized microfold (M) cells, and can replicate intracellularly during infection (54). In fact, *Salmonella* are capable of transforming intestinal epithelial cells into M cells, thereby promoting their own colonization and invasion (55). *Salmonella* is known for its many pathogenicity islands found to be conserved among multiple serovars; these represent gene clusters that encode various virulence factors important for adhesion, invasion, intracellular survival, and immune evasion (56). *Shigella* can also invade the intestinal epithelium resulting in high levels of intestinal inflammation (57). Although *Shigella* can invade host cells, it does not enter the intestinal epithelium at the apical side, but rather invades M cells and enters epithelial cells from the basolateral side (58, 59). The ability to spread adjacently through the epithelium prevents exposing *Shigella* and other pathogens to extracellular immune components, which can prolong infection (57).

By contrast, STEC behaves differently from these other enteric pathogens in that it does not invade the intestinal epithelium. This diverse group of pathogens is known for its severity of infection, as a minimal infectious dose can cause hemorrhagic colitis and hemolytic uremic syndrome (HUS) in some cases, which can lead to acute renal failure and death (60). The notable virulence trait of STEC is due to its production of one or more Shiga toxins, which are AB<sub>5</sub>



toxins that halt protein synthesis in host cells leading to cell death (61). Although STEC does not invade, it uses an adherence mechanism that contributes to massive rearrangements of the intestinal epithelium, resulting in inflammation, and Shiga toxin production has been linked to upregulation of host immune genes (62). While STEC is not invasive, the Shiga toxins are translocated across into the intestinal epithelium where they are trafficked to the ribosome to irreversibly inhibit host protein synthesis (63). Despite the differences in pathogenic mechanisms, these four enteric pathogens can result in similar conditions and alterations in the human gut. Moreover, the rising prevalence of each is of great concern and novel prevention and treatment strategies are needed. Defining how specific pathogens alter the composition and function of the gut microbiome may lead to the development of such strategies to impact gastrointestinal health.

## **MECHANISMS OF COLONIZATION RESISTANCE PROTECT THE GUT FROM ENTERIC INFECTION**

The commensal gut microbiota use multiple mechanisms to control and counteract invading enteric pathogens (64). One of these mechanisms is direct interaction through which commensals secrete antimicrobial compounds, alter the gut environment, outcompete pathogenic bacteria for common nutrients, or produce compounds which downregulate pathogen virulence. Certain commensals produce bacteriocins, small peptides with bacteriostatic or bactericidal activity against a narrow range of species often related to the producing strain (65). For example, *Enterococcus faecium* mediates protective expression of mucin and antimicrobial peptides via a secreted peptidoglycan hydrolase, SagA, in mice (66), suggesting that similar mechanisms take place in humans. Additionally, various residential gut microbes can produce short-chain fatty acids (SCFAs) such as butyrate, propionate, and acetate, which alter the gut pH and have been shown to suppress proliferation of enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 (67), a

more virulent form of STEC. Similarly, Fukuda et al. (2011) (68) found that *Bifidobacterium* spp. produced acetate that provided defense against cellular translocation of *E. coli* O157:H7 in mice, prolonging survival. Additionally, butyrate has been shown to down-regulate expression of pathogenicity island 1 (SPI1), which contains a set of genes necessary for epithelial invasion in *Salmonella enterica* (69). Commensal gut microbes will also outcompete invading pathogens for various nutrients; this has been shown in multiple studies involving *E. coli* O157:H7 through competition for proline (70), organic acids (71), as well as various carbohydrates (72). Resource competition was also implicated as an important factor in reducing *S. enterica* serovar Typhimurium colonization, as the commensal *E. coli* Nissle outcompeted this pathogen for iron, an important micronutrient (73). Perhaps one of the most sophisticated methods of colonization resistance is the ability of host commensals to mitigate virulence of enteric pathogens. For example, certain enteric pathogens such as *Shigella flexneri* are dependent on oxygenation in a predominantly anaerobic environment (74); hence, utilization of oxygen by commensal bacteria may contribute to reduced pathogenicity or survival of this pathogen.

The commensal gut bacteria also indirectly counteract pathogens by priming and communicating with the host immune system. Vaishnava et al. (2008) (75) found that Paneth cells in the small intestine interact directly with the gut microbiota, specifically through MyD88-dependent toll-like receptor (TLR) activation. This sensing triggers production of host-excreted antimicrobial peptides, such as  $\alpha$ -defensins, which can influence the composition of the commensal gut microbiota, regulate the mucosal immune response (76), and stymie pathogen colonization (75). Furthermore, commensal bacteria condition the innate immune system's production of natural killer (NK) cells that express the natural cytotoxicity receptor NKp46 and the transcription factor ROR $\gamma$ t, which regulate production of the interleukin-22 (IL-22) cytokine

(77). These factors were found to be important for antimicrobial protein-mediated immune defense against *Citrobacter rodentium* infection in mice (78, 79), which mimics EHEC infections in humans. NOD2, or the nucleotide-binding oligomerization domain 2, is another important factor, as it senses components of the bacterial peptidoglycan and can interact with other immune machinery such as TLRs (80). Indeed, NOD2 and the commensal gut microbiota were previously shown to be involved in an intricate regulation feedback loop (81). This study demonstrated that commensals positively regulated NOD2 signaling, priming it for activation upon pathogen invasion, while this increased signaling also limited overgrowth of commensal microbes (81). In another study, intestinal mononuclear phagocytes (iMPs) differentially responded to commensal and pathogenic bacteria; specifically, they mediated NLRC4, a Nod-like receptor (NLR)-containing inflammasome responsible for production of IL-1 $\beta$ , a key inflammatory cytokine (82). It is therefore evident that, in addition to direct antagonism against invading pathogens, commensal gut microbiota play important roles in readying the host immune response against these invaders.

Despite the efforts of commensal microbes to reduce risk of infection, pathogenic isolates have developed methods to work around these defenses. For example, in causing intestinal inflammation, enteric pathogens alter the gut environment and shift the composition of microbiota, thereby altering available nutrients which can enable further colonization of pathogenic strains (83). Another notable example of nutrient pilfering is the ability of pathogenic *Escherichia* and *Salmonella* strains to uptake Fe(III) despite host-mediated iron sequestration methods designed to prevent such occurrences (84). More broadly, nutrient use among gastrointestinal pathogens enhances their ability to partition resources, outcompete commensal gut residents, evade immune responses, and increase cell-to-cell signaling which could promote

virulence (85). Some pathogens, such as *S. enterica* serovar Typhimurium, have evolved to thrive in an inflamed gut environment. Interestingly, this pathogen can use tetrathionate, a compound generated from host-produced reactive oxygen species during an inflammatory response, as an alternative electron acceptor during respiration (86). Despite the many mechanisms displayed by both commensal microbes and the human host to prevent infection, enteric pathogens continue to cause widespread foodborne disease.

### **OTHER FACTORS INFLUENCING THE TRAJECTORY AND SEVERITY OF ENTERIC INFECTIONS**

Microbiome health at the time of infection also plays a role in determining the outcome of pathogen invasion. The ability of a community to return to its pre-disturbance state is indicative of its resilience. In fact, Moya and Ferrer (2016) (87) describe metabolic plasticity and functional redundancy as the greatest influencers of gut microbiome trajectories induced by differences in age, diet, and disease. Metabolic plasticity has been described as a change in single-cell properties that adjust depending on the needs of the overall community (88). Functional redundancy, on the other hand, refers to a community's capacity for metabolic function regardless of taxonomic or genetic composition. In other words, different species have the capacity to complete the same function(s) within a community (89). Comte et al. (2013) (88) explored the interplay of metabolic plasticity and functional redundancy in aquatic microbial communities after transplantation to new nutrient sources. Their results suggest that these phenomena may not dictate community response to environmental factors, but rather modulate factors such as composition and diversity, which are likely playing key roles in this response. Indeed, functional redundancy among gut communities has been documented during and after antibiotic treatment, as original functionality was shown to be restored upon treatment cessation (90).

A gut environment that is already experiencing periods of distress or dysbiosis is at greater risk of further disruptions by pathogen presence, an occurrence which could have lasting effects on gut dysbiosis (27). Kampmann et al. (2016) (91), for instance, demonstrated that taxonomic diversity of the gut microbiota was lower among individuals who eventually became infected with *Campylobacter* compared to people who remained uninfected. A similar outcome was observed in mice as those with relative microbial imbalance prior to infection were more susceptible to infection by *S. enterica* (92). Even stressors external to the gut, such as those affecting organ health, can negatively influence the body's ability to react to infection. Hyperglycemia, for example, which is caused by increased levels of glucose in the body, was found to disrupt gut epithelial integrity in mice, leading to weakened barrier function that increased susceptibility to enteric infection and pathogen spread (93). Interestingly, there are various markers that can indicate whether a human gut environment is trending towards dysbiosis. For example, one of the most documented taxonomic shifts related to dysbiosis is a substantial "bloom" in Proteobacteria (94). Although members of this phylum are normal residents of a healthy microbiome in relatively small numbers, increased proportions have been implicated in obesity and diabetes (95) in addition to more transient disturbances such as enteric infection (46). It is clear, however, that more comprehensive analytical methods such as shotgun metagenomic sequencing and metabolomics, are needed to fully understand how enteric pathogens impact gut communities.

## **RELEVANCE OF ANTIBIOTIC RESISTANCE TO THE HUMAN GUT MICROBIOME AND ENTERIC PATHOGENS**

Antibiotic resistance among bacterial pathogens is an imminent global concern. While antibiotics have historically served as lifesaving treatments used to combat bacterial infections, resistance has been documented for nearly all antibiotics developed (96). The ubiquity of

resistance has arisen due to multiple factors including rampant overuse of antibiotics and improper use and prescription of these drugs in veterinary and human medicine as well as extensive application in the agricultural industry. Additionally, research into the development of new drugs is lacking, as is the funding (96). Despite recent decreases in the overall number of hospitalizations and deaths associated with resistant bacterial infections, the CDC emphasizes ongoing challenges in combatting resistance (97). The most recent estimates indicate that over 2.8 million antibiotic-resistant infections occur annually in the United States, resulting in greater than 35,000 deaths (97). In addition, infections caused by resistant bacteria were estimated to result in significantly higher financial and physiological costs to the patient (98). Importantly, the burden of antibiotic resistance is more prevalent in low- and middle-income countries (LMICs) which, on average, have fewer regulations for antibiotic use, higher rates of infectious disease, and less programming that emphasizes the importance of antibiotic resistance monitoring and prevention (99).

### **Antibiotics and Antibiotic Resistance**

Antibiotics have been referred to as “wonder drugs” for their ability to combat bacterial infection (100) and have greatly changed the trajectory of human health throughout history (101). Compounds which exert antibiotic activity are either naturally occurring or synthetically produced. While naturally occurring antibiotics are secondary metabolites produced by bacteria or fungi to combat other microbes sharing the same environment, synthetic antibiotics have been developed in the laboratory to optimize antimicrobial activity (102). Importantly, the most common mechanisms of antibiotic compounds include inhibition of cell wall synthesis, destruction of the cell membrane, inhibition of nucleic acid or protein synthesis, and disruption of crucial metabolic pathways (102, 103). These mechanisms have bacteriostatic effects, which

halt bacterial growth, or bactericidal effects that cause bacterial death; this range of mechanisms is relevant to clinical applications (104). Certainly, the development and application of antibiotics have substantially altered patient outcomes related to infectious disease.

Despite the wide-reaching efficacy of antibiotics, resistance to these compounds has been documented for nearly every drug in circulation (96, 100). Antimicrobial resistance (AMR) is defined as the ability to withstand the molecular impacts of an antimicrobial or antibiotic, allowing continued survival and growth (100, 101). AMR is a natural evolutionary response to the selective pressure of antibiotic exposure; the presence of antibiotics in a microbial community will select for members with mutations and genetic components that promote survival (105). A recent study demonstrated that although selection is the predominant force at play in the development and maintenance of AMR, both history and chance also influenced resistance phenotypes, evolved resistance, and sensitivity networks (106).

Antibiotic resistance can be innate or acquired. Innate resistance refers to a bacterium's ability to withstand antibiotic exposure due to its natural genetic makeup or structure. For example, there is intrinsic resistance among Gram-positive pathogens to the drug aztreonam, a beta-lactam antibiotic that cannot bind sufficiently to Gram-positive cell wall components (107). Acquired resistance, however, refers to a bacterium acquiring the ability to resist antibiotics via genetic exchange with other microbes or spontaneous mutation (101). The most common mechanisms of AMR include limiting drug uptake, modifying the molecular target of the drug, directly inactivating the drug itself, and removing antibiotics from the cell via efflux (101). Drug uptake can be modulated by the lipopolysaccharide (LPS) layer in Gram-negative bacteria, thickening of the cell wall in Gram-positive bacteria, formation of biofilms (107, 108), or reduced production and activity of porin proteins, which is common for members of

*Enterobacteriaceae* (109, 110). Drug target modification is a resistance mechanism used to prevent antibiotic activity against biosynthetic pathways, in particular. For example, target modification can take place in the bacterial cell wall (preventing attack on cell wall synthesis pathways), the ribosome (protecting the cell's protein synthesis capabilities), DNA gyrase (preventing binding of nucleic acid synthesis inhibitors) and metabolic intermediates (allowing reactions to continue, such as in the folate biosynthesis pathway) (107). Inactivation of the drug occurs by destroying the drug, as is the case with beta-lactamase proteins, which cleave the beta-lactam ring rendering the antibiotic inviable (101, 107, 111). Drugs can also be inactivated through chemical modification, in which bacterial enzymes attach chemical groups to the antibiotic compound; these can include acetyl, adenylyl, and phosphoryl groups, all of which have been shown to inactivate aminoglycosides via aminoglycoside-modifying enzymes, in addition to other antibiotic classes (101, 112).

Drug efflux is a resistance mechanism that is particularly concerning as many efflux pumps can extrude multiple compounds, an ability that heightens the likelihood of multidrug resistance (MDR) in some organisms (107). There are five main families of bacterial efflux pumps including the ATP-binding cassette (ABC) family, the small multidrug resistance (SMR) family, the major facilitator superfamily (MFS), the multidrug and toxic compound extrusion (MATE) family, and the resistance-nodulation-cell division (RND) family (107, 113). While Gram-positive bacteria have single-component efflux pumps that span the cytoplasmic membrane, Gram-negative microbes employ multi-component pumps to traverse the inner membrane, periplasm, and outer membrane (114). The most common efflux pumps among Gram-negative bacteria are tripartite systems in the RND family that confer resistance to various biocides, detergents, and solvents in addition to many clinically-relevant antibiotics (114).



Although most antimicrobial resistance genes (ARGs) for MDR efflux systems are located on the bacterial chromosome, genes for specific efflux transporters (such as those relevant to macrolide, lincosamides, and streptogramins (MLS) or tetracyclines), have been associated with mobile genetic elements (MGEs), highlighting their ability for exchange among bacteria (115).

### **Mobility of antibiotic resistance**

Many bacteria acquire antibiotic resistance genes or determinants through horizontal gene transfer (HGT), which results in resistance phenotypes faster than via the introduction of spontaneous mutations (116). HGT occurs through three main mechanisms: conjugation, transformation, and transduction. While transformation involves the uptake of exogenous DNA from the surrounding environment, conjugation and transduction require more machinery. In conjugation, a donor and a recipient cell are directly attached by a sex pilus through which genetic material is exchanged; this represents the most common form of HGT among enteric pathogens (117). Transduction, on the other hand, relies on bacteriophage infection to transfer genetic material from one bacterial cell to another. Indeed, bacteriophages can transfer ARGs across bacterial communities, which is important to consider when developing and applying therapeutics (118).

ARGs are not the only important genetic components that can be transferred across bacterial populations via conjugation. Rather, a variety of MGEs including conjugative plasmids, conjugative transposons, and integrative and conjugative elements (ICEs) can all be passed from one cell to another through conjugation (119). Therefore, the capacity for drug resistance is enhanced because of the ease by which they spread among bacteria within and across species. Conjugative plasmids are specialized plasmids which contain the genetic information needed to induce conjugation (120). In addition to the genetic infrastructure enabling spread, conjugative

plasmids typically possess accessory genes, such as ARGs, which may confer an adaptive advantage to the host cell (120). Alarming, conjugative plasmids were shown to disseminate among microbial populations even in the absence of antibiotic-related selective pressures, suggesting that cessation of antibiotic treatment, even for extended periods, may not mitigate AMR in a population (121).

Meanwhile, conjugative transposons, which are plasmid-like and transposon-like representing discrete genetic elements that can integrate into the bacterial genome, are also important. In conjugative transfer, the portions of DNA that contain ARGs are excised, nicked to form single strands (one of which is transferred to the recipient cell), constructed back to double-stranded DNA, then integrated back into the bacterial genome (122). Finally, ICEs are self-transmissible elements that combine features of transposons, plasmids, and even bacteriophages. Excision of these MGEs can be induced by certain conditions, after which ICEs circularize, replicate, are passed to a recipient cell via conjugation, and ultimately integrate back into each cell's respective chromosome (119). The ubiquity and diversity of MGEs emphasize the heightened mobility of antibiotic resistance within and across microbial communities, which highly complicates the mitigation of AMR across environments.

### **The ecology of antimicrobial resistance**

The mobility of ARGs is not confined within discrete environmental systems. Rather, AMR is highly prevalent in a variety of environments including soils, organic materials, aquatic systems, industrialized areas, wildlife, and the human gut (123). The antibiotic “resistome” is a collection of all ARGs and their genetic precursors in a community (124). While we may consider separate resistomes for different environments (such as human vs. soil vs. animal), these respective resistomes can shape each other (123). Forsberg et al. (2012) (125) identified multiple

genomic regions in environmental soil isolates that had >99% similarity to five relevant human pathogens. Importantly, each region contained one or more ARGs in addition to various MGEs, a finding that highlights the shared resistome across environments (125). Similarly, another study found that resistomes across habitats were primarily structured by bacterial phylogeny along an ecological gradient (126). However, an association between the proportion of antibiotic resistance contiguous genomic sequences (contigs) containing a MGE or multidrug resistance cluster (MDRC) and the number of habitats in which the ARG was found was observed, highlighting the importance of MGEs in facilitating transfer across environmental boundaries (126).

Many studies have explored the resistomes of various environments to better understand ARG composition, affiliation with members of the microbiota, and potential for transmission across taxa and habitats (127). Most work on resistomes is separated into three overarching categories: environmental, animal, and human. Environmental resistomes include both natural (which can be subdivided into terrestrial and aquatic ecosystems) and built (such as agriculture, wastewater treatment, aquaculture, and hospitals), whereas animal and human resistomes primarily consider ARGs harbored by these reservoir hosts (127).

Freshwater ecosystems such as rivers, were previously found to be an important reservoir and transmitter of ARGs. Multiple studies have identified the presence of ARGs in river systems and, importantly, have noted marked increases in ARG abundances within rivers under greater anthropogenic influence (e.g., urban rivers, greater contact with humans, industrialized areas, etc.) (128, 129). More specifically, Lee et al. (2020) (130) showed that anthropogenic fecal contamination resulted in a bloom of ARGs in the Han River in South Korea. This bloom was not attributed to the proliferation of drug resistant microbes, but rather an increased prevalence

of environmental microbes which had integrated ARGs due to high prevalence of MGEs in the contaminating matter (130). Notably, many ARGs found in freshwater systems subsist during water treatment and can be identified in drinking water. For instance, Ma et al. (2017) (131) identified ARGs conferring bacitracin, MDR, aminoglycoside, sulfonamide, and beta-lactam resistance. This study also characterized the microbial hosts harboring ARGs of interest. Strikingly, 80% of contigs carrying an ARG in *Pseudomonas* spp. were important for MDR, demonstrating the ubiquity of MDR even within our drinking water (131).

Another relevant system that serves as a potent resistance reservoir is agriculture, including husbandry of livestock. One reason for the high frequencies of AMR in agricultural settings is due to the overuse of antibiotics in animal food production (132). Although exact amounts can vary by country, the average annual consumption of antibiotics per kilogram of animal produced was estimated to be highest in swine (172mg/kg) followed by cattle (148mg/kg) and chickens (45mg/kg) (133). Indeed, these antimicrobial additives were found to shift the gut microbiome of swine and significantly enhance the diversity of ARGs and MGEs identified (134). Application of metals such as copper or zinc, is also common in the livestock industry to promote growth and prevent disease (132). Notably, use of these compounds not only results in increased metal resistance among animals, but also enhanced AMR, as these two types of resistance have been found to co-occur (135). Increased levels of resistance in agricultural settings pose a risk for widespread transmission to humans directly through animal contact or indirectly through food consumption (99). In fact, Sun et al. (2020) (136) demonstrated that veterinary students working on swine farms experienced a shift in their gut microbiomes and resistomes. After three months of exposure, these individuals' microbiomes and resistomes became more similar to those of resident farm workers and even environmental samples. These

data demonstrate that the agricultural industry has contributed to the persistence of resistance in different hosts and environments, emphasizing our need to address ARG spread.

Similarly, the human gut is also an important reservoir for ARGs, a system that also has implications for human health (127, 137). In a prior study, Feng et al. (2018) (138) characterized the gut resistome for 180 healthy individuals from 11 different countries. They found ARGs for aminoglycosides, bacitracin, MLS, MDR, tetracyclines, and vancomycin to be ubiquitous across all samples and subsequent network analyses identified 12 bacterial species to serve as hosts for 58 distinct ARG subtypes (based on co-occurrence and correlation) (138). Furthermore, resistome composition appears to be dependent on geography since individuals residing in different countries registered distinct resistomes (138, 139), an outcome that was suggested to be linked to differences in antibiotic applications and usage (140).

The resistome has also been explored in individuals whose microbiomes are in a state of flux. For example, antibiotic treatment drastically impacted the composition and diversity of the human gut microbiome in one study, which was linked to ARG composition (90). Specifically, an increase in members of *Proteobacteria* and associated ARGs was documented, a result that was replicated in this dissertation (141). Interestingly, Raymond et al. (2019) (142) performed culture-enrichment to explore associations between ARGs and lower abundance taxa in the human gut, specifically members of *Enterobacteriaceae*. It was found that *E. coli* had a large accessory genome composed of proximal ARGs and MGEs (142). These findings corroborate those described herein and of those presented by Perez-Cobas et al. (2013), as increases in the abundance of *Escherichia* would inherently increase the abundance of these associated accessory ARGs. Importantly, previous functional characterization of ARGs in the human gut resistome revealed that a large proportion of these genes were identical to those harbored by human

pathogens (143). It is important to note, however, that many ARGs identified with these methods had not yet been characterized, suggesting even greater diversity among human and pathogen resistomes than previously thought (143). The role of the human gut resistome in perpetuating the spread of AMR must be strongly considered when exploring methods of resistance mitigation.

### **Antibiotic resistance among enteric pathogens**

Of great importance to this research is the high prevalence of antibiotic resistance among enteric pathogens (144, 145) and the potential for transfer to commensals or opportunistic pathogens during infection. Each of the four enteric pathogens examined in this study (*Campylobacter*, *Salmonella*, *Shigella*, and *E. coli*) have been classified as serious antibiotic resistant threats (97). Antibiotic-resistant *Campylobacter* spp., for instance, were estimated to cause nearly 450,000 infections per year in the United States; this number accounts for ~30% of the total number of *Campylobacter* infections annually (97). Increasing levels of resistance to azithromycin, a macrolide often used to treat *Campylobacter* infections, and ciprofloxacin, a quinolone, have been reported in *Campylobacter* spp. (97, 146). Resistance to ciprofloxacin is more common, with about 28% of all *Campylobacter* isolates demonstrating reduced susceptibility to this drug in the last twenty years (97). Nonetheless, considerable variation in resistance frequencies has been reported across geographic locations (147). The mechanism of resistance to ciprofloxacin and other fluoroquinolones is due to point mutations in the quinolone-resistance-determining regions of *gyrA* encoding the DNA gyrase (116, 148). *Campylobacter* is also naturally competent and can take up exogenous DNA from the surrounding environment, a feature that enhances its ability to acquire ARG-containing genetic material and subsequently survive in the presence of antibiotics (116, 149).

Nontyphoidal *Salmonella* (NTS) causes approximately 1.35 million infections each year, with ~15% (212,500) of these being caused by drug-resistant *Salmonella* (97). Since most *Salmonella* infections are self-limiting, it was suggested to avoid antibiotics for treating NTS infections, as treatment can result in prolonged shedding of drug-resistant pathogens (150). In fact, inappropriate application of antibiotics for NTS infection is a key driver of increased resistance, a development that has led to more adverse clinical outcomes (151) and an enhanced risk of hospitalization (152). Notably, nearly 10% of all *Salmonella* were shown to have resistance to ciprofloxacin, with documented increases in resistance to ceftriaxone (a third-generation cephalosporin in the beta-lactam class) and azithromycin as well (97). Just as *Salmonella* is known for its pathogenicity islands, this pathogen also contains genomic islands capable of conferring antibiotic resistance. Specifically, *Salmonella* genomic island 1 (SGI1) contains a region called ACSSuT that confers resistance to ampicillin, chloramphenicol, streptomycin, sulfamethoxazole, and tetracycline (153). Similarly, another genomic island (*Salmonella* streptomycin and azithromycin resistance island (SASARI)), which contains ARGs for resistance to these antibiotics, was discovered (154). The identification of these genomic islands harboring ARGs for multiple antibiotic classes emphasizes the growing importance of MDR among pathogenic *Salmonella* isolates.

Although *Shigella* spp. are responsible for fewer infections overall (n=450,000) in the United States, a high percentage of these infections are caused by antibiotic-resistant strains (77,000; 17%) (97). According to the CDC, ~25% of all *Shigella* isolates recovered from infected cases displayed reduced susceptibility to azithromycin, and at least 10% have demonstrated resistance to ciprofloxacin (97, 146). HGT is important for *Shigella*, which has been shown to acquire resistance to multiple antibiotic classes via the acquisition of plasmid-

borne or mobilized ARGs (155). Similar to NTS, *Shigella* can harbor the ACSSuT genomic region conferring MDR, an important consideration when attempting to treat *Shigella* infection (156).

Finally, resistance in STEC has been increasingly documented even though antibiotics are not recommended for treatment (157, 158). In 2019, the incidence of STEC was 6.3 per 100,000 people, the third highest ranking for enteric pathogens behind *Campylobacter* and *Salmonella* (48). At the family level, members of *Enterobacteriaceae*, which include genera such as *Escherichia*, *Salmonella*, *Klebsiella*, and *Shigella*, often harbor clinically-relevant ARGs with phenotypic resistance. Indeed, the CDC has classified carbapenem-resistant *Enterobacteriaceae* as an urgent threat, while extended-spectrum beta-lactamase (ESBL) producing *Enterobacteriaceae* represent a serious threat to public health (97). Of great importance is the ability of different members of *Enterobacteriaceae* to share genetic material across genera, particularly on plasmids, via HGT (159). This realization becomes more important as we continue to identify the breadth of ARGs that can be transmitted horizontally; for example, ESBLs are highly mobilized across *Enterobacteriaceae*, posing a great threat to the efficacy of this antibiotic class in the clinic (160).

Indeed, the increasing prevalence of drug resistance among enteric pathogens coupled with rising mobility of these ARGs throughout microbial communities is great cause for concern. In addition to acquiring resistance outside of the host, some enteric pathogens can acquire resistance from the reservoir of ARGs in human gut communities. The exchange of ARGs between pathogens and commensal bacteria via HGT was previously documented in the gut (116). Therefore, not only are the ARGs harbored by these enteric pathogens of great concern,



but the community-level impacts of invasion by drug-resistance pathogens appear to be widespread.

## **ROLES OF THE GUT MICROBIOME AND RELATED METABOLIC CAPACITY IN SHAPING HUMAN HEALTH**

### **Interplay between the gut microbiome and metabolism**

Impacts of disease, both chronic and acute, on the human gut microbiome and metabolome have increasingly been characterized, with particular emphasis on the use of multi-omics to better define host-microbiota metabolic interactions (161). Numerous studies have explored microbial differences between ‘healthy’ individuals and those with various health conditions including obesity, liver disease, metabolic disease, diabetes, and inflammatory bowel disease (IBD) (162-164). However, a standard definition for a “healthy gut” is flexible, as one’s microbial and metabolic profile can vary due to day-to-day diet, exercise, and even genetics (165). Diet is especially important in driving the composition of the gut. Although, diet-related shifts in the composition of gut microbiota are personalized, and even sustained intake of a monotonous diet does not result in a “standard” microbial profile (166). Despite this, another study of humanized mice documented that prolonged consumption of low-fiber diets resulted in the removal and eventual extinction of some beneficial microbes in the gut, primarily those in the order Bacteroidales (167). Altered nutrient load, a metric dictated by dietary intake, was found to influence changes in microbial composition, a consequence that also influenced the gut’s ability to absorb nutrients (168). Colonic transit time, a measure of how frequently the colon is cleared via defecation, is a potential indicator of gut microbiome health as well since transit times correlate with various forms of metabolism. Indeed, shorter transit times were linked to

carbohydrate fermentation, whereas longer times were associated with protein catabolism, a less preferred energy source (169).

Diet-related shifts in gut microbial composition also inherently impact metabolism since community members have a direct effect on host metabolic health (170). In a systematic review, Wolters et al. (2019) (171) explored the impact of dietary fat intake on gut metabolic health and found that diets rich in fats and monounsaturated and saturated fatty acids resulted in reduced bacterial richness and diversity, while diets high in polyunsaturated fatty acids did not exhibit this negative effect. Relatedly, Le Chatelier et al. (2013) (172) demonstrated that individuals with reduced gut richness and diversity are distinguished by greater adiposity, insulin resistance, and inflammation. Moreover, Cani et al. (2007) (173) showed that ingestion of long-term high-fat diets selected for lipopolysaccharide (LPS)-producing bacteria, an effect contributing to endotoxemia-related inflammation that has been correlated with obesity and insulin resistance. Interestingly, intraspecies variation can have differential effects on host metabolism; indeed, genetic variation exhibited by these ‘structural variants’ results in differences of metabolic capacity of the microbiome (174).

### **The importance of short-chain fatty acids (SCFAs) in gut health**

Many members of the commensal gut are responsible for breaking down complex carbohydrates after ingestion (175) or generating vitamins that are critical to human health (176). Complex carbohydrate degradation by gut microbes leads to generation of beneficial SCFAs, compounds composed of a carboxylic acid and small hydrocarbon chains, which influence the host’s inflammatory response and autoimmunity (177). In addition to diet, another factor found to influence SCFA production is place of residence; Jacobson et al. (2021) (178) found that individuals residing in non-industrial areas had greater SCFA-related resilience, whereas those

from industrial areas were at greater risk of disruption due to limited SCFA production by just a few genera. In particular, three SCFAs, acetate, butyrate, and propionate, which are synthesized by gut microbes, have been extensively studied and linked to gut health. Acetate is often the most abundant of these SCFAs and is primarily produced by members of the Firmicutes phylum (179); radioisotope analysis performed by Miller and Wolin (1996) (180) revealed that acetate production occurs most often via the Wood-Ljungdahl pathway (also called the reductive Acetyl-CoA pathway) following glucose catabolism and conversion of acetyl-CoA to acetate. Butyrate is synthesized in the human colon by anaerobic Gram-positive bacteria, with the most prominent butyrate producers being *Faecalibacterium prausnitzii*, *Eubacterium rectale*, and *Roseburia* spp (181). Acetate has been shown to serve as a precursor for butyrate production in the gut (182), but butyrate is primarily produced via conversion of acetyl-CoA, glutarate, 4-aminobutarate, or lysine, with the acetyl-CoA pathway being most prevalent among butyrate producers (183). Propionate, on the other hand, is produced by multiple phylogenetic groups including Bacteroidetes, Verrucomicrobia, and Negativicutes via the succinate pathway, and Lachnospiraceae and Ruminococcaceae, which use the acrylate or propanediol pathways (184).

These SCFAs continue to be explored for their specific impacts on human gut health. Because SCFAs are taken up from the lumen by intestinal epithelial cells (IECs), they are metabolized as a source of energy for these colonocytes while playing a vital role in fatty acid metabolism, glucose metabolism, and cholesterol metabolism (185). Once absorbed, these SCFAs also improve aspects of immune defense and development, contributing to homeostasis of the intestinal epithelium (186). It has been shown, for example, that SCFAs produced via fermentation of dietary fiber bind to the G-protein coupled receptors, GPR43 and GPR109A, to prime and activate the NLRP3 inflammasome in colonocytes. This activity facilitates recovery

from inflammation, overall gut homeostasis, and further protection from disease (187, 188). Butyrate, in particular, is the SCFA most frequently metabolized by IECs and directly influences the growth and apoptosis of healthy colonocytes while also demonstrating an ability to inhibit unregulated cancerous growth (189). Butyrate, in addition to acetate and propionate, has been documented to increase various defensive functions of the intestinal epithelium, including production of  $\beta$ -defensins and cathelicidins, which are host-secreted peptides that aid in antimicrobial defense (190).

These SCFAs, along with the microbes that produce them, have also been implicated for their role in combating various metabolism-related diseases. For example, an increased Firmicutes:Bacteroidetes ratio as well as lower overall bacterial diversity has been associated with obesity (191). Nogal et al. (2021) (192) demonstrated that gut microbiome alpha diversity positively correlated with higher circulation of acetate, a SCFA that was negatively associated with the quantity of visceral fat. Acetate has also been explored for its therapeutic potential in counteracting obesity with promising results (193), while SCFAs have been shown to counteract decreased metabolic functions associated with Type 2 diabetes. Uptake of these molecules indirectly leads to secretion of GLP-1, a hormone that aids in maintaining glucose homeostasis, and promoting  $\beta$ -cell development that benefits insulin production (194). These beneficial outcomes, combined with the aforementioned management of intestinal inflammatory pathways, aid in combatting effects of Type 2 diabetes.

Levels of SCFA production have also been linked to inflammatory bowel diseases (IBD) such as ulcerative colitis and Crohn's disease; individuals experiencing IBD have shown reduced concentrations of SCFAs and SCFA-producing taxa in the gut (195). Relatedly, IBD patients contained different compositions of bile acids in their guts compared to healthy individuals

(196). These conditions likely influence how the host manages intestinal inflammation and may even exacerbate gut epithelial inflammation and decreased metabolic health. In another study, patients with colitis had reduced representation of the phylum Firmicutes, specifically the species *Faecalibacterium prausnitzii* (197), which is known for its role in SCFA generation in the gut (198). Together, these studies suggest that adequate levels of SCFAs can mediate metabolic conditions responsible for multiple conditions including obesity, Type 2 diabetes, and IBD. While SCFAs contribute greatly to the overall management of gut metabolic health, little is known about their role in or recovery from enteric infections.

### **Other microbially-mediated metabolites of importance**

In addition to SCFAs, the gut microbiota synthesize and modulate other relevant metabolites like vitamins. Indeed, members of gut microbial communities were demonstrated to construct vitamin K and various B group vitamins, which have been implicated in human health (199). Magnusdottir et al. (2015) (200), for instance, explored 256 common gut bacteria for the presence of B-vitamin biosynthesis pathways. They found the distribution of these synthesis pathways to be diverse but noted evident exchange of vitamins among organisms whose vitamin pathways were complementary (200). Similarly, Soto-Martin et al. (2020) (201) used *in silico* methods supplemented by *in vitro* testing to characterize vitamin requirements among butyrate-producing auxotrophs and prototrophs in the gut. They found evidence of cross-feeding among these groups, specifically for vitamins B<sub>1</sub> and B<sub>9</sub> (thiamine and folate, respectively), suggesting cooperation among beneficial commensal gut bacteria (201).

Another group of compounds important for intestinal health is bile acids. While microbiota in the gut can modify the structure and chemical properties of bile acids, the latter also exert antimicrobial properties that can shape gut communities (199, 202). Bile acids are

normal gut metabolites that aid in the digestion and absorption of lipids; these molecules also promote homeostasis of the intestinal epithelium (203). In the liver, primary bile acids undergo conjugation, typically with glycine or taurine, and are subsequently stored in the gall bladder before release into the small intestine (203). While many conjugated bile acids are destined for enterohepatic circulation and are reabsorbed, a small fraction are metabolized by the gut microbiota (204). Function-based metagenomics was used to show how bacteria employ a widely conserved bile salt hydrolase (BSH) to catalyze deconjugation of many conjugated bile acids (CBAs) and liberate primary bile acids into the gut lumen (205). These primary bile acids can be used by other microbes in the gut to generate secondary or tertiary bile acids. The most common transformation among gut bacteria is the conversion of the primary acids, chenodeoxycholic acid (CDA) and cholic acid (CA), to the secondary acids, lithocholic acid (LCA) and deoxycholic acid (DA), respectively, via 7-dehydroxylation (206). Importantly, certain secondary bile acids such as DA can generate reactive oxygen species (ROS) and reactive nitrogen species (RNS), which can lead to DNA damage (207). These secondary bile acids have therefore been implicated for a role in colonic and liver cancers. Interestingly, these bile acids also have potent antimicrobial properties; for example, DCA is hydrophobic and acts as a strong detergent that can disrupt bacterial membranes, highlighting the importance of feedback between bile acids and gut microbes (208). In fact, the gut microbiota not only plays a role in generating secondary bile acids, but also influences bile acid pool size by regulating synthesis pathways in the liver (209). Certainly, maintaining a healthy balance between gut microbiota activity and bile acid production and cycling is crucial to human gut health.

## **COMPUTATIONAL METHODS FOR STUDYING THE HUMAN GUT MICROBIOME, RESISTOME, AND METABOLOME**

As discussed in previous sections, there are myriad aspects of the human gut microbiome that continue to be explored and studied. To glean information from this host-associated ecosystem, multiple tools have been developed to appropriately capture, condense, and characterize the complexity in the gut (210). Despite the rapid advancements in the fields of bioinformatics and computational biology, many tools in circulation can only detect associations between various microbiome features and not establish causation (210). Although these methods provide incredible insight into the genomic and metabolic landscape of the human gut microbiome, these computational methods require complementation with *in vitro* and *in vivo* studies to better define microbial interactions in the gut. Nevertheless, bioinformatic and computational analyses are invaluable resources for this crucial stage of microbiome research. Indeed, results from these computationally-based studies supply hypothesis-generating data from which more targeted and nuanced cause-and-effect studies can be designed.

### **16S rRNA and ITS sequencing**

One of the most ubiquitous forms of sequencing analysis involves the 16S rRNA gene in bacteria and the internal transcribed spacer (ITS) region in fungi. These marker genes are highly conserved across microbial lineages, though taxa-specific divergence can promote differentiation (211). Importantly, many studies sequence only a specific portion of the gene, such as specific variable regions like V3-V4 in the 16S rRNA gene, because widely used sequencing platforms (e.g., Illumina) can only accommodate short sequences of  $\leq 300$  base pairs (212). Although targeting these sub-regions has allowed confident taxonomic identification at the genus-level or above, restricting one's view to these regions cannot sufficiently discriminate between more

closely related taxa. Additionally, the 16S rRNA and ITS genes each contain hypervariable regions that can change over a relatively short time period (210); the variation in these regions can be detected in closely related taxa or even within a single genome, an aspect that greatly complicates sequence-based diversity analyses (213). Because the entire 16S rRNA gene is approximately 1,500 bp long, sequencing the entire gene (via long read technologies) provides much greater resolution at the species- and strain-level (212, 214). However, long-read sequencers currently display higher error rates and comprehensive databases for these full-length marker genes are limited (210). Therefore, the most common marker gene analyses still rely on short-read, sub-region sequencing.

Due to the evident variation within and among marker genes across taxa, many investigators have adopted use of operational taxonomic units (OTUs). This method clusters bins of sequences based on pairwise alignment sequence dissimilarity (PSD), with a common similarity threshold of 97% (210, 215). A representative sequence is then selected for each OTU cluster. Although this method's ability to simplify a complex sequencing dataset increases applicability, choosing a single sequence to represent an entire OTU cluster was suggested to be misleading, especially if OTUs are poorly clustered (215). After identifying a representative sequence, the sequence is annotated (usually with taxonomic information) using a naïve Bayesian classifier like the Ribosomal Database Project (RDP) Classifier (216) or extrapolating taxonomy via mapping to reference databases like Greengenes (217), SILVA (218), or UNITE (219). Any other sequences within the OTU cluster will inherit the annotations attributed to the representative sequence. Multiple analytical pipelines have been developed for 16S rRNA gene analysis including QIIME (220, 221), Mothur (222), and DADA2 (223), each of which provides information on taxonomic assignment and enables users to perform relevant diversity and



statistical analyses. Indeed, 16S rRNA and ITS sequencing analysis are powerful tools to explore microbiome composition and diversity. Despite some of their limitations, these methods have been optimized and are easy to use, an advantage that not all computational approaches yet share.

### **Shotgun metagenomics sequencing**

Despite their utility, marker gene analyses are restricted to a single aspect of the microbial genome, while shotgun metagenomics methods perform untargeted sequencing of all microbial genomes per sample (224). This method can be used to determine the taxonomic composition of a microbial community as well as its functional potential (210, 224). Prior to analysis, however, metagenome sequences must first undergo computational pre-processing in which adapters, poor quality sequences, and non-target DNA sequences are removed (225-227). There are two main analysis methods used to explore metagenomes: assembly-based profiling and read-based profiling. Each of these methods has advantages and pitfalls; hence, it is often recommended that a combination of these two approaches be used when investigating the microbiome (224).

Assembly-based profiling first requires construction of contiguous sequences or “contigs” from the raw sequencing reads. There are multiple algorithms that have been developed to construct these contigs, but the de Bruijn graph approach is currently most popular (224). A de Bruijn graph is a directed graph that shows overlaps between sequences. This assembly method breaks sequencing reads into overlapping sequences of a fixed length ( $k$ ); these overlapping  $k$ -mers are then clarified to produce a directional graph with a single path that accurately reconstructs genomic regions (228). Some highly used assemblers that employ the de Bruijn graph method include IDBA-UD (229), MEGAHIT (230), metaSPAdes (231),

MetaVelvet (232), and Ray Meta (233). Multiple groups have sought to benchmark these different assembly methods to improve investigators' ability to choose the best method to address their research questions (234-236).

One notable advantage of assembly, if of sufficient sequence quality, is to have the ability to explore the genomic architecture of microbes in the community. For example, contig construction can assemble multiple short reads into a coherent section of the genome, allowing greater confidence when analyzing genomic regions that may exceed the length of short reads alone (237). Additionally, this insight into genomic architecture also has the power to associate functional genes with taxonomy, as genes of interest will likely co-occur on contigs that can be taxonomically annotated (237). Despite these advantages, there are notable limitations of metagenomic assembly. For example, the quality of assembly strongly depends on the number of sequences obtained, the quality of sequencing and library preparation, and the diversity of the community being sampled (238), although these aspects are also limitations relevant to read-based analysis. Another potential limitation is the relatively high computational expense of assembly, as most assemblers require significant memory and computing power (234). It is also notable that different assembly tools can differ markedly in their usability, assembly method, and use of computational resources, aspects that can drastically influence assembly output and downstream interpretations (234). Assembly also reduces the complexity of the sampled microbial community since only a fraction of microbial genomes can typically be resolved; unfortunately, the genomes first excluded are usually those representing low-abundance taxa (224). Although metagenomic assembly is a powerful and evolving approach, researchers must consider which assembly methods are best suited for their system of interest and their research questions before performing this step in the analytical pipeline.

By contrast, read-based analyses use unassembled metagenomic sequencing reads to obtain taxonomic or functional profiles following alignment to publicly available reference databases (224, 237). This process determines the abundance of various taxa and/or functional genes based on sequence similarity between raw reads and reference genes in the chosen database. One advantage is that it can capture more complex communities if sequencing depth is adequate and the reference database is sufficiently large and well curated. Unlike assembly-based methods, which may not be capable of resolving low-abundance reads, the read-based pipelines consider all sequencing data provided (224). Moreover, the use of reads is often less computationally taxing than performing assemblies, which require binning, read-mapping, and the generation of metagenome-assembled genomes (MAGs). Some limitations associated with a read-based analysis, however, are important to note. Although this method can provide insight into community structure and function, it can only capture information relevant to reads that successfully map to the reference database, limiting these analyses to known taxa or genes. Despite this, ongoing improvements are being made to reference databases as more and more microbes are cultivated and sequenced (239, 240). In contrast to read-based assessments, assembly-based methods can construct genomes of entirely novel organisms (224). Another limitation associated with read-based analysis is based on the truncated nature of short-read data (~250-300bp fragments), which can complicate annotation, particularly when attempting to elucidate function (237). Previously, short-read sequencing has failed to properly detect relatively distant homologs of bacterial and viral genes, especially in a plankton community that was presumably less characterized, though this study is now somewhat dated (241). More recently, Treiber et al. (2020) (242) sought to determine the influence of various parameters on proper functional annotation of human fecal microbiomes. They found that read length, E-value

threshold, and choice of protein database strongly influenced the outcome of functional annotation and recommended using short reads between 180-250 bp for more accurate annotation. Importantly, however, another study found conflicting evidence that precision and recall of functional annotation tools can decrease for longer (>200 bp) short reads (243). The inconsistencies across these studies may be attributed to choice of annotation tool or other pre-processing steps. In any case, this incongruence emphasizes the importance of researchers choosing analysis tools that most appropriately align with their study system.

In both assembly-based and read-based methods, the subsequent step(s) in a metagenomic pipeline is taxonomic or functional annotation. To date, many computational tools have been developed to estimate taxonomic or functional gene abundance. However, each of these tools present subtle nuances in how they assign annotations to sequencing data. Hence, it is critically important that investigators choose the method most appropriate for their study system (244). Taxonomic classifiers use one of three methods to assign taxonomy to sequencing data: DNA-based methods, translated protein methods, or marker gene methods. DNA-based methods assign taxonomy based on nucleotide sequence alignment and include tools such as Kraken (245), CLARK (246), and Centrifuge (247), among others. Protein-based methods have also been developed to address various shortcomings associated with DNA-based classification, such as increased computational burdens and reduced sensitivity for sequences that are evolutionary related but divergent (a situation in which the amino acid sequence may garner an assignment but a nucleotide sequence would not) (248). Examples of protein-based classifiers include Kaiju (248) and DIAMOND (249). Finally, marker-based taxonomic classification uses clade-specific marker genes to accurately assign taxonomy (250); a popular tool that employs this method is MetaPhlAn (250-252). As these tools continue to develop, multiple groups have sought to

benchmark methods, giving investigators some guidance regarding relative advantages and disadvantages of each (244, 253). However, direct comparison among different classification methods can be challenging, as these techniques often employ different considerations of abundance; some tools report relative sequence abundance while others report relative taxonomic abundance, an important distinction when integrating these numbers into downstream diversity and statistical analyses (254).

Unlike taxonomic classification, functional annotation of metagenomic sequences first requires gene prediction (255). Predicted genes can then be searched against functionally characterized protein families to glean information related to metabolism, antibiotic resistance, or virulence (224). Popular protein family databases for functional annotation of metagenomes include the Kyoto Encyclopedia of Genes and Genomes (KEGG) (256), the Clusters of Orthologous Genes (COG) database (257), and UniProt (258). Additionally, pipelines such as HUMAnN (252, 259) have been developed to generate information on predicted metabolic pathways including pathway presence/absence, abundance, and taxonomic association. Antibiotic resistance capacity can also be explored by aligning metagenomes to resistance gene databases such as the Comprehensive Antibiotic Resistance Database (CARD) (260), MEGARes (261, 262), and Resfams (263), among others. Virulence capacity among members of microbial communities is also of great interest, and can be annotated via alignment to virulence gene databases such as the Virulence Factor Database (VFDB) (264).

Shotgun metagenomics sequencing is an incredibly valuable tool that enables researchers to evaluate the entire genomic landscape of a microbial community. Computational tools continue to evolve rapidly in this sector of microbiology, and both taxonomic and functional annotation will continue to improve as developers refine their software. Notably, metagenomics

is a relatively young aspect of this field, and a key limitation is the current lack of a standardized pipeline. The sheer number of computational tools at researchers' disposal can be overwhelming, especially for investigators who are new to the field. As previously discussed, choice of tool can greatly influence downstream interpretation, and therefore, decisions regarding correct tool choice can be quite daunting. Nevertheless, if approached with the correct study design, knowledge of computational tools, and insight on various parameters, shotgun metagenomics greatly augments our ability to explore the intricacies of microbial communities.

### **Metabolomics and other 'omics technologies**

In addition to metagenomics, there are other 'omics technologies that can provide information on different aspects of the gut microbiome. These include metabolomics, metatranscriptomics, and metaproteomics. Metabolomics seeks to quantify small metabolic compounds derived from both microbes and the human host (265). Most often, these metabolites are identified and quantified using processes such as gas chromatography-mass spectrometry (GC-MS) or liquid chromatography-mass spectrometry (LC-MS) (210, 266, 267).

Chromatography methods are used to separate metabolites within a given sample prior to passing them through a mass spectrometer, a tool which characterizes the specific mass-to-charge ratio ( $m/z$ ) of each compound (266). MS spectra produced via GC- or LC-MS can then be used for feature detection and aligned to known compounds to extrapolate chemical identity and structure. Intensities associated with these spectra can also be normalized and used for downstream statistical analysis (266). Multiple pipelines have been developed to aid in processing metabolomics data. One incredibly useful tool is Global Natural Product Social Molecular Networking (GNPS), an online resource which facilitates small molecule-focused tandem mass spectrometry ( $MS^2$ ) data curation and analysis (268). Not only does GNPS capture

the molecular diversity present in metabolomics samples, it also streamlines compound annotation via alignment to known spectral libraries. Additionally, a valuable feature of GNPS is its implementation of molecular networking, a method which can identify similarities among found metabolites and extrapolate meaning for unknown metabolites through their relationship to known compounds (268, 269). Notably, metabolomics has previously been used to explore links between human gut microbial composition and metabolic output and health (267, 270) while enabling more comprehensive characterization of the human gut environment.

Comparatively, metatranscriptomics characterizes RNA that has been transcribed in microbial cells, providing insight into which genes are being expressed in given conditions (210, 271). RNA sequencing reads can be mapped to microbial genomes or metabolic pathways to deduce taxonomy and function, respectively (210). For example, Jorth et al. (2014) (272) used metatranscriptomics to assess which genes were up- or down-regulated in the oral microbiome during periods of health and periodontal disease. Metaproteomics can nicely complement both metatranscriptomics and metagenomics, as this method allows for an investigation of the protein composition of a sample (273). Importantly, this approach can detect relevant post-translational modifications of various proteins and enzymes in the gut space, which can be incredibly useful when elucidating function (273).

In isolation, each of these ‘omics techniques can provide meaningful insight into the many workings of the gut environment from different perspectives. However, more recently, greater emphasis has been put upon integrating the analyses of multiple ‘omics technologies to more comprehensively assess changes within communities (274). Integration of these ‘omics datasets can occur after performing separate analyses (post-analysis data integration) or the data analysis itself can include all ‘omics outputs from the beginning (integrated data analysis) (275).

Despite the evident benefits that these integrated assessments can offer, use of multiple ‘omics datasets at one time comes with many challenges. Examples include high levels of variability among datasets, bias or lack of operability between analytical tools, limitations associated with annotation and visualization resources, and the potential for failing to justify the necessity of integrative techniques (275). In addition to these challenges, integrative ‘omics approaches can be quite daunting, as working with multiple sets of “Big Data” and having the knowledge needed to perform analyses properly takes time and practice. Advantageously, multiple groups have sought to recapitulate computational approaches, useful tools, common setbacks, and important considerations associated with multi-omics analyses (274-276). The ongoing curation of these ‘omics technologies is exciting, as continued characterization of complex microbial systems will become more approachable, manageable, and interpretable as advancement progresses.

### **Downstream interpretation and statistical analysis**

Upon retrieving taxonomic or functional gene abundances, researchers can employ a host of analysis methods to further investigate and clarify patterns within sampled microbial communities. Studies that compare community composition between two or more groups will often investigate microbial diversity. Microbiome differences can be evaluated using both alpha and beta diversity metrics. Alpha diversity quantifies the level of within-sample diversity using metrics such as species richness (i.e., how many species are present in each sample), evenness (i.e., of the species present, how equally represented are they), and indices that measure both richness and evenness such as the Shannon or Simpson Index (210, 277). Statistical tests used to compare alpha diversity often include nonparametric univariate statistical methods such as the Wilcoxon rank-sum test for nonpaired samples (also called the Mann-Whitney  $U$  test) or the Wilcoxon signed-rank test for paired samples (278). Because microbiome analysis often involves



multiple comparisons among samples due to the inherent high dimensionality of the data, methods to adjust  $P$  values are needed. The most common  $p$ -value adjustment methods include Bonferroni adjustment (to control family-wise error rate) or Benjamini-Hochberg (which controls the false discovery rate (FDR) and is less conservative than Bonferroni) (278). Beta diversity, on the other hand, compares between-sample diversity among groups; this is achieved by investigating dissimilarity between samples by generating a pairwise distance matrix in which samples with comparable compositions will be less dissimilar (210). Common metrics used for beta diversity analyses include Bray-Curtis dissimilarity, which accounts for taxa abundance among samples (277), and weighted and unweighted unique fraction (UniFrac) distances that consider phylogenetic relatedness in addition to taxonomic abundance (279). Beta diversity can be analyzed using multivariate statistics methods such as permutational analysis of variance (PERMANOVA), which is one of the most widely used multivariate methods. This method assesses differences in centroids (center point of a cluster of samples within a group) and dispersion (level of spread within a group) between samples groups in a study (278, 280). The Mantel test is similar to PERMANOVA in that it evaluates differences using permutations; however, this test is most often used to investigate associations between metadata or environmental factors and microbiome abundance data (278, 280). In addition to performing statistical analyses, beta diversity can also be visualized. Due to the high dimensionality and complexity of microbiome data, however, investigators often use dimension-reduction-based ordination methods for visualization (210). Common examples used in microbiome analyses are principal component analysis (PCA), principal coordinate analysis (PCoA), correspondence analysis (CA), and nonmetric multidimensional scaling (NMDS) (281, 282). Each of these

methods have their own respective assumptions about the data, and therefore investigators must consider the appropriateness of these techniques for their study system.

In addition to diversity analyses and their related statistics, researchers often seek to characterize discrete differences in taxonomic or functional composition between study groups. One popular method is a differential abundance (DA) analysis, which identifies taxa that are more highly represented in one group versus another (277). Many DA algorithms use linear models with the relative abundance of taxa or functional genes (i.e., proportion of a sample attributed to a particular feature) to assess these differences (277, 283). Importantly, however, DA analysis methods differ in their algorithms, normalization methods, transformation methods, inclusion of covariates and random effects, as well as base hypothesis test (283). In fact, a handful of studies have assessed differences in the results output by various DA methods available. Each of these studies found high levels of discordance among these tools, with just a couple of methods demonstrating consistency across various datasets (283, 284). Limitations associated with these methods may be related to difficulties with the incorporation of zero-inflated data; microbiome abundance data is inherently zero-inflated as many features (taxa or functional genes) appear in low or inconsistent abundance (277). These findings suggest that multiple DA methods should be used when assessing taxa or genes of interest and that greater work must be done to improve the accuracy of DA analyses for microbiome studies.

Another popular approach to characterize the microbiome is via network analysis, which allows investigators to visualize interactions as they pertain to taxa, relevant genes, metabolites, environmental factors, and more (210, 282). Networks are often constructed using pairwise similarity to deduce correlations between microbial features. These networks can then be visualized using software such as Gephi (285) or Cytoscape (286), which allow users to highlight

or filter various aspects of their networks and obtain relevant network statistics. Importantly, correlation networks used to assess co-occurrence of various microbial features can be used to generate predictions that may be tested with more robust computational or laboratory methods. For example, Ma et al. (2017) (131) performed network analysis to investigate co-occurrence patterns between various taxa and ARG subtypes. Next, this group assessed actual associations between key taxa and ARGs by isolating ARG-carrying contigs (ACCs) and annotating these specific contigs with taxonomic information (131). Indeed, network analysis can be a powerful tool when attempting to interpret the myriad associations and patterns present in microbiome data.

## **SUMMARY**

In summary, the human gut is an environment replete with host-microbe and microbe-microbe interactions. This system presents a unique opportunity for researchers to investigate various ecological processes on a relatively manageable scale. Notable phenomena such as community assembly, disturbance ecology, and community response to invasion as well as resource production, sharing, and use can all be assessed in the gut environment. Additionally, the human gut microbiota have been implicated for their role in promoting health via immune system priming and development, mediation of host metabolic health, and colonization resistance against invading pathogens. Although microbiota composition appears to fluctuate over time, the functional capacity of these microbial communities can be indicative of both health and disease.

The human gut microbiome is also widely regarded as a notable reservoir for harboring antimicrobial resistance. The ubiquity of resistance is alarming and multiple human pathogens have demonstrated reduced susceptibility to clinically relevant antibiotics in recent years, a trend

that is appearing globally. Widespread disease prevalence coupled with rising rates of resistance among these pathogens suggests that bacterial infections will become more severe and difficult to treat in the coming years with fewer treatment options. Among those microbes showing increasing levels of AMR are enteric pathogens such as *Campylobacter*, *Salmonella*, *Shigella*, and STEC, which are responsible for millions of foodborne infections every year. Increasing incidence of antibiotic resistant enteric infections substantiates the need to further characterize these pathogens' role in the curation and dissemination of AMR across environments. Importantly, these pathogens interact directly with the resident microbiota of the human gut during infection, thereby presenting an interesting ecology that deserves further clarification.

Despite the incredible complexity of systems such as the human gut microbiome, many computational tools have been generated to allow investigation, characterization, and analysis of microbial communities. Shotgun metagenomics sequencing, in particular, allows researchers to capture the genetic diversity of all microbes present within a given sample. Coupling metagenomics with other 'omics technologies such as metabolomics or metatranscriptomics, further enables investigators to identify informative connections between microbial taxa, community function, metabolic output, or enzymatic activity. Although challenges still exist in establishing a widely accepted, standardized analysis pipeline for these 'omics approaches, their power in capturing the diversity of microbial systems is unmatched.

Given these considerations, the primary objective of this dissertation was to characterize community changes in the human gut microbiome associated with enteric infection. Specifically, this study used metagenomics sequencing and metabolomics methods to characterize the composition of microbial taxa, ARGs, and metabolites in the gut environment of healthy, infected, and recovered individuals. Overall, this study will improve understanding of the

ecological consequences related to enteric infection, specifically concerning antimicrobial resistance spread and host metabolic health.

## **REFERENCES**

## REFERENCES

1. Sender R, Fuchs S, Milo R. 2016. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* 164:337-340.
2. Gilbert JA, Blaser MJ, Gregory Caporaso J, Jansson JK, Lynch SV, Knight R. 2018. Current understanding of the human microbiome. *Nature Medicine* 24:392-400.
3. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The Human Microbiome Project. *Nature* 449:804-810.
4. Shreiner AB, Kao JY, Young VB. 2015. The gut microbiome in health and in disease. *Current Opinion in Gastroenterology* 31:69-75.
5. Coyte KZ, Rao C, Rakoff-Nahoum S, Foster KR. 2021. Ecological rules for the assembly of microbiome communities. *PLOS Biology* 19:e3001116.
6. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. 2012. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science* 336:1255-1262.
7. (ed). 2001. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press, ProQuest Ebook Central. Accessed
8. Hubbell SP. 2006. Neutral theory and the evolution of ecological equivalence. *Ecology* 87:1387-1398.
9. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7:601-613.
10. Mihaljevic JR. 2012. Linking metacommunity theory and symbiont evolutionary ecology. *Trends in Ecology & Evolution* 27:323-329.
11. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences* 107.
12. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2010. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences* 108:4578-4585.
13. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan T, Muhammad, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee S, Ying, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L,

- Kristiansen K, Dahlgren J, Wang J. 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe* 17:690-703.
14. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, Stevens HJ, Bennet Jr. WE, Shaikh N, Linneman LA, Hoffmann JA, Hamvas A, Deych E, Shands BA, Shannon WD, Tarr PI. 2014. Patterned progression of bacterial populations in the premature infant gut. *Proceedings of the National Academy of Sciences* 111:12522-12527.
  15. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the Human Intestinal Microbial Flora. *Science* 308:1635-1638.
  16. Lozupone C. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* 489:220-230.
  17. Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet J-P, Ugarte E, Munoz-Tamayo R, Paslier DLE, Nalin R, Dore J, Leclerc M. 2009. Towards the human intestinal microbiota phylogenetic core. *Environmental Microbiology* 11:2574-2584.
  18. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto J-M, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59-65.
  19. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. A core gut microbiome in obese and lean twins. *Nature* 457:480-484.
  20. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Human gut microbes associated with obesity. *Nature* 444:1022-1023.
  21. Valdes AM, Walter J, Segal E, Spector TD. 2018. Role of the gut microbiota in nutrition and health. *BMJ*:k2179.
  22. Walters WA, Xu Z, Knight R. 2014. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Letters* 588:4223-4233.
  23. Sartor RB, Mazmanian SK. 2012. Intestinal Microbes in Inflammatory Bowel Diseases. *The American Journal of Gastroenterology Supplements* 1:15-21.
  24. Larsen N, Vogensen FK, Van Den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sørensen SJ, Hansen LH, Jakobsen M. 2010. Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. *PLoS ONE* 5:e9085.



25. Consortium THMP. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-214.
26. Allison SD, Martiny JBH. 2008. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* 105:11512-11519.
27. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. 2017. The resilience of the intestinal microbiota influences health and disease. *Nature Reviews Microbiology* 15:630-638.
28. Dethlefsen L, Huse S, Sogin ML, Relman DA. 2008. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biology* 6:e280.
29. Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PNAS* 108:4554-4561.
30. Young VB, Schmidt TM. 2004. Antibiotic-Associated Diarrhea Accompanies by Large-Scale Alterations in the Composition of the Fecal Microbiota. *Journal of Clinical Microbiology* 42:1203-1206.
31. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, Hansen TH, Liang S, Feng Q, Zhang C, Pyl PT, Coelho LP, Yang H, Wang J, Typas A, Nielsen MF, Nielsen HB, Bork P, Wang J, Vilsbøll T, Hansen T, Knop FK, Arumugam M, Pedersen O. 2018. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology* 3:1255-1265.
32. Johanesen PA, Mackin KE, Hutton ML, Awad MM, Larcombe S, Amy JM, Lyras D. 2015. Disruption of the gut microbiome: *Clostridium difficile* infection and the threat of antibiotic resistance.
33. Vincent C, Manges A. 2015. Antimicrobial Use, Human Gut Microbiota and *Clostridium difficile* Colonization and Infection. *Antibiotics* 4:230-253.
34. Barra-Carrasco J, Paredes-Sabja D. 2014. *Clostridium difficile* spores: a major threat to the hospital environment. *Future Microbiology* 9.
35. Johnson S. 2009. Recurrent *Clostridium difficile* infection: a review of risk factors, treatments, and outcomes. *Journal of Infection* 58:403-410.
36. Rohlke F, Stollman N. 2012. Fecal microbiota transplantation in relapsing *Clostridium difficile* infection. *Therapeutic Advances in Gastroenterology* 5:403-420.
37. Mallon CA, Elsas JDV, Salles JF. 2015. Microbial Invasions: The Process, Patterns, and Mechanisms. *Trends in Microbiology* 23:719-729.

38. Shea K, Chesson P. 2002. Community ecology theory as a framework for biological invasions. *Trends in Ecology & Evolution* 17:170-176.
39. Kinnunen M, Dechesne A, Proctor C, Hammes F, Johnson D, Quintela-Baluja M, Graham D, Daffonchio D, Fodelianakis S, Hahn N, Boon N, Smets BF. 2016. A conceptual framework for invasion in microbial communities. *The ISME Journal* 10:2773-2779.
40. Vila JCC, Jones ML, Patel M, Bell T, Rosindell J. 2019. Uncovering the rules of microbial community invasions. *Nature Ecology & Evolution* 3:1162-1171.
41. Rivett DW, Jones ML, Ramoneda J, Mombrikotb SB, Ransome E, Bell T. 2018. Elevated success of multispecies bacterial invasions impacts community composition during ecological succession. *Ecology Letters* 21:516-524.
42. Ketola T, Saarinen K, Lindström L. 2017. Propagule pressure increase and phylogenetic diversity decrease community's susceptibility to invasion. *BMC Ecology* 17.
43. Acosta F, Zamor RM, Najar FZ, Roe BA, Hambright KD. 2015. Dynamics of an experimental microbial invasion. *Proceedings of the National Academy of Sciences* 112:11594-11599.
44. Baumgartner M, Pfrunder-Cardozo KR, Hall AR. 2021. Microbial community composition interacts with local abiotic conditions to drive colonization resistance in human gut microbiome samples. *Proceedings of the Royal Society B: Biological Sciences* 288.
45. Kelly CP, Lamont JT. 2008. *Clostridium difficile*— More Difficult Than Ever. *New England Journal of Medicine* 359:1932-1940.
46. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* 3:45-45.
47. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne Illness Acquired in the United States -- Major Pathogens. *Emerging Infectious Disease* 17:7-15.
48. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, Jervis R, Lathrop S, Muse A, Duwell M, Smith K, Tobin-D'Angelo M, Vugia DJ, Zablotsky Kufel J, Wolpert BJ, Tauxe R, Payne DC. 2020. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food -- Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016-2019. *MMWR Morbidity and Mortality Weekly Report* 69:509-514.
49. Koynucu Ozyurt O, Bertocco ALFVB, Pereira LAB, Jimenes LP, Yazisiz H, Ozhak B, Ogunc D, Donmez L, Gunseren F, Yilmaz A, Ongut G. 2019. Detection of *Salmonella*,

- Campylobacter*, Shiga toxin-producing *E. coli* and *Shigella*/EIEC by culture and a multiplex PCR panel in pediatric patients with acute diarrheal illness. *Journal of Laboratory Medicine* 43:211-215.
50. Young KT, Davis LM, DiRita VJ. 2007. *Campylobacter jejuni*: Molecular biology and pathogenesis. *Nature Reviews Microbiology* 5:665-679.
  51. Grassl GA, Finlay BB. 2008. Pathogenesis of enteric *Salmonella* infections. *Current Opinion in Gastroenterology* 24:22-26.
  52. Niyogi SK. 2005. Shigellosis. *The Journal of Microbiology* 43:133-143.
  53. O'Loughlin EV, Robins-Browne RM. 2001. Effect of Shiga toxin and Shiga-like toxins on eukaryotic cells. *Microbes and Infection* 3:493-507.
  54. Jajere SM. 2019. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Veterinary World* 12:504-521.
  55. Tahoun A, Mahajan S, Paxton E, Malterer G, Donaldson S, David, Wang D, Tan A, Gillespie L, Trudi, O'Shea M, Roe J, Andrew, Shaw J, Darren, Gally L, David, Lengeling A, Mabbott A, Neil, Haas J, Mahajan A. 2012. *Salmonella* Transforms Follicle-Associated Epithelial Cells into M Cells to Promote Intestinal Invasion. *Cell Host & Microbe* 12:645-656.
  56. Amavisit P, Lightfoot D, Browning GF, Markham PF. 2003. Variation between Pathogenic Serovars within *Salmonella* Pathogenicity Islands. *Journal of Bacteriology* 185:3624-3635.
  57. Schroeder GN, Hilbi H. 2008. Molecular Pathogenesis of *Shigella* spp.: Controlling Host Cell Signaling, Invasion, and Death by Type III Secretion. *Clinical Microbiology Reviews* 21.
  58. Wassef JS, Keren DF, Mailloux JL. 1989. Role of M Cells in Initial Antigen Uptake and in Ulcer Formation in the Rabbit Intestinal Loop Model of Shigellosis. *Infection and Immunity* 57:858-863.
  59. Mounier J, Vasselon T, Hellio R, Lesourd M, Sansonetti PJ. 1992. *Shigella flexneri* Enters Human Colonic Caco-2 Epithelial Cells through the Basolateral Pole. *Infection and Immunity* 60:237-248.
  60. Melton-Celsa A, Mohawk K, Teel L, O'Brien A. 2012. Pathogenesis of Shig-Toxin Producing *Escherichia coli*. *Current Topics in Microbiology and Immunology* 357:67-103.
  61. Melton-Celsa AR, Smith MJ, O'Brien AD. 2005. Shiga Toxins: Potent Poisons, Pathogenicity Determinants, and Pharmacological Agents. *EcoSal Plus* 1.

62. Warr AR, Kuehl CJ, Waldor MK. 2021. Shiga toxin remodels the intestinal epithelial transcriptional response to Enterohemorrhagic *Escherichia coli*. *PLOS Pathogens* 17:e1009290.
63. Thorpe CM. 2004. Shiga Toxin-Producing *Escherichia coli* Infection. *Clinical Infectious Diseases* 38:1298-1303.
64. Kamada N, Chen GY, Inohara N, Núñez G. 2013. Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology* 14:685-690.
65. Hammami R, Fernandez B, Lacroix C, Fliss I. 2013. Anti-infective properties of bacteriocins: an update. *Cellular and Molecular Life Sciences* 70:2947-2967.
66. Pedicord VA, Lockhart A, A.K., Rangan KJ, Craig JW, Loschko J, Rogoz A, Hang HC, Mucida D. 2016. Exploiting a host-commensal interaction to promote intestinal barrier function and enteric pathogen tolerance. *Science Immunology* 1:eaai7732.
67. Shin R, Suzuki M, Morishita Y. 2002. Influence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic *Escherichia coli* O157:H7. *Journal of Medical Microbiology* 51:201-206.
68. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469:543-547.
69. Gantois I, Ducatelle R, Pasmans F, Haesebrouck F, Hautefort I, Thompson A, Hinton JC, Van Immerseel F. 2006. Butyrate Specifically Down-Regulates Salmonella Pathogenicity Island 1 Gene Expression. *Applied and Environmental Microbiology* 72:946-949.
70. Momose Y, Hirayama K, Itoh K. 2008. Competition for proline between indigenous *Escherichia coli* and *E. coli* O157:H7 in gnotobiotic mice associated with infant intestinal microbiota and its contribution to the colonization resistance against *E. coli* O157:H7. *Antonie van Leeuwenhoek* 94:165-171.
71. Momose Y, Hirayama K, Itoh K. 2008. Effect of organic acids on inhibition of *Escherichia coli* O157:H7 colonization in gnotobiotic mice associated with infant intestinal microbiota. *Antonie van Leeuwenhoek* 93:141-149.
72. Leatham MP, Banerjee S, Autieri SM, Mercado-Lubo R, Conway T, Cohen PS. 2009. Precolonized Human Commensal *Escherichia coli* Strains Serve as a Barrier to *E. coli* O157:H7 Growth in the Streptomycin-Treated Mouse Intestine. *Infection and Immunity* 77:2876-2886.
73. Deriu E, Liu Z, Janet, Pezeshki M, Edwards A, Robert, Ochoa J, Roxanna, Contreras H, Libby J, Stephen, Fang C, Ferric, Raffatellu M. 2013. Probiotic Bacteria Reduce

- Salmonella Typhimurium Intestinal Colonization by Competing for Iron. *Cell Host & Microbe* 14:26-37.
74. Marteyn B, West NP, Browning DF, Cole JA, Shaw JG, Palm F, Mounier J, Prévost M-C, Sansonetti P, Tang CM. 2010. Modulation of *Shigella* virulence in response to available oxygen in vivo. *Nature* 465:355-358.
  75. Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper LV. 2008. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proceedings of the National Academy of Sciences* 105:20858-20863.
  76. Salzman NH, Hung K, Haribhai D, Chu H, Karlsson-Sjoberg J, Amir E, Tegatz P, Barman M, Hayward M, Eastwood D, Stoel M, Zhou Y, Sodergren E, Weinstock GM, Bevins CL, Williams CB, Bos NA. 2010. Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunology* 11:76-82.
  77. Sanos SL, Bui VL, Mortha A, Oberle K, Heners C, Johner C, Diefenbach A. 2009. ROR $\gamma$ t and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46+ cells. *Nature Immunology* 10:83-91.
  78. Satoh-Takayama N, Vosshenrich CAJ, Lesjean-Pottier S, Sawa S, Lochner M, Rattis F, Mention J-J, Thiam K, Cerf-Bensussan N, Mandelboim O, Eberl G, Di Santo JP. 2008. Microbial Flora Drives Interleukin 22 Production in Intestinal NKp46+ Cells that Provide Innate Mucosal Immune Defense. *Immunity* 29:958-970.
  79. Zheng Y, Valdez PA, Danilenko DM, Hu Y, Sa SM, Gong Q, Abbas AR, Modrusan Z, Ghilardi N, de Sauvage FJ, Ouyang W. 2008. Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Medicine* 14:282-289.
  80. Strober W, Watanabe T. 2011. NOD2, an intracellular innate immune sensor involved in host defense and Crohn's disease. *Mucosal Immunology* 4:484-495.
  81. Petnicki-Ocwieja T, Hrnčir T, Liu Y-J, Biswas A, Hudcovic T, Tlaskalova-Hogenova H, Kobayashi KS. 2009. Nod2 is required for the regulation of commensal microbiota in the intestine. *Proceedings of the National Academy of Sciences* 106:15813-15818.
  82. Franchi L, Kamada N, Nakamura Y, Burberry A, Kuffa P, Suzuki S, Shaw MH, Kim Y-G, Nunez G. 2012. NLRC4-driven production of IL-1 $\beta$  discriminates between pathogenic and commensal bacteria and promotes host intestinal defense. *Nature Immunology* 13:449-456.
  83. Keeney KM, Finlay BB. 2011. Enteric pathogen exploitation of the microbiota-generated nutrient environment of the gut. *Current Opinion in Microbiology* 14:92-98.
  84. Fischbach MA, Lin H, Liu DR, Walsh CT. 2006. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nature Chemical Biology* 2:132-138.

85. Brown SA, Palmer KL, Whiteley M. 2008. Revisiting the host as a growth medium. *Nature Reviews Microbiology* 6:657-666.
86. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, Russell JM, Bevins CL, Adams LG, Tsois RM, Roth JR, Baumler AJ. 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467:426-429.
87. Moya A, Ferrer M. 2016. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology* 24:402-413.
88. Comte J, Fauteaux L, del Giordio PA. 2013. Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities. *Frontiers in Microbiology* 4.
89. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, Doebeli M, Wegener Parfrey L. 2018. Function and functional redundancy in microbial systems. *Nature Ecology & Evolution* 2:936-943.
90. Pérez-Cobas AE, Artacho A, Knecht H, Ferrús ML, Friedrichs A, Ott SJ, Moya A, Latorre A, Gosalbes MJ. 2013. Differential Effects of Antibiotic Therapy on the Structure and Function of Human Gut Microbiota. *PLoS ONE* 8:e80201.
91. Kampmann C, Dicksved J, Engstrand L, Rautelin H. 2016. Composition of human faecal microbiota in resistance to *Campylobacter* infection. *Clinical Microbiology and Infection* 22:61.e1-61.e8.
92. Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, Brett Finlay B. 2008. Antibiotic-Induced Perturbations of the Intestinal Microbiota Alter Host Susceptibility to Enteric Infection. *INFECTION AND IMMUNITY* 76:4726-4736.
93. Thaiss CA, Levy M, Grosheva I, Zheng D, Soffer E, Blacher E, Braverman S, Tengeler AC, Barak O, Elazar M, Ben-Zeev R, Lehavi-Regev D, Katz MN, Pevsner-Fischer M, Gertler A, Halpern Z, Harmelin A, Amar S, Serradas P, Grosfeld A, Shapiro H, Geiger B, Elinav E. 2018. Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. *Science* 359:1376-1383.
94. Shin N-R, Whon TW, Bae J-W. 2015. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology* 33:496-503.
95. Thingholm LB, Ruhlemann MC, Koch M, Fuqua B, Laucke G, Boehm R, Bang C, Franzosa EA, Hubenthal M, Rahnavard A, Frost F, Lloyd-Price J, Schirmer M, Lusi AJ, Vulpe CD, Lerch MM, Homuth G, Kacprowski T, Schmidt CO, Nothlings U, Karlsen TH, Lieb W, Laudes M, Franke A, Huttenhower C. 2019. Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition. *Cell Host & Microbe* 26:252-264.
96. Ventola CL. 2015. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics* 40:277-283.

97. CDC. 2019. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA.
98. Cosgrove SE. 2006. The Relationship between Antimicrobial Resistance and Patient Outcomes: Mortality, Length of Hospital Stay, and Health Care Costs. *Clinical Infectious Diseases* 42:S82-S89.
99. Founou LL, Founou RC, Essack SY. 2016. Antibiotic Resistance in the Food Chain: A Developing Country-Perspective. *Frontiers in Microbiology* 7.
100. Zaman SB, Hussain MA, Nye R, Mehta V, Mamun KT, Hossain N. 2017. A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* 9:e1403.
101. Uddin TM, Chakraborty AJ, Khusro A, Zidan BRM, Mitra S, Emran TB, Dhama K, Ripon KH, Mario G, Sahibzada MUK, Hossain J, Koirala N. 2021. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *Journal of Infection and Public Health* 14:1750-1766.
102. Walsh C, Wencewicz T. 2016. *Antibiotics: Challenges, Mechanisms, Opportunities*. ASM Press, Washington D.C.
103. Etebu E, Ariekpar I. 2016. Antibiotics: Classification and mechanisms of action with emphasis on molecular perspectives. *International Journal of Applied Microbiology and Biotechnology Research* 4:90-101.
104. Pankey GA, Sabath LD. 2004. Clinical Relevance of Bacteriostatic versus Bactericidal Mechanisms of Action in the Treatment of Gram-Positive Bacterial Infections. *Clinical Infectious Diseases* 38:864-870.
105. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, Guerin PJ, Piddock LJV. 2016. Antimicrobials: access and sustainable effectiveness 2. *Lancet* 387:176-187.
106. Santos-Lopez A, Marshall CW, Haas AL, Turner C, Rasero J, Cooper VS. 2021. The roles of history, chance, and natural selection in the evolution of antibiotic resistance. *eLife* 10:e70676.
107. Reygaert WC. 2018. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiology* 4:482-501.
108. Van Acker H, Van Dijck P, Coenye T. 2014. Molecular mechanisms of antimicrobial tolerance and resistance in bacterial and fungal biofilms. *Trends in Microbiology* 22:326-333.
109. Cornaglia G, Mazzariol A, Fontana R, Satta G. 1996. Diffusion of Carbapenems Through the Outer Membrane of Enterobacteriaceae and Correlation of Their Activities with Their Periplasmic Concentrations. *Microbial Drug Resistance* 2:273-276.

110. Chow JW, Shlaes DM. 1991. Imipenem resistance associated with the loss of a 40 kDa outer membrane protein in *Enterobacter aerogenes*. *Journal of Antimicrobial Chemotherapy* 28:499-504.
111. Bush K, Bradford PA. 2016.  $\beta$ -Lactams and  $\beta$ -Lactamase Inhibitors: An Overview. *Cold Spring Harbor Perspectives in Medicine* 6:a025247.
112. Lin J, Nishino K, Roberts MC, Tolmasky M, Aminov RI, Zhang L. 2015. Mechanisms of antibiotic resistance. *Frontiers in Microbiology* 6.
113. Piddock LJV. 2006. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria.
114. Kumar A, Schweizer HP. 2005. Bacterial resistance to antibiotics: Active efflux and reduced uptake. *Advanced Drug Delivery Reviews* 57:1486-1513.
115. Butaye P, Cloeckaert A, Schwarz S. 2003. Mobile genes coding for efflux-mediated antimicrobial resistance in Gram-positive and Gram-negative bacteria. *International Journal of Antimicrobial Agents* 22:205-210.
116. Wallace MJ, Fishbein SRS, Dantas G. 2020. Antimicrobial resistance in enteric bacteria: current state and next-generation solutions. *Gut Microbes* 12:e1799654.
117. Sun D. 2018. Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria. *Frontiers in Microbiology* 9.
118. Balcazar JL. 2014. Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. *PLoS Pathogens* 10:e1004219.
119. Wozniak RAF, Waldor MK. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology* 8:552-563.
120. Norman A, Hansen LH, Sørensen SJ. 2009. Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:2275-2289.
121. Lopatkin AJ, Meredith HR, Srimani JK, Pfeiffer C, Durrett R, You L. 2017. Persistence and reversal of plasmid-mediated antibiotic resistance. *Nature Communications* 8.
122. Salyers AA, Shoemaker NB, Stevens AM, Li L-Y. 1995. Conjugative Transposons: an Unusual and Diverse Set of Integrated Gene Transfer Elements. *Microbiology Reviews* 59:579-590.
123. Surette MD, Wright GD. 2017. Lessons from the Environmental Antibiotic Resistome. *Annual Review of Microbiology* 71:309-329.



124. Wright GD. 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology* 5:175-186.
125. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. *Science (New York, NY)* 337:1107-11.
126. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, Calderon M, Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G. 2016. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533:212-216.
127. Kim D-W, Cha C-J. 2021. Antibiotic resistome from the One-Health perspective: understanding and controlling antimicrobial resistance transmission. *Experimental & Molecular Medicine* 53:301-309.
128. Ouyang W-Y, Huang F-Y, Zhao Y, Li H, Su J-Q. 2015. Increased levels of antibiotic resistance in urban stream of Jiulongjiang River, China. *Applied Microbiology and Biotechnology* 99:5697-5707.
129. Chen B, Yang Y, Liang X, Yu K, Zhang T, Li X. 2013. Metagenomic Profiles of Antibiotic Resistance Genes (ARGs) between Human Impacted Estuary and Deep Ocean Sediments. *Environmental Science & Technology* 47:12753-12760.
130. Lee K, Kim D-W, Lee D-H, Kim Y-S, Bu J-H, Cha J-H, Thawng CN, Hwang E-M, Seong HJ, Sul WJ, Wellington EMH, Quince C, Cha C-J. 2020. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic resistance. *Microbiome* 8.
131. Ma L, Li B, Jiang X-T, Wang Y-L, Xia Y, Li A-D, Zhang T. 2017. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* 5:154-154.
132. Zhao Y, Yang QE, Zhou X, Wang F-H, Muurinen J, Virta MP, Brandt KK, Zhu Y-G. 2021. Antibiotic resistome in the livestock and aquaculture industries: Status and solutions. *Critical Reviews in Environmental Science and Technology* 51:2159-2196.
133. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, Teillant A, Laxminarayan R. 2015. Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences* 112:5649-5654.
134. Zhao Y, Su J-Q, An X-L, Huang F-Y, Rensing C, Brandt KK, Zhu Y-G. 2018. Feed additives shift gut microbiota and enrich antibiotic resistance in swine gut. *Science of the Total Environment* 621:1224-1232.
135. Yazdankhah S, Rudi K, Berhoft A. 2014. Zinc and copper in animal feed - development of resistance and co-resistance to antimicrobial agents in bacteria of animal origin. *Microbial Ecology in Health and Disease* 25.

136. Sun J, Liao X-P, D'Souza AW, Boolchandani M, Li S-H, Cheng K, Luis Martínez J, Li L, Feng Y-J, Fang L-X, Huang T, Xia J, Yu Y, Zhou Y-F, Sun Y-X, Deng X-B, Zeng Z-L, Jiang H-X, Fang B-H, Tang Y-Z, Lian X-L, Zhang R-M, Fang Z-W, Yan Q-L, Dantas G, Liu Y-H. 2020. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nature Communications* 11.
137. Van Schaik W. 2015. The human gut resistome. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20140087.
138. Feng J, Li B, Jiang X, Yang Y, Wells GF, Zhang T, Li X. 2018. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environmental Microbiology* 20:355-368.
139. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, Pan Y, Li J, Zhu L, Wang X, Meng Z, Zhao F, Liu D, Ma J, Qin N, Xiang C, Xiao Y, Li L, Yang H, Wang J, Yang R, Gao GF, Wang J, Zhu B. 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications* 4.
140. Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A, Bork P. 2013. Country-specific antibiotic use practices impact the human gut resistome. *Genome Research* 23:1163-1169.
141. Hansen ZA, Cha W, Nohomovich B, Newton DW, Lephart P, Salimnia H, Khalife W, Shade A, Rudrik JT, Manning SD. 2021. Comparing gut resistome composition among patients with acute *Campylobacter* infections and healthy family members. *Scientific Reports* 11.
142. Raymond F, Boissinot M, Ouameur AA, Déraspe M, Plante P-L, Kpanou SR, Bérubé È, Huletsky A, Roy PH, Ouellette M, Bergeron MG, Corbeil J. 2019. Culture-enriched human gut microbiomes reveal core and accessory resistance genes. *Microbiome* 7.
143. Sommer MOA, Dantas G, Church GM. 2009. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science (New York, NY)* 325:1128-1131.
144. Pickering LK. 2004. Antimicrobial resistance among enteric pathogens. *Seminars in Pediatric Infectious Diseases* 15:71-77.
145. Ballal M. 2016. Chapter 4 - Trends in Antimicrobial Resistance Among Enteric Pathogens: A Global Concern, p 63-92. *In* Kon K, Rai M (ed), *Antibiotic Resistance*. Academic Press.
146. CDC. 2018. National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Surveillance Report for 2015 (Final Report). U.S. Department of Health and Human Services, Atlanta, Georgia.
147. Rodrigues JA, Cha W, Mosci RE, Mukherjee S, Newton DW, Lephart P, Salimnia H, Khalife W, Rudrik JT, Manning SD. 2021. Epidemiologic Associations Vary Between

- Tetracycline and Fluoroquinolone Resistant *Campylobacter jejuni* Infections. *Frontiers in Public Health* 9.
148. Sproston EL, Wimalarathna HML, Sheppard SK. 2018. Trends in fluoroquinolone resistance in *Campylobacter*. *Microbial Genomics* 4.
  149. Bae J, Oh E, Jeon B. 2014. Enhanced Transmission of Antibiotic Resistance in *Campylobacter jejuni* Biofilms by Natural Transformation. *Antimicrobial Agents and Chemotherapy* 58:7573-7575.
  150. Gal-Mor O, Boyle EC, Grassl GA. 2014. Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella enterica* serovars differ. *Frontiers in Microbiology* 5.
  151. Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. 2015. Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive *Salmonella* Infections. *Clinical Microbiology Reviews* 28:901-937.
  152. Mukherjee S, Anderson CM, Mosci RE, Newton DW, Lephart P, Salimnia H, Khalife W, Rudrik JT, Manning SD. 2019. Increasing Frequencies of Antibiotic Resistant Non-typhoidal *Salmonella* Infections in Michigan and Risk Factors for Disease. *Frontiers in Medicine* 6.
  153. Carraro N, Durand R, Rivard N, Anquetil C, Barrette C, Humbert M, Burrus V. 2017. *Salmonella* genomic island 1 (SGI1) reshapes the mating apparatus of IncC conjugative plasmids to promote self-propagation. *PLOS Genetics* 13:e1006705.
  154. Cohen E, Davidovich M, Rokney A, Valinsky L, Rahav G, Gal-Mor O. 2019. Emergence of new variants of antibiotic resistance genomic islands among multidrug-resistant *Salmonella enterica* in poultry. *Environmental Microbiology* 22:413-432.
  155. Ranjbar R, Farahani A. 2019. *Shigella*: Antibiotic-Resistance Mechanisms And New Horizons For Treatment.
  156. Shiferaw B, Solghan S, Palmer A, Joyce K, Barzilay EJ, Krueger A, Cieslak P. 2012. Antimicrobial Susceptibility Patterns of *Shigella* Isolates in Foodborne Diseases Active Surveillance Network (FoodNet) Sites, 2000–2010. *Clinical Infectious Diseases* 54:S458-S463.
  157. Mukherjee S, Blankenship HM, Rodrigues JA, Mosci RE, Rudrik JT, Manning SD. 2021. Antibiotic Susceptibility Profiles and Frequency of Resistance Genes in Clinical Shiga Toxin-Producing *Escherichia coli* Isolates from Michigan over a 14-Year Period. *Antimicrobial Agents and Chemotherapies* 65.
  158. Freedman SB, Xie J, Neufeld MS, Hamilton WL, Hartling L, Tarr PI. 2016. Shiga Toxin–Producing *Escherichia coli* Infection, Antibiotics, and Risk of Developing

- Hemolytic Uremic Syndrome: A Meta-analysis. *Clinical Infectious Diseases* 62:1251-1258.
159. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* 73:1121-1137.
  160. Brolund A, Rajer F, Giske CG, Melefors Ö, Titelman E, Sandegren L. 2019. Dynamics of Resistance Plasmids in Extended-Spectrum- $\beta$ -Lactamase-Producing Enterobacteriaceae during Postinfection Colonization. *Antimicrobial Agents and Chemotherapy* 63.
  161. Fan Y, Pedersen O. 2021. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology* 19:55-71.
  162. Gomaa EZ. 2020. Human gut microbiota/microbiome in health and diseases: a review. *Antonie van Leeuwenhoek* 113:2019-2040.
  163. Lin L, Zhang J. 2017. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC Immunology* 18.
  164. Schippa S, Conte M. 2014. Dysbiotic Events in Gut Microbiota: Impact on Human Health. *Nutrients* 6:5786-5805.
  165. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. 2018. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210-215.
  166. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Kim AD, Shmagel AK, Syed AN, Walter J, Menon R, Koecher K, Knights D. 2019. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host & Microbe* 25:789-802.e5.
  167. Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. 2016. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529:212-215.
  168. Jumpertz R, Le DS, Turnbaugh PJ, Trinidad C, Bogardus C, Gordon JI, Krakoff J. 2011. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *The American Journal of Clinical Nutrition* 94:58-65.
  169. Roager HM, Hansen LBS, Bahl MI, Frandsen HL, Carvalho V, Gøbel RJ, Dalgaard MD, Plichta DR, Sparholt MH, Vestergaard H, Hansen T, Sicheritz-Pontén T, Nielsen HB, Pedersen O, Lauritzen L, Kristensen M, Gupta R, Licht TR. 2016. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nature Microbiology* 1:16093.

170. Janssen AWF, Kersten S. 2015. The role of the gut microbiota in metabolic health. *The FASEB Journal* 29:3111-3123.
171. Wolters M, Ahrens J, Romaní-Pérez M, Watkins C, Sanz Y, Benítez-Páez A, Stanton C, Günther K. 2019. Dietary fat, the gut microbiota, and metabolic health – A systematic review conducted within the MyNewGut project. *Clinical Nutrition* 38:2504-2520.
172. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, De Vos WM, Zucker J-D, Raes J, Hansen T, Bork P, Wang J, Ehrlich SD, Pedersen O. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500:541-546.
173. Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget AL, DelméE E, Cousin BA, Sulpice T, Chamontin B, Ferrière J, Tanti J-FO, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin RM. 2007. Metabolic Endotoxemia Initiates Obesity and Insulin Resistance. *Diabetes* 56:1761-1772.
174. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, Segal E. 2019. Structural variation in the gut microbiome associates with host health. *Nature* 568:43-48.
175. Louis P, Flint HJ. 2017. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology* 19:29-41.
176. Hill M. 1997. Intestinal flora and endogenous vitamin synthesis. *European Journal of Cancer Prevention* 6:S43-5.
177. Richards JL, Yap YA, Mcleod KH, Mackay CR, Mariño E. 2016. Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases. *Clinical & Translational Immunology* 5:e82.
178. Jacobson DK, Honap TP, Ozga AT, Meda N, Kagoné TS, Carabin H, Spicer P, Tito RY, Obregon-Tito AJ, Reyes LM, Troncoso-Corzo L, Guija-Poma E, Sankaranarayanan K, Lewis CM. 2021. Analysis of global human gut metagenomes shows that metabolic resilience potential for short-chain fatty acid production is strongly influenced by lifestyle. *Scientific Reports* 11.
179. Vital M, Howe A, Bergeron N, Krauss RM, Jansson JK, Tiedje JM. 2018. Metagenomic Insights into the Degradation of Resistant Starch by Human Gut Microbiota. *Applied and Environmental Microbiology* 84:e01562-18.
180. Miller TL, Wolin MJ. 1996. Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Applied and Environmental Microbiology* 62:1589-1592.

181. Louis P, Flint HJ. 2009. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters* 294:1-8.
182. Duncan SH, Holtrop G, Lobley GE, Calder AG, Stewart CS, Flint HJ. 2004. Contribution of acetate to butyrate formation by human faecal bacteria. *British Journal of Nutrition* 91:915-923.
183. Vital M, Howe AC, Tiedje JM. 2014. Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data. *mBio* 5:e00889-14-e00889.
184. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, Flint HJ, Louis P. 2014. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME Journal* 8:1323-1335.
185. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. 2013. The role of short-chain fatty acids in the interplay between diet, gut, microbiota, and host energy metabolism. *Journal of Lipid Research* 54:2325-2340.
186. Correa-Oliveira R, Fachi JL, Vieira A, Sato FT, Vinolo MAR. 2016. Regulation of immune cell function by short-chain fatty acids. *Clinical & Translational Immunology* 5.
187. Macia L, Tan J, Vieira AT, Leach K, Stanley D, Luong S, Maruya M, Ian Mckenzie C, Hijikata A, Wong C, Binge L, Thorburn AN, Chevalier N, Ang C, Marino E, Robert R, Offermanns S, Teixeira MM, Moore RJ, Flavell RA, Fagarasan S, Mackay CR. 2015. Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nature Communications* 6:6734.
188. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, Schilter HC, Rolph MS, Mackay F, Artis D, Xavier RJ, Teixeira MM, Mackay CR. 2009. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* 461:1282-1286.
189. Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ. 2012. The Warburg Effect Dictates the Mechanism of Butyrate Mediated Histone Acetylation and Cell Proliferation. *Molecular Cell* 48:612-626.
190. Zeng X, Sunkara LT, Jiang W, Bible M, Carter S, Ma X, Qiao S, Zhang G. 2013. Induction of Porcine Host Defense Peptide Gene Expression by Short-Chain Fatty Acids and Their Analogs. *PLoS ONE* 8:e72922.
191. Cândido FG, Valente FX, Grześkowiak ŁM, Moreira APB, Rocha DMUP, Alfenas RDCG. 2018. Impact of dietary fat on gut microbiota and low-grade systemic inflammation: mechanisms and clinical implications on obesity. *International Journal of Food Sciences and Nutrition* 69:125-143.
192. Nogal A, Louca P, Zhang X, Wells PM, Steves CJ, Spector TD, Falchi M, Valdes AM, Menni C. 2021. Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat. *Frontiers in Microbiology* 12.

193. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, Petersen KF, Kibbey RG, Goodman AL, Shulman GI. 2016. Acetate mediates a microbiome–brain– $\beta$ -cell axis to promote metabolic syndrome. *Nature* 534:213-217.
194. Puddu A, Sanguineti R, Montecucco F, Viviani GL. 2014. Evidence for the Gut Microbiota Short-Chain Fatty Acids as Key Pathophysiological Molecules Improving Diabetes. *Mediators of Inflammation* 2014:1-9.
195. Parada Venegas D, De la Fuente MK, Landskron G, Gonzalez MJ, Quera R, Dijkstra G, Harmsen HJM, Faber KN, Hermoso MA. 2019. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Frontiers in Microbiology* 10.
196. Duboc H, Rainteau D, Rajca S, Humbert L, Farabos D, Maubert M, Grondin V, Jouet P, Bouhassira D, Seksik P, Sokol H, Coffin B, Sabate JM. 2012. Increase in fecal primary bile acids and dysbiosis in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterology & Motility* 24:513-e247.
197. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Doré J. 2009. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflammatory Bowel Diseases* 15:1183-1189.
198. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. 2002. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 52:2141-2146.
199. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K. 2018. Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition* 57:1-24.
200. Magnusdottir S, Ravcheev D, de Crecy-Lagard V, Thiele I. 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Frontiers in Genetics* 6.
201. Soto-Martin EC, Warnke I, Farquharson FM, Christodoulou M, Horgan G, Derrien M, Faurie J-M, Flint HJ, Duncan SH, Louis P. 2020. Vitamin Biosynthesis by Human Gut Butyrate-Producing Bacteria and Cross-Feeding in Synthetic Microbial Communities. *mBio* 11.
202. Ridlon JM, Kang DJ, Hylemon PB, Bajaj JS. 2014. Bile Acids and the Gut Microbiome. *Current Opinion in Gastroenterology* 30:332-338.
203. Chiang JY. 2009. Bile acids: regulation of synthesis. *Journal of Lipid Research* 50:1955-1966.
204. Zeng H, Umar S, Rust B, Lazarova D, Bordonaro M. 2019. Secondary Bile Acids and Short Chain Fatty Acids in the Colon: A Focus on Colonic Microbiome, Cell

- Proliferation, Inflammation, and Cancer. *International Journal of Molecular Sciences* 20:1214.
205. 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. *Proceedings of the National Academy of Sciences* 105:13580-13585.
  206. Hofmann AF, Hagey LR. 2008. Bile Acids: Chemistry, Pathochemistry, Biology, Pathobiology, and Therapeutics. *Cellular and Molecular Life Sciences* 65:2461-2483.
  207. Bernstein H, Bernstein C, Payne CM, Dvorakova K, Garewal H. 2005. Bile acids as carcinogens in human gastrointestinal cancers. *Mutation Research/Reviews in Mutation Research* 589:47-65.
  208. Begley M, Gahan CGM, Hill C. 2005. The interaction between bacteria and bile. *FEMS Microbiology Reviews* 29:625-651.
  209. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall H-U, Bamberg K, Angelin B, Hyotylainen T, Oresic M, Backhed F. 2013. Gut Microbiota Regulates Bile Acid Metabolism by Reducing Levels of Tauro-beta-muricholic Acid, a Naturally Occurring FXR Antagonist. *Cell Metabolism* 17:225-235.
  210. Galloway-Peña J, Hanson B. 2020. Tools for Analysis of the Microbiome. *Digestive Diseases and Sciences* 65:674-685.
  211. Clarridge III JE. 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews* 17.
  212. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM. 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications* 10.
  213. Větrovský T, Baldrian P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE* 8:e57923.
  214. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK. 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research* 47:e103-e103.
  215. Nguyen N-P, Warnow T, Pop M, White B. 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms and Microbiomes* 2:16004.



216. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* 73.
217. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* 6:610-618.
218. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse 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 Research* 42:D643-D648.
219. Nilsson RH, Larsson K-H, Taylor S, Andy F, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K. 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* 47:D259-D264.
220. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7:335-336.
221. 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, Duvallet C, Edwardson 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, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37:852-857.
222. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75:7537-7541.
223. Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581-583.
224. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* 35:833-844.

225. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
226. Anonymous. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.
227. Yeoh YK. 2021. Removing Host-derived DNA Sequences from Microbial Metagenomes via Mapping to Reference Genomes, p 147-153, *The Plant Microbiome*. Springer US.
228. Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences* 98.
229. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420-1428.
230. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674-1676.
231. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome research* 27:824-834.
232. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* 40:e155-e155.
233. Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology* 13:R122.
234. Vollmers J, Wiegand S, Kaster A-K. 2017. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *PLOS ONE* 12:e0169662-e0169662.
235. Breitwieser FP, Lu J, Salzberg SL. 2019. A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics* 20:1125-1136.
236. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, Demare MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaitre C, Peterlongo P, Rizk G, Lavenier D, Wu YW, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin HH, Liao YC, et al. 2017. Critical Assessment of Metagenome Interpretation - A benchmark of metagenomics software. *Nature Methods* 14:1063-1071.

237. Tamames J, Cobo-Simón M, Puente-Sánchez F. 2019. Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics* 20.
238. Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. 2012. Individual genome assembly from complex community short-read metagenomic datasets. *The ISME Journal* 6:898-901.
239. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431-437.
240. Stewart EJ. 2012. Growing Unculturable Bacteria. *Journal of Bacteriology* 194:4151-4160.
241. Wommack KE, Bhavsar J, Ravel J. 2008. Metagenomics: Read Length Matters. *Applied Environmental Microbiology* 74.
242. Treiber ML, Taft DH, Korf I, Mills DA, Lemay DG. 2020. Pre- and post-sequencing recommendations for functional annotation of human fecal metagenomes. *BMC Bioinformatics* 21.
243. Carr R, Borenstein E. 2014. Comparative analysis of functional metagenomic annotation and the mappability of short reads. *PloS one* 9:e105776-e105776.
244. Ye SH, Siddle KJ, Park DJ, Sabeti PC. 2019. Benchmarking Metagenomics Tools for Taxonomic Classification, vol 178, p 779-794. Cell Press.
245. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15:R46.
246. Ounit R, Wanamaker S, Close TJ, Lonardi S. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* 16.
247. Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* 26:1721-1729.
248. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7:11257-11257.
249. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12:59-60.

250. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* 9:811-814.
251. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods* 12:902-903.
252. Beghini F, Mciver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10.
253. Lindgreen S, Adair KL, Gardner PP. 2016. An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports* 6.
254. Sun Z, Huang S, Zhang M, Zhu Q, Haiminen N, Carrieri AP, Vázquez-Baeza Y, Parida L, Kim H-C, Knight R, Liu Y-Y. 2021. Challenges in benchmarking metagenomic profilers. *Nature Methods* 18:618-626.
255. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research* 38:e132-e132.
256. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* 42:D199-D205.
257. Galperin MY, Wolf YI, Makarova KS, Alvarez V, Roberto, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research* 49:D274-D281.
258. Anonymous. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research* 42:D191-D198.
259. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, Huttenhower C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods* 15.
260. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2019. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*.
261. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: A database for classification of antimicrobial drug,

- biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research* 48:D561-D569.
262. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Research* 45:D574-D580.
263. Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME journal* 9:207-16.
264. Chen L. 2004. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* 33:D325-D328.
265. Fiehn O. 2002. Metabolomics — the link between genotypes and phenotypes, p 155-171, *Functional Genomics*. Springer Netherlands.
266. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D’Auria J, Ewald J, C. Ewald J, Fraser PD, Giavalisco P, Hall RD, Heinemann M, Link H, Luo J, Neumann S, Nielsen J, Perez De Souza L, Saito K, Sauer U, Schroeder FC, Schuster S, Siuzdak G, Skirycz A, Sumner LW, Snyder MP, Tang H, Tohge T, Wang Y, Wen W, Wu S, Xu G, Zamboni N, Fernie AR. 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods* 18:747-756.
267. Lamichhane S, Sen P, Dickens AM, Oresic M, Bertram HC. 2018. Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. *Methods* 149:3-12.
268. Aron AT, Gentry EC, Mcphail KL, Nothias L-F, Nothias-Esposito M, Bouslimani A, Petras D, Gauglitz JM, Sikora N, Vargas F, Van Der Hooft JJJ, Ernst M, Kang KB, Aceves CM, Caraballo-Rodríguez AM, Koester I, Weldon KC, Bertrand S, Roullier C, Sun K, Tehan RM, Boya P. CA, Christian MH, Gutiérrez M, Ulloa AM, Tejeda Mora JA, Mojica-Flores R, Lakey-Beitia J, Vásquez-Chaves V, Zhang Y, Calderón AI, Tayler N, Keyzers RA, Tugizimana F, Ndlovu N, Aksenov AA, Jarmusch AK, Schmid R, Truman AW, Bandeira N, Wang M, Dorrestein PC. 2020. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols* 15:1954-1991.
269. Vincenti F, Montesano C, Di Ottavio F, Gregori A, Compagnone D, Sergi M, Dorrestein P. 2020. Molecular Networking: A Useful Tool for the Identification of New Psychoactive Substances in Seizures by LC-HRMS. *Frontiers in Chemistry* 8.
270. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, Mohny RP, Small KS, Bell JT, Steves CJ, Valdes AM, Spector TD, Menni C. 2018. The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics* 50:790-795.
271. Bashiardes S, Zilberman-Schapira G, Elinav E. 2016. Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and Biology Insights* 10:BBIS34610.

272. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the Human Oral Microbiome during Health and Disease. *mBio* 5:e01012-14-e01012.
273. Lai LA, Tong Z, Chen R, Pan S. 2019. Metaproteomics Study of the Gut Microbiome. *In* Wang X, Kuruc M (ed), *Functional Proteomics Methods in Molecular Biology*, vol 1871. Humana Press, New York, NY.
274. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. 2020. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* 14:117793221989905.
275. Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, Wishart D. 2019. Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites* 9:76.
276. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, Del Pozo-Yauner L, Ochoa-Leyva A. 2015. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome.
277. Li H. 2015. Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annual Review of Statistics and Its Application* 2:73-94.
278. Qian X-B, Chen T, Xu Y-P, Chen L, Sun F-X, Lu M-P, Liu Y-X. 2021. A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. *Chinese Medical Journal* 133:1844-1855.
279. Lozupone C, Knight R. 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 71:8228-8235.
280. Xia Y, Sun J. 2017. Hypothesis testing and statistical analysis of microbiome. *Genes and Diseases* 4:138-148.
281. Ramette A. 2007. Multivariate analyses in microbial ecology. *FEMS microbiology ecology* 62:142-60.
282. Sudarikov K, Tyakht A, Alexeev D. 2017. Methods for The Metagenomic Data Visualization and Analysis. *Current Issues in Molecular Biology*:37-58.
283. Nearing JT, Douglas GM, Hayes MG, Macdonald J, Desai DK, Allward N, Jones CMA, Wright RJ, Dhanani AS, Comeau AM, Langille MGI. 2022. Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications* 13.
284. Wallen ZD. 2021. Comparison study of differential abundance testing methods using two large Parkinson disease gut microbiome datasets derived from 16S amplicon sequencing. *BMC Bioinformatics* 22.

285. Bastian M, Heymann S. 2009. Gephi : An Open Source Software for Exploring and Manipulating Networks.361-362.
286. 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 Research 13:2498-2504.

## CHAPTER 2

Comparing gut resistome composition among patients with acute *Campylobacter* infections and healthy family members

This chapter is from an open-access published manuscript:

Hansen ZA, Cha W, Nohomovich B, Newton DW, Lephart P, Hossein S, Khalife W, Shade A, Rudrik JT, Manning SD. 2021. Comparing gut resistome composition among patients with acute *Campylobacter* infections and healthy family members. *Scientific Reports* 11.

DOI: 10.1038/s41598-021-01927-7



## ABSTRACT

*Campylobacter* commonly causes foodborne infections and antibiotic resistance is an imminent concern. It is not clear, however, if *Campylobacter* affects the human gut ‘resistome’ during infection. Application of shotgun metagenomics on stools from 26 cases with *Campylobacter* infections and 44 healthy family members (controls) identified 406 unique antibiotic resistance genes (ARGs) representing 153 genes/operons, 40 mechanisms, and 18 classes. Cases had greater ARG richness ( $p < 0.0001$ ) and Shannon diversity ( $p < 0.0001$ ) than controls with distinct compositions ( $p = 0.000999$ ; PERMANOVA). Cases were defined by multidrug resistance genes and dominated by Proteobacteria species (40.8%), specifically those in *Escherichia* (20.9%). Tetracycline resistance genes ARGs were most abundant in controls with Bacteroidetes (45.3%) and Firmicutes (44.4%) dominating. Hierarchical clustering of cases identified three clusters with distinct resistomes. Case clusters 1 and 3 differed from controls containing more urban and hospitalized patients. Relative to family members of the same household, ARG composition among matched cases was mostly distinct, though some familial controls had similar profiles that could be explained by a shorter time since exposure to the case. Together, these data indicate that *Campylobacter* infection is associated with an altered resistome composition and increased ARG diversity, raising concerns about the role of infection in the spread of resistance determinants.

## INTRODUCTION

Enteric pathogens are common causes of foodborne illness affecting 9.4 million individuals each year in the United States; 3.6 million of these enteric infections are caused by bacteria (1). In 2018, the Centers for Disease Control and Prevention (CDC) reported that the incidence of foodborne infection was highest for *Campylobacter* and *Salmonella*, with the incidence of both pathogens increasing relative to the frequencies reported in 2015-2017 (2). The pathogenesis and virulence of these enteric pathogens have been well characterized and, more recently, several studies have examined how enteric pathogens influence the gut microbiota. For example, our prior study showed that infection by one of four enteric pathogens resulted in decreased diversity of the gut microbiota, specifically resulting in an increase in the relative abundance of *Proteobacteria* (3). Thus, further consideration and characterization of the ecological consequences of enteric infection in the human gut microbiome is needed.

In addition to their role in causing foodborne illness, *Campylobacter* spp., are progressively found to be drug-resistant, which has led to their classification as a serious public health threat by the CDC (4). The increasing prevalence of *Campylobacter* spp. linked to human infections plus their enhanced ability to evade modern antibiotics substantiate the need to further understand their total impact on health. Generally, antibiotic resistance increasingly results in adverse health and economic outcomes due to the growing prevalence and emergence of drug-resistant infections (1, 5). Growing awareness of these burdens has led to a rise in the number of studies investigating the resistome, or the compilation of antimicrobial resistance genes (ARGs), within microbial communities (6). Several studies have investigated resistomes across different environments including the guts of humans, cattle, poultry, and swine (7-9). These environments do not exist in isolation; one study found similar genetic regions containing ARGs among environmental soil isolates and five relevant human pathogens (10), while another identified

ARGs that could cross habitat boundaries (11). Many of these genes are co-localized with mobile genetic elements and other ARGs, suggesting significant potential for transmission of multiple resistance genes via horizontal gene transfer. This spread of antimicrobial resistance across environments illuminates our need to further clarify the ecological mechanisms facilitating such exchange.

As in other ecosystems, the human gut microbial community exhibits microbe-microbe and microbe-host interactions, temporal and spatial dynamics, and has varied community responses to disturbance or species invasion (12). Multiple studies have explored the ability of the human gut microbiota to recover after a disturbance such as antibiotic treatment (13, 14). One longitudinal study, for instance, investigated the effects of repeated antibiotic exposure in infants and found that antibiotic use contributed to a loss of species- and strain-level diversity (15). Just as disturbance has the capacity to uproot stable communities, so, too, does microbial invasion. Microbial invasion ecology involves the introduction of a foreign microbe to a stable environment and follows a trajectory from establishment to growth and spread, leading to downstream ecological consequences (16). Previous studies have examined the importance of microbial invasion in various environmental contexts such as soil, plant, and agricultural settings (17, 18). However, investigation of microbial invasion as it pertains to the human gut microbiome and resistome via infection has yet to be fully explored. Elucidating the impacts of ecological invasion on the composition and mobility of ARGs in the human gut is crucial to advancing our fight against the spread of drug resistance.

Given the health and economic burden of foodborne pathogens and the ubiquity of antimicrobial resistance across environments, further understanding the impacts of infection on ARGs and their dissemination is needed. This study therefore aims to understand the impact of

enteric infection by a bacterial pathogen, *Campylobacter*, on the human gut resistome using shotgun metagenome analyses.

## **MATERIALS AND METHODS**

### **Study population**

Between 2011 and 2015, 26 stools were obtained from patients with *Campylobacter* infections prior to treatment. Most infections were caused by *C. jejuni*, although one patient had *C. coli* and three isolates could not be classified. Samples were collected via the Michigan Department of Health and Human Services (MDHHS) as described (3). Briefly, stools were added to Cary-Blair transport media, cultured for *Campylobacter* spp., and transported to Michigan State University (MSU). Upon receipt at MSU, stool samples were homogenized, centrifuged and aliquoted for analysis and/or storage at -80°C. Metagenomic stool DNA was extracted using the QIAamp DNA Stool Mini Kit (QIAGEN; Valencia, CA) as described (19). Epidemiological data about demographics, exposures, hospitalization, and symptoms were extracted from the Michigan Disease Surveillance System (MDSS) and household members were contacted for inclusion as study controls. Forty-four healthy household family members submitted a stool 5-21 weeks after the cases' infection and completed a questionnaire about exposures and symptoms. Sixteen families were included. While 10 cases and 7 controls were not matched to a shared household, they were included in the overall comparative case versus control analyses. County of residence was classified as 'rural' or 'urban' based on the classification scheme developed by the National Center for Health Statistics (NCHS) (20). These classifications were based upon 2010 census data while considering the 2013 county designation assigned by the Office of Management and Budget as metropolitan, micropolitan, or noncore, as well as the specific population sizes and city location for metropolitan areas.

Study protocols and consent procedures were performed as described (3) in accordance with the relevant guidelines and regulations. Approval to conduct the study was granted by the Institutional Review Boards at MSU (IRB #10-736SM), the MDHHS (842-PHALAB), and the four hospital laboratories. Each participant and/or their legal guardian was required to provide informed consent prior to enrollment and was given a monetary incentive after each sample was submitted. Data were stripped of all personal identifiable information prior to use.

### **Sample preparation and sequencing analysis**

Metagenomic DNA from the 70 stools was extracted, sheared, and normalized as described (3). Library construction was completed using a TruSeq Nano library kit (Illumina, Inc., San Diego, CA, USA). Shotgun sequencing was performed in a series of four sequencing runs on an Illumina HiSeq 2500. Reads were demultiplexed at the MSU Research Technology Support Facility (RTSF). Sequencing run was investigated as a potential source of batch effects prior to analysis of the data; considerable overlap was observed across runs (Figure A.1).

AmrPlusPlus v2.0 was used to perform quality control and align and annotate the metagenomic fragments using the MEGARes 1.0 database (21). This database was chosen for its comprehensive, hand-curated compilation of ARGs and associated annotation structure containing three hierarchical levels that maximizes the number of representative sequences and lacks cycles or statistical dependencies. Trimmomatic (22) was used to remove adapters and poor-quality reads. Specifically, the reads were trimmed by removing the three leading and trailing nucleotides, followed by trimming of the 5' end of the sequence until an average Phred score of >15 was attained in a sliding window of size four. Short sequences <36 nucleotides were discarded. If reads matched to adapter sequences with less than or equal to two mismatches, then they were eligible for clipping to ensure adapter removal; successful clipping was dependent on a

match score of  $\geq 30$  using a publicly-accessible adapters file provided on GitHub (<https://github.com/BioInfoTools/BBMap/blob/master/resources/adapters.fa>).

Metagenomic reads were mapped to the human genome (GRCh38\_latest\_genomic.fna.gz, downloaded December 2020 from RefSeq) using Burrows-Wheeler Aligner (BWA) (23); SAMTools (24) and BEDTools (25) were used to remove these human genomic sequences from each sample. Following trimming, quality filtering, and host genome removal, 176,686,501 of the 217,104,781 raw paired-end reads were used for downstream analyses. The number of paired-end reads used in the analysis did not significantly differ between cases and controls ( $p=0.051$ ). The estimated and actual sequencing coverage were determined using Nonpareil (26); the average coverage was estimated to be 83.0% (Figure A.2). Average Genome Size (AGS) and the number of genome equivalents (GE) within each sample were quantified using MicrobeCensus (27). Because AGS analyses have been considered a potential source of bias in gene-based metagenomic comparisons (28), comparing communities across different sample types may be confounded by varying AGS. Additionally, AGS analyses may provide insight into the ecological capacity of samples; those with a larger AGS may represent generalist taxa, while those with a smaller AGS may represent more specialist species (29). In our study, AGS was higher in cases (4,406,749.57 bp) versus controls (4,004,525.52 bp) ( $p=0.02$ , Wilcoxon rank sum test; Figure A.3). Because no difference in the number of GE was observed between cases ( $n=238.1$ ) and controls ( $n=273.5$ ) ( $p=0.23$ , Wilcoxon rank sum test), raw ARG abundance counts were normalized across samples using GE metrics.

### **Identification of antimicrobial resistance genes (ARGs)**

Non-host FASTQ files resulting from human genome removal were aligned to MEGARes 1.0 (21) using BWA (23) and SAMTools (24) with default parameters to classify

ARGs present in each sample. Reads were deduplicated and annotated using ResistomeAnalyzer with an identity threshold of  $\geq 80\%$  to quantify ARG abundance per sample. RarefactionAnalyzer was performed to obtain the data necessary to assess the adequacy of our sequencing depth. SNPs known to be important for antibiotic resistance were also extracted from the metagenomes using the AmrPlusPlus pipeline (30). These SNPs were analyzed with the Resistance Gene Identifier (RGI) created in conjunction with The Comprehensive Antibiotic Resistance Database (CARD) (31) to confirm or reject their presence in ARGs within our samples. In this analysis, however, all ARGs were considered, including those without confirmation of SNP presence. These ARGs were included because they were within a single point mutation and remain relevant even if they serve as a resistance precursor. In future studies, a more in-depth analysis including these SNP data may further illuminate differences between study groups.

Output at the gene level included the target gene, its sequence identity, and putative function; however, output at the group level, or the overall gene- or operon-level group for a given sequence was used. The mechanism level, which indicates the biological mechanism of resistance encoded by each sequence, was also provided as well as the class level representing the antibiotic class relevant to each ARG.

### **Identification of microbial taxa**

FASTQ reads with the human genome removed via AmrPlusPlus v2.0 were taxonomically annotated using the classifier Kaiju (32). The NCBI BLAST nr database including sequences for bacteria, archaea, viruses, fungi, and microbial eukaryotes was used as a reference. The alignment mode used in Kaiju was 'greedy', meaning that a maximum of three mismatches were allowed when identifying taxonomic signatures in sequences. A match length cutoff of 11 nucleotides and the default match score of 65 was used when classifying sequences as well. Raw

abundances of reads assigned to taxa were normalized by the estimated number of genome equivalents calculated by MicrobeCensus (27).

### **Ecological analyses**

Resistome composition was determined by investigating the identity and diversity of ARGs across samples at the gene, group, and class levels. The relative abundance of each ARG was determined per sample by dividing the number of GE-normalized reads for a specific ARG gene, group, or class by the total number of GE-normalized reads for that sample. Alpha and beta diversity metrics, including ordination plots (PCoA) based on Bray-Curtis dissimilarity at the gene level, were determined using the vegan package (33) in R (34). The Wilcoxon rank-sum test was used to test for statistical significance between case and control samples (alpha diversity), while PERMANOVA and PERMDISP were used to detect differences in the centroid (mean) and dispersion (degree of spread) across case and control groups (beta diversity).

For the family case-control pairs, the ‘envfit’ function of the vegan package was used to fit environmental variables onto the ordination generated via the PCoA. MaAsLin2 (35) was used to generate log-transformed linear models exploring multivariate associations among resistome features and relevant metadata. Default values were used for all significance cutoffs as well as normalization (total sum scaling; TSS), transformation (log transform), and multiple hypothesis testing correction (Benjamini-Hochberg; BH) with a target False Discovery Rate of 0.05.

### **Hierarchical clustering and epidemiological analysis**

Case clusters were defined based on the Bray-Curtis dissimilarity among cases at the gene and group levels using the ‘ape’ package (36) in R and were examined using PCoA and plotted using



vegan. The Wilcoxon rank-sum test was used to test for statistical significance between case clusters (alpha diversity), while MaAsLin2 (35) was used to identify differentially abundant ARGs at the group and class level across clusters. For epidemiological analyses, Chi-square tests were used to detect significant differences in epidemiological variables (e.g., patient sex, age, residence (rural vs. urban), and symptoms) between cases and controls and identify associations with case clusters.

## **RESULTS**

### **Characteristics of the study population**

Stools from 26 patients with acute campylobacteriosis (cases) and 44 related healthy family members (controls) from the same household were compared. Controls belonged to 16 different families with two to eight participating members. Although 10 cases and seven controls were not matched to a family, they were included in the comparative case versus control analyses. Among cases, 17 (65.4%) were female with 13 (50%) between 19-64 years of age, 8 (30.7%) between 0-9 years old, and 5 (19.2%) greater than 65 years (Table A.1). Controls had a slightly different demographic distribution in which 18 (40.9%) were female; 17 (38.7%) were between 0-9 years old, 4 (9.1%) were 10-18 years of age, 21 (47.7%) were 19-64 years, and 2 (4.5%) were greater than 65 years old. No significant differences were observed in the age or sex distribution between groups. Although more controls resided in urban areas, the difference was not significant and is likely due to the recruitment of more than one control per case depending on the household. Most cases self-identified as Caucasian (n=22; 88.0%) and reported abdominal pain (n=21; 80.1%) and diarrhea (n=24; 92.3%). Nausea (n=9; 34.6%), vomiting (n=6; 23.1%), and bloody stool (n=10; 38.5%) were also reported with 20 cases (76.0%) receiving outpatient care and six (23%) requiring hospitalization. Among the latter, four (66.7%) were hospitalized

for two days, one for three days, and another for six. Three of the 26 cases (11.5%) and three of the 44 controls (6.8%) received antibiotics two weeks prior to sample collection.

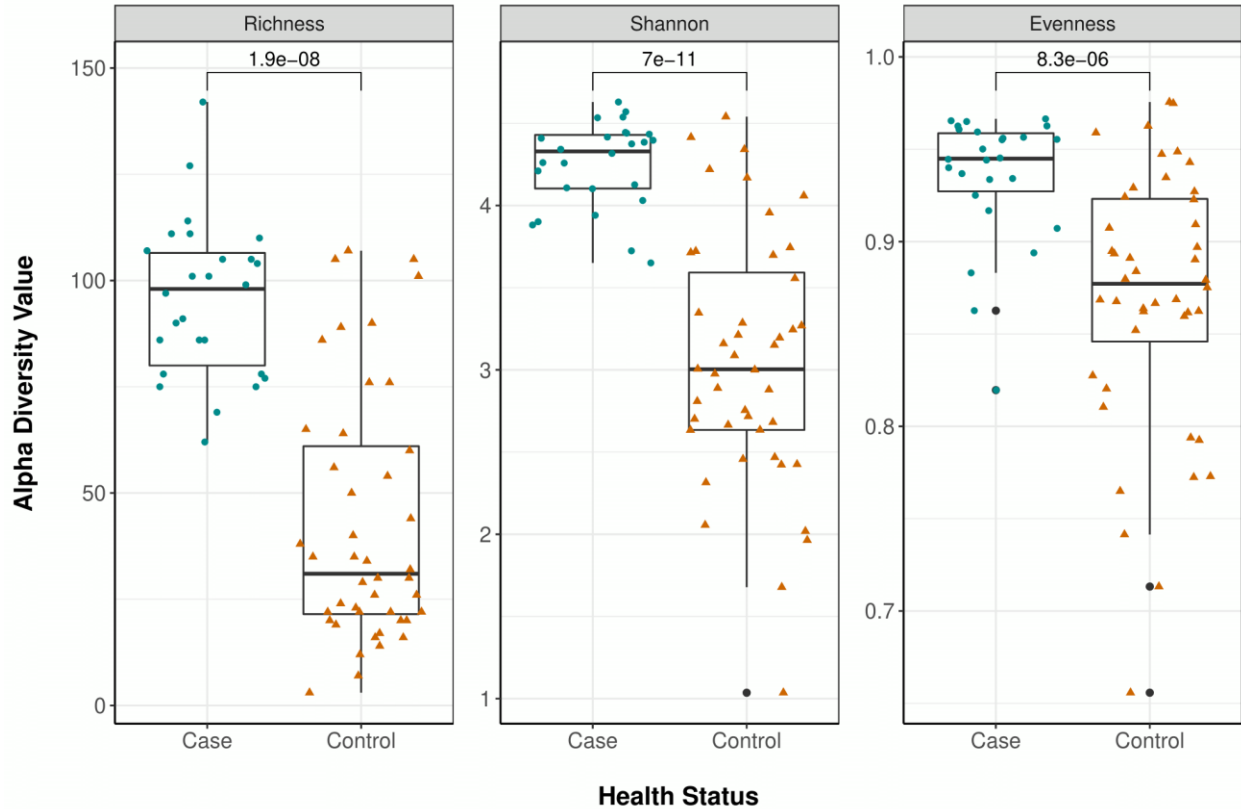
### **Number and diversity of ARGs vary depending on health status**

In total, 406 unique genes representing 153 ARG groups or operons for 18 antibiotic classes and 40 resistance mechanisms were detected. Three measures of alpha diversity (ARG richness, Shannon diversity, and evenness) differed significantly between groups (Figure 2.1). The mean richness, or unique ARGs per sample, was greater in cases ( $S=95.7$ ;  $\text{min}=62$ ,  $\text{max}=142$ ) than controls ( $S=42.8$ ;  $\text{min}=3$ ,  $\text{max}=107$ ;  $p<0.0001$ ) as were the mean Shannon Diversity Index (cases= $4.25$  vs. controls= $3.05$ ;  $p<0.0001$ ) and resistome evenness ( $J'=0.935$  (cases) vs.  $J'=0.869$  (controls);  $p<0.0001$ ).

Principal Coordinate Analysis (PCoA) using Bray-Curtis dissimilarity also revealed separation between case and control resistomes (Figure 2.2). Indeed, health status, or identity as a case or control, had a significant effect on the centroid of each group as assessed using Permutational Analysis of Variance (PERMANOVA  $p=0.000999$ ;  $F=14.083$ ). The dispersion of points within each cluster evaluated using Permutational Analysis of Multivariate Dispersion, however, was not significantly different (PERMDISP  $p=0.115$ ;  $F=2.6264$ ), suggesting that the comparison between group centroids is valid. Participants reporting antibiotic use two weeks prior to sample collection did not cluster separately from other samples within each group.

### **Specific ARGs define case and control samples**

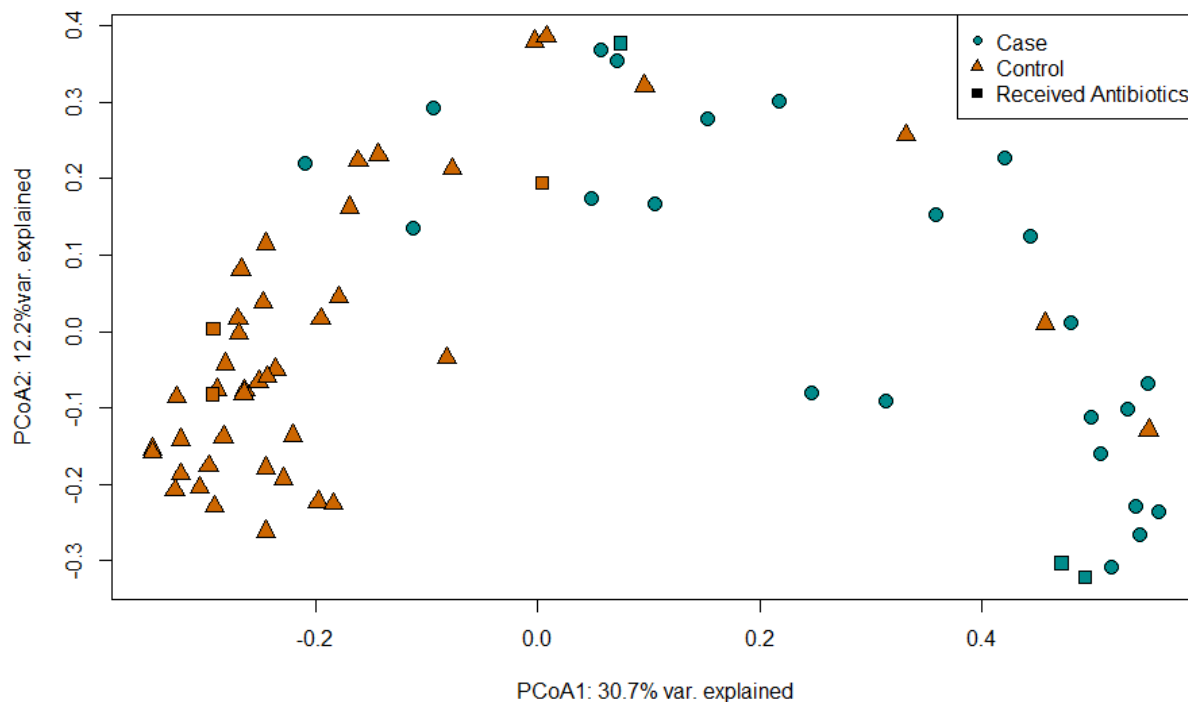
At the antibiotic class level, ARGs for multidrug resistance (MDR), defined as phenotypic resistance to  $\geq 1$  antibiotic belonging to more than 3 drug classes, had the highest average relative abundance (42.6%) in cases followed by tetracycline (11.0%) and



**Figure 2.1. Resistomes in cases are more diverse than resistomes of controls.**

Three measures of alpha diversity (Richness, Shannon diversity, and Pielou's Evenness) are shown stratified by health status. The median of each measure is indicated by the thick black bar in each box and the first and third quartiles are represented by the bottom and top of the box, respectively; points (circles and triangles) show variation within each sample type. Outlying points within each group are indicated by the black dots associated with each boxplot. P-values were calculated using the Wilcoxon rank-sum test and are shown above the comparison bar within each plot.

fluoroquinolone (9.5%) (Figure 2.3). For controls, tetracycline (54.4%) and beta-lactam (16.0%) ARGs were most represented. At the group level (i.e., gene or operon), *tetQ* encoding tetracycline resistance was most abundant in both cases (8.0%) and controls (33.0%). In cases, the next highest groups were *mdtC* (5.9%) encoding a MDR efflux pump subunit in MdtBC and *rpoB* (5.4%), the beta 30S RNA polymerase subunit gene important for rifampin, glycopeptide and lipopeptide resistance. Controls had a greater relative abundance of *tetW* (11.7%) encoding a ribosomal protection protein important for tetracycline resistance and the class A beta-lactamase *cfx* (11.1%). Among both sets of samples, the three respective predominant ARG groups



**Figure 2.2. Resistomes of cases and controls are distinct.**

Principal Coordinates Analysis (PCoA) plot of case (cyan, circles) and control (orange, triangles) resistomes based on Bray-Curtis dissimilarity. The first and second coordinate are shown with their respective percentage of explained variance. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by square data points.

represented ~60% of ARGs in controls compared to <20% of the ARGs in case resistomes, further highlighting the increased resistome diversity within case communities.

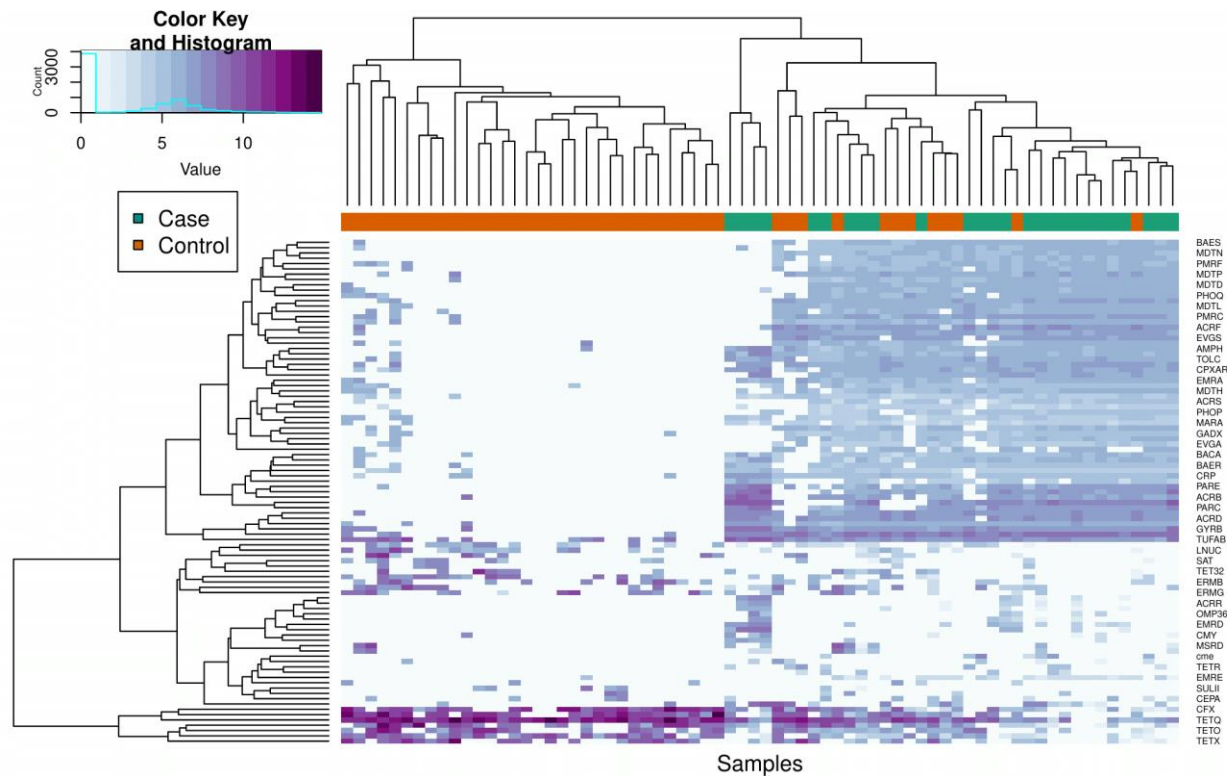
Normalizing by the number of genome equivalents per sample also detected differences in actual ARG abundance. Roughly 1,216 MDR genes were detected in cases versus 160 in controls. Cases also had more fluoroquinolone (n=254) and aminoglycoside (n=204) resistance genes than controls (n=26, n=31, respectively), while controls had more tetracycline ARGs (n=270) than cases (n=101). Moreover, clear differences in ARG abundance were observed across samples at the group level and hierarchical clustering revealed two distinct resistome clusters (Figure 2.4). Of these clusters, one is comprised entirely of controls (n=28) and the other contains samples from all 26 cases and 12 controls.

MaAsLin2 (35) was used to generate log-transformed linear models to identify differentially abundant ARGs among cases and controls. These models used health status as a fixed effect and residence type, age, and sex as random effects. At the class level, tetracycline, MLS, and beta-lactam ARGs were significantly associated with controls (adjusted p-values=2.2E-11; 0.004; 0.021, respectively). In cases, the greatest association was observed for MDR (C=-4.69; adjusted p-value=4.01E-11) and fluoroquinolone resistance (C=-4.37; adjusted p-value=2.3E-12) relative to controls. At the gene level, increased abundance of MDR and fluoroquinolone resistance genes such as *cpxAR*, *mdtC*, *parE*, and *parC*, was observed among cases after adjusting for demographic variables (Table A.2).



**Figure 2.3. Relative abundance of ARGs differs among cases and controls.**

The relative abundance of ARGs assigned to 18 different antibiotic classes is shown with each column representing the resistome from one individual. Relative abundances were determined using raw ARG abundances that had been normalized by the approximate number of genome equivalents in the sample as determined using MicrobeCensus. CAP = cationic antimicrobial peptides; MLS = Macrolide, Lincosamide, Streptogramin; MDR = Multidrug resistance.



**Figure 2.4. Hierarchical clustering illustrates group level ARG abundance differences between cases and controls.**

The columns represent the resistome communities per sample, which are ordered based on similarity in the top X-axis dendrogram that displays two resistome clusters. Case and control samples are indicated by the color bar below the dendrogram (cases = cyan; controls = orange). The Y-axis shows the hierarchical clustering of ARG groups as they appear in sample resistomes; ARG group names are indicated in small print on the right. Those ARG groups with a cumulative normalized abundance value <5 across all samples were excluded from the analysis. Relative abundance is indicated by the color key; a value of 15 (deep purple) indicates that there are approximately 15 normalized copies of that ARG in a sample, while a value of 0 (light blue/white) indicates a very low or negligible abundance.

Comparatively, tetracycline (*tetQ* and *tetW*) and beta-lactam resistance genes (*cfx* and *bla*) were associated with controls. Similar ARG classes (Figure A.4) and ARGs (Figure A.5) were also found to differentiate case and control samples using Linear Discriminant Analysis (LDA) Effect Size (LEfSe) (37).

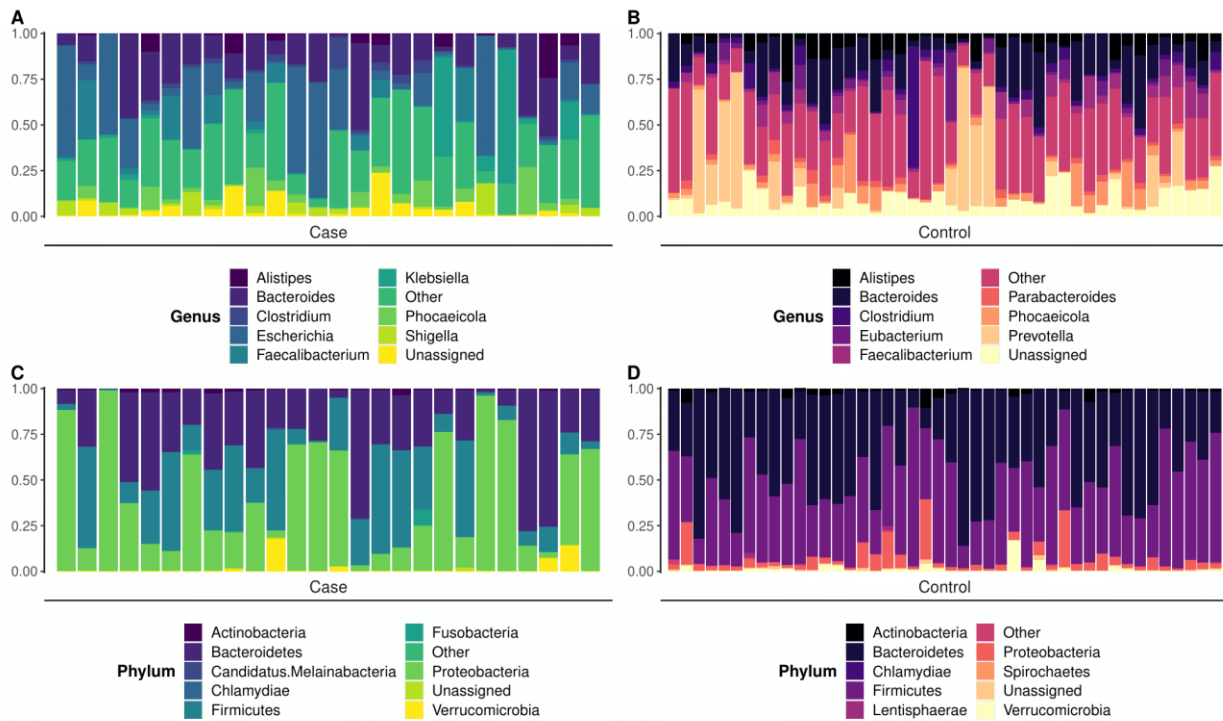
## Taxonomic diversity differs between cases and controls

A total of 40,227 species were detected among the case and control samples including bacteria, archaea, fungi, and viruses. Mean taxonomic richness was significantly greater in controls (S=6,374; min=1,506, max=15,548) compared to cases (S=3,605; min=1,499, max=11,612;  $p < 0.0001$ ), a trend that was also observed for Shannon diversity (case=3.36, control=4.24;  $p = 0.00014$ ) (Figure A.6). Expectedly, taxonomic composition was also distinct among cases and controls (Figure 2.5). Cases were mostly comprised of *Proteobacteria* (average relative abundance = 40.8%) followed by *Bacteroidetes* (30.8%) represented primarily by the genera *Escherichia* (20.9%) and *Bacteroides* (18.1%), respectively. Conversely, controls were dominated by *Bacteroidetes* and *Firmicutes* with average relative abundances of 45.3% and 44.4%, respectively. The most highly represented genera in controls were *Bacteroides* (15.5%) and *Prevotella* (12.8%), both members of the *Bacteroidetes* phylum. Notably, a single control sample contained a high proportion of *Prevotella*, which accounted for 78% of its taxonomic abundance; with this outlying sample removed, the average relative abundance across controls for *Prevotella* was 11.2%.

The actual abundances of these bacterial groups also differed in taxonomic composition among cases and controls (Figure A.7). While cases had an average of 472 reads assigned to the genus *Escherichia*, controls had an average of just 37 *Escherichia* reads. Controls were dominated by *Prevotella* with an average of 1,626 reads; this incredibly high number was due to the outlier sample, which had a high abundance of *Prevotella*. With this sample removed, the average number of *Prevotella* reads across controls was 411. Conversely, cases had an average of 75 *Prevotella* reads per sample. Among all cases, *Campylobacter* only comprised an average relative abundance of 0.28% at the time of sample collection. When considering actual abundance, cases had an average of only 4.0 *Campylobacter* reads per sample.

## Specific ARGs are not strongly associated with *Campylobacter* in the case samples

An analysis exploring correlations between the genus *Campylobacter* and ARG groups was pursued to investigate the potential role of the invading pathogens in shaping case resistomes. Spearman rank correlations between ARG and taxonomic abundances in cases were taken with a cutoff value of  $\rho \geq 0.75$ . Although no significant correlations were observed between *Campylobacter* and other taxa or ARGs above this cutoff, statistically significant correlations with lower coefficients were detected. Namely, *Campylobacter* was positively correlated with



### Figure 2.5. Taxonomic relative abundance notably differs between cases and controls.

The relative abundance of bacterial genera and phyla in each sample are displayed as columns for cases (A, C) and controls (B, D). Similar to ARG relative abundance, taxonomic relative abundances were determined using raw abundances that had been normalized by the approximate number of genome equivalents in the sample as determined using MicrobeCensus. For the phylum and genus levels, the top-10 phyla and genera were chosen, respectively, based on the highest average relative abundance assigned to a specific phylum or genus among cases or controls (which were considered separately). The remaining read abundances for phyla or genera in samples were summed and are shown in the category “Other.” Note: plots for cases (A, C) and controls (B, D) contain the same respective color schemes but that these refer to genera (A, B) and phyla (C, D), respectively.



*cme*, a gene encoding a class A beta-lactamase (coeff=0.585; p=0.00169), and *cmeR* that encodes a MDR efflux pump (coeff=0.505; p=0.00857). No other correlations with coefficients >0.50 were observed between *Campylobacter* and ARGs in case samples.

Intriguingly, ARGs that best defined case samples were correlated with other taxa identified in case and control metagenomes (Table A.3). For example, *mdtC* was highly abundant and positively correlated with *Shigella* (coeff=0.886; p<0.0001), *Pseudoalteromonas* (coeff=0.789; p<0.0001), *Rhodococcus* (coeff=0.785; p<0.0001), and *Phytobacter* (coeff=0.756; p<0.0001). Although most of these genera were not overly abundant in cases, *Shigella* was among the top-10 most abundant genera. In addition, *cpxAR*, which encodes a regulatory system for a MDR efflux pump and was highly abundant in cases, was positively correlated with *Pseudoalteromonas* (coeff=0.839; p<0.0001), *Phytobacter* (coeff=0.793; p<0.0001), and *Siccibacter* (coeff=0.784; p<0.0001); none of these were among the top-100 most abundant genera.

### **Clusters 1 and 3 have more diverse resistomes than Cluster 2**

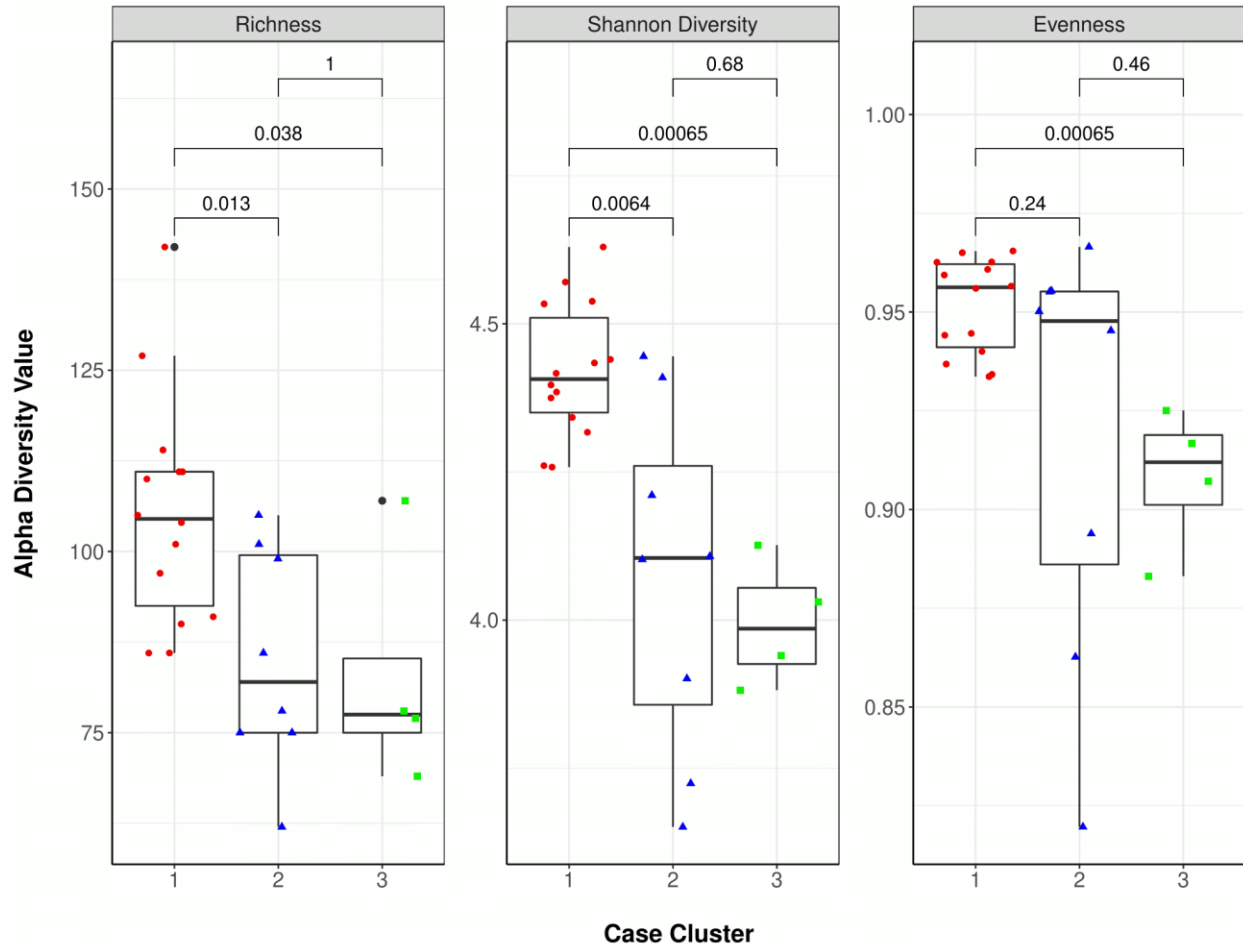
Hierarchical clustering of case resistomes at the gene level using Bray-Curtis dissimilarity identified three separate clusters among the 26 case samples (Figure A.8). The Cluster 1 cases had a significantly greater mean ARG richness (S=105) than Cluster 2 (S=85.1) and Cluster 3 (S=82.8) cases (p=0.01 and p=0.04, respectively; Wilcoxon rank-sum test) (Figure 2.6). Cluster 1 resistomes also had a significantly greater Shannon Diversity Index than Clusters 2 (p=0.006) and 3 (p=0.0007) and a greater Pielou's evenness score than Cluster 3 (p=0.0007). Evenness did not differ between Clusters 1 and 2 (p=0.24). To visualize each case cluster in relation to controls, a PCoA was generated using Bray-Curtis dissimilarity (Figure 2.7). In this analysis, Cluster 2 resistomes were more similar to controls, whereas Cluster 1 resistomes

separated along the first and second coordinate with Cluster 3 oriented in between. The difference between the centroids of each case cluster was significant (PERMANOVA  $p=0.000999$ ;  $F=8.7401$ ).

### **Case epidemiological data is linked to specific resistome profiles**

Among 25 of the 26 cases with data available, those residing in urban versus rural settings were significantly more likely to have resistome profiles belonging to Clusters 1 or 3 than Cluster 2 (Fisher's Exact test  $p=0.0007$ ). While no significant differences were observed for any of the symptoms across the three clusters, eight of 10 (80.0%) cases reporting bloody stool and five of six (83.3%) cases requiring hospitalization had resistome profiles belonging to Clusters 1 or 3. In addition, 12 of the 17 (70.6%) cases with Cluster 1 or Cluster 3 profiles reported animal contact within one week of illness.

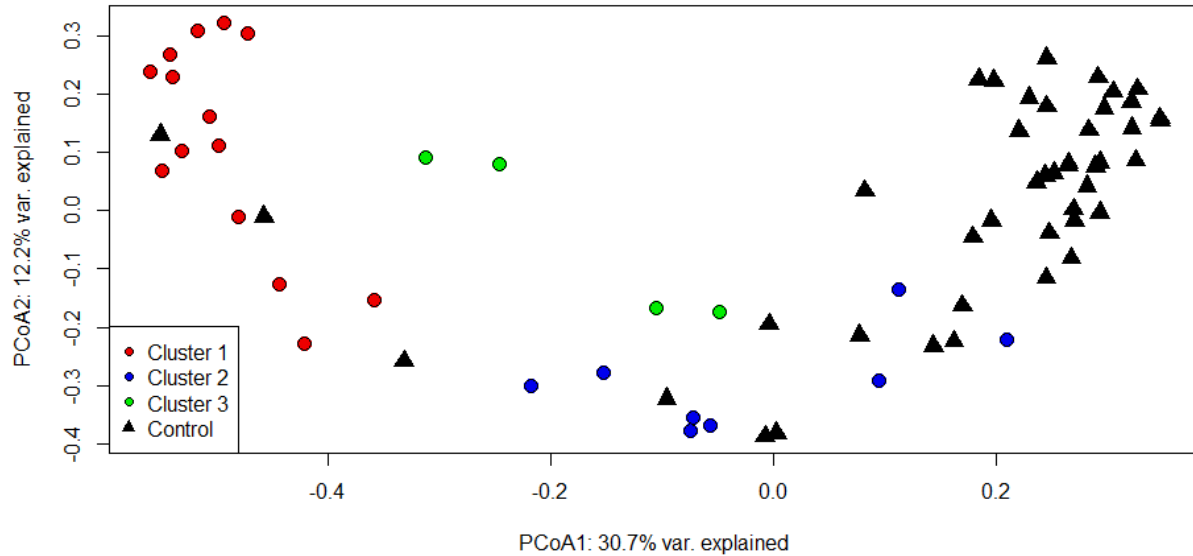
To further explore associations between ARGs, case clusters and epidemiological data, MaAsLin2 was used with cluster and residence type as fixed effects and age and sex as random effects (Table A.4). Relative to Cluster 1, significant associations were identified between Cluster 2 profiles and MLS and tetracycline ARGs ( $C= 5.276026, 2.692487$ ; adjusted  $p$ -value= $0.03564, 0.048547$ , respectively), whereas fosfomycin ( $C= 3.426063$ ; adjusted  $p$ -value= $8.61E-05$ ), aminocoumarin ( $C= 1.481023$ ; adjusted  $p$ -value= $0.001451$ ), and elfamycin ( $C= 1.303181$ ; adjusted  $p$ -value= $0.000107$ ) ARGs were associated with Cluster 3. By contrast, Cluster 1 communities were associated with aminoglycoside ( $C= -1.66939$ ; adjusted  $p$ -value= $1.30E-09$ ), cationic antimicrobial peptides ( $C= -5.03738$ ; adjusted  $p$ -value= $1.60E-17$ ) and MDR ( $C= -0.47969$ ; adjusted  $p$ -value= $0.002346$ ) classes relative to both Clusters 2 and 3. Similar results were determined when using an alternative method, LeFSe, for identifying



**Figure 2.6. Case Cluster 1 resistomes are more diverse than resistomes of Clusters 2 and 3 combined.**

Boxplot displaying alpha diversity metrics for resistomes of case Cluster 1 (red circles), Cluster 2 (blue triangles) and Cluster 3 (green squares). Resistome richness, diversity (Shannon), and evenness are indicated. The median of each measure is shown by the black bar within each box and the first and third quartiles are indicated by the bottom and top of the box, respectively; points show variation within sample types. P-values were calculated using the Wilcoxon rank-sum test and are shown above the comparison bar within each plot.

differentially abundant ARGs among clusters at the class level (Figure A.9). Intriguingly, trimethoprim was the only ARG class associated with rural residence, while the group level analysis identified *dfhR*, which is important for trimethoprim resistance, to be more common in rural cases (Table A.5). Five additional ARGs including *tetA* and *tetB* (tetracycline), *mphA* (macrolides), *aac3* (aminoglycoside), and *ANT3-DPRIME* (aminoglycoside), were also significantly more common in rural versus urban residents.



**Figure 2.7. Case resistomes cluster separately and case Cluster 2 is more similar to control samples.**

Principal Coordinates Analysis (PCoA) plot of the three case clusters (red, blue, green; circles) compared to the control resistomes (black; triangles) based on Bray-Curtis dissimilarity at the gene level. The first and second coordinate are shown with their respective percentage of explained variance. Case Cluster 1 separates clearly from Clusters 2 and 3 along the first and second coordinate, while case Cluster 2 aligns with a handful of control samples.

### **Family relation is less influential than health status in shaping the gut resistome during enteric infection**

An analysis of 16 families was pursued by comparing case samples to 1-7 family members (controls) who submitted stools 5-21 weeks following the case's infection. Although no significant differences were observed for richness, evenness, and Shannon diversity by family (Figure 2.8), the resistome composition varied considerably. The latter result is supported by an examination of beta diversity metrics since a PCoA revealed little separation among the families with health status contributing to most of the separation (Figure 2.9). Some cases and controls within a family, however, were in closer proximity than expected, which is in line with the ARG distribution and abundances by family. With the exception of a few families, resistome

composition among cases was clearly distinct from those observed among their related controls (Figure A.10).

To explain the variance observed among families, environmental factors and vectors were fit onto the ordination. These variables included gender, age in years, residence type, and number of days since exposure to the infected family member (controls only). There was a significant correlation between the number of days since exposure and the ordination values ( $p=0.001$ ,  $R^2=0.543$ ). Interestingly, the directionality of this influence corresponds with controls that were less similar than their associated cases. In other words, the longer the time between a case being infected and a control submitting a stool sample, the less similar the control resistome appeared to its corresponding case. A significant correlation was also observed between residence type and the ordination values ( $p=0.019$ ,  $R^2=0.113$ ).

Next, we used MaAsLin2 to model family (fixed effect) with health status, residence type, sex, and age as random effects to identify ARG classes and groups associated with specific households while controlling for demographic factors. At the class level, fosfomycin ARGs were significantly associated with family #14 ( $C=4.2$ ; adjusted- $p=1.1E-07$ ), while the group level analysis identified three ARGs to be associated with four different families. *fosA* and *acrB*, for example, were significantly associated with family #14 ( $C=4.3, 2.0$ ; adjusted- $p=1.1E-06, 9.2E-05$ , respectively), whereas *acrB* was associated with family #15 ( $C=1.3$ ; adjusted- $p=0.007$ ) and *mel* was linked to families #4 and #8 ( $C=5.6, 4.4$ ; adjusted- $p=1.8E-06, 5.7E-05$ , respectively).

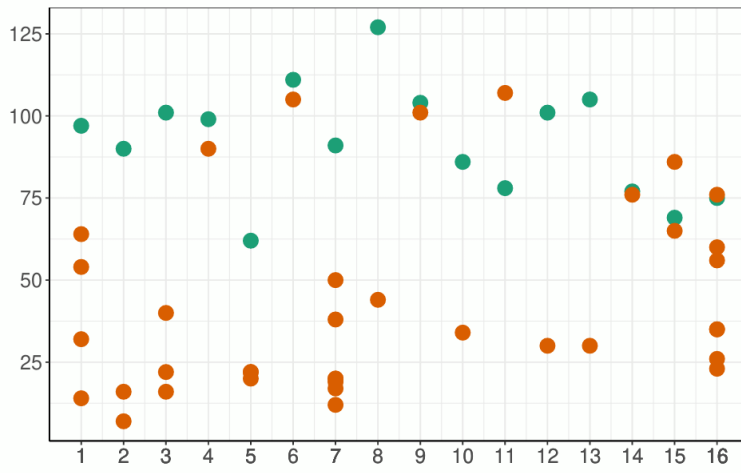
## DISCUSSION

Gastrointestinal dysbiosis has been shown to influence and be influenced by the gut microbiota (3, 38). Disease state as it pertains to dysbiosis not only impacts microbial taxa in the gut but can influence the functional composition of this environment as well (39). Herein, we

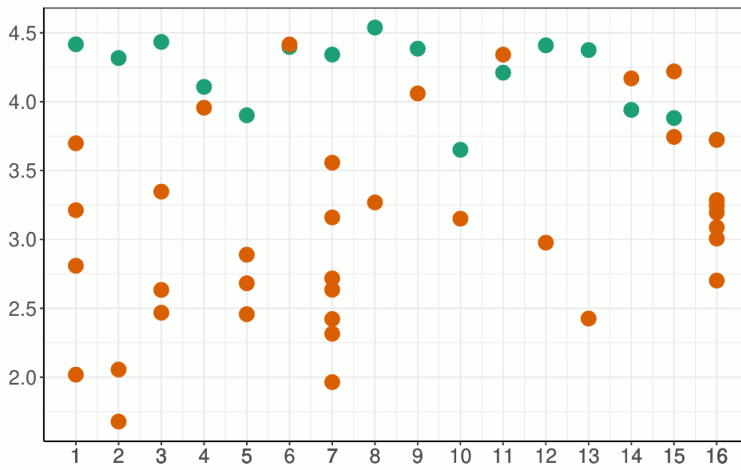
found that gut communities characterized from the stools of patients with *Campylobacter* infections (cases) had increased resistome diversity relative to healthy family members (controls). The differences observed between cases and controls in this metagenome analysis are consistent with our prior study which used 16S rRNA sequencing to demonstrate discrepancies in microbiota diversity between study groups (3). It is probable that fluctuations observed in the microbiome and resistome following enteric infection are linked, as changes in microbial composition will inherently shift the relative presence/absence of associated genes. Therefore, the role of enteric infection in driving these fluctuations is of great interest.

The identification of multiple differentially abundant ARG classes and groups in case samples suggests that *Campylobacter* infection influences the composition and diversity of the resistome. Most notable is the relative increase in MDR and fluoroquinolone resistance genes in case samples. *Campylobacter* strains can often harbor these genes (40), highlighting the possibility that pathogens can transport them into the gut community. Our taxa analysis of case samples, however, estimated the relative abundance of the *Campylobacter* genus to be low (0.28%), while genera such as *Escherichia* were much more abundant. Interestingly, however, many of the MDR ARGs detected among cases were correlated with other taxa such as *Shigella*; for example, *Shigella* was highly correlated with the MDR gene *mdtC*. Multidrug efflux is a common resistance mechanism and therefore carriage of *mdtC* by *Shigella* spp. is consistent with previous findings (41). Nonetheless, detection of *Shigella* also raises questions regarding co-infections, which require confirmation via culture or other diagnostic tests. It is also important to note that we did not directly explore genetic architecture, a technique which would more clearly elucidate which microbes harbor specific ARGs of interest. Despite providing preliminary information regarding ARG-harboring taxa, this method assumes that ARGs and taxa co-

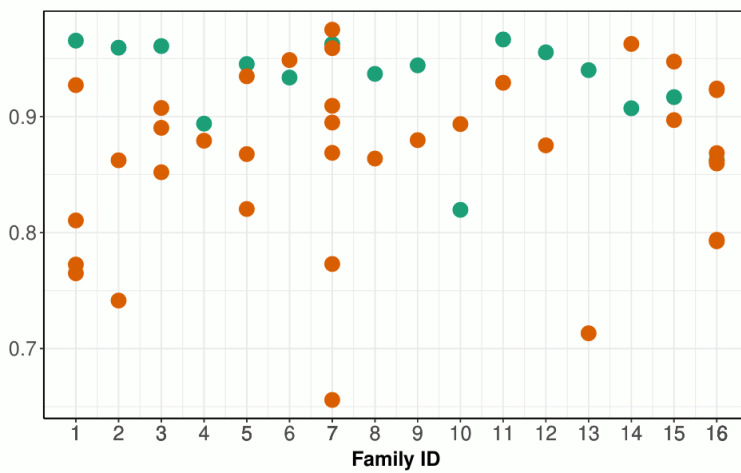
**ARG Richness**



**Shannon Diversity Index**



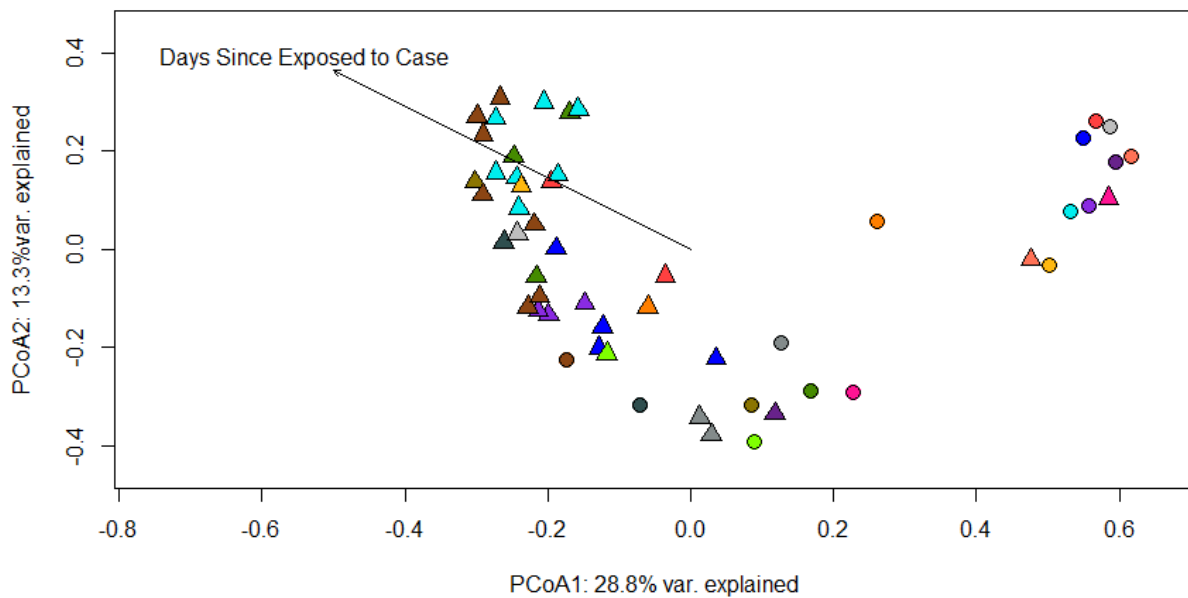
**Pielou's Evenness**



**Figure 2.8. Diversity among different families is not significantly different.**

**Figure 2.8 (cont'd)**

The alpha diversity measures of richness, evenness, and Shannon diversity do not significantly differ by family. Notably, however, there are differing levels of variance among families, particularly when comparing families with one control sample vs. many. Each boxplot represents a single family (i.e., one case sample and one or more control samples). Each sample is designated by a point on the plot; cases are represented by cyan circles while controls are designated by orange triangles. The median value for each family is depicted as a thick black line within the boxplot; the first and third quartiles are indicated by the lower and upper ranges of the box, respectively. P-values were determined using the Wilcoxon rank-sum test; significant p-values were not found and so are omitted from the plot.



**Figure 2.9. Beta diversity analyses do not reveal clear similarities among families.**

Principal coordinates analysis (PCoA) plot using Bray-Curtis dissimilarity of case and control samples shows clustering of 16 separate families. Case samples are depicted as circles and controls as triangles. Environmental factors with potential to influence this ordination were also examined; the variable exploring the number of days since exposure to the case was significantly correlated with the observed ordination. These data were fitted to the ordination using the ‘envfit’ function in R and displayed with a labeled arrow below. The number of days since being exposed to an ill family member is correlated to more “normal” looking controls (i.e., controls that are most dissimilar from their corresponding infected cases).

occurring in similar abundances indicate a correlation, potentially leading to inaccurate

associations. Indeed, an association between ARGs and their microbial hosts was observed in a

prior study using methods that measured both ARGs and taxa abundance (7). While fluctuations



in the abundance of specific taxa during infection likely change the abundance of ARGs harbored by these taxa, future work employing a more rigorous test assessing key taxa-associated ARGs is needed. It will also be important to examine these communities in a phylogenetic context using tools like UniFrac (42) when they become more readily available for use with metagenomics data.

It is notable that cases with *Campylobacter* infections had three distinct clusters with differentially abundant ARG profiles. Resistomes belonging to Cluster 2 were more similar to control resistomes, a finding that could point to less perturbed gut communities with a greater initial resilience or partially recovered communities at the time of sampling. Cluster 1 and Cluster 3 resistomes, however, were either more perturbed by infection or were distinct at the start of the infection. Indeed, it is possible that the trajectory of an individual's resistome during infection is contingent upon the microbiome composition before infection. Support for this possibility comes from a prior study of *Campylobacter* patients, who had significantly lower taxonomic diversity in their gut communities before infection (43), an outcome that could be related to varying levels of microbiome resilience (44). Indeed, studies in mice with varying degrees of microbial imbalance prior to infection demonstrated that disturbed gut communities were more susceptible to infection by *Salmonella enterica* serovar Typhimurium (45). Another study in chickens observed *Campylobacter* invasion of the cecal microbiome only after substantial changes to the metabolic profile were detected (46). Direct interactions between the normal gut microbiota and invading pathogens via resource competition, metabolite production, and direct antagonism, coupled with the complexity of pathogen-induced inflammation, have also been shown to influence enteric infections (47, 48). Variable perturbations among

individuals is consistent with prior studies showing distinct shifts in the gut microbiota and resistome following antibiotic treatment (49, 50).

Because we could not evaluate patient communities prior to infection onset, an approach that would require a costly and lengthy longitudinal study of healthy individuals, we cannot rule out the possibility that the sampled communities were already distinct. We have utilized a single sample taken during infection (cases) and during a self-described “healthy” period (controls) and therefore, an assessment of microbiome changes over time could not be performed. Longitudinal studies are needed to define the trajectory of microbiome fluctuations in the gut and determine how *Campylobacter* impacts these alterations. Another limitation is the assumption that stool is representative of the human gut microbiome. Previous work has shown that microbial signatures in the stool differ from other gut-related samples from the same individual (51, 52). Since the abundance of different bacterial populations differs in stool, our findings likely represent an underestimate of the actual abundance of taxa and ARGs within the gut. Future studies should also examine additional sample types, such as mucosal tissues, to better define how the gut microbiome is impacted by *Campylobacter*.

Specific factors responsible for observed differences between resistome profiles of cases remain elusive. It is possible, however, that geographic location as well as variable exposures and host responses play a role. For example, a prior study reported differences in ARG composition and abundance across land-use sites (rural, urban, and industrial), with ARGs fluctuating seasonally and in accordance with a relevant mobile genetic element (MGE), *int1* (53). Another study suggested that local anthropic activities, regardless of rural or urban identity, play a role in determining ARG profiles (54). These findings are consistent with our observation that urban residents were significantly more likely to have a resistome profile belonging to

Clusters 1 or 3 than Cluster 2, indicating that unique factors may be important for the expansion of specific ARGs during acute infection. Variation in host immune responses could offer another potential explanation for the observed differences among cases. *Campylobacter* infection elicits activation of NFK-B and the production of pro-inflammatory cytokines such as interleukin (IL)-8 (55), though cytokine responses can vary among strains (56). This variation may contribute to dissimilar levels of inflammation that differentially influence the resident gut microbiota and, at times, benefit the invading pathogen (47, 57, 58). Of importance, too, is that specific pathogen features such as the lipooligosaccharide and polysaccharide capsule can impact virulence and host responses (59, 60) while repeated exposure to *Campylobacter* has been linked to local and systemic inflammation in children (61). Host responses and *Campylobacter* strain characteristics were not evaluated in our study and, hence, we cannot rule out the possibility that some of these factors impacted resistome profiles. Future studies should therefore utilize a larger sample size to further clarify factors that contribute to more perturbed or variable resistomes.

Resistome profiles detected in control samples also provide important information about background levels of resistance due to the presence of specific ARGs and ARG classes. The finding that tetracycline, MLS, and beta-lactam ARGs were more abundant in controls is consistent with two global studies, though gene prevalence was somewhat impacted by geography (9, 62). In the United States, healthy individuals harbored MLS and beta-lactam resistance genes (63). While the reasons behind the increased abundance of these ARGs in healthy individuals is not clear, it is possible that historic circulation of these drugs in agriculture as well as veterinary and human health has had a long-term impact on the gut microbiome. Comparatively, historical use of tetracyclines in the medical field could also have long-term effects on gut microbes, which has been shown for group B *Streptococcus* (64), an opportunistic

pathogen that commonly resides in the gut. Because ARGs can be horizontally transferred to commensal gut bacteria and are stably maintained in this environment regardless of recent exposure or antibiotic use (65), detection in healthy control stools could be attributed to prior exposure to antibiotics or acquisition of antibiotic resistant bacteria. This suggestion is consistent with ARGs identified in other cohorts (9, 62, 63) and strengthens our assumption that uninfected control samples can be used as a baseline for comparison when analyzing resistomes following pathogen infection.

It is important to note that directly comparing case and control samples presented some challenges herein. First, control samples were obtained weeks after the related patients had recovered, which prevented an assessment of other factors that may be linked to resistome differences. Indeed, some of the observed variation between cases and controls could be due to factors such as diet and exercise level, which were not measured in this study but have been shown to influence gut communities. Secondly, the sample size of our cases and controls differed. Multiple controls were associated with a single case in some circumstances, while other cases lacked corresponding controls. Regardless of these limitations, however, we were provided with a novel opportunity to conduct a family-based analysis to explore how familial relations may influence the resistome.

Among all 16 families examined, family relation did not appear to outweigh the effect of disease state on the resistome as most cases had different profiles than the controls within each family. Variation in ARG distribution and abundance, however, was observed across families with four families having specific ARGs that were more likely to be shared among their family members. The mismatched number of controls per case, however, makes interpreting these data difficult, as more controls per case may have overestimated the importance of some ARGs.

Nonetheless, the analysis exploring environmental associations was notable. Specifically, time since exposure to the infected family member significantly influenced control resistomes; the longer the time period between a case being infected and a control submitting a stool sample, the less similar the resistomes are. Intuitively, this is expected since the longer the period following exposure to a case, the less likely a healthy family member will show signatures of potential infection/crossover.

It is also possible that the level of social closeness among family members played a role in the similarity of their resistomes. A prior study, for instance, noted that the closer the social interaction between two family members (such as between married partners), the more similar their gut microbiome compositions were (66). Unfortunately, we did not consistently receive information about the relational status of each control, and hence, conclusions about these relationships could not be made. In addition, due to the differing number of household members available per family as well as our hesitancy to exclude samples on a nonrandom basis, the uneven distribution of controls:cases limits our interpretation of these data. Regardless, the provision of multiple control samples enables us to observe similarities/differences between healthy members of a family in relation to each other and their infected relative, a tenet of this study that may prove useful in future analyses when considering how pathogens impact the gut microbiome.

Collectively, these data demonstrate that patients with *Campylobacter* infections have key differences in the human gut resistome relative to healthy, uninfected individuals. Of great interest, we observed an increase in specific taxa, the diversity of ARGs, and ARGs related to MDR in the patients. These findings substantiate the need for further characterizing the microbiome and resistome in response to perturbations such as those caused by enteric

pathogens. Future work should also involve examining bacterial genes found to be differentially abundant between groups or that possessed SNPs within genes linked to antibiotic resistance previously. Indeed, it is likely that periods of flux not only influence the composition of the microbiome, but also its capacity for horizontal gene transfer, which can play a role in the persistence and transmissibility of ARGs and emergence of resistant pathogens.

## **APPENDIX**

**Table A.1. Characteristics of 26 patients with *Campylobacter* infections (cases) and 44 healthy individuals (controls).**

Characteristic	Cases No (%)	Controls No (%)	<i>p</i> -value‡
<b>Demographics</b>			
Age			0.093
0-9 years	8 (30.7)	17 (38.7)	
10-18 years	0 (0.0)	4 (9.1)	
19-64 years	13 (50.1)	21 (47.7)	
≥65 years	5 (19.2)	2 (4.5)	
Sex			0.083
Male	9 (34.6)	26 (59.1)	
Female	17 (65.4)	18 (40.9)	
Residence			0.378
Rural	11 (44.0)	11 (29.7)	
Urban	14 (56.0)	26 (70.3)	

Note: Not all variables in each column added up to the total number of individuals because of missing data for some variables.

‡ *p*-values were calculated using the Chi-Square test or Fisher's exact test for variables with  $n \leq 5$  in at least one cell.



**Table A.2. Differentially abundant antimicrobial resistance genes (ARGs) detected in stool samples from cases and controls.**

<b>Group (gene)</b>	<b>Association</b>	<b>Coefficient</b>	<b>Standard Error</b>	<b>p-value</b>	<b>Adjusted p-value</b>	<b>ARG class</b>
<i>cpxAR</i>	Controls	-4.2583446	0.513936708	1.01E-11	5.17E-11	Multidrug resistance
<i>mdtC</i>	Controls	-3.8261348	0.439477581	1.15E-12	9.08E-12	Multidrug resistance
<i>parE</i>	Controls	-3.1825883	0.361985735	8.02E-13	6.98E-12	Fluoroquinolone resistance
<i>parC</i>	Controls	-3.014637	0.248540229	1.21E-18	1.06E-16	Fluoroquinolone resistance
<i>tetQ</i>	Controls	3.02600618	0.490406412	4.27E-08	8.65E-08	Tetracycline resistance
<i>cfx</i>	Controls	3.72226841	1.012164326	0.000466	0.000654	Class A beta-lactamase
<i>cbla</i>	Controls	4.17780963	0.686690753	6.06E-08	1.15E-07	Class A beta-lactamase
<i>tetW</i>	Controls	4.77784686	0.754448744	2.21E-08	5.49E-08	Tetracycline resistance

Gene groups identified with a coefficient  $\geq |3.0|$  using MaAsLin2 (Mallick et al. *bioRxiv* 2021, doi:10.1101/2021.01.20.427420) with health status (case vs. control) included as a fixed effect and residence type, age, and sex as random effects. Genes with negative coefficients are more abundant in cases, while positive coefficients are more abundant in control samples.

**Table A.3. Correlation values between highly abundant antimicrobial resistant genes (ARGs) and specific taxa detected in *Campylobacter* cases.**

<b>ARG</b>	<b>Target taxa</b>	<b>Correlation</b>	<b>P value</b>
<i>mdtC</i>	<i>Shigella</i>	0.885812	1.79E-09
<i>parE</i>	<i>Pseudoalteromonas</i>	0.881176	2.82E-09
<i>gyrA</i>	<i>Pseudoalteromonas</i>	0.873312	5.82E-09
<i>gyrB</i>	<i>Pseudoalteromonas</i>	0.870085	7.74E-09
<i>gyrA</i>	<i>Trabulsiella</i>	0.851064	3.6E-08
<i>parC</i>	<i>Pseudoalteromonas</i>	0.85094	3.63E-08
<i>cpxAR</i>	<i>Pseudoalteromonas</i>	0.839316	8.4E-08
<i>parE</i>	<i>Siccibacter</i>	0.82492	2.17E-07
<i>gyrA</i>	<i>Siccibacter</i>	0.818727	3.19E-07
<i>pare</i>	<i>Trabulsiella</i>	0.813223	4.43E-07
<i>gyrB</i>	<i>Kosakonia</i>	0.812522	4.61E-07
<i>gyrA</i>	<i>Phytobacter</i>	0.811464	4.91E-07
<i>gyrB</i>	<i>Phytobacter</i>	0.809267	5.57E-07
<i>parC</i>	<i>Siccibacter</i>	0.80311	7.88E-07
<i>gyrB</i>	<i>Siccibacter</i>	0.796919	1.1E-06
<i>rpoB</i>	<i>Trabulsiella</i>	0.796432	1.13E-06
<i>pare</i>	<i>Phytobacter</i>	0.796024	1.16E-06
<i>gyrB</i>	<i>Trabulsiella</i>	0.795543	1.19E-06
<i>cpxAR</i>	<i>Phytobacter</i>	0.793487	1.32E-06
<i>mdtC</i>	<i>Pseudoalteromonas</i>	0.789402	1.64E-06
<i>gyrA</i>	<i>Pluralibacter</i>	0.789195	1.66E-06
<i>gyrA</i>	<i>Klebsiella</i>	0.787485	1.81E-06
<i>mdtC</i>	<i>Rhodococcus</i>	0.785091	2.04E-06
<i>cpxAR</i>	<i>Siccibacter</i>	0.784193	2.13E-06
<i>rpoB</i>	<i>Siccibacter</i>	0.783009	2.26E-06
<i>parC</i>	<i>Phytobacter</i>	0.781823	2.4E-06
<i>parC</i>	<i>Trabulsiella</i>	0.775251	3.3E-06
<i>parE</i>	<i>Serratia</i>	0.773807	3.54E-06
<i>gyrB</i>	<i>Pluralibacter</i>	0.766154	5.05E-06
<i>gyrB</i>	<i>Pantoea</i>	0.76547	5.21E-06
<i>gyrA</i>	<i>Serratia</i>	0.763207	5.77E-06
<i>parC</i>	<i>Pseudodescherichia</i>	0.762122	6.06E-06
<i>parE</i>	<i>Pluralibacter</i>	0.761156	6.33E-06
<i>mdtC</i>	<i>Phytobacter</i>	0.755751	8.03E-06
<i>gyrB</i>	<i>Klebsiella</i>	0.755214	8.22E-06
<i>gyrA</i>	<i>Lelliottia</i>	0.755175	8.23E-06
<i>gyrA</i>	<i>Yersinia</i>	0.752949	9.06E-06
<i>parC</i>	<i>Pluralibacter</i>	0.751795	9.52E-06

**Table A.4. Differentially abundant antimicrobial resistance gene (ARG) classes detected in 25 cases with specific resistome profiles (clusters) determined by hierarchical clustering.**

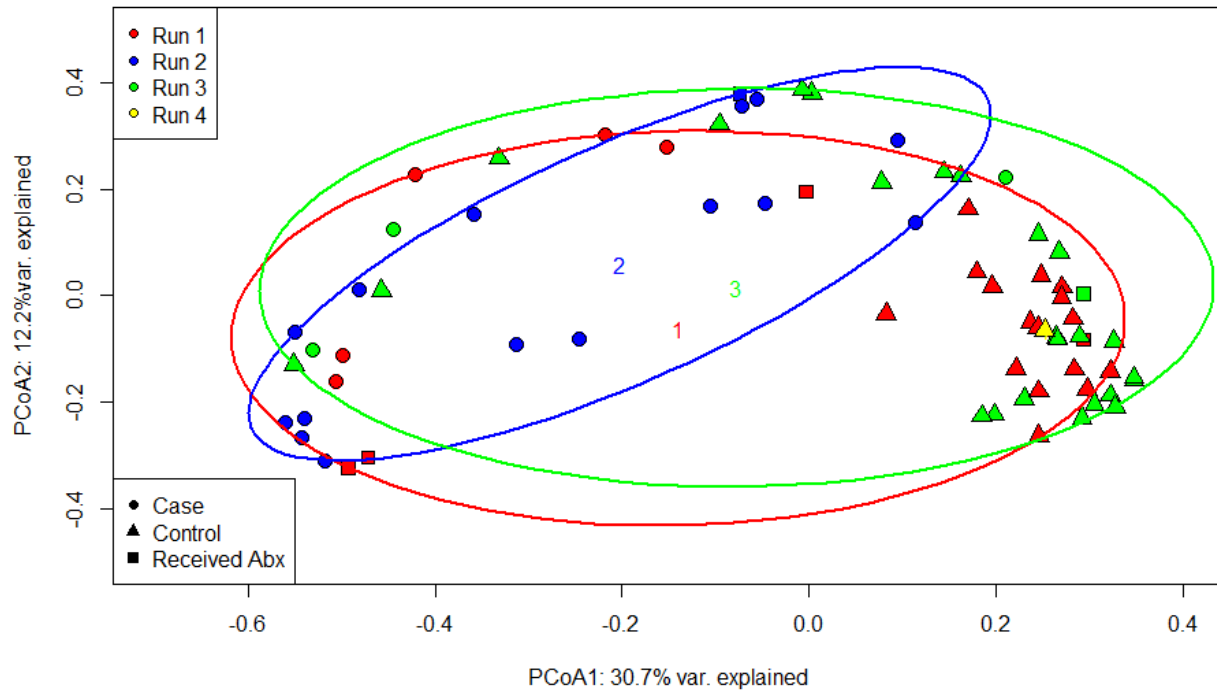
ARG Class	Association	Coefficient	Standard Error	p-value	Adjusted p-value
MLS	Cluster 2	5.276026	1.888096	0.011031	0.03564
Tetracyclines	Cluster 2	2.692487	1.027064	0.016182	0.048547
Multidrug resistance	Cluster 2	-0.45724	0.146904	0.005269	0.018441
CAP	Cluster 2	-0.48417	0.203713	0.027054	0.066838
Aminoglycosides	Cluster 2	-0.59723	0.172927	0.002378	0.009989
Fluoroquinolones	Cluster 2	-0.72234	0.206928	0.002273	0.009989
Aminocoumarins	Cluster 2	-0.98069	0.428388	0.032525	0.075891
Rifampin	Cluster 2	-1.03029	0.428546	0.025522	0.066838
Sulfonamides	Cluster 2	-1.59707	0.493373	0.004139	0.015803
Fosfomycin	Cluster 3	3.426063	0.691953	8.61E-05	0.001127
Aminocoumarins	Cluster 3	1.481023	0.404347	0.001451	0.009989
Elfamycins	Cluster 3	1.303181	0.274124	0.000107	0.001127
Fluoroquinolones	Cluster 3	0.682634	0.195489	0.002225	0.009989
Multidrug resistance	Cluster 3	-0.47969	0.13866	0.002346	0.009989
Bacitracin	Cluster 3	-0.93482	0.389323	0.026071	0.066838
Aminoglycosides	Cluster 3	-1.66939	0.163222	1.30E-09	2.74E-08
CAP	Cluster 3	-5.03738	0.19228	1.60E-17	6.72E-16
Beta-lactams	Urban	0.460124	0.233892	0.062505	0.135027
Trimethoprim	Urban	-3.5159	0.905585	0.000912	0.007657

ARG classes were identified using MaAsLin2 (Mallick et al. *bioRxiv* 2021, doi:10.1101/2021.01.20.427420) with case cluster and residence type as fixed effects and age and sex as random effects. Coefficients for the Cluster association were calculated using Cluster 1 as the reference groups, while the urban association used rural residence as the reference. A negative coefficient for Cluster 3, for instance, indicates that Cluster 1 is positively associated with a given class (e.g., aminoglycosides). Some classes were negatively associated with both Clusters 2 and 3 indicating a positive association with Cluster 1. CAP = cationic antimicrobial peptides

**Table A.5. Differentially abundant genes detected among cases living in urban versus rural settings.**

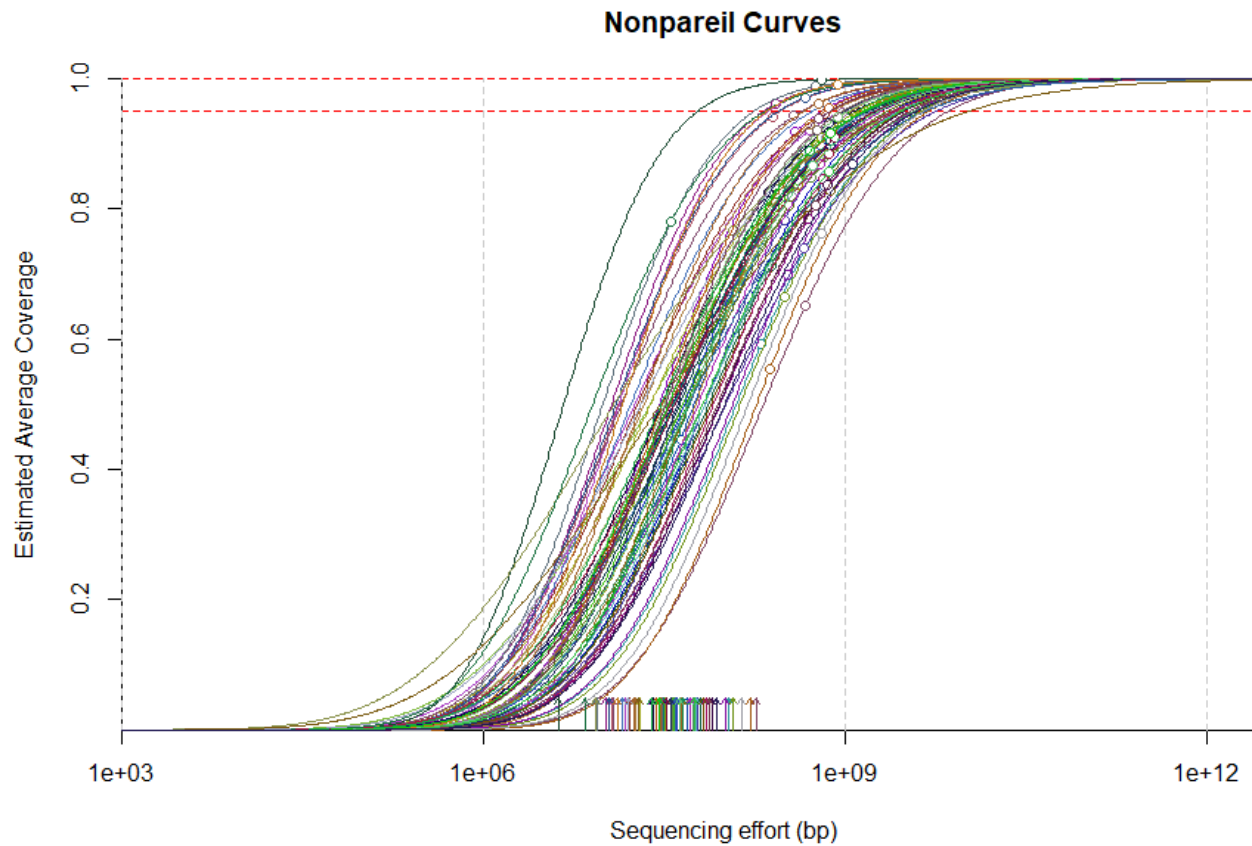
Group (gene)	Association	Coefficient	Standard Error	p-value	Adjusted p-value	ARG class
<i>pbp4B</i>	Urban	1.44670	0.632259	0.033228	0.110403	Beta-lactam resistance
<i>tetA</i>	Urban	-0.8197	0.29973	0.012413	0.049812	Tetracycline resistance
<i>tetB</i>	Urban	-0.8725	0.305991	0.010208	0.04321	Tetracycline resistance
<i>mphA</i>	Urban	-1.20645	0.341498	0.002378	0.013606	Macrolide resistance
<i>tetR</i>	Urban	-1.55327	0.63068	0.02351	0.084223	Tetracycline resistance
<i>dhfR</i>	Urban	-1.64111	0.425635	0.000917	0.006028	Trimethoprim resistance
<i>aac3</i>	Urban	-1.80729	0.517081	0.002157	0.012816	Aminoglycoside resistance
<i>ANT3-DPRIME</i>	Urban	-1.95451	0.627014	0.005213	0.025979	Aminoglycoside resistance

Gene groups identified using MaAsLin2 (Mallick et al. *bioRxiv* 2021, doi:10.1101/2021.01.20.427420) with Cluster and residence type included as fixed effects and age and sex as random effects. Rural residence is the reference group and hence, genes with negative coefficients are more abundant in rural cases, while positive coefficients are more abundant in urban cases.



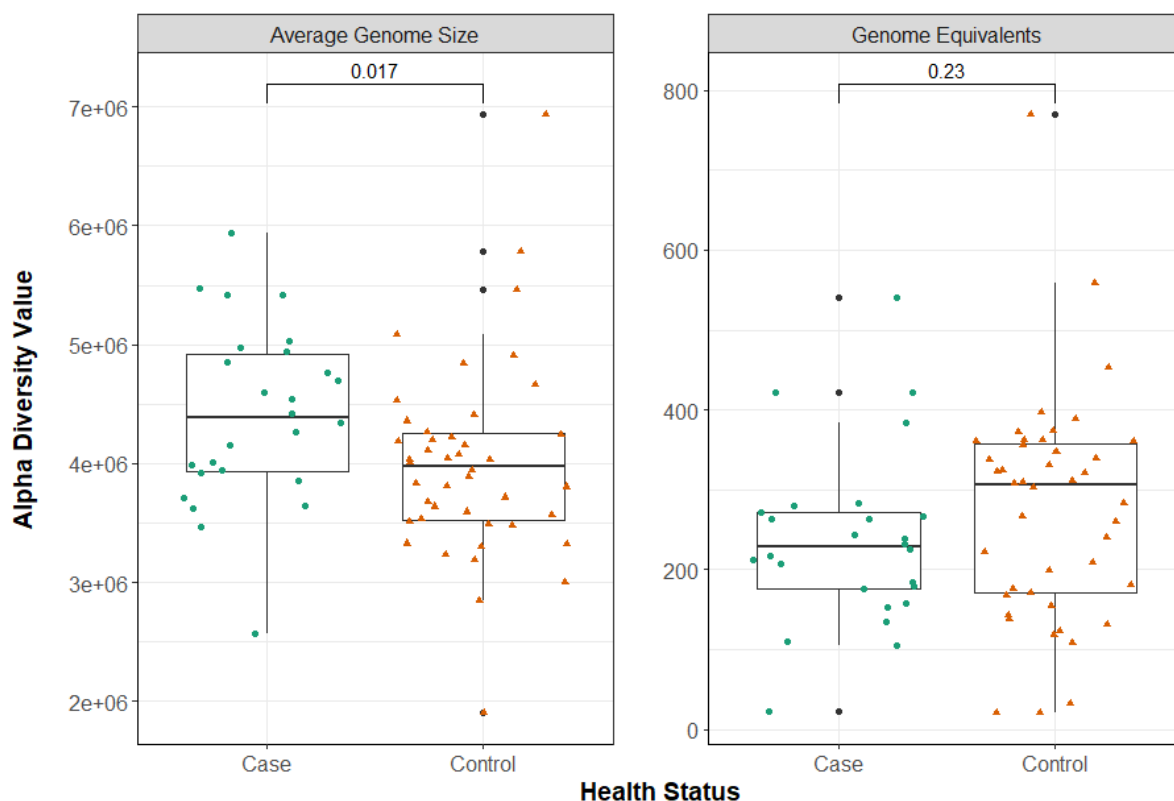
**Figure A.1. Sequencing run does not appear to impact resistome similarity among cases and controls.**

A Principal Coordinates Analysis (PCoA) plot of case (circles) and control (triangles) resistomes based on Bray-Curtis dissimilarity at the ARG gene level. The first and second coordinate are shown with their respective percentage of explained variance. Sequencing run is denoted by color: Red=Run 1; Blue=Run 2; Green=Run 3; Yellow=Run 4, while patients reporting use of antibiotics are indicated by square data points. Notably, there is considerable overlap among all four sequencing runs. Although a test indicated that the centroids of each run were different (PERMANOVA  $p=0.000999$ ;  $F=3.3029$ ) as well as the dispersion of points within each run (PERMDISP  $p=0.001$ ;  $F=10.152$ ), this result is attributed to the unequal sample sizes across runs. Run 4, for instance, contains just one sample, whereas Runs 1-3 contain 25, 16, and 28 samples, respectively. Therefore, the difference in centroid and dispersion is expected.



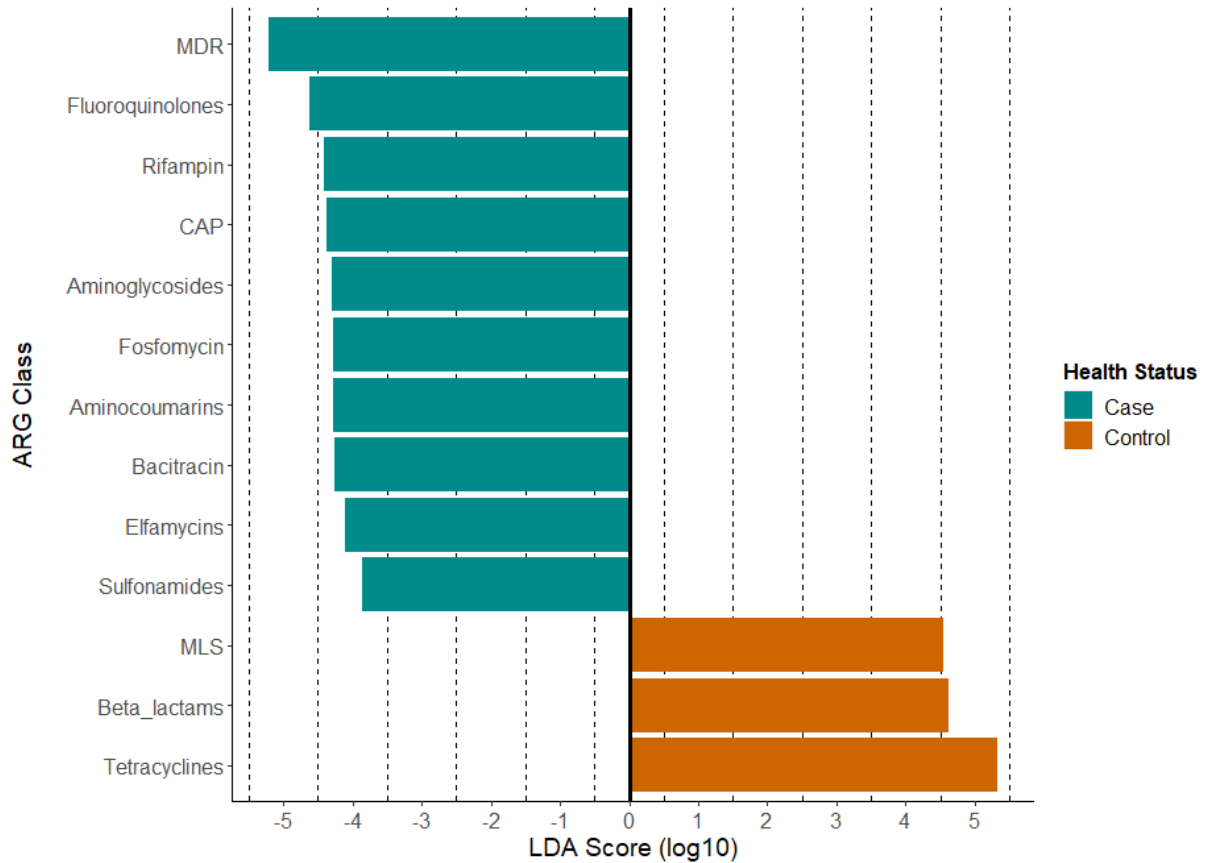
**Figure A.2. Estimated sequencing coverage curves for cases and controls.**

The estimated coverage (S-curves) and actual coverage (open circles) for case (n=26) and control (n=44) samples evaluated in this study. Each colored S-curve represents a single sample. Arrows at the bottom of the graph represent the Nonpareil index of sequence diversity, which is a measure of community complexity in sequence space; the mean Nonpareil diversity was 17.32 consistent with other stool samples documented with this tool. Dotted red lines represent 100% coverage and 95% coverage, respectively. The overall mean coverage for cases and controls was 83.0%.



**Figure A.3. Comparing the average genome size and number of genome equivalents among case and control samples.**

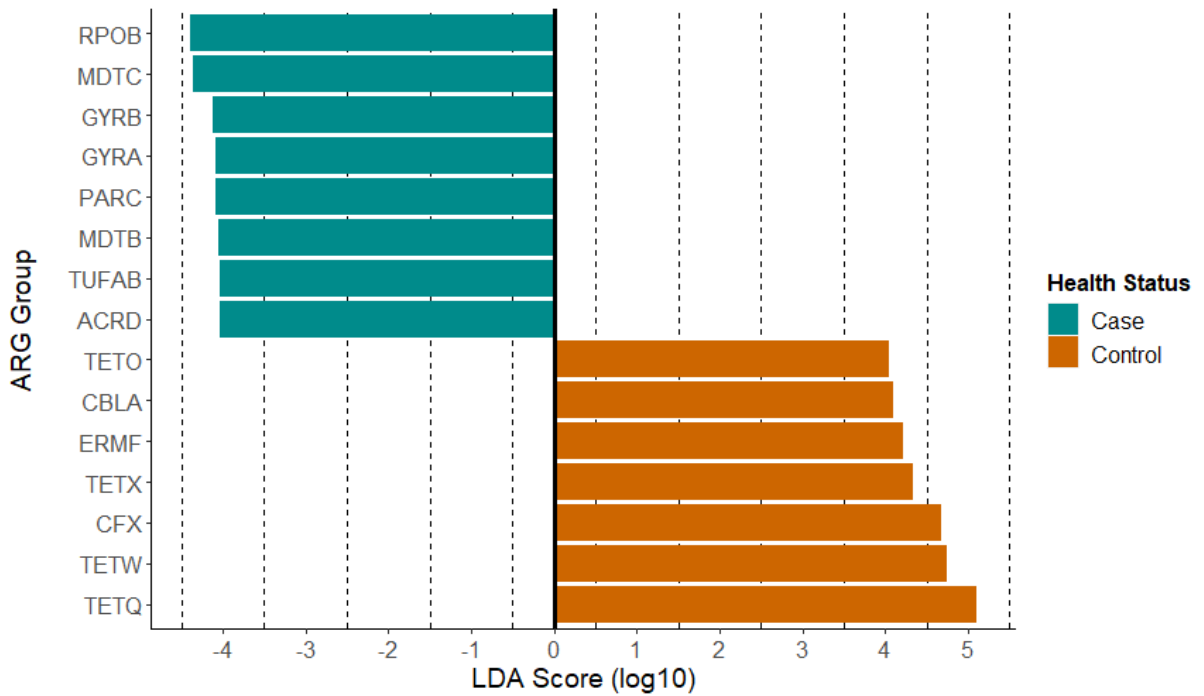
The median of each measure is shown by the black horizontal bar in each box. The first and third quartiles are indicated by the bottom and top of each box, respectively. Points (circles and triangles) are displayed to show variation within the sample types. Outliers within each group are indicated by the black dots. P-values comparing the difference between cases and controls were calculated using a Wilcoxon rank sum test and are shown above the comparison bar for each metric. Cases = cyan; controls = orange.



**Figure A.4. Linear discriminant analysis (LDA) scores showing differentially abundant antimicrobial resistance gene (ARG) classes by health status.**

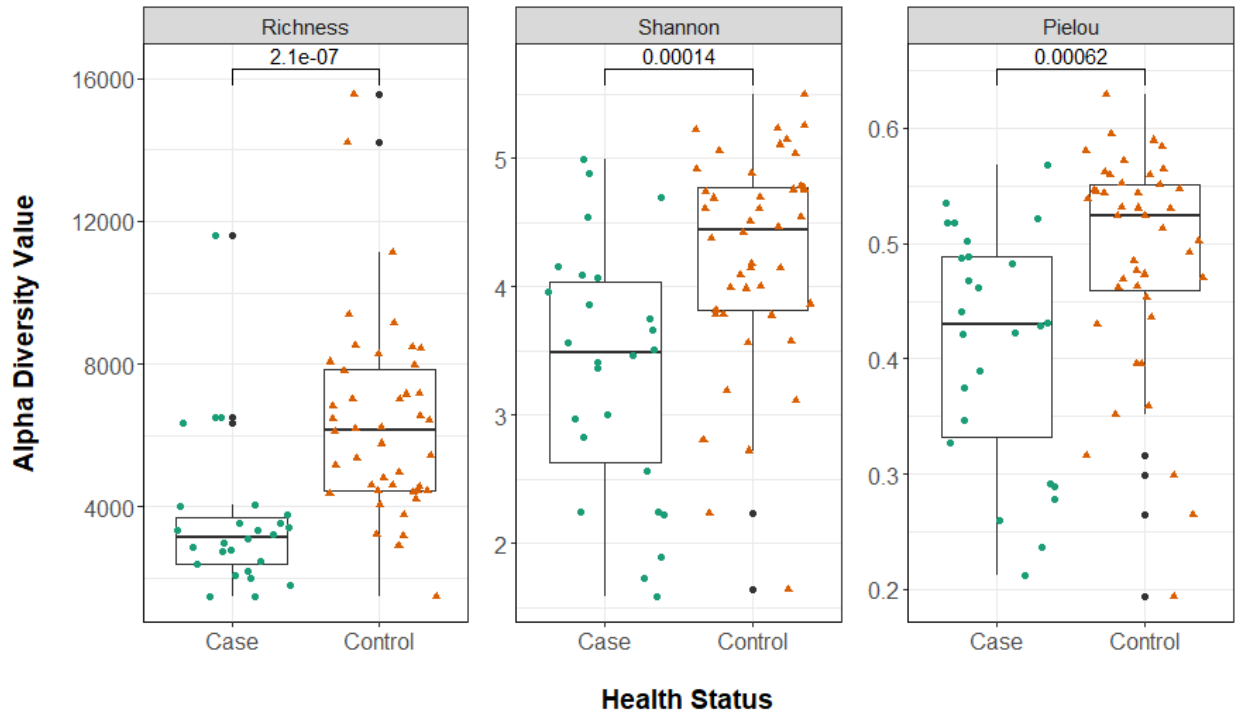
The classes shown registered an LDA score  $>2.0$ . The bars shown in orange indicate ARG classes that were more abundant in controls, while green bars show ARG classes that were more abundant in cases. In controls, tetracycline ARGs had the greatest LDA score (5.3;  $p=8.34e-10$ ) followed by beta-lactam and the Macrolide, Lincosamide, Streptogramin (MLS) ARG classes (LDA=4.6, 4.6;  $p=0.002$ ,  $0.0002$ , respectively). Ten classes were more abundant in cases, with MDR (LDA=5.2;  $p=1.68e-09$ ), fluoroquinolones (LDA=4.6;  $p=8.18e-11$ ), and rifampin ARGs (LDA=4.4;  $p=3.07e-10$ ) having the highest scores. CAP = cationic antimicrobial peptides; MDR = multidrug resistance.





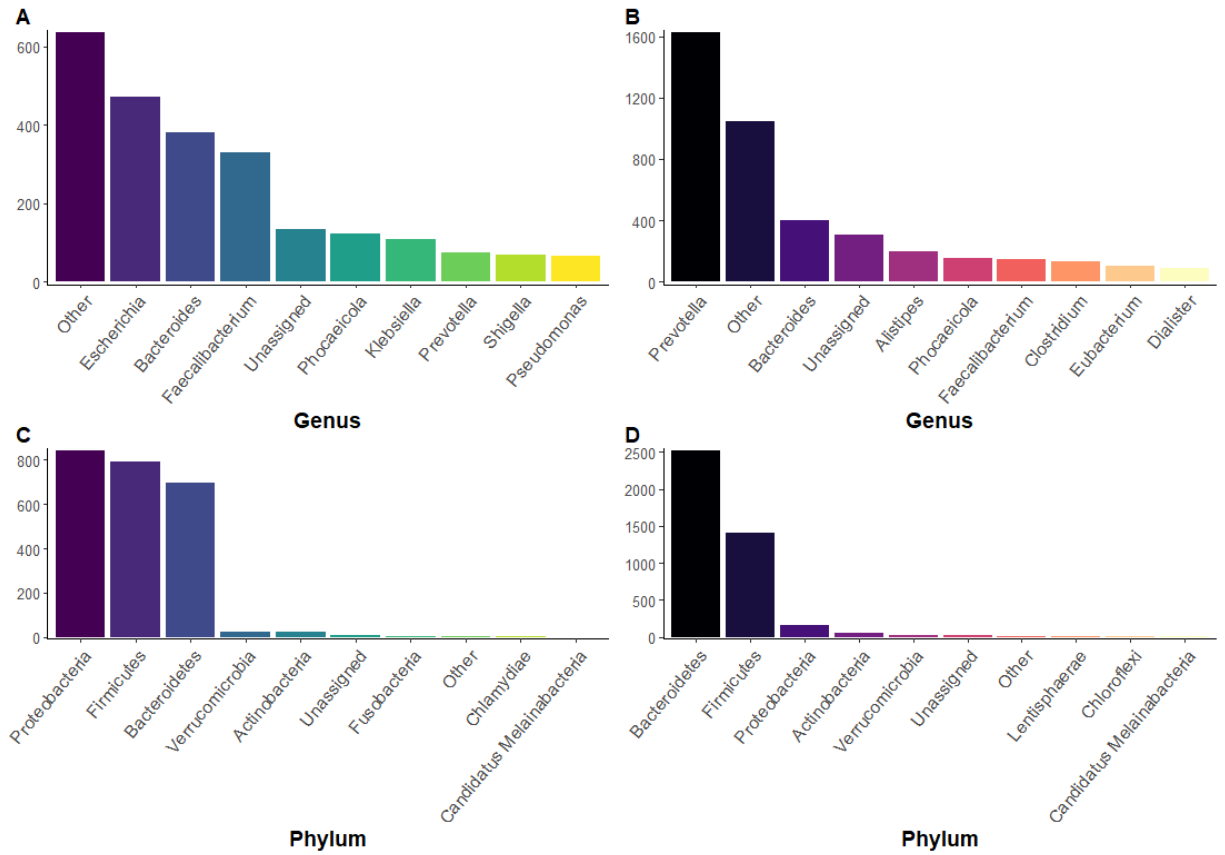
**Figure A.5. Linear discriminant analysis (LDA) scores for differentially abundant antimicrobial resistance genes (ARGs) at the group (gene) level by health status.**

Each ARG gene included in this plot registered an LDA score  $>4.0$ . The orange bars show ARG genes that were more abundant in controls, whereas green bars show genes that were more abundant in cases. In all, 93 of 153 features were differentially abundant between cases and controls. Of these, 12 were more abundant in controls with *tetQ*, *tetW*, and *cfx* predominating, while 81 were more abundant in cases; *rpoB*, *mdtC* and DNA gyrase genes, *gyrB* and *gyrA* predominated in the latter.



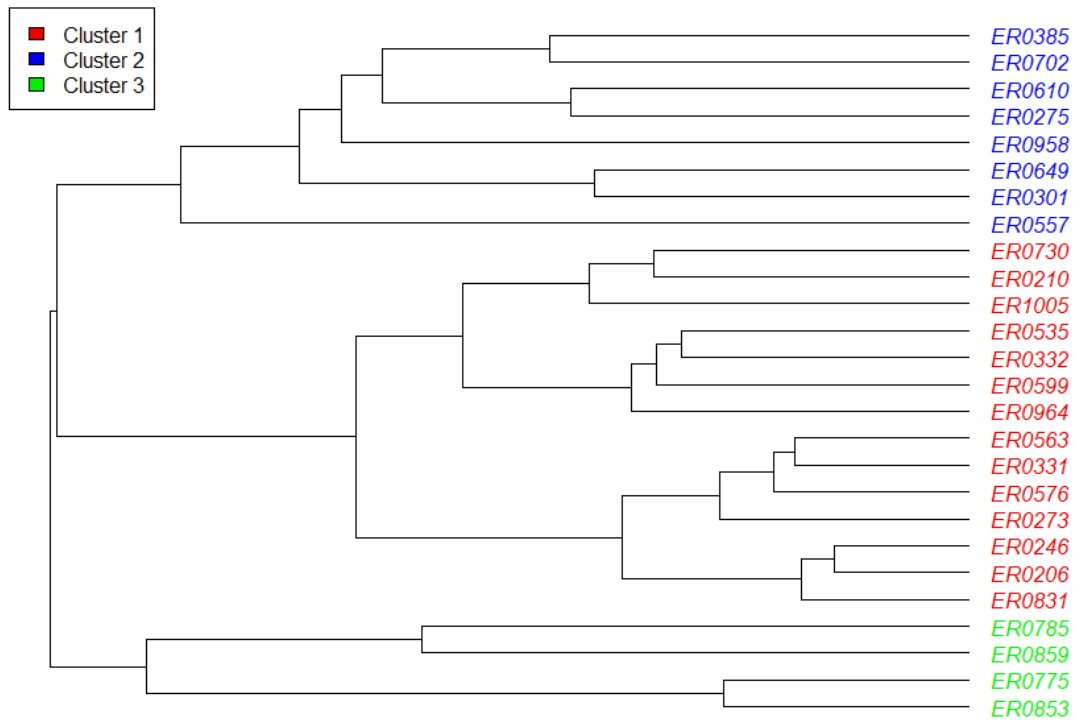
**Figure A.6. Controls display higher taxonomic diversity than cases.**

Three measures of alpha diversity (Richness, Shannon diversity, and Pielou's Evenness, respectively) are shown for microbial taxonomy among samples. The median of each measure is indicated by the thick black bar in each box and the first and third quartiles are represented by the bottom and top of the box, respectively; jittered points (circles and triangles) show variation within each sample type. Outlying points within each group are indicated by the black dots associated with each boxplot. P-values were calculated using the Wilcoxon rank-sum test and are shown above the comparison bar within each plot.



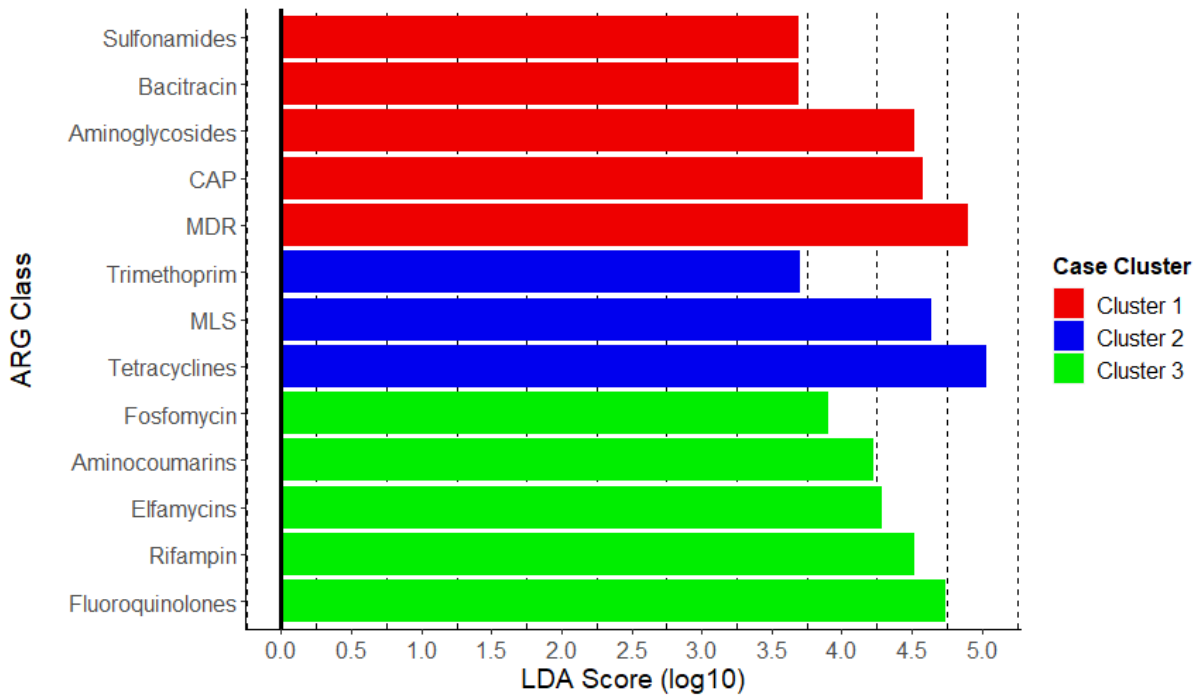
**Figure A.7. Actual abundances of bacterial taxa differ considerably between cases and controls.**

Rank abundance plots display the average number of reads assigned to bacterial genera and phyla for cases (A, C) and controls (B, D) in decreasing order. The top-10 genera and phyla were determined using the highest average number of reads assigned among cases or controls. All remaining genera or phyla were combined and summed to comprise the group “Other”, shown in the plots below. Note: the y-axis has different scales in each abundance plot.



**Figure A.8. Hierarchical clustering reveals three distinct resistome profiles among the cases.**

Average linkage hierarchical clustering at the gene level was performed based on the Bray-Curtis dissimilarity. Two primary clusters, Cluster 1 and Cluster 2, were identified as well as one outgroup (Cluster 3). Case sample numbers are indicated and colored based on the resistome cluster.



**Figure A.9. Linear discriminant analysis (LDA) showing differentially abundant antimicrobial resistance gene (ARG) classes between case clusters.**

The classes shown here each registered an LDA score >2.0. The bars shown in red indicate ARG classes that were more abundant in case Cluster 1; blue bars show ARG classes that were more abundant in case Cluster 2; green bars indicate ARG classes more abundant in Cluster 3. MLS = Macrolide, Lincosamide, Streptogramin; CAP = cationic antimicrobial peptides; MDR = multidrug resistance.



**Figure A.10. Relative abundance of ARG classes varies across families but maintains the case versus control dichotomy in most circumstances.**

The relative abundance of ARGs assigned to 18 different antibiotic classes is shown with each column representing the resistome from one individual. Each set of numbered plots is faceted by family ID with the left-most column representing the infected individual (cases) in each family; the remaining columns in a family represent 1-7 healthy controls. Relative abundances were determined using raw ARG abundances normalized by the approximate number of genome equivalents in the sample. CAP = cationic antimicrobial peptides; MLS = Macrolide, Lincosamide, Streptogramin; MDR = Multidrug resistance.

## **REFERENCES**

## REFERENCES

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne Illness Acquired in the United States -- Major Pathogens. *Emerging Infectious Disease* 17:7-15.
2. Tack DM, Marder EP, Griffin PM, Cieslak PR, Dunn J, Hurd S, Scallan E, Lathrop S, Muse A, Ryan P, Smith K, Tobin-D'Angelo M, Vugia DJ, Holt KG, Wolpert BJ, Tauxe R, Geissler AL. 2019. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2015–2018. *MMWR Morbidity and Mortality Weekly Report* 68:369-373.
3. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* 3:45-45.
4. CDC. 2019. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA.
5. Cosgrove SE. 2006. The Relationship between Antimicrobial Resistance and Patient Outcomes: Mortality, Length of Hospital Stay, and Health Care Costs. *Clinical Infectious Diseases* 42:S82-S89.
6. Wright GD. 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology* 5:175-186.
7. Ma L, Li B, Jiang X-T, Wang Y-L, Xia Y, Li A-D, Zhang T. 2017. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* 5:154-154.
8. Surette MD, Wright GD. 2017. Lessons from the Environmental Antibiotic Resistome. *Annual Review of Microbiology* 71:309-329.
9. Feng J, Li B, Jiang X, Yang Y, Wells GF, Zhang T, Li X. 2018. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environmental Microbiology* 20:355-368.
10. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* 337:1107-1111.
11. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, Calderon M, Cabrera L, Hoyos-Arango W, Teresita Bertoli M, Berg DE, Gilman RH, Dantas G, Salvador E. 2016. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533:212-216.



12. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* 489:220-230.
13. Antonopoulos DA, Huse SM, Morrison HG, Schmidt TM, Sogin ML, Young VB. 2009. Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infection and Immunity*.
14. Dethlefsen L, Huse S, Sogin ML, Relman DA. 2008. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biology* 6:e280.
15. Yassour M, Vatanen T, Siljander H, Hämäläinen A-m. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatments on strain-level diversity and stability. 8.
16. Mallon CA, Van Elsas JD, Salles JF. 2015. Microbial invasions: The process, patterns, and mechanisms. *Trends in Microbiology* 23:719-729.
17. Van Der Putten WH, Klironomos JN, Wardle DA. 2007. Microbial ecology of biological invasions. *ISME Journal* 1:28-37.
18. Litchman E. 2010. Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecology Letters* 13:1560-1572.
19. Singh P, Manning SD. 2016. Impact of age and sex on the composition and abundance of the intestinal microbiota in individuals with and without enteric infections. *Annals of Epidemiology* 26:380-385.
20. Ingram DD, Franco SJ. 2014. 2013 NCHS Urban-rural Classification Scheme for Counties, 166 ed, vol Stat 2. National Center for Health Statistics.
21. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Research* 45:D574-D580.
22. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
23. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25:2078-9.
25. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *BIOINFORMATICS APPLICATIONS NOTE* 26:841-842.

26. Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* 30:629-635.
27. Nayfach S, Pollard KS. 2015. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biology* 16:51-51.
28. Beszteri B, Temperton B, Frickenhaus S, Giovannoni SJ. 2010. Average genome size: a potential source of bias in comparative metagenomics. *The ISME Journal* 4:1075-1077.
29. Walter J, Ley R. 2011. The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. *Annual Review of Microbiology* 65:411-429.
30. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research* 48:D561-D569.
31. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy* 57:3348-3357.
32. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7:11257-11257.
33. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Maintainer HW. 2019. Package 'vegan' Title Community Ecology Package. *Community ecology package* 2.
34. The R Foundation. *R: The R Project for Statistical Computing*.
35. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Tickle TL, Weingart G, Ren B, Schwager EH, Thompson KN, Wilkinson JE, Subramanian A, Lu Y, Paulson JN, Franzosa EA, Corrada Bravo H, Huttenhower C. 2021. Multivariable Association Discovery in Population-scale Meta-omics Studies 3. *bioRxiv:2021.01.20.427420-2021.01.20.427420*.
36. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526-528.
37. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biology* 12.

38. Lynch SV, Pedersen O. 2016. The human intestinal microbiome in health and disease. *New England Journal of Medicine* 375:2369-2379.
39. Armour CR, Nayfach S, Pollard KS, Sharpston TJ. 2019. A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome.
40. Cha W, Mosci R, Wengert SL, Singh P, Newton DW, Salimnia H, Lephart P, Khalife W, Mansfield LS, Rudrik JT, Manning SD. 2016. Antimicrobial susceptibility profiles of human *Campylobacter jejuni* isolates and association with phylogenetic lineages. *Frontiers in Microbiology* 7:589-589.
41. Ranjbar R, Farahani A. 2019. *Shigella*: Antibiotic-Resistance Mechanisms And New Horizons For Treatment.
42. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2010. UniFrac: an effective distance metric for microbial community comparison. *ISME Journal* 5:169-172.
43. Kampmann C, Dicksved J, Engstrand L, Rautelin H. 2016. Composition of human faecal microbiota in resistance to *Campylobacter* infection. *Clinical Microbiology and Infection* 22:61.e1-61.e8.
44. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. 2017. The resilience of the intestinal microbiota influences health and disease. *Nature Reviews Microbiology* 15:630-638.
45. Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, Brett Finlay B. 2008. Antibiotic-Induced Perturbations of the Intestinal Microbiota Alter Host Susceptibility to Enteric Infection. *INFECTION AND IMMUNITY* 76:4726-4736.
46. Ijaz UZ, Sivaloganathan L, McKenna A, Richmond A, Kelly C, Linton M, Stratakos AC, Lavery U, Elmi A, Wren BW, Dorrell N, Corcionivoschi N, Gundogdu O. 2018. Comprehensive Longitudinal Microbiome Analysis of the Chicken Cecum Reveals a Shift From Competitive to Environmental Drivers and a Window of Opportunity for *Campylobacter*. *Frontiers in Microbiology* 9.
47. Barman M, Unold D, Shifley K, Amir E, Hung K, Bos N, Salzman N. 2008. Enteric Salmonellosis Disrupts the Microbial Ecology of the Murine Gastrointestinal Tract. *Infection and Immunity* 76:907-915.
48. Omurwa Masanta W, Heimesaat MM, Bereswill S, Tareen AM, Lugert R, Groß U, Zautner AE. 2013. Modification of Intestinal Microbiota and Its Consequences for Innate Immune Response in the Pathogenesis of *Campylobacteriosis*. *Clinical and Developmental Immunology* 2013.
49. Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PNAS* 108:4554-4561.

50. Duan Y, Chen Z, Tan L, Wang X, Xue Y, Wang S, Wang Q, Das R, Lin H, Hou J, Li L, Mao D, Luo Y. 2020. Gut resistomes, microbiota and antibiotic residues in Chinese patients undergoing antibiotic administration and healthy individuals. *Science of the Total Environment* 705:135674-135674.
51. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans ADL, de Vos WM. 2002. Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Applied and Environmental Microbiology* 68:3401-3407.
52. Carroll IM, Chang Y-H, Park J, Sartor RB, Ringel Y. 2010. Luminal and mucosal-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Gut Pathogens* 2:19.
53. Xie J, Jin L, Luo X, Zhao Z, Li X. 2018. Seasonal Disparities in Airborne Bacteria and Associated Antibiotic Resistance Genes in PM 2.5 between Urban and Rural Sites. *Environmental Science & Technology Letters* 5:74-79.
54. Szekeres E, Chiriac CM, Baricz A, Sz T, Oke N, Lung I, Soran M-L, Rudi K, Dragos N, Coman C. 2018. Investigating antibiotics, antibiotic resistance genes, and microbial contaminants in groundwater in relation to the proximity of urban areas \*. *236:734-744*.
55. Young KT, Davis LM, DiRita VJ. 2007. *Campylobacter jejuni*: Molecular biology and pathogenesis. *Nature Reviews Microbiology* 5:665-679.
56. John DA, Williams LK, Kanamarlapudi V, Humphrey TJ, Wilkinson TS. 2017. The Bacterial Species *Campylobacter jejuni* Induce Diverse Innate Immune Responses in Human and Avian Intestinal Epithelial Cells. *Frontiers in Microbiology* 8:1840-1840.
57. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host & Microbe* 2:119-129.
58. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, Chaffron S, Macpherson AJ, Buer J, Parkhill J, Dougan G, Von Mering C, Hardt W-D. 2007. *Salmonella enterica* Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. *PLoS Biology* 5.
59. Louwen R, Heikema A, van Belkum A, Ott A, Gilbert M, Ang W, Endtz HP, Bergman MP, Nieuwenhuis EE. 2008. The Sialylated Lipooligosaccharide Outer Core in *Campylobacter jejuni* Is an Important Determinant for Epithelial Cell Invasion. *Infection and Immunity* 76:4431-4438.
60. Maue AC, Mohawk KL, Giles DK, Poly F, Ewing CP, Jiao Y, Lee G, Ma Z, Monteiro MA, Hill CL, Ferderber JS, Porter CK, Trent MS, Guerry P. 2013. The Polysaccharide Capsule of *Campylobacter jejuni* Modulates the Host Immune Response. *Infection and Immunity* 81:665-672.

61. Amour C, Gratz J, Mduma E, Svensen E, Rogawski ET, McGrath M, Seidman JC, J McCormick BJ, Shrestha S, Samie A, Mahfuz M, Qureshi S, Hotwani A, Babji S, Rengifo Trigos D, M Lima AA, Bodhidatta L, Bessong P, Ahmed T, Shakoor S, Kang G, Kosek M, Guerrant RL, Lang D, Gottlieb M, Houpt ER, Platts-Mills JA. 2016. Epidemiology and Impact of *Campylobacter* Infection in Children in 8 Low-Resource Settings: Results From the MAL-ED Study. *Clinical Infectious Diseases* 63:1171-1179.
62. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, Pan Y, Li J, Zhu L, Wang X, Meng Z, Zhao F, Liu D, Ma J, Qin N, Xiang C, Xiao Y, Li L, Yang H, Wang J, Yang R, Gao GF, Wang J, Zhu B. 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications* 4.
63. Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A, Bork P. 2013. Country-specific antibiotic use practices impact the human gut resistome. *Genome Research* 23:1163-1169.
64. Da Cunha V, Davies MR, Douarre P-E, Rosinski-Chupin I, Margarit I, Spinali S, Perkins T, Lechat P, Dmytruk N, Sauvage E, Ma L, Romi B, Tichit M, Lopez-Sanchez M-J, Descorps-Declere S, Souche E, Buchrieser C, Trieu-Cuot P, Moszer I, Clermont D, Maione D, Bouchier C, McMillan DJ, Parkhill J, Telford JL, Dougan G, Walker MJ, Holden MTG, Poyart C, Glaser P. 2014. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nature Communications* 5.
65. Salyers AA, Gupta A, Wang Y. 2004. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology* 12:412-416.
66. Dill-McFarland KA, Tang ZZ, Kemis JH, Kerby RL, Chen G, Palloni A, Sorenson T, Rey FE, Herd P. 2019. Close social relationships correlate with human gut microbiota composition. *Scientific Reports* 9:1-10.

### **CHAPTER 3**

Exploring recovery of the gut microbiome following enteric infection and the persistence of resistance genes in specific microbial hosts

## ABSTRACT

Enteric pathogens cause widespread foodborne illness and are increasingly found to harbor antimicrobial resistance. The ecological impact of these pathogens on the human gut microbiome and resistome, however, has yet to be fully elucidated. This study pursued shotgun metagenome analyses on stools collected from 60 patients during (cases) and after (follow-ups) infection caused by *Campylobacter*, *Salmonella*, *Shigella*, or Shiga toxin-producing *E. coli* (STEC). Overall, cases harbored more antimicrobial resistance genes (ARGs) and had greater resistome diversity than follow-ups ( $p < 0.001$ ). Conversely, follow-ups had much more diverse microbiomes ( $p < 0.001$ ). While cases were primarily defined by genera in Proteobacteria such as *Escherichia*, *Salmonella*, and *Shigella* and ARGs relevant to multi-compound and multi-drug resistance, follow-ups had much greater abundance of notoriously beneficial bacteria in the Bacteroidetes and Firmicutes phyla, with ARGs for tetracycline, MLS, and aminoglycoside resistance. Correlation networks were constructed to predict relevant ARG-taxa associations; these hypotheses were followed by a host-tracking analysis designed to investigate whether various ARGs were indeed affiliated with certain taxa. Host-tracking revealed that *Escherichia* was the primary carrier of ARGs in both cases and follow-ups, with greater abundance demonstrated during infection. Patterns relevant to extended spectrum beta-lactamases (ESBLs) and other clinically relevant beta-lactam resistance genes were also investigated, with findings suggesting the potential for transmission among microbes within the gut. Considered together, these data highlight the importance of further studying and understanding the impacts of enteric infection on human gut ecology, specifically as it pertains to antimicrobial resistance.

## INTRODUCTION

Foodborne illness caused by enteric pathogens impacts approximately 9.4 million people annually in the United States, with over one-third of these infections being bacterial in nature (1). In 2019, the Centers for Disease Control and Prevention (CDC) documented the marked increase in incidence of foodborne infection among various pathogens including *Campylobacter* and Shiga toxin-producing *Escherichia coli* (STEC), with the incidence of *Salmonella* and *Shigella* infections remaining relatively high but unchanged based on previous years (2). The consequences of enteric infection on the overall health of the human gut microbiome continues to be elucidated. Previously, studies conducted in our lab displayed a marked decrease in overall microbiome diversity attributed to enteric infection (3) as well as notable shifts in the gut resistome, or the compilation of antimicrobial resistance genes (ARGs) (4), of infected patients compared to healthy family members (5). Additional studies have demonstrated an increase in the proportion of Proteobacteria upon infection by *Salmonella*, *Campylobacter*, and *Shigella*, among other pathogens in a range of host organisms (6-9). However, the potential ecological repercussions relevant to recovery from enteric infection have yet to be explored.

In addition to their roles in enteric disease, four bacterial culprits come to the forefront at the intersection of foodborne illness and antimicrobial resistance. In their 2019 AMR Report, the CDC classified *Campylobacter*, non-Typhoidal *Salmonella*, *Shigella*, and various members of Enterobacteriaceae (which includes *Escherichia*) as serious threats for harboring and transmitting antimicrobial resistance (10). Moreover, each of these pathogens has demonstrated capability for transmission of ARGs via horizontal gene transfer (HGT) not only intra-specifically but inter-specifically as well (11). Transmission of antimicrobial resistance does not occur in a closed system, however. Rather, these genetic elements have been shown to cross environmental boundaries (12, 13). The increasing incidence of disease caused by these enteric pathogens,



coupled with their evolving role in harboring antimicrobial resistance and relatively high ability to transfer ARGs across communities justifies further examination of the repercussions of enteric infection on the human gut microbial community.

Just as we consider the microbial signatures of waterways, soils, and plants through an ecological lens, so too can we perceive the human gut microbiome. Of primary relevance, here, is our cogitation of the ecological consequences of enteric pathogens unwelcomely invading a healthy gut environment. Previous work has shown that enteric infection can result in decreased diversity of the microbiome, a state which can result in reduction of beneficial microbially-driven metabolism and potential increase in gut inflammation (14). In addition, we have previously documented differences in the composition of the resistome in cases with *Campylobacter* infections relative to healthy family members (5). If the microbiome demonstrates a certain degree of resilience, these impairments may not be felt with such amplitude and are typically resolved over time (15). Multiple studies, for example, have evinced the human microbiome's trajectory of recovery following administration of antibiotics, a known disruptor of gut microbial community homeostasis (16-18). In the context of pathogen invasion, various ecological interactions must be considered including direct antagonism from commensal microbes, resource competition and competitive exclusion, as well as secondary metabolite production (19-22); each of these factors may influence the success of an enteric pathogen in the gut environment as well as the ability of the human host to recovery from infection.

These interactions among gut microbiota comprise just one ecological facet regarding these pathogens' introduction to the gut environment. Consideration must also be given to the potentiality of these invading pathogens to introduce various mechanisms of antimicrobial resistance to the gut community. It is possible that the pathogens themselves harbor ARGs which

can be spread to other gut microbes or vice versa, resulting in formation or maintenance of a resistance reservoir (23, 24). This reservoir is particularly concerning when considering the phenomenon of pathobionts in which common commensal microbes develop pathogenic properties (25); in the context of acquired antimicrobial resistance, these microbial players become much more sinister. Additionally, given that introduction of these pathogens appears to alter the overall relative abundance of various microbes (3), it is also probable that ARGs harbored by microbes that “bloom” during these infections will also increase in abundance. Various approaches have been developed to identify the microbial hosts of ARGs in different environments; these include both physical linking of genetic components to their host (26) as well as *in silico* analyses, such as co-occurrence correlation networks (27, 28) and ARG-carrying contig analysis (29), which use sequencing data to inform microbe-ARG relationships. The elucidation of microbial hosts harboring ARGs may provide useful information regarding potential spread of drug resistance both within the gut and among different environments.

The ongoing plight of antimicrobial resistance, namely among enteric pathogens, is cause for great concern. Further understanding the impacts of these pathogens on the makeup and function of the human gut microbiome is necessary for our fight against continued dissemination of drug resistance. In addition to exploring how infection by and recovery from enteric pathogens influences the composition of the human gut resistome and microbiome, this study also aims to elucidate the roles of specific taxa in harboring ARGs both during and after infection using shotgun metagenome data.

## **METHODS**

### **Study population**

Between 2011 and 2015, 60 stool samples were obtained from patients with enteric bacterial infections prior to treatment (cases). Of these patients, 24 (40.0%) had infections caused by *Campylobacter*, 29 (48.3%) had *Salmonella* infections, and 4 (6.7%) and 3 (5.0%) experienced *Shigella* or Shiga toxin-producing *E. coli* (STEC) infections, respectively. Data from the *Campylobacter* patients were examined previously to examine differences in the resistome relative to healthy family members (5). Patient stools were collected by the Michigan Department of Health and Human Services (MDHHS) as described previously (3) and transported to Michigan State University (MSU) in Cary-Blair transport media following de-identification and pathogen culture. Patients were interviewed about demographics, exposures, and symptoms for reporting through the Michigan Disease Surveillance System (MDSS). They also provided names of household members for inclusion as study controls. After providing informed consent, 125 healthy household members submitted a stool sample to MSU between 5 and 29 weeks following the cases' infection and completed a questionnaire about exposures and symptoms.

In addition, 60 patients submitted a follow-up sample between 1 and 29 weeks after they recovered from their initial infection. Hence, a total of 120 paired samples from 60 patients during and after infection were available for analysis. Ninety-one household controls associated with 38 of the 60 patients were included for comparison in a subset of the analyses. For the epidemiological analysis, county of residence was classified as 'rural' or 'urban' based on the classification scheme developed by the National Center for Health Statistics (30).

## **Sample preparation and sequencing analysis**

Metagenomic DNA from the 120 fecal samples was extracted, sheared, and normalized as described previously (3). Library construction was completed using a TruSeq Nano library kit (Illumina, Inc., San Diego, CA, USA). Shotgun metagenomics sequencing was performed in a series of four sequencing runs on an Illumina HiSeq 2500. Reads were demultiplexed at the MSU Research Technology Support Facility (RTSF). Upon filtering poor quality or heavily contaminated sequences, certain samples were removed from analysis; if a case sample was removed from analysis, the corresponding control and follow-up samples (if present) were also removed to maintain pairedness of the data.

## **AmrPlusPlus – Read-based pipeline**

The AmrPlusPlus v2.0 pipeline was used to perform quality control and align and annotate our metagenomic fragments directly using the MEGARes 2.0 database (31). Briefly, the pipeline employs Trimmomatic (32); parameters supplied to Trimmomatic were followed as described in (5). Metagenomic reads were mapped to the GRCh38 human genome in RefSeq (GRCh38\_latest\_genomic.fna.gz, downloaded December 2020) using Burrows-Wheeler Aligner (BWA) (33) and removed using SAMTools (34) and BEDTools (35). The non-host FASTQ files were stored and aligned to MEGARes 2.0 to identify the ARGs present in each sample using BWA and SAMTools with default values. Metagenomic reads were deduplicated and annotated using ResistomeAnalyzer with an identity threshold of  $\geq 80\%$  to obtain the ARG abundances in each sample, while RarefactionAnalyzer was used to estimate sequencing depth. The final step in the AmrPlusPlus pipeline extracted SNPs pertaining to those ARGs that require specific haplotypes to be classified as resistance genes. The pipeline was designed to confirm these SNPs using the Resistance Gene Identifier (RGI) created in conjunction with The Comprehensive

Antibiotic Resistance Database (CARD) (36). In this analysis, however, all ARGs were included regardless of SNP status as any ARG requiring SNP confirmation was within one point mutation of conferring resistance; therefore, its role as a resistance precursor is likely and relevant to this analysis.

Following annotation and determination of ARG abundances, the average genome size (AGS) and number of genome equivalents (GE) was investigated per sample using MicrobeCensus (37) (Figure B.1). The number of GE was used to normalize ARG and taxonomic abundances. Nonpareil, an assembly- and database-independent tool used to estimate metagenomic coverage, (38) was used to assess the degree of coverage for our short paired end reads (Figure B.2).

### **Identification of microbial taxa**

Non-host paired end reads were taxonomically annotated using Kaiju (39). Kaiju is a protein-based classifier which provides taxonomic annotations by translating metagenomic reads to amino acid sequences and searching for maximum exact matches (MEMs) among microbial reference genomes (39). The reference database used was the NCBI BLAST *nr* database including sequences for bacteria, archaea, viruses, fungi, and microbial eukaryotes. Parameters used when running Kaiju were described previously (5). Raw abundances of reads assigned to taxa were normalized by the estimated number of GE calculated by MicrobeCensus (37).

### **Ecological analyses**

#### *Abundance and diversity analyses*

Resistome and microbiome composition were determined by investigating the identity and diversity of ARGs and taxa across infected cases and recovered follow-ups. For the

resistome, analyses were completed using the gene, group, mechanism, class, and type levels denoted by MEGARes v2.0 and the ResistomeAnalyzer tool in AmrPlusPlus v2.0 (31). Actual estimated abundance of ARGs and taxa was determined by normalizing raw abundance counts to the number of GE per sample. Relative abundance was calculated by dividing the number of GE-normalized reads assigned to a specific feature by the total number of GE-normalized reads for that sample. Alpha diversity metrics such as richness estimates, Shannon diversity, and Pielou's evenness score were obtained using the vegan package (40) in R (41). Nonparametric tests were used for alpha diversity significance testing because the data were presumed to be non-normal. To test for normality, the Shapiro-Wilk test was used to assess the metrics of richness, Shannon diversity, and evenness, each of which registered significant p-values for both the resistome and microbiome data (Table B.1) The Wilcoxon signed-rank test was used to detect significant differences between paired case and follow-up samples, while the Wilcoxon rank-sum test was applied to unpaired samples. Beta diversity metrics and ordination plots (e.g., Principal Coordinate Analysis (PCoA)) based on Bray-Curtis dissimilarity at the gene and group (ARGs) or species and genus (taxa) levels were also determined using vegan. Upon generation of Bray-Curtis dissimilarity matrices, the overall mean dissimilarity among cases and follow-ups was compared to the mean dissimilarity between all paired case-follow samples. A Welch's t-test was used to determine whether these means were statistically significant; means were also plotted onto a histogram demonstrating the distribution of dissimilarity measures across samples (Figure B.3). A Permutational Analysis of Variance (PERMANOVA) was completed on the Bray-Curtis dissimilarities in R to assess differences in centroids (mean) between cases and follow-ups for resistome and microbiome composition; Permutational Analysis of Multivariate Dispersion (PERMDISP) was used to detect differences in dispersion (degree of spread) of these groups.

### *Differential abundance of taxa and ARGs*

To assess representative features in cases and follow-ups, MMUPHIn was used to construct general linear models relating various sample features to microbiome and resistome relative abundances (42). First, MMUPHIn was used to perform batch adjustment of relative abundance data based on sequencing run since this variable was significantly influencing the distribution of points in our microbiome ordination (Figure B.4). Next, a linear model was constructed to identify differentially abundant ARGs and taxa among cases and follow-ups; follow-ups were used as the reference for the fixed effect, while age in years, average genome size, number of genome equivalents, year of collection, and use of antibiotics were included as covariates. Significance values were adjusted using the Benjamini-Hochberg method of correction for multiple hypothesis testing (q-value representing False Discovery Rate (FDR)). Since a prior study showed that different abundance testing methods can result in skewed data interpretations (43), the Analysis of compositions of microbiomes with bias correction (ANCOM-BC) method (44) was also used for comparison. ANCOM-BC considers absolute abundances (which we included as GE-normalized counts) as input and cannot currently implement a mixed model in which fixed and random effects are considered. This lack of additional random effects or covariates may explain the variation observed. Nevertheless, ANCOM has been cited as one of the most reliable methods for differential abundance testing (43), and its overall concordance with the findings of MMUPHIn increase our confidence of the respective findings.

### *Identification of continuous population structure*

MMUPHIn was also used to further characterize the intrinsic drivers of point distributions observed in our beta diversity analyses (ordination). To do this, the

'continuous\_discover()' function was applied to our microbiome and resistome abundance data. This function performs unsupervised continuous structure discovery using Principal Components Analysis (PCA); continuous structure scores (called "loadings") that comprise the top principal components are compared across batches to identify "consensus" loadings assigned to certain microbial features. A parameter, 'var\_perc\_cutoff()', which instructs the method to filter out the top components accounting for at set proportion of the variability within the samples, was set to 0.75 for phylum and ARG class levels and 0.50 for genus and ARG group levels. At the species level, 'var\_perc\_cutoff()' was set to 0.40. The different filters are needed because levels with broader characterization (e.g., phylum and ARG class) have fewer categories, and therefore, each category accounts for greater variability by default. Upon generation of these loadings, we constructed respective plots to visualize the main drivers of continuous data structure and overlaid the data onto ordination plots which displayed the Bray-Curtis dissimilarity of microbiome or resistome relative abundances. In nearly every comparison, the distribution of points could be attributed to a taxonomic and/or ARG tradeoff.

### **Co-occurrence network construction**

Co-occurrence network analysis was completed on ARG and taxonomic data to explore feature associations among cases and follow-ups. Prior to network construction, a subset of ARG and taxonomic GE-normalized abundances was taken to obtain approximately 50% of the most prevalent features in each dataset (ARG groups = 251; genera = 2,282). This was pursued because network performance has been shown to decrease markedly with high sparsity of data (high proportion of zeroes, usually among rare taxa or genes), and removal of such data results in a higher true positive-to-false positive ratio (45). These subsets were used to calculate Spearman's Rank correlation coefficients ( $\rho$ ). Potential correlations were explored for ARG-



ARG, ARG-taxa and taxa-taxa associations among cases and follow-ups. A more refined analysis was also performed to investigate whether there were differing trends in ARG-taxa co-occurrence among pathogens during and after infection. The number of cases/follow-ups infected with *Shigella* (n=4/4) or STEC (n=3/3) were not included in analysis, however, since a Spearman Rank correlation matrix cannot be constructed with fewer than four observations. Since these networks are constructed separately for cases and follow-ups, there were not enough samples to use this approach for these two genera. However, correlation matrices and networks were successfully constructed for those infected by *Salmonella* and *Campylobacter*.

A correlation between two features (called “nodes” in the network) was considered significant if  $\rho \geq 0.80$  and  $p \leq 0.01$ . Spearman correlations were determined using the “rcorr” function from the Hmisc package v4.5-0 (46) in R. The output from Hmisc was formatted to compose a nodes file and an edges file, which were imported into Gephi for visualization (47); the Fruchterman-Reingold layout was chosen to display all associations among ARGs and taxa. Filters were applied so all nodes required a degree (e.g., number of connections) greater than 1. To detect associations between ARGs and specific taxa, the MASK setting was used to isolate the Partition Type “ARG”, meaning that the only connections displayed in the network involved ARGs directly (i.e., ARG-ARG or ARG-taxa associations only). Two separate analyses were performed: a global analysis and a targeted analysis exploring associations related to the beta-lactam class. For each analysis, separate correlation networks were constructed for cases and follow-ups; the results of the correlation matrix construction as well as the visual network were compared as a method of prediction for taxa-associated ARGs.

### **Anvi'o – Assembly-based pipeline**

The non-host FASTQ files generated with AmrPlusPlus v2.0 were used for metagenome assembly. Prior to assembly, BBTools was used for paired end read merging using the 'bbmerge-auto.sh' script; if reads initially failed merging they were error-corrected using Tadpole (48) and reexamined. If merging continued to fail, reads were extended 20bp and merging was iterated up to five additional times until complete. If merging failed, unmerged original reads were included. Assembly was performed with MEGAHIT (49) using the forward and reverse paired end reads in addition to the merged reads. The Quality Assessment Tool for Genome Assemblies (QUAST) (50) was used to assess assembly quality and coverage (Figure B.5).

Following assembly, a custom workflow was composed using tools provided in anvi'o to analyze and visualize microbial genomes from metagenomes (51). First, assembled contigs were reformatted using 'anvi-script-reformat-fasta' to generate a contigs database for each sample using 'anvi-gen-contigs-database'. The script 'anvi-run-hmms' was run to populate the contigs database with hits found using Hidden Markov Models, a strategy that can improve assembly annotation. Prodigal (52) was used in the script 'anvi-get-sequences-for-gene-calls' to obtain the amino acid sequences of genes present in our assemblies for use in the ARG-carrying contigs analysis.

### **ARG-carrying contigs host-tracking analysis**

Gene calls obtained from anvi'o were used to identify ARG-carrying contigs (ACCs) by aligning our amino acid sequences to the HMD-ARG database (53) using DIAMOND (54). The resulting SAM files were filtered to identify contigs with a hit listed as "antibiotic"; these contig IDs were stored in a list. Seqtk (55) was used to extract these ARG-carrying contigs from the original list of gene calls and store them in a separate FASTA file. Finally, the resulting FASTA

files were aligned to the BLAST database v5.0 (56, 57) using blastp to identify microbial taxa represented by the ACCs and confirm the presence of ARGs; an E-value of 0.00001 was used as a cutoff with a maximum number of 50 target sequences (i.e., up to 50 matches were allowed per contig).

Of the 60 case-follow-up pairs, one pair that had been infected with *Campylobacter* could not be properly annotated and was excluded from this analysis, resulting in 59 pairs (118 samples). The output from our BLASTP alignment was used to identify the most likely taxon associated with each ARG identified on a contig. Since up to 50 matches (hsps) were allowed per contig, a custom Python script was composed to quantify the proportions of each genus comprising a contig. The script then determined the most prevalent genus (via maximum number of hits and highest calculated percentage) per contig. In other words, any taxon representing the greatest percentage of hits per contig was considered the most likely taxon to be present in association with that particular ARG. The custom Python script output two different types of files: one which quantified the average proportion of each genus per sample on the ACCs and one which quantified the average percentage of different ARGs per genus within each sample's ACCs. The former was used to determine which genera, on average, most commonly harbor ACCs in our samples. The latter was used to identify which ARGs are found in these prominent genera and whether they differed among cases and follow-ups.

## **RESULTS**

### **Study population characteristics**

The 60 cases were infected with one of four different enteric pathogens (*Campylobacter* (n=24), *Salmonella* (n=29), *Shigella* (n=4), or Shiga toxin-producing *E. coli* (n=3)). Stools were obtained from each case during acute infection as well as 8 to 205 days post-recovery (i.e.,

“follow-ups”), yielding 120 stool samples in all. The average number of days to follow-up was 107.9, though this information was absent for one individual. Most follow-up samples were submitted between 101-150 days after the initial sample (n=28; 47.5%), followed by 51-100 days (n=20; 33.9%). Fewer follow-up samples were taken  $\leq 50$  or  $>150$  days after initial sample (n=4; 6.7% and n=7; 11.9%, respectively).

Of the 60 individuals, 28 were male (46.7%) and 32 were female (53.3%). The age range in years was between 1.5 and 90, with many (n=16; 26.7%) representing 0-9 years, followed by 10-18 years (n=6; 10.0%), 19-64 years (n=26; 43.3%), and  $\geq 65$  years (n=12; 20.0%). Forty-eight (80.0%) cases self-reported as Caucasian, whereas five (8.3%) self-identified as African American and 2 (3.3%) as Asian; one individual (1.7%) reported more than one race, and seven individuals (6.7%) failed to respond. No difference in the proportion of stool submissions was observed by year, though the lowest frequency (n=13.3%) was recovered in 2011. Sixteen were recovered (26.7%) in 2012, 22 (36.7%) in 2013, and 14 (23.3%) in 2014. Fifty-nine of the cases responded to prompts regarding symptoms experienced during infection, with 50 (84.8%) reporting abdominal pain and 57 (96.6%) reporting diarrhea. Twenty (33.9%) and 28 (47.5%) patients reported vomiting or nausea, respectively, while 22 cases (37.3%) reported bloody stool. While infected, just two people (3.4%) described their stool as being “Solid”, while a majority described their stool as either “Loose” (n=8; 13.6%) or “Watery” (n=49; 83.0%). Fortunately, upon recovery from infection, most follow-ups described their stool as “Solid” (n=45; 75.0%). Most of the 60 cases received care in an outpatient setting (n=40; 66.7%); however, 17 people (28.3%) required hospitalization. While just two cases (3.3%) reported taking amoxicillin within two weeks of stool collection, five (8.3%) reported antibiotic use during their recovery. These

five cases had taken either amoxicillin (n=2), azithromycin (n=1), ciprofloxacin (n=1), or an unknown antibiotic (n=1) up to two weeks before sample collection.

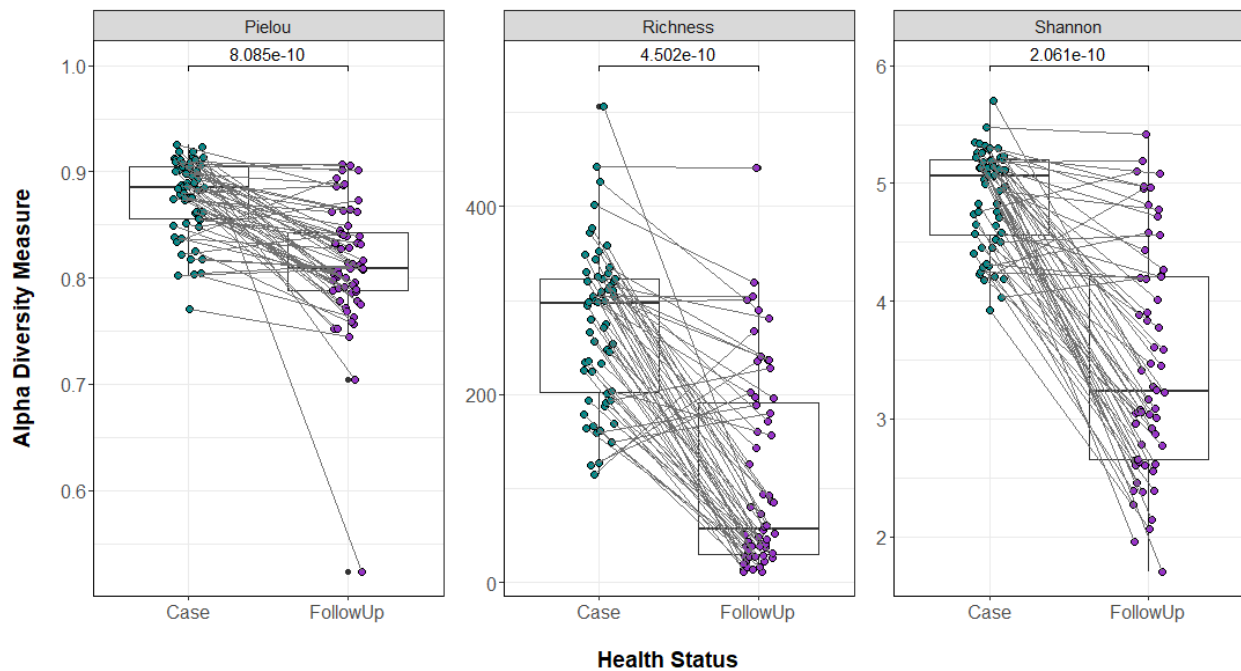
Thirty-three (55.0%) of the cases lived in rural areas and 26 (43.3%) lived in urban counties with one individual failing to respond. Most individuals had access to municipal water (n=38; 63.3%), though a subset reported well water (n=10; 16.7%), bottled water (n=4; 6.7%), or both municipal and bottled water (n=1; 1.67%) as their main water source; seven people did not respond. Fifty-eight people responded to prompts regarding recent travel. Nineteen cases (32.76%) reported traveling within the last month, with 12 (20.7%) indicating travel within the United States and 8 (13.8%) reporting non-domestic travel.

### **Changes in the composition and diversity of the resistome and microbiome after recovery from enteric infection**

#### *Resistome diversity*

Our resistome analysis identified 1,212 resistance genes among the 120 stool samples. These genes encode resistance to four different overarching types of compounds: biocides, antibiotic drugs, metals, and multi-compound substrates. Among the resistance genes, 474 distinct gene groups or operons are represented that translate into 120 mechanisms conferring resistance to 44 different classes of compounds. Infected cases had significantly more diverse resistomes than follow-ups with a greater mean ARG richness ( $S_{\text{cases}}=254$  vs.  $S_{\text{follow-ups}}=103$ , respectively;  $p=4.5e-10$ ) (Figure 3.1). The Shannon diversity index was also greater in cases than follow-ups ( $H_{\text{cases}}=4.79$  vs.  $H_{\text{follow-ups}}=3.36$ ;  $p=2.1e-10$ ) as was the Pielou's evenness index ( $J'_{\text{cases}}=0.87$  vs.  $J'_{\text{follow-ups}}=0.80$ ;  $p=8.1e-10$ ). Notably, the family member controls did not significantly differ from follow-up samples, suggesting recovery to a relatively "normal" state following infection (Figure B.6A).

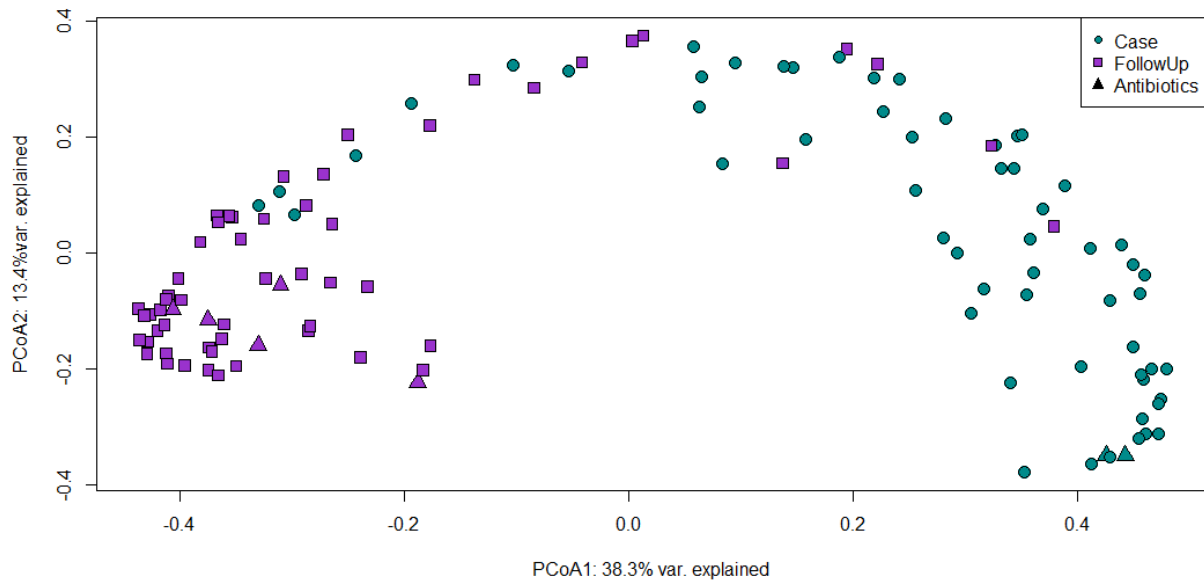
Beta-diversity analysis revealed that the composition of these paired case and follow-up resistomes also differed. Principal Coordinate Analysis (PCoA) was performed on the Bray-Curtis dissimilarity of cases and follow-ups (Figure 3.2). PERMANOVA revealed notable separation of these two groups ( $p=0.000999$ ;  $F=38.75$ ). PERMDISP did not identify a significant difference in the level of dispersion among cases and follow-ups ( $p=0.52$ ;  $F=0.468$ ). Importantly, the five individuals who self-reported antibiotic use prior to sampling did not cluster separately



**Figure 3.1. Resistome diversity is greater during enteric infection than after recovery.** Three alpha diversity measures are shown above (Richness, Shannon's Diversity Index, and Pielou's Evenness Index); these are stratified by health status, with samples represented by circles (cases=green; follow-ups=purple). Points are slightly offset from the vertical to allow interpretation of all samples. The median of each measure is indicated by the thick bar within each box (green for cases; purple for follow-ups) and the first and third quartiles are represented by the bottom and top of the box, respectively. The gray lines between points connect both of an individual's samples: the sample taken during infection (case) and the sample taken during recovery (follow-up). P-values were calculated using the Wilcoxon signed-rank test for paired samples and are shown above the comparison bar within each plot.

from those that did not receive antibiotics or did not disclose their antibiotic use. Data for residence type, antibiotic use, gender, age, hospital, county of origin, stool type, sequencing run,

and number of days between the initial sampling and follow-up samples (“Follow-Up Days”) were fit to the ordination. Of these variables, age in years ( $p=0.013$ ), and year of collection ( $p=0.043$ ) were found to significantly and independently influence the distribution of points. However, residence location, hospital, and the number of days since infection trended toward significance ( $p=0.097$ ,  $p=0.099$ , and  $p=0.092$ , respectively).



**Figure 3.2. Resistomes during infection differ significantly from those of recovered samples.**

A Principal Coordinates Analysis (PCoA) plot of case (green, circles) and follow-up (purple, squares) resistomes based on Bray-Curtis dissimilarity calculated from gene-level abundances. The first and second coordinate are shown and include the corresponding percentage of similarity explained. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points.

We re-plotted the ordination using group-level data (which includes the gene- or operon-level group) rather than gene-level data (which includes sequence-level gene information), for increased clarity due to fewer features. After fitting intrinsic variables (i.e., resistance gene information) to the ordination, 30 groups registered an  $R^2$ -value  $>0.75$  and a  $p$ -value  $\leq 0.001$ . Those with the greatest  $R^2$  included *bacA* (bacitracin resistance), *cpxAR* (drug and biocide resistance), *glpT* (fosfomycin resistance), and *copA* (copper resistance). Based upon the

directionality of the fitted vectors in the ordination plot, these four ARG groups are primarily driving separation of cases and follow-ups along the first coordinate (Figure B.7). As was observed for alpha diversity, follow-ups had a similar resistome composition to controls in the ordination plot (Figure B.6B). Notably, the pathogen responsible for infection did not have a significant effect on alpha or beta diversity trends (Figure B.8).

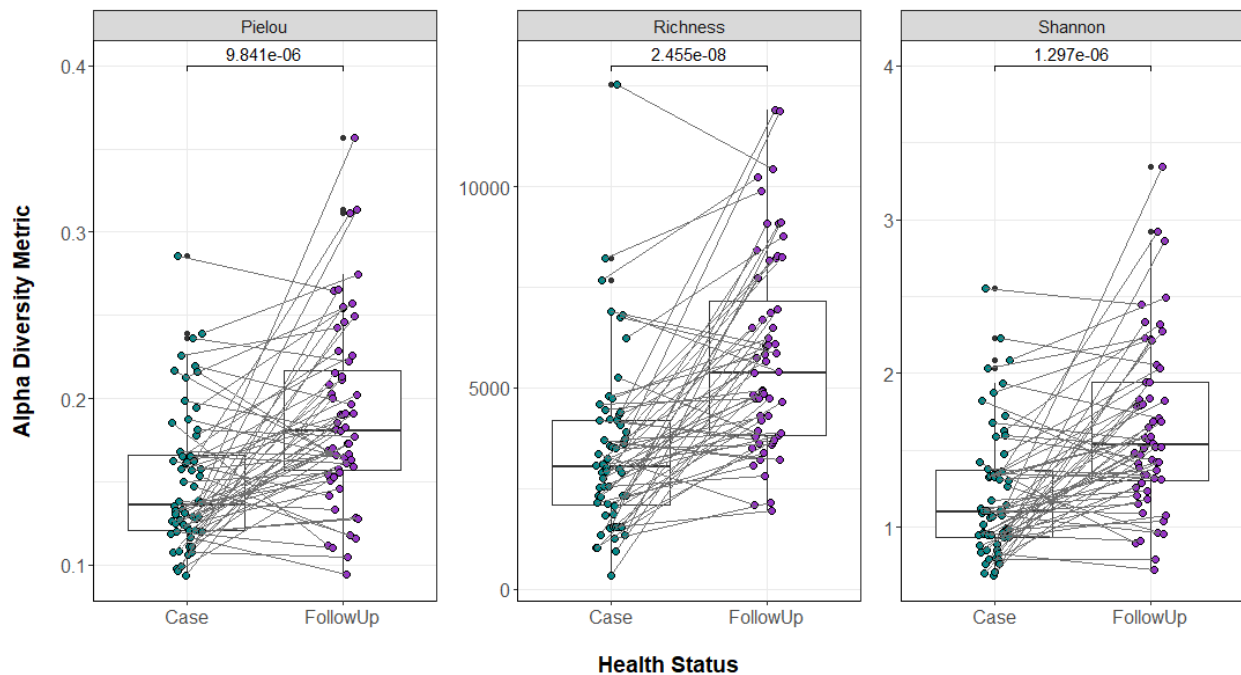
### *Microbiome diversity*

Among cases and follow-ups, a total of 40,022 species, 4,851 genera, 1,157 families, 537 orders, 236 classes, and 224 phyla were found. The trends for microbiome diversity were opposite that of the resistomes; follow-ups had more diverse gut microbiomes than their corresponding cases (Figure 3.3). Not only was the mean species richness significantly greater after recovery from infection ( $S_{\text{cases}}=3,426$ ,  $S_{\text{follow-ups}}=5,789$ ;  $p=2.5e-08$ ), but the recovered microbiomes had greater mean evenness ( $J'_{\text{case}}=0.150$ ,  $J'_{\text{follow-up}}=0.190$ ;  $p=9.8e-06$ ) and a higher mean Shannon Diversity index ( $H_{\text{cases}}=1.21$ ,  $H_{\text{follow-ups}}=1.65$ ;  $p=1.3e-06$ ). When compared to controls, follow-ups had similar levels for Shannon Diversity and evenness, though the measure of richness ( $S_{\text{follow-ups}}=5,789$ ,  $S_{\text{controls}}=6,872$ ;  $p=0.012$ , Wilcoxon rank-sum test (unpaired)) differed significantly between the groups (Figure B.6C).

Principal Coordinate Analysis (PCoA) of the Bray-Curtis dissimilarity among cases and follow-ups showed considerable overlap in the microbiome composition at the species level, yet significant differences between the groups were found (Figure 3.4; PERMANOVA ( $p=0.000999$ ;  $F=7.31$ )). PERMDISP, however, did not detect a significant difference in the dispersion of points between cases and follow-ups ( $p=0.086$ ;  $F=2.86$ ). The same extrinsic covariates described prior were fitted to the PCoA and three significantly impacted the distribution of points ( $p\leq 0.01$ ). These included age in years ( $p=0.008$ ), sequencing run ( $p=0.001$ ), average genome size



( $p=0.001$ ), number of genome equivalents ( $p=0.001$ ), year of sampling ( $p=0.005$ ), and antibiotic use (Yes vs. No;  $p=0.008$ ). Notably, the number of follow-up days ( $p=0.013$ ) and hospital ( $p=0.030$ ) also met the significance cutoff of  $p=0.05$ . Intrinsic variables (e.g., species and genus) were also fitted. At the genus level, the  $R^2$ -values were lower than those found in the resistome analysis. Of the genera that had an effect on the ordination, *Cronobacter*, *Pseudoalteromonas*, and *Cedecea*

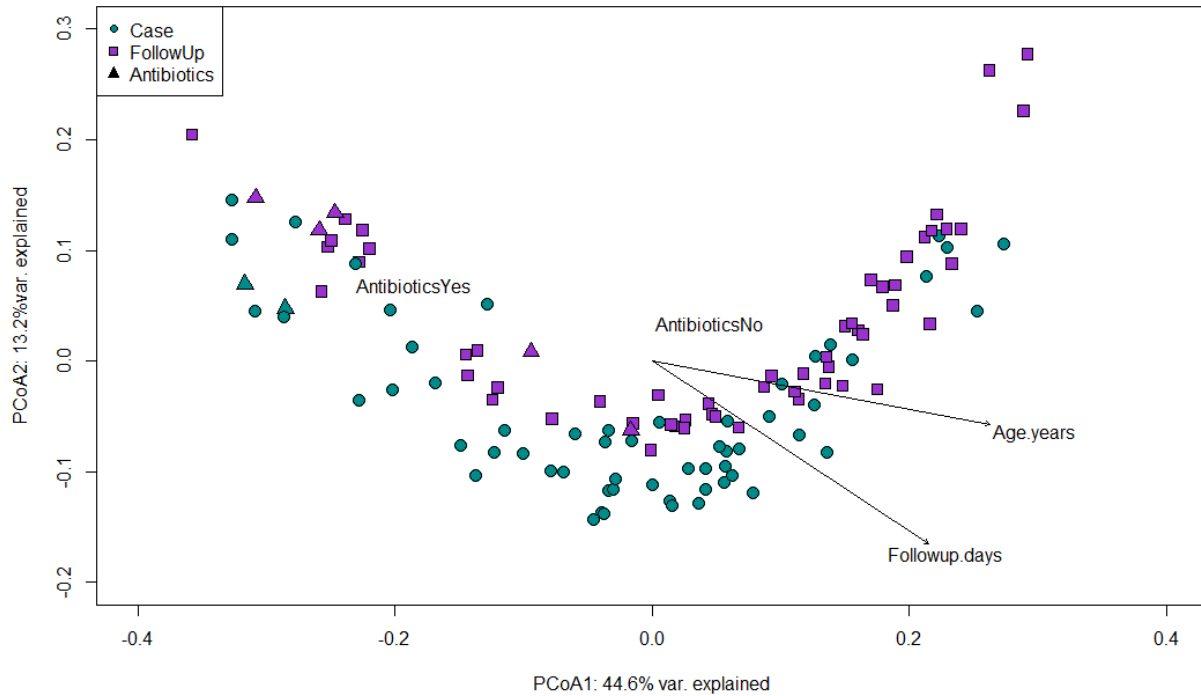


**Figure 3.3. Microbiome diversity is greater after recovering from enteric infection.**

Three alpha diversity measures are shown above to represent microbiome diversity (Richness, Shannon's Diversity Index, and Pielou's Evenness Index); these are stratified by health status, with samples represented by circles (cases=green; follow-ups=purple). Points are slightly offset from the vertical to allow interpretation of all samples. The median of each measure is indicated by the thick bar within each box (green for cases; purple for follow-ups) and the first and third quartiles are represented by the bottom and top of the box, respectively. The gray lines between points connect both of an individual's samples: the sample taken during infection (case) and the sample taken during recovery (follow-up). P-values were calculated using the Wilcoxon signed-rank test for paired samples and are shown above the comparison bar within each plot.

had the highest  $R^2$ -values ( $p \leq 0.001$ ); these, along with a cluster of other genera including *Escherichia*, *Salmonella*, and *Shigella*, among others, were the key contributors to the separation

of cases from follow-ups. *Bacteroides* also demonstrated a significant  $R^2$ -value of 0.547 ( $p=0.001$ ); contrary to the other genera explored, this genus directed points higher on the y-axis (Figure B.9). In all, the diversity metrics for cases and follow-ups did not differ after stratifying by infecting pathogen (Figure B.10).



**Figure 3.4. Compositional differences between case and follow-up microbiomes are nuanced.**

A Principal Coordinates Analysis (PCoA) plot of case (green, circles) and follow-up (purple, squares) microbiomes based on Bray-Curtis dissimilarity at the species level. A biplot was overlaid to display variables that were found to have a significant influence on the distribution of points in the ordination. Notably, Age-in-years and the number of follow-up days were influential vectors while the binary variable for receiving antibiotics (Yes/No) was an influential factor. The first and second coordinate are shown and include the corresponding percentage of similarity explained. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points.

*Exploring potential for continuous structure of resistome and microbiome compositions using MMUPHin*

In addition to its use in differential abundance testing, MMUPHin was used to identify continuous population structure from the microbiome and resistome abundance data. Continuous

structure discovery can be useful when attempting to identify taxonomic or, in this case, resistance gene tradeoffs that could be driving observed structure of data. For each level of investigation (e.g., phylum, genus, ARG class, ARG group), the top contributing features were determined and relevant continuous structure scores were overlaid onto ordination plots. As suggested by the results of the `envfit()` function, various taxonomic and resistance features were associated with point distribution. Namely, when considering taxonomy, we observed a notable tradeoff between the phyla Proteobacteria (primarily displayed by cases) and Bacteroides and Firmicutes (mostly in follow-ups and some overlapping cases) (Figure 3.5A). Additionally, there is a taxonomic tradeoff at the genus level, with an evident gradient between *Escherichia-Salmonella-Klebsiella-Shigella-Pseudomonas*-containing samples vs. those dominated by *Bacteroides* and *Alistipes* (Figure 3.5B). Indeed, these differences are visible when overlaid onto ordination as a gradient relevant to loading score can be observed (Figure 3.5C/D). At the species level, which reveals gradients at the greatest resolution, we observed a tradeoff between harboring *Escherichia coli*, *Klebsiella pneumoniae*, and *Shigella sonnei* vs. many *Bacteroides* species (including *B. fragilis*, *B. stercoris*, *B. uniformis*, and more) and *Phocaeicola* species (namely *P. vulgatus* and *P. plebeius*) (Figure B.11).

Tradeoffs were also observed among resistance genes. At the class level, there is a continuous gradient relative to tetracycline-, macrolide, lincosamide, and streptogramin (MLS)-aminoglycoside-dominant resistomes vs. ARGs for multiple classes such as multi-metal resistance, drug and biocide resistance, and drug, metal, and biocide resistance (Figure 3.6A). . At the ARG group level, *tetQ* represents a dominant driver of continuous structure scoring for follow-ups while resistance genes such as *rpoB*, *acrA*, *acrB*, *mdtC*, and *mdtB* were defining for the opposite side of the PCoA axis (Figure 3.6B). Overlaying these loading scores onto

ordination further reveals the taxonomic gradients identified among cases and follow-ups (Figure 3.6C/D).

### *Resistome composition*

The relative abundance of ARGs comprising the resistome differed between cases and follow-up samples. Genes conferring drug resistance accounted for an average of 44.8% of the total genes annotated in cases but represented 84.9% of ARGs in follow-ups (Figure B.12). Nonetheless, the actual abundance (which considers the average number of reads assigned to a Type) was higher in cases (n=71.5 reads) than follow-ups (n=51.4). Interestingly, genes for multi-compound and metal resistance were more highly represented in cases, with a relative abundance of 24.8% and 22.0%, respectively; these types were far less prevalent in follow-ups, with respective relative abundances of 6.9% and 6.2%. These trends also hold true when considering actual abundances; cases had an average abundance of 47.6 and 42.3 reads for multi-compound resistance and metal resistance, respectively, while follow-ups contained just 30.0 and 27.4 reads for these respective Types. In both cases and follow-ups, genes relevant to biocide resistance were least represented with relative abundances of 8.4% and 2.0%, respectively. This was also reflected in their actual abundances (cases=15.8; follow-ups=10.3).

At the Class level, the compositional differences were even more pronounced (Figure 3.7). The resistance classes with the greatest relative abundance in cases were for drugs and biocides (15.1%), MLS (13.3%), and multi-metals (11.3%). These classes also have the greatest actual abundance among cases, registering average read counts of 46.4, 25.4, and 33.9, respectively. The top-three most relatively abundant classes in follow-ups were MLS (33.5%), tetracycline (22.0%), and aminoglycoside (15.5%) resistance, a trend also reflected in the actual abundances (read counts = 8.7, 5.4, 4.3, respectively). Notably, the top-three resistance classes in

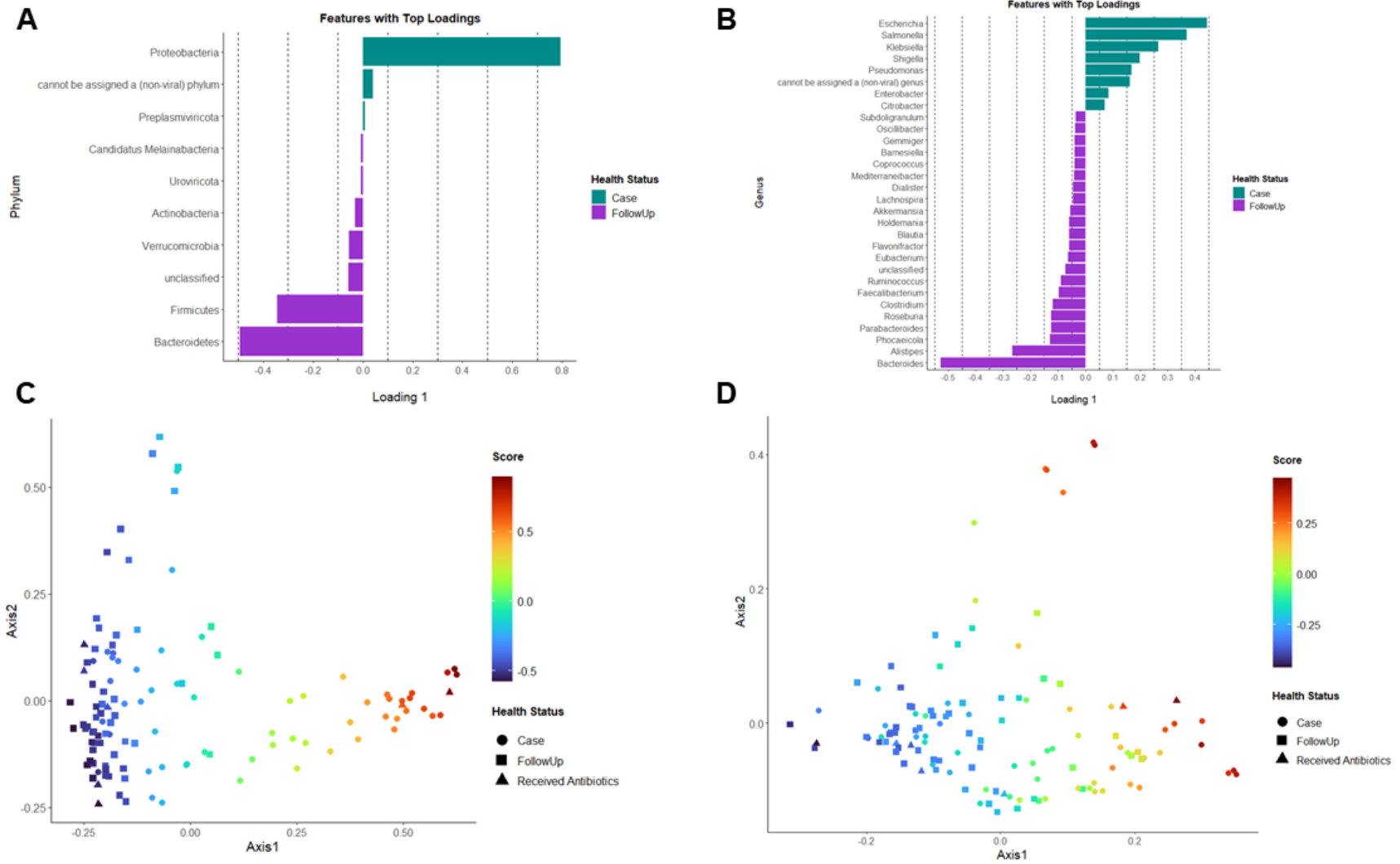


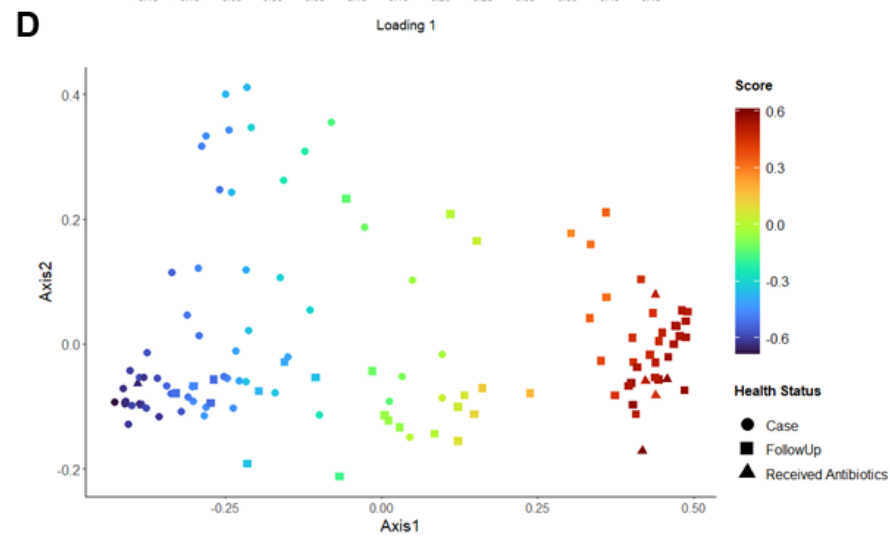
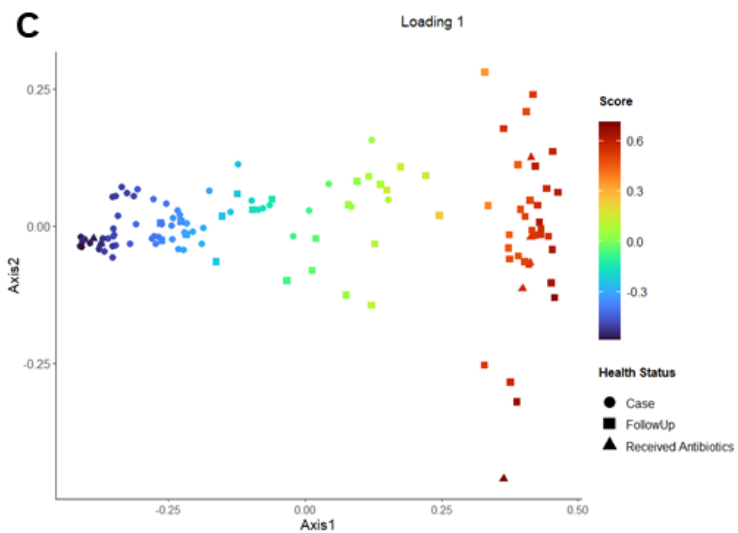
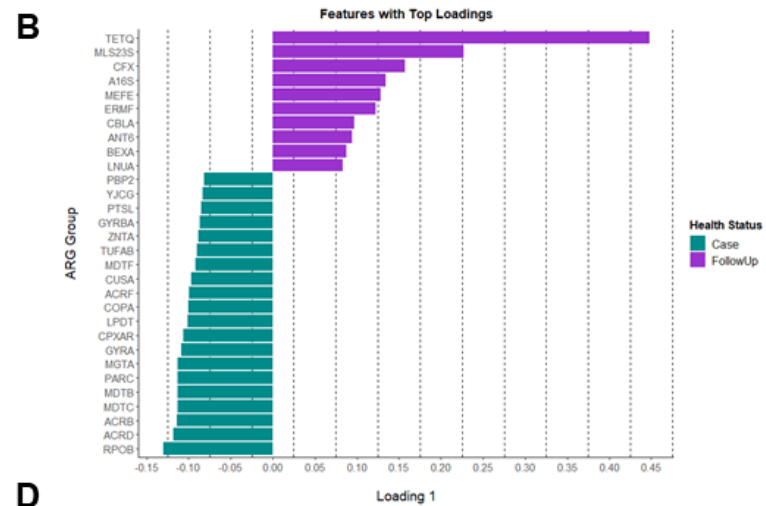
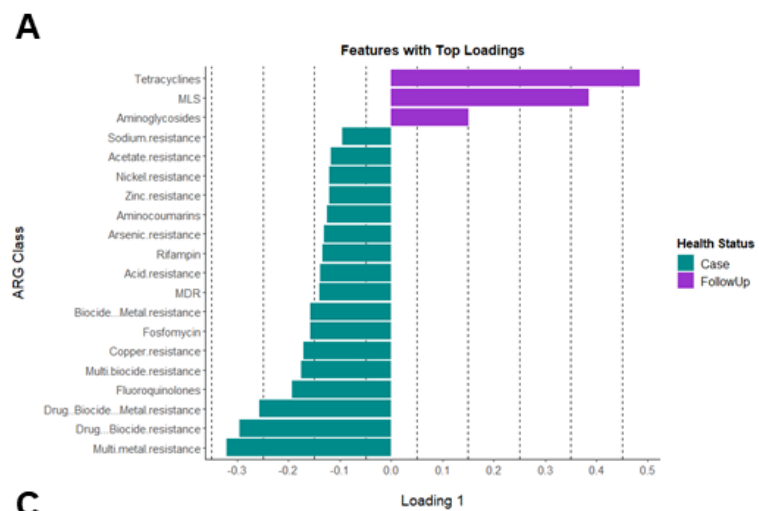
Figure 3.5. Continuous structure analysis reveals taxonomic gradients driving distribution of samples across the population.

### Figure 3.5 (cont'd)

MMUPHin was used to investigate potential continuous structure of the microbiome composition among cases and follow-ups. Phyla (**A**) and genera (**B**) determined to comprise top consensus loadings of the PCA are shown; colors have been assigned to the loadings based on primary health status affiliated with each loading (drawn from differential abundance analyses; cases (green) and follow-ups (purple)). The phylum (**C**) and genus (**D**) composition gradients are shown overlaid onto respective ordination plot based on Bray-Curtis dissimilarity of case and follow-up microbiomes at the phylum and genus level, respectively. Cases (circles), follow-ups (squares), and individuals who received antibiotics (triangles) are shown. The color gradient ("Score") refers to the continuous structure score affiliated with Loading 1 for phyla and genera, respectively. Juxtaposition of (**A**)-(**C**) and (**B**)-(**D**) allow interpretation of phyla and genera tradeoffs, respectively, that occur within the samples. For example, at the phylum level, we see a stark tradeoff between Proteobacteria-dominant and Bacteroidetes/Firmicutes-dominant samples. At the genus level, tradeoffs between *Escherichia*, *Salmonella*, *Klebsiella*, *Shigella*, among other Proteobacteria and *Bacteroides*, *Alistipes*, and various Firmicutes are evident.

cases account for just 39.8% of the total resistance genes, while in follow-ups, the top-three represent 71.0%, a trend that reiterates the greater resistome diversity of patients during infection. It is also important to consider the incredible difference in actual abundances of these classes; while cases register average read counts above 25, the average abundance of these genes is below 10 in follow-ups.

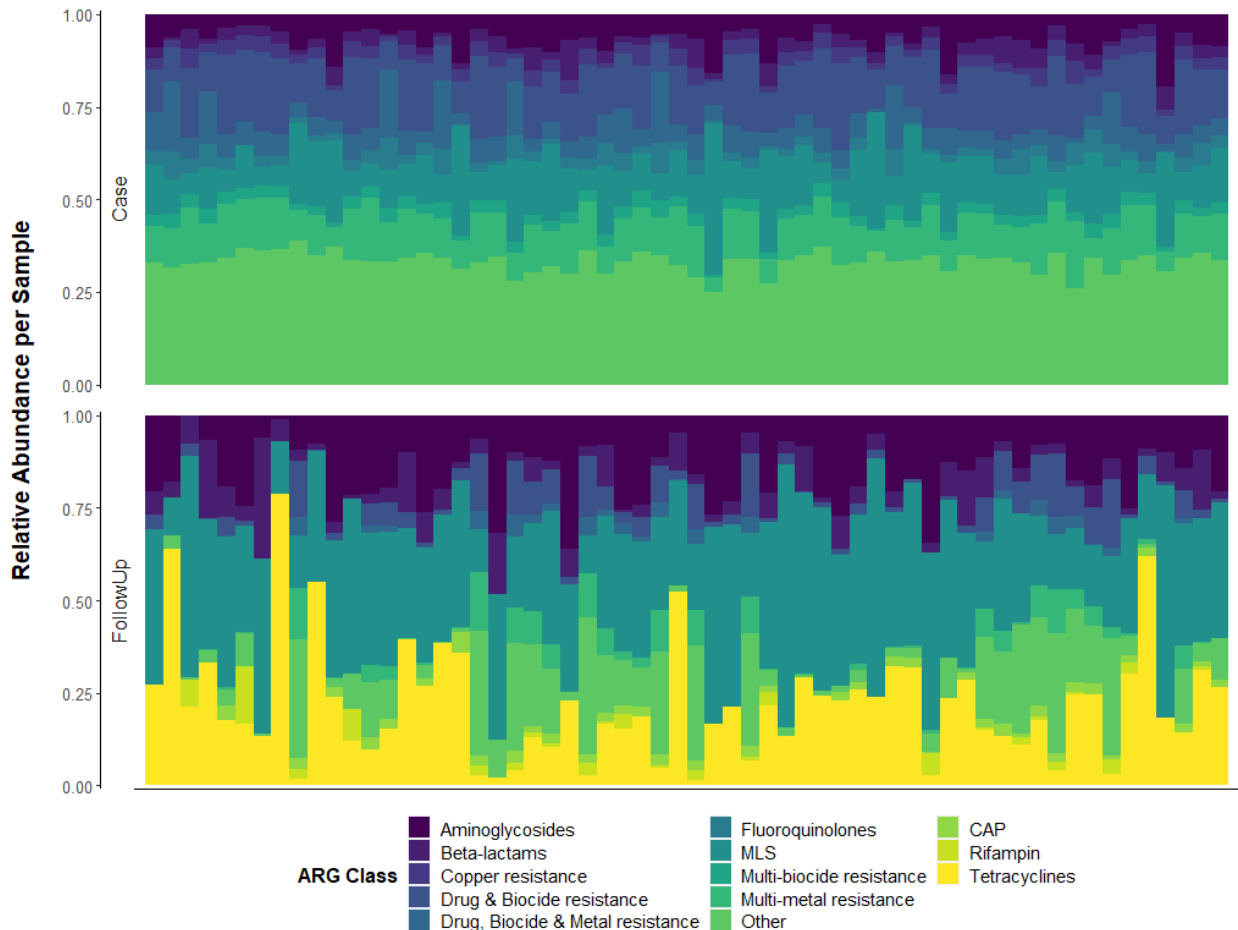
Mechanistically, the macrolide-resistant 23S rRNA mutation was most abundant in both cases (n=24.0; 11.9%) and follow-ups (n=6.5; 24.4%), respectively. The next most abundant mechanisms in cases were Resistance-Nodulation-Division (RND) efflux pumps relevant to drug and biocide resistance and drug, metal, and biocide resistance (average n=19.6 and 17.2 reads, respectively) which, together, comprised 12.2% of all mechanisms detected in cases. As anticipated, the next most prevalent mechanisms in follow-ups related to tetracycline and aminoglycoside resistance; tetracycline resistant ribosomal protection proteins had an average abundance of 4.8 (19.9%) and the aminoglycoside resistant 16S ribosomal subunit protein averaged 3.5 reads per sample (12.6%).



**Figure 3.6. Continuous structure analysis highlights ARG abundance gradients driving differences among cases and follow-ups.**

**Figure 3.6 (cont'd)**

MMUPHin was used to investigate potential continuous structure of the resistome composition among cases and follow-ups. ARG class (A) and group (B) determined to comprise top consensus loadings of the PCA are shown; colors have been assigned to the loadings based on primary health status affiliated with each loading (drawn from differential abundance analyses; cases (green) and follow-ups (purple)). Class (C) and group (D) composition gradients are shown overlaid onto respective ordination plot based on Bray-Curtis dissimilarity of case and follow-up resistomes at the class and group level, respectively. Cases (circles), follow-ups (squares), and individuals who received antibiotics (triangles) are shown. The color gradient (“Score”) refers to the continuous structure score affiliated with Loading 1 for class and group, respectively. Juxtaposition of (A)-(C) and (B)-(D) allow interpretation of ARG class and group tradeoffs, respectively that occur within the samples. At the ARG class level, we observe tradeoffs between resistomes dominated by tetracycline and MLS ARGs and those containing primarily multi-compound and multi-drug ARGs. At the group level, similar patterns are observed: we see tradeoffs between multi-compound resistance genes such as the *acr* and *mdt* gene families and those which confer tetracycline and MLS resistance (*tetQ*, *mIs23*, etc.).



**Figure 3.7. Relative abundance of the Top-10 resistance gene classes notably differs between case and follow-up samples.**



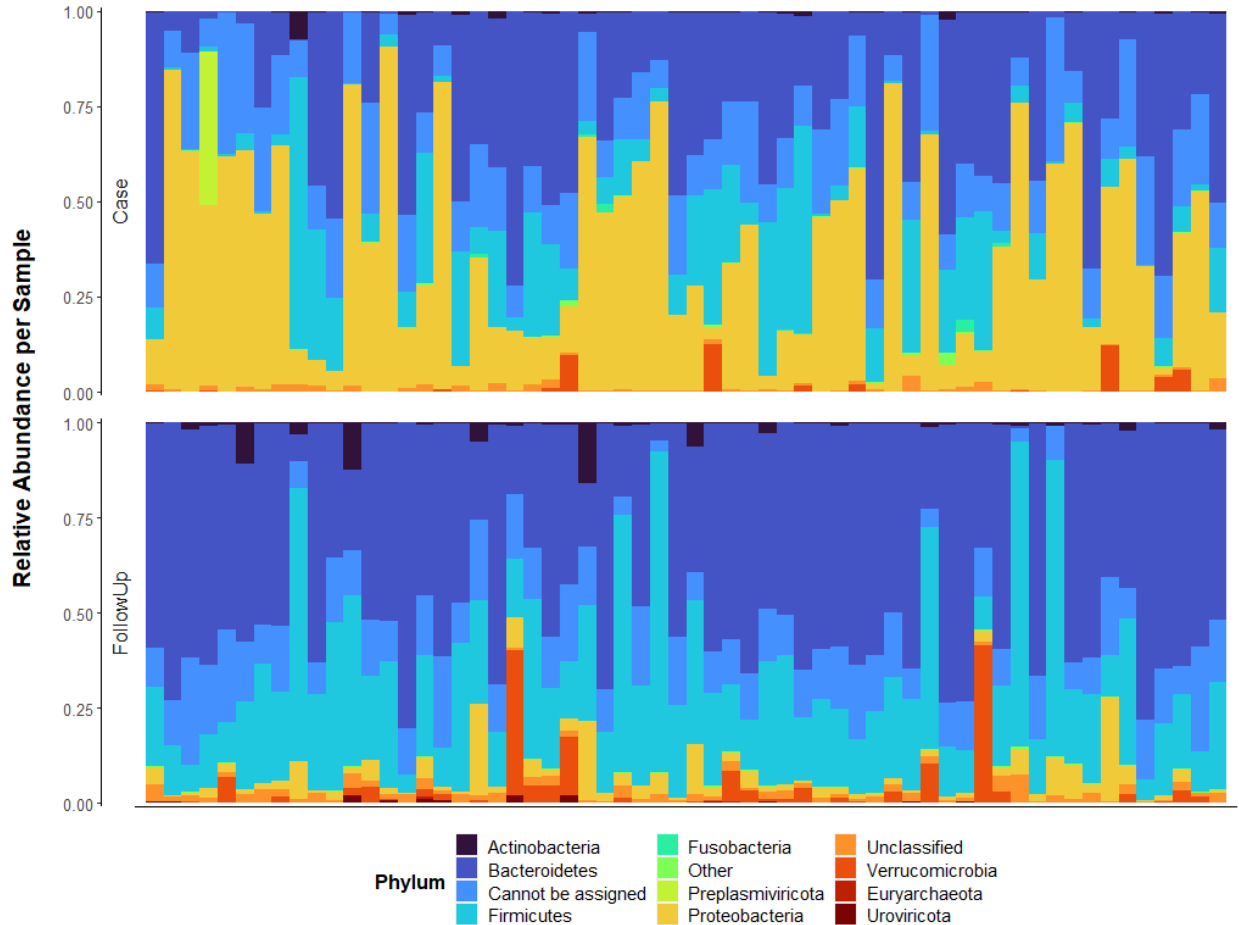
### Figure 3.7 (cont'd)

The relative abundance of resistance genes assigned to 44 different compound classes is shown for each health status, with each column representing the resistome from one individual. Columns are ordered by their sample pairing, meaning that the column position in each side of the plot refers to the same individual either during (Case; Left) or after (FollowUp; Right) enteric infection. Relative abundances were determined using raw gene abundances that had been normalized by the approximate number of genome equivalents in the sample as determined using MicrobeCensus. CAP = cationic antimicrobial peptides; MLS = Macrolide, Lincosamide, Streptogramin; MDR = Multidrug resistance; QACs = Quaternary Ammonium Compounds.

Finally, when investigating the group level, the trends observed in other levels of annotation continue to show. The most abundant groups in cases were MLS23S (average  $n=24.0$ ; 11.9%), *rpoB* ( $n=7.2$ ; 2.8%), and A16S ( $n=6.2$ ; 3.8%). The groups which facilitate drug and biocide and drug, biocide, and metal resistance (*mdtB* and *mdtC*) identified prior registered average abundances of 4.2 (1.4%) and 3.9 (1.3%) in cases. In follow-ups, the most abundant groups were relevant to MLS, tetracyclines, and aminoglycosides; MLS23S ( $n=6.6$ ; 24.3%), *tetQ* ( $n=4.0$ ; 17.0%), and A16S ( $n=2.4$ ; 9.5%) all dominate.

### Microbiome composition

The composition of infected and recovered microbiomes differed markedly. Although both cases and follow-ups were dominated by Bacteria (relative abundance = 82.0% and 84.4%, respectively) with fewer Archaea or Eukarya, the members of this kingdom comprising the respective microbiomes were distinct. For example, during infection, cases contained a high proportion of members in Proteobacteria (37.1%) and displayed a starkly decreased relative abundance of phyla known to contain beneficial commensals such as Bacteroidetes (29.6%) and Firmicutes (13.7%) (Figure 3.8). Contrastingly, these beneficial phyla re-establish themselves during recovery, and are much more prevalent in follow-ups (Bacteroidetes, 49.3%; Firmicutes, 26.9%). Notably, the relative abundance of Proteobacteria in follow-ups is much lower than in infected cases (3.9%).



**Figure 3.8. Relative abundance of microbial phyla notably differs between cases and follow-ups.**

The top-10 microbial phylum with the greatest average relative abundance among cases or follow-ups is shown for each health status with each column representing the microbiome from one individual. Columns are ordered by their sample pairing, meaning that the column position for each facet of the plot refers to the same individual either during (Case; Top) or after (FollowUp; Bottom) enteric infection. Relative abundances were determined using raw gene abundances that had been normalized by the approximate number of genome equivalents in the sample as determined using MicrobeCensus.

Unfortunately, for both cases and follow-ups, approximately 50% of the reads assigned to genera could not be fully classified at this level; these reads, rather, are likely accounted for at a higher taxonomic level (e.g., class or phylum), but were too ambiguous to assign. However, when we consider those that could be assigned to specific genera, key differences were noted among cases and follow-ups (Figure B.13). Interestingly, the most prevalent genus in both cases

and follow-ups was *Bacteroides* (14.5% and 18.7%, respectively). In cases, this was followed by two prominent members of the *Enterobacteriaceae* family within Proteobacteria: *Salmonella* (7.1%) and *Escherichia* (5.0%). The next highest relatively abundant genera in cases also belonged to Proteobacteria: *Pseudomonas* (2.8%). Follow-ups, on the other hand, contained larger proportions of beneficial genera from the Bacteroidetes phylum such as *Alistipes* (5.0%) and *Prevotella* (2.5%). They also had high signatures of *Akkermansia* (2.8%), a beneficial microbe within the phylum Verrucomicrobia.

#### *Covariate-controlled batch effect adjustment and differential abundance testing with MMUPHin*

Beginning with ARGs, multiple differentially abundant classes and groups were identified among cases and follow-ups (Figure B.14). Cases were primarily represented by ARGs conferring resistance to multiple classes such as multi-metal resistance (coef= -0.243; adjusted p=1.04e-04), drug and biocide resistance genes (coef= -0.243; adjusted p= 1.46e-03), and drug, metal, and biocide resistance ARGs (coef=-0.212; adjusted p=7.86e-09). Interestingly, at the group level, the most differentiating ARG group for cases was *rpoB* (coef= -0.123; adjusted p=6.30e-05), which confers resistance to rifampin, followed by *mdtC* (coef= -0.103; adjusted p=4.97e-09) for MDR. Fluoroquinolone resistance genes were also more common in cases, both at the class level (coef= -0.168; adjusted p= 8.19e-10) and via groups such as *parC* (coef= -0.102; adjusted p= 3.90e-11) and *gyrA* (coef= -0.101; adjusted p=7.38e-08). Follow-ups, on the other hand, were defined by a different set of resistance genes. Specifically, tetracycline resistance genes were strongly represented at the class level (coef=0.352; adjusted p=2.26e-05), with much of this trend being driven by *tetQ* (coef=0.30; adjusted p=6.56e-05). MLS resistance genes were the next most representative class for follow-ups (coef=0.251; adjusted p=1.49e-25), with MLS groups such as MLS23S (coef=0.172; adjusted p=5.54e-06), *mefE* (coef=0.08;

adjusted  $p=3.54e-07$ ) and *ermF* (coef=0.07; adjusted  $p=3.68e-08$ ) serving as primary contributors. Lastly, aminoglycoside resistance differentiated follow-ups (coef=0.118; adjusted  $p=7.86e-09$ ) with genes such as *ant(6)* (coef= 0.103; adjusted  $p=5.23e-04$ ) and A16S (coef= 0.092; adjusted  $p=5.14e-04$ ).

Various taxa were also identified as defining features of cases or follow-ups (Figure B.15). Only one phylum defined the cases: Proteobacteria (coef= -0.461; adjusted  $p=9.35e-28$ ). The connection between Proteobacteria and case status was strongest among all associations. At the genus level, for instance, *Escherichia* (coef= -0.156; adjusted  $p=0.0021$ ) was dominant among cases, which is a trend driven primarily by *Escherichia coli* (coef=-0.146; adjusted  $p=0.0082$ ). *Escherichia* was followed by multiple members of Proteobacteria such as *Shigella* (coef= -0.057; adjusted  $p=0.0059$ ), which was represented by three species (*Shigella sonnei*, *Shigella flexneri*, and *Shigella dysenteriae*), as well as *Enterobacter* (coef= -0.020; adjusted  $p=1.10e-08$ ), and *Citrobacter* (coef= -0.017; adjusted  $p=8.07e-06$ ). As expected, follow-ups had a greater number of taxa with higher abundance. At the phylum level, follow-ups were heavily defined by Bacteroidetes (coef=0.305; adjusted  $p=1.87e-05$ ) and Firmicutes (coef=0.199; adjusted  $p=4.61e-07$ ). At the genus level, Firmicutes comprised the most differentially abundant taxa that included *Roseburia* (coef=0.050; adjusted  $p=6.28e-05$ ), *Dialister* (coef=0.038; adjusted  $p=0.0036$ ), and *Ruminococcus* (coef=0.037; adjusted  $p=2.83e-06$ ). *Phocaeicola*, a member of Bacteroidetes, was also highly represented in the follow-ups (coef=0.037; adjusted  $p=1.82e-08$ ) and was primarily represented by the species *Phocaeicola vulgatus* and *Phocaeicola dorei*. *Akkermansia* (coef=0.033; adjusted  $p=0.0069$ ), a member of the Verrucomicrobia phylum, was also a defining genus for follow-up samples. For the most part, the overarching patterns observed in MMUPHin were reflected by the ANCOM-BC method at each level of comparison (Phylum,

Genus, ARG Class, ARG Group) (Figure B.16) with observable differences in rank of correlation among various features (e.g., *Salmonella* was identified as a defining genus for cases in each method, but with differing status in comparison to *Escherichia*).

## **Co-occurrence network analysis reveals connections between taxa and ARGs**

### *Global network construction*

First, a global network was constructed considering ARGs and taxa in cases and follow-ups separately. Among cases, the network constructed of genera and ARG groups was first subset to include only nodes that had at least one connection (Degree  $\geq 1$ ); with this setting, there were a total of 587 nodes and 9,996 connections. Within the global network, multiple smaller, localized networks appeared, primarily among ARGs or genera, with various connections bridging these two features (Figure B.17A). Overall, the strongest correlations were observed among various ARG groups and just a handful of genera (indicated by warmer colored edges in the graph). To investigate the associations between ARGs and taxa in greater detail, a filter was applied in which all ARG-ARG and ARG-taxa connections were included (excluding taxa-taxa connections (Figure 3.9A). Specifically, *Escherichia* and *Salmonella* were of higher abundance (denoted by node size) with connections between multiple ARGs. The strongest ARG associations for *Escherichia* included multiple drug and biocide resistance genes such as *mdtF*, *mdtN*, *mdtP*, and *gadX* (Table B.2). Additionally, a handful of acid resistance (*gadC* and *hdeB*) and metal resistance genes (*rcnB* and *rcnA*) were also highly correlated with *Escherichia*. *Salmonella* displayed strong associations primarily with ARGs related to metal and biocide resistance; these included *cueP*, *golt*, *sitA* and *sitD*, and *smvA*. Another genus very relevant to cases, *Shigella*, was also a main player in the ARG-taxa connections. Similar to *Escherichia*, this genus was highly correlated with drug and biocide resistance genes such as *mdtN*, *mdtF*, and

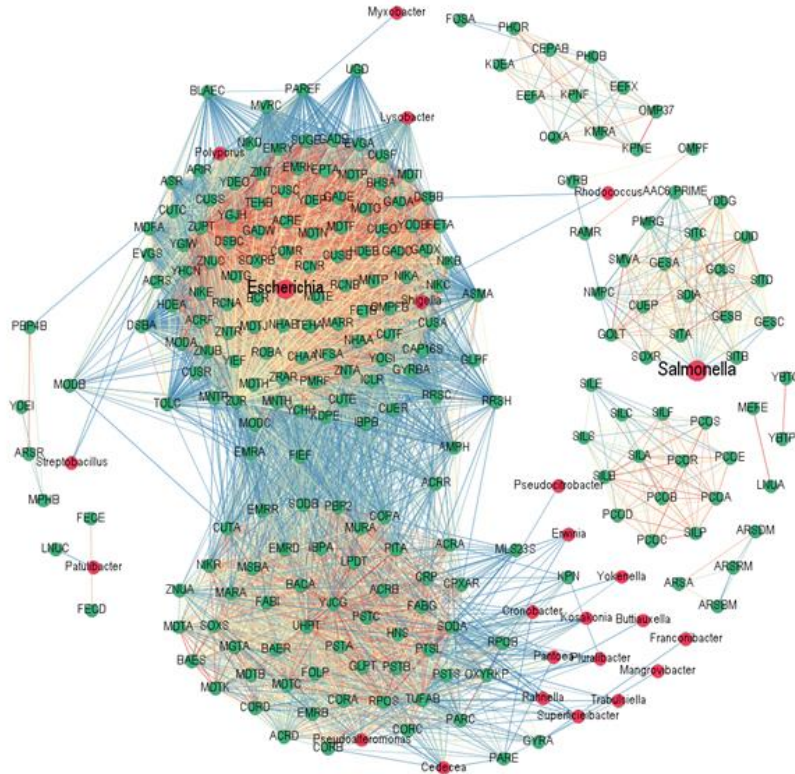
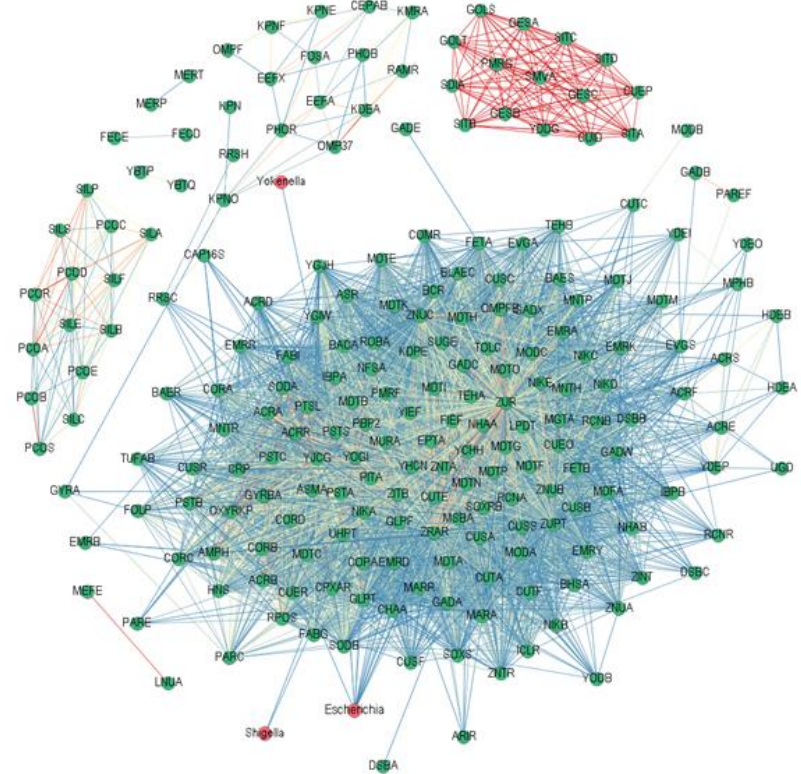
*mdtO*. Various metal resistance genes were also highly correlated with *Shigella* including *mntP*, *tehA*, and *ygiW*. In fact, the co-occurrence of *Escherichia* and *Shigella* was also highly correlated (coef=0.946; adjusted-p < 0.0001). Other genera that were found to be associated with ARGs in this subsetted co-occurrence network among cases included *Pseudoalteromonas*, *Lysobacter*, *Cronobacter*, and *Cedecea*, all of which are also members of Proteobacteria.

At the global level, follow-ups displayed somewhat different connections among ARGs and taxa. When considering the initial network with a degree cutoff  $\geq 1$ , we observed fewer overall nodes and edges compared to cases (nodes=400; edges=7,914). In addition, there are fewer large sub-networks displayed among follow-up samples than cases (Figure B.17B), though these local networks between ARGs and taxa are still evident. In follow-ups, the strongest correlations were displayed among resistance genes. When isolating the sub-networks that are strictly relevant to ARG-ARG and ARG-taxa connections, the degree cutoff of 1 was sufficient to observe relevant connections. In follow-ups, three genera had ARG connections including *Escherichia*, *Shigella*, and *Yokenella* (Figure 3.9B). *Escherichia* displayed the most connections (Table B.3), the most highly correlated of which was *Shigella* (coef=0.937; adjusted-p < 0.0001). The ARGs with which *Escherichia* co-occurred included *uhpT*, a fosfomycin resistance gene, *nika*, which confers nickel resistance, *ychh* and *yhcn*, both of which confer biocide and metal resistance, and *mdtN*, a drug and biocide resistance ARG that was also associated with *Escherichia* in cases. *Shigella*, in addition to its connection to *Escherichia*, also displayed notable correlations with *nika* (coef=0.805; adjusted-p < 0.0001) and *uhpT* (coef=0.801; adjusted-p < 0.0001). Intriguingly, network analysis classified just one connection between *Yokenella* and ARGs; *copA*, which plays a role in copper resistance, was found to co-occur with this genus.

Intriguingly, various patterns emerge for various ARG-ARG connections across both cases and follow-ups. For example, each group contains a subnetwork composed of *pco* cluster genes (e.g., *pcoA*, *pcoB*, *pcoE*, etc.) that confer copper resistance as well as *sil* cluster genes (e.g., *silA*, *silB*, *silC*, etc.), which have demonstrated multi-metal resistance. Additionally, both cases and follow-ups have a highly correlated sub-network containing ARGs for multiple classes (e.g., drug and biocide resistance, multi-metal resistance, biocide and metal resistance, etc.). In cases, this cluster is associated with *Salmonella*; conversely, in follow-ups, this cluster is not associated with any taxa but displays high levels of correlation between ARGs. Another important finding related to both networks is the lack of association attributed to the genus *Bacteroides*. Although members of this genus are quite prevalent in both the cases and follow-ups, no ARG-*Bacteroides* associations were identified in this analysis at the  $\rho \geq 0.80$  level of significance.

#### *Co-occurrence networks relevant to beta-lactam ARGs*

The widespread prevalence of beta-lactam resistance motivated our analysis of beta-lactam resistance gene connections within our networks. For cases and follow-ups, all network nodes were subset based on their association with beta-lactam ARGs; any genera or other ARGs that correlated with beta-lactam resistance genes were thus included. For both cases and follow-ups, a degree cutoff of 2 was established to clarify the most prominent connections among features. In cases, six beta-lactam ARGs contributed to the network, with four of these (*blaEC*, *ompFB*, *pbp2*, and *ampH*) demonstrating high levels of interconnectivity among ARGs and genera (Figure B.18A). Of these, *ompFB*, which encodes a mutant porin protein relevant to

**A****B**

**Figure 3.9. Global network analysis highlighting ARG connections among cases and follow-ups.**

Correlation co-occurrence networks were constructed in Gephi 0.9.2 using Spearman's Rank correlation coefficients generated with the R-package 'Hmisc' (v4.5-0) for cases (A) and follow-ups (B). These networks display all ARG-ARG and ARG-taxa connections; taxa-taxa connections were excluded for clarity. Correlations included in the network all passed a cutoff of  $p > 0.80$  and  $q\text{-value} < 0.05$ . Nodes are colored by their identity as a taxonomic genus (red) or ARG group (green). Nodes are sized based on their overall abundance among samples; the larger the node, the more abundant. Nodes with  $\geq 1$  connection were included (i.e. degree cutoff=1). The edge color displays the strength of correlation, with blue demonstrating relatively weaker correlations (yet still  $> 0.80$ ), yellow showing medium correlation, and red showing strong correlation. Nodes are labeled with their corresponding genus or ARG group.



beta-lactams, displayed the most connections (n=114) with the highest average correlation coefficient (0.926). *pbp2* was next highest, with 98 total connections and an average correlation of 0.886 among its connections. Notably the non-beta-lactam ARGs associated with these genes ranged greatly in class, though it is noteworthy that ARGs resistant to multiple classes (*mdt* and *acr* gene clusters, for example) were included in the network. Two genera, *Escherichia* and *Shigella*, were also found to be correlated with *blaEC*, *ompFB*, and *pbp2* and *ompFB* and *ampH*, respectively.

A similar pattern was observed in the follow-ups as the same four beta-lactam ARGs are highly connected within the network (Figure B.18B). However, in this group, *pbp2* contains the highest number of connections (n=127) and the greatest average correlation coefficient across connections (coef=0.883). Multiple ARGs relevant to different classes were also detected in addition to ARGs conferring resistance to different types of metals such as nickel resistance (e.g. *nikA* and *nikC*) and copper resistance (e.g. *cutE* and *copA*). Various multi-metal resistance genes are also prevalent (e.g., members of the *cor* and *mnt* gene clusters). Notably, the beta-lactam-specific network did not identify any taxa with correlations that met the cutoff of 0.80, a finding that differs from that of cases.

#### *Investigating global co-occurrence networks related to infectious pathogen*

For patients infected with *Salmonella*, some interesting patterns of ARG-ARG and ARG-taxa co-occurrence were observed. Notably, *Salmonella* itself was a prominent node within the network connected to multiple ARGs of interest (Figure B.19A). These ARGs spanned multiple classes including paraquat resistance (*yddG*, *nmpC*), gold resistance (*golT*, *golS*), copper resistance (*cuiD*, *cueP*), and multi-compound resistance which included resistance to biocides, metals, and drugs (*ges* and *sit* gene clusters, among others). The association of the pathogen with

these varied groups of ARGs is an interesting finding and may suggest a relatively involved role of the pathogenic microbe in resistance dissemination. Among *Salmonella*-infected cases, we also observed high connectivity for *Escherichia* and *Shigella* as well, each of which appeared highly associated with various multi-drug and multi-compound resistance genes. Another noteworthy finding among *Salmonella* cases is the prevalence of connections between *tet* genes and various taxa, a finding not as prominent in the overall global analysis.

For follow-ups recovering from *Salmonella* infection, ARG connections appear more diffuse and, overall, less prevalent (Figure B.19B). Another noticeable deviation from the network observed among cases is the lack of the *Salmonella*-centered subnetwork; in fact, *Salmonella* does not appear as a genus in the follow-up network at all. Interestingly, many of the ARGs that had previously been associated with *Salmonella* are still present and highly correlated to one another. Additionally, *Escherichia* and *Shigella* still appear as genera connected to various ARGs, though the strength and number of correlations is notably lower than in that of cases.

Among cases who were infected with *Campylobacter*, we observed a dense network with an even more dense subnetwork composed primarily of ARGs (Figure B.20A). Some important patterns include the presence of *Escherichia* and *Shigella* in this tightly packed network of resistance. The ARGs in this subnetwork, again, are primarily associated with multi-drug or multi-compound resistance, with many belonging to the *mdt* family of genes, among others. Also of note is the prevalence, again, of various tetracycline resistance genes such as *tetW*, *tet32*, and *tet40*, each of which displayed many connections to various taxa in the network. A prominent deviation from the network of *Salmonella*-infected cases is the absence of *Salmonella* in this network, again suggesting that the pathogen was playing a more involved role in those cases.

Similar to those recovering from *Salmonella* infections, follow-ups recovering from *Campylobacter* infections displayed a more diffuse correlation network among taxa and ARGs than cases (Figure B.20B). Follow-ups still contained a very dense inner network composed of mostly ARGs, with *zur*, a zinc resistance gene, appearing to serve as a hub. Interestingly, *Escherichia* is still present and connected to various ARGs, but in lower numbers and reduced correlation compared to cases. *Shigella*, on the other hand, is absent from the follow-ups' network.

### **Host-tracking analysis**

#### *ARG-harboring microbial hosts detected in cases vs. follow-ups*

In cases, ACCs, on average, were primarily attributed to *Escherichia* (38.05% of case-associated ACCs) followed by *Salmonella* (18.31%) and *Klebsiella* (9.92%) (Figure 3.10). Interestingly, the most prominent genus represented in follow-up ACCs was also *Escherichia* (19.81%); however, the next most prevalent genera were *Bacteroides* (15.12%) and *Faecalibacterium* (5.99%) (Figure 3.11).

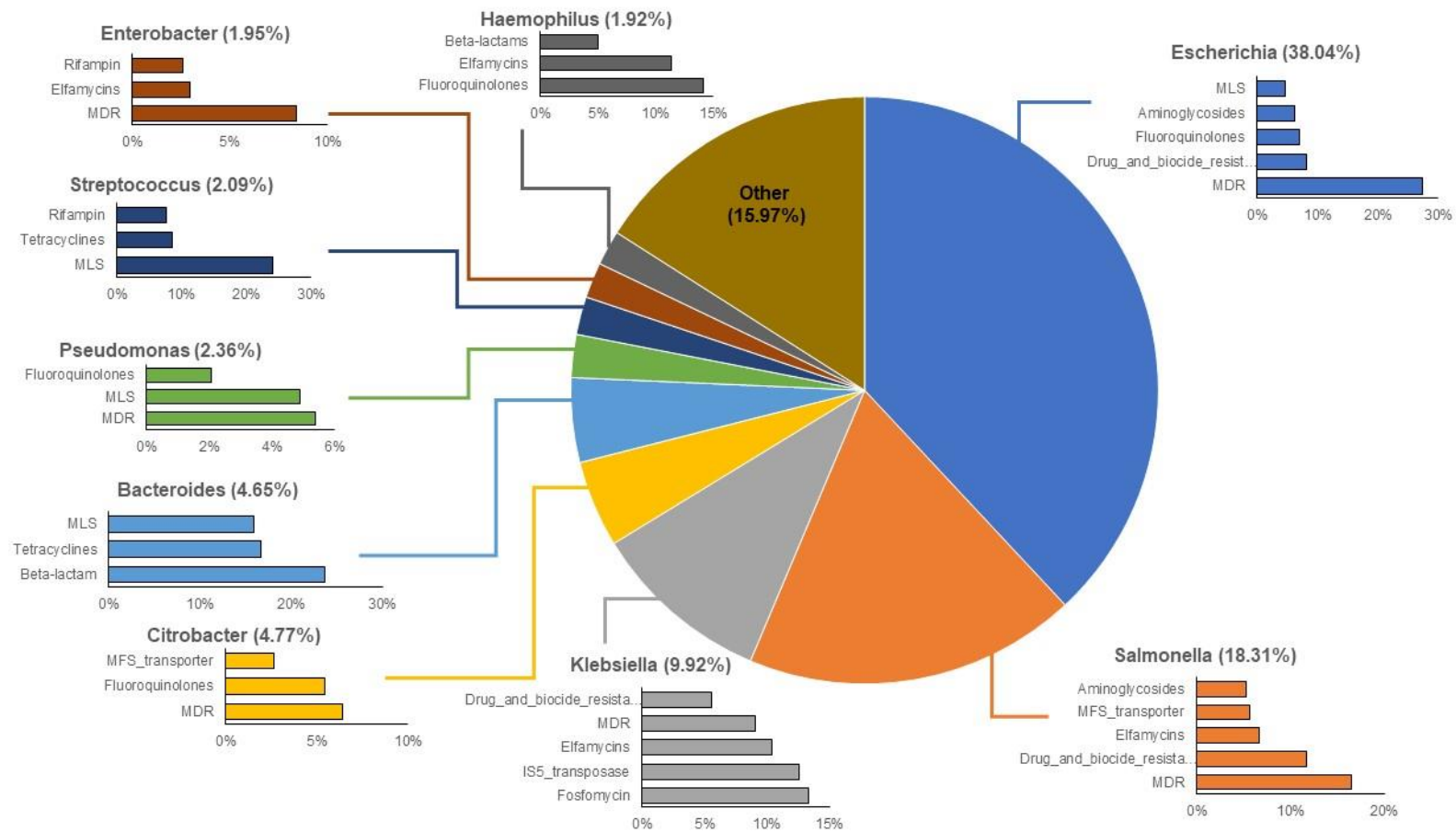
Following the identification of genera on the ACCs among cases and follow-ups, the most prevalent ARG classes attributed to these genera were determined. Of all ARGs assigned to *Escherichia* in cases, 27.4% were assigned to MDR on average. *Escherichia* also harbored ARGs relevant to drug and biocide resistance (8.12%), fluoroquinolone resistance (7.06%), and aminoglycoside resistance (6.21%). Of the ARGs on *Salmonella*-associated ACCs, MDR and drug and biocide resistance were most highly represented (16.5% and 11.7%, respectively). *Klebsiella* harbored an array of fosfomycin resistance genes (13.3%) followed by relevant transposase genes in the IS5 family (12.6%). This high representation of transposases suggests

increased potential for ARG mobility to and from this genus. In addition to these ARGs, *Klebsiella* contained ARGs for elfamycin resistance (10.4%) and MDR (9.08%).

Although, similar to cases, ACCs among follow-ups were primarily attributed to *Escherichia* (19.81%), the other genera harboring ARGs in recovered patients were more divergent. For example, *Bacteroides* contained 15.12% of all identified ACCs, carrying ARGs related to MLS, beta-lactam, and tetracycline resistance. *Faecalibacterium* harbored 5.99% of all ACCs, with an “Uncultured” taxon holding 5.21%. In *Escherichia*-associated ACCs among follow-ups, the array of ARGs harbored was nearly identical to cases; MDR genes predominated (25.1%), followed by resistance to drugs and biocides (4.71%), fluoroquinolones (4.70%), and aminoglycosides (3.84%).

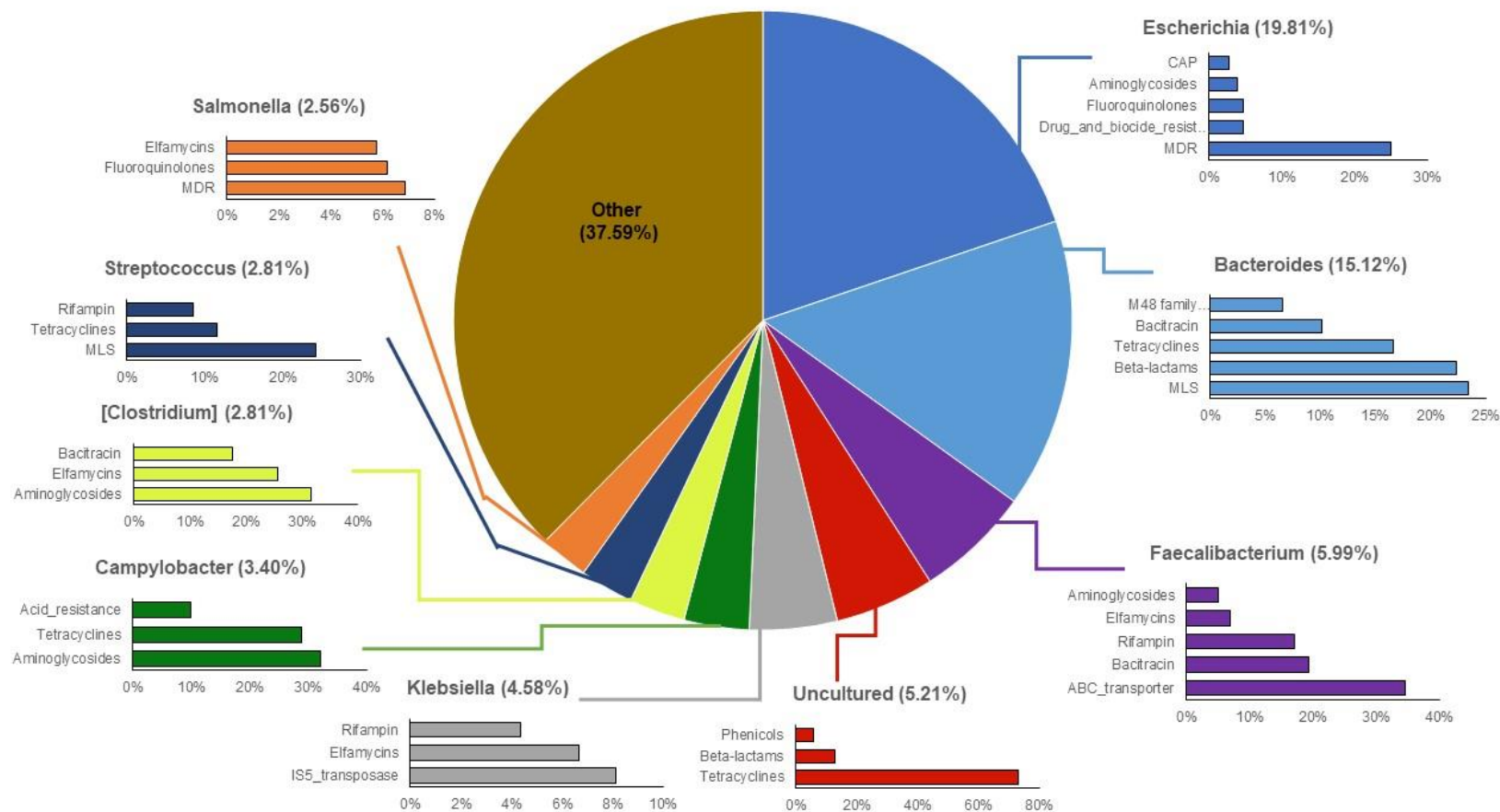
#### *Comparing across enteric pathogens*

When considering these results in the context of infecting pathogen, some interesting trends were observed. Among the cases infected with *Campylobacter* (n=23), the genera comprising the greatest proportion of ACCs were *Escherichia* (42.84%), *Klebsiella* (10.01%), and *Salmonella* (7.09%). Upon recovery, however, these proportions changed markedly. Among *Campylobacter* follow-ups, *Bacteroides* was most highly represented on ACCs (18.34%), followed by *Escherichia* (17.31%) and *Faecalibacterium* (6.76%). Interestingly, both *Klebsiella* and *Salmonella* remained in the top-20 genera comprising ACCs in follow-ups, but their average proportion was markedly reduced (3.59% and 1.28%, respectively). It is notable, too, that *Campylobacter* registered in the top-20 genera represented on ACCs as well, with proportions of 1.96% and 3.81% in cases and follow-ups, respectively. In cases, the ARGs harbored by *Campylobacter* were primarily tetracycline resistance genes (27.6%) followed by genes for aminoglycoside (9.92%) and rifampin resistance (8.31%). In the follow-ups, tetracycline



**Figure 3.10. Host-tracking via investigation of ARG-carrying contigs reveals genera responsible for harboring ARGs among cases.**

The top-10 genera assigned to ACCs for cases are indicated in the respective pie charts. The percentages associated with each genus name indicate the percent of ACCs that were assigned to that genus. For example, *Escherichia* (38.04%) indicates that 38.04% of all ACCs among cases were annotated as *Escherichia*. Each bar chart associated with a genus displays the top-5 or top-3 ARG classes affiliated with that particular genus on the ACCs. E.g., nearly 30% of all ARGs attributed to *Escherichia* ACCs in cases were classified as multi-drug resistance genes or “MDR”.



**Figure 3.11. Host-tracking via investigation of ARG-carrying contigs reveals genera responsible for harboring ARGs among follow-ups.**

The top-10 genera assigned to ACCs follow-ups are indicated in the respective pie charts. The percentages associated with each genus name indicate the percent of ACCs that were assigned to that genus. For example, *Escherichia* (19.81%) indicates that 19.81% of all ACCs among follow-ups were annotated as *Escherichia*. Each bar chart associated with a genus displays the top-5 or top-3 ARG classes affiliated with that particular genus on the ACCs. E.g., about 25% of all ARGs attributed to *Escherichia* ACCs in follow-ups were classified as multi-drug resistance genes or “MDR”.

resistance genes (29.0%) predominated on *Campylobacter*-associated ACCs, followed by genes for aminoglycoside resistance (27.0%). The third most prevalent class of ARGs were linked to MLS resistance (10.3%), representing a notable shift from cases. Of note, too, is the relative increase in resistance to aminoglycosides among *Campylobacter* ACCs in follow-ups; this class experienced a 172% increase throughout the duration of recovery.

In people infected by *Salmonella* (n=29), the most highly represented genera on ACCs included *Escherichia* (32.39%), *Salmonella* (30.96%), and *Klebsiella* (7.89%). The ACCs in follow-ups indicated a similar trend as patients recovering from *Campylobacter* infection mostly had *Escherichia* (20.66%), *Bacteroides* (14.16%), and *Faecalibacterium* (6.29%), which were the top three most prominent genera. As noted, the genus relevant to the infecting pathogen, *Salmonella*, was prominent among ACCs in the sample taken during infection. In cases, the ARGs found in *Salmonella* represented multiple classes relevant to multi-compound resistance: drug and biocide resistance (14.1%), MFS transporters (13.1%), which can have MDR effects or high specificity to certain classes as well as drug, biocide, and metal resistance (7.61%). Among follow-ups, the most prevalent class harbored by *Salmonella*-associated ACCs was RND efflux transporters (9.29%), which, like MFS, can either confer resistance to multiple classes or a specific antibiotic class. MFS transporters (6.84%) and fluoroquinolone resistance genes (6.31%) were also prevalent.

Patients with *Shigella* infections (n=4) displayed a comparable list of the most represented taxa; ACCs in cases were dominated by *Escherichia* (60.5%), with less prevalent signatures of *Klebsiella* (17.43%) and *Bacteroides* (4.77%). By contrast, the follow-ups (n=4) had more similar proportions of *Escherichia* (16.84%), *Bacteroides* (13.57%) and *Citrobacter*

(7.31%). *Shigella* made up just 0.50% of genera attributed to ACCs in cases and was not found among ACCs in follow-ups (0%).

Finally, among the individuals infected with STEC (n=3), *Escherichia* predominated among ACCs (25.93%), followed by *Klebsiella* (18.78%) and *Pseudomonas* (14.34%). In follow-up samples, *Escherichia* and *Klebsiella* were also among the most common ACCs, registering proportions of 28.07% and 17.20%, respectively. Unlike the case samples, *Bacteroides* was the third most represented genus among the follow-up ACCs (7.23%).

#### *Investigating the potential persistence of clinically relevant ESBLs post-recovery*

The host-tracking analysis enabled us to investigate various genes of interest across paired cases and follow-ups. Of paramount concern are the extended-spectrum beta-lactamases (ESBLs), which are highly mobile ARGs conferring resistance to a broad range of beta-lactam antibiotics (58). Our analysis detected multiple beta-lactamase genes (n=37), some of which were classified as ESBLs (n=14; 37.8%). In some cases, unfortunately, the resolution of identification for the specific beta-lactamase was inadequate, resulting in general hits entitled, simply “beta-lactamase.” In other instances, the particular class of beta-lactamase was provided, but with no further information about the ARG detected in the contig (e.g. “class A beta-lactamase”). Regardless of this irregular resolution among genes, comparisons were pursued between paired case and follow-up samples to explore the potential transmission and persistence of these genes among different taxa and identify patterns relating to type of infection (i.e., connecting pathogen type to beta-lactamases).

Among the 14 ESBL genes detected, one was classified as the TEM-1 variant, which was linked to *Escherichia* but was found in just a single sample from a case infected with *Salmonella*. Over the course of recovery, this gene was “lost” (meaning it only appeared in the case sample



but was not present in the paired follow-up). In addition, the gene encoding an ADC family of class C ESBL was harbored and subsequently lost by two cases. Among all samples, the gene encoding a CepA family ESBL, which was highly prevalent in 19 cases and 13 follow-up samples, was linked to *Bacteroides*. Of these 19 cases, 9 lost the gene by follow-up, 10 maintained it as it was detected in both paired samples, and 3 individuals acquired it during the recovery period (present in follow-up sample but not case).

Moreover, genes representing the OXY family of class A ESBLs were detected in *Klebsiella* among four cases; this gene family was not found in any follow-up samples, indicating that it was “lost” during recovery. While multiple OXA genes were detected, the family of OXA genes varied substantially and each was attributed to a different microbial host. For example, the gene encoding an OXA-1 class D beta-lactamase was carried by *Klebsiella* in two cases, both of which lost this gene during recovery. The OXA-50 family of genes, however, was detected in *Pseudomonas* that was present in two cases but was lost in the paired follow-up samples. Two different cases harbored genes for the OXA-51 family of carbapenem-hydrolyzing class D beta-lactamases, which were also lost. Notably, the OXA-61 family of class D beta-lactamases was harbored by *Campylobacter* but was only found in two cases with infections caused by *Campylobacter*, insinuating the pathogens role in harboring ARGs. Finally, *Klebsiella* was also observed to possess genes for the SHV family of class A beta-lactamases, which were found in eight cases and two unpaired follow-ups, indicating that all eight cases lost the gene and two follow-ups acquired it. Of the ESBLs explored, there was not enough evidence to infer transmission of these genes between various taxa. Since many of the ESBLs were present in cases but not follow-ups (i.e., “lost”), we could not assess whether these ESBLs were transferring among bacteria during recovery from infection.

In addition to the observed ESBLs, other relevant beta-lactamases were also identified among our samples (Table B.4). A prevalent beta-lactamase gene belonging to the BlaEC family of class C beta-lactamases, for instance, were primarily attributed to the genus *Escherichia*. A total of 49 cases and 19 follow-ups harbored this ARG; the ARG was lost in 35 cases, maintained in 14, and acquired in 5 follow-ups. The BlaEC family was found in samples from patients with all four enteric pathogens; 28 were detected in *Campylobacter* cases and 30 were found in cases with *Salmonella* infections along with 5 in *Shigella* and 5 in STEC case samples. This gene family was also detected in *Shigella* in one case which was infected by *Shigella*; however, this gene was lost during recovery, providing no evidence of inter-genus transmission.

Another prominent beta-lactamase of clinical importance was the CfxA family of class A broad-spectrum beta-lactamases. These genes were primarily harbored by *Bacteroides*, but also appeared in *Prevotella*. Among these *Bacteroides*-associated ARGs, 46 were found in cases and 48 in follow-ups. Although only 7 of these genes were lost by cases, 39 were maintained and 9 were acquired during recovery. Most samples related to each of the infecting pathogens contained this gene: 37 for *Campylobacter*, 45 for *Salmonella*, and 6 for both *Shigella* and STEC infections. A somewhat similar trend was found for the CfxA genes harbored by *Prevotella*, though at a smaller scale. In this case, 5 cases contained these ARGs compared to 9 follow-ups. Three of the cases lost these genes while 2 maintained them; meanwhile, 7 follow-ups acquired these ARGs during recovery. Seven of these were attributed to *Campylobacter* infections, 6 to *Salmonella*, and one to *Shigella*. Interestingly, there is evidence of potential transmission of these CfxA genes between *Bacteroides* and *Prevotella*. For example, there are 6 separate case-follow-up pairs in which the CfxA gene(s) appear as “acquired” by *Prevotella* in follow-up samples and also maintained by *Bacteroides* in follow-up samples, suggesting potential *Bacteroides*-to-

*Prevotella* transmission. There are also two case-follow-up pairs in which these CfxA genes are maintained in both *Bacteroides* and *Prevotella* during recovery, which may or may not suggest potential transfer. Interestingly, there are also three instances in which the *Prevotella*-harbored ARG is “lost” and the *Bacteroides*-harbored CfxA is maintained during recovery, also showing potential of *Prevotella*-to-*Bacteroides* transmission.

The CMY-2 family of class C beta-lactamases was also identified and was harbored by *Citrobacter* and *Salmonella*. Among these ARGs harbored by *Citrobacter*, 8 were found in cases and 3 in follow-ups; 6 cases lost the gene while 2 maintained it and 1 follow-up acquired it. Five of these instances were related to *Campylobacter* infection, 5 to *Salmonella*, and one to STEC. Of the CMY-2 ARGs harbored by *Salmonella*, 2 were found in cases (each of which were lost) and 1 was acquired in a follow-up sample. One of these cases was infected with *Campylobacter* while the other two individuals had *Salmonella* infections. Our findings do not suggest transmission of the CMY-2 beta-lactamases between *Citrobacter* and *Salmonella*, as these occurred in separate case-follow-up pairs. However, the more broadly defined CMY-family of class C betalactamase was also identified and was assigned to *Salmonella* in 3 cases (all lost) and 2 follow-ups (both acquired). Although the CMY family is a broader category than the CMY-2 family of beta-lactamases, it is possible that the CMY family defined in our study contains CMY-2 genes relevant to this analysis. For example, there is one case-follow-up pair in which the CMY-2 family was maintained in *Citrobacter* and the CMY family was acquired in *Salmonella*; yet another case-follow-up pair indicated loss of the CMY family of beta-lactamases in *Salmonella* but maintenance and noted increase of the CMY-2 family in *Citrobacter*. Although loosely inferred, these data indicate the potential for transmission of CMY-family genes across genera.

Finally, genes for the general subclass A2 of class A beta-lactamases were found in *Bacteroides* among a majority of cases (n=45) and follow-ups (n=47); 7 cases lost the gene during recovery, while 38 maintained it and 9 follow-ups acquired it. In all, 39 were related to *Campylobacter* infections, 44 to *Salmonella*, 5 to *Shigella*, and 4 to STEC. The more general “class A beta-lactamase” gene was also found in nine other genera including *Atlantibacter*, *Bacillus*, *Burkholderia*, *Clostridium*, *Proteus*, *Salmonella*, *Yersinia*, *Escherichia*, and *Klebsiella*. Although there is a slight difference in resolution of these identified features, it is still interesting to consider the potential for transmission across genera. For example, 8 paired samples demonstrated potential transmission of class A beta-lactamase genes; six of these pairs indicate transfer between two different genera, 4 of which had *Campylobacter* infections and 2 had *Salmonella* infections. The remaining 3 showed acquisition of class A beta-lactamase genes in *Clostridium*, while the subclass A2 genes were maintained in *Bacteroides*. Of the other three, one contains an *Escherichia*-acquired ARG and a *Bacteroides*-maintained ARG, one has a *Burkholderia*-acquired beta-lactamase and a *Bacteroides*-maintained gene, and the last holds a *Proteus*-lost gene and a *Bacteroides*-maintained gene. The final two case-follow-up pairs, each of which had a *Salmonella* infection, had loss-maintained-acquisition patterns that involve multiple genera. One pair demonstrates loss of class A beta-lactamase genes from *Atlantibacter* and *Salmonella*, with acquisition in *Escherichia* and maintenance by *Bacteroides*. The other case-follow-up pair shows loss of the class A betalactamase in *Klebsiella* and *Clostridium*, with acquisition by *Bacillus* and maintenance by *Bacteroides*.

## DISCUSSION

The human gut microbiome, when disrupted by an infectious pathogen, can drastically change in its composition taxonomically, genetically, and even functionally (59). In most

instances, pathogen invasion leads to a state of dysbiosis in which the infected individual experiences a state of gastrointestinal distress linked to a dramatic decrease in gut microbiota diversity (3, 60). The findings of our present study confirm this, as stools of patients infected with an enteric pathogen (*Campylobacter*, *Salmonella*, *Shigella*, or STEC; ‘cases’) displayed markedly lower microbiome diversity than stools of these same individuals after recovery from infection (follow-ups). However, this study also sought to characterize the inverse impact of enteric infection on the gut resistome, as infected cases showed much greater diversity of resistance genes compared to recovered follow-ups. These findings were presumed to be linked, as shifts in microbial composition inherently influence the presence and abundance of ARGs harbored by microbes within the gut. As our later analyses exploring microbial hosts of ARGs demonstrate, this hypothesis was indeed correct.

In concordance with our previous findings that investigated the impact of *Campylobacter* infections on the human gut resistome (5), this study displayed noticeable shifts in the microbiome and resistome during enteric infection and after recovery. Namely, cases had more multi-compound and multi-drug resistance genes than follow-ups, who, in general, had more tetracycline, MLS, and aminoglycoside resistance genes. This range of resistance observed among follow-ups is consistent with patterns of resistance documented in individuals deemed “healthy” across a range of studies (5, 28, 61), suggesting that, on average, these follow-up samples demonstrated a return or near-return to pre-infection gut health. The observed shifts in microbiome composition, too, suggest that most follow-ups were well on their way to recovery. This was evidenced by the dramatic decrease in microbiome diversity among cases during infection, and the demonstrated increase in diversity during or after recovery. Some specific taxonomic signatures also suggested that follow-ups were returning to pre-infection health, as

these individuals contained higher abundances of notoriously beneficial commensals in the Bacteroidetes and Firmicutes phyla, namely *Bacteroides*, *Prevotella*, and *Phocaeicola* and *Faecalibacterium*, *Roseburia*, and *Ruminococcus*, respectively. These microbes have been shown to play influential roles in maintaining gut homeostasis and metabolic health (62-64). Conversely, cases were defined primarily by members of Proteobacteria such as *Escherichia*, *Salmonella*, *Shigella*, and *Klebsiella* which have been associated not only with acute enteric disturbance but also prolonged dysbiosis and disease (20, 65, 66).

Intriguingly, several follow-up samples (n~5) clustered more closely with cases based on both the resistome and microbiome composition analyses (Figures 3.2 & 3.4). Although number of follow-days was explored as a potential driver of these clustering patterns, the average number of days since infection among these five samples (n=110) was very similar to the overall mean (n=108). Additionally, notable trends were not observed when stratifying by sex, age, pathogen, residence type, and care status (e.g., hospitalized vs. outpatient); however, two of the five individuals were 10 years of age or under and two were  $\geq 53$  years or older. As these populations, historically, can be considered at higher risk of more severe disease (67-69), it is likely that these patients were predisposed to experiencing longer effects of infection than other members of the sample cohort. In addition, two of these 5 follow-up samples were from cases who were hospitalized at the time of their acute infection, suggesting they may have had more severe disease, which could contribute to a longer recovery time. It is also not clear whether most follow-up samples came from patients after a full recovery, as we were unable to evaluate the gut composition of patients prior to infection. It is likely that the state of the microbiome prior to infection, in addition to its resilience to disturbances, will greatly impact the trajectory of disease and subsequent recovery among individuals (70). Indeed, implementation of a more rigorous

longitudinal study is needed to further understand the intricacies of recovery from enteric infection.

When first attempting to understand the interplay between microbes and ARGs within the gut environment, co-occurrence networks served as a useful predictive tool. There were many significant correlations with taxa-taxa associations dominating. This can partially be explained by the level of resolution which was used in our analysis; we considered taxonomic genus with ARG group; since there are far more taxa than ARGs, these global networks appeared skewed. Although our focus on the most prevalent ARGs and taxa may not fully represent the co-occurrences among samples, especially among rare microbial features, these analyses provided useful insight into ARG-taxa co-occurrence. Indeed, our correlation networks suggested that these microbial features shift together. Of particular interest are the increased levels of taxa-ARG connections observed among cases during enteric infection; although some of these connections were conserved in the follow-up samples, it is intriguing that fewer of these correlations were identified overall. Some of this result may be due to the inherently higher abundance of specific taxa within cases such as *Escherichia*, *Salmonella*, and *Shigella*, each of which appear to play significant roles in the global network. These members of Proteobacteria, for example, harbored resistance genes to separate classes including beta-lactams, trimethoprim, tetracyclines, fluoroquinolones, and aminoglycosides, among others, but have also been documented to harbor MDR genes (71-75). Of utmost concern is the widespread presence and spread of ESBL genes, which are commonly found in members of Proteobacteria, especially in the family Enterobacteriaceae (76, 77). Many ARGs harbored by genera such as *Escherichia*, *Klebsiella*, and *Shigella* are commonly found on multiple plasmid families specific to Enterobacteriaceae (78). *Escherichia*, specifically, serves as an important carrier of resistance to multiple classes of

antibiotics in a variety of organisms (79-82). Despite the role of *Escherichia* as a commensal in the gut that blooms after enteric pathogen invasion, its potential to acquire clinically important ARGs and emerge as a resistant pathobiont is concerning (11, 25).

An interesting aspect of the global network constructed for cases was the presence of a *Salmonella*-specific subnetwork comprised of multiple metal, biocide, and multi-drug resistance genes (Figure B.19). To investigate whether these *Salmonella* signatures could be evidence of the pathogen itself, separate networks were constructed for individuals infected with *Salmonella* and *Campylobacter*, respectively; unfortunately, there were not enough samples relevant to *Shigella* or STEC infections to generate correlations matrices (n=4, n=3, respectively). Notably, *Salmonella* was only a prominent contributor to the network associated with *Salmonella*-infected cases; this subnetwork did not appear in the *Campylobacter*-infected group, suggesting the pathogen itself was at play in harboring and potentially disseminating ARGs. Previously, pathogenic *Salmonella* was indeed detectable via metagenome analyses (6), and so the detection of pathogenic *Salmonella* among these individuals is precedented. Additionally, *Salmonella* has previously been shown to harbor resistance genes for antibiotics, disinfectants, and heavy metals (83). Unfortunately, *Salmonella* is not alone; in fact, previous documentation has indicated co-selection for resistance to these three compound types (antibiotics, metals, and biocides) across many genera (84) including multiple foodborne pathogens (85), a trend which we see within our own networks as well. There is also evidence that this broad array of resistance likely develops in the environment, where metal pollution is most common, and subsequently spreads to human pathogens via HGT (86). Indeed, the relevant co-occurrence of such resistance with pathogens or potentially pathogenic pathobionts among our samples raises concerns regarding the role of enteric infection in furthering the spread of these resistance genes.



A noteworthy result of our co-occurrence networks was the absence of some of the most abundant taxa among our samples, namely members of Bacteroidetes and Firmicutes such as *Bacteroides*, *Prevotella* and *Alistipes* and *Faecalibacterium* and *Roseburia*, respectively. These microbes were more abundant in follow-up samples, yet still had measurable signatures within our cases. Historically, members of Bacteroidetes and Firmicutes have been associated with high levels of tetracycline and erythromycin resistance, carrying resistance genes *tetQ*, *tetO*, *tetW*, and *ermF* and *ermT* (87, 88). The trends observed in our relative and differential abundance analyses would suggest that we would observe co-occurrence of these ARGs and taxonomic groups as well. Interestingly, we do observe *Bacteroides* to be connected to *tetQ* in the network of cases infected with *Salmonella* (Figure B.19A). This network, too, contains the most connections relevant to tetracycline resistance genes like *tetO* and *tet16S* as well; the tetracycline resistance genes *tetW* and *tet32* were also observed in the network of cases infected with *Campylobacter* (Figure B.20A). It is odd, however, that these ARG-taxa connections are not present in our global networks considering cases and follow-ups, particularly in follow-ups where their abundance is relatively much higher. This notable absence may highlight some of the limitations of correlation co-occurrence analysis. For example, our correlations were based upon abundances normalized by the number of genome equivalents. However, increasing attention is being paid to the inherent compositionality of high-throughput sequencing data, an artifact which can taint interpretation of various statistical methods if not accounted for correctly (89). However, various methods, including data transformations, can be used to correct for compositionality (90). Although we sought to normalize our abundances by the number of genome equivalents per sample, thereby accounting for any discrepancies in microbial richness, further measures to enact data transformations and thus protect against compositional artifacts

were not taken. However, such methods similar to ours have been previously used to characterize co-occurrence of ARGs and taxa without such consideration for compositionality (27-29).

Additionally, the results observed in our co-occurrence network analysis were designed to be hypothesis-generating, rather than conclusive. By no means do we presume to state that all ARGs connected to taxa within the networks are indeed harbored by such taxa. Rather, the simple correlation between various ARGs and taxa is of interest given the ecological complexity of enteric infection. In any case, however, it is reassuring to note that various methods have been developed to perform more robust network analyses (91). Sparse Correlations for Compositional data (SparCC), for example, uses a log-transform approach to infer correlations between components in a dataset (92) and would be a sensible next-step if we were to follow up our present network analysis. Another potential explanation for the unexpected absence of Bacteroidetes and Firmicutes microbes in our networks is that the abundances of these taxa were not found to correlate with ARGs or other taxa that met our network cutoff. Again, these networks were constructed after first excluding 50% of the most sparse features, thereby removing genes and taxa that were not present in 10% and 20% of samples, respectively. This is a relatively low cutoff, compared to other studies which include more stringent thresholds for inclusion in their networks (28, 29). Therefore, it is possible that some of these less abundant genes co-occur with Bacteroidetes and Firmicutes but were not observed due to exclusion. Finally, it is also possible that comparison of abundances across ARGs and taxa is inherently skewed. For example, even after excluding sparse features, the number of genera was nearly one order of magnitude higher than that of ARG groups (ARG groups = 251; genera = 2,282). Although our interpretation of the networks focused on ARG-taxa connections and so reduced the “noise” associated with many potential nodes, it is likely that the distribution of abundance

across ARG and taxonomic features was quite different. Therefore, any associations observed in this correlation network analysis must be considered with caution. As mentioned before, these networks were designed to serve as hypothesis-generating practices, the results of which were further investigated via ARG-carrying contig analysis.

Host-tracking analysis was pursued with the goal of more concretely capturing connections among taxa and ARGs observed in the co-occurrence networks. This method of identifying ARG-carrying contigs (ACCs) and annotating them with taxonomic information has been used in multiple studies (29, 93-95). Our findings suggest that many of the associations observed in the co-occurrence network analysis were indeed relevant. For example, *Escherichia* was found to be a prominent host to ARGs in the infected cases, comprising an average of 38% of all ACCs among samples. As suggested by our network analysis, a large proportion of these resistance genes were relevant to multi-drug resistance and multi-compound resistance. Indeed, as stated prior, *Escherichia* have historically been found to harbor multi-drug resistance. The development of such resistance has often been attributed to high potential for HGT in the human gut as well as continued application of antibiotics (96). Recently, it was also found that multi-drug resistance among *Escherichia* may also be linked to non-antibiotic pharmaceutical use in addition to antibiotics (97). Alarmingly, it has been found that as the level of multi-drug resistance increases, so too does the number of integrons, highlighting *Escherichia*'s role in the increasing mobility of multi-drug resistance (98, 99).

The prominence of *Salmonella* as an ARG-carrying microbe also aligns with our network findings. Intriguingly, *Salmonella* was once again most represented among patients who were infected with *Salmonella*, accounting for approximately 31% of all ACCs in these individuals compared to the overall case average of 18.3%; however, there was also evidence of *Salmonella*

being in the top-10 most abundant microbial hosts to ARGs within cases infected with each of the four enteric pathogens (data not shown). A noticeable deviation from our network findings was the identification of *Klebsiella* as a prominent ARG carrier among cases (9.92% of all ACCs) and even follow-ups (4.58%). Of great interest is the relatively high occurrence of the IS5 family of transposases in *Klebsiella*. Although the role of this family of transposases in disseminating resistance genes has not yet been fully characterized (100), the identification of a genomic element with the potential to transfer ARGs among members of the gut microbial community is quite relevant. This is especially true as we increasingly consider the human gut to be a reservoir for antimicrobial resistance, a status that has only been exacerbated by the wide array of MGEs present in this environment (101, 102).

The trends observed among follow-ups in the ACC analysis are also quite interesting. It is meaningful, for example, that *Escherichia* still accounts for nearly 20% of all ARG-carrying contigs among recovered patients. Contrary to the pattern seen in cases, *Bacteroides* was next most represented on ACCs, comprising approximately 15%. Of interest are the ARGs harbored by *Bacteroides* which included genes conferring MLS, beta-lactam, and tetracycline resistance, primarily. Resistance of this kind has previously been documented in *Bacteroides* (103, 104). More broadly, resistance to these three classes in addition to aminoglycoside resistance was previously characterized among healthy individuals across various countries, suggesting that resistance to these antibiotics is increasingly ubiquitous regardless of disease status or geography (28, 105).

When investigating beta-lactam resistance genes and ESBLs, in particular, certain findings were of greater concern than others. The documentation of 14 different ESBLs among cases, for example, was alarming. Although many of these were found in just one or two samples

of the cohort, the diversity of ESBLs among patients with enteric infection is concerning, especially considering the relatively high abundance of various members of *Enterobacteriaceae* for which ESBL carriage and dissemination is a grave concern (10). A handful of these resistance genes were prevalent across our sample set, with CepA and SHV families occurring in 19 and 8 cases, respectively. The OXA genes were also widely present in our samples, though there was little consensus within this grouping. In most cases except for the CepA family genes, the signatures for these ESBLs were “lost” during recovery. Levels of persistence and acquisition have been measured for various ESBLs, with certain CTX and SHV being more easily lost, though this was found to depend on the identity of the bacterial host (106). The noted roles of *Klebsiella* and *Escherichia* in carrying ESBLs among our samples calls attention to the documented capacity of these genera to horizontally transfer such genes despite being of different species or clonal lineages (107, 108). Indeed, great attention must be paid to the trajectory of such genes during and after enteric infection.

Patterns observed among non-ESBL beta-lactam resistance genes were also illuminating. One of the most prevalent beta-lactam resistance genes was represented by the CfxA family, harbored by *Bacteroides* and *Prevotella*. In a handful of paired case-follow samples, the occurrence of CfxA appeared as though it may have undergone a transfer from one genus to the other; indeed, the exchange of genetic components has been documented between *Bacteroides* and *Prevotella* (109). However, the confidence with which we weight these speculations must remain in check, as our assertion of “transfer” in this circumstance relies solely on the appearance of a gene type in two different genera at two different time points. Much more rigorous methods, such as characterizing the sequence-level similarity of the gene(s) in question in both the case and follow-up samples, would be required to confirm such a statement.

However, observation of this beta-lactamase family among two different genera still holds great interest as is.

We also observed widespread prevalence of the BlaEC family among cases, which were primarily harbored by *Escherichia*. Other members of *Enterobacteriaceae* also harbored non-ESBL beta-lactam ARGs such as the CMY family, which was carried by *Salmonella*, and class A and D beta-lactamases, which were assigned to multiple members of this family. The marked increase observed in these genera as well as genes they were presumed to harbor (beta-lactam ARGs in addition to other multi-compound and multi-drug resistance genes) may be explained by a phenomenon in which gut inflammation creates an environment conducive for the growth of *Enterobacteriaceae* (110). These conditions have also been shown to augment levels of HGT between both commensal and pathogenic members of this family (111).

Despite our extended interest in the findings of the ARG-carrying contig analysis, one noteworthy limitation of this pipeline is the potential for ARGs to be located on plasmids. While many resistance genes occur on plasmids and are spread via HGT, these genomic entities can still contain taxonomic information regarding their microbe of origin (112). It is possible, therefore, that various ARGs identified in our ACC analysis are indeed associated with plasmids. However, it has been found that assembly, especially of short-read sequencing, can fall short when it comes to characterizing plasmids and other mobile genetic elements (MGEs) (113). Additionally, it is possible that limitations related to sequencing depth, incomplete assemblies, redundancy, or other sequencing-based concerns do present an inaccurate representation of microbial hosts in our study. To address these types of limitations, a tool, PlasFlow, has been designed to specifically extract plasmid sequences from metagenome assemblies, allowing direct taxonomic classification of the plasmids themselves (114). A more direct method, which

employs physical isolation of plasmids from metagenomic DNA is called high-throughput transposon-aided capture (TRACA); this protocol is designed to extract circular plasmids from metagenomic DNA and subsequently maintain and select for these plasmids after transferring to an *E. coli* host (115).

Another limitation of our ACC analysis is that it is strictly reliant on inferring microbial hosts from ARG-taxa co-occurrence on the same contig. Alternative methods, rather, are better able to confirm whether microbes harbor ARGs by using more targeted protocols when extracting and isolating metagenomic DNA. One example of this is Single-molecule Real-time (SMRT) sequencing, which is a sequencing method that bins metagenomic reads based on the methylation status (116). Since methylation of both chromosomal and plasmid-based nucleotides are consistent within a microbe, these motifs can confidently be assigned to a taxon. A second lab-based technique for linking various genomic components to their microbial host is high-throughput chromosomal conformation capture, dubbed “Hi-C.” Hi-C is a method in which DNA molecules which are proximal to each other will be covalently bonded together and subsequently ligated to form a contiguous DNA strand (26). The Hi-C method has been used in multiple human gut microbiome studies, particularly to explore the linkages between ARGs, various mobile genetic elements, and host microbes within the gut (117, 118). Future studies of this dataset should consider usage of these laboratory techniques to more accurately portray which members of the human gut microbiome are responsible for harboring ARGs of interest.

In conclusion, enteric infection has been shown to severely alter the human gut environment, both taxonomically and genetically. Importantly, we have shown that invasion by an enteric pathogen can result in notable shifts of certain microbiota and their associated ARGs. The findings included here aid in further elucidating the interplay between microbes and ARGs

in an infected gut environment. However, much work is needed to advance our understanding of the trajectory of recovery from enteric infection as it pertains to the presence and dissemination of drug resistance. Future work should indeed focus on characterizing the interaction of microbial hosts, ARGs, and mobile genetic elements during recovery. Illuminating these interactions will help to facilitate our growing understanding of this unique intersection between microbial ecology, antimicrobial resistance, and enteric disease.



## **APPENDIX**

**Table B.1. Shapiro-Wilk Test results for case and follow-up samples in the microbiome and resistome datasets.**

Dataset	Alpha Diversity Metric	Test Statistic (W)	p-value
Microbiome	Richness	0.936	2.35E-05
	Shannon Diversity	0.933	1.45E-05
	Pielou's Evenness	0.940	4.25E-05
Resistome	Richness	0.940	4.18E-05
	Shannon Diversity	0.905	3.63E-07
	Pielou's Evenness	0.885	3.68E-08

**Table B.2. Top-25 co-occurrence associations between *Escherichia* and ARG groups in cases.**

Label	Target	Correlation	q-value (FDR)
Escherichia	MDTF	0.981353998	0
Escherichia	MDTN	0.980327032	0
Escherichia	MDTP	0.979157357	0
Escherichia	GADX	0.978488971	0
Escherichia	TEHA	0.978403813	0
Escherichia	MDTO	0.978377574	0
ACRE	Escherichia	0.978126929	0
Escherichia	GADC	0.977336127	0
Escherichia	HDEB	0.976927668	0
Escherichia	RCNB	0.976595211	0
Escherichia	RCNA	0.976428115	0
Escherichia	HDEA	0.975493044	0
Escherichia	MDTE	0.974363553	0
Escherichia	YGIW	0.974229023	0
CUTF	Escherichia	0.974062032	0
Escherichia	GADA	0.973949518	0
CUSB	Escherichia	0.97252408	0
Escherichia	GADE	0.97244565	0
Escherichia	OMPFB	0.971793253	0
DSBB	Escherichia	0.969242968	0
Escherichia	MNTP	0.968538549	0
Escherichia	YODB	0.968272989	0
Escherichia	NFSA	0.967404907	0
Escherichia	NIKA	0.967132266	0
Escherichia	GADW	0.966947253	0

**Table B.3. Co-occurrence associations between *Escherichia* and other taxa and ARGs in follow-ups.**

<b>Label</b>	<b>Target</b>	<b>Correlation</b>	<b>q-value (FDR)</b>
Escherichia	Shigella	0.936649069	0
Escherichia	UHPT	0.833315741	0
Escherichia	NIKA	0.831335752	1.50E-13
Escherichia	YCHH	0.822717477	5.29E-13
Escherichia	YHCN	0.819259584	7.67E-13
Escherichia	MDTN	0.816175487	9.98E-13
Escherichia	FIEF	0.814374059	1.34E-12
Escherichia	ZRAR	0.813508268	1.44E-12
Escherichia	ZNTA	0.810051112	2.30E-12
EPTA	Escherichia	0.808197735	2.91E-12
Escherichia	NHAA	0.80771521	3.12E-12
Escherichia	IBPA	0.806446575	3.72E-12
Escherichia	PITA	0.806047925	3.92E-12
Escherichia	MDTH	0.80594996	3.92E-12
Escherichia	GYRBA	0.803715593	5.09E-12
Escherichia	ZITB	0.802731572	5.75E-12
CUTE	Escherichia	0.800860986	7.17E-12

**Table B.4A. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups.**

This table includes information about the prevalence of these resistance genes in cases vs. follow-ups (A), the number of genes “lost” (i.e., present in the case sample but absent in the follow-up sample), “maintained” (i.e., present in both samples), or “acquired” (absent in case sample, present in follow-up sample) (B), as well as the number of each gene found among people infected by each enteric pathogen (C).

\*identified as an extended-spectrum beta-lactamase (ESBL)

Host	Beta-lactamase	Cases	Follow-ups
uncultured	ACI family class A beta-lactamase	0	0
Enterobacter	ACT family cephalosporin-hydrolyzing class C beta-lactamase	5	0
Acinetobacter	ADC family extended-spectrum class C beta-lactamase*	2	0
Escherichia	beta-lactamase	3	6
Klebsiella	Beta-lactamase	0	1
Klebsiella	beta-lactamase	4	0
Salmonella	beta-lactamase	0	0
uncultured	beta-lactamase	0	1
Enterobacter	Beta-lactamase class C and other penicillin binding proteins	0	0
Escherichia	beta-lactamase TEM-1 variant*	1	0
synthetic	beta-lactamase TEM-1 variant*	19	4
Klebsiella	beta-lactamase, partial	3	0
Escherichia	BlaEC family class C beta-lactamase	49	19
Shigella	BlaEC family class C beta-lactamase	1	0
Bacteroides	CepA family class A extended-spectrum beta-lactamase*	19	13
Bacteroides	CfxA family class A broad-spectrum beta-lactamase	46	48
Prevotella	CfxA family class A broad-spectrum beta-lactamase	5	9
Atlantibacter	class A beta-lactamase	1	0
Bacillus	class A beta-lactamase	1	1
Burkholderia	class A beta-lactamase	0	1
Clostridium	class A beta-lactamase	1	3
Proteus	class A beta-lactamase	1	0
Salmonella	class A beta-lactamase	1	0
Yersinia	class A beta-lactamase	0	1
Escherichia	class A beta-lactamase, partial	0	2

**Table B.4A (cont'd)**

Klebsiella	class A beta-lactamase, partial	1	0
Bacteroides	class A beta-lactamase, subclass A2	45	47
Escherichia	class A broad-spectrum beta-lactamase TEM-1, partial*	0	0
Acinetobacter	class C beta-lactamase	0	0
Hafnia	class C beta-lactamase	3	0
Providencia	class C beta-lactamase	1	0
Pseudomonas	class C beta-lactamase	7	0
Achromobacter	class D beta-lactamase	0	0
Acinetobacter	class D beta-lactamase	0	0
Flavobacterium	class D beta-lactamase	4	6
Klebsiella	class D beta-lactamase, partial	1	0
Salmonella	CMY family class C beta-lactamase	3	2
Citrobacter	CMY-2 family class C beta-lactamase	8	3
Salmonella	CMY-2 family class C beta-lactamase	2	1
Cronobacter	CSA family class C beta-lactamase	0	0
Escherichia	CTX-M family class A extended-spectrum beta-lactamase*	0	0
Morganella	DHA family class C beta-lactamase	1	0
uncultured	extended spectrum beta-lactamase CTX-M*	0	0
Escherichia	extended-spectrum beta-lactamase CTX-M-2, partial*	0	0
Stenotrophomonas	L1 family subclass B3 metallo-beta-lactamase	0	0
Klebsiella	LEN family class A beta-lactamase	1	2
Bacteroides	metallo-beta-lactamase, partial	0	0
Enterobacter	MIR family cephalosporin-hydrolyzing class C beta-lactamase	1	0
Klebsiella	MULTISPECIES: OXY family class A extended-spectrum beta-lactamase*	2	0
Alistipes	MULTISPECIES: subclass B1 metallo-beta-lactamase	0	0
Klebsiella	OXA-1 family class D beta-lactamase*	2	0
Acinetobacter	OXA-134 family carbapenem-hydrolyzing class D beta-lactamase*	0	0
Acinetobacter	OXA-211 family carbapenem-hydrolyzing class D beta-lactamase*	0	0

**Table B.4A (cont'd)**

Pseudomonas	OXA-50 family oxacillin-hydrolyzing class D beta-lactamase*	2	0
Acinetobacter	OXA-51 family carbapenem-hydrolyzing class D beta-lactamase*	2	0
Campylobacter	OXA-61 family class D beta-lactamase*	2	0
Klebsiella	OXY family class A extended-spectrum beta-lactamase*	2	0
Pseudomonas	PDC family class C beta-lactamase	2	0
Burkholderia	PenA family class A beta-lactamase	1	0
Raoultella	PLA/ORN/TER family class A beta-lactamase	2	0
Citrobacter	SED family class A beta-lactamase	4	3
Klebsiella	SHV family class A beta-lactamase*	8	2
Serratia	SRT/SST family class C beta-lactamase	0	0
Bacteroides	subclass B1 metallo-beta-lactamase	8	14
Desulfovibrio	subclass B1 metallo-beta-lactamase	0	1
Myxococcus	subclass B1 metallo-beta-lactamase	1	0

**Table B.4B. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups.**

This table includes information about the prevalence of these resistance genes in cases vs. follow-ups (A), the number of genes “lost” (i.e., present in the case sample but absent in the follow-up sample), “maintained” (i.e., present in both samples), or “acquired” (absent in case sample, present in follow-up sample) (B), as well as the number of each gene found among people infected by each enteric pathogen (C).

\*identified as an extended-spectrum beta-lactamase (ESBL)

Host	Beta-lactamase	Lost	Maintained	Acquired
uncultured	ACI family class A beta-lactamase	0	0	0
Enterobacter	ACT family cephalosporin-hydrolyzing class C beta-lactamase	5	0	0
Acinetobacter	ADC family extended-spectrum class C beta-lactamase*	3	0	6
Escherichia	beta-lactamase	0	0	1
Klebsiella	Beta-lactamase	4	0	0
Klebsiella	beta-lactamase	0	0	0
Salmonella	beta-lactamase	0	0	1
uncultured	beta-lactamase	0	0	0
Enterobacter	Beta-lactamase class C and other penicillin binding proteins	3	0	0
Escherichia	beta-lactamase TEM-1 variant*	35	14	5
synthetic	beta-lactamase TEM-1 variant*	1	0	0
Klebsiella	beta-lactamase, partial	7	39	9
Escherichia	BlaEC family class C beta-lactamase	3	2	7
Shigella	BlaEC family class C beta-lactamase	1	0	0
Bacteroides	CepA family class A extended-spectrum beta-lactamase*	1	0	1
Bacteroides	CfxA family class A broad-spectrum beta-lactamase	0	0	1
Prevotella	CfxA family class A broad-spectrum beta-lactamase	1	0	3
Atlantibacter	class A beta-lactamase	1	0	0
Bacillus	class A beta-lactamase	1	0	0
Burkholderia	class A beta-lactamase	0	0	1
Clostridium	class A beta-lactamase	0	0	2
Proteus	class A beta-lactamase	1	0	0
Salmonella	class A beta-lactamase	7	38	9

**Table B.4B (cont'd)**

Yersinia	class A beta-lactamase	0	0	0
Escherichia	class A beta-lactamase, partial	3	0	0
Klebsiella	class A beta-lactamase, partial	1	0	0
Bacteroides	class A beta-lactamase, subclass A2	7	0	0
Escherichia	class A broad-spectrum beta-lactamase TEM-1, partial*	0	0	0
Acinetobacter	class C beta-lactamase	0	0	0
Hafnia	class C beta-lactamase	0	4	2
Providencia	class C beta-lactamase	1	0	0
Pseudomonas	class C beta-lactamase	3	0	2
Achromobacter	class D beta-lactamase	6	2	1
Acinetobacter	class D beta-lactamase	2	0	1
Flavobacterium	class D beta-lactamase	0	0	0
Klebsiella	class D beta-lactamase, partial	1	0	0
Salmonella	CMY family class C beta-lactamase	0	0	0
Citrobacter	CMY-2 family class C beta-lactamase	1	0	2
Salmonella	CMY-2 family class C beta-lactamase	0	0	0
Cronobacter	CSA family class C beta-lactamase	1	0	0
Escherichia	CTX-M family class A extended-spectrum beta-lactamase*	0	0	0
Morganella	DHA family class C beta-lactamase	2	0	0
uncultured	extended spectrum beta-lactamase CTX-M*	1	0	0
Escherichia	extended-spectrum beta-lactamase CTX-M-2, partial*	2	0	0
Stenotrophomonas	L1 family subclass B3 metallo-beta-lactamase	4	0	3
Klebsiella	LEN family class A beta-lactamase	0	0	0
Bacteroides	metallo-beta-lactamase, partial	3	5	9
Enterobacter	MIR family cephalosporin-hydrolyzing class C beta-lactamase	0	0	1
Klebsiella	MULTISPECIES: OXY family class A extended-spectrum beta-lactamase*	1	0	0
Alistipes	MULTISPECIES: subclass B1 metallo-beta-lactamase	2	0	0
Klebsiella	OXA-1 family class D beta-lactamase*	1	0	0
Acinetobacter	OXA-134 family carbapenem-hydrolyzing class D beta-lactamase*	18	1	3



**Table B.4B (cont'd)**

Acinetobacter	OXA-211 family carbapenem-hydrolyzing class D beta-lactamase*	9	10	3
Pseudomonas	OXA-50 family oxacillin-hydrolyzing class D beta-lactamase*	0	0	0
Acinetobacter	OXA-51 family carbapenem-hydrolyzing class D beta-lactamase*	0	0	0
Campylobacter	OXA-61 family class D beta-lactamase*	0	0	0
Klebsiella	OXY family class A extended-spectrum beta-lactamase*	0	0	0
Pseudomonas	PDC family class C beta-lactamase	2	0	0
Burkholderia	PenA family class A beta-lactamase	2	0	0
Raoultella	PLA/ORN/TER family class A beta-lactamase	0	0	0
Citrobacter	SED family class A beta-lactamase	0	0	0
Klebsiella	SHV family class A beta-lactamase*	2	0	0
Serratia	SRT/SST family class C beta-lactamase	2	0	0
Bacteroides	subclass B1 metallo-beta-lactamase	2	0	0
Desulfovibrio	subclass B1 metallo-beta-lactamase	2	0	0
Myxococcus	subclass B1 metallo-beta-lactamase	8	0	2

**Table B.4C. Summary of beta-lactamase genes and their corresponding microbial hosts in cases and follow-ups.**

This table includes information about the prevalence of these resistance genes in cases vs. follow-ups (A), the number of genes “lost” (i.e., present in the case sample but absent in the follow-up sample), “maintained” (i.e., present in both samples), or “acquired” (absent in case sample, present in follow-up sample) (B), as well as the number of each gene found among people infected by each enteric pathogen (C).

\*identified as an extended-spectrum beta-lactamase (ESBL)

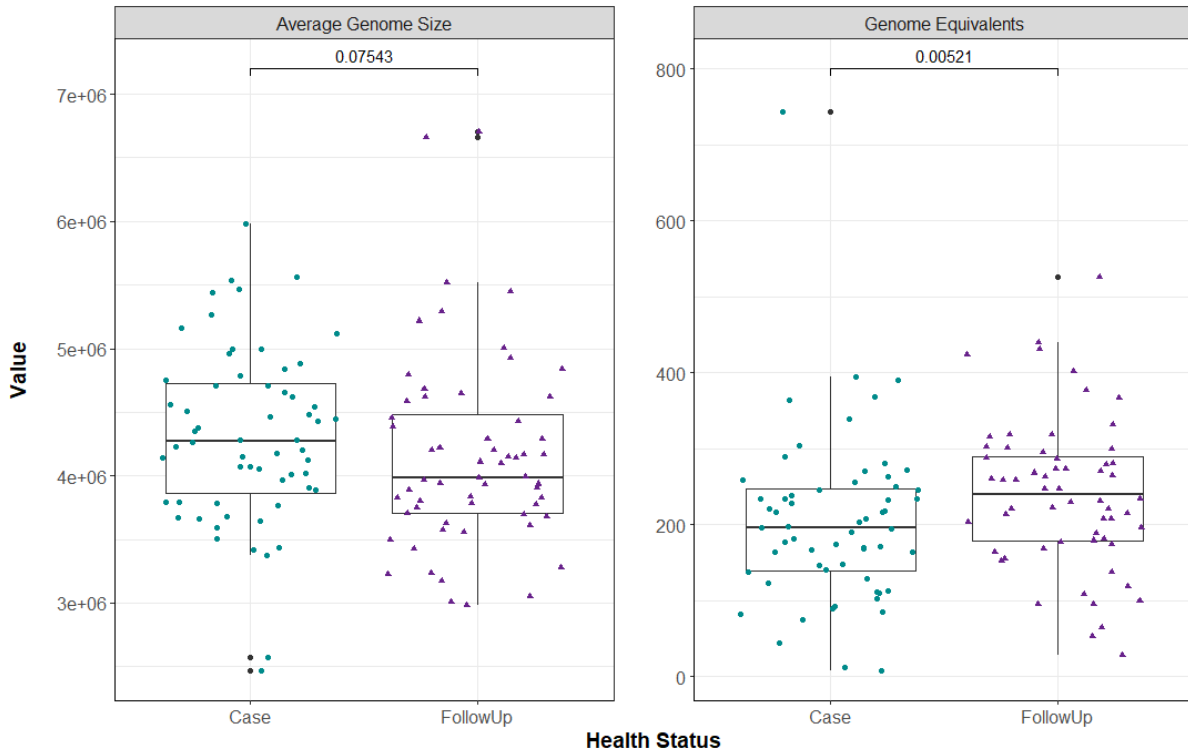
Host	Beta-lactamase	Campylobacter	Salmonella	Shigella	STEC
uncultured	ACI family class A beta-lactamase	0	0	0	0
Enterobacter	ACT family cephalosporin-hydrolyzing class C beta-lactamase	1	4	0	0
Acinetobacter	ADC family extended-spectrum class C beta-lactamase*	3	6	0	0
Escherichia	beta-lactamase	0	0	0	1
Klebsiella	Beta-lactamase	2	1	0	1
Klebsiella	beta-lactamase	0	0	0	0
Salmonella	beta-lactamase	0	0	1	0
uncultured	beta-lactamase	0	0	0	0
Enterobacter	Beta-lactamase class C and other penicillin binding proteins	1	1	0	1
Escherichia	beta-lactamase TEM-1 variant*	28	30	5	5
synthetic	beta-lactamase TEM-1 variant*	0	0	1	0
Klebsiella	beta-lactamase, partial	37	45	6	6
Escherichia	BlaEC family class C beta-lactamase	7	6	1	0
Shigella	BlaEC family class C beta-lactamase	0	1	0	0
Bacteroides	CepA family class A extended-spectrum beta-lactamase*	0	1	1	0
Bacteroides	CfxA family class A broad-spectrum beta-lactamase	1	0	0	0
Prevotella	CfxA family class A broad-spectrum beta-lactamase	2	2	0	0
Atlantibacter	class A beta-lactamase	1	0	0	0
Bacillus	class A beta-lactamase	0	1	0	0
Burkholderia	class A beta-lactamase	0	0	1	0

**Table B.4C (cont'd)**

Clostridium	class A beta-lactamase	0	2	0	0
Proteus	class A beta-lactamase	0	1	0	0
Salmonella	class A beta-lactamase	39	44	5	4
Yersinia	class A beta-lactamase	0	0	0	0
Escherichia	class A beta-lactamase, partial	3	0	0	0
Klebsiella	class A beta-lactamase, partial	1	0	0	0
Bacteroides	class A beta-lactamase, subclass A2	2	3	1	1
Escherichia	class A broad-spectrum beta-lactamase TEM-1, partial*	0	0	0	0
Acinetobacter	class C beta-lactamase	0	0	0	0
Hafnia	class C beta-lactamase	3	5	0	2
Providencia	class C beta-lactamase	0	0	1	0
Pseudomonas	class C beta-lactamase	1	3	0	1
Achromobacter	class D beta-lactamase	5	5	0	1
Acinetobacter	class D beta-lactamase	1	2	0	0
Flavobacterium	class D beta-lactamase	0	0	0	0
Klebsiella	class D beta-lactamase, partial	1	0	0	0
Salmonella	CMY family class C beta-lactamase	0	0	0	0
Citrobacter	CMY-2 family class C beta-lactamase	0	3	0	0
Salmonella	CMY-2 family class C beta-lactamase	0	0	0	0
Cronobacter	CSA family class C beta-lactamase	0	1	0	0
Escherichia	CTX-M family class A extended-spectrum beta-lactamase*	0	0	0	0
Morganella	DHA family class C beta-lactamase	0	1	0	1
uncultured	extended spectrum beta-lactamase CTX-M*	1	0	0	0
Escherichia	extended-spectrum beta-lactamase CTX-M-2, partial*	1	1	0	0
Stenotrophomonas	L1 family subclass B3 metallo-beta-lactamase	3	4	0	0
Klebsiella	LEN family class A beta-lactamase	0	0	0	0
Bacteroides	metallo-beta-lactamase, partial	9	10	2	1
Enterobacter	MIR family cephalosporin-hydrolyzing class C beta-lactamase	0	1	0	0

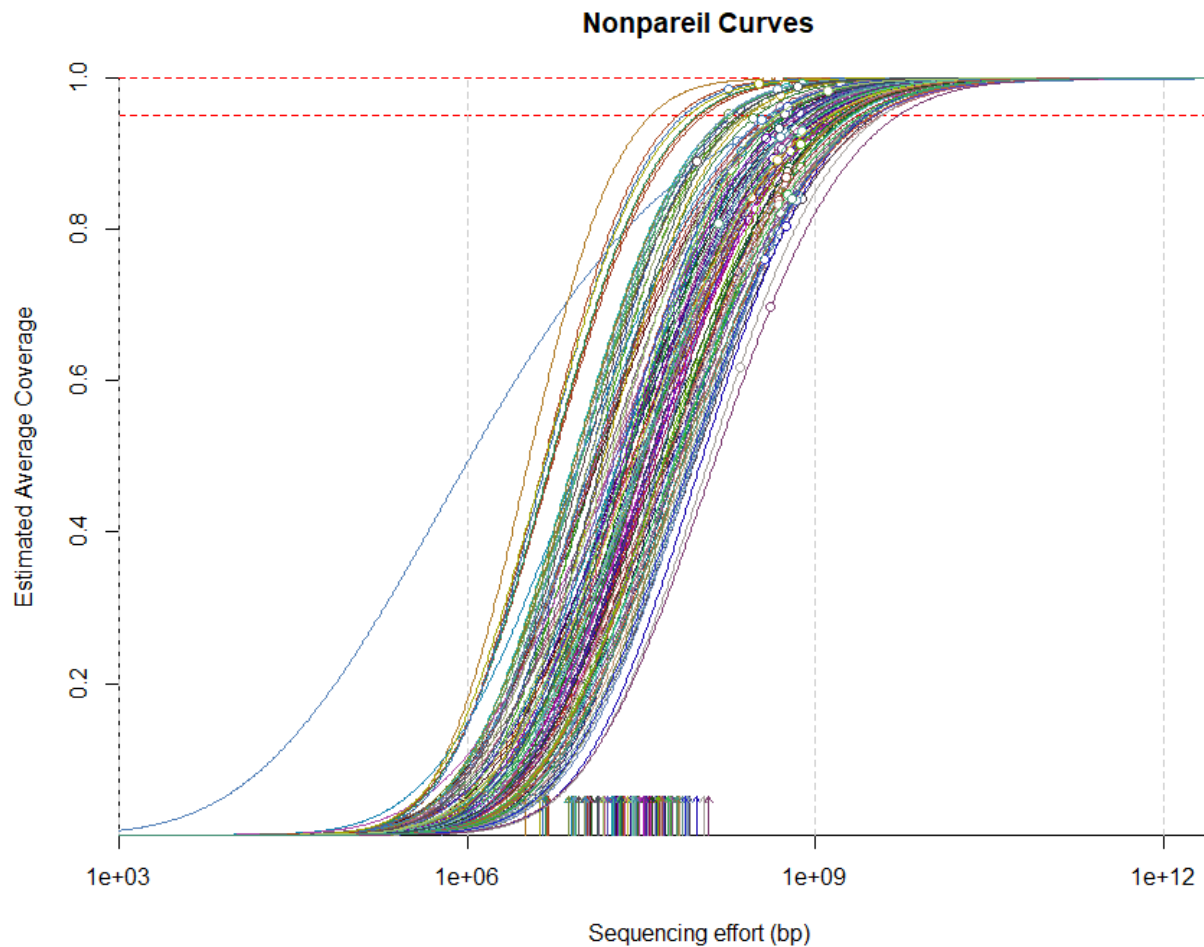
**Table B.4C (cont'd)**

Klebsiella	MULTISPECIES: OXY family class A extended-spectrum beta-lactamase*	0	1	0	0
Alistipes	MULTISPECIES: subclass B1 metallo-beta-lactamase	1	1	0	0
Klebsiella	OXA-1 family class D beta-lactamase*	0	1	0	0
Acinetobacter	OXA-134 family carbapenem-hydrolyzing class D beta-lactamase*	8	10	3	2
Acinetobacter	OXA-211 family carbapenem-hydrolyzing class D beta-lactamase*	14	12	3	4
Pseudomonas	OXA-50 family oxacillin-hydrolyzing class D beta-lactamase*	0	0	0	0
Acinetobacter	OXA-51 family carbapenem-hydrolyzing class D beta-lactamase*	0	0	0	0
Campylobacter	OXA-61 family class D beta-lactamase*	0	0	0	0
Klebsiella	OXY family class A extended-spectrum beta-lactamase*	0	0	0	0
Pseudomonas	PDC family class C beta-lactamase	1	1	0	0
Burkholderia	PenA family class A beta-lactamase	1	0	1	0
Raoultella	PLA/ORN/TER family class A beta-lactamase	0	0	0	0
Citrobacter	SED family class A beta-lactamase	0	0	0	0
Klebsiella	SHV family class A beta-lactamase*	0	1	0	1
Serratia	SRT/SST family class C beta-lactamase	1	1	0	0
Bacteroides	subclass B1 metallo-beta-lactamase	2	0	0	0
Desulfovibrio	subclass B1 metallo-beta-lactamase	2	0	0	0
Myxococcus	subclass B1 metallo-beta-lactamase	3	4	2	1



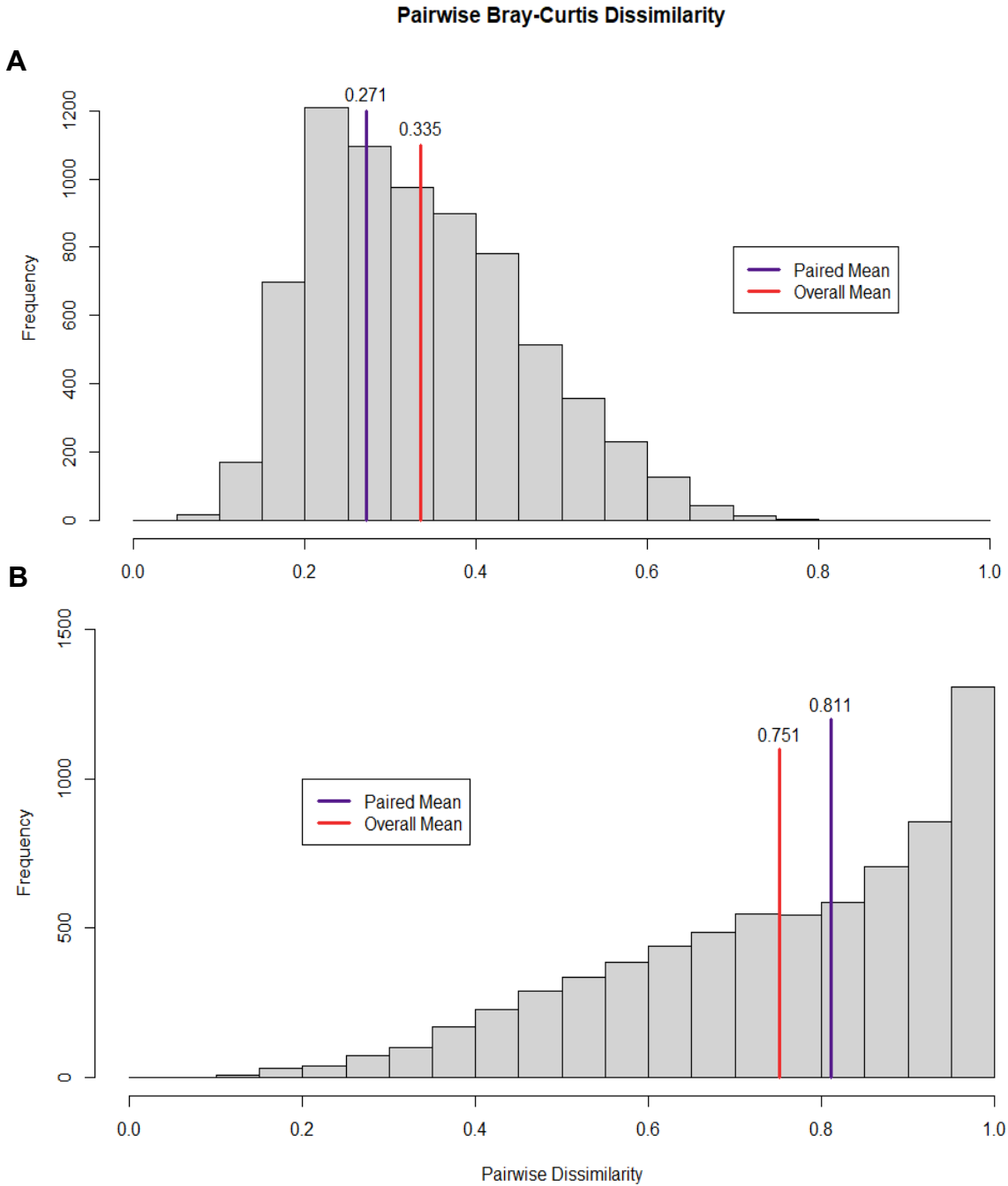
**Figure B.1. Average Genome Size (AGS) and estimated number of Genome Equivalents (GE) for paired case and follow-up samples.**

AGS (left) and GE (right) are displayed and stratified by health status, with samples represented by circles (cases, green) or squares (follow-ups, purple). Points are offset from the vertical to allow interpretation of all samples. The median of each measure is shown as a thick bar within the box (green for cases; purple for follow-ups) and the first and third quartiles are represented by the bottom and top of the box, respectively. P-values were calculated using the Wilcoxon signed-rank test for paired samples and are shown above the comparison bar within each plot.



**Figure B.2. Metagenomic sequencing coverage of short paired-end reads was determined using Nonpareil.**

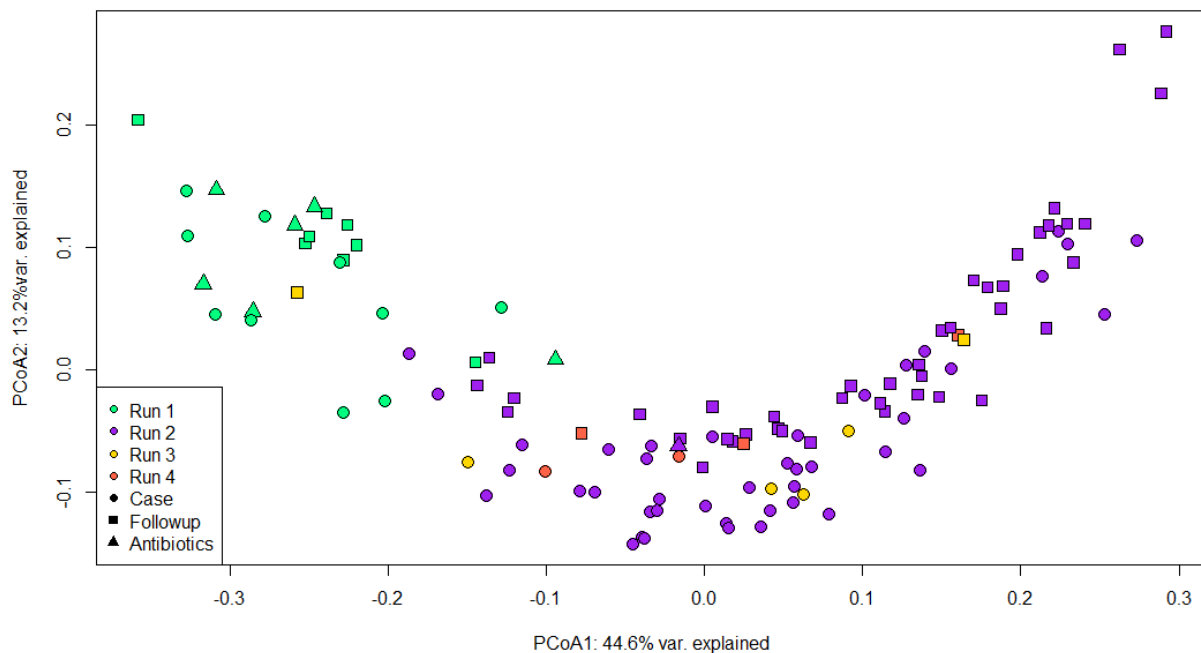
The estimated sequencing coverage at increasing sequencing effort (s-curves) and actual coverage of each sample (open circles) for case (n=60) and follow-up (n=60) samples. Each s-curve represents a single sample. Arrows aligned on the x-axis represent the Nonpareil index of sequence diversity, a metric capturing the complexity of microbial communities in sequencing space. The box bordered by dotted red lines encapsulated coverage ranging from 95-100%. The overall mean coverage among cases and follow-ups was 86.27%, with a Nonpareil diversity score of 17.0. Additionally, the estimated sequencing effort among all cases and follow-ups was  $4.77e+08$  base pairs.



**Figure B.3. Pairwise comparison of Bray-Curtis dissimilarity among cases and follow-ups.** Bray-Curtis dissimilarity was calculated for A) the microbiome (species-level) and B) resistome (gene-level) for cases and follow-ups. Upon construction of the dissimilarity matrix, pairwise dissimilarity scores were output into two separate data-frames: one which contained all pairwise comparisons and a second which contained only pairwise comparisons among relevant case-follow paired samples. The mean pairwise dissimilarity was calculated across all samples, then subsequently across paired samples only. Histograms were generated for dissimilarity measures

**Figure B.3 (cont'd)**

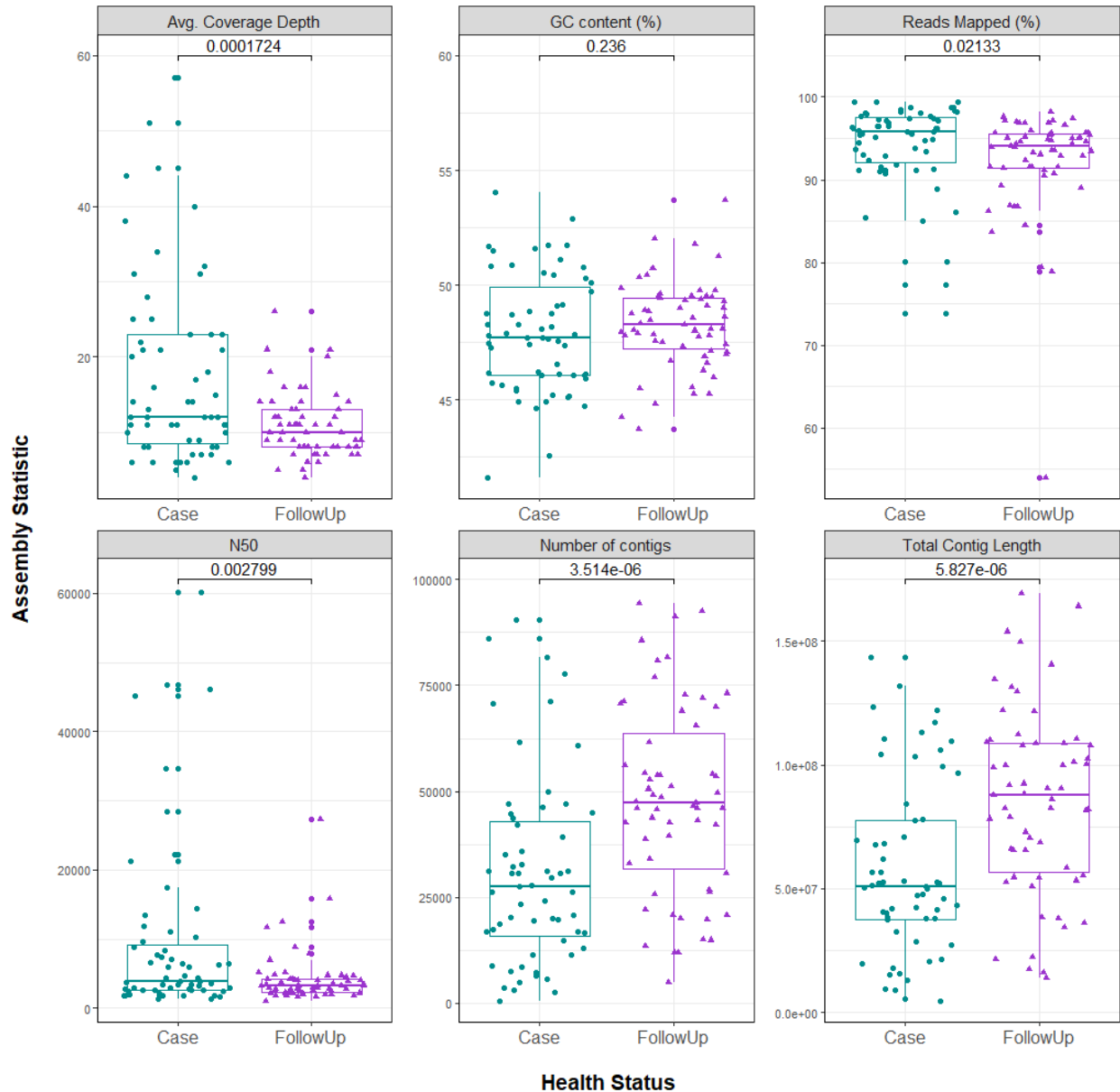
for the microbiome (top) and resistome (bottom). In each plot, the overall mean dissimilarity across all samples is displayed as a red line with the value of the mean oriented above the line (microbiome=0.335; resistome=0.751). The mean among paired samples is shown as a purple line in each respective plot (microbiome=0.271; resistome=0.811). A Welch's t-test was performed to test whether these means were significantly different; indeed, the dissimilarity among paired samples was significantly lower than the overall mean in regards to microbiome composition ( $p=2.58e-05$ ; two-sided) and significantly higher than the overall mean in regards to the resistome ( $p=0.013$ ; two-sided).



**Figure B.4. Exploring potential batch effects related to sequencing run using principal coordinate analysis of cases and follow-ups.**

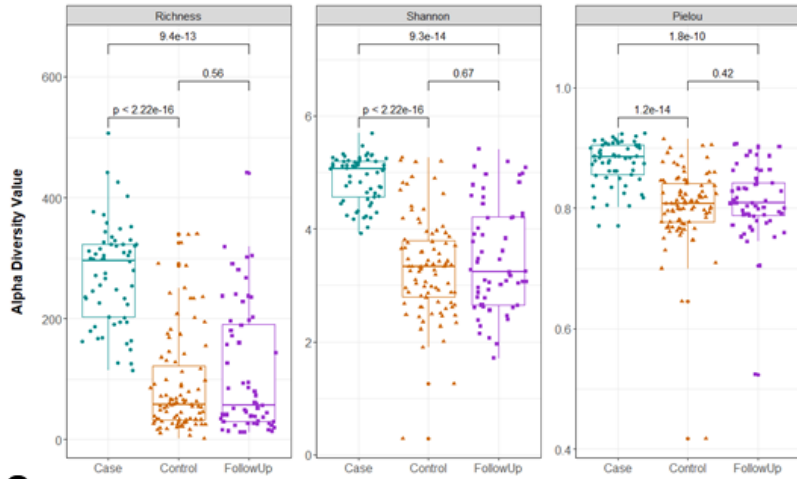
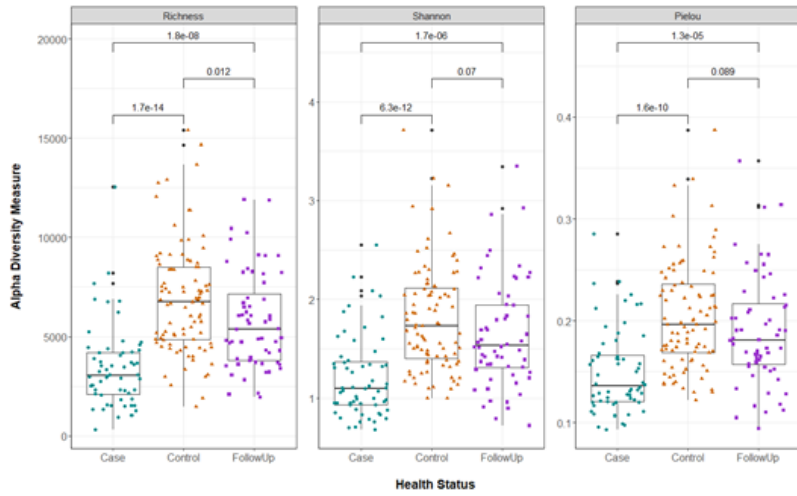
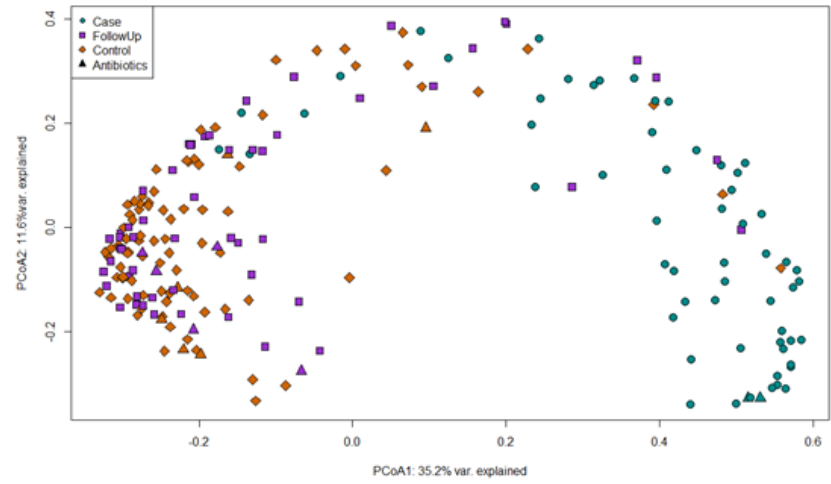
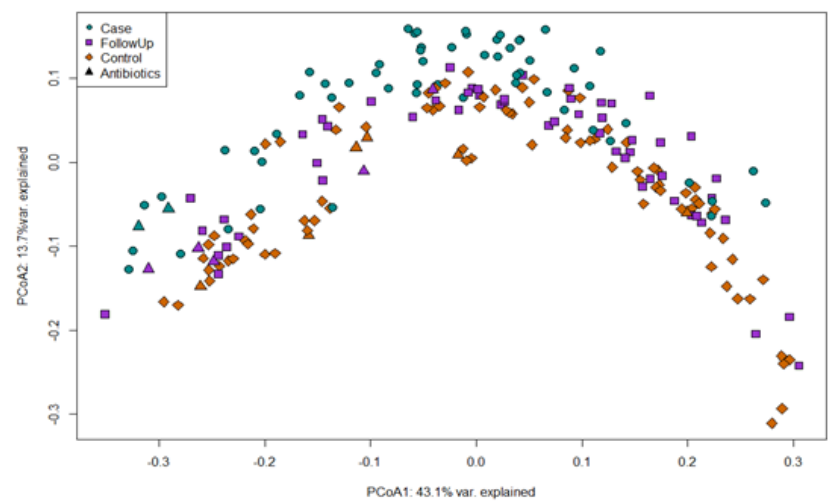
Principal Coordinates Analysis (PCoA) plot shows the distribution of cases (circles) and follow-ups (squares) based on Bray-Curtis dissimilarity calculated from species-level abundances of the gut microbiota. The first and second coordinate display the percentage of similarity explained. Samples sequenced in Run 1 appear to cluster separately from those in Runs 2, 3, and 4 along the first axis. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangles. Points are colored by their corresponding sequencing run: Run 1 (green), Run 2 (purple), Run 3 (yellow), and Run 4 (orange).





**Figure B.5. Assembly coverage statistics quantified by the Quality Assessment Tool for Genome Assemblies (QUAST).**

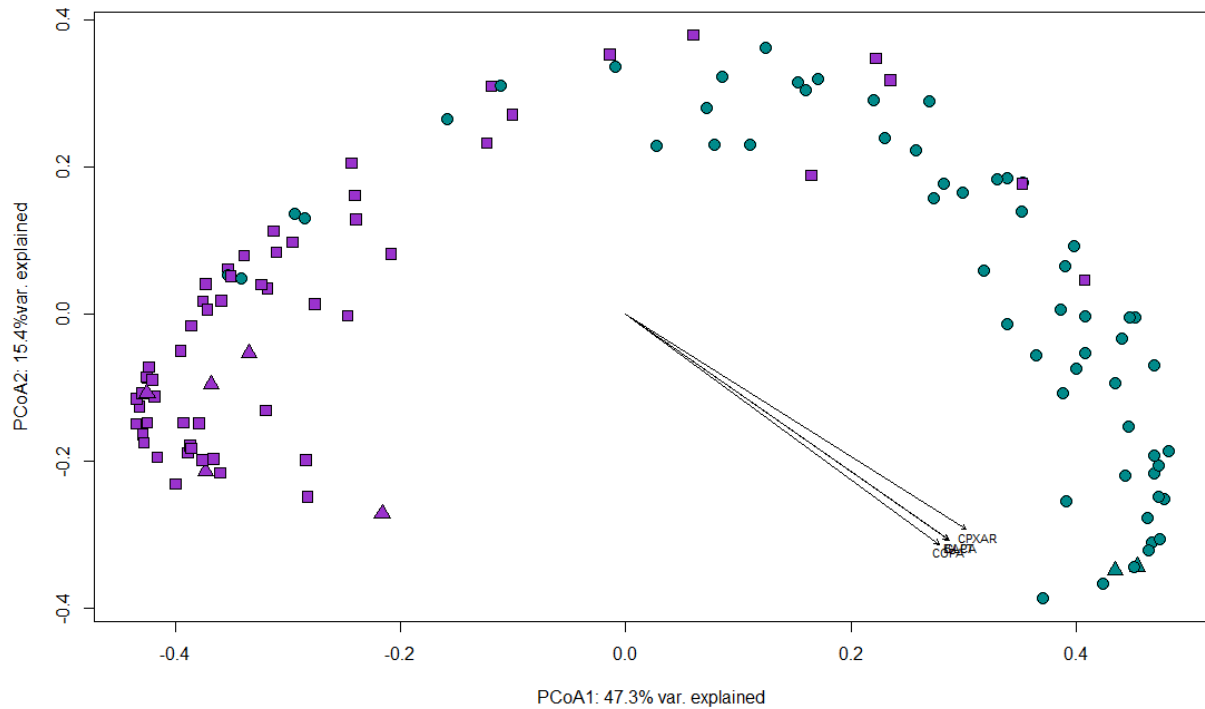
Six different assembly statistics are shown stratified by health status, with samples represented by circles (cases, green) or squares (follow-ups, purple). The six statistics indicated include (clockwise from top-left) average depth of coverage, percent of GC base pairs among samples, percentage of reads mapped to contigs, total contig length, number of contigs, and N50 value. In each plot, points are offset from the vertical to allow interpretation of all samples. The median of each measure is shown as a thick bar within the box (green for cases; purple for follow-ups) and the first and third quartiles are represented by the bottom and top of the box, respectively. P-values were calculated using the Wilcoxon signed-rank test for paired samples and are shown above the comparison bar within each plot.

**A****C****B****D**

**Figure B.6. Resistome and microbiome diversity analyses between cases, healthy household member (controls) and recovered cases (FollowUp).**

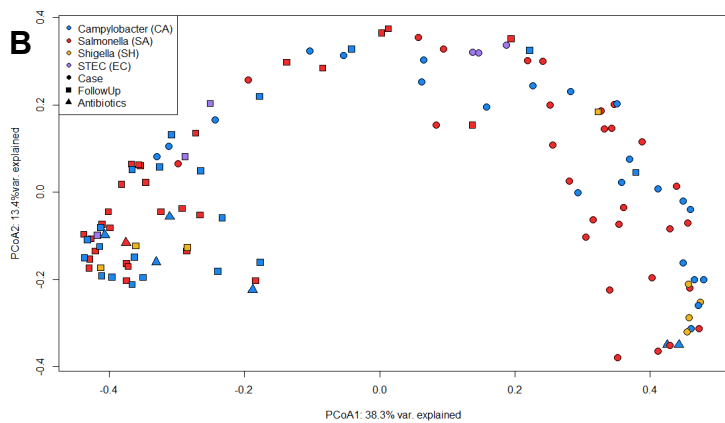
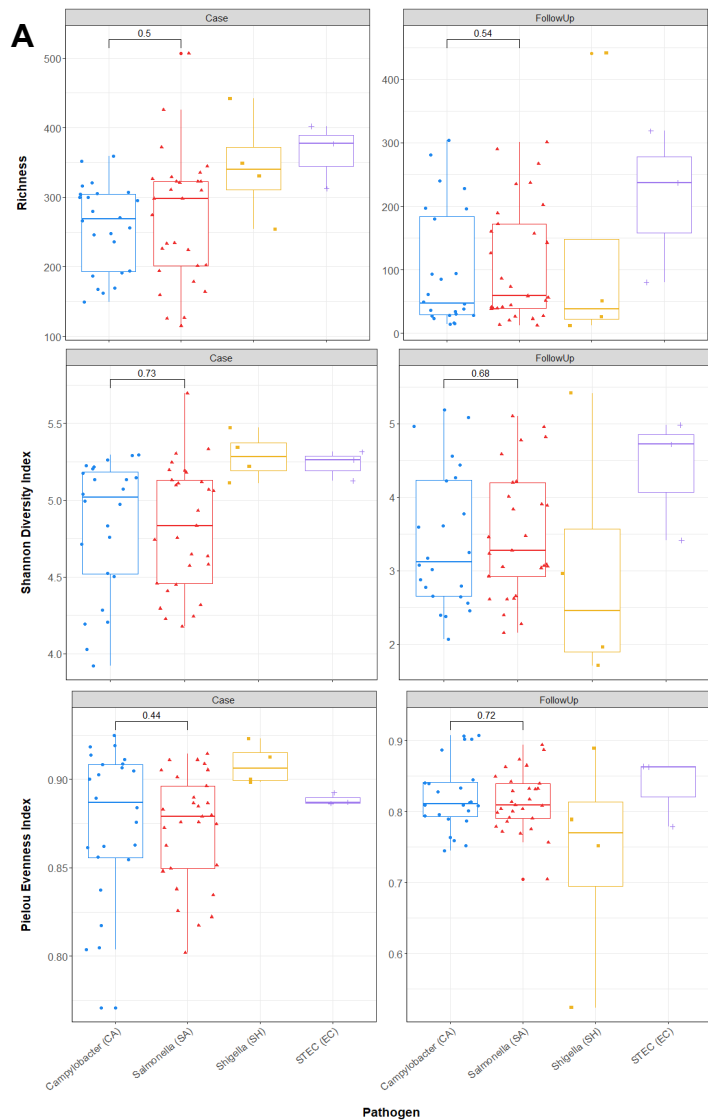
**Figure B.6 (cont'd)**

**A and C)** Three alpha diversity measures are shown (Richness, Shannon's Diversity Index, and Pielou's Evenness Index) for the resistome (**A**) and microbiome (**C**); these are stratified by health status, with samples represented by circles (cases, green), squares (follow-ups, purple), or triangles (controls, orange). Points are slightly offset from the vertical to allow interpretation of all samples. The median is indicated by the thick bar (green for cases; purple for follow-ups) and the first and third quartiles are represented by the bottom and top of the box, respectively. P-values were calculated using the Wilcoxon signed-rank test for paired samples and are shown above the comparison bar within each plot. **B and D)** Beta-diversity is displayed via a Principal Coordinates Analysis (PCoA) plot of cases (green, circles), follow-ups (purple, squares), and controls (orange, diamonds) for the resistome (**B**) and microbiome (**D**) based on Bray-Curtis dissimilarity calculated from gene-level (resistome) or species-level (microbiome) abundances. The first and second coordinate are shown and include the corresponding percentage of similarity explained. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points.



**Figure B.7. Various intrinsic factors influence case and follow-up resistomes.**

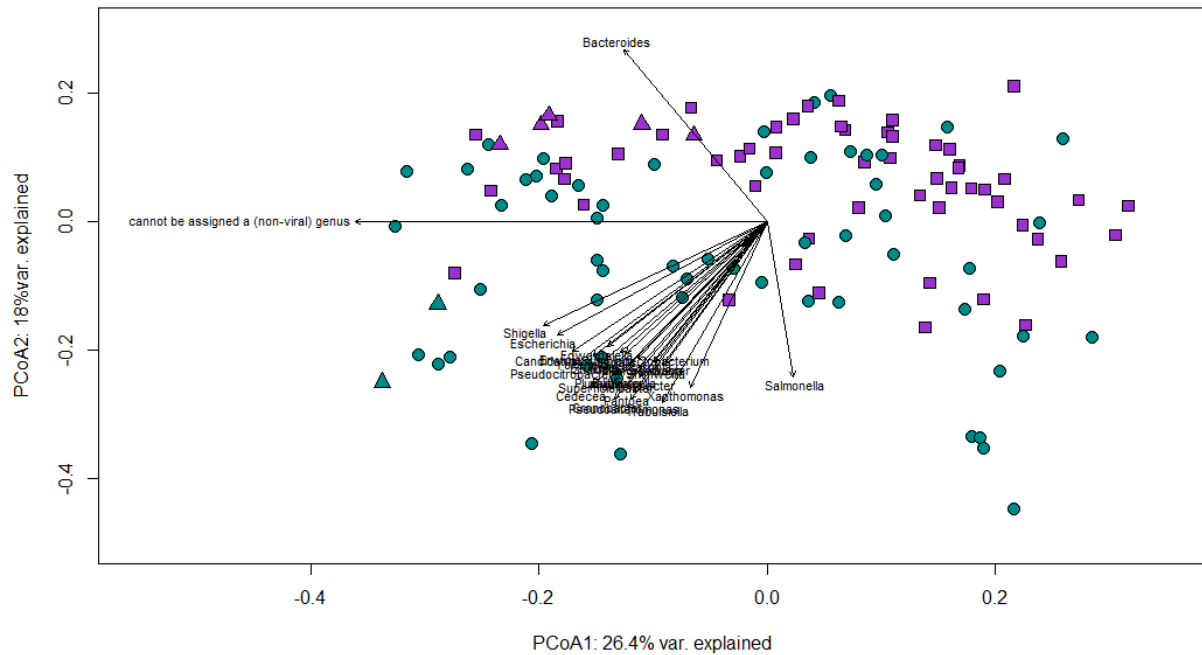
Principal Coordinates Analysis (PCoA) plot shows the distribution of case (green, circles) and follow-up (purple, squares) resistomes based on Bray-Curtis dissimilarity calculated from group-level abundances. The first and second coordinate are shown and include the corresponding percentage of similarity explained. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points. Intrinsic and extrinsic environmental factors were fitted to the ordination using the ‘envfit()’ function from the R-package ‘vegan’ (see Methods). The intrinsic factors displayed on the plot include the ARG groups *bacA*, *cpxAR*, *glpT*, and *copA*, all of which had  $R^2 \geq 0.82$  and  $p \leq 0.001$ .



**Figure B.8. Alpha and beta diversity of the resistome do not appear to differ across the four different enteric pathogens.**

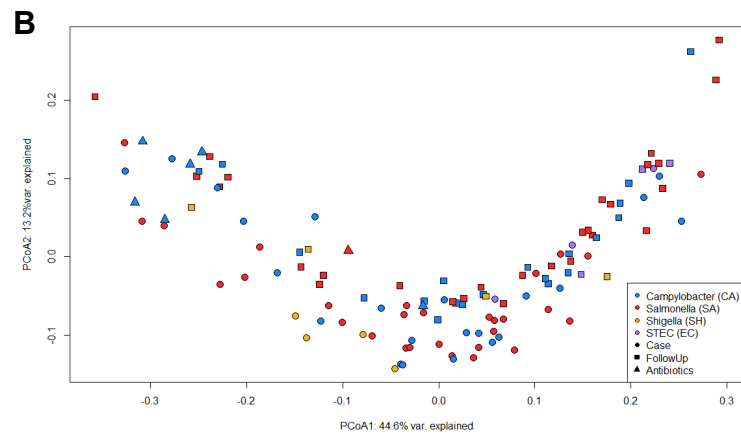
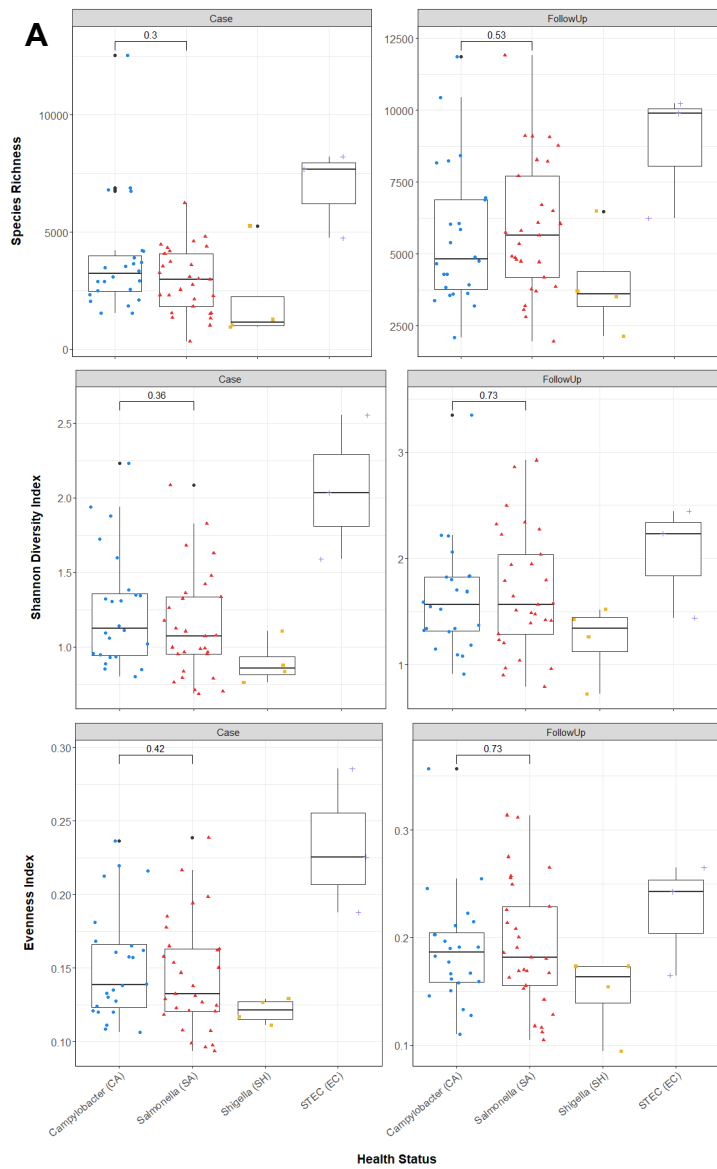
**Figure B.8 (cont'd)**

(A) Three alpha diversity measures are shown (Richness, Shannon's Diversity Index, and Pielou's Evenness Index) for the resistome of cases and follow-ups. These are stratified by infectious pathogen: *Campylobacter* (blue circles), *Salmonella* (red triangles), *Shigella* (yellow squares), and STEC (purple pluses). Points are slightly offset from the vertical to allow interpretation of all samples. In the boxplots, the median of each measure is indicated by the thick bar and the first and third quartiles are represented by the bottom and top of the box, respectively. P-values comparing *Campylobacter* and *Salmonella* were calculated using the Wilcoxon rank-sum test and are shown above the comparison bar within each plot; statistical test involving individuals infected with *Shigella* and STEC were not pursued due to low N (n=4, n=3, respectively). (B) Beta-diversity is displayed via a Principal Coordinates Analysis (PCoA) plot of case (circles) and follow-up (squares) resistome compositions based on Bray-Curtis dissimilarity calculated from gene-level abundances. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points. Points are colored by the infecting pathogen: *Campylobacter* (blue), *Salmonella* (red), *Shigella* (yellow), STEC (purple). The first and second coordinate are shown and include the corresponding percentage of similarity explained.



**Figure B.9. Multiple intrinsic factors influence case and follow-up microbiomes.**

Principal Coordinates Analysis (PCoA) plot shows the distribution of case (green, circles) and follow-up (purple, squares) microbiomes based on Bray-Curtis dissimilarity calculated from genus-level taxonomic abundances. The first and second coordinate are shown and include the corresponding percentage of similarity explained. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points. Intrinsic and extrinsic environmental factors were fitted to the ordination using the ‘envfit()’ function from the R-package ‘vegan’ (see Methods). The intrinsic factors displayed on the plot include the genera *Bacteroides*, *Cronobacter*, *Pseudoalteromonas*, and *Cedecea*, among others.

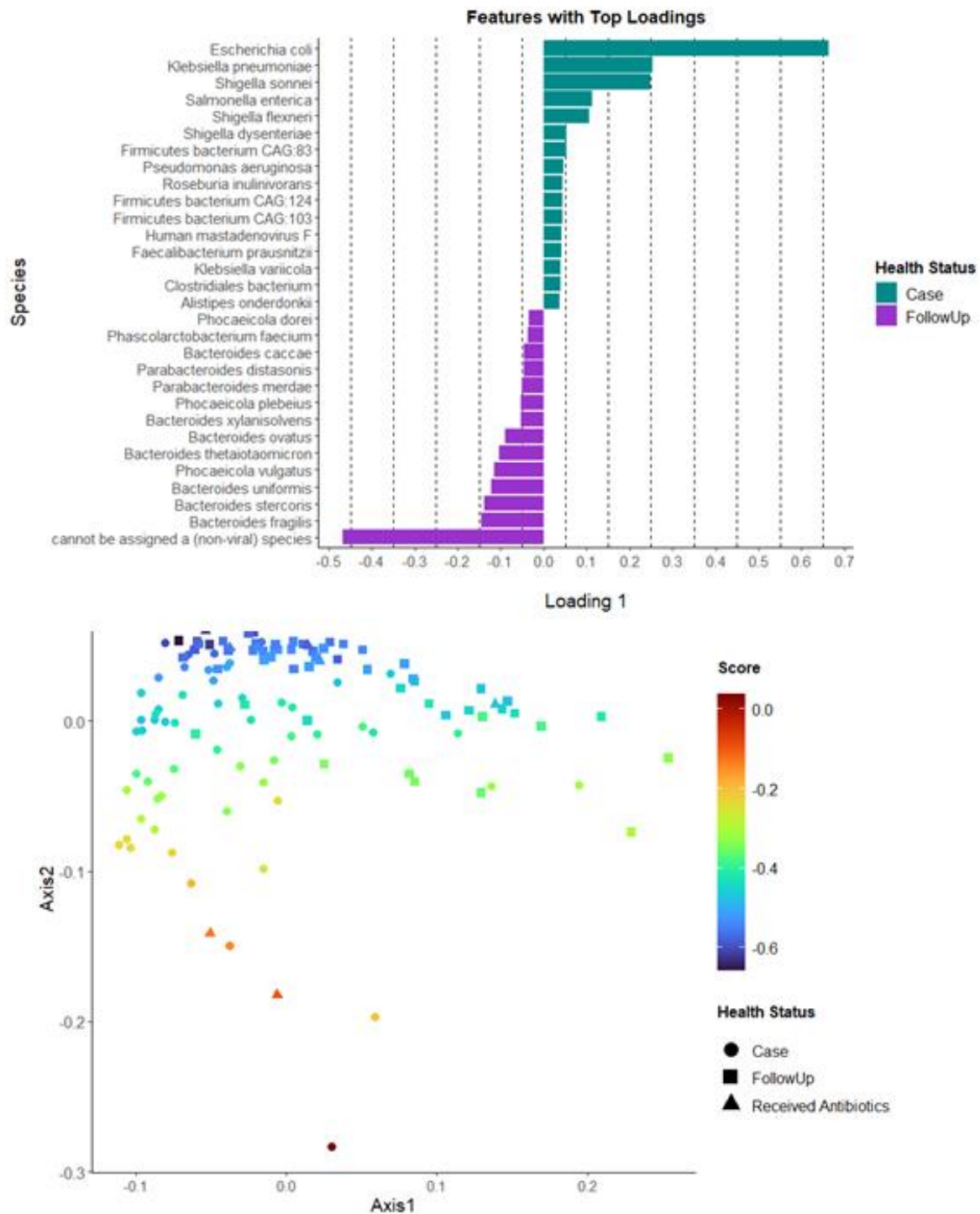


**Figure B.10. Alpha and beta diversity of the microbiome do not appear to differ across the four different enteric pathogens.**



**Figure B.10 (cont'd)**

(A) Three alpha diversity measures are shown (Richness, Shannon's Diversity Index, and Pielou's Evenness Index) for the microbiome of cases and follow-ups. These are stratified by infectious pathogen: *Campylobacter* (blue circles), *Salmonella* (red triangles), *Shigella* (yellow squares), and STEC (purple pluses). Points are slightly offset from the vertical to allow interpretation of all samples. In the boxplots, the median of each measure is indicated by the thick bar and the first and third quartiles are represented by the bottom and top of the box, respectively. P-values comparing *Campylobacter* and *Salmonella* were calculated using the Wilcoxon rank-sum test and are shown above the comparison bar within each plot; statistical test involving individuals infected with *Shigella* and STEC were not pursued due to low N (n=4, n=3, respectively). (B) Beta-diversity is displayed via a Principal Coordinates Analysis (PCoA) plot of case (circles) and follow-up (squares) microbiome compositions based on Bray-Curtis dissimilarity calculated from species-level abundances. Patients that self-reported use of antibiotics two weeks prior to sample collection are indicated by triangular data points. Points are colored by the infecting pathogen: *Campylobacter* (blue), *Salmonella* (red), *Shigella* (yellow), STEC (purple). The first and second coordinate are shown and include the corresponding percentage of similarity explained.

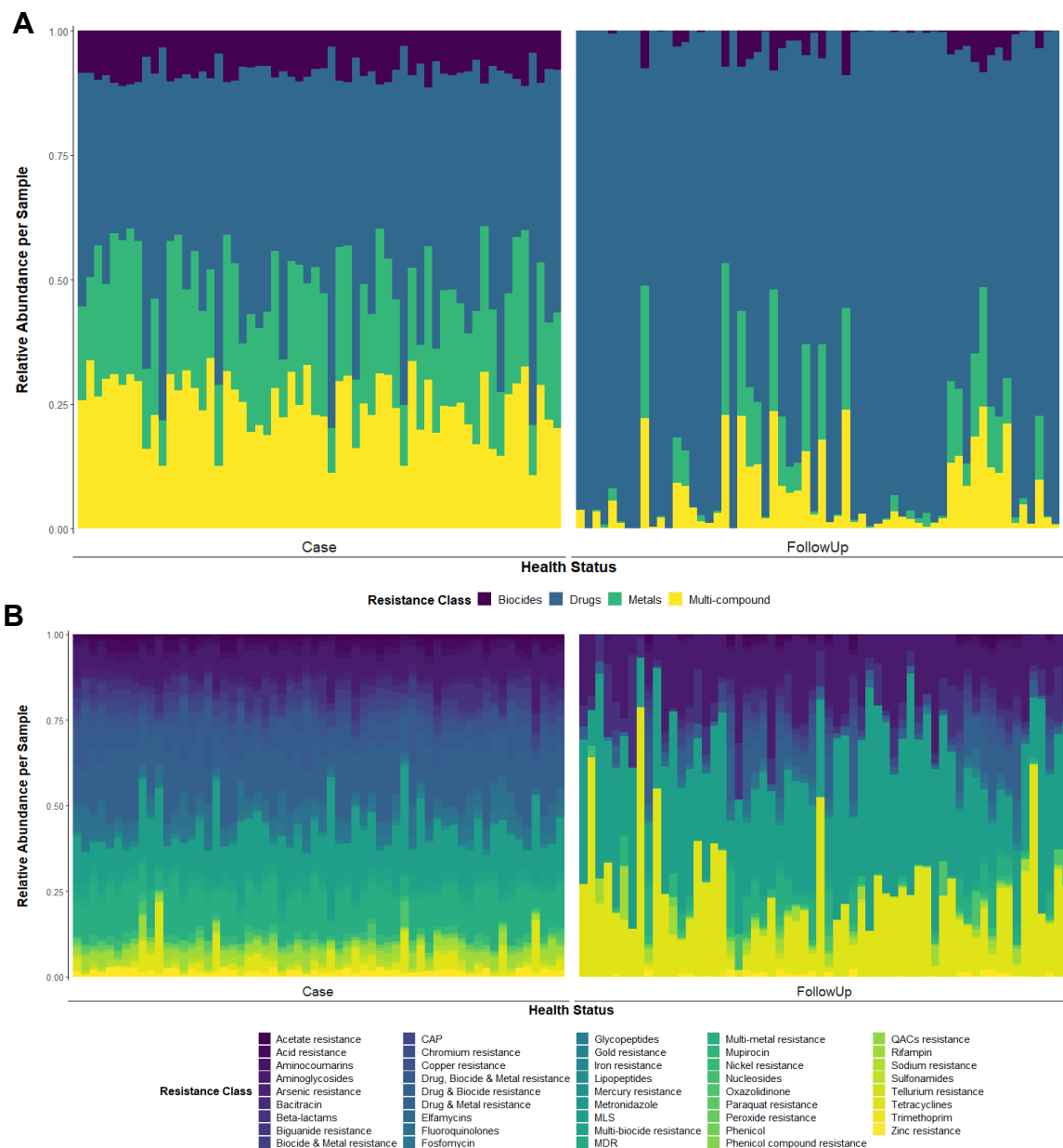


**Figure B.11. Continuous structure analysis reveals species gradients among cases and follow-ups.**

MMUPHin was used to investigate potential continuous structure among the microbiome composition of cases and follow-ups at the species level. (A) Species determined to comprise the top consensus loadings are shown; colors have been assigned to the loadings based on primary

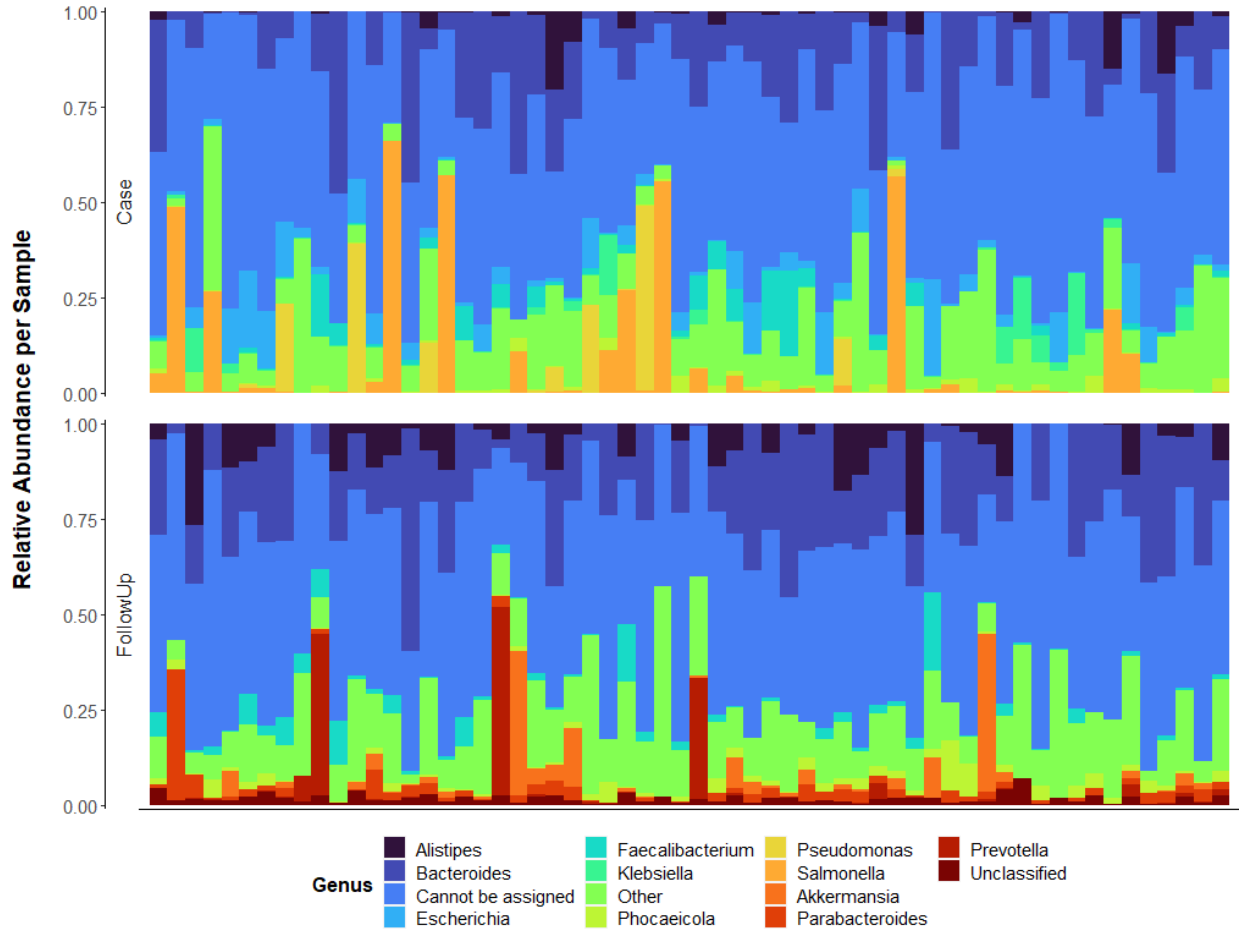
**Figure B.11 (cont'd)**

health status affiliated with each loading (drawn from differential abundance analyses; cases (green) and follow-ups (purple)). **(B)** The species composition gradient is shown overlaid onto an ordination plot based on Bray-Curtis dissimilarity of case and follow-up microbiomes at the species level. Cases (circles), follow-ups (squares), and individuals who received antibiotics (triangles) are shown. The color gradient (“Score”) refers to the continuous structure score affiliated with Loading 1. Juxtaposition of (A) and (B) allow interpretation of species tradeoffs that occur within the sample set (e.g., many cases contain higher levels of *Escherichia coli* at the expense of more beneficial bacteria such as *Bacteroides* species associated with the opposite direction).



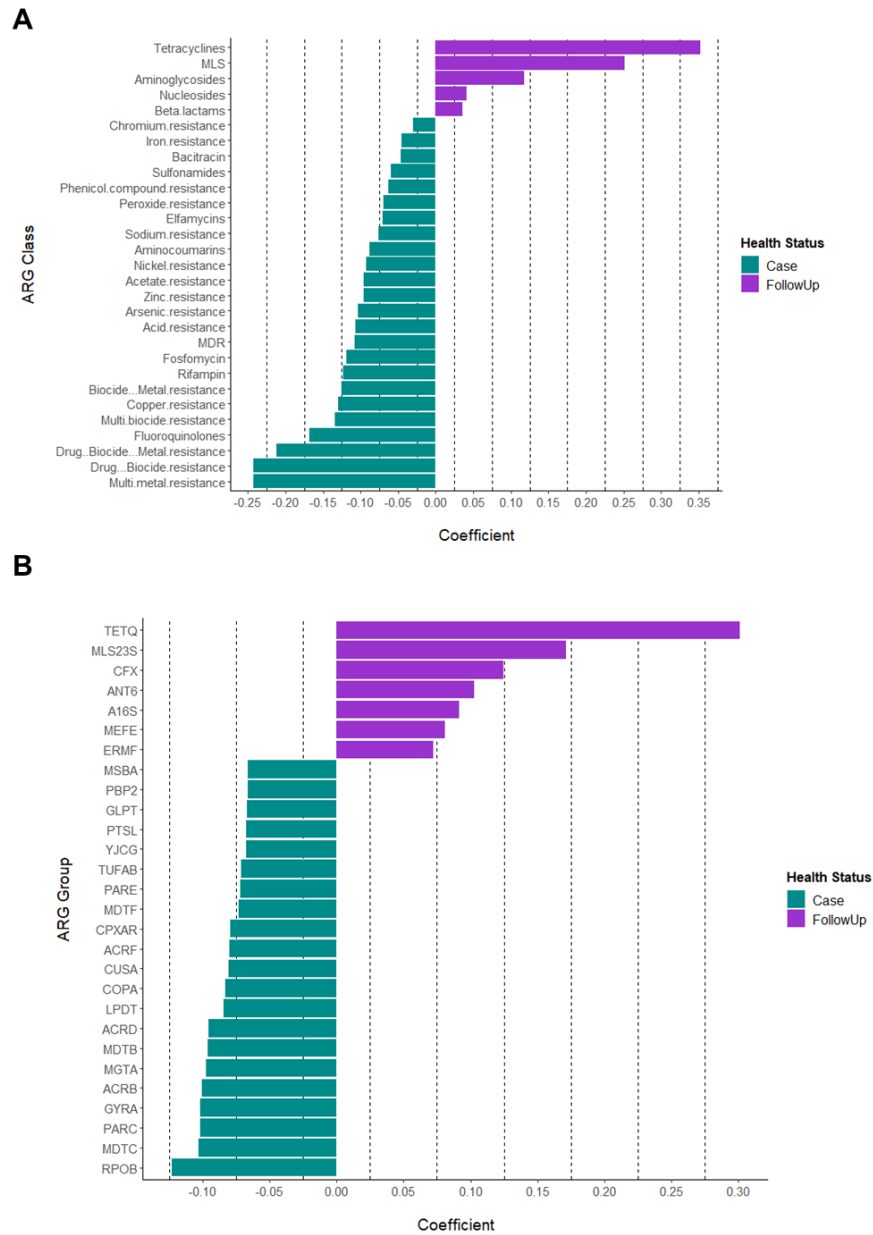
**Figure B.12. Relative abundance of resistance gene types and classes among cases and follow-ups.**

The relative abundance of resistance genes assigned to 4 different compound types (**A**) and 44 resistance classes (**B**) is shown for each health status, with each column representing the resistome from one individual. Columns are ordered by their sample pairing, meaning that the column position in each side of the plot refers to the same individual either during (Case; Left) or after (FollowUp; Right) enteric infection. Relative abundances were determined using raw gene abundances that had been normalized by the estimated number of genome equivalents in the sample as determined using MicrobeCensus.

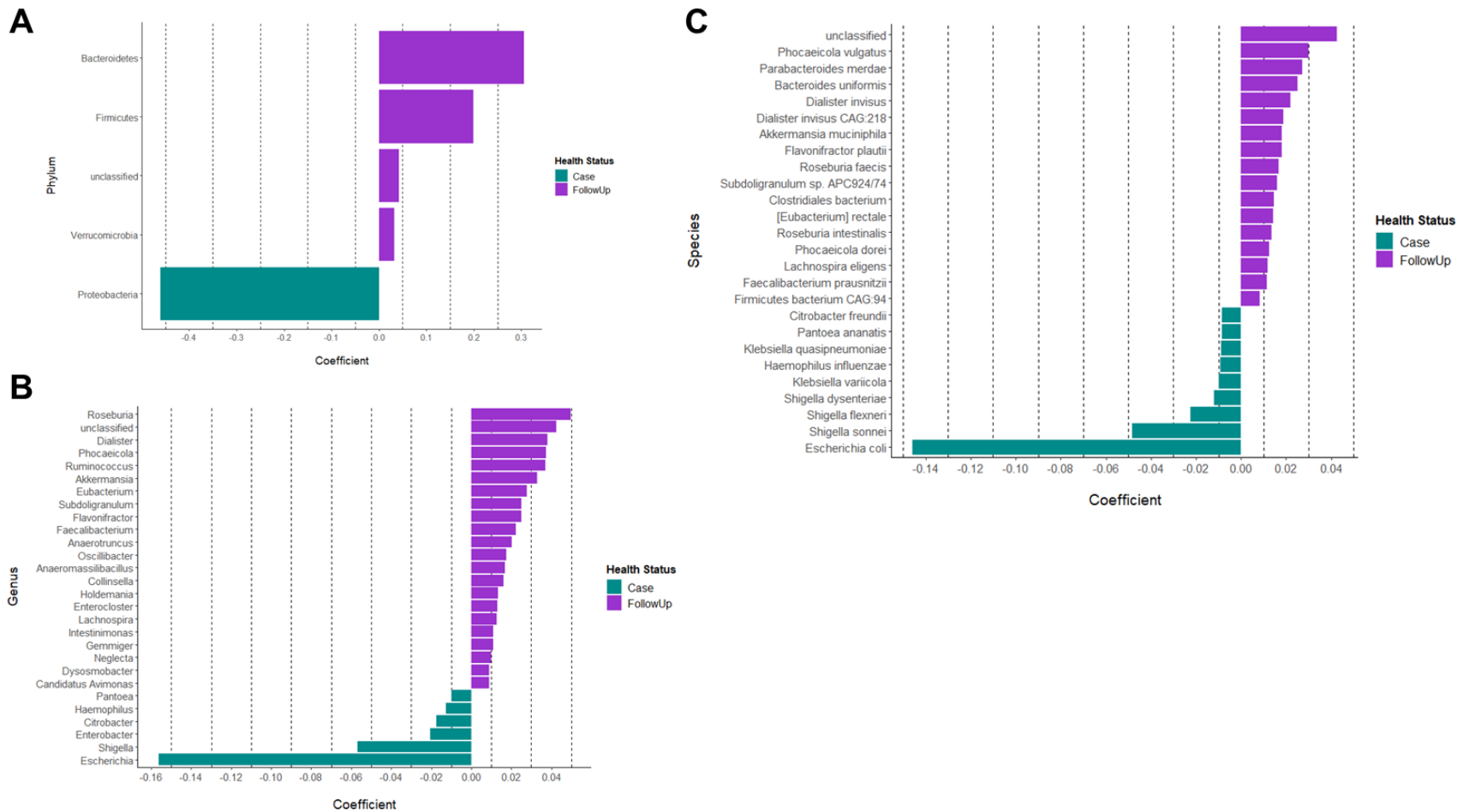


**Figure B.13. Relative abundance of microbial genera notably differ between cases and follow-ups.**

The top-10 microbial genus with the greatest average relative abundance among cases or follow-ups is shown for each health status with each column representing the microbiome from one individual. Columns are ordered by their sample pairing, meaning that the column position for each facet of the plot refers to the same individual either during (Case; Top) or after (FollowUp; Bottom) enteric infection. Relative abundances were determined using raw gene abundances that had been normalized by the approximate number of genome equivalents in the sample as determined using MicrobeCensus.

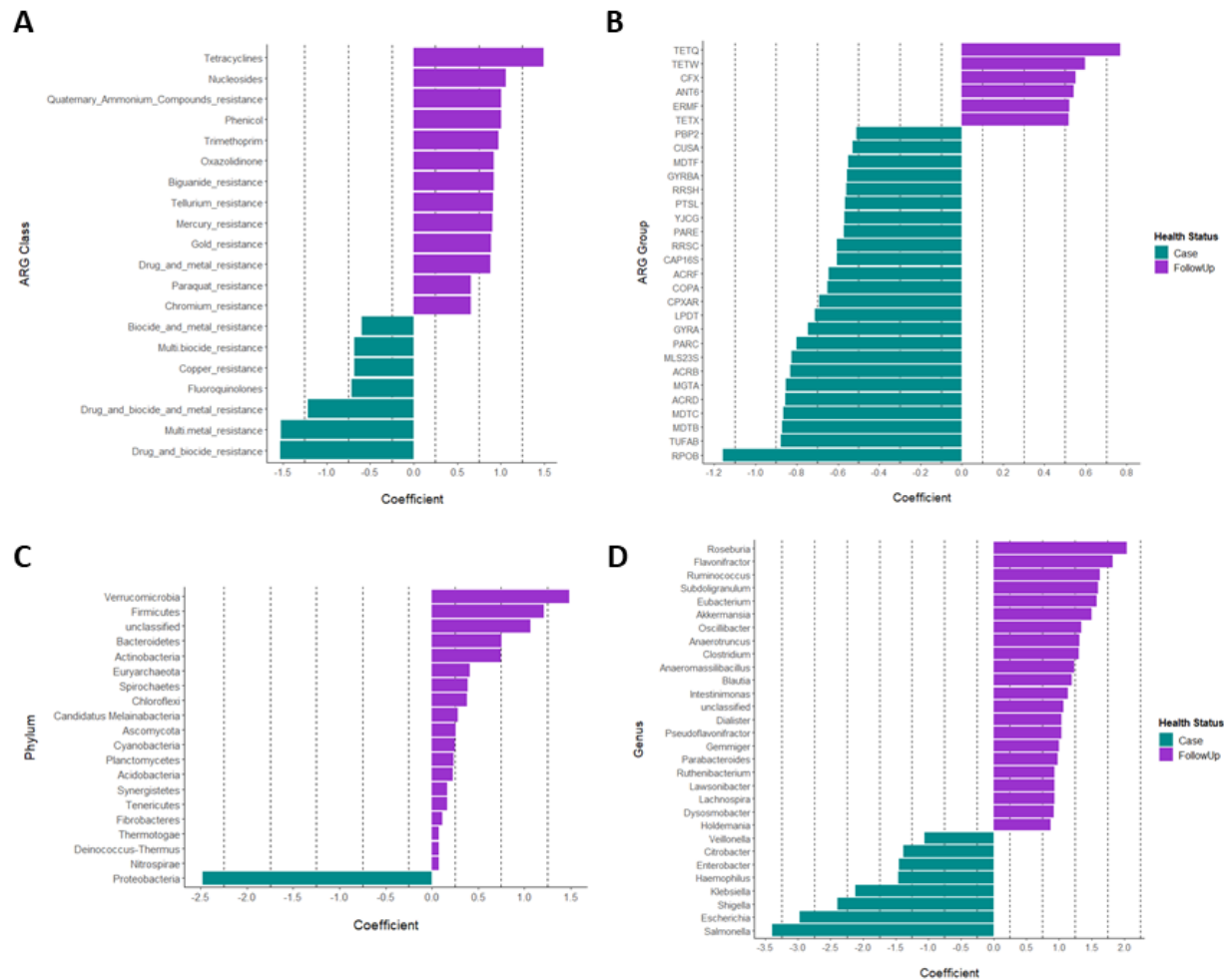


**Figure B.14. Differentially abundant ARG classes and groups among cases and follow-ups.** MMUPHin was used to identify ARG features of differential abundance. Coefficients for each ARG class (**A**) or ARG group (**B**) are shown on the x-axis with a cutoff of absolute value = 0.05; positive coefficients indicate ARG features which were more abundant among follow-ups, while those with negative coefficients were more prominent in cases. The specific ARG features are displayed on the y-axis. The bars in the plot are colored by health status with which that particular feature is associated (cases=green, follow-ups=purple).



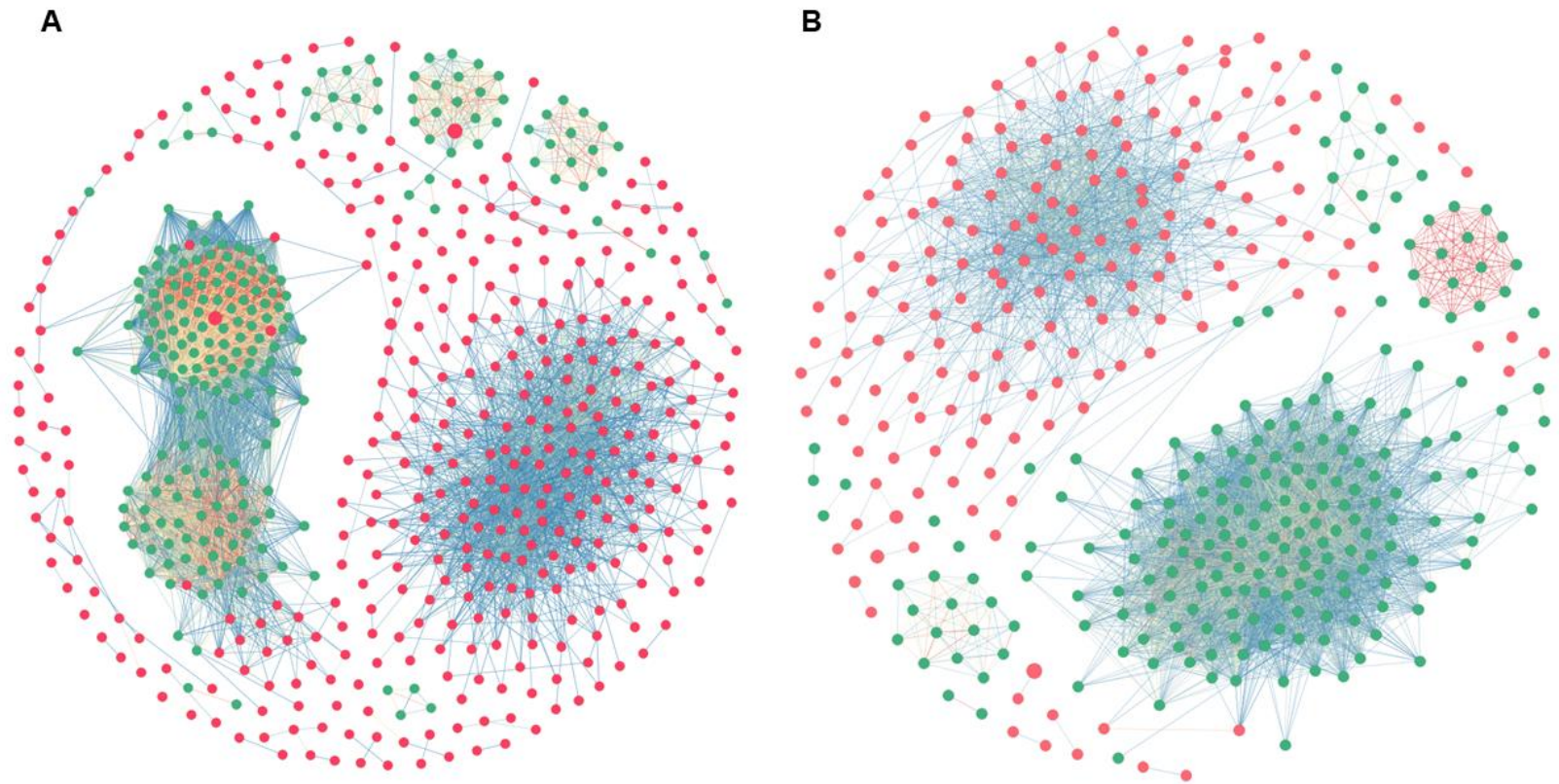
**Figure B.15. Differential abundance of phyla, genera, and species among cases and follow-ups.**

MMUPHin was used to identify microbial features of differential abundance. Coefficients for each phylum (A) genus (B) or species (C) are shown on the x-axis with a cutoff of absolute value=0.05, 0.008, and 0.008, respectively; positive coefficients indicate features which were more abundant among follow-ups, while those with negative coefficients were more prominent in cases. The specific taxonomic features are displayed on the y-axis. Bars in the plot are colored by health status with which that particular feature is associated (cases=green, follow-ups=purple).



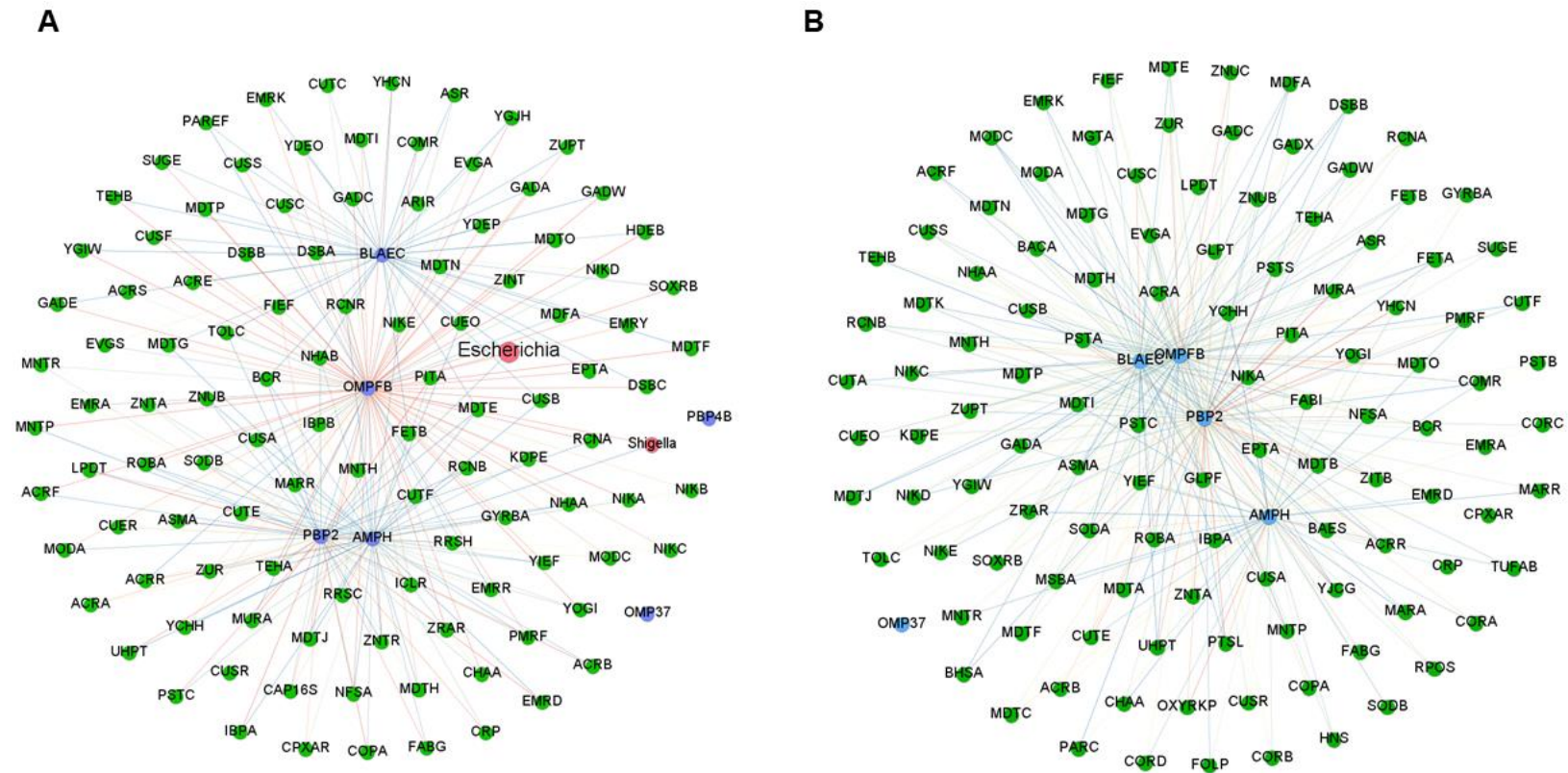
**Figure B.16. Differential abundance of taxonomic and resistance gene features for cases and follow-ups with ANCOM-BC.** ANCOM-BC was used as a secondary method to compare and confirm the findings from our differential abundance analysis using MMUPHin. Differentially abundance ARG classes (A), ARG groups (B), taxonomic phyla (C) and genera (D) are shown. The x-axis displays the range of coefficients associated with each feature, which is shown on the y-axis. In each sub-plot, purple bars designate resistome or microbiome features that were found to be more abundant in follow-ups and green bars show those that were more represented in cases





**Figure B.17. Global co-occurrence network analysis reveals interesting patterns among cases and follow-ups.**

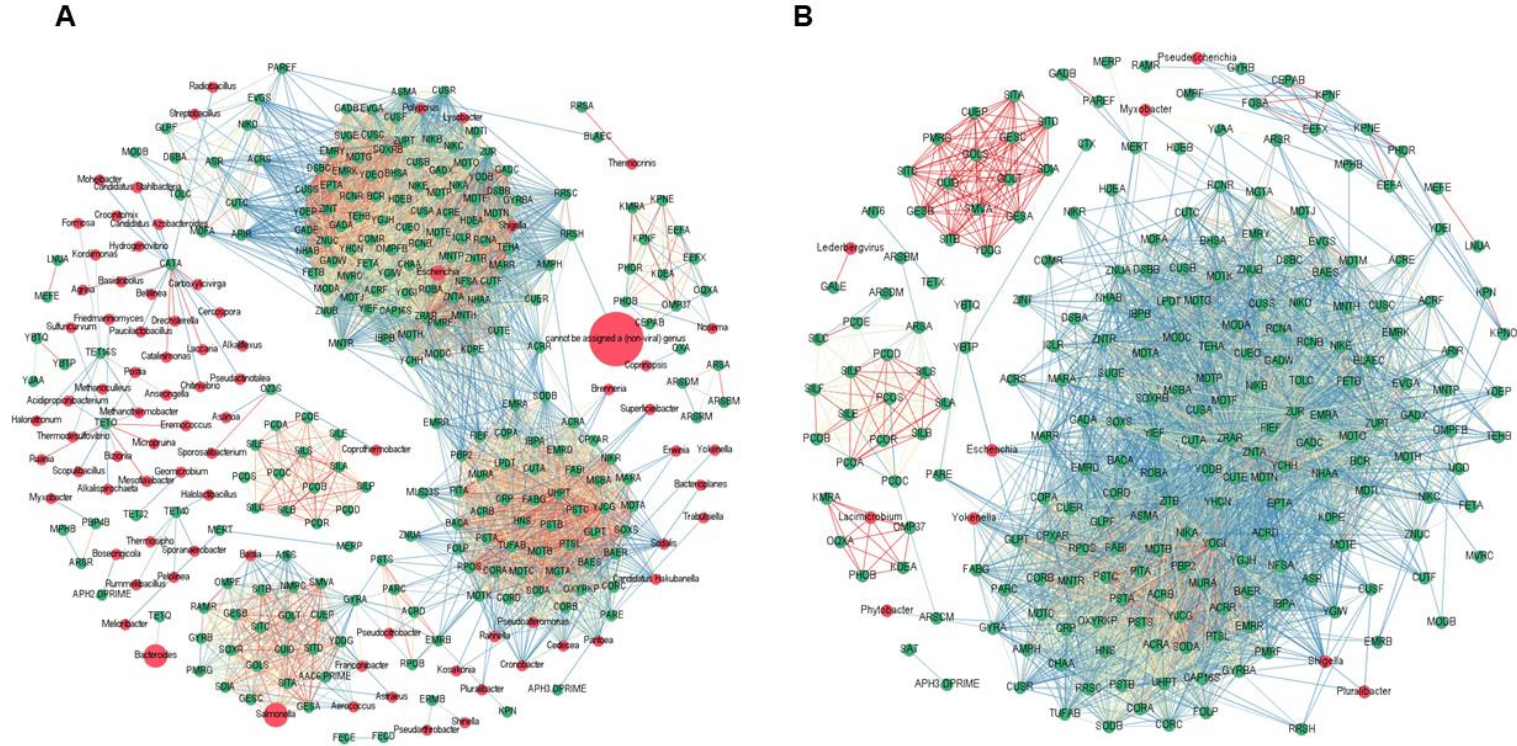
Correlation co-occurrence networks were constructed in Gephi 0.9.2 using Spearman's Rank correlation coefficients generated with the R-package 'Hmisc' (v4.5-0) for cases (A) and follow-ups (B). Correlations included in the network all passed a cutoff of  $\rho > 0.80$  and  $q\text{-value} < 0.05$ . Nodes are colored by their identity as a taxonomic genus (red) or ARG group (green). Nodes are sized based on their overall abundance among samples; the larger the node, the more abundant. Nodes with  $\geq 1$  connection were included (i.e. degree cutoff=1). The edge color displays the strength of correlation, with blue demonstrating relatively weaker correlations (yet still  $> 0.80$ ), yellow showing medium correlation, and red showing strong correlation.



**Figure B.18. Co-occurrence of Beta-lactam ARGs with other ARGs and taxa varies between cases and follow-ups.**

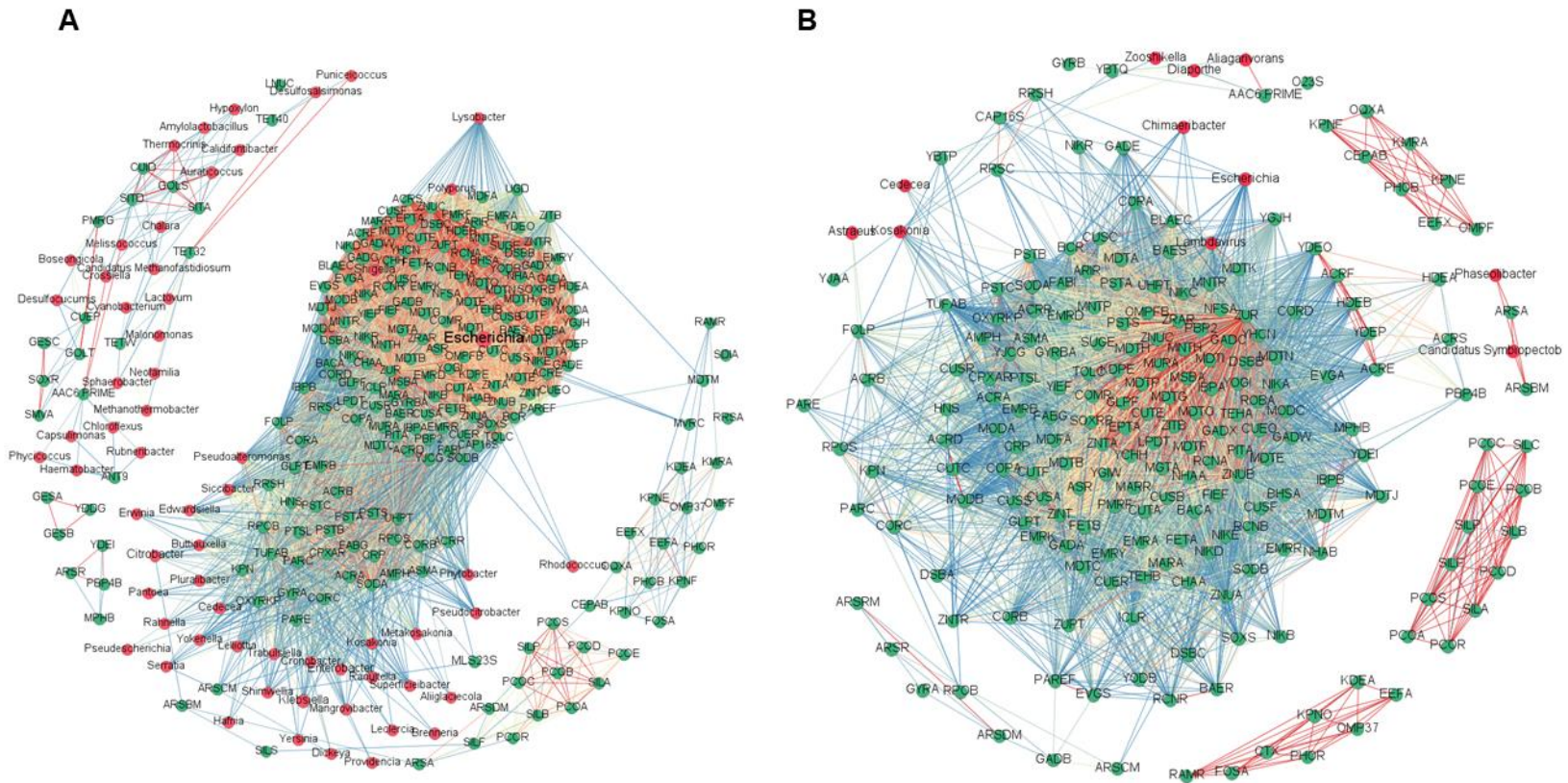
Correlation co-occurrence networks were constructed in Gephi 0.9.2 using Spearman's Rank correlation coefficients generated with the R-package 'Hmisc' (v4.5-0) for cases (A) and follow-ups (B). These networks display all relevant ARG-ARG and ARG-taxa connections for beta-lactam ARGs, specifically. Correlations included in the network all passed a cutoff of  $\rho > 0.80$  and  $q\text{-value} < 0.05$ . Nodes are colored by their identity as a taxonomic genus (red) or ARG group (green). Nodes are sized based on their overall abundance among samples; the larger the node, the more abundant. Nodes with  $\geq 2$  connection were included (i.e. degree cutoff=2). The edge color displays the strength of correlation, with blue demonstrating relatively weaker correlations (yet still  $> 0.80$ ), yellow showing medium correlation, and red showing strong correlation. Nodes are labeled with their corresponding genus or ARG group.





**Figure B.19. ARG-ARG and ARG-taxa connections are different for individuals infected with and recovering from *Salmonella*.**

Correlation co-occurrence networks were constructed in Gephi 0.9.2 using Spearman's Rank correlation coefficients generated with the R-package 'Hmisc' (v4.5-0) for cases infected with *Salmonella* (A) or recovering (follow-ups) (B). These networks display all ARG-ARG and ARG-taxa connections; taxa-taxa connections were excluded for clarity. Correlations included in the network all passed a cutoff of  $\rho > 0.80$  and  $q\text{-value} < 0.05$ . Nodes are colored by their identity as a taxonomic genus (red) or ARG group (green). Nodes are sized based on their overall abundance among samples; the larger the node, the more abundant. Nodes with  $\geq 1$  connection were included (i.e. degree cutoff=1). The edge color displays the strength of correlation, with blue demonstrating relatively weaker correlations (yet still  $> 0.80$ ), yellow showing medium correlation, and red showing strong correlation. Nodes are labeled with their corresponding genus or ARG group.



**Figure B.20. ARG-ARG and ARG-taxa connections are different for individuals infected with and recovering from *Campylobacter*.**

Correlation co-occurrence networks were constructed in Gephi 0.9.2 using Spearman's Rank correlation coefficients generated with the R-package 'Hmisc' (v4.5-0) for cases infected with *Campylobacter* (A) and recovering follow-ups (B). These networks display all ARG-ARG and ARG-taxa connections; taxa-taxa connections were excluded for clarity. Correlations included in the network all passed a cutoff of  $\rho > 0.80$  and  $q\text{-value} < 0.05$ . Nodes are colored by their identity as a taxonomic genus (red) or ARG group (green). Nodes are sized based on their overall abundance among samples; the larger the node, the more abundant. Nodes with  $\geq 2$  connections were included (i.e. degree cutoff=2). The edge color displays the strength of correlation, with blue demonstrating relatively weaker correlations (yet still  $> 0.80$ ), yellow showing medium correlation, and red showing strong correlation. Nodes are labeled with their corresponding genus or ARG group.

## **REFERENCES**

## REFERENCES

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne Illness Acquired in the United States -- Major Pathogens. *Emerging Infectious Disease* 17:7-15.
2. Tack DM, Marder EP, Griffin PM, Cieslak PR, Dunn J, Hurd S, Scallan E, Lathrop S, Muse A, Ryan P, Smith K, Tobin-D'Angelo M, Vugia DJ, Holt KG, Wolpert BJ, Tauxe R, Geissler AL. 2019. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2015–2018. *MMWR Morbidity and Mortality Weekly Report* 68:369-373.
3. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* 3:45-45.
4. Van Schaik W. 2015. The human gut resistome. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20140087.
5. Hansen ZA, Cha W, Nohomovich B, Newton DW, Lephart P, Salimnia H, Khalife W, Shade A, Rudrik JT, Manning SD. 2021. Comparing gut resistome composition among patients with acute *Campylobacter* infections and healthy family members. *Scientific Reports* 11.
6. Huang AD, Luo C, Pena-Gonzalez A, Weigand MR, Tarr CL, Konstantinidis KT. 2017. Metagenomics of Two Severe Foodborne Outbreaks Provides Diagnostic Signatures and Signs of Coinfection Not Attainable by Traditional Methods. *Applied and Environmental Microbiology* 83:e02577-16.
7. Argüello H, Estellé J, Zaldívar-López S, Jiménez-Marín Á, Carvajal A, López-Bascón MA, Crispie F, O'Sullivan O, Cotter PD, Priego-Capote F, Morera L, Garrido JJ. 2018. Early *Salmonella* Typhimurium infection in pigs disrupts Microbiome composition and functionality principally at the ileum mucosa. *Scientific Reports* 8.
8. Haag L-M, Fischer A, Otto B, Plickert R, Kühl AA, Göbel UB, Bereswill S, Heimesaat MM. 2012. Intestinal Microbiota Shifts towards Elevated Commensal *Escherichia coli* Loads Abrogate Colonization Resistance against *Campylobacter jejuni* in Mice. *PLoS ONE* 7:e35988.
9. Yang J, Chen W, Xia P, Zhang W. 2020. Dynamic comparison of gut microbiota of mice infected with *Shigella flexneri* via two different infective routes. *Experimental and Therapeutic Medicine*.
10. CDC. 2019. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA.

11. Wallace MJ, Fishbein SRS, Dantas G. 2020. Antimicrobial resistance in enteric bacteria: current state and next-generation solutions. *Gut Microbes* 12:e1799654.
12. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The shared antibiotic resistome of soil bacteria and human pathogens. *Science (New York, NY)* 337:1107-11.
13. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, Calderon M, Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G. 2016. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533:212-216.
14. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, De Vos WM, Zucker J-D, Raes J, Hansen T, Bork P, Wang J, Ehrlich SD, Pedersen O. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500:541-546.
15. Lozupone C. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* 489:220-230.
16. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, Hansen TH, Liang S, Feng Q, Zhang C, Pyl PT, Coelho LP, Yang H, Wang J, Typas A, Nielsen MF, Nielsen HB, Bork P, Wang J, Vilsbøll T, Hansen T, Knop FK, Arumugam M, Pedersen O. 2018. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology* 3:1255-1265.
17. Dethlefsen L, Huse S, Sogin ML, Relman DA. 2008. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biology* 6:e280.
18. Antonopoulos DA, Huse SM, Morrison HG, Schmidt TM, Sogin ML, Young VB. 2009. Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation. *Infection and Immunity* 77:2367-2375.
19. Reid G, Howard J, Siang Gan B. 2001. Can bacterial interference prevent infection? *Trends in Microbiology* 9:424-428.
20. Khosravi A, Mazmanian SK. 2013. Disruption of the gut microbiome as a risk factor for microbial infections. *Current Opinion in Microbiology* 16:221-227.
21. Sassone-Corsi M, Raffatellu M. 2015. No Vacancy: How Beneficial Microbes Cooperate with Immunity To Provide Colonization Resistance to Pathogens. *The Journal of Immunology* 194:4081-4087.



22. Lustrì BC, Sperandio V, Moreira CG. 2017. Bacterial Chat: Intestinal Metabolites and Signals in Host-Microbiota Pathogen Interactions. *Infection and Immunity* 85:e00476-17.
23. Sitaraman R. 2018. Prokaryotic horizontal gene transfer within the human holobiont: ecological-evolutionary inferences, implications and possibilities. *Microbiome* 6.
24. Stokes HW, Gillings MR. 2011. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiology Reviews* 35:790-819.
25. Chow J, Tang H, Mazmanian SK. 2011. Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Current Opinion in Immunology* 23:473-480.
26. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326:289-293.
27. Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, Zhang T. 2015. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME Journal* 9:2490-2502.
28. Feng J, Li B, Jiang X, Yang Y, Wells GF, Zhang T, Li X. 2018. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environmental Microbiology* 20:355-368.
29. Ma L, Li B, Jiang X-T, Wang Y-L, Xia Y, Li A-D, Zhang T. 2017. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* 5:154-154.
30. Ingram DD, Franco SJ. 2014. 2013 NCHS Urban-rural Classification Scheme for Counties, 166 ed, vol Stat 2. National Center for Health Statistics.
31. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, Belk KE, Noyes NR, Morley PS. 2020. MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research* 48:D561-D569.
32. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
33. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25:2078-9.



35. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *BIOINFORMATICS APPLICATIONS NOTE* 26:841-842.
36. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Pidcock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. *Antimicrobial Agents and Chemotherapy* 57:3348-3357.
37. Nayfach S, Pollard KS. 2015. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biology* 16:51-51.
38. Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* 30:629-635.
39. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7:11257-11257.
40. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Maintainer HW. 2019. Package 'vegan' Title Community Ecology Package. *Community ecology package* 2.
41. Team RC. 2017. R: A language and environment for statistical computing., *on R Foundation for Statistical Computing*. <https://www.R-project.org/>. Accessed
42. Ma S. 2021. MMUPHin: Meta-analysis Methods with Uniform Pipeline for Heterogeneity in Microbiome Studies, vol R package version 1.8.0.
43. Nearing JT, Douglas GM, Hayes MG, Macdonald J, Desai DK, Allward N, Jones CMA, Wright RJ, Dhanani AS, Comeau AM, Langille MGI. 2022. Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications* 13.
44. Lin H, Peddada SD. 2020. Analysis of compositions of microbiomes with bias correction. *Nature Communications* 11.
45. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Sun F, Zhou J, Knight R. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal* 10:1669-1681.
46. Harrell Jr FE. 2021. Harrell Miscellaneous. <https://hbiostat.org/R/Hmisc/>. Accessed
47. Bastian M, Heymann S. 2009. Gephi : An Open Source Software for Exploring and Manipulating Networks.361-362.

48. Bushnell B. BBMAP. [sourceforge.net/projects/bbmap/](https://sourceforge.net/projects/bbmap/). Accessed
49. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674-1676.
50. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* 32.
51. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, Mckay LJ, Mcdaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* 6:3-6.
52. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification.
53. Li Y, Xu Z, Han W, Cao H, Umarov R, Yan A, Fan M, Chen H, Duarte CM, Li L, Ho P-L, Gao X. 2021. HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes. *Microbiome* 9.
54. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12:59-60.
55. Li H. seqtk. <https://github.com/lh3/seqtk>. Accessed
56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
58. Shaikh S, Fatima J, Shakil S, Danish Rizvi SM, Kamal MA. 2015. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. *Saudi Journal of Biological Sciences* 22:90-101.
59. Kriss M, Hazleton KZ, Nusbacher NM, Martin CG, Lozupone CA. 2018. Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Current Opinion in Microbiology* 44:34-40.
60. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications* 8.
61. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, Pan Y, Li J, Zhu L, Wang X, Meng Z, Zhao F, Liu D, Ma J, Qin N, Xiang C, Xiao Y, Li L, Yang H, Wang J, Yang R, Gao GF,

- Wang J, Zhu B. 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications* 4.
62. Clemente C, Jose, Ursell K, Luke, Parfrey W, Laura, Knight R. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* 148:1258-1270.
  63. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, De Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174-180.
  64. Gibiino G, Lopetuso LR, Scaldaferri F, Rizzatti G, Binda C, Gasbarrini A. 2018. Exploring Bacteroidetes: Metabolic key points and immunological tricks of our gut commensals. *Digestive and Liver Disease* 50:635-639.
  65. Spor A, Koren O, Ley R. 2011. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology* 9:279-290.
  66. Peterson DA, Frank DN, Pace NR, Gordon JI. 2008. Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases. *Cell Host & Microbe* 3:417-427.
  67. Sockett PN, Rodgers FG. 2001. Enteric and foodborne disease in children: A review of the influence of food- and environment-related risk factors. *Paediatric Child Health* 6.
  68. Smith JL. 1998. Foodborne Illness in the Elderly†. *Journal of Food Protection* 61:1229-1239.
  69. Scallan E, Crim SM, Runkle A, Henao OL, Mahon BE, Hoekstra RM, Griffin PM. 2015. Bacterial Enteric Infections Among Older Adults in the United States: Foodborne Diseases Active Surveillance Network, 1996–2012. *Foodborne Pathogens and Disease* 12:492-499.
  70. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature* 489:220-230.
  71. CDC. 2018. National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Surveillance Report for 2015 (Final Report). U.S. Department of Health and Human Services, Atlanta, Georgia.
  72. Ranjbar R, Farahani A. 2019. Shigella: Antibiotic-Resistance Mechanisms And New Horizons For Treatment.
  73. Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. 2015. Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial

- Management of Invasive Salmonella Infections. *Clinical Microbiology Reviews* 28:901-937.
74. Dominguez E, Zarazaga M, Saenz Y, Brinas L, Torres C. 2002. Mechanisms of Antibiotic Resistance in *Escherichia coli* Isolates Obtained from Healthy Children in Spain. *Microbial Drug Resistance* 8.
  75. Bartoloni A, Pallecchi L, Benedetti M, Fernandez C, Vallejos Y, Guzman E, Villagran AL, Mantella A, Lucchetti C, Bartalesi F, Strohmeyer M, Bechini A, Gamboa H, Rodríguez H, Falkenberg T, Kronvall G, Gotuzzo E, Paradisi F, Rossolini GM. 2006. Multidrug-resistant Commensal *Escherichia coli* in Children, Peru and Bolivia. *Emerging Infectious Diseases* 12:907-913.
  76. Brolund A, Rajer F, Giske CG, Melefors Ö, Titelman E, Sandegren L. 2019. Dynamics of Resistance Plasmids in Extended-Spectrum- $\beta$ -Lactamase-Producing Enterobacteriaceae during Postinfection Colonization. *Antimicrobial Agents and Chemotherapy* 63.
  77. Titelman E, Hasan CM, Iversen A, Nauc ler P, Kais M, Kalin M, Giske CG. 2014. Faecal carriage of extended-spectrum  $\beta$ -lactamase-producing Enterobacteriaceae is common 12 months after infection and is related to strain factors. *Clinical Microbiology and Infection* 20:O508-O515.
  78. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* 73:1121-1137.
  79. Bailey JK, Pinyon JL, Anantham S, Hall RM. 2010. Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *Journal of Medical Microbiology* 59:1331-1339.
  80. Ghorbani-Dalini S, Kargar M, Doosti A, Abbasi P, Sarshar M. 2015. Molecular Epidemiology of ESBLs Genes and Multi-Drug Resistance in Diarrheagenic *Escherichia coli* Strains Isolated from Adults in Iran. *Iranian Journal of Pharmaceutical Research* 14:1257-1262.
  81. Chuppava B, Keller B, Abd El-Wahab A, Surie C, Visscher C. 2019. Resistance Reservoirs and Multi-Drug Resistance of Commensal *Escherichia coli* From Excreta and Manure Isolated in Broiler Houses With Different Flooring Designs. *Frontiers in Microbiology* 10.
  82. Miles TD, McLaughlin W, Brown PD. 2006. Antimicrobial resistance of *Escherichia coli* isolates from broiler chickens and humans. *BMC Veterinary Research* 2:7.
  83. Deng W, Quan Y, Yang S, Guo L, Zhang X, Liu S, Chen S, Zhou K, He L, Li B, Gu Y, Zhao S, Zou L. 2018. Antibiotic Resistance in *Salmonella* from Retail Foods of Animal Origin and Its Association with Disinfectant and Heavy Metal Resistance. *Microbial Drug Resistance* 24:782-791.

84. Pal C, Asiani K, Arya S, Rensing C, Stekel DJ, Larsson DGJ, Hobman JL. 2017. Metal Resistance and Its Association With Antibiotic Resistance, p 261-313, *Microbiology of Metal Ions*. Elsevier.
85. Wales A, Davies R. 2015. Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. *Antibiotics* 4:567-604.
86. Imran M, Das KR, Naik MM. 2019. Co-selection of multi-antibiotic resistance in bacterial pathogens in metal and microplastic contaminated environments: An emerging health threat. *Chemosphere* 215:846-857.
87. Shoemaker NB, Vlamakis H, Hayes K, Salyers AA. 2001. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Applied and Environmental Microbiology* 67:561-568.
88. De Vries LE, Vallès Y, Agersø Y, Vaishampayan PA, García-Montaner A, Kuehl JV, Christensen H, Barlow M, Francino MP. 2011. The Gut as Reservoir of Antibiotic Resistance: Microbial Diversity of Tetracycline Resistance in Mother and Infant. *PLoS ONE* 6:e21644.
89. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology* 8:1-6.
90. Aitchison J. 1982. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society* 44:139-177.
91. Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, List M. 2021. Network analysis methods for studying microbial communities: A mini review. *Computational and Structural Biotechnology Journal* 19:2687-2698.
92. Friedman J, Alm EJ. 2012. Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology* 8:e1002687.
93. Yin X, Deng Y, Ma L, Wang Y, Chan LYL, Zhang T. 2019. Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. *Environmental International* 133.
94. Dang C, Xia Y, Zheng M, Liu T, Liu W, Chen Q, Ni J. 2020. Metagenomic insights into the profile of antibiotic resistomes in a large drinking water reservoir. *Environment International* 136.
95. Raza S, Jo H, Kim J, Shin H, Hur H-G, Unno T. 2021. Metagenomic exploration of antibiotic resistome in treated wastewater effluents and their receiving water. *Science of The Total Environment* 765.
96. McInnes RS, McCallum GE, Lamberte LE, Van Schaik W. 2020. Horizontal transfer of antibiotic resistance genes in the human gut microbiome. *Current Opinion in Microbiology* 53:35-43.

97. Shi D, Hao H, Wei Z, Yang D, Yin J, Li H, Chen Z, Yang Z, Chen T, Zhou S, Wu H, Li J, Jin M. 2022. Combined exposure to non-antibiotic pharmaceuticals and antibiotics in the gut synergistically promote the development of multi-drug-resistance in *Escherichia coli*. *Gut Microbes* 14.
98. Hall LV, A., Maurine, Blok M, E., Hetty, Donders T, Rogier, A., Paauw A, Fluit C, Ad, Verhoef J. 2003. Multidrug Resistance among Enterobacteriaceae Is Strongly Associated with the Presence of Integrons and Is Independent of Species or Isolate Origin. *The Journal of Infectious Diseases* 187:251-259.
99. Skurnik D, Le Menac'H A, Zurakowski D, Mazel D, Courvalin P, Denamur E, Andremont A, Ruimy R. 2005. Integron-Associated Antibiotic Resistance and Phylogenetic Grouping of *Escherichia coli* Isolates from Healthy Subjects Free of Recent Antibiotic Exposure. *Antimicrobial Agents and Chemotherapy* 49:3062-3065.
100. TnPedla. 2021. IS Families/IS5 and related IS1182 families. [https://tnpedia.fcav.unesp.br/index.php/IS\\_Families/IS5\\_and\\_related\\_IS1182\\_families#:~:text=of%20this%20element,-,Mechanism%20IS903,addressed%20at%20present%20is%20IS903](https://tnpedia.fcav.unesp.br/index.php/IS_Families/IS5_and_related_IS1182_families#:~:text=of%20this%20element,-,Mechanism%20IS903,addressed%20at%20present%20is%20IS903). Accessed
101. Broaders E, Gahan CGM, Marchesi JR. 2013. Mobile genetic elements of the human gastrointestinal tract. *Gut Microbes* 4:271-280.
102. Francino MP. 2016. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Frontiers in Microbiology* 6.
103. Snyderman DR, Jacobus NV, Mcdermott LA, Goldstein EJC, Harrell L, Jenkins SG, Newton D, Patel R, Hecht DW. 2017. Trends in antimicrobial resistance among *Bacteroides* species and *Parabacteroides* species in the United States from 2010–2012 with comparison to 2008–2009. *Anaerobe* 43:21-26.
104. Kierzkowska M, Majewska A, Mlynarczyk G. 2020. Trends and Impact in Antimicrobial Resistance Among *Bacteroides* and *Parabacteroides* Species in 2007-2012 Compared to 2013-2017. *Microbial Drug Resistance* 26.
105. Qiu Q, Wang J, Yan Y, Roy B, Chen Y, Shang X, Dou T, Han L. 2020. Metagenomic Analysis Reveals the Distribution of Antibiotic Resistance Genes in a Large-Scale Population of Healthy Individuals and Patients with Varied Diseases. *Frontiers in Molecular Biosciences* 7.
106. Teunis PFM, Evers EG, Hengeveld PD, Dierikx CM, Wielders CCCH, Van Duijkeren E. 2018. Time to acquire and lose carriage of ESBL/pAmpC producing *E. coli* in humans in the Netherlands. *PLOS ONE* 13:e0193834.
107. Doi Y, Adams-Haduch JM, Peleg AY, D'Agata EMC. 2012. The role of horizontal gene transfer in the dissemination of extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolates in an endemic setting. *Diagnostic Microbiology and Infectious Disease* 74:34-38.

108. Dunn SJ, Connor C, McNally A. 2019. The evolution and transmission of multi-drug resistant *Escherichia coli* and *Klebsiella pneumoniae*: the complexity of clones and plasmids. *Current Opinion in Microbiology* 51:51-56.
109. Whittle G, Shoemaker NB, Salyers AA. 2002. The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cellular and Molecular Life Sciences (CMLS)* 59:2044-2054.
110. Zeng MY, Inohara N, Nuñez G. 2017. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunology* 10:18-26.
111. Stecher B, Denzler R, Maier L, Bernet F, Sanders MJ, Pickard DJ, Barthel M, Westendorf AM, Krogfelt KA, Walker AW, Ackermann M, Dobrindt U, Thomson NR, Hardt W-D. 2012. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal *Enterobacteriaceae*. *Proceedings of the National Academy of Sciences* 109:1269-1274.
112. Shintani M, Sanchez ZK, Kimbara K. 2015. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy, vol 6.
113. Carr VR, Shkoporov A, Hill C, Mullany P, Moyes DL. 2021. Probing the Mobilome: Discoveries in the Dynamic Microbiome. *Trends in Microbiology* 29.
114. Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic acids research* 46:e35-e35.
115. Jones BV, Marchesi JR. 2007. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nature Methods* 4:55-61.
116. Beaulaurier J, Zhu S, Deikus G, Mogno I, Zhang XS, Davis-Richardson A, Canepa R, Triplett EW, Faith JJ, Sebra R, Schadt EE, Fang G. 2018. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nature Biotechnology* 36.
117. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. *The ISME Journal*:1-1.
118. Yaffe E, Relman DA. 2020. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nature Microbiology* 5:343-353.

## **CHAPTER 4**

Recovery from enteric infection demonstrates a shift in functional capacity and metabolite composition



## ABSTRACT

While enteric pathogens have been widely studied for their roles in causing foodborne infection, their impacts on the gut microbial community have yet to be fully characterized. Previous work has identified notable changes in the gut microbiome related to pathogen invasion, both taxonomically and genetically. However, characterization of the metabolic landscape during and after enteric infection has not yet been broadly explored. In this study, we investigated the metabolome of paired stools recovered from 60 patients during (cases) and after (follow-ups) enteric infection. To do this, we performed functional pathway prediction of metagenomes and untargeted metabolomics. Pathway prediction methods indicated that cases had a greater overall microbially-mediated metabolic capacity, as these individuals demonstrated significantly higher pathway richness and evenness relative to the follow-up samples ( $p < 0.05$ ). Metabolic pathways more highly represented in cases included variations to central carbon metabolism, amino acid metabolism, lipid and fatty acid biosynthesis, as well as distinct signatures of menaquinone production. Follow-up samples, however, showed greater diversity of the actual metabolic landscape; untargeted metabolomics resulted in significantly greater richness of polar metabolites ( $p < 0.0001$ ) as well as significantly greater richness, evenness, and overall diversity among nonpolar metabolites ( $p < 0.0001$ ). Many of the metabolites identified in this analysis were unknown and unable to be annotated with existing databases. Despite this limitation, we observed marked increases in certain clusters of metabolites among recovered vs. infected patients, implicating their importance in gut health and recovery. Collectively, these data aid in further understanding metabolic fluctuations important to the ecology of the gut during enteric infection.

## INTRODUCTION

Microbes in the human gut have long been understood to contribute to host metabolic health. Generally, gut microbiota are critical for breaking down complex carbohydrates and converting various compounds into forms usable by the body (1). More specifically, these microorganisms generate beneficial short-chain fatty acids (SCFAs) such as butyrate, acetate, and propionate which play crucial roles in counteracting inflammation and immune disorders (2). Butyrate, for example, influences gut epithelial health and can be derived from acetate, another SCFA (3). Acetate has also been identified as a potential therapeutic that may counteract the detrimental effects of obesity (4). Notably, however, the successful production of these compounds relies on specific members of the gut community, namely microbes in the Bacteroidetes and Firmicutes phyla (5, 6), whose abundance can change with diet (7) and other perturbations.

Indeed, gut microbial composition is associated with disease state. Impacts of disease, both chronic and acute, on the human gut microbiome and metabolome have increasingly been characterized, with particular emphasis on the use of multi-omics approaches to identify differences in host-microbiota metabolic interactions (8). Interestingly, defining a “healthy gut” is not straightforward; even within individuals, there is considerable variation in microbial and metabolic profiles due to changes in diet and exercise, for instance, as well as host genetics (9). Jumpertz et al. (2011) (10) demonstrated that altered nutrient load resulted in compositional changes to the gut microbiota that yielded a shift in nutrient absorption capacity. Colonic transit time, too, has been studied as a potential indicator of gut microbiome health, as transit times were correlated with various metabolic products (11). Adding to the complexity of analyzing these systems is the strong influence of the gut microbiota on host metabolic health (12), with more diverse microbiomes demonstrating a reduced risk of metabolism-related disease (13).

However, even members of the same microbial groups can have varying effects on human metabolism. For example, a previous study explored the impacts of ‘structural variants’ among microbes of the same species and found that genetic variation among these organisms can lead to differences in metabolic capacity and, thus, host health (14). Indeed, the interplay between the fluctuating human gut microbiome and related metabolic consequences is of continued interest in the context of disease.

Numerous studies have explored microbial differences between ‘healthy’ individuals and those with various health conditions including obesity, liver disease, metabolic disease, diabetes, and inflammatory bowel disease (IBD) (15-17). For example, obesity has been found to be associated with an increased Firmicutes:Bacteroidetes ratio and lower overall bacterial diversity (18). As noted, diet can play a crucial role in developing certain disease states such as obesity or Type 2 diabetes; in fact, prolonged consumption of foods low in dietary fiber was linked to the extinction of beneficial microbes (19). Additionally, another study demonstrated that long-term high-fat diets selected for microbes that produce lipopolysaccharide (LPS), an endotoxin that contributes to inflammation linked to obesity and insulin resistance (20). Sokol et al. (2009) (21) found that patients experiencing different forms of colitis harbored fewer members of the phylum Firmicutes, with notable underrepresentation of the species *Faecalibacterium prausnitzii*; this species is known to contribute heavily to SCFA generation in the gut (6). Moreover, IBD patients were found to contain different compositions of gut-derived bile acids compared to healthy individuals, a difference that may directly influence the host response to intestinal inflammation (22). Indeed, the intersection of gut microbiota and metabolism and their impact on human health is of great relevance.

Of particular interest to this study are metabolic shifts due to enteric disease. Indeed, it is estimated that enteric pathogens cause more than 9.4 million foodborne infections in the United States each year (23). The Centers for Disease Control and Prevention (CDC) has recently reported increased incidence of infections caused by *Campylobacter* and Shiga toxin-producing *Escherichia coli* (STEC), while pathogens such as *Salmonella* and *Shigella* maintained a high level of incidence (24). While extensive work continues to explore the intersection of chronic disease, the gut microbiome, and metabolic consequences, few works have explored the impact of acute enteric infection on metabolic shifts within the gut. Indeed, further exploration of the human gut microbiome and its influence on host metabolic health during periods of acute infection is needed. Because previous studies in our lab have documented shifts in microbial composition using 16S rRNA sequencing (25) and metagenomics (Chapter 3) as well as changes in resistance gene composition and diversity (26) following enteric infections, an examination of metabolic dynamics is also warranted. This study, therefore, aims to characterize metabolic trajectories of individuals during and after enteric infection using metagenome analysis and untargeted metabolomics.

## **METHODS**

### **Study population**

Between 2011 and 2015, 61 stool samples were obtained from patients experiencing enteric bacterial infections prior to treatment and designated for metagenome analyses. Of these patients, 25 (41.0%) had infections caused by *Campylobacter*, 29 (47.5%) had *Salmonella* infections, and 4 (6.6%) and 3 (4.9%) experienced *Shigella* or Shiga toxin-producing *E. coli* (STEC) infections, respectively. Stools from these patients were collected through the Michigan Department of Health and Human Services (MDHHS) as described previously (25). Follow-up

samples (n=61) were also collected by each patient after they recovered from the infection and provided informed consent; these samples were sent directly to Michigan State University (MSU). Patients reported information about demographics, exposures, hospitalization, and symptoms through the Michigan Disease Surveillance System (MDSS) and answered a follow-up questionnaire at the time the post-recovery sample was submitted. County of residence was classified as 'rural' or 'urban' based on the classification scheme developed by the National Center for Health Statistics (27).

### **Sample preparation and metagenome sequencing analysis**

Metagenomic DNA from the 122 fecal samples was extracted, sheared, and normalized as described previously (25). Briefly, libraries were constructed using a TruSeq Nano library kit (Illumina, Inc., San Diego, CA, USA) and shotgun metagenomics sequencing was performed in a series of four runs on an Illumina HiSeq 2500. Reads were demultiplexed at the MSU Research Technology Support Facility (RTSF). Metagenomic sequencing reads were processed using the AmrPlusPlus v2.0 pipeline as described in our previous work (Chapter 3). Non-host FASTQ files generated in this workflow were used for metagenome assembly and input to the HUMAnN 3.0 program to characterize metabolic profiles. Of note, two samples (a case and follow-up pair) were removed from the metagenomics pipeline due to poor annotation and assembly resulting in a total of 118 samples. However, metabolomics was completed on all samples (n=122), explaining the small discrepancy in sample numbers between these two methods.

### **Metagenome assembly and metabolic prediction profiling with Anvi'o**

Pre-assembly processing, metagenome assembly with MEGAHIT, and generation of contig databases with Anvi'o (28) was performed as described previously (Chapter 3). To begin

metabolic profiling with Anvi'o, the function 'anvi-run-kegg-kofams' was used to annotate the contig databases with HMM hits from the KOfam database which houses KEGG Orthologs (KOs). Then, 'anvi-estimate-metabolism' was run; this program uses the annotated contig databases to determine which metabolic enzymes are present, thus defining the metabolic functions within each sample in addition to module completeness. Two output files were generated per sample, one which contained specific KOfam hits and their assignment to relevant KEGG modules (if applicable) and one which contained KEGG module names, subcategories, categories, and completeness scores. A custom Python script was constructed to filter KEGG modules based on a module completeness cutoff of 0.70. The resulting data tables for each sample were merged to form a comprehensive dataset. Subsequently, frequency of module occurrence was determined by summing the number of contigs to which a module was assigned. For example, if a module was found on 50 different contigs within a sample, that module registered a frequency of 50. The metric of frequency was used in place of abundance for these analyses. Prior to statistical analysis, module frequencies were normalized by the number of genome equivalents (GE) per sample as determined by MicrobeCensus (29). Notably, these module frequencies were used as hypothesis-generating observations which were followed by more rigorous metabolic prediction analyses and untargeted metabolomics.

### **Metabolic prediction profiling with HUMAnN 3.0**

The third iteration of the HMP Unified Metabolic Analysis Network (HUMAnN 3.0) was used to profile the abundance of microbial metabolic pathways and other relevant functions from our 120 metagenome samples (30). Non-host paired end reads were first merged for input into HUMAnN 3.0, which was run using the UniRef90 database. The program generates three output files for each sample: gene family abundances, pathway abundances, and pathway coverage

estimates. The gene families file for each sample was fed into the program ‘humann\_infer\_taxonomy’, which retrieves approximate taxonomic information for translated search results that had prior to been assigned “unclassified.” This function was run at the genus level.

Upon inference of genera associated with gene families and pathways, the resulting files were re-run through HUMAnN 3.0 to compute pathway abundances and coverage affiliated with these newly inferred taxonomic assignments. The resulting gene families and pathway abundance files, which inherently report reads-per-kilobase (RPK) values, were normalized to relative abundances using the ‘humann\_renorm\_table’ function. Separate sample tables were then joined to create a comprehensive dataset. The ‘humann\_regroup\_table’ function was used to modify the gene family abundance table to display gene families with MetaCyc reaction annotations rather than UniRef90 (the default output based on our earlier parameters). Next, ‘humann\_rename\_table’ was used to assign MetaCyc pathway names to the pathway abundance and coverage files for easier interpretation; the modified pathways abundance file was used for downstream interpretation and analysis. The ‘humann\_barplot’ function was used to produce plots of stratified metabolic features; these plots displayed taxa abundances assigned to a specific MetaCyc pathway. Various parameters were used to sort by other variables (namely case status and pathogen type) and scale the data (original vs. logstack scaling). Plots were sorted in two different ways, either by metadata assignment or Bray-Curtis dissimilarity.

## **Ecological analyses of metagenome and metabolome data**

### *Abundance and diversity analyses*

The diversity of predicted metabolic profiles was determined by investigating the composition of metabolic pathways across infected cases and recovered follow-ups. Abundances

for KEGG modules (Anvi'o pipeline) or MetaCyc pathways (HUMAN 3.0 pipeline) were used as input for diversity analyses. Alpha and beta diversity metrics were calculated and plotted in R (31) as described in previous work (26). When comparing diversity metrics across infecting pathogen, only comparisons between individuals infected with *Campylobacter* and *Salmonella* were included, as the samples sizes associated with *Shigella* and STEC infections were markedly smaller.

#### *Differential abundance of metabolic pathways*

To assess representative features in cases and follow-ups, the R-package Meta-analysis Methods with Uniform Pipeline for Heterogeneity in Microbiome Studies (MMUPHin) was used to construct general linear models exploring module and pathway abundances (32). MMUPHin required relative abundance as input. First, we performed batch adjustment of relative abundance data based on sequencing run since this variable was previously identified as driving a degree of stratification among samples. Next, a linear model was constructed to identify differentially abundant pathways among cases and follow-ups; follow-ups were used as the reference for the fixed effect, while age in years, number of genome equivalents, gender, and use of antibiotics were included as covariates in the model. Significance values were adjusted using the Benjamini-Hochberg method of correction for multiple hypothesis testing (q-value representing False Discovery Rate (FDR)).

#### *Identification of continuous population structure*

MMUPHin was also used to further characterize intrinsic drivers of point distributions observed in beta diversity analyses (ordination). A gradient of module and pathway abundance was suspected due to the distribution of points observed in our PCoA plots. To classify this



potential gradient, the ‘continuous\_discover()’ function was applied to our metabolic profile abundance data. This function performs unsupervised continuous structure discovery using Principal Components Analysis (PCA). Upon generation of these continuous structure scores (called “loadings”), we constructed respective plots to visualize the main drivers of continuous data structure. Loadings that comprise the top principal components were compared across batches to identify “consensus” loadings assigned to certain microbial features. These data were overlaid onto ordination plots which displayed the Bray-Curtis dissimilarity. Continuous structure scores were indicated by a color gradient and cases and follow-ups were shown as differently shaped points. In nearly every comparison, the distribution of points could be attributed to metabolic tradeoff identified in the loadings scores.

### **Metabolite extraction**

Metabolite extractions were performed for 122 human stool samples (cases=61; follow-ups=61). Prior to extraction, the following five internal standard solutions were prepared for downstream quality control and normalization: 1)  $^{13}\text{C}$ -labeled short-chain fatty acids (SCFAs) (10  $\mu\text{M}$  each of  $^{13}\text{C}$ sodium formate,  $^{13}\text{C}_2$ sodium acetate,  $^{13}\text{C}_3$ sodium propionate, and  $^{13}\text{C}_4$ sodium butyrate in 50:50 (v/v) methanol/water); 2) phenylalanine- $d_7$  (10  $\mu\text{M}$  in 50:50 methanol/water); 3) succinic acid- $d_4$  (10  $\mu\text{M}$  in 50:50 methanol/water); 4)  $^{13}\text{C}_{16}$ palmitic acid (10  $\mu\text{M}$  in 100% isopropanol); and 5) labeled bile acids (10  $\mu\text{M}$  each of glycocholic acid- $d_4$  and glyoursodeoxycholic acid- $d_4$  in 50:50 methanol/water). 20mg of each fecal specimen was aliquoted into microcentrifuge tubes for metabolite extraction and stored on ice. Next, 350 $\mu\text{l}$  of methanol containing 0.1% BHT was added to the fecal sample and mixed. 10 $\mu\text{l}$  of each internal standard solution (50 $\mu\text{l}$  total) was added to each sample tube and mixed. Samples were then centrifuged at 10,000  $\times g$  at 4°C for 10 minutes. The resulting supernatant was transferred to a

clean microcentrifuge tube; the pellet was washed with 200µl of HPLC-grade isopropanol then centrifuged at 10,000 x g at 4°C for 10 minutes. The supernatant was removed and combined with the previously collected supernatant to form the 'Total Extract' (TE). 100 µl aliquots of the TE for each sample was aliquoted into amber glass autosampler vials in preparation for liquid chromatography-mass spectrometry. A separate microcentrifuge tube was designated for long-term storage of the remaining TE. All extracts were stored at -80°C until further use.

### **Liquid Chromatography Mass Spectrometry (LC/MS)**

Each sample was analyzed using separate reverse phase and hydrophilic interaction liquid chromatography (HILIC) methods to cover a wider range of metabolite space in the study. A Thermo Q-Exactive and Vanquish Ultra High-Performance Liquid Chromatography (UHPLC) system was used for the analysis. For the reverse phase separation, 10µL of sample was injected onto a Waters Acquity Ethylene Bridged Hybrid (BEH)-C18 UPLC column (2.1x100mm) held at 60°C. Compounds were separated using the following gradient with a 0.4 ml/min flow rate: initial conditions were 98% mobile phase A (water + 0.1% formic acid) and 2% mobile phase B (acetonitrile + 0.1% formic acid), hold at 2% B until 1 min, ramp to 100% B at 8 min, hold at 100% B until 10 min, return to 2% B at 10.01 min and hold at 2% B until 12 min. For the HILIC separation, 10 µL of sample was injected onto a Waters BEH-Amide UPLC column (2.1x100mm) held at 60°C. The following gradient run at 0.4 ml/min was used: initial conditions were 100% mobile phase B (10 mM ammonium formate/10 mM ammonium hydroxide in 95:5 acetonitrile/water (v/v) and 0% mobile phase A (10 mM ammonium formate/10 mM ammonium hydroxide in water), hold until 1 min at 100% B, ramp to 40% B at 8 min, hold at 40% B until 10 min, return to 100% B at 10.01 min and hold at 100% B until 12 min.

Mass spectra were acquired for both chromatography methods using the same MS settings. Compounds were ionized by electrospray ionization operating in positive ion mode with a capillary voltage of 3.5 kV, transfer capillary temp at 262.5°C, sheath gas at 50, auxiliary gas at 12.5, probe heater at 425°C, and S-lens RF level at 50. A data-dependent MS/MS method was used to acquire spectra with survey scan settings of 35,000 resolution, AGC target 1e6, maximum inject time 100 ms, and m/z range 100-1500. MS/MS spectra were acquired for the top 5 ions at a resolution setting of 17,500, AGC target 1e5, minimum AGC 5e3, maximum inject time 50 ms, isolation window of 1.5, and fixed first mass at m/z 50, dynamic exclusion setting of 3 s and stepped normalized collision energy settings of 20, 40 and 60.

### **Feature-based Molecular Networking (FBMN) of metabolites**

A molecular network was created with the Feature-Based Molecular Networking (FBMN) workflow (33) on the Global Natural Product Social Molecular Networking (GNPS) site (34). The MS data were first processed with MZMINE2 (35, 36) and the results were exported to GNPS for FBMN analysis. The data were filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and the MS/MS fragment ion tolerance to 0.02 Da. A molecular network was then created where edges were filtered to have a cosine score above 0.7 (nonpolar metabolites) or 0.65 (polar metabolites) and > 4 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in both of the top-10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the size was below this threshold. The spectra in the network were then searched against GNPS spectral libraries (34,

37). The 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 DEREPLICATOR was used to annotate MS/MS spectra (38). The molecular networks were visualized internally in GNPS (39) and externally using Cytoscape software (40).

### **Intensity normalization and Random Forest in R**

Metabolic intensities output by FBMN through GNPS were used for downstream analysis. The cluster index assigned by GNPS was used to associate peak intensities with known metadata. Prior to use in statistical comparisons, all metabolites identified in blank samples were removed from the dataset. Additionally, peak intensities were normalized via sum-scaling. To identify the most important polar and nonpolar metabolites, the random forest method was used; this is a classification algorithm that combines multiple decision trees into an ensemble, thereby reducing error and increasing accuracy of assignments (41). Random forest was completed using the randomForest package (version 4.6-14) in R (42). The algorithm was fed information about the dichotomous variable “health status” and subsequently classified samples based on intensities generated with GNPS; 5,000 decision trees were generated to enhance classification accuracy. For clusters that distinguish the case vs. follow-up samples, molecular networks and structures were explored. In many cases, the compound was uncharacterized and did not have a library ID; in this situation, other clusters in the related molecular network (if existent) were explored for annotation. If one of these other clusters contained compound characterization, the unnamed cluster of interest was discussed in reference to this known compound.

## **Statistical analysis of metabolites using MetaboAnalyst 5.0**

Normalized peak intensity tables for polar and nonpolar metabolites were used as input to complete statistical analysis with MetaboAnalyst 5.0 (43). No further filtering or normalization was pursued in MetaboAnalyst; the sum-scaled intensities were used. The ‘Statistical Analysis [one factor]’ approach was used to complete paired analysis for paired Case-Follow-up samples. First, a paired fold-change (FC) analysis was completed with a FC cutoff of 5.0. A volcano plot exploring nonparametric paired fold-change was generated using a FC cutoff of 5.0 and a false-discovery rate threshold of 0.05 with an assumption of equal variance among groups.

Correlations between features were explored to infer potential co-occurrence. Spearman rank correlation was used with a correlation cutoff of 0.5; a correlation heatmap was also generated using these values. A heatmap displaying the distribution of feature intensity across samples was also generated. Hierarchical clustering of samples was performed by using Euclidean distance with the Ward clustering method and was fitted to the heatmap. A filtered heatmap was constructed to highlight features of importance, displaying the top-50 features based on t-test results.

The ‘Statistical Analysis [metadata table]’ was used for more detailed analysis with consideration of multiple covariates. Correlations between various metabolic features and metadata were explored; health status was used as the primary metadata of interest while controlling for the covariates of Pair ID, infecting pathogen, sequencing run, age, sex, residence type (urban vs. rural), average genome size among microbes in the sample, and number of microbial genome equivalents per sample. The correlation metric used was Spearman Rank correlation. Heatmaps were generated using Euclidean distance and the Ward clustering method for features which were scaled for viewing. Health status and pathogen were plotted on the

heatmap to view distributions of features across these variables. A subset of the top-50 features based on average intensity value was generated.

## **RESULTS**

### **Characteristics of the study population**

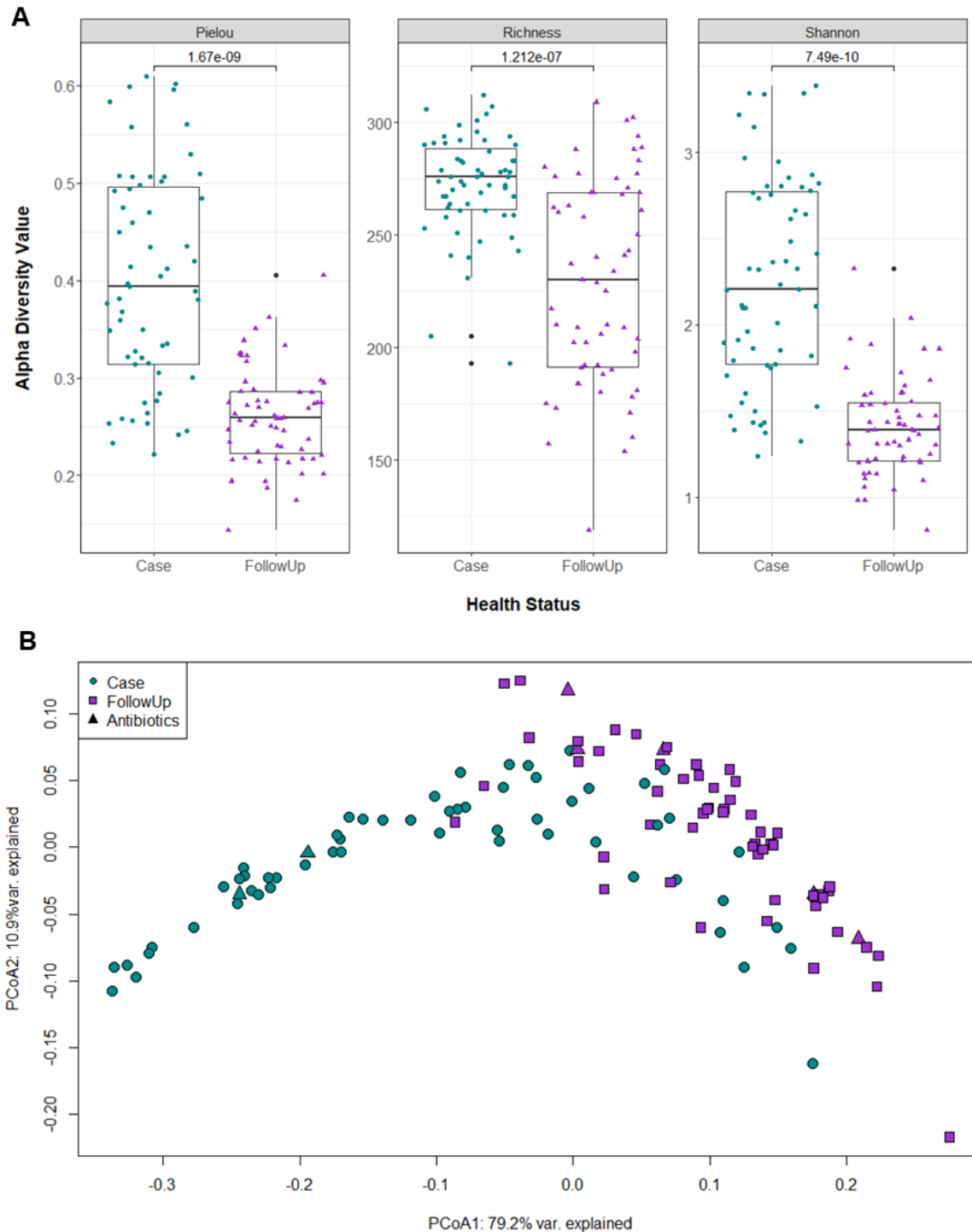
Our study analyzed stools from 61 patients presenting with enteric infection. The same 61 patients (cases) also submitted stools after they recovered from their acute infection (i.e., “follow-ups”). While all samples were included when performing untargeted metabolomics, one case-follow-up pair (which was originally infected with *Campylobacter*) was removed from the metagenome analyses due to poor sequencing quality. The remaining 60 cases were infected with one of the following enteric pathogens: *Salmonella* (n=29), *Campylobacter* (n=24), *Shigella* (n=4), or STEC (n=3). The follow-up period ranged from 8-205 days after the initial infection with an average of 107.9 days; the follow-up period was not known for one case. As described in our prior analysis (Chapter 3), 28 (46.7%) of the cases were male and most were between 19-64 years (n=26; 43.3%), self-identified as Caucasian (n=48; 80.0%) and lived in urban areas (n=33; 55%). Most cases also reported abdominal pain (n=50; 84.8%) and/or diarrhea (n=57; 96.6%) with 17 (28.3%) requiring hospitalization. Only two cases (3.3%) reported antibiotic use within the two weeks prior to stool submission, while five (8.3%) reported using antibiotics within two weeks of submitting their follow-up sample. Moreover, 18 (32.8%) of the 58 cases with data available reported traveling in the month prior to their infection.

### **Variation in the metabolic potential of the gut during and after enteric infection**

When assessing metabolic diversity based on the metagenomics data, one sample was an obvious outlier; this sample and its paired counterpart were removed from the analysis, resulting

in 59 pairs (118 samples). Based on KEGG module frequency predicted with Anvi'o, 338 modules within 41 subcategories and 12 categories were identified among case and follow-up samples. Results generated with HUMAnN exhibit a comparable degree of annotation, and 389 MetaCyc pathways were identified at the community level (i.e., pathways not explicitly assigned to a specific genus). Case samples contained significantly more metabolic pathway signatures than those from follow-ups ( $S_{\text{case}}=272$ ,  $S_{\text{follow}}=230$   $p=1.212e-07$ ; Wilcoxon signed-rank test). Cases also demonstrated more diverse and even metabolic pathways ( $H'_{\text{case}}=2.25$ ,  $H'_{\text{follow}}=1.41$ ;  $p=7.49e-10$  and  $J'_{\text{case}}=0.402$ ,  $J'_{\text{follow}}=0.260$ ;  $p=1.67e-09$ , respectively) (Figure 4.1A), a trend that was reflected in the Anvi'o analysis as well (Figure C.1A). Beta-diversity analysis showed a significant difference in the pathway compositions between cases and follow-ups (PERMANOVA,  $F=62.73$ ;  $p=0.000999$ ); however, the level of dispersion within these groups was also significantly different (PERMDISP,  $F=20.10$ ;  $p=0.001$ ), an indicator that the PERMANOVA results may not be reliable. This nuance is observable when metabolic module composition is captured through ordination, as generous overlap was observed among the two sample types (Figure 4.1B). The KEGG module prediction with Anvi'o produced similar results, with significant differences identified for group centroids and dispersion (Figure C.1B).

The extensive overlap and presence of an arch effect in each ordination plot suggests a gradient of metabolic pathway abundance across cases and follow-ups. To explore this potential continuous structure, MMUPHin was used to correct for batch effects relevant to sequencing run; although this factor had not appeared as an influential driver of sample separation, previous work with these data required batch adjustment (Chapter 3). Unfortunately, the most significant feature was "UNMAPPED", which is a category assigned by HUMAnN to all reads that failed to map to known sequences. This feature was primarily observed in follow-up samples, however,



**Figure 4.1. Predicted MetaCyc pathways identified via HUMAnN 3.0 indicate significant differences in metabolic potential between cases and follow-ups.**

(A) Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou’s Evenness) are displayed. Each boxplot is stratified by health status with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the



**Figure 4.1 (cont'd)**

first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. **(B)** Principal coordinates analysis (PCoA) was performed and plotted for cases (green, circles) and follow-ups (purple, squares) based on Bray-Curtis dissimilarity of community level pathway abundances. The first and second coordinate are displayed with their respective percentage of variance explained. Individuals who reported use of antibiotics  $\geq 2$  weeks prior to sample collection are shown as triangular data points.

suggesting that other non-microbially-mediated pathways are potentially at play in recovering individuals. To investigate differences in annotated pathways more comprehensively, the UNMAPPED category was omitted. In this analysis, rhamnose biosynthesis and histidine degradation were at odds with a superpathway for glycolysis, TCA, and glyoxylate bypass, palmitate biosynthesis, and ornithine degradation (Figure C.2A). Indeed, the continuous structure scores assigned to various features aid in interpreting the distribution of case and follow-up samples within the ordination (Figure C.2B). The KEGG modules identified by Anvi'o also demonstrate notable tradeoffs, with the most striking divergence appearing between subcategories for glycosaminoglycan, cofactor and vitamin metabolism and nitrogen metabolism versus polyamine biosynthesis (Figure C.3A&B). At the KEGG module level, a distinct tradeoff between several nitrogen metabolism pathways such as denitrification and dissimilatory nitrate reduction, and vancomycin resistance pathways were identified (Figure C.3C&D).

**Functional differences in metabolic pathways during and after infection**

Differential abundance of various MetaCyc pathways identified via HUMAnN was explored among cases and follow-ups to further assess differences in metabolic potential. As noted, follow-ups were defined by a high abundance of UNMAPPED reads (coef=0.13; q-value=3.78e-14). No other affiliated features registered a coefficient above 0.01; however, when the UNMAPPED feature was removed, pathway differences were observed between the groups.

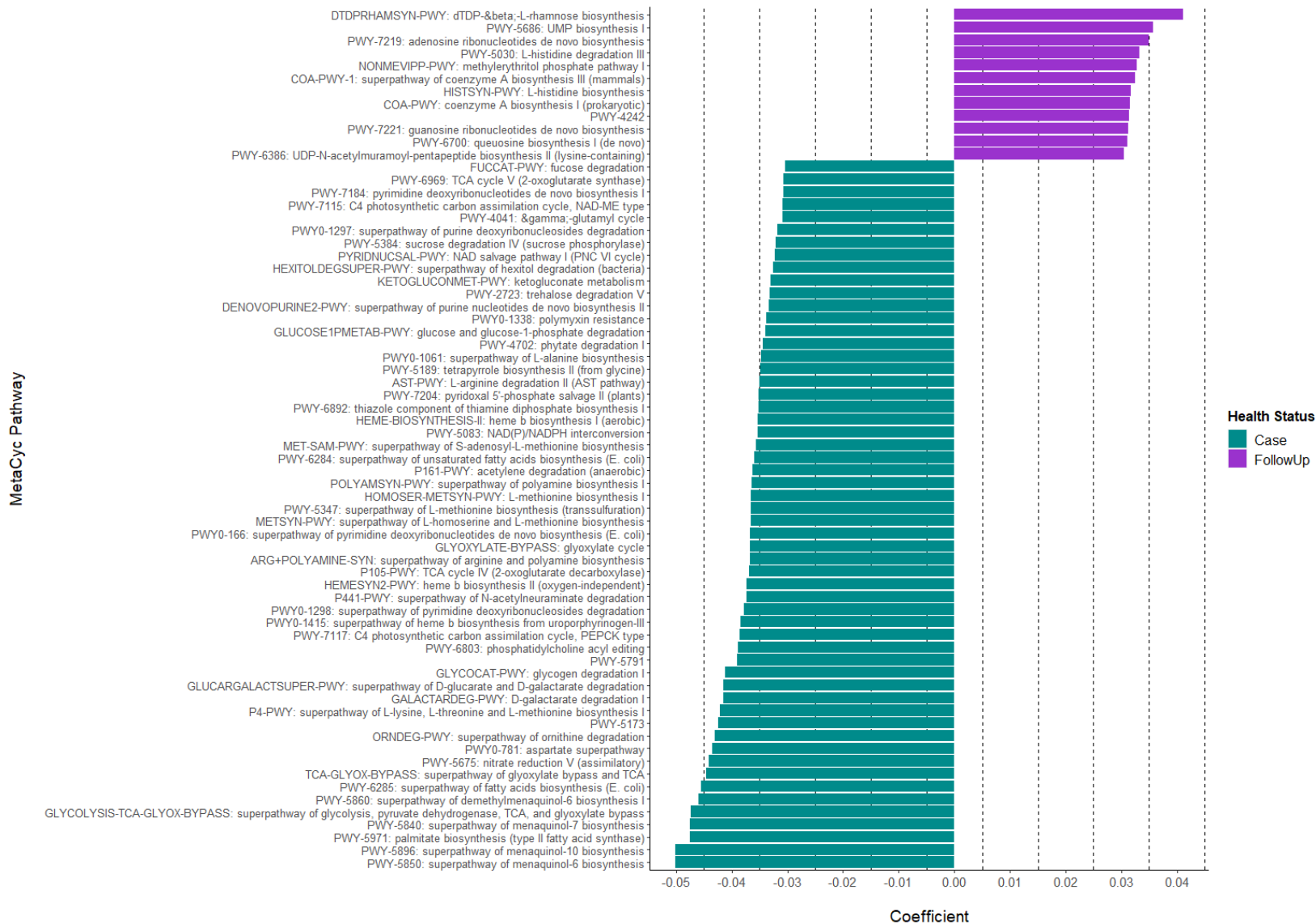
Cases, for instance, had a high abundance of multiple menaquinol biosynthesis pathways including menquinol-10, -6, and -7 (Figure 4.2; Table C.1) as well as palmitate biosynthesis (coef= -0.048; q-value= 0.0061) and glycolysis, TCA, and glyoxylate bypass (coeff= -0.047; q-value= 1.76e-07). Follow-ups, on the other hand, had a high abundance of the L-rhamnose biosynthesis pathway (coef= 0.041; q-value=2.12e-10) and UMP biosynthesis (coef= 0.036; q-value=8.06e-08).

Differentially abundant KEGG pathway subcategories and modules were also explored (Figure C.4). At the subcategory level, cases registered greater frequency of nitrogen metabolism (coef= -0.032 (relative to follow-ups); q-value=1.35e-12), lipopolysaccharide metabolism (coef= -0.021; q-value=0.0015), and “Other” amino acid metabolism (coef= -0.016; q-value=0.00018). Interestingly, beta-lactam biosynthesis was also a defining subcategory among cases (coef= -0.0066; q-value=0.030). Follow-ups, on the other hand, contained a higher frequency of cysteine and methionine metabolism (coef= 0.016; q-value=0.00116), purine metabolism (coef=0.010; q-value=0.020), and central carbohydrate metabolism (coef=0.010; q-value=7.79e-07). The subcategory “Biosynthesis of other antibiotics” was also observed in follow-ups, suggesting residual antibiotic resistance functions upon recovery from infection (coef=0.0070; q-value=0.046). Information specific to subcategories is discussed in Figure C.4. Additionally, relative abundance information for KEGG pathways is explored in Figures C.5, C.6, C.7. At the KEGG module level, cases were heavily represented by the denitrification pathway which reduces nitrate to nitrogen gas (coef= -0.024; q-value=1.22e-11). Nitrate assimilation was also more common in cases (coef= -0.022; q-value=5.11e-09), in addition to dissimilatory nitrate reduction (nitrate → ammonia) (coef= -0.022; q-value=7.25e-08). Notably, three drug resistance pathways (imipenem resistance via OprD, multi-drug resistance via BpeEF-OprC and MexPQ-

OpmE) were also significantly more abundant in cases, a finding in concordance with our earlier work (Chapter 3). At the module level, follow-ups contained higher frequency of vancomycin resistance pathways, specifically the D-Ala-D-Lac type (coef=0.016; q-value=4.82e-07). They also demonstrated increased frequency of trehalose biosynthesis (D-glucose-1P → trehalose) (coef=0.015; q-value=0.0011) and beta-lactam resistance via the Bla system (coef=0.013; q-value=9.95e-06).

### **Specific metabolic pathways differ between sample groups**

As SCFA production has been indicated to play a role in host metabolic health, various MetaCyc pathways related to SCFA production or degradation were explored. Compounds relevant to butyrate, propionate, and acetate were prioritized since these have been implicated as beneficial compounds contributing to host health. Although the MetaCyc PWY-5100: pyruvate fermentation to acetate and lactate II pathway was most abundant in both sample types, it was associated primarily with cases (Figure 4.3; coef= -0.0069; q-value=0.027). The overall relative abundance was comparably low (cases=1.68e-04; follow-ups=1.22e-04); notably, though, these values are consistent with other relative abundance analyses completed with HUMAnN 3.0. Another pathway potentially involved in acetate synthesis was PWY-7254: TCA cycle VII (acetate-producers), which was associated with cases (Figure C.8; coef= -0.029; q-value= 0.00069). By contrast, P163-PWY: L-lysine fermentation to acetate and butanoate, which is relevant to butyrate production, was associated with follow-ups (Figure C.9; coeff=0.0032; q-value=0.022) despite the low abundance in both sample types (cases=1.27e-06; follow-ups=3.14e-06). Other butyrate-specific pathways included PWY-5676: acetyl-CoA fermentation



**Figure 4.2. Differentially abundant MetaCyc pathways among cases and follow-ups with UNMAPPED reads removed.**

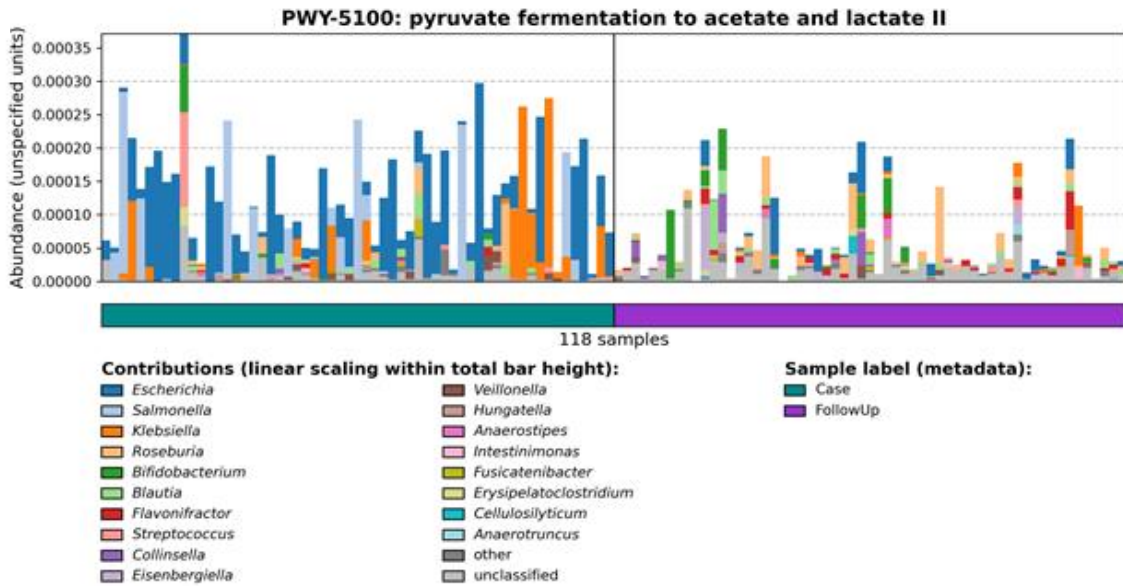
**Figure 4.2 (cont'd)**

Using MMUPHin, differentially abundant metabolic features were detected in cases and follow-ups. Coefficients for each MetaCyc pathway, which are shown on the y-axis, are displayed on the x-axis with an absolute value cutoff of  $>0.030$ . Positive coefficients indicate metabolic pathways with higher abundances in follow-ups (purple) while negative coefficients show pathways more represented among cases (green).

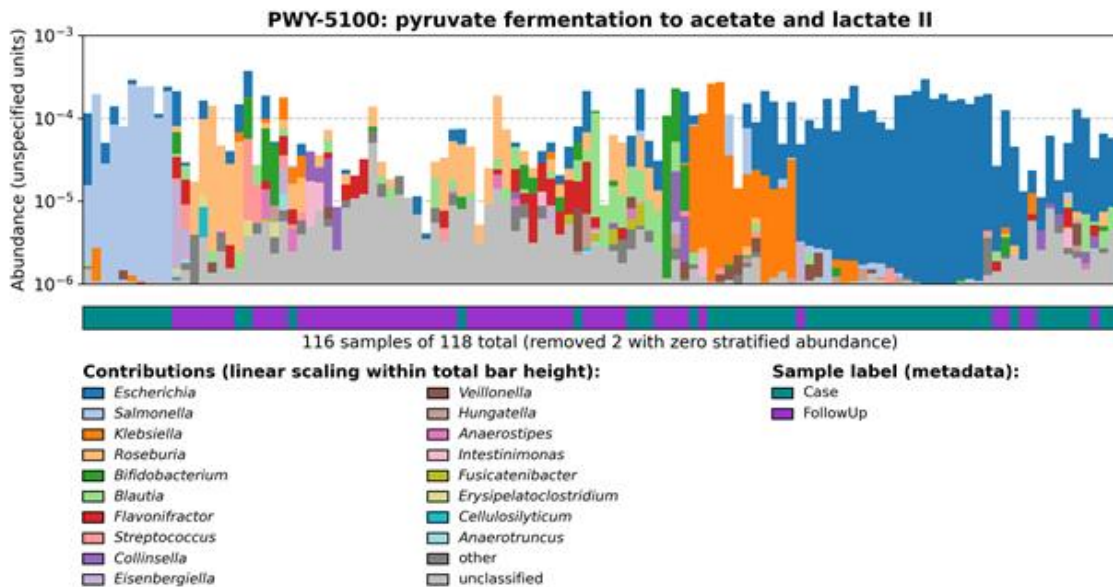
to butanoate II, CENTFERM-PWY: pyruvate fermentation to butanoate, and PWY-5677: succinate fermentation to butanoate (Figure C.10). Although these pathways registered low relative abundance and were not differentially abundant between cases and follow-ups, the distribution of pathways among samples suggests various patterns cannot be captured by statistical analysis alone. Propionate production, for instance, was only identified in one pathway, P108-PWY: pyruvate fermentation to propanoate I, but was not associated with either sample type based on differential abundance. Nonetheless, some interesting patterns in distribution were observed for this pathway among samples and taxa (Figure C.11). All other pathways related to butyrate, acetate, and propionate involved degradation of these compounds. In agreement with our earlier results, a pathway involved in the production of palmitate, another relevant fatty acid in the human body, (PWY-5971: palmitate biosynthesis (type II fatty acid synthase)) was prevalent primarily among cases (Figure C.12).

Various metabolites that have been linked to gut dysbiosis were also explored. For example, the production of lipopolysaccharide (LPS) historically connects to health issues related to this endotoxin. In our samples, we observed the LPSSYN-PWY: superpathway of LPS biosynthesis to be more abundant in cases (coef=  $-0.021$ ; q-value= $2.95e-14$ ), which was somewhat expected due to the activity of gram-negative pathogens. When stratifying by pathogen linked to the acute infections, interesting differences in taxa associated with this pathway were observed (Figure C.13). Presence of *p*-Cresol, a derivative of toluene that has carcinogenic properties, has also been linked to reduced health in the gut. One pathway related to

**A**



**B**



**Figure 4.3. Relative abundances of PWY-5100: pyruvate fermentation to acetate and lactate II among cases and follow-ups.**

Barplots show the A) relative abundance of PWY-5100 calculated by HUMAnN 3.0, and B) the relative abundance clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance. The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest as determined by MetaPhlan 3.0; colors in the stacked barplots show the proportion of relative abundances for PWY-5100 attributed to that specific genus.

*p*-Cresol production, PWY-5181: toluene degradation III (aerobic) (via *p*-cresol) was found (Figure C.14). This pathway demonstrated very low abundances, however, (case=2.09e-05; follow-up=2.50e-06) and had a slight affiliation with cases (coef= -0.0093; q-value=0.0055).

### **Untargeted metabolomics of polar metabolites reveal crucial differences between samples**

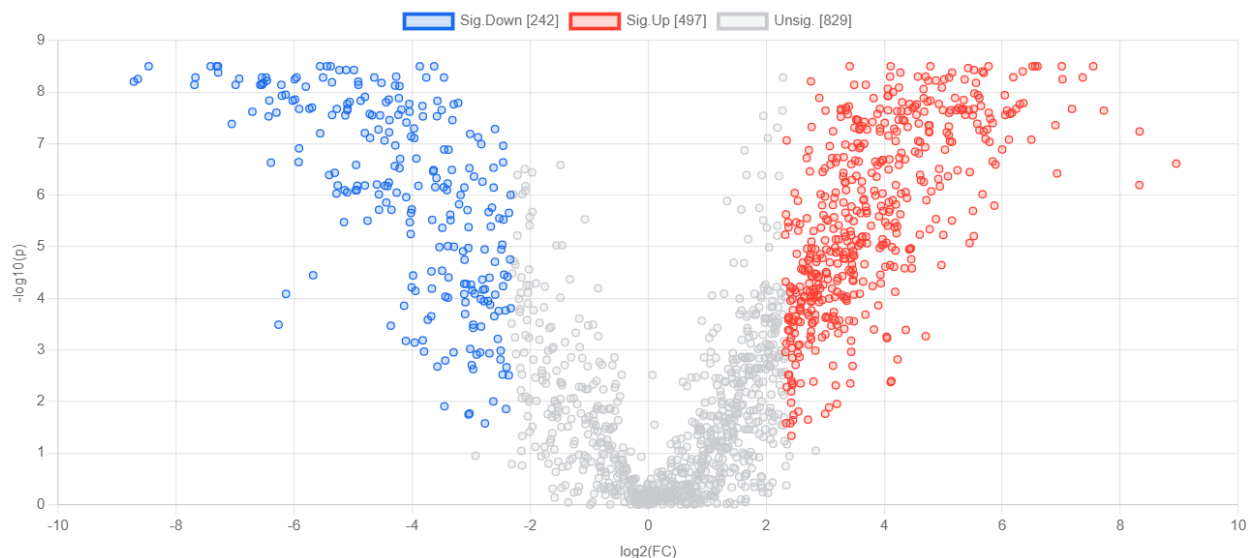
After filtering and normalization, a total of 7,916 polar features were identified among our infected and recovered samples. Overall, follow-ups displayed significantly greater richness of polar metabolites than cases ( $S_{\text{case}}=875$ ,  $S_{\text{follow}}=1024$   $p=2.28e-07$ ; Wilcoxon signed-rank test), though no significant difference in Shannon diversity was observed ( $H'_{\text{case}}=5.00$ ,  $H'_{\text{follow}}=5.07$ ;  $p=0.8971$ ). Intriguingly, cases showed greater metabolite evenness ( $J'_{\text{case}}=0.739$ ,  $J'_{\text{follow}}=0.731$ ;  $p=0.008211$ , respectively) (Figure C.15A). No significant differences in diversity of polar metabolites were observed when samples were stratified by the two predominant pathogens (*Campylobacter* and *Salmonella*) linked to the acute infection (Figure C.16A). PCoA based on Bray-Curtis dissimilarity revealed distinct clustering of polar metabolites between case and follow-up samples (PERMANOVA F-value=26.27; p-value=0.000999; Figure C.15B), with greater dispersion among cases (PERMDISP p-value=0.026). Additionally, no distinction was observed between the four pathogens (PERMANOVA F-value=1.260; p-value=0.1209 | PERMDISP p-value=0.013; Figure C.16B).

Random forest of normalized peak intensities for polar metabolites identified various features that could distinguish between case and follow-up sample types (Figure C.17). The top-30 clusters most important to health status classification are shown in Table C.2 with library IDs (if found). The out-of-bag estimate of error rate for our random forest classification was 5.74%, suggesting high accuracy in assigning health status to our sample based on metabolite composition (Table C.3). Cluster 313 was deemed most important in distinguishing cases from

follow-ups and registered a mean decrease in accuracy (MDA) score of 13.86; this compound was found to be elevated among cases, specifically, but did not appear to be affiliated with a specific pathogen (Figure C.18). The next most important compound was Cluster 2705 (MDA=13.27), which was elevated among follow-ups with no differences in the abundances after stratifying by the pathogen linked to the initial infection (Figure C.19).

Investigation of paired statistical analysis using MetaboAnalyst v5.0 further characterized associations between different polar features and health status. A fold-change (FC) analysis detected metabolites present in one group or the other (Figure 4.4; Table C.4). Of the polar metabolites considered, 497 were increased in follow-ups relative to cases (i.e., a positive  $\log_2\text{FC}$  value). Fewer ( $n=242$ ) experienced a negative  $\log_2\text{FC}$  with regards to follow-ups, suggesting that these polar compounds play a role primarily during infection only. Notably, there were three clusters in the top-10 most positive FC values (associated with follow-ups) that were located in a molecular network with tomatidine (Figure C.20). These included Cluster 326 ( $\log_2\text{FC}=8.95$ ;  $p\text{-value}=2.42\text{e-}07$ ), Cluster 7558 ( $\log_2\text{FC}=8.32$ ;  $p\text{-value}=6.33\text{e-}07$ ), and Cluster 1593 ( $\log_2\text{FC}=6.93$ ;  $p\text{-value}=3.76\text{e-}07$ ). Another cluster increased in follow-ups, Cluster 2113 ( $\log_2\text{FC}=7.72$ ;  $p\text{-value}=2.27\text{e-}08$ ), was part of a dense molecular network including the annotated compounds desmethylenylnocardamine and a spectral match to Nonaethylene glycol (Figure C.21). Cluster 2666 was also elevated in follow-ups ( $\log_2\text{FC}=6.51$ ;  $p\text{-value}=3.14\text{e-}09$ ) and was annotated as 1-(1Z-Hexadecenyl)-sn-glycero-3-phosphocholine (Figure C.22). Of the clusters displaying a negative  $\log_2\text{FC}$  (i.e., affiliated with cases), the strongest signals were from Clusters 970 ( $\log_2\text{FC}= -8.72$ ;  $p\text{-value}=6.22\text{e-}09$ ) and 221 ( $\log_2\text{FC}= -8.65$ ;  $p\text{-value}=5.56\text{e-}09$ ), which were located in the same molecular network. This network contained multiple annotated compounds, but Clusters 970 (Figure C.23) and 221 (Figure 4.5) were both directly connected to





**Figure 4.4. Volcano plot demonstrating fold-change of polar metabolites in cases and follow-ups.**

Fold-change (FC) analysis was performed to explore differentially abundant metabolites among samples. A FC cutoff of 5.0 and false discovery rate (FDR) threshold of 0.05 were set to identify the strongest signals. The volcano plot displays metabolites that were significantly more represented in follow-ups (“Sig.Up”, red, positive  $\log_2FC$ ) and those in cases (“Sig.Down”, blue; negative  $\log_2FC$ ). The x-axis indicates the  $\log_2FC$  value; the y-axis shows the  $-\log_{10}(P)$  value. Metabolites that lacked significant associations with either group are shown as gray dots (“Unsig.”). The legend at the top of the plot indicates the number of metabolites in each category.

Cluster 318, which was a spectral match for 1-(1Z-Octadecenyl)-sn-glycero-3-phosphocholine.

Cluster 221 also had connections to four other nodes annotated as variations of

glycerophosphocholine compounds including Cluster 227 (Lyso-PAF C-18), Cluster 1337 (1-

Heptadecanoyl-sn-glycero-3-phosphocholine), Cluster 6245 (1-Hexadecyl-sn-glycero-3-

phosphocholine), and Cluster 259 (sn-glycero-3-phosphocholine). Notably, each of these clusters

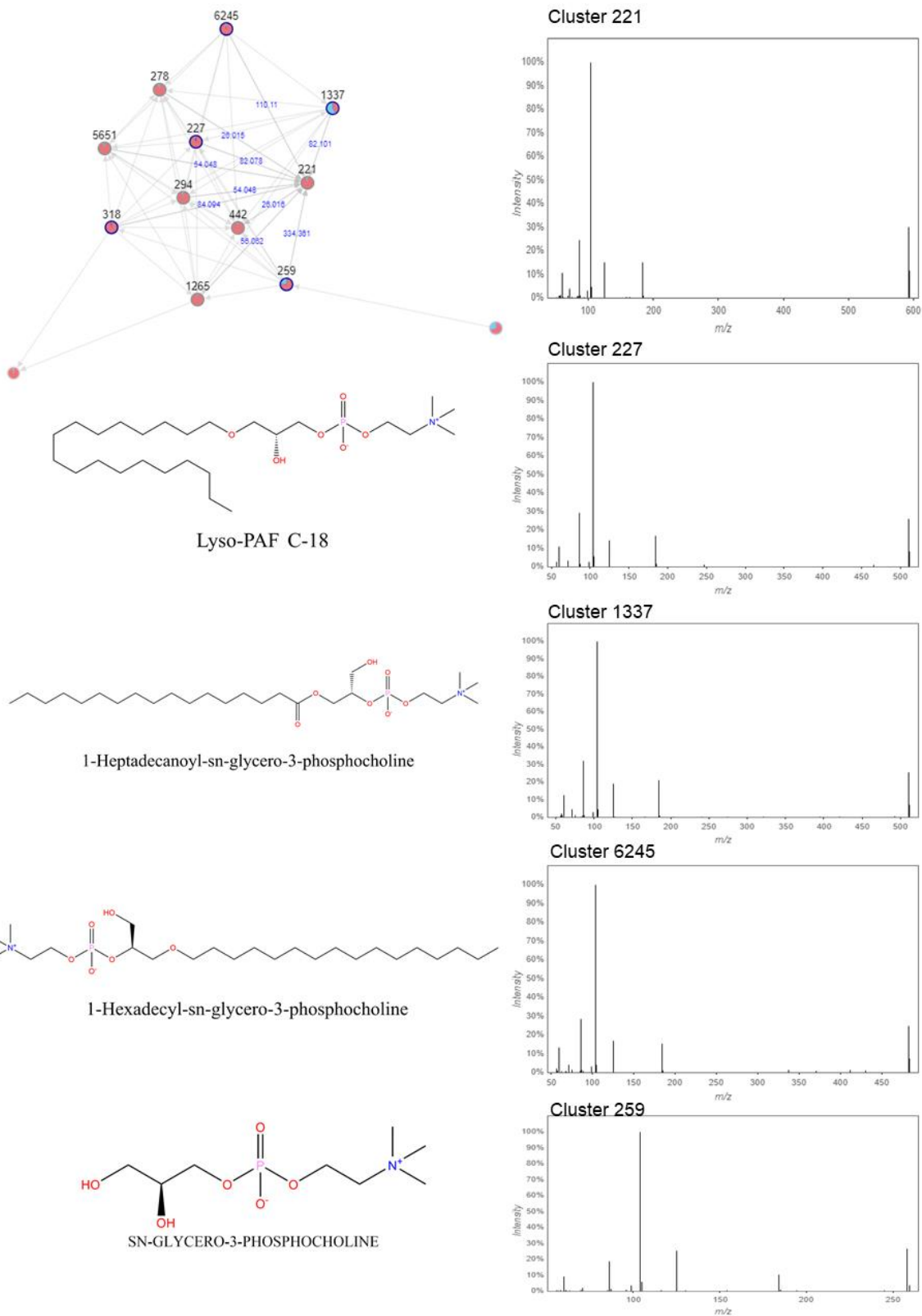
were also in the top-30 most important features identified in the Random Forest classification.

Cluster 806 ( $\log_2FC=-8.47$ ;  $p\text{-value}=3.14e-09$ ) was also case-related, and was annotated as 3-

hydroxy-2-(tetracosamido)octadecyl (2-(trimethylammonio)ethyl) phosphate

(Figure C.24). Another annotated cluster strongly affiliated with cases was Cluster 313 ( $\log_2FC=$

$-7.29$ ;  $p\text{-value}=3.14e-09$ ), which represented [2-hexadecanamido-3-hydroxyoctadec-4-en-1-



**Figure 4.5. Molecular network and MS<sup>2</sup> spectra for polar Cluster 221 and related clusters in case samples.**

**Figure 4.5 (cont'd)**

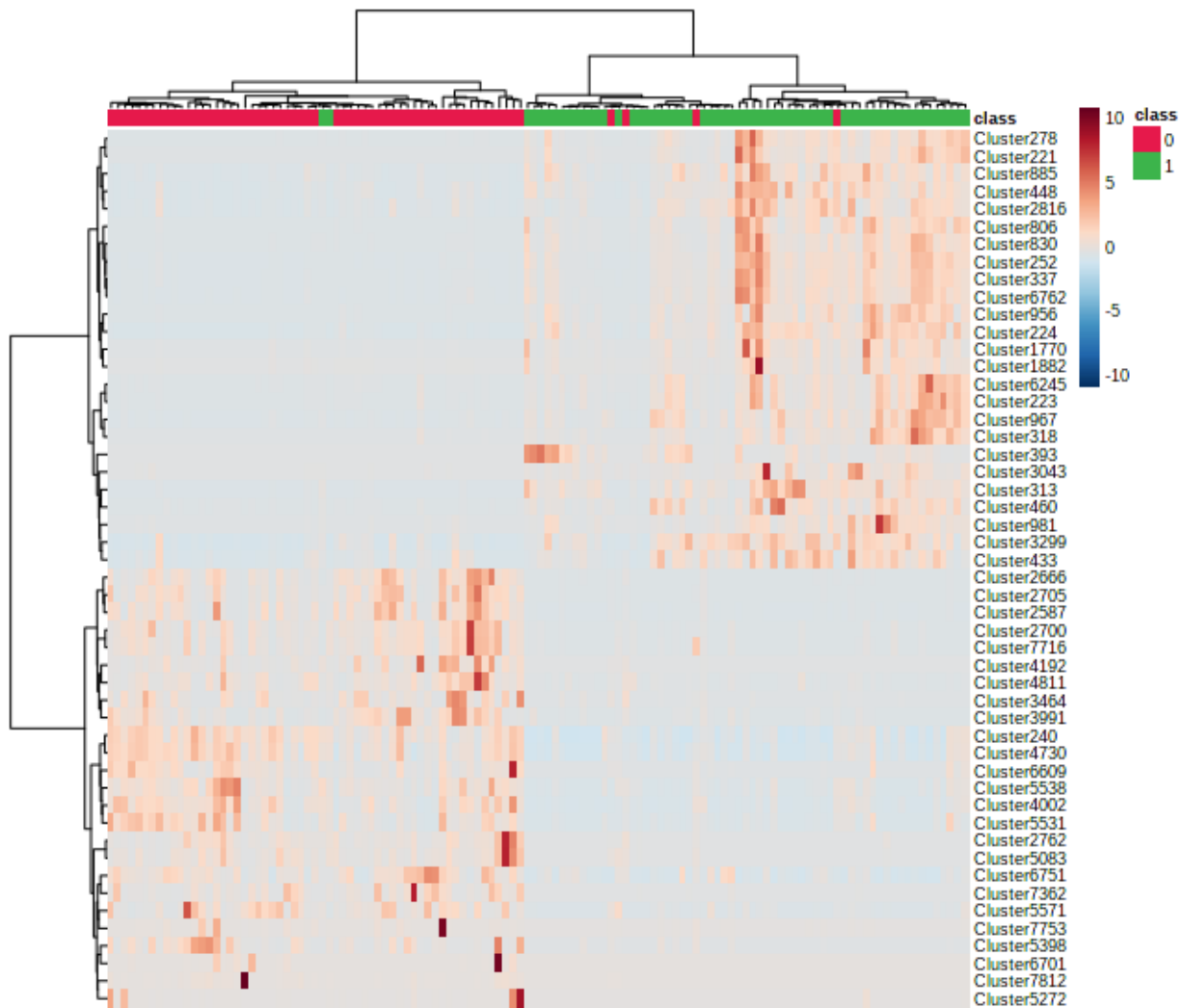
A molecular network constructed in GNPS (top, left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 221 and related clusters 227, 1337, 6245, and 259 are shown (right). Four of these clusters were successfully annotated as a series of phosphocholines; structures for these compounds were generated in ChemDraw 20.1 and are shown (left).

yl]oxy[2-(trimethylazaniumyl)ethoxy]phosphinic acid (Figure C.25). Cluster 313 was the strongest feature to distinguish between cases and follow-ups in the random forest model, highlighting the congruency between methods. However, no associations were observed between these clusters and relevant epidemiological data (such as hospitalization or bloody stool) using a Chi-square test (data not shown).

Following FC analysis, heatmaps were generated to view not only the distribution of peak intensities across samples, but also clustering based on the polar metabolite compositions. Among the top-50 metabolic features, distinct differences in polar metabolite composition were observed among cases and follow-ups (Figure 4.6). Of note, too, is the clustering and separation of follow-ups and cases based on their metabolic composition. Importantly, there is agreement between the clusters affiliated with cases or follow-ups across statistical measures. For example, many metabolites with a high fold-change relevant to follow-ups were in high abundance in the follow-up cluster on the heatmap, while those with negative log<sub>2</sub>FC (case-affiliated) were in higher abundance among the case cluster.

**Nonpolar metabolites are distinct between infected and recovered metabolomes**

A total of 13,940 nonpolar metabolites were identified among all samples after filtering and normalization. In contrast to the polar compounds, nonpolar metabolites had significantly greater diversity across all three metrics of alpha diversity (Figure C.26A; S<sub>case</sub>=1790,



**Figure 4.6. Heatmap displaying abundance of the top-50 polar metabolites based on significance determined by paired Wilcoxon tests in cases and follow-ups.**

A heatmap generated in MetaboAnalyst 5.0 indicates the abundance of the top-50 metabolites among cases and follow-ups. Intensity values were scaled by feature (metabolite). The color of each cell represents the abundance; the darker the red, the more abundant the metabolite. Each column represents one sample; the color bar at the top of the heatmap designates cases (1; green) and follow-ups (0; red). A dendrogram was generated using the Ward algorithm to display sample clustering based on Euclidean distance. Rows represent metabolic features or “clusters” which are named on the right y-axis. A dendrogram was generated for these clusters based on their distribution across samples and is shown on the left y-axis.

$S_{\text{follow}}=2832$ ,  $p=1.53e-11$ ;  $H'_{\text{case}}=4.89$ ,  $H'_{\text{follow}}=5.96$ ,  $p=1.48e-10$ ;  $J'_{\text{case}}=0.656$ ,  $J'_{\text{follow}}=0.750$ ,

$p=6.49e-09$ ) in the follow-up samples. No difference, however, was observed between cases

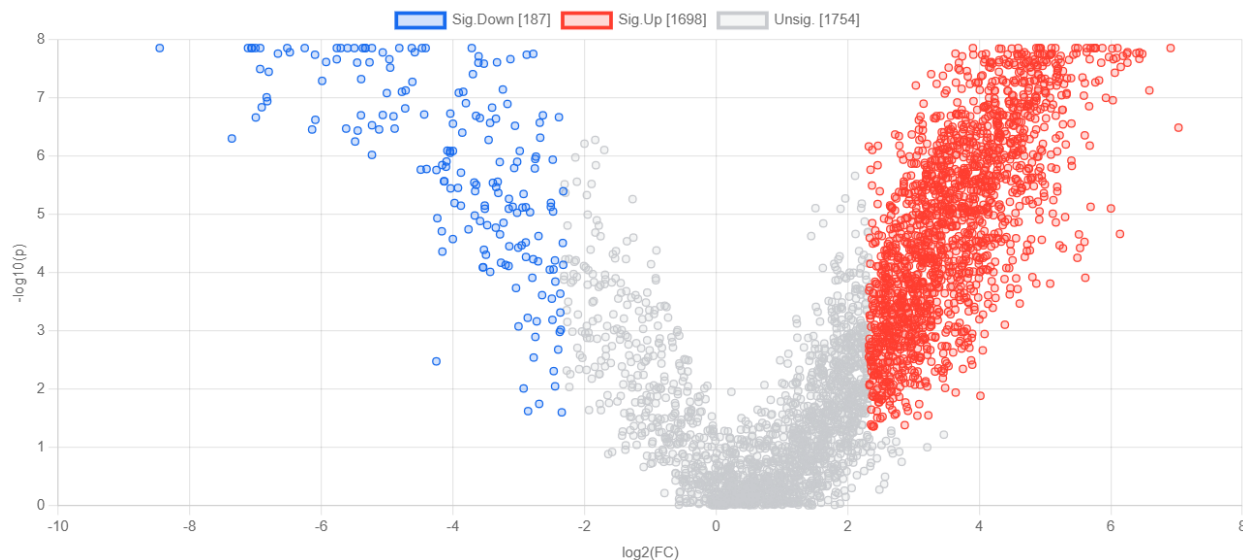
infected with *Campylobacter* relative to *Salmonella* (Figure C.27A;  $S_C=2021$ ,  $S_S=1627$ ,  $p=0.032$ ;

$H'_c=5.18$ ,  $H'_s=4.64$ ,  $p=0.0085$ ;  $J'_c=0.683$ ,  $J'_s=0.631$ ,  $p=0.024$ ). Similar to the polar metabolites, the nonpolar metabolite composition was distinct for cases and follow-ups based on Bray-Curtis dissimilarity (PERMANOVA F-value=19.607; p-value=0.000999; Figure C.26B), and PERMDISP indicated a significant difference in the dispersion of points (F-value=14.903; p-value=0.001). No clear clustering of nonpolar metabolite composition was observed after stratifying by pathogen (Figure C.27B; PERMANOVA F-value=1.2301, p-value=0.1189; PERMDISP F-value=3.263, p-value=0.019).

The random forest analysis on filtered, normalized intensities for nonpolar metabolites indicated an out-of-bag estimation of error rate of 4.92% with a confusion matrix similar to polar compounds (Tables C.5 and C.6). Based on the mean decrease in accuracy metric (Figure C.28), Cluster 2659 was most important in distinguishing between cases and follow-ups during classification (MDA=12.34) and was more abundant in cases (Figure C.29). The next most important compounds were clusters 321 (MDA=11.70) and 299 (MDA=11.58), both of which were more highly represented in cases (Figure C.30 and C.31). Because nonpolar metabolite diversity differed by pathogen, a random forest analysis was also used to explore accuracy of metabolite classification. Notably, the out-of-bag estimation of error rate was much higher for this model (41.8%), which may be partially explained by the difference in sample sizes among all 60 cases infected with the different pathogens. Nevertheless, various metabolites could distinguish among infectious agents (Figure C.32). For example, Cluster 2964 had the highest mean decrease in accuracy (6.05) and was elevated among cases infected with *Salmonella* (Figure C.33). Clusters 6581 and 8369 also had relatively high MDA scores for this model (6.02 and 5.69, respectively) and were each more abundant in cases infected with *Campylobacter* (Figure C.34).

Fold-change (FC) analysis in MetaboAnalyst v5.0 identified multiple nonpolar metabolites to be affiliated with cases or follow-ups (Figure 4.7; Table C.7). Of the nonpolar metabolites included in our FC analysis, 1,698 were increased in the recovered gut metabolomes (follow-ups) relative to the cases. Contrastingly, just 187 nonpolar metabolites demonstrated a negative  $\log_2\text{FC}$ , suggesting their presence solely during acute infection. The strongest association was for Cluster 321 (Figure C.35;  $\log_2\text{FC} = -8.46$ ;  $p\text{-value} = 1.38\text{e-}08$ ), which was affiliated with case samples, similar to our findings generated by random forest. Interestingly, Cluster 321 was a singleton without a molecular network; this singularity, coupled with its lack of annotation, suggests it may be an important, novel metabolite connected to infection. The next strongest signals were for Cluster 1618 ( $\log_2\text{FC} = -7.36$ ;  $p\text{-value} = 4.97\text{e-}07$ ) and Cluster 244 ( $\log_2\text{FC} = -7.11$ ;  $p\text{-value} = 1.38\text{e-}08$ ), each of which was also represented among cases. Cluster 1618 contributed to a small molecular network of four compounds, though Cluster 244 was also a singleton (Figure C.36 and C.37).

The clusters with the highest positive  $\log_2\text{FC}$  values, which were more abundant in follow-ups, included Clusters 2756 ( $\log_2\text{FC} = 7.03$ ;  $p\text{-value} = 3.23\text{e-}07$ ), 4470 ( $\log_2\text{FC} = 6.91$ ;  $p\text{-value} = 1.38\text{e-}09$ ), and 5193 ( $\log_2\text{FC} = 6.59$ ;  $p\text{-value} = 7.41\text{e-}08$ ). Cluster 2756 was a part of an extensive molecular network comprising ten different connections. Two of these connections, Clusters 2739 and 4512, were annotated as chenodeoxycholic acid, suggesting that Cluster 2756 may be involved in the metabolism of this primary bile acid (Figure 4.8). Clusters 4470 and 5193 each had less connectivity in their respective networks (Figure C.38); additionally, none of the compounds in these networks could be annotated, making inference of the roles of these metabolites difficult.

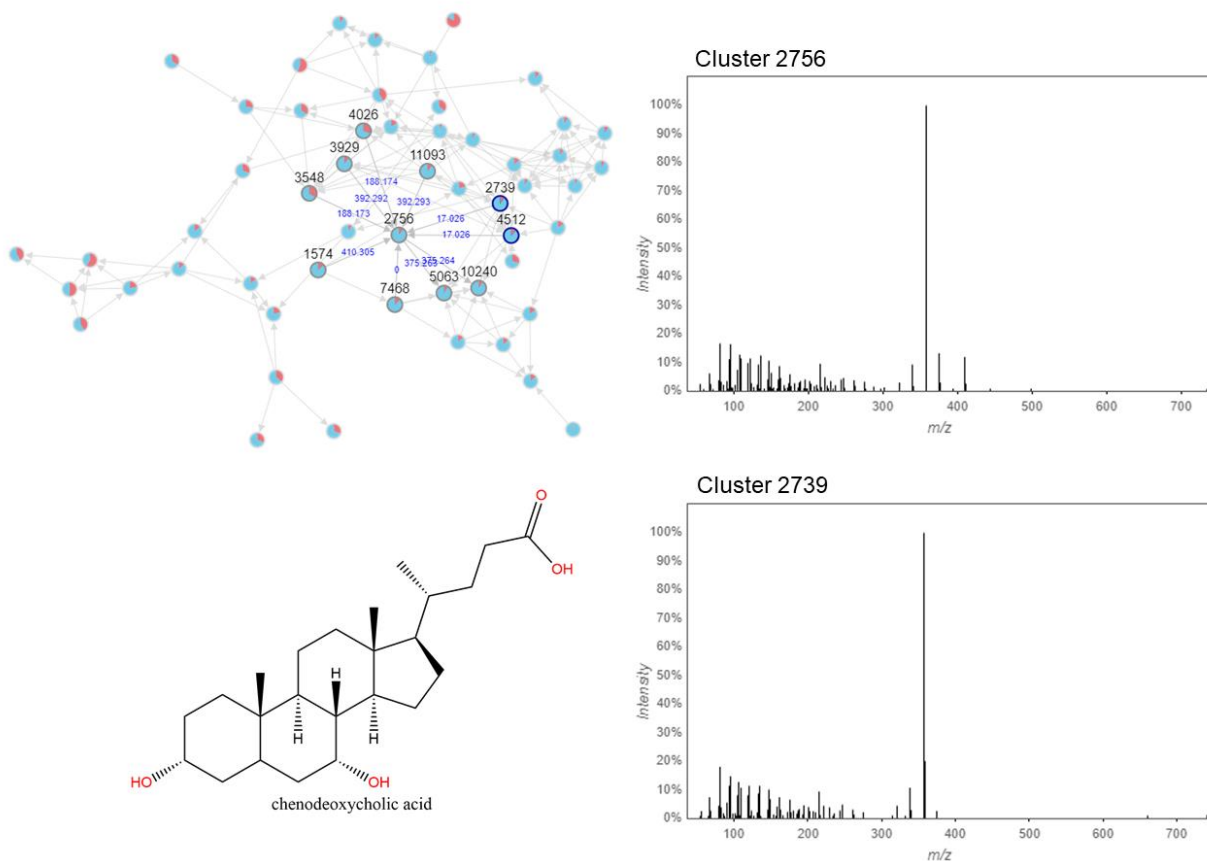


**Figure 4.7. Volcano plot demonstrating fold-change of nonpolar metabolites in cases and follow-ups.**

Fold-change (FC) analysis was performed to explore differentially abundant metabolites among samples. A FC cutoff of 5.0 and FDR threshold of 0.05 were set to identify the strongest signals. The volcano plot shows metabolites that were significantly more represented in follow-ups (“Sig.Up”, red, positive  $\log_2\text{FC}$ ) and those in cases (“Sig.Down”, blue; negative  $\log_2\text{FC}$ ). The x-axis indicates the  $\log_2\text{FC}$  value; the y-axis shows the  $-\log_{10}(\text{P})$  value. Metabolites that lacked significance with these parameter cutoffs are shown as gray dots (“Unsig.”). The legend at the top of the plot indicates the number of metabolites in each category.

Molecular networks for the metabolites suggested to be associated with specific enteric infections were also explored. Cluster 2964, which was affiliated with *Salmonella* infections, is part of a small molecular network containing similarly related compounds (Figure C.39). Comparatively, Clusters 6581 and 8369 were elevated in *Campylobacter* cases and are part of an extensive molecular network (Figure C.40). Interestingly, a subnetwork that is more distantly related to these compounds was present almost exclusively in *Shigella* patients, though only 4 cases were included in this analysis.

Construction of a heatmap highlighted the most striking differences in peak intensity among nonpolar metabolic features in cases and follow-ups (Figure 4.9). Notably, the clustering of samples based on nonpolar metabolite composition nearly perfectly separates cases and



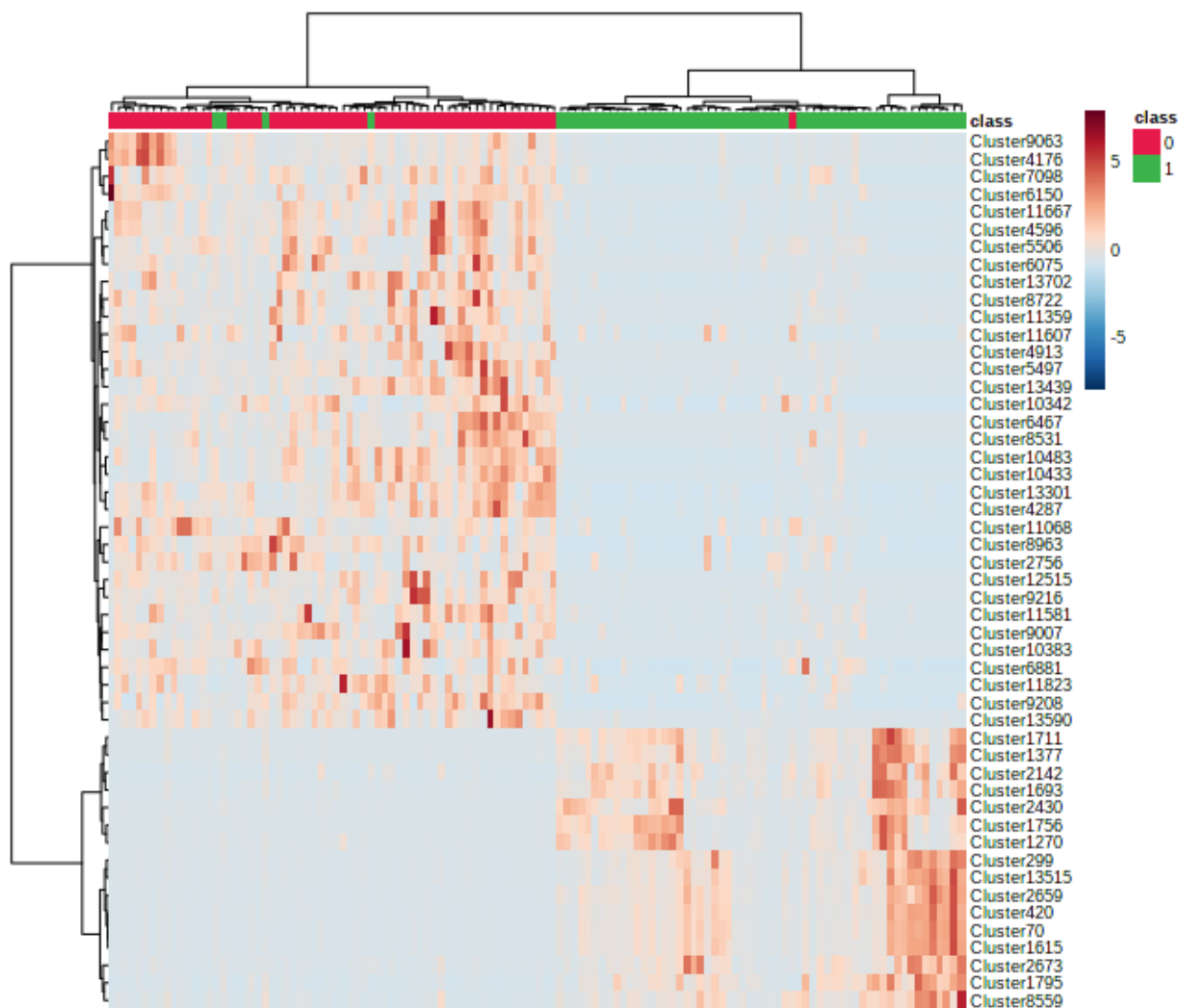
**Figure 4.8. Molecular network and MS<sup>2</sup> spectra for Cluster 2756 and related cluster (2739), which were greatly increased in follow-ups.**

A molecular network constructed in GNPS (top, left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 2756 and a closely related cluster, 2739, are shown (right). Clusters 2739 and 4512 (spectra not shown) were successfully annotated in GNPS as chenodeoxycholic acid; the structure for this compound was generated in ChemDraw 20.1 and is also shown (bottom, left).

follow-ups with minimal overlap. The distribution of metabolic features, too, indicates a stark difference in composition among these sample groups. Among the metabolites included in the heatmap, there was only moderate agreement with the important metabolites detected via the random forest and FC analyses. These findings suggest that it may be difficult to attribute ubiquitous importance to a handful of metabolites related to recovery and intestinal health. For



those metabolites associated with cases, however, there was more agreement across analytical methods.



**Figure 4.9. Heatmap displaying abundance of the top-50 nonpolar metabolites based on significance determined by paired Wilcoxon tests in cases and follow-ups.**

A heatmap generated in MetaboAnalyst 5.0 indicates the abundance of the top-50 metabolites among cases and follow-ups. Intensity values were scaled by feature (metabolite). The color of each cell represents the abundance; the darker the red, the more abundant the metabolite. Each column represents one sample; the color bar at the top of the heatmap designates cases (1; green) and follow-ups (0; red). A dendrogram was generated using the Ward algorithm to display sample clustering based on Euclidean distance. Rows represent metabolic features or “clusters” which are named on the right y-axis. A dendrogram, which was generated for these clusters based on their distribution across samples, is shown on the left y-axis.

## DISCUSSION

Metabolic health of the human gut is undoubtedly linked to the microbiome. Additionally, environmental flux of the human gut related to disease state, diet, antibiotic use, or exercise can also greatly influence the composition of microbially-mediated metabolic pathways (44). It is known that gut communities demonstrating greater taxonomic diversity, which typically represents a healthy, homeostatic gut environment (45), would have greater metabolic functionality than a community with fewer members. In our analysis, functional prediction of microbial metabolic pathways using metagenomics data examined using two methods showed that patients with acute enteric infections had greater metabolic capacity during infection than post-recovery. This finding differs from our hypothesis that the overall metabolic capacity, or number of pathways, would be similar or lower during an infection. Similarly, Dash et al. (2021) (46) showed that individuals with Type 2 diabetes mellitus displayed significantly higher metabolic pathway richness than healthy people, though the connection between disease state, microbiota composition, and pathway abundance was not clear. Our corresponding LC/MS analysis of actual metabolite profiles on the same set of samples, however, showed increased metabolic diversity among the recovered follow-up samples, which is opposite the trend shown in our pathway prediction data. In addition to these discrepancies, the overlap among predicted microbial metabolic pathways and the identified metabolites was relatively scant, which is somewhat expected. While functional prediction of microbial metagenomes allows us to visualize metabolic capacities among gut microbes, untargeted metabolomics captures the entire metabolic chemistry of the gut environment, microbially-related or not. Because untargeted metabolomics picks up human-, drug- and food-derived compounds along with microbial-derived molecules, this method will inherently provide differing results than our metagenome analysis, which is solely based on microbial composition. Yet, the comparison of functional

prediction with known metabolite signatures enables us to further characterize the importance (or lack thereof) of these microbial functions in the gut metabolome. In addition to highlighting the importance of comparing multiple characterization methods, these observed differences also indicate that enhanced diversity of host-derived metabolites is important for human health.

It is likely that the change in microbial composition during infection influences the abundance of metabolic pathway genes detected as was demonstrated through our predictive search. For example, we previously showed that individuals with enteric infection exhibit a marked increase in ARGs harbored by members of *Enterobacteriaceae* such as *Escherichia* and *Klebsiella* (Chapter 3). Moreover, this increase was correlated with enhanced abundance of these genera during infection regardless of the pathogen (25, 26). The overgrowth of *Enterobacteriaceae* and *E. coli*, in particular, has previously been documented as a result of the host-mediated inflammatory response (47). In fact, Winter et al. (2013) (48) demonstrated that nitrate, which is host-generated during inflammation, confers a growth advantage to members of *Enterobacteriaceae*, which are capable of degrading non-fermentable substrates unlike many commensal anaerobes. Indeed, many pathways elevated among cases include signatures of nitrogen metabolism related to nitrate reduction (Figure C.41), amino acid regulation, and amino acid biosynthesis. For instance, multiple arginine and ornithine pathways were detected (Figure C.42 and C.43); these compounds are precursors for nitric oxide (NO) and polyamines (such as putrescine), respectively (49). Both nitric oxide and ornithine have been implicated for their role in altering gut microbiota composition as well as gut metabolism in a prior study (50). Specifically, the presence of NO was shown to favor the overgrowth of *Enterobacteriaceae* and modified amino acid composition and concentration; NO also led to decreased abundance of beneficial SCFA-producing bacteria such as *F. prausnitzii* (50). In concordance with these

findings, the overrepresentation of metabolic pathways related to enhanced *Enterobacteriaceae* growth in case samples highlights the importance of these pathways during infection.

Some other notable pathways among infected cases included multiple menaquinol synthesis pathways. Menaquinols are the reduced form of menaquinones, which play crucial roles in the bacterial cell membrane to facilitate electron transfer and oxidative phosphorylation (51). Menaquinones are synonymous with vitamin K<sub>2</sub>, an essential nutrient in humans. Despite its importance in human health, increased levels of vitamin K have been implicated in various disease states such as Type 2 diabetes (46). Other studies have explored the association of elevated menaquinone synthesis to intestinal inflammation, particularly related to obesity (52, 53); however, these studies failed to find a connection between increased menaquinol concentrations and inflammation. It has been shown that protein families involved in vitamin K synthesis are noticeably reduced in patients with IBD (54). This lack of consensus regarding the myriad roles of menaquinone in the gut environment make interpretation of our results difficult. However, from a microbial perspective rather than through the lens of human health, the increase in menaquinol synthesis during enteric infection seems plausible. As mentioned prior, menaquinones enable respiration via electron transfer. It has also been found that these molecules can serve as antioxidants and may protect bacterial cell membranes from harmful oxidation (55). Given that enteric infection typically results in inflammation that can lead to increased luminal oxygen (56), the enhanced prevalence of pathways producing compounds to assist in survival of such conditions is comprehensible.

Another interesting finding among cases was the increased prevalence of the glycolysis, pyruvate dehydrogenase, tricarboxylic acid cycle, and glyoxylate bypass superpathways, which were detected using both predictive methods. Accordingly, Perez-Cobas et al. (2013) (57) found

that proteins for these metabolic pathways were enhanced in individuals who had recently received antibiotics; they hypothesized that this enhancement was due to a community response to fluctuating nutrient supply and stress, resulting in an overcompensation in carbohydrate metabolism. In particular, the increased prevalence of glyoxylate bypass is of great relevance, as this pathway enables microbes to use a variety of substrates for central carbon metabolism including fatty acids, alcohols, esters, alkenes, and other compounds (51). Given that antibiotic treatment is a known factor causing disturbance in the gut community, it is plausible that disruption caused by an infectious pathogen would result in a similar effect. Indeed, our metabolic pathway prediction harbored many results in line with previous findings relevant to a disrupted gut environment.

Compared to the metabolic pathway prediction analyses, our untargeted metabolomics on polar and nonpolar metabolites uncovered distinct metabolic profiles among cases and follow-ups. One prevalent group of compounds identified was a series of glycerophosphocholines, which were assigned to at least six annotated molecules (e.g., Clusters 806, 318, 227, 259, 1337, and 6245). Although these compounds were detected in all samples regardless of health status, 1-(1Z-Hexadecenyl)-sn-glycero-3-phosphocholine was found primarily in follow-ups. Others including 1-(1Z-Octadecenyl)-sn-glycero-3-phosphocholine, Lyso-PAF C-18, 1-Heptadecanoyl-sn-glycero-3-phosphocholine, 1-Hexadecyl-sn-glycero-3-phosphocholine, and sn-glycero-3-phosphocholine, were more highly represented among cases. In support of these findings, our metagenomics pathway prediction pipeline also identified a phosphatidyl choline acyl editing pathway, PWY-6803, to be more abundant in cases along with an overall enhanced capacity for lipid and fatty acid metabolism (Figure C.44). Indeed, glycerophosphocholines are required in the synthesis of phosphatidylcholine, an abundant phospholipid that plays an important role in

lipid metabolism throughout the body (58); hence, enhanced abundance of these factors during infection is plausible. While choline is an essential nutrient that assists with healthy brain function, cell signaling, lipid movement, and metabolism (59), it can also be metabolized by anaerobic bacteria in the gut, resulting in the generation of trimethylamine (TMA) (60). TMA can subsequently be metabolized by the host to form trimethylamine N-oxide (TMAO), a compound linked to various human pathologies such as cardiovascular disease (61). Although TMA or TMAO were not detected, two trimethyl-ammonium-related products were identified in the polar metabolite analysis. One of these compounds, 3-hydroxy-2-(tetracosyl 11.13.15)octadecyl (2-(trimethylammonio)ethyl) phosphate, represented by Cluster 806, was elevated in cases. The other compound, [2-hexadecanamido-3-hydroxyoctadec-4-en-1-yl]oxy[2-(trimethylazaniumyl)ethoxy]phosphinic acid, assigned as Cluster 313, was important for differentiating cases from follow-ups, suggesting its significance in the infected gut. Since Cluster 313 was found in virtually all (n=58; 95.1%) samples, no associations were observed for markers of disease severity (e.g., presence of bloody stool or hospitalization) as well as demographic variation (e.g., age, sex, residence type). Although further characterization of these compounds is needed to define a potential role in TMA(O) metabolism, their presence in most cases and in only 19 of the follow-ups highlights an association with acute infection that requires investigation in the future.

Other notable findings in sample metabolomes include Clusters 326, 7558, and 5193, which contributed to an overlapping molecular network that included a distant cluster with the annotation for tomatidine. Each of these clusters, including the annotated cluster, were found predominantly in follow-ups and may be indicators for a healthy gut. Specifically, tomatidine was found in a majority of recovered patients (66%) with elevated average relative intensity

compared to cases (0.048% vs. 0.0017%, respectively). Tomatidine is a glycoalkaloid compound produced by members of the Solanaceae family which includes tomatoes. Tomatidine is notorious for its benefits to human health and has widely been studied for its role in inhibiting atrophy of skeletal muscle (62, 63). However, this compound has also been implicated for its antimicrobial effects, particularly against *Staphylococcus aureus* (64-66). Moreover, Guthrie et al. (2019) (67) determined that tomatidine is structurally similar to taurochenodeoxycholic acid (TCDCA), a conjugated bile acid with antimicrobial activity in the gut, and hence, they hypothesized that the antimicrobial nature of tomatidine may include acidification of bacterial cells. Although the impact of tomatidine in the gut is not known, its presence in a majority of recovered patients is intriguing and may indicate that this compound is a facet of a healthy, homeostatic gut environment, a hypothesis that requires further investigation.

The identification of Cluster 2113, which was related to desmethylenylnocardamine, was also found primarily among follow-up samples. This compound is a cyclic peptide that was originally isolated from marine species of *Streptomyces* and found to demonstrate slight inhibition of sortase B, an enzyme responsible for modifying cell surface proteins (68). Another study by Shaaban et al. (2014) (69) isolated desmethylenylnocardamine from a different *Streptomyces* strain while searching for large antifungal macrolide compounds known as venturicidins. Isolation of this compound in conjunction with known macrolides potentially implicates its use as an antimicrobial secondary metabolite in the gut. Although *Streptomyces* is primarily recognized as a soil bacterium, the presence of this microbe in the human gut is not unprecedented; in fact, *Streptomyces* have been shown to benefit the human host through production of immune-related regulatory metabolites (70). Therefore, our finding of

desmethylenylcardamine among follow-ups suggests that production of various antimicrobial secondary metabolites could also be a facet of a recovered, healthy gut.

The prevalence of Cluster 2756 among follow-ups is also intriguing. Although this cluster could not be annotated, our analysis showed connections to Clusters 2739 and 4512, which represent chenodeoxycholic acid (CDCA), a naturally occurring bile acid (BA) produced in the liver that assists with cholesterol breakdown (71). Microbes in the gut are known to facilitate important biotransformations of bile acids, among other molecules (72). Specifically, members of *Eubacterium* and *Clostridium* can perform  $7\alpha$ -dehydroxylation, the process that converts CDCA to the secondary bile acid lithocholic acid (73). Duboc et al. (2012) (22) found that individuals experiencing diarrhea-predominant irritable bowel syndrome (IBS) had significantly more primary BAs than their healthy counterparts, emphasizing the importance of microbial conversion of primary to secondary BAs. In our analysis, Clusters 2756, 2739, and 4512 were all elevated among recovered patients, a finding that is somewhat contradictory to the results of Duboc et al. (2012); indeed, Cluster 2756 was present in most follow-ups (n=58; 95.1%) with a much higher average relative intensity than cases (3.0% vs. 0.62%). Additionally, the number of days between infection and follow-up was investigated to determine if timing was associated with the observed intensity of this compound. A relative intensity threshold of 0.05 was established; ten follow-ups contained Cluster 2756 with intensities at or above this value. Of these 11 individuals, the average number of follow-up days was 93.5, whereas the average for the remaining 50 follow-ups (relative intensity < 0.05) was 110 days. Although this association was not explored statistically, these trends provide interesting connections worth greater investigation. Furthermore, lithocholic acid, the secondary BA produced from CDCA, was also more commonly detected among follow-ups, albeit modestly (data not shown). Therefore, it is



possible that the entirety of this biotransformation pathway is more prevalent among follow-ups relative to cases may be related to time since infection. Future work is required to determine the relationship between these compounds and recovery from enteric infection.

It is important to note that this study along with other studies that utilize untargeted metabolomics via LC/MS are limited in that many polar and nonpolar compounds identified are unknown and have yet to be characterized (74). While this lack of annotation limits our ability to make biologically sound conclusions, particularly about infected vs. recovered gut states, observing the compositional differences among infected and recovered metabolomes still holds meaning. For example, each compound isolated through this study registered unique MS<sup>2</sup> spectra, and hence, they may be characterized in the future. Even without structural characterization, these unknown compounds' relationship to known metabolites via molecular networking analysis enables us to infer their contribution to the human gut metabolome, possibly serving as precursors or intermediates in known pathways. And, more abstractly, comparing metabolite compositions from a birds-eye view further enhances our understanding of how enteric infection can influence the gut microbial community; though we may not know specifically which metabolites are changing in abundance, we can confidently assert that infection does play an important role in dictating the gut's metabolic capacity.

While this study focuses primarily on exploring potential functions of the gut microbiota and characterizing the overall gut metabolome, further investigation of these data is encouraged. A plausible future direction for this work is to integrate microbiome data such as taxonomic classification or gene annotation with the metabolomics data described. Previous work has demonstrated that integration of these two 'omics techniques (metagenomics and metabolomics) can provide a much more comprehensive understanding of the human gut environment. For

example, The Human Microbiome Consortium (2012) (75) demonstrated that individuals with differing microbiome compositions shared a majority of metabolic pathways identified.

Similarly, Visconti et al. (2019) (76) captured metabolic similarities among people while also characterizing associations between various microbial taxa, predicted pathways, and fecal metabolite frequency. Further clarifying these links between microbial composition, metabolic pathway prediction, and metabolite abundance will be an important addition to advance our understanding of changes in the gut environment related to enteric infection.

Indeed, gut microbial communities are known to undergo notable change after experiencing a disturbance such as antibiotic treatment, modified diet, or enteric infection. Previously, we demonstrated that enteric infection is associated with severe changes in both taxonomic and resistance gene composition. In this study, we have shown that stools of patients during enteric infection not only display different functional potential via pathway prediction, but also contain markedly different metabolites than stools collected from the same patients upon recovery. While cases registered higher diversity of metabolic pathways, recovered communities appeared to have greater overall diversity of metabolites. Our use of functional prediction via metagenome analyses coupled with untargeted LC/MS metabolomics strengthens our ability to comprehensively define the impacts of enteric infection on microbiota within the human gut. While interpretation of metabolites identified via untargeted metabolomics is currently difficult due to limited compound annotation, observing patterns of diversity as well as abundance and intensity among infected and recovering individuals is quite meaningful. Indeed, future work with these data should consider integrating other ‘omics techniques, as connecting specific microbial features to metabolite composition will further our comprehension of this complex interplay between pathogens, resident gut microbiota, and the human host.

## **APPENDIX**

**Table C.1. Differentially abundant metabolic pathways in cases and follow-ups predicted by HUMAnN 3.0.**

Feature	Class	Coefficient	STD_error	p_value	q_value
DTDPRHAMSYN-PWY: dTDP-&beta;-L-rhamnose biosynthesis	FollowUp	0.041156056	0.006211572	3.46E-11	2.12E-10
PWY-5686: UMP biosynthesis I	FollowUp	0.035655917	0.006356094	2.03E-08	8.06E-08
PWY-7219: adenosine ribonucleotides de novo biosynthesis	FollowUp	0.034887861	0.003781554	2.81E-20	7.24E-19
PWY-5030: L-histidine degradation III	FollowUp	0.033188933	0.005492083	1.51E-09	7.48E-09
NONMEVIPP-PWY: methylerythritol phosphate pathway I	FollowUp	0.032789846	0.004873582	1.72E-11	1.09E-10
COA-PWY-1: superpathway of coenzyme A biosynthesis III (mammals)	FollowUp	0.032567147	0.002940853	1.68E-28	2.16E-26
HISTSYN-PWY: L-histidine biosynthesis	FollowUp	0.031755177	0.003809123	7.64E-17	1.13E-15
COA-PWY: coenzyme A biosynthesis I (prokaryotic)	FollowUp	0.031638381	0.004681204	1.39E-11	9.12E-11
PWY-4242	FollowUp	0.031355838	0.003948359	2.00E-15	2.08E-14
PWY-7221: guanosine ribonucleotides de novo biosynthesis	FollowUp	0.031225219	0.004598069	1.11E-11	7.41E-11
PWY0-781: aspartate superpathway	Case	0.043439605	0.006944825	3.98E-10	2.10E-09
PWY-5675: nitrate reduction V (assimilatory)	Case	0.044108903	0.008529998	2.33E-07	7.62E-07
TCA-GLYOX-BYPASS: superpathway of glyoxylate bypass and TCA	Case	0.044534629	0.007730492	8.37E-09	3.51E-08
PWY-6285: superpathway of fatty acids biosynthesis (E. coli)	Case	0.045517033	0.01202773	0.000154116	0.000345866
PWY-5860: superpathway of demethylmenaquinol-6 biosynthesis I	Case	0.045902337	0.005512658	8.31E-17	1.19E-15
GLYCOLYSIS-TCA-GLYOX-BYPASS: superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass	Case	0.047280305	0.008658842	4.75E-08	1.76E-07
PWY-5840: superpathway of menaquinol-7 biosynthesis	Case	-0.04744992	0.007936726	2.25E-09	1.06E-08

**Table C.1. (cont'd)**

PWY-5971: palmitate biosynthesis (type II fatty acid synthase)	Case	-0.04757056	0.016197457	0.003314946	0.006151775
PWY-5850: superpathway of menaquinol-6 biosynthesis	Case	-0.05010986	0.006165949	4.41E-16	5.00E-15
PWY-5896: superpathway of menaquinol-10 biosynthesis	Case	-0.05010986	0.006165949	4.41E-16	5.00E-15

**Table C.2. Top-30 polar clusters most important to health status classification via random forest.**

Cluster	Library ID	RT	MZ	Mean Decrease Accuracy
Cluster313	{[2-hexadecanamido-3-hydroxyoctadec-4-en-1-yl]oxy}[2-(trimethylazaniumyl)ethoxy]phosphinic acid	3.9226	725.5561	13.86255853
Cluster2705	N/A	4.4761	480.3523	13.26895851
Cluster6376	N/A	4.638	321.1442	12.89567889
Cluster830	N/A	3.9304	701.5587	12.86589776
Cluster7812	N/A	1.6002	289.1178	12.76648721
Cluster221	N/A	4.056	592.4691	12.24319532
Cluster2762	N/A	3.7288	299.1246	11.87208286
Cluster2587	N/A	4.4918	480.3526	11.28914703
Cluster2666	Spectral Match to 1-(1Z-Hexadecenyl)-sn-glycero-3-phosphocholine from NIST14	4.4685	480.3522	11.09808904
Cluster6701	N/A	4.1263	336.1561	11.08651731
Cluster5083	N/A	3.7115	299.1246	11.05404386
Cluster970	N/A	4.0823	564.4384	10.72343845
Cluster6130	N/A	4.5821	313.186	10.71071175
Cluster5571	N/A	7.7265	626.6993	10.61969179
Cluster7788	Spectral Match to N-Tetracosenoyl-4-sphingenyl-1-O-phosphorylcholine from NIST14	3.7916	813.6829	10.03043212
Cluster3575	N/A	1.8936	972.7333	9.172266376
Cluster7753	N/A	1.1968	354.1845	9.041261073
Cluster2039	N/A	1.7524	720.5929	8.924073237
Cluster224	Spectral Match to N-Tetracosenoyl-4-sphingenyl-1-O-phosphorylcholine from NIST14	3.781	813.6829	8.911454958
Cluster252	Sphingomyelin (18:1/14:0)	3.9588	675.5429	8.766568483
Cluster4192	N/A	5.1719	335.178	8.762194276
Cluster6762	N/A	3.9427	689.5583	8.738426824
Cluster442	N/A	4.0508	618.4853	8.572093642
Cluster393	N/A	7.6641	310.8303	8.527643158
Cluster337	N/A	3.9406	689.5582	8.463004595

**Table C.2 (cont'd)**

Cluster956	3-hydroxy-2-(tetracos-11.13.15-trienamido)octadecyl (2-(trimethylammonio)ethyl) phosphate	3.7938	811.6677	8.416119788
Cluster4775	N/A	1.1349	294.0923	8.375957668
Cluster5399	N/A	7.7335	264.8498	8.186101641
Cluster2700	N/A	4.8821	286.1401	8.04330702
Cluster1988	N/A	4.8702	296.0985	8.028849738

**Table C.3. Confusion matrix for classification of samples by health status generated by random forest on polar metabolites.**

OOB estimate of error rate:

5.74%

	CASE	FOLLOW	Class Error
CASE	56	5	0.08196721
FOLLOW	2	59	0.03278689

**Table C.4. Output generated by fold-change analysis exploring differentially abundant polar metabolites among cases and follow-ups.**

Fifty clusters displaying the most positive (n=25) and most negative (n=25) FC are displayed.

Feature	FC	log2(FC)	q-value	Log(10)P
Cluster326	494.41	8.9496	2.42E-07	6.6165
Cluster3942	321.24	8.3275	5.77E-08	7.2391
Cluster7558	320.2	8.3228	6.33E-07	6.1989
Cluster2113	211.01	7.7211	2.27E-08	7.6435
Cluster2762	186.29	7.5414	3.14E-09	8.5037
Cluster5083	164.61	7.3629	5.17E-09	8.2865
Cluster6871	145.09	7.1808	2.11E-08	7.6765
Cluster7362	129.9	7.0213	5.62E-09	8.2506
Cluster2700	128.7	7.0079	3.14E-09	8.5037
Cluster1593	121.81	6.9285	3.76E-07	6.4248
Cluster7597	119.53	6.9012	4.35E-08	7.3612
Cluster2705	96.772	6.5965	3.14E-09	8.5037
Cluster2587	94.185	6.5574	3.14E-09	8.5037
Cluster2666	91.314	6.5128	3.14E-09	8.5037
Cluster1988	90.051	6.4927	8.29E-08	7.0813
Cluster4096	82.075	6.3589	1.63E-08	7.7877
Cluster7716	81.402	6.347	3.95E-09	8.4029
Cluster4682	78.59	6.2963	1.72E-08	7.7649
Cluster6067	75.364	6.2358	2.11E-08	7.6765
Cluster7753	72.902	6.1879	5.11E-09	8.2918
Cluster6879	71.764	6.1652	2.50E-08	7.6013
Cluster6679	70.057	6.1304	2.65E-08	7.5768
Cluster6208	69.231	6.1133	8.29E-08	7.0814
Cluster4872	67.704	6.0812	2.26E-08	7.645
Cluster6696	66.263	6.0501	2.77E-08	7.5572
Cluster5651	0.013454	-6.2158	7.19E-09	8.1432
Cluster548	0.013009	-6.2644	0.00032363	3.49
Cluster662	0.012686	-6.3006	2.48E-08	7.605
Cluster3566	0.011898	-6.3932	2.32E-07	6.6339
Cluster1265	0.01167	-6.421	1.45E-08	7.8386
Cluster442	0.011572	-6.4332	2.94E-08	7.5313
Cluster7788	0.011335	-6.4631	5.98E-09	8.223
Cluster224	0.01124	-6.4752	5.17E-09	8.2865
Cluster336	0.010857	-6.5252	6.66E-09	8.1764
Cluster278	0.010694	-6.547	5.17E-09	8.2865
Cluster4329	0.010688	-6.5479	7.19E-09	8.1432
Cluster294	0.010484	-6.5756	7.19E-09	8.1432
Cluster307	0.009564	-6.7082	2.39E-08	7.6218
Cluster393	0.008196	-6.9309	5.47E-09	8.2623
Cluster3575	0.007853	-6.9926	7.19E-09	8.1432



**Table C.4 (cont'd)**

Cluster2039	0.007502	-7.0586	4.13E-08	7.3845
Cluster1882	0.006402	-7.2873	4.15E-09	8.3823
Cluster313	0.006393	-7.2892	3.14E-09	8.5037
Cluster830	0.006279	-7.3153	3.14E-09	8.5037
Cluster1770	0.00586	-7.4149	3.14E-09	8.5037
Cluster956	0.0049	-7.673	5.17E-09	8.2865
Cluster500	0.004843	-7.6898	7.19E-09	8.1432
Cluster806	0.002829	-8.4656	3.14E-09	8.5037
Cluster221	0.002491	-8.6493	5.56E-09	8.2551
Cluster970	0.002375	-8.7177	6.22E-09	8.206

**Table C.5. Top-30 nonpolar clusters most important to health status classification via random forest.**

Cluster	LibraryID	RT	MZ	MeanDecreaseAccuracy
Cluster2659	N/A	4.5341	400.1219	12.34109358
Cluster321	N/A	1.1114	398.0681	11.69941273
Cluster299	N/A	2.8292	419.8553	11.58190455
Cluster244	N/A	5.2053	273.0909	11.26608259
Cluster70	N/A	4.5345	356.0663	10.9058924
Cluster11528	N/A	3.6763	327.0995	10.75545454
Cluster2634	N/A	6.2441	355.063	10.67215235
Cluster1615	384	4.5333	355.0632	10.6557956
Cluster420	N/A	4.5342	471.9906	10.33994613
Cluster13515	N/A	2.8299	419.8553	9.504268644
Cluster4470	N/A	4.3358	353.2064	8.74183389
Cluster4988	384	4.7124	355.0633	8.661024763
Cluster4596	N/A	5.5299	450.3426	8.563009156
Cluster10112	N/A	6.2456	372.0897	8.357792042
Cluster1795	N/A	2.8266	322.9212	8.351389966
Cluster2430	N/A	7.5465	598.2675	7.876945288
Cluster1244	N/A	9.1037	520.3754	7.84221535
Cluster6467	N/A	5.0792	417.2017	7.77855293
Cluster4492	N/A	3.7583	327.13	7.720679607
Cluster8559	N/A	2.6672	507.9757	7.71737409
Cluster10787	N/A	3.3183	341.1744	7.679961302
Cluster12515	N/A	2.8462	285.1079	7.610383406
Cluster7782	N/A	4.5287	321.2899	7.594279345
Cluster3630	N/A	4.661	337.1754	7.574973534
Cluster1711	N/A	8.2844	626.2965	7.498791559
Cluster2683	N/A	3.7665	429.9303	7.403677586
Cluster2673	N/A	3.7657	215.9474	7.193631343
Cluster11359	N/A	7.6155	542.403	7.088795545
Cluster4287	N/A	5.1179	470.2282	7.064229725
Cluster9063	N/A	4.662	337.1755	7.035665411

**Table C.6. Confusion matrix for classification of samples by health status generated by random forest on nonpolar metabolites.**

OOB estimate of error rate:

4.92%

	CASE	FOLLOW	Class Error
CASE	56	5	0.08196721
FOLLOW	1	60	0.01639344

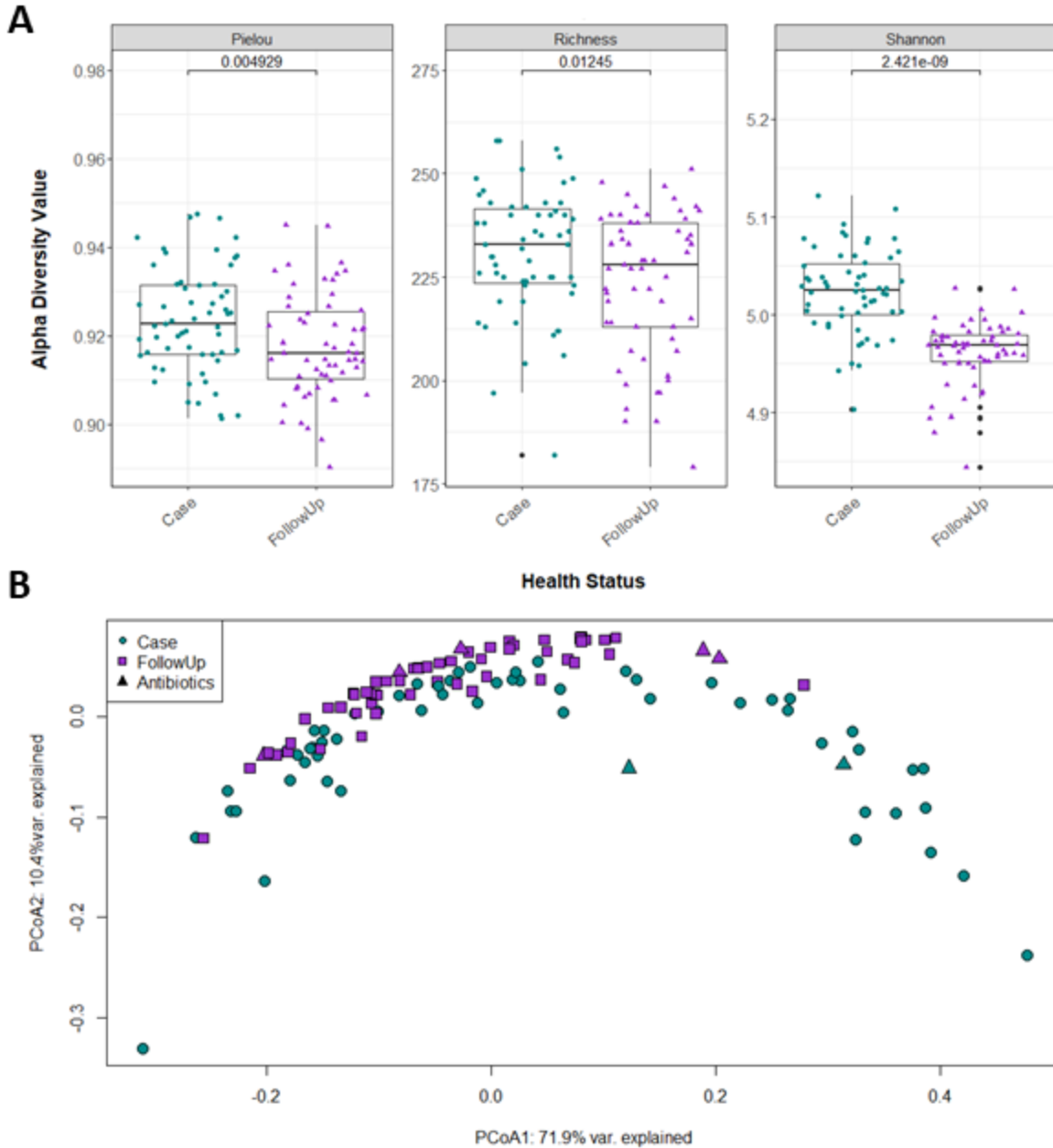
**Table C.7. Output generated by fold-change analysis exploring differentially abundant nonpolar metabolites among cases and follow-ups.**

Fifty clusters displaying the most positive (n=25) and most negative (n=25) FC are displayed.

Feature	FC	log2(FC)	q-value	Log(10)P
Cluster2756	130.7	7.0301	3.23E-07	6.4908
Cluster4470	120.25	6.9098	1.38E-08	7.8587
Cluster5193	96.167	6.5875	7.41E-08	7.1304
Cluster9173	89.224	6.4794	1.74E-08	7.7582
Cluster9161	86.561	6.4357	1.65E-08	7.7834
Cluster9433	86.488	6.4344	2.12E-08	7.6734
Cluster4097	83.748	6.388	1.71E-08	7.7683
Cluster12321	76.312	6.2538	1.63E-08	7.787
Cluster3182	76.29	6.2534	2.48E-08	7.6061
Cluster4146	75.869	6.2454	1.74E-08	7.7582
Cluster3981	72.755	6.185	1.89E-08	7.7235
Cluster10015	71.128	6.1524	2.93E-08	7.5338
Cluster4104	70.555	6.1407	2.17E-05	4.6637
Cluster4268	68.718	6.1026	1.38E-08	7.8587
Cluster8854	67.822	6.0837	2.05E-08	7.6888
Cluster3975	66.808	6.062	3.68E-08	7.4343
Cluster6467	65.861	6.0414	2.16E-08	7.666
Cluster5095	65.406	6.0313	1.09E-07	6.9621
Cluster5140	64.994	6.0222	1.41E-08	7.85
Cluster4360	64.12	6.0027	7.90E-06	5.1024
Cluster9185	62.779	5.9722	1.98E-08	7.7041
Cluster13564	60.809	5.9262	9.94E-08	7.0028
Cluster6841	58.589	5.8726	1.38E-08	7.8587
Cluster4438	57.866	5.8546	2.48E-08	7.6061
Cluster8582	56.592	5.8225	4.97E-08	7.3034

**Table C.7 (cont'd)**

Cluster10112	0.019187	-5.7037	1.38E-08	7.8587
Cluster70	0.018407	-5.7636	1.38E-08	7.8587
Cluster2134	0.018401	-5.7641	2.16E-08	7.666
Cluster1966	0.016446	-5.9261	2.41E-08	7.6188
Cluster2248	0.015764	-5.9872	5.07E-08	7.2947
Cluster7969	0.014679	-6.0901	2.37E-07	6.6255
Cluster6701	0.014623	-6.0956	1.80E-08	7.7438
Cluster4203	0.014186	-6.1394	3.48E-07	6.4587
Cluster2659	0.013127	-6.2514	1.38E-08	7.8587
Cluster1229	0.011217	-6.4781	1.64E-08	7.786
Cluster1377	0.010914	-6.5177	1.38E-08	7.8587
Cluster2683	0.009897	-6.6588	1.72E-08	7.7633
Cluster1756	0.008984	-6.7984	3.55E-08	7.4504
Cluster2137	0.008848	-6.8205	1.13E-07	6.9454
Cluster11997	0.008789	-6.8301	9.72E-08	7.0123
Cluster1232	0.008342	-6.9054	1.44E-07	6.8412
Cluster420	0.00821	-6.9283	1.38E-08	7.8587
Cluster1279	0.008205	-6.9293	3.17E-08	7.4986
Cluster2029	0.007836	-6.9956	2.16E-07	6.6658
Cluster1711	0.007814	-6.9998	1.38E-08	7.8587
Cluster1244	0.007526	-7.054	1.38E-08	7.8587
Cluster4988	0.007483	-7.0622	1.38E-08	7.8587
Cluster244	0.007219	-7.114	1.38E-08	7.8587
Cluster1618	0.00609	-7.3593	4.97E-07	6.304
Cluster321	0.002848	-8.4558	1.38E-08	7.8587

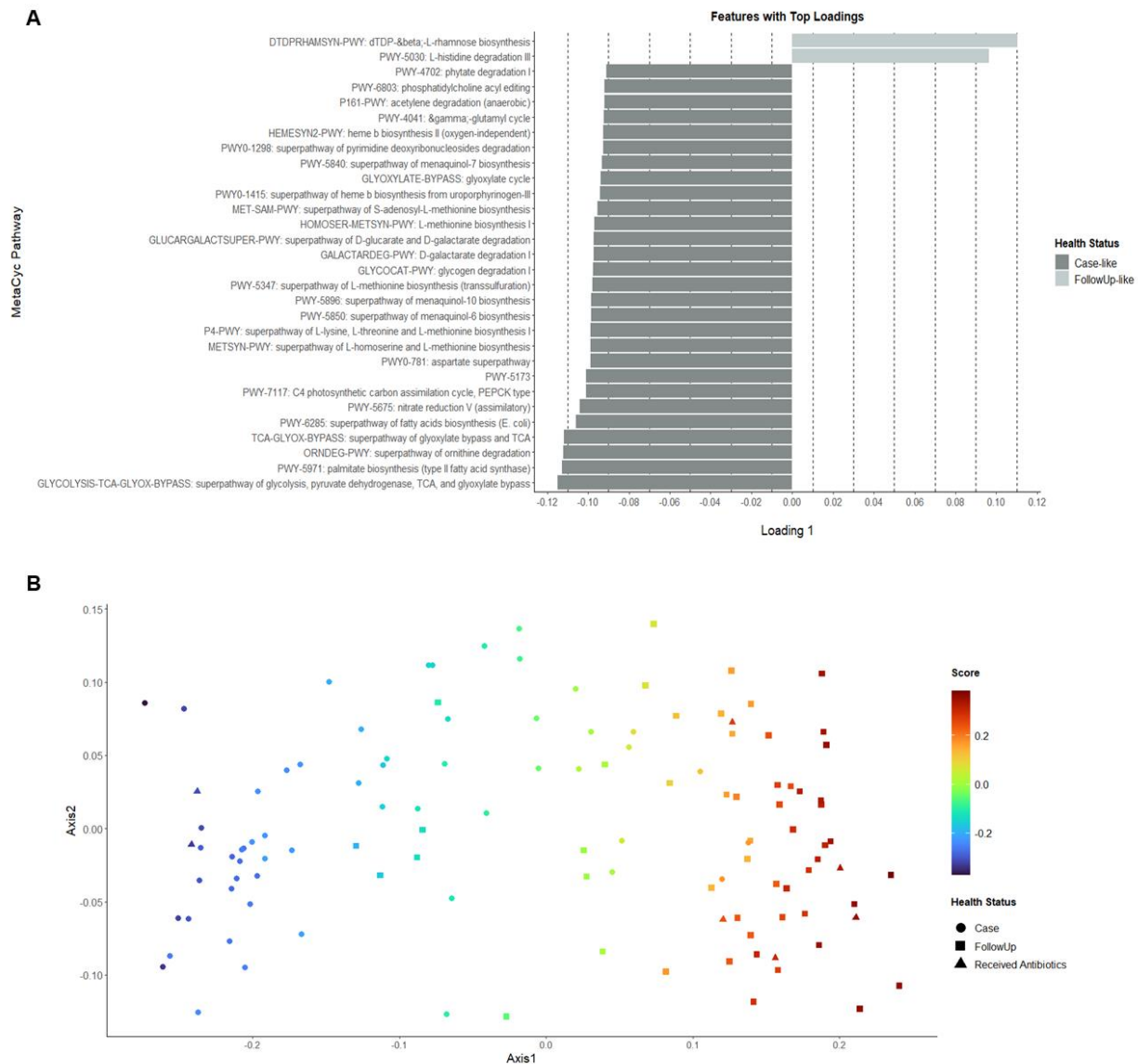


**Figure C.1. Metabolic diversity of KEGG modules significantly differs among patients during infection and after recovery.**

(A) Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou's Evenness) are displayed. The total number of metabolic modules was significantly greater in cases than follow-ups ( $S_{\text{case}}=231$ ,  $S_{\text{follow}}=224$   $p=0.0124$ ; Wilcoxon signed-rank test), with cases displaying a more diverse and even metabolic capacity ( $H'_{\text{case}}=5.02$ ,  $H'_{\text{follow-up}}=4.96$ ;  $p=2.41\text{e-}09$  and  $J'_{\text{case}}=0.923$ ,  $J'_{\text{follow-up}}=0.918$ ;  $p=0.0049$ , respectively). Each boxplot is stratified by health status with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top

**Figure C.1 (cont'd)**

of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. **(B)** Principal coordinates analysis (PCoA) was performed and plotted for cases (green, circles) and follow-ups (purple, squares) based on Bray-Curtis dissimilarity of KEGG module frequencies. A difference in the metabolic module composition (beta-diversity) was observed (PERMANOVA,  $F=9.33$ ;  $p=0.000999$ ) but the level of dispersion was also significantly different between case and follow-up samples (PERMDISP,  $F=29.52$ ;  $p=0.001$ ). The first and second coordinate of the PCoA are displayed with their respective percentage of variance explained. Individuals who reported use of antibiotics  $\geq 2$  weeks prior to sample collection are shown as triangular data points.



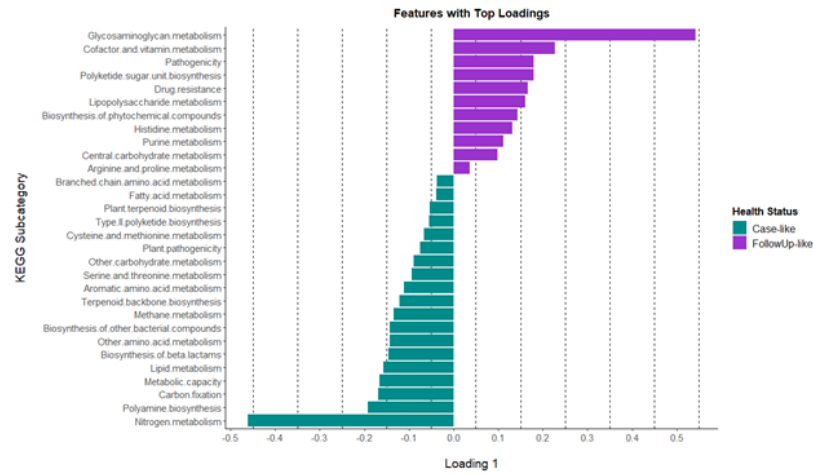
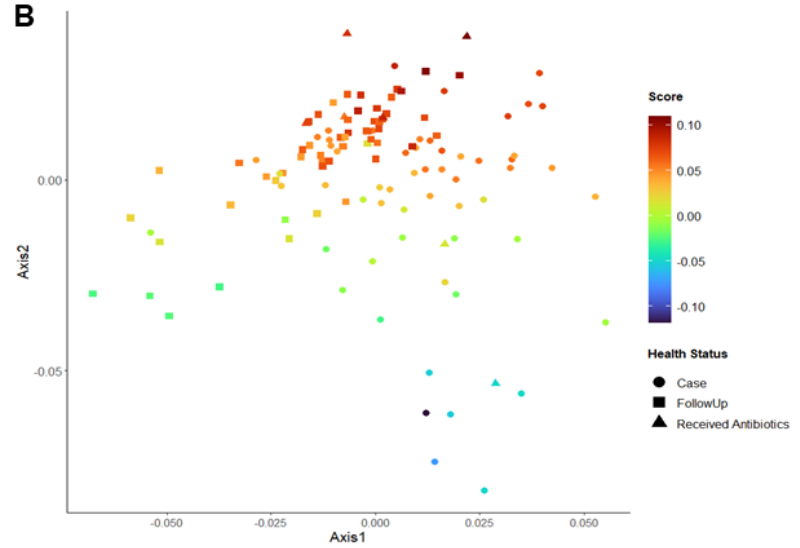
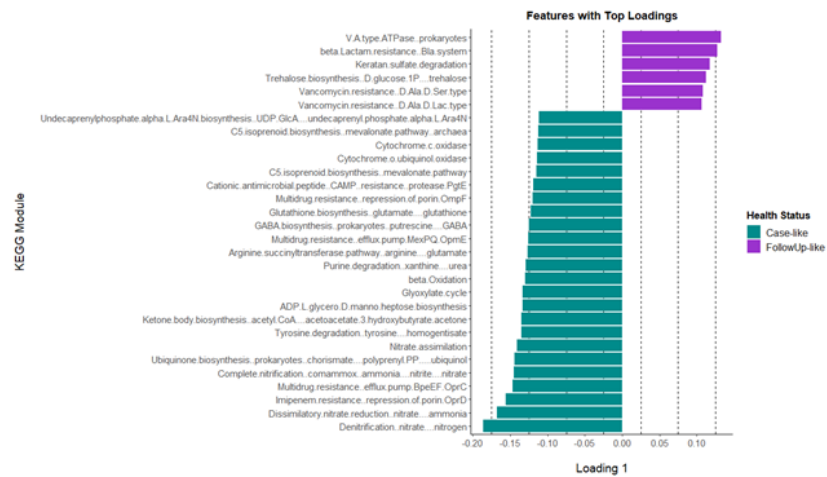
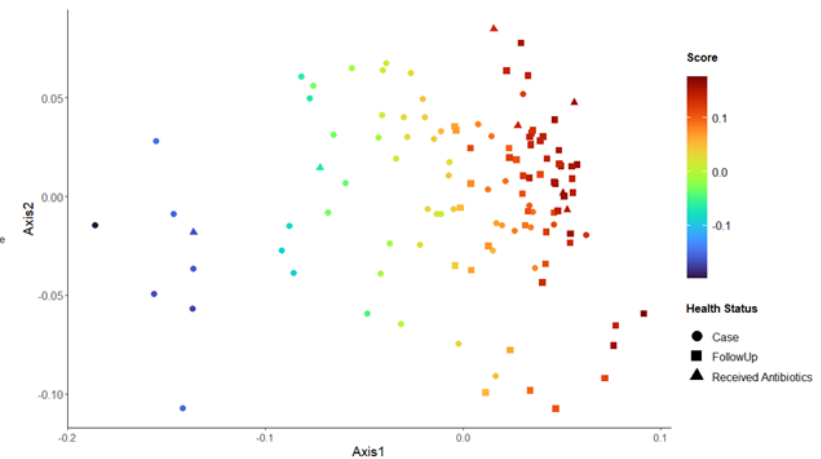
**Figure C.2. Investigation of continuous structure within pathways identified by HUMAnN 3.0 reveals metabolic tradeoffs.**

MMUPHin was used to identify metabolic features contributing to continuous structure among cases and follow-ups. **(A)** MetaCyc pathways determined to comprise the top consensus loadings of the multidimensional scaling plot are shown; colors have been assigned to the loadings based on their presumed “likeness” related to findings via differential abundance analysis (case-like=dark grey; follow-up-like=light grey). **(B)** Composition gradients are shown overlaid on ordination plots based on Bray-Curtis dissimilarity of pathway relative abundances among cases and follow-ups. Cases (circles) and follow-ups (squares) are shown in addition to individuals reporting use of antibiotics (triangles). The color gradient (“Score”) indicates the continuous structure score related to “Loading 1”, the top loading affiliated with the PCA. Considering the top loadings plots with the gradient-labeled ordination plots enable interpretation of metabolic

**Figure C.2 (cont'd)**

tradeoffs driving the observed distribution of points. For example, we observe a notable tradeoff between metabolic profiles dominated by rhamnose biosynthesis and histidine degradation and those with heavy signatures of glycolysis and glyoxylate bypass, ornithine degradation, and palmitate synthesis.

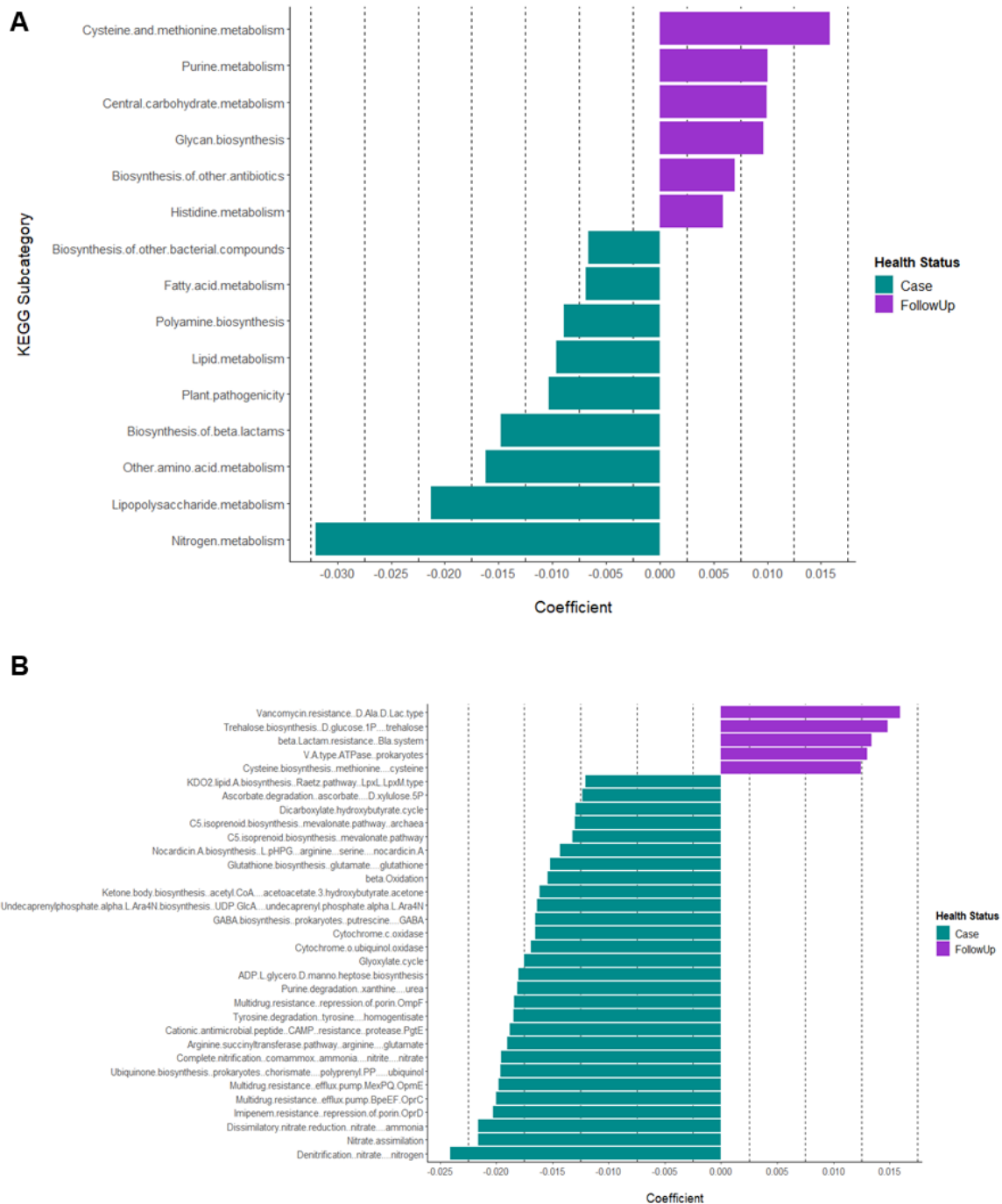


**A****B****C****D**

**Figure C.3. Investigation of continuous structure within module compositions reveals metabolic tradeoffs.**

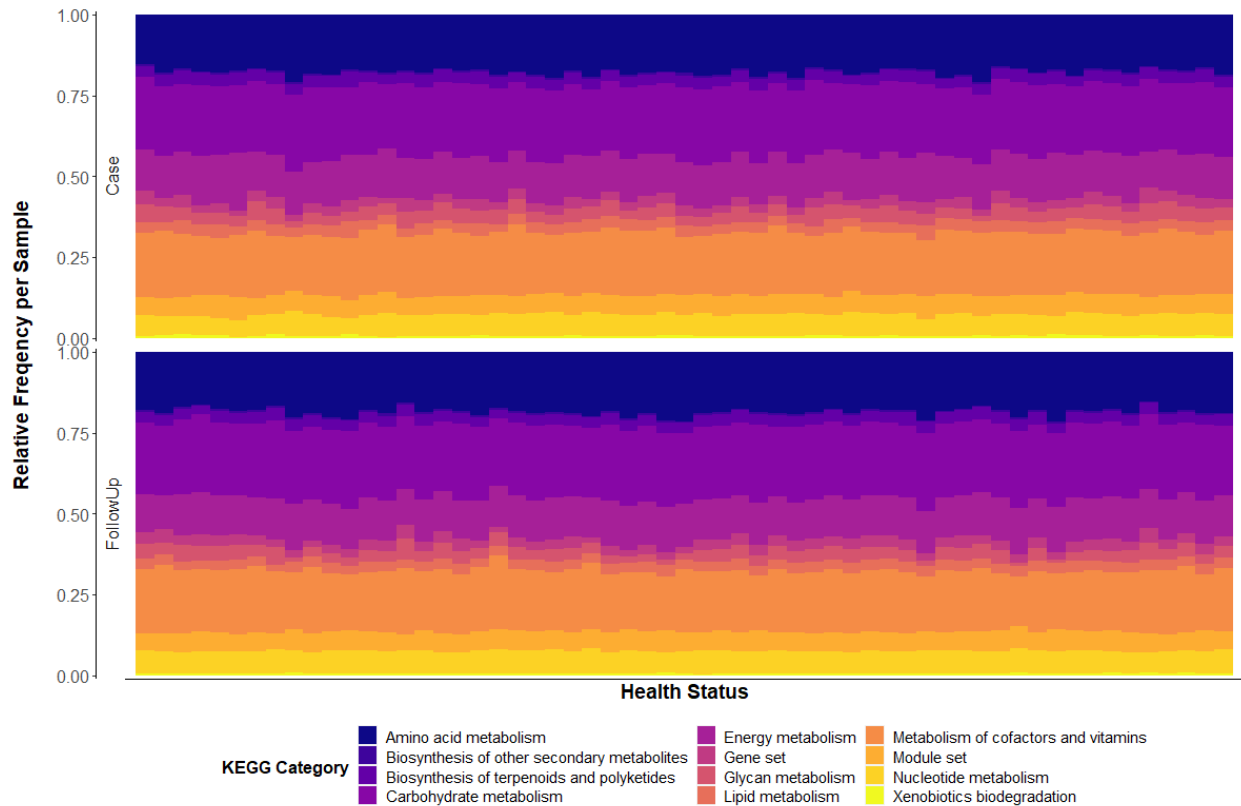
### Figure C.3 (cont'd)

MMUPHin was used to identify metabolic features influencing the distribution of points observed in the earlier ordination plots. KEGG subcategories (**A**) and modules (**C**) determined to comprise the top consensus loadings of the multidimensional scaling plot are shown; colors have been assigned to the loadings based on their presumed “likeness” related to findings via differential abundance analysis (case-like=green; follow-up-like=purple). Composition gradients for subcategory (**B**) and module (**D**) are shown overlaid on ordination plots based on Bray-Curtis dissimilarity at the subcategory or module level, respectively. Cases (circles) and follow-ups (squares) are shown in addition to individuals reporting use of antibiotics (triangles). The color gradient (“Score”) indicates the continuous structure score related to “Loading 1”, the top loading affiliated with the PCA. Considering the top loadings plots with the gradient-labeled ordination plots enable interpretation of metabolic tradeoffs driving the observed distribution of points. For example, at the subcategory level, we observe a notable tradeoff between metabolic profiles dominated by glycosaminoglycan metabolism and those dominated by nitrogen metabolism.



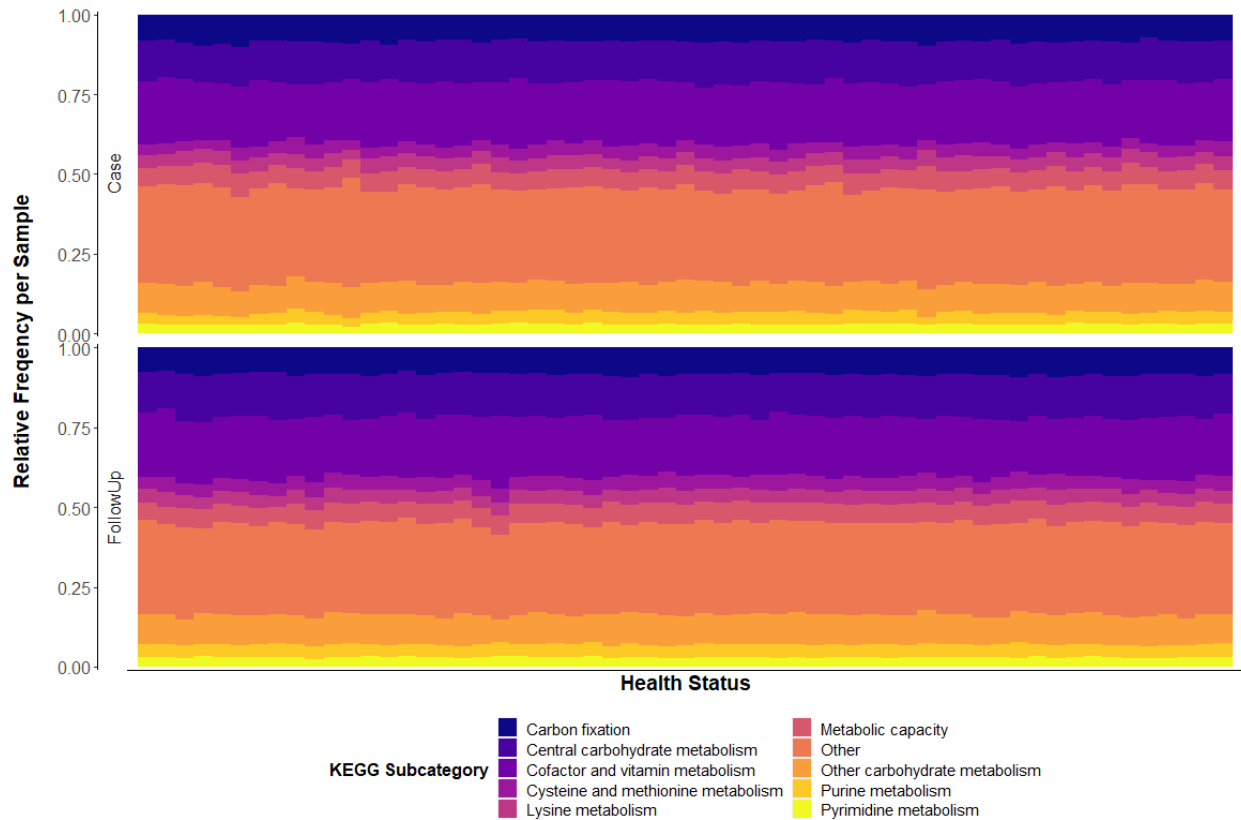
**Figure C.4. Metabolic subcategories and modules demonstrate different frequencies among cases and follow-ups.**

Differentially frequent metabolic features were defined using MMUPHin. Coefficients for KEGG subcategories (A) or modules (B) are displayed on the x-axis with an absolute value cutoff of  $>0.012$ . Positive coefficients indicate metabolic pathways with higher frequencies in follow-ups (purple) while negative coefficients show pathways more represented among cases (green). The metabolic subcategory or module is shown on the y-axis.



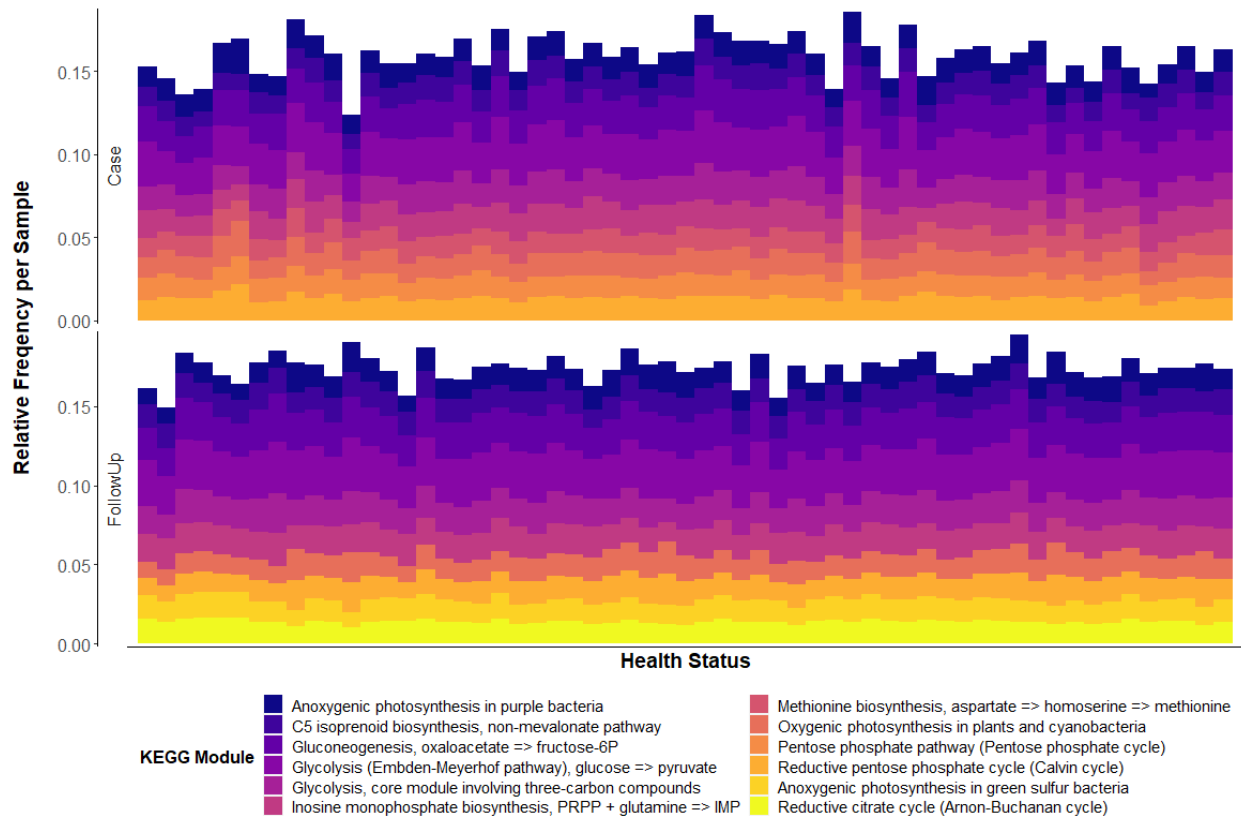
**Figure C.5. Relative abundance of KEGG metabolic categories is consistent across cases and follow-ups.**

Twelve metabolic categories are depicted in the relative abundance plots for cases (top) and follow-ups (bottom). Metabolic categories were consistent among the two sample types, with carbohydrate metabolism displaying the highest relative abundance (cases=22.1%; follow-ups=22.7%), followed by metabolism of cofactors and vitamins (19.4% and 18.9%, respectively). Each column represents one sample; columns are ordered by their sample pairing, meaning that the same column position within each facet represents the same individual at two different time points. Relative abundances were determined using category frequencies (i.e., number of contigs containing the relevant category) normalized by number of genome equivalents for each sample.



**Figure C.6. Relative abundance of KEGG metabolic subcategories also demonstrate consistency among cases and follow-ups.**

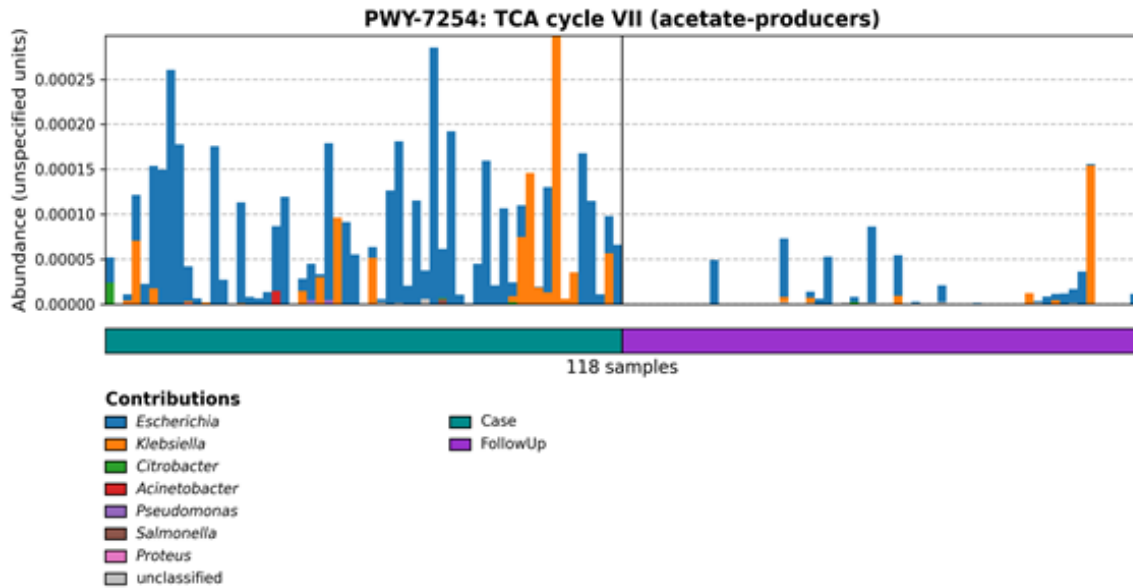
The top-ten metabolic subcategories are depicted in the relative abundance plots for cases (top) and follow-ups (bottom); notably, each health status contained the same top-ten subcategories overall. The highest relative frequencies were assigned to cofactor and vitamin metabolism (cases=19.4%; follow-ups=8.9%), central carbohydrate metabolism (12.8%; 13.4%), and “other carbohydrate metabolism” (9.3%; 9.3%). Each column represents one sample; columns are ordered by their sample pairing, meaning that the same column position within each facet represents the same individual at two different time points. Relative abundances were determined using subcategory frequencies (i.e., number of contigs containing the relevant subcategory) normalized by number of genome equivalents for each sample.



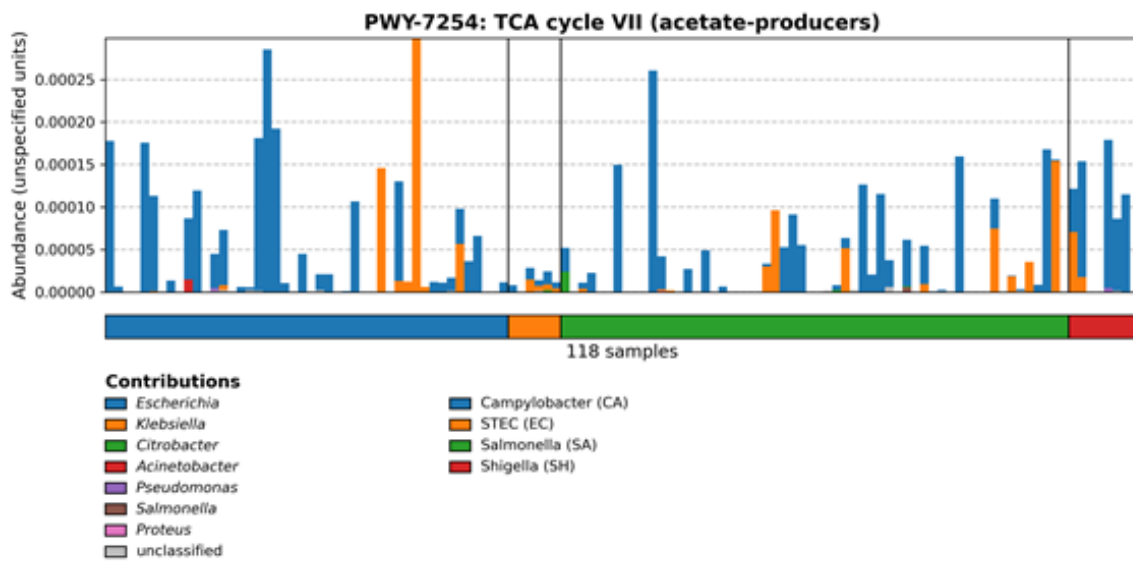
**Figure C.7. Investigating relative abundance of KEGG metabolic modules reveals slight discrepancies between cases and follow-ups.**

The top-ten metabolic modules are depicted in the relative abundance plots for cases (top) and follow-ups (bottom). Glycolysis was most frequent among both health statuses (cases=2.7%; follow-ups=2.9%) along with gluconeogenesis (2.1%; 2.3%). Each column represents one sample; columns are ordered by their sample pairing, meaning that the same column position within each facet represents the same individual at two different time points. Relative abundances were determined using subcategory frequencies (i.e., number of contigs containing the relevant subcategory) normalized by number of genome equivalents for each sample.

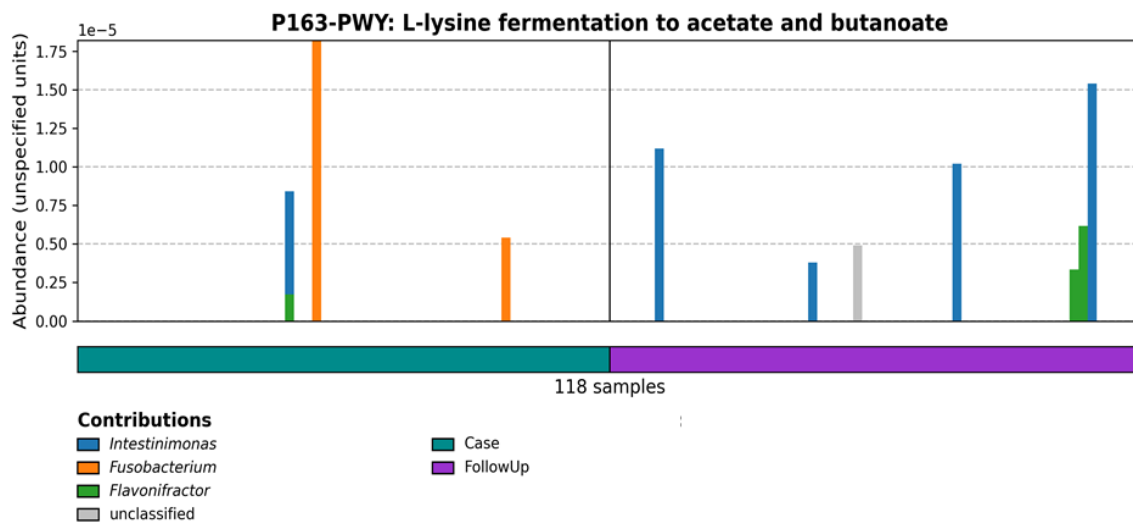
A



B



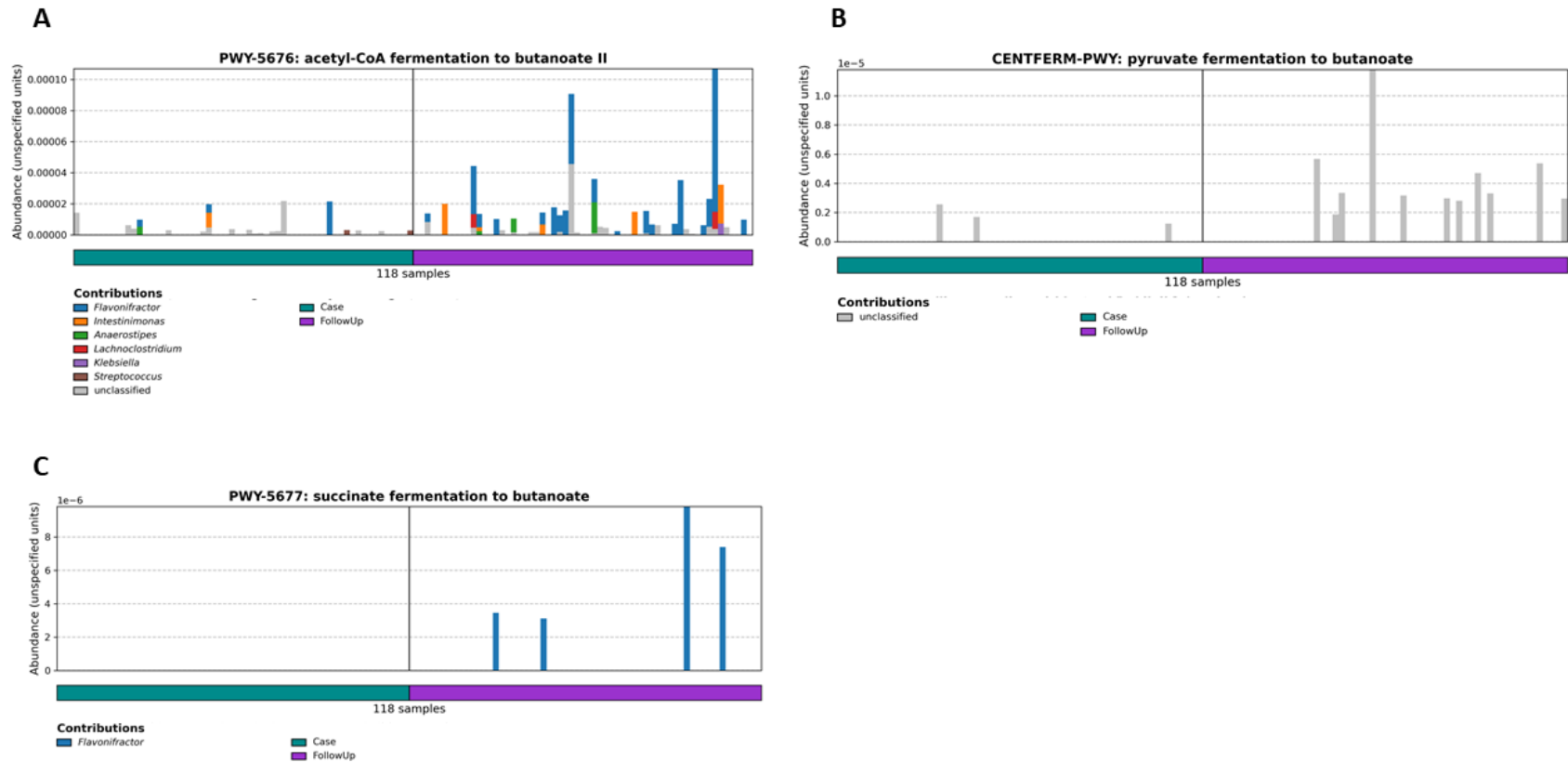
**Figure C.8. Relative abundances of PWY-7254: TCA cycle VII (acetate-producers).** Barplots show the relative abundance of PWY-7254 calculated by HUMAnN 3.0 stratified by health status (A). Samples were also stratified by infecting pathogen to observe potential genera-related associations affiliated with type of infection (B). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples when stratified by health status; for infecting pathogen, samples are grouped by *Campylobacter* (blue), *Salmonella* (green), *Shigella* (red), or STEC (orange). The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for PWY-7254 attributed to that specific genus.



**Figure C.9. Relative abundances of P163-PWY: L-lysine fermentation to acetate and butanoate.**

Barplots show the relative abundance of P163-PWY calculated by HUMAnN 3.0. The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for P163-PWY attributed to that specific genus.

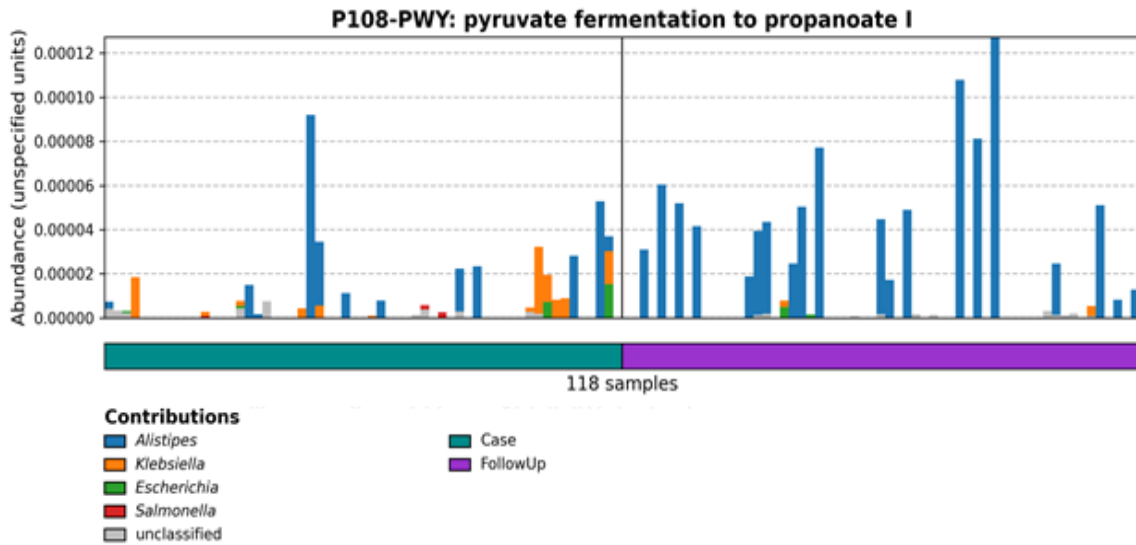




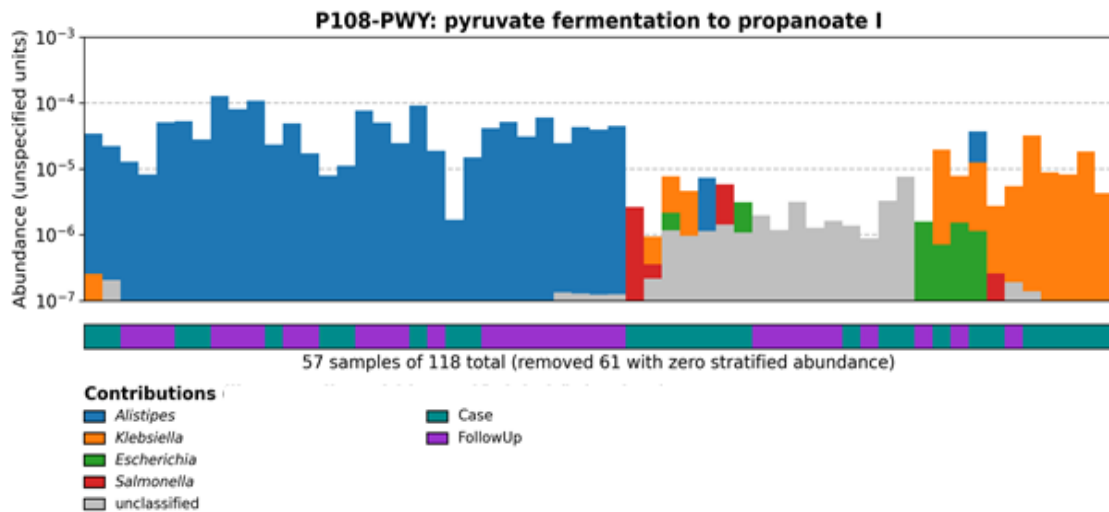
**Figure C.10. Relative abundances of three butanoate production pathways.**

Barplots show the relative abundance of relevant butanoate calculated by HUMAnN 3.0. PWY-5676: acetyl-CoA fermentation to butanoate II (A) was the most widely distributed of these pathways and appeared primarily in follow-ups. CENTFERM-PWY: pyruvate fermentation to butanoate (B) was most abundant, though was only sporadic among samples and assigned solely to ‘unclassified’ bacteria. PWY-5677: succinate fermentation to butanoate (C) was also found. The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances attributed to that specific genus. Samples were also stratified by infecting pathogen to observe potential genera-related associations affiliated with type of infection (bottom).

A

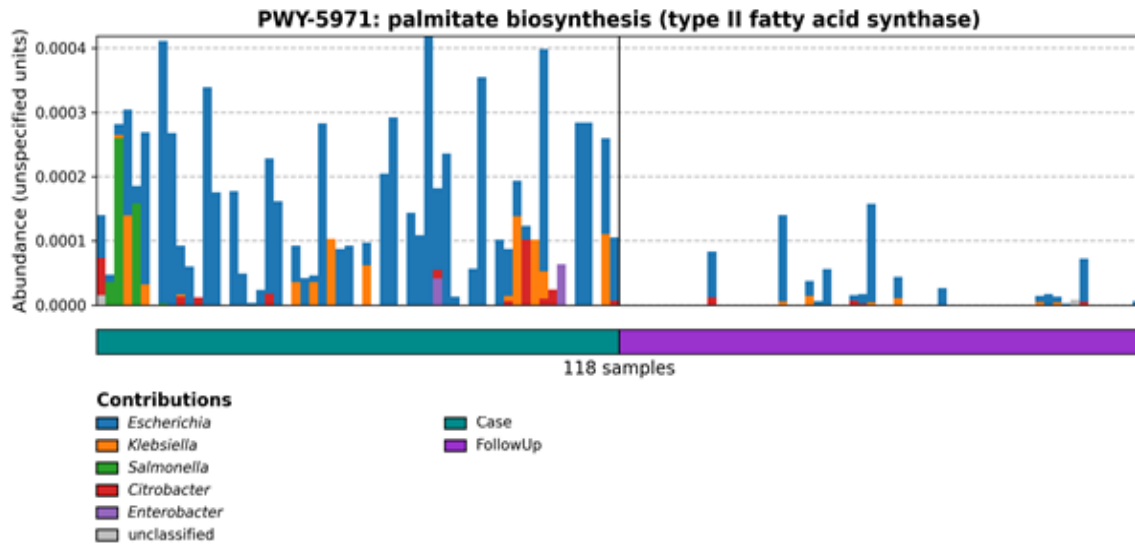


B

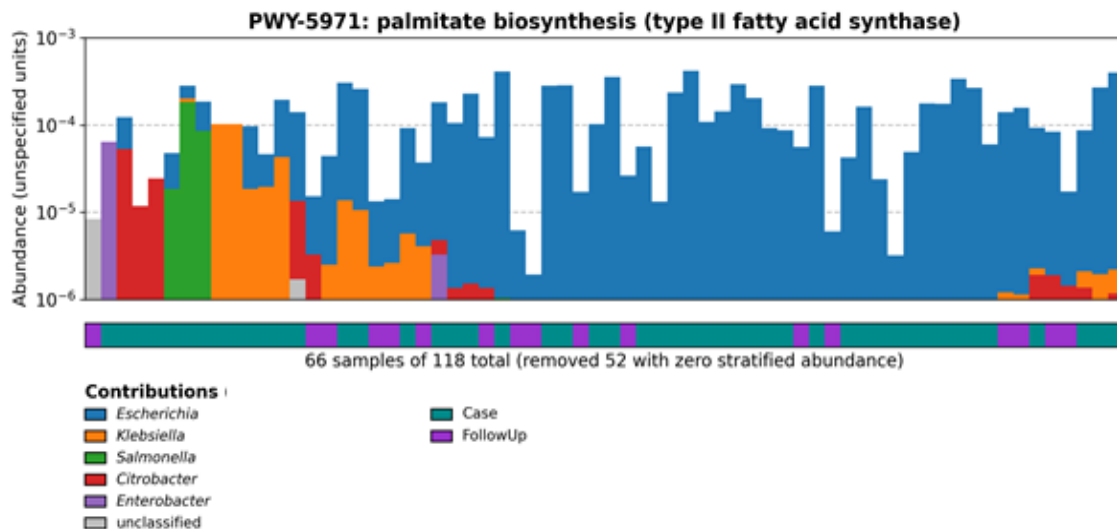


**Figure C.11. Relative abundances of P108-PWY: pyruvate fermentation to propanoate I.** Barplots show the relative abundance of P108-PWY calculated by HUMAnN 3.0 (A). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for P108-PWY attributed to that specific genus. Samples were also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (B).

A

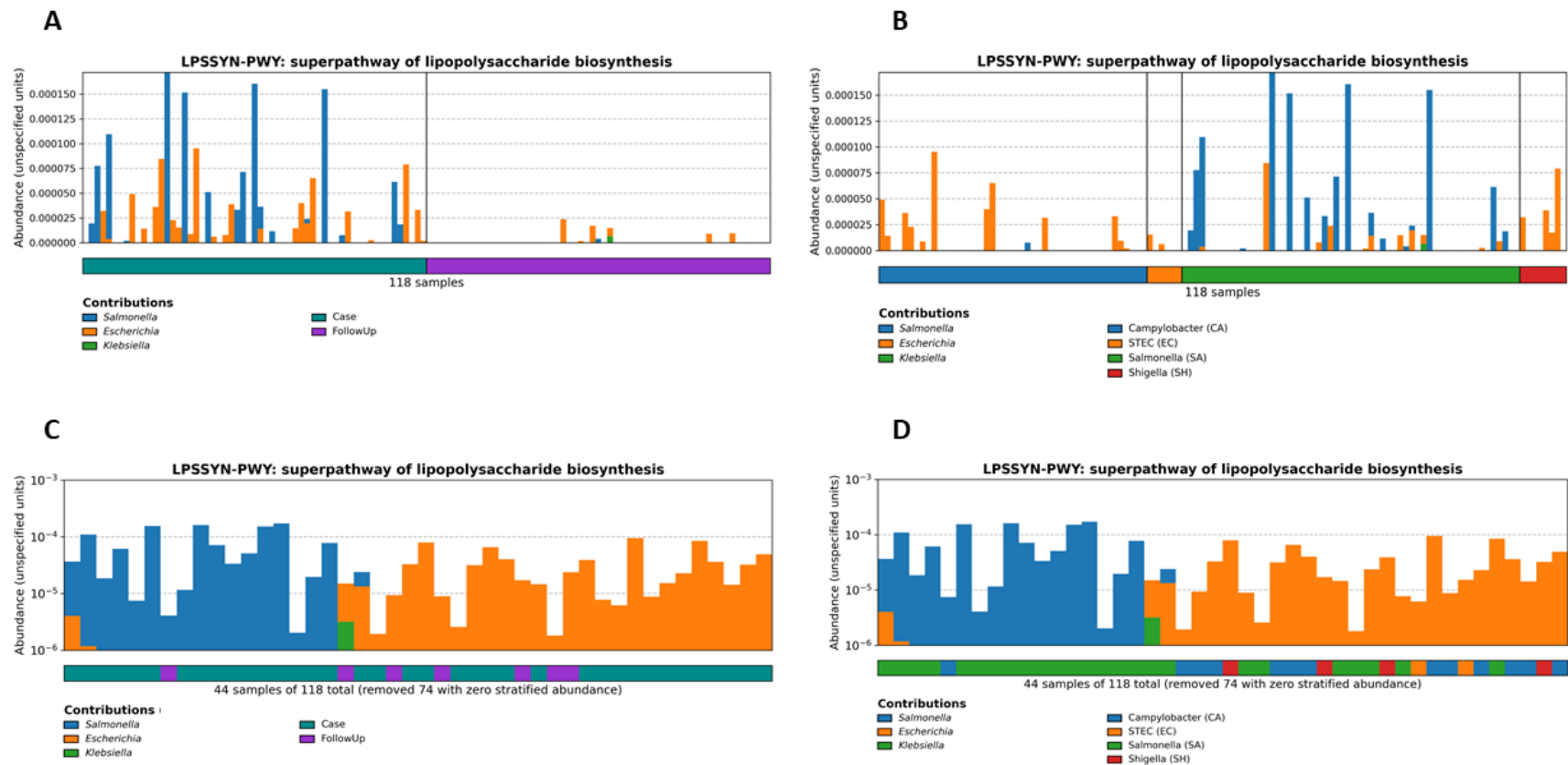


B



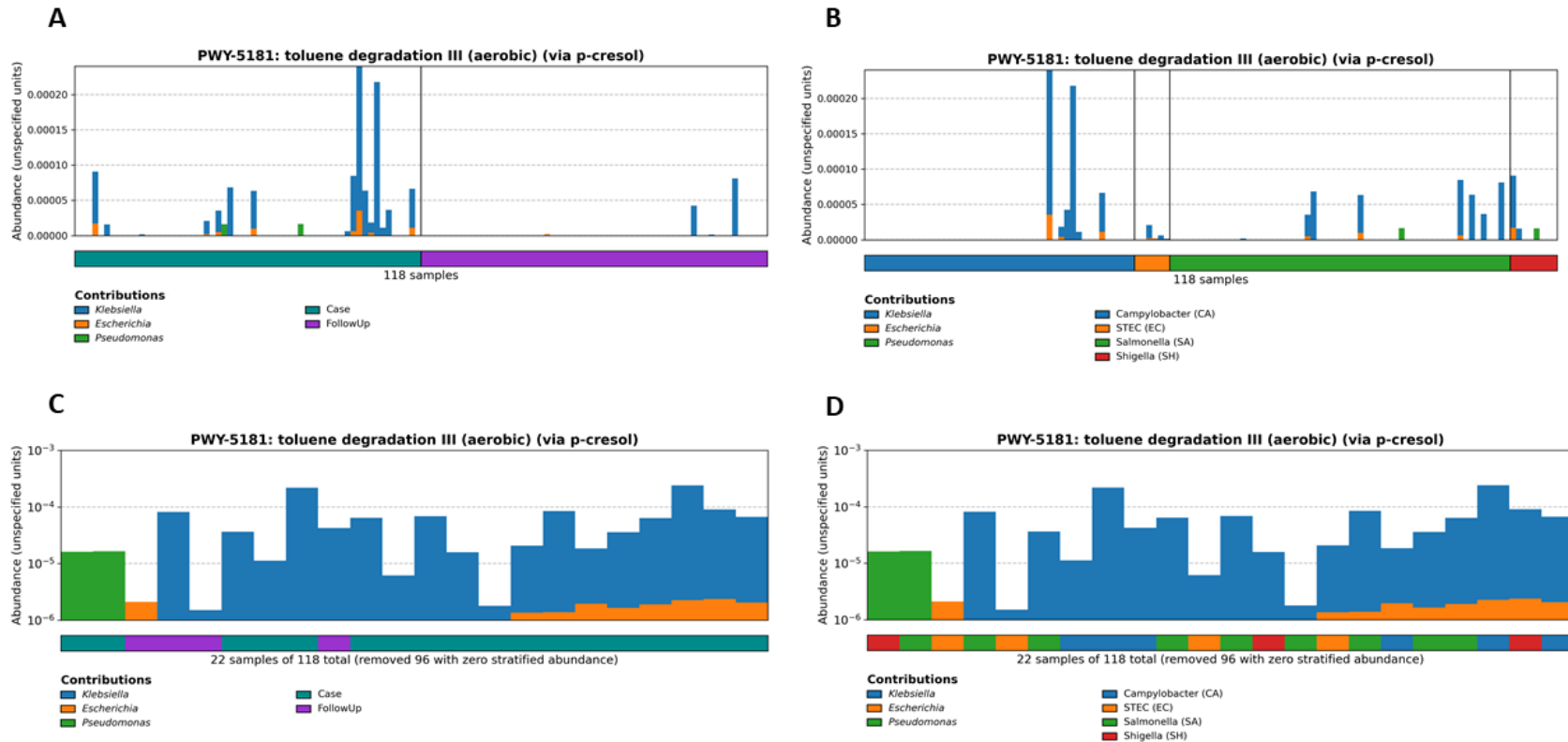
**Figure C.12. Relative abundances of PWY-5971: palmitate biosynthesis (type II fatty acid synthase).**

Barplots show the relative abundance of PWY-5971 calculated by HUMAnN 3.0 (A). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for PWY-5971 attributed to that specific genus. Samples were also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (B).



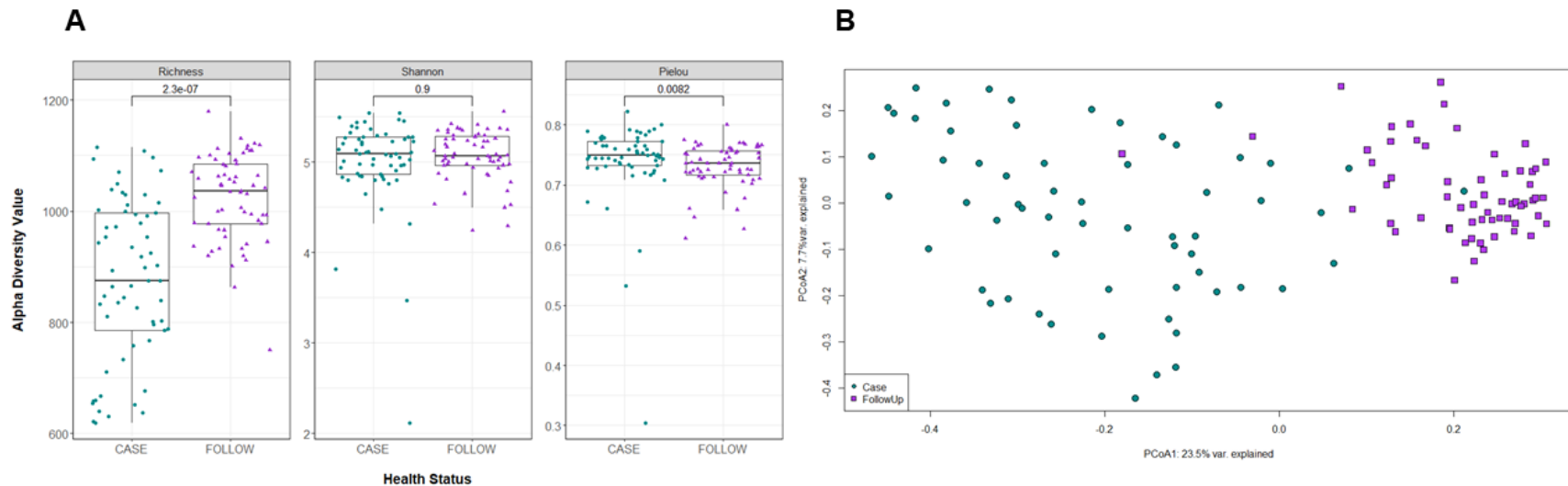
**Figure C.13. Relative abundances of lipopolysaccharide (LPS) biosynthesis among cases and follow-ups.**

Barplots show the relative abundance of the LPSSYN-PWY: superpathway of lipopolysaccharide biosynthesis calculated by HUMAnN 3.0 stratified by health status (A) and infecting pathogen (B). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples when stratified by health status; for infecting pathogen, samples are grouped by *Campylobacter* (blue), *Salmonella* (green), *Shigella* (red), or STEC (orange). The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for LPSSYN-PWY attributed to that specific genus. Samples were also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C and D).



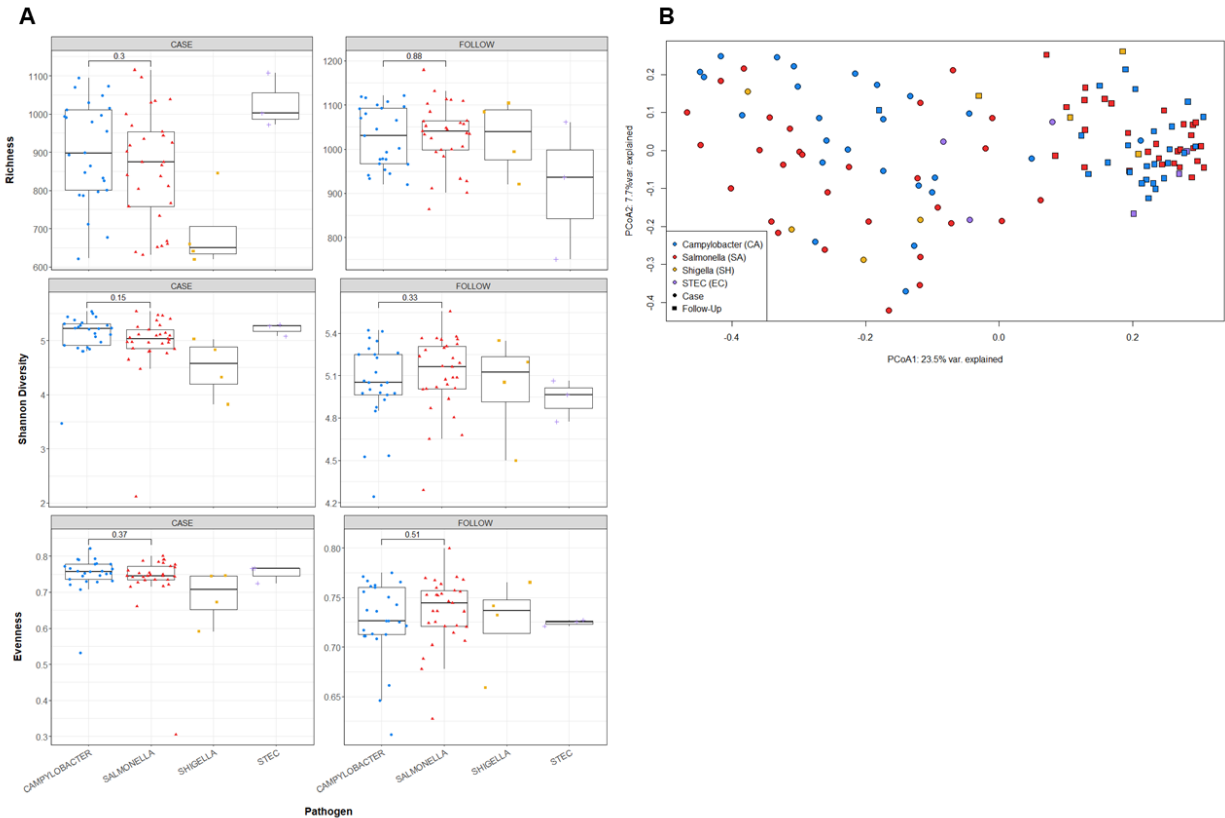
**Figure C.14. Relative abundances of toluene degradation via p-cresol among cases and follow-ups.**

Barplots show the relative abundance of the PWY-5181: toluene degradation III (aerobic) (via p-cresol) calculated by HUMAnN 3.0 stratified by health status (A) and infecting pathogen (B). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples when stratified by health status; for infecting pathogen, samples are grouped by *Campylobacter* (blue), *Salmonella* (green), *Shigella* (red), or STEC (orange). The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for PWY-5181 attributed to that specific genus. Samples were also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C and D).



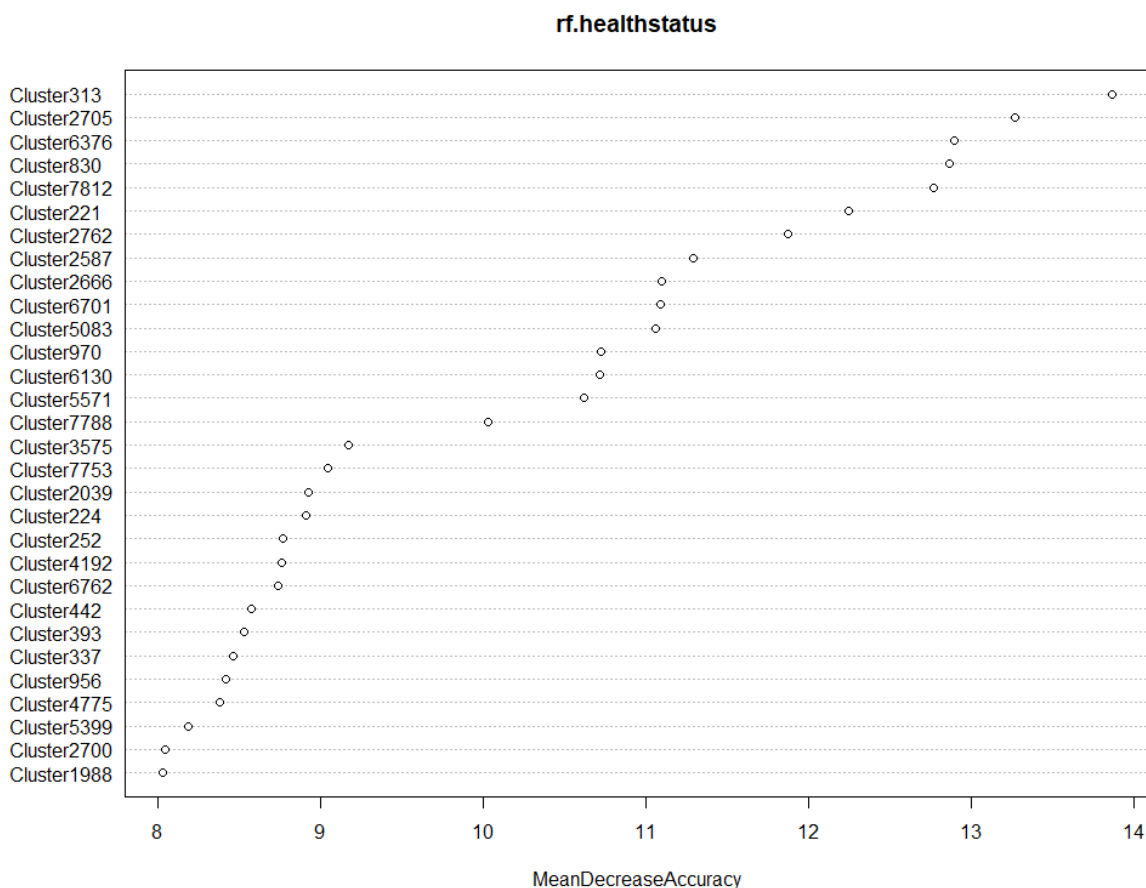
**Figure C.15. Richness and composition of polar metabolites significantly differs among patients during infection and after recovery.**

**(A)** Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou's Evenness) are displayed. Each box-plot is stratified by health status with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. **(B)** Principal coordinates analysis (PCoA) was performed and plotted for cases (green, circles) and follow-ups (purple, squares) based on Bray-Curtis dissimilarity of polar metabolite quantification. The first and second coordinate are displayed with their respective percentage of variance explained.



**Figure C.16. Richness and composition of polar metabolites does not appear to be influenced by infecting pathogen.**

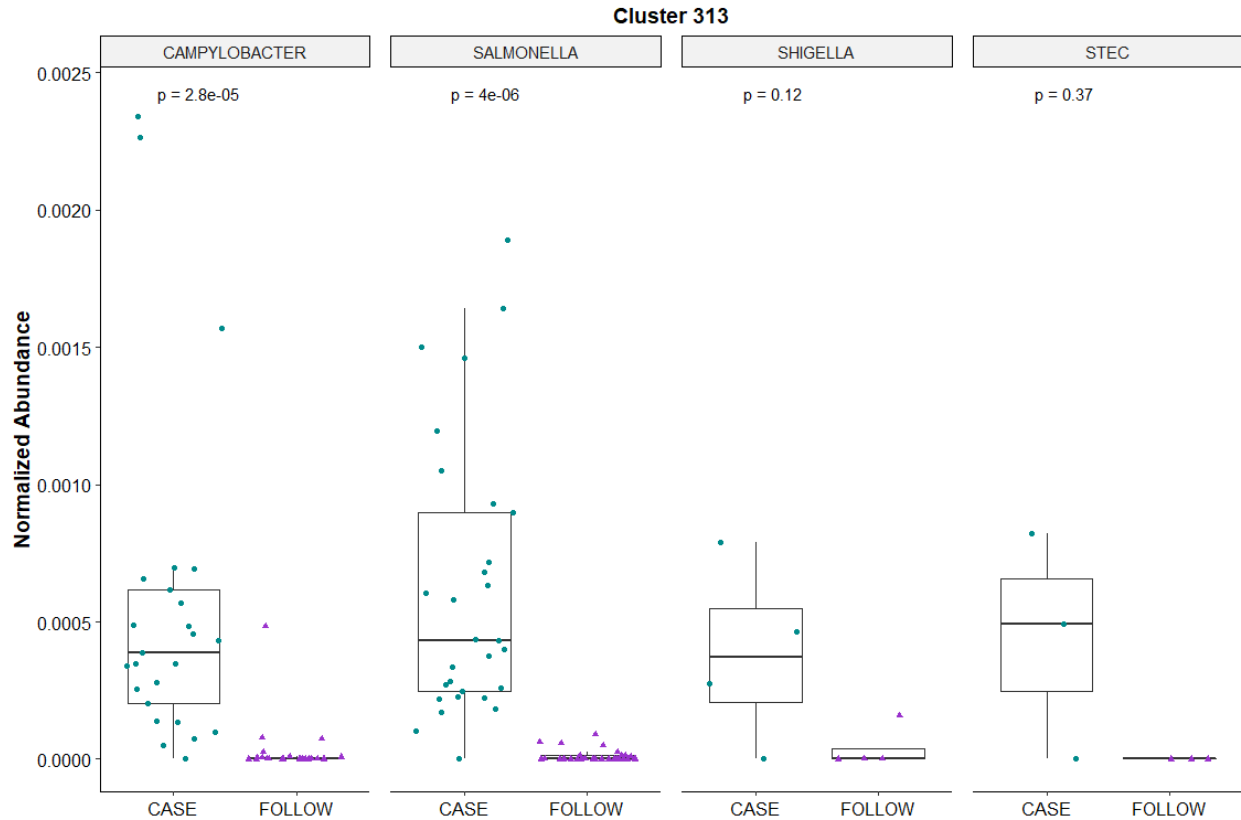
(A) Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou's Evenness) are displayed. Each box-plot is stratified by infecting pathogen with samples represented by circles (*Campylobacter*, blue), triangles (*Salmonella*, red), squares (*Shigella*, yellow), or crosses (STEC, violet). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. (B) Principal coordinates analysis (PCoA) was performed and plotted for cases (circles) and follow-ups (squares) based on Bray-Curtis dissimilarity of polar metabolite quantification. Points were colored based on the source of enteric infection (*Campylobacter*, blue; *Salmonella*, red; *Shigella*, yellow; or STEC, violet) The first and second coordinate are displayed with their respective percentage of variance explained.



**Figure C.17. Mean decrease in accuracy plot from random forest analysis of polar metabolites.**

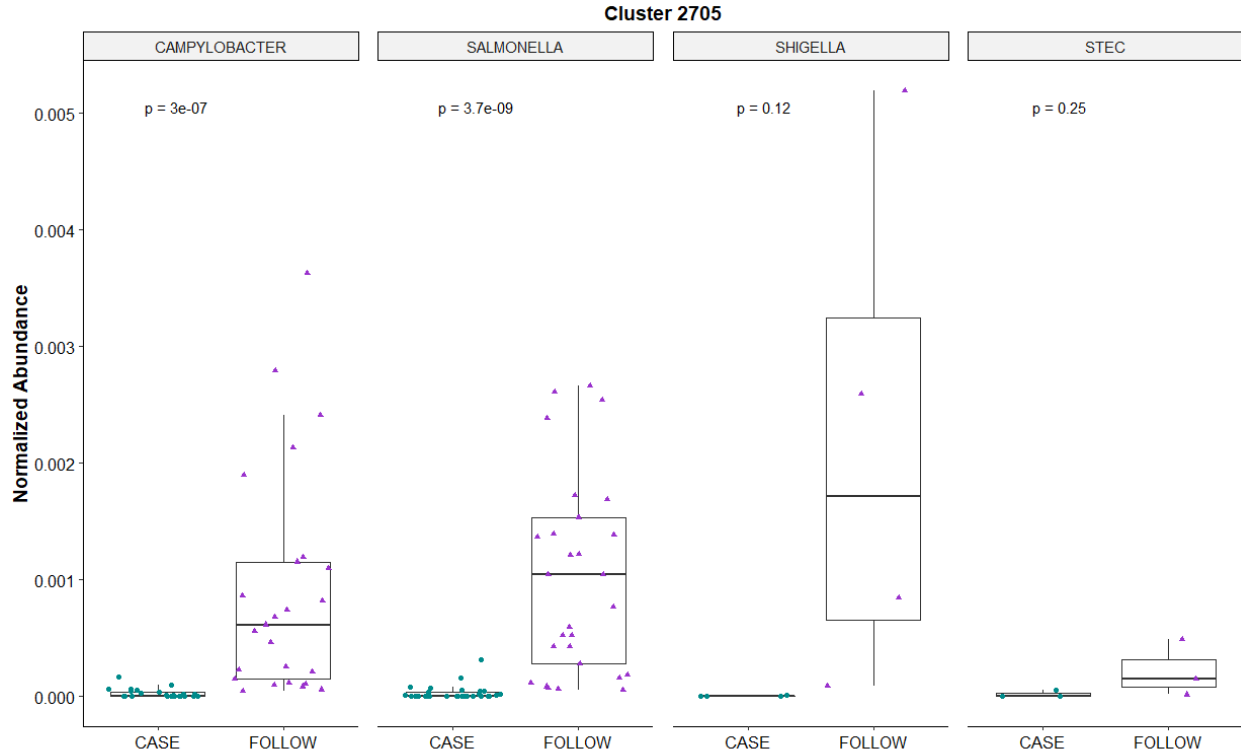
Random forest was run on polar metabolite intensities and set to classify samples by health status. The dot plot displays the top-30 clusters (metabolites) which were found to be of highest importance when distinguishing cases and follow-ups. Cluster index is noted on the y-axis, while mean decrease in accuracy is plotted on the x-axis. A higher mean decrease in accuracy metric indicates greater importance in the classifying algorithm.





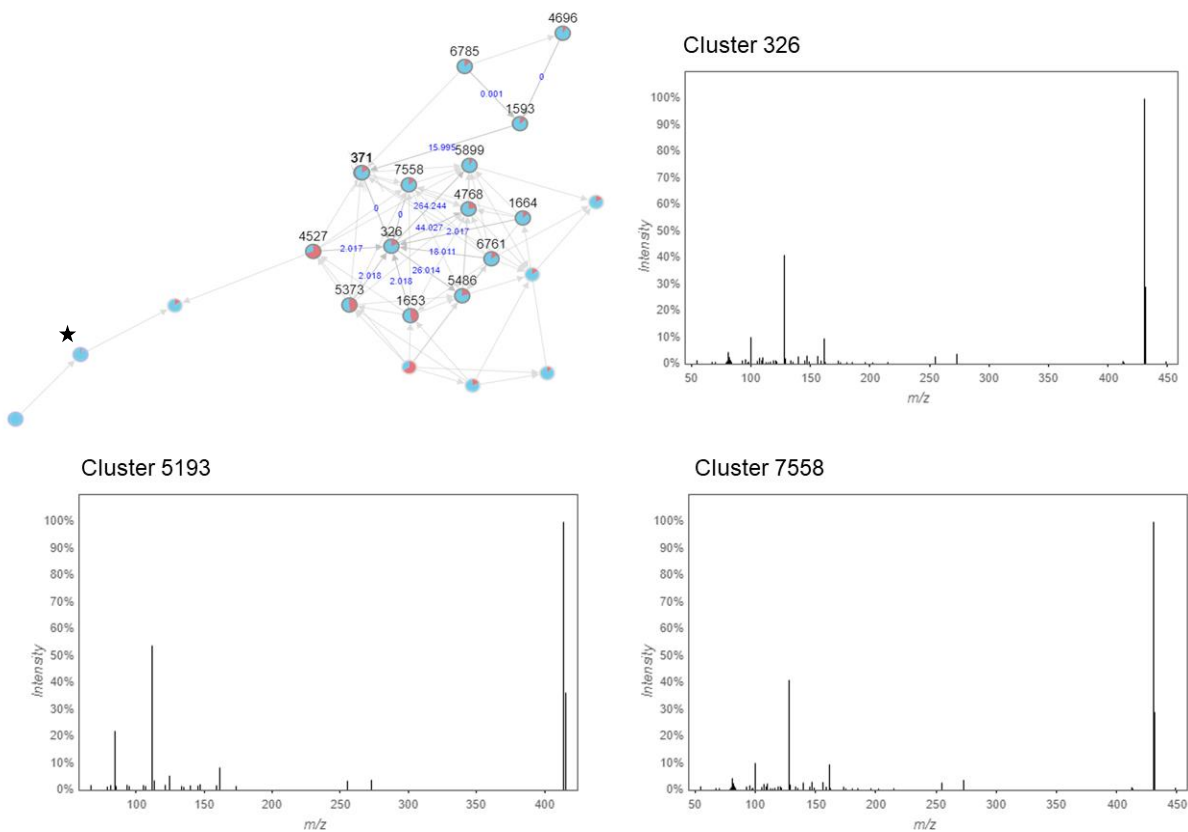
**Figure C.18. Normalized abundance of Cluster 313 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 313 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



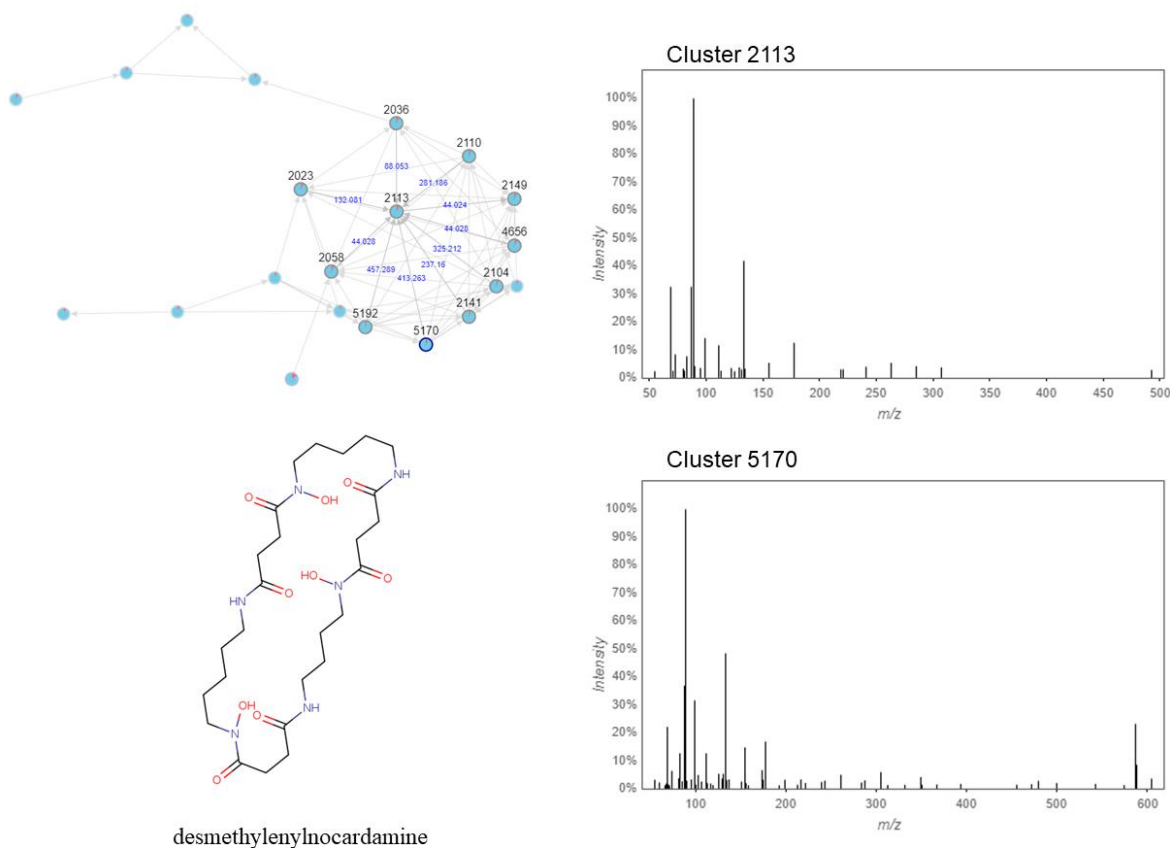
**Figure C.19. Normalized abundance of Cluster 2705 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 2705 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



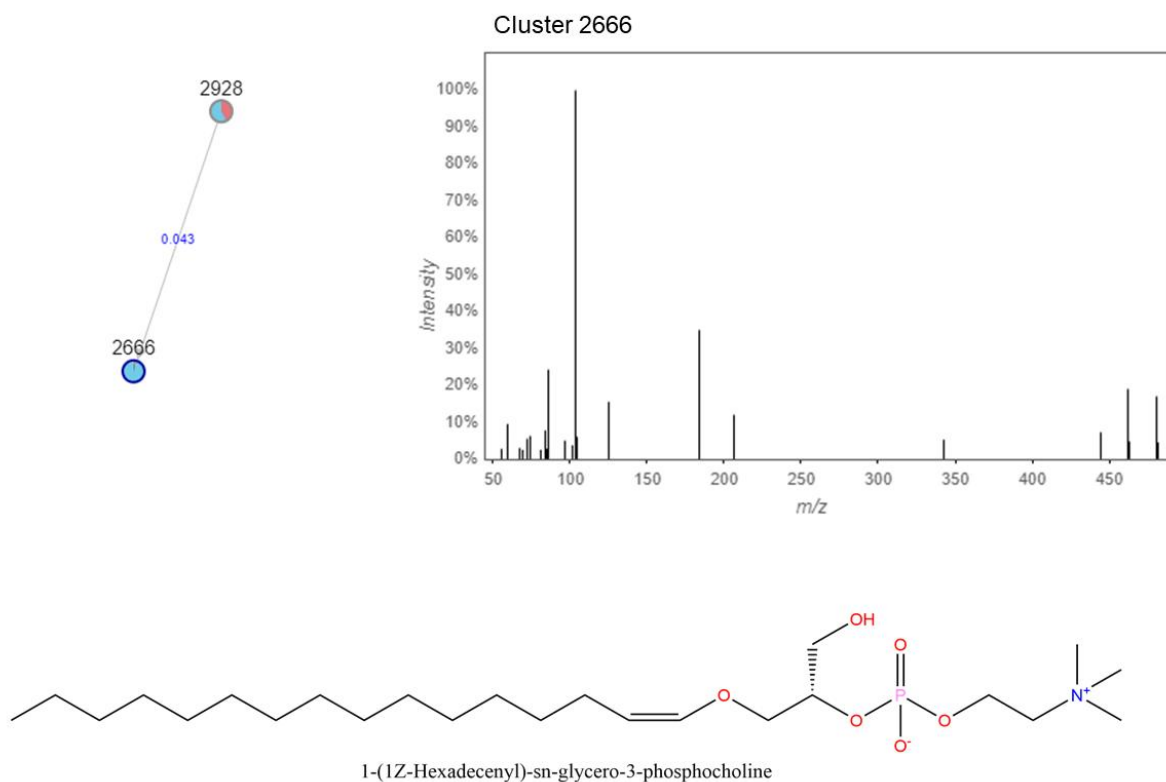
**Figure C.20. Molecular network and MS<sup>2</sup> spectra for three related clusters prevalent in follow-ups.**

A molecular network constructed in GNPS (top, left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra for clusters found to be important indicators of follow-ups (Clusters 326, 7558, and 5193) are shown. Notably, this molecular network contained metabolites related to tomatidine, which is the cluster designated with a star.



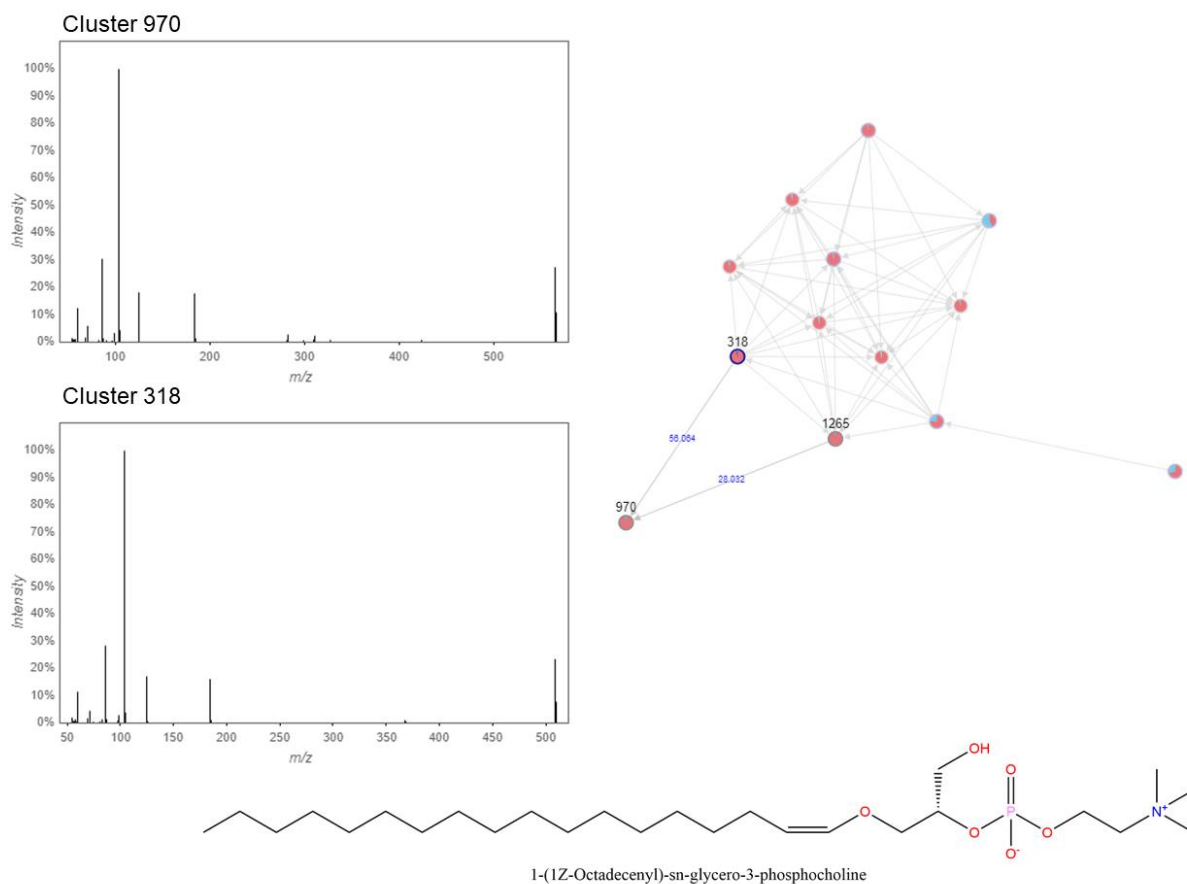
**Figure C.21. Molecular network and MS<sup>2</sup> spectra for Cluster 2113, which was highly represented in follow-ups.**

A molecular network constructed in GNPS (top, left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 2113 and a closely related cluster, 5170, are shown. Cluster 5170 was successfully annotated in GNPS as desmethylenylnocardamine; the structure for this compound was generated in ChemDraw 20.1 and is also shown (bottom, left).



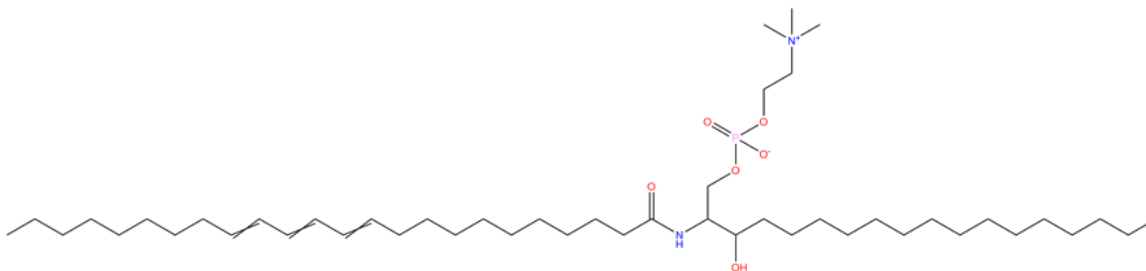
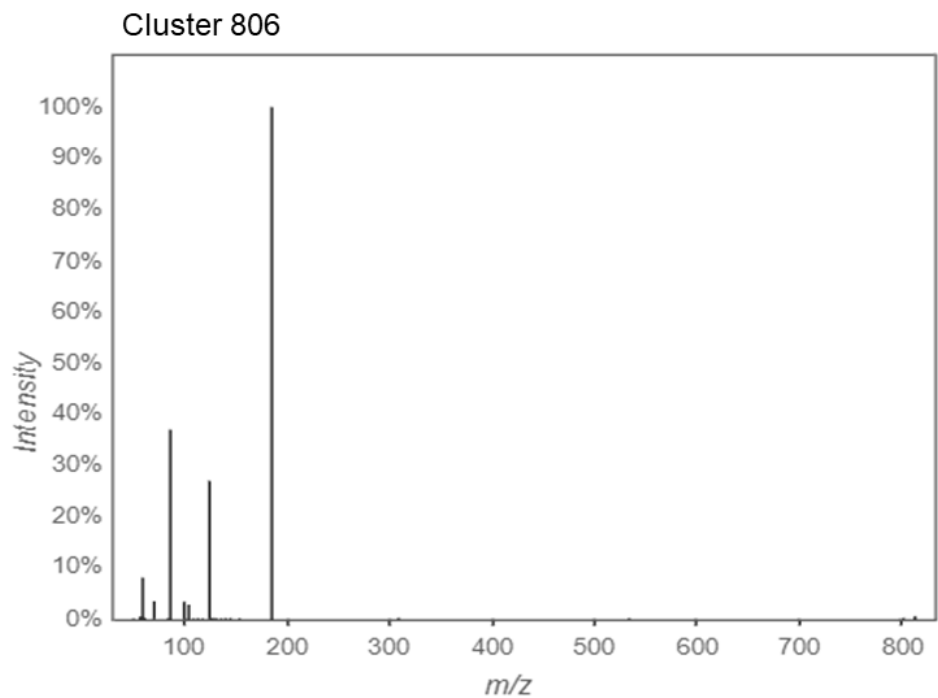
**Figure C.22. Molecular network and MS<sup>2</sup> spectra for Cluster 2666, which was prevalent in follow-ups.**

A molecular network constructed in GNPS (top, left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 2666 is shown. This cluster was successfully annotated in GNPS as 1-(1Z-Hexadecenyl)-sn-glycero-3-phosphocholine; the structure was generated in ChemDraw 20.1 and is included (bottom).



**Figure C.23. Molecular network and MS<sup>2</sup> spectra for Cluster 970 and related Cluster 318 which were present primarily in cases.**

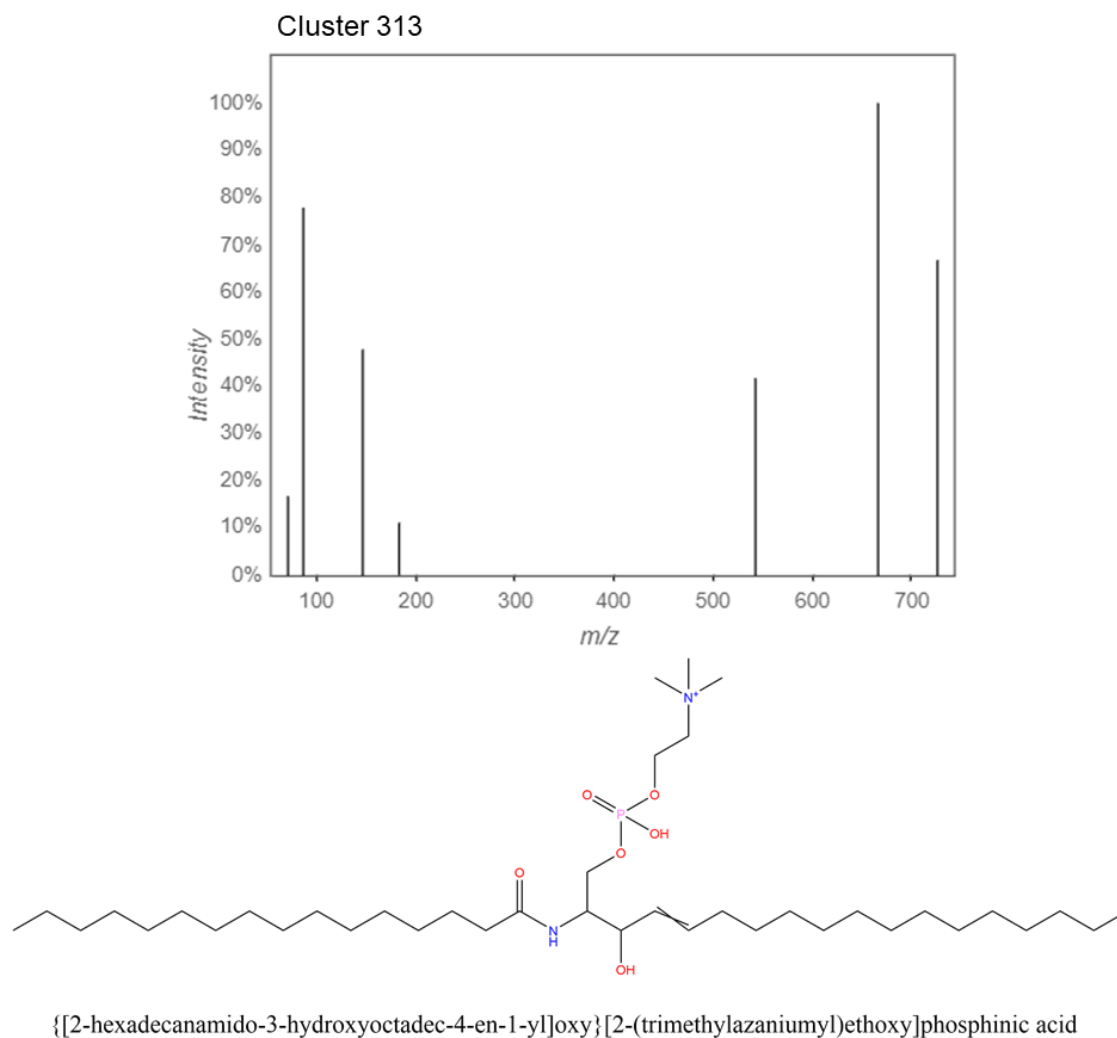
A molecular network constructed in GNPS (right) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 970 and related Cluster 318 is shown. Cluster 318 was successfully annotated in GNPS as 1-(1Z-Octadecenyl)-sn-glycero-3-phosphocholine; the structure was generated in ChemDraw 20.1 and is included (bottom).



3-hydroxy-2-(tetracos-11.13.15-trienamido)octadecyl (2-(trimethylammonio)ethyl) phosphate

**Figure C.24. Molecular network and MS<sup>2</sup> spectra for Cluster 806, which was abundant in cases.**

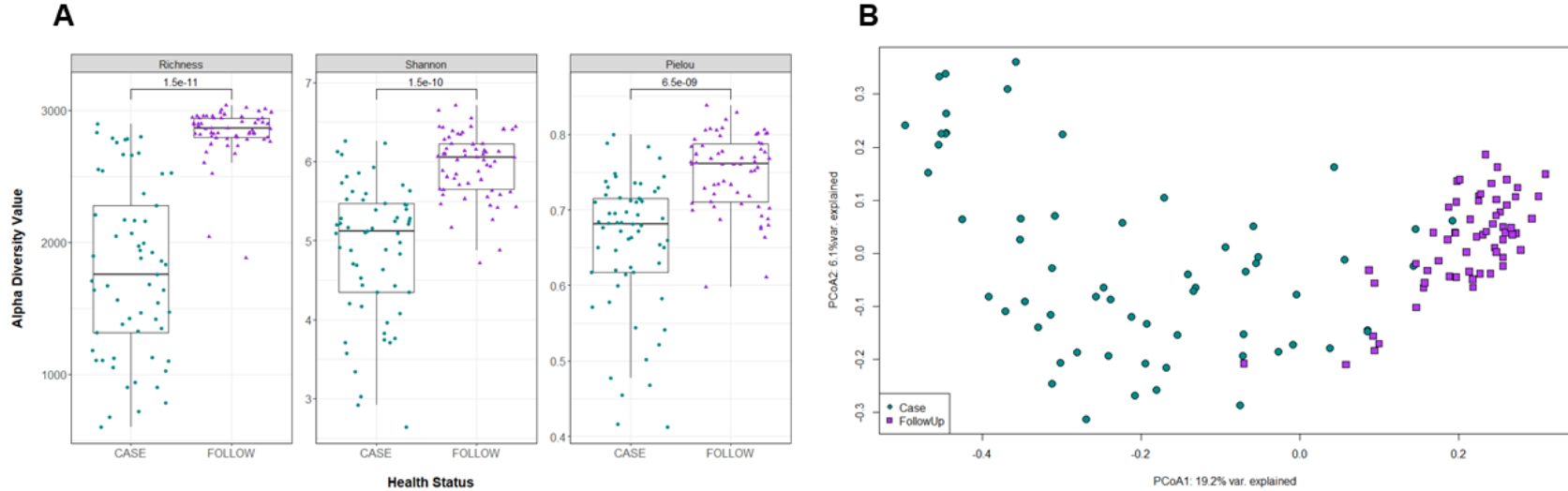
The MS<sup>2</sup> spectra Cluster 806 is shown (top). This cluster was successfully annotated as 3-hydroxy-2(tetracos-11.13.15-trienamido)octadecyl (2-(trimethylammonio)ethyl) phosphate; the chemical structure was generated in ChemDraw 20.1 and is shown (bottom).



**Figure C.25. Molecular network and MS<sup>2</sup> spectra for Cluster 313, which was abundant in cases.**

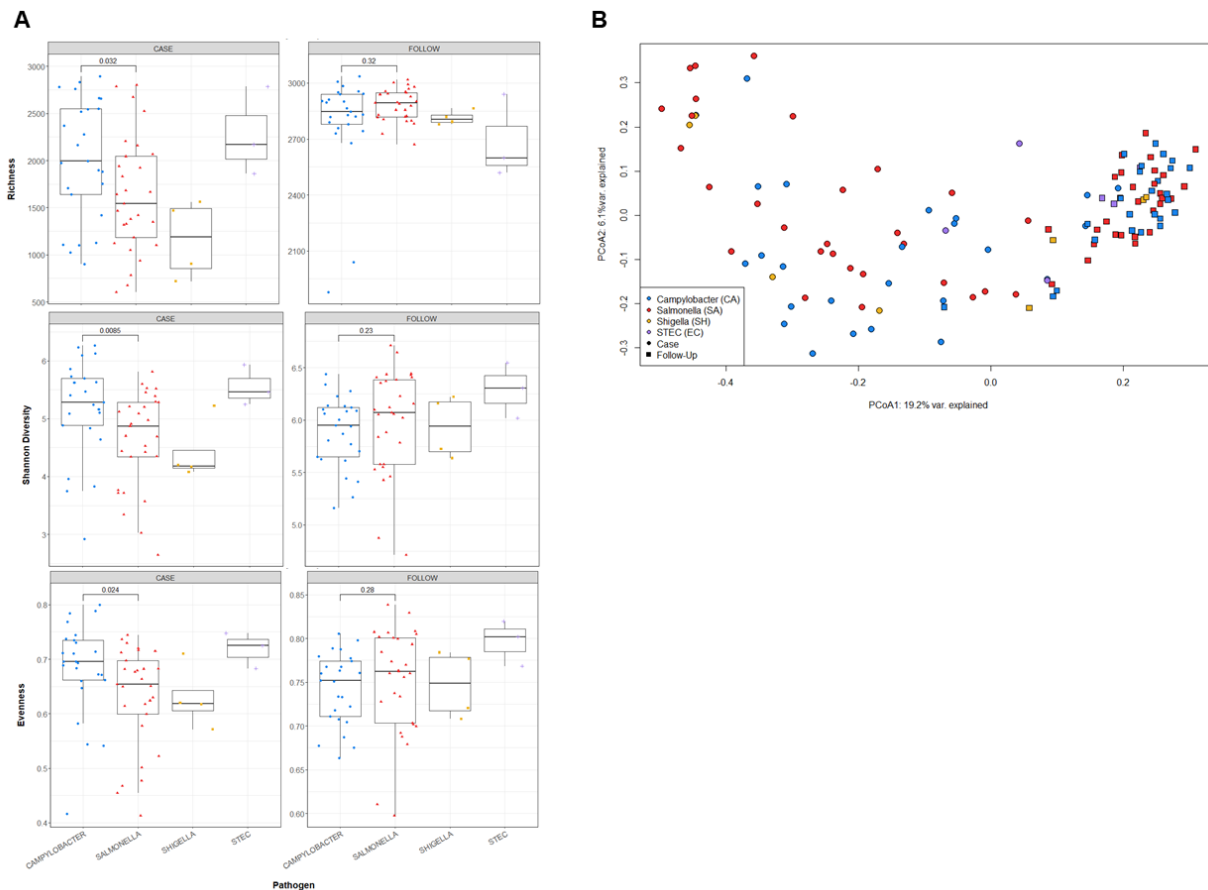
The MS<sup>2</sup> spectra Cluster 313 is shown (top). This cluster was successfully annotated as {[2-hexadecanamido-3-hydroxyoctadec-4-en-1-yl]oxy}[2-(trimethylazaniumyl)ethoxy]phosphinic acid; the chemical structure was generated in ChemDraw 20.1 and is shown (bottom).





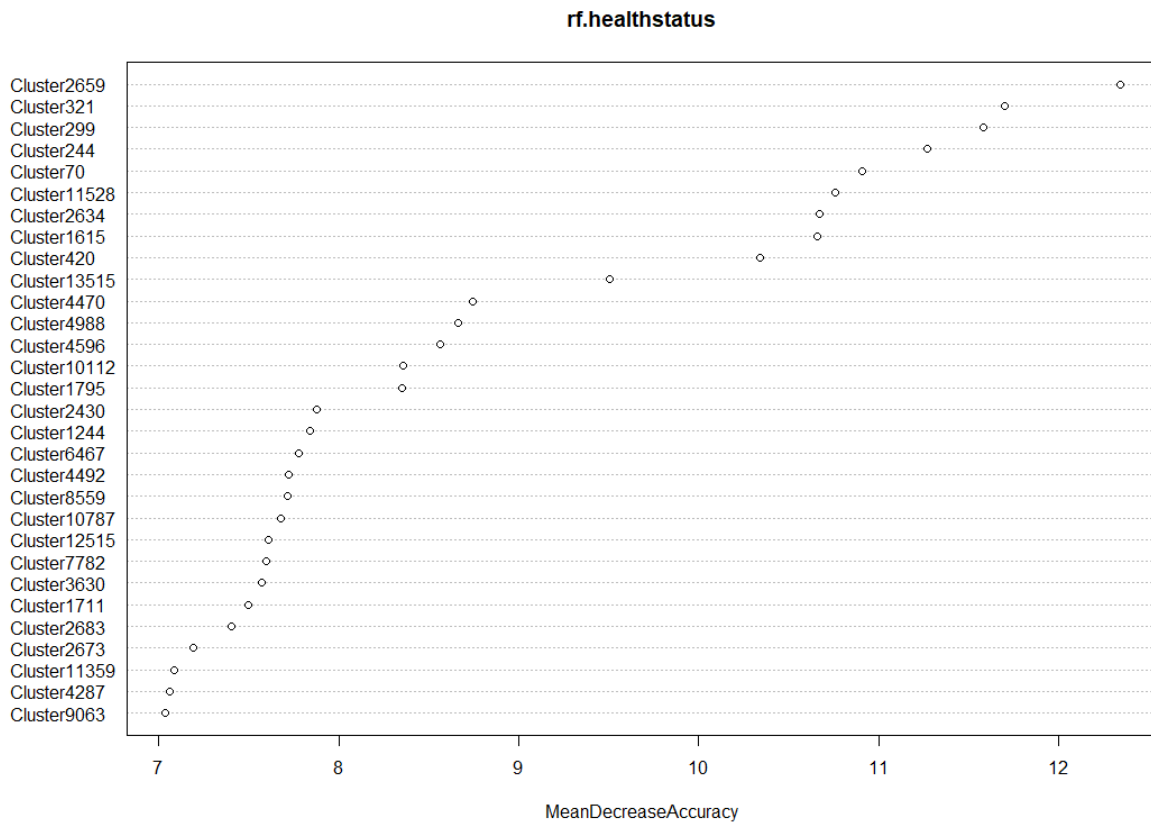
**Figure C.26. Richness and composition of nonpolar metabolites significantly differs among patients during infection and after recovery.**

**(A)** Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou's Evenness) are displayed. Each box-plot is stratified by health status with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. **(B)** Principal coordinates analysis (PCoA) was performed and plotted for cases (green, circles) and follow-ups (purple, squares) based on Bray-Curtis dissimilarity of nonpolar metabolite quantification. The first and second coordinate are displayed with their respective percentage of variance explained.



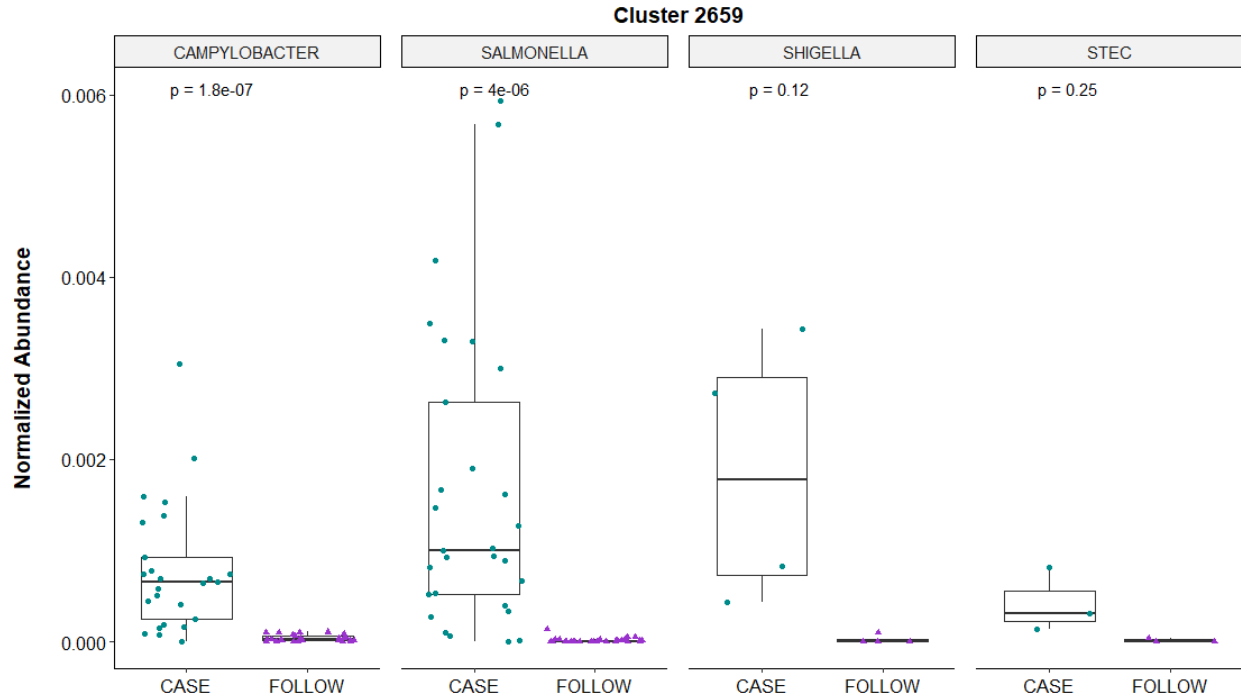
**Figure C.27. Richness and composition of nonpolar metabolites does not appear to be influenced by infecting pathogen.**

(A) Three measures of alpha diversity (Richness, Shannon Diversity, and Pielou's Evenness) are displayed. Each box-plot is stratified by infecting pathogen with samples represented by circles (*Campylobacter*, blue), triangles (*Salmonella*, red), squares (*Shigella*, yellow), or crosses (STEC, violet). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples; these values are indicated above the comparison bar within each boxplot. (B) Principal coordinates analysis (PCoA) was performed and plotted for cases (circles) and follow-ups (squares) based on Bray-Curtis dissimilarity of polar metabolite quantification. Points were colored based on the source of enteric infection (*Campylobacter*, blue; *Salmonella*, red; *Shigella*, yellow; or STEC, violet) The first and second coordinate are displayed with their respective percentage of variance explained.



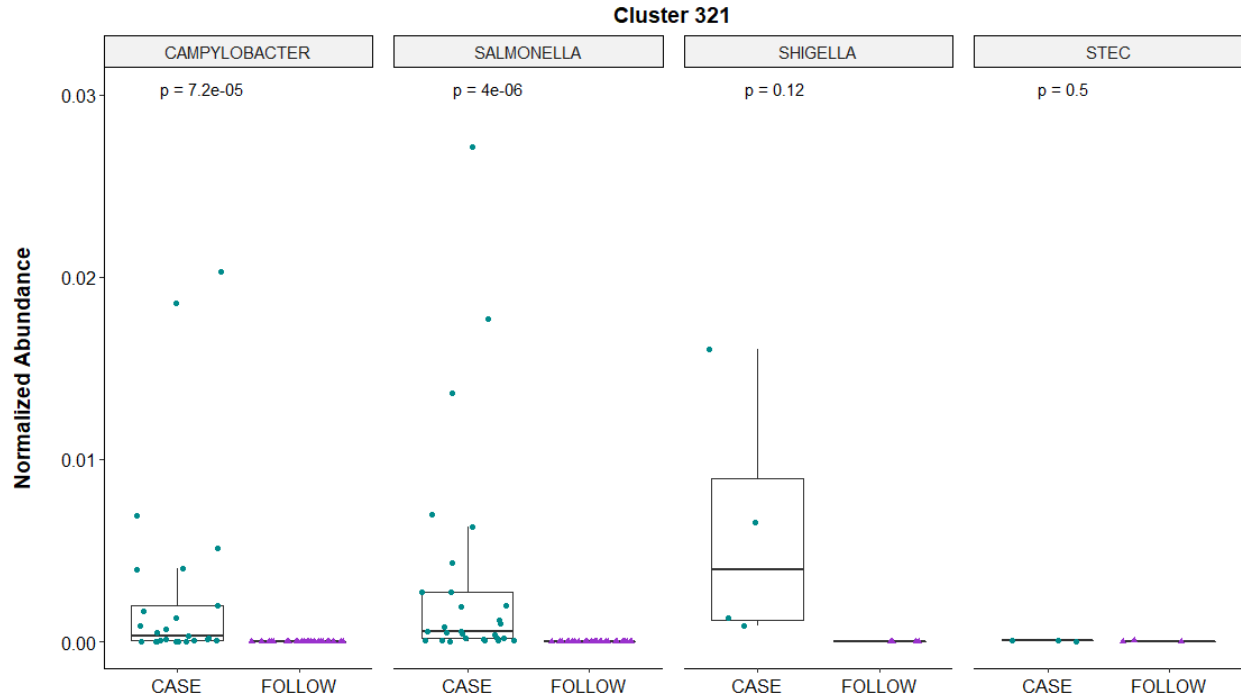
**Figure C.28. Mean decrease in accuracy plot from random forest analysis of nonpolar metabolites.**

Random forest was run on nonpolar metabolite intensities and set to classify samples by health status. The dot plot displays the top-30 clusters (metabolites) which were found to be of highest importance when distinguishing cases and follow-ups. Cluster index is noted on the y-axis, while mean decrease in accuracy is plotted on the x-axis. A higher mean decrease in accuracy metric indicates greater importance in the classifying algorithm.



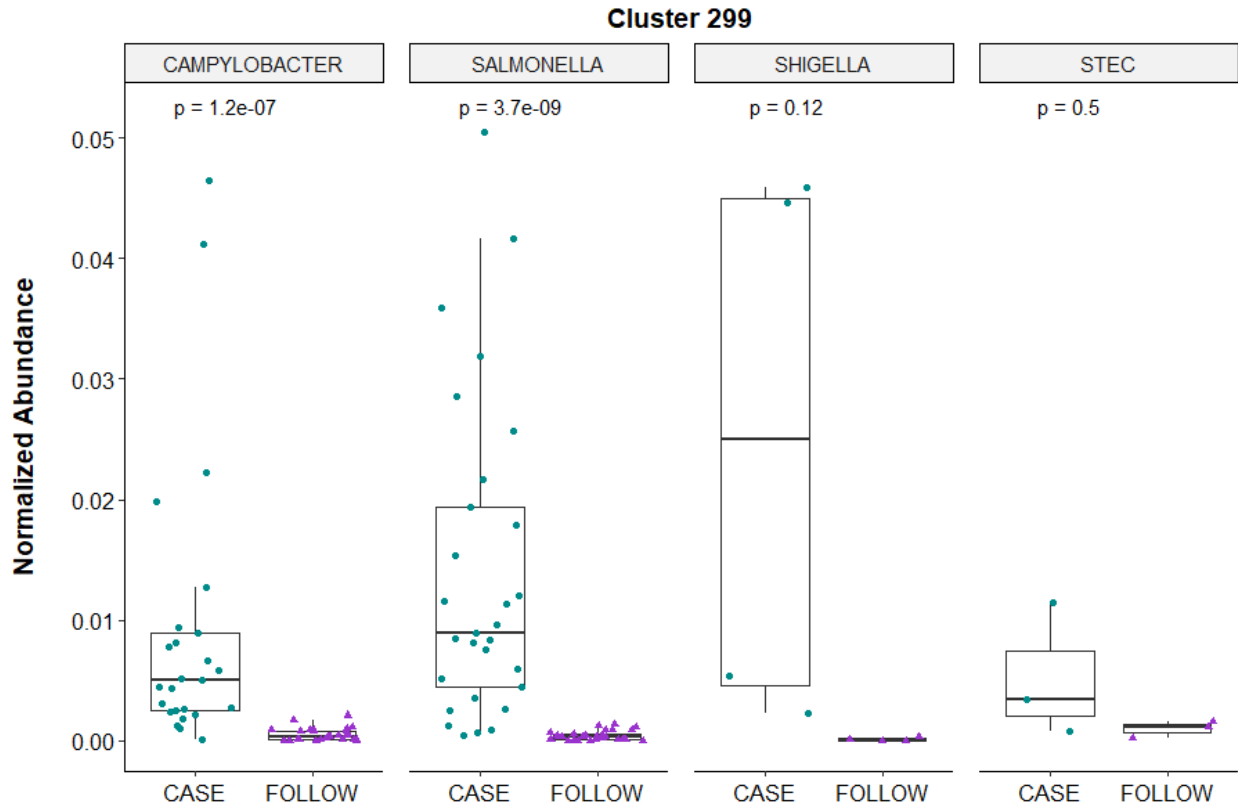
**Figure C.29. Normalized abundance of Cluster 2659 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 2659 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



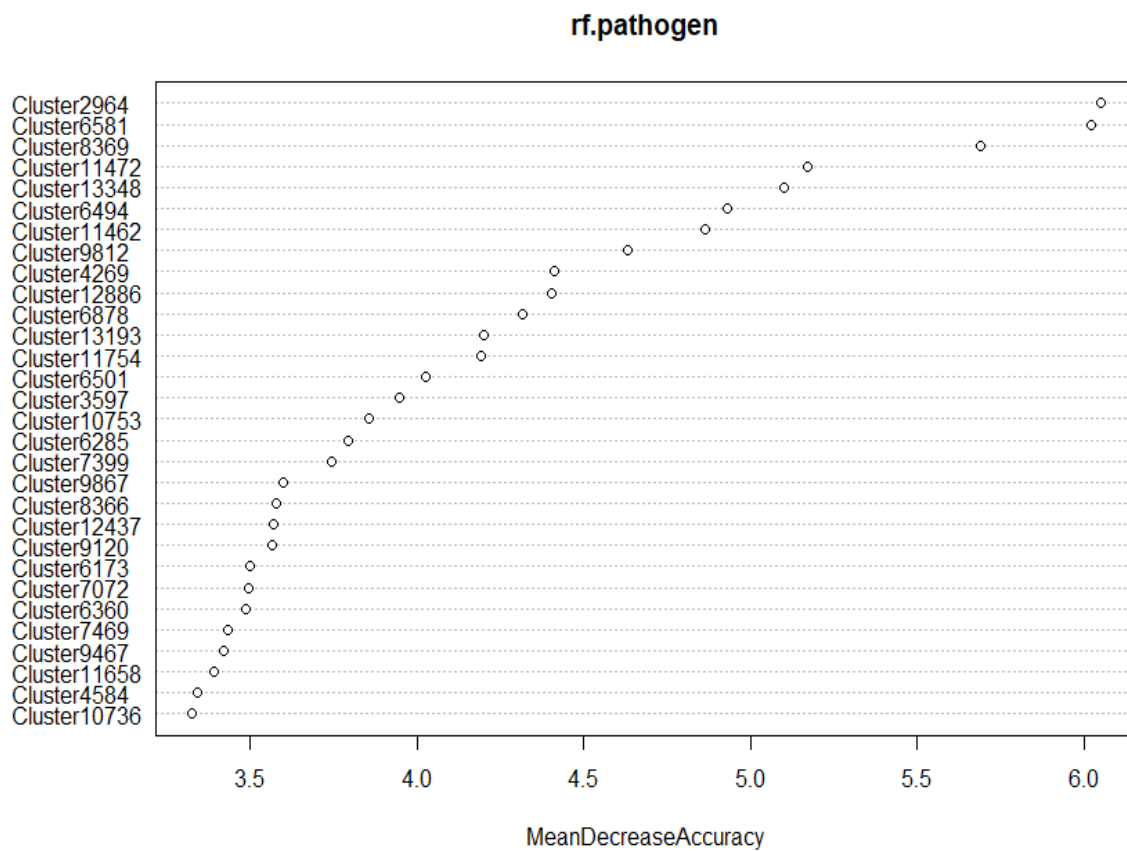
**Figure C.30. Normalized abundance of Cluster 321 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 321 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



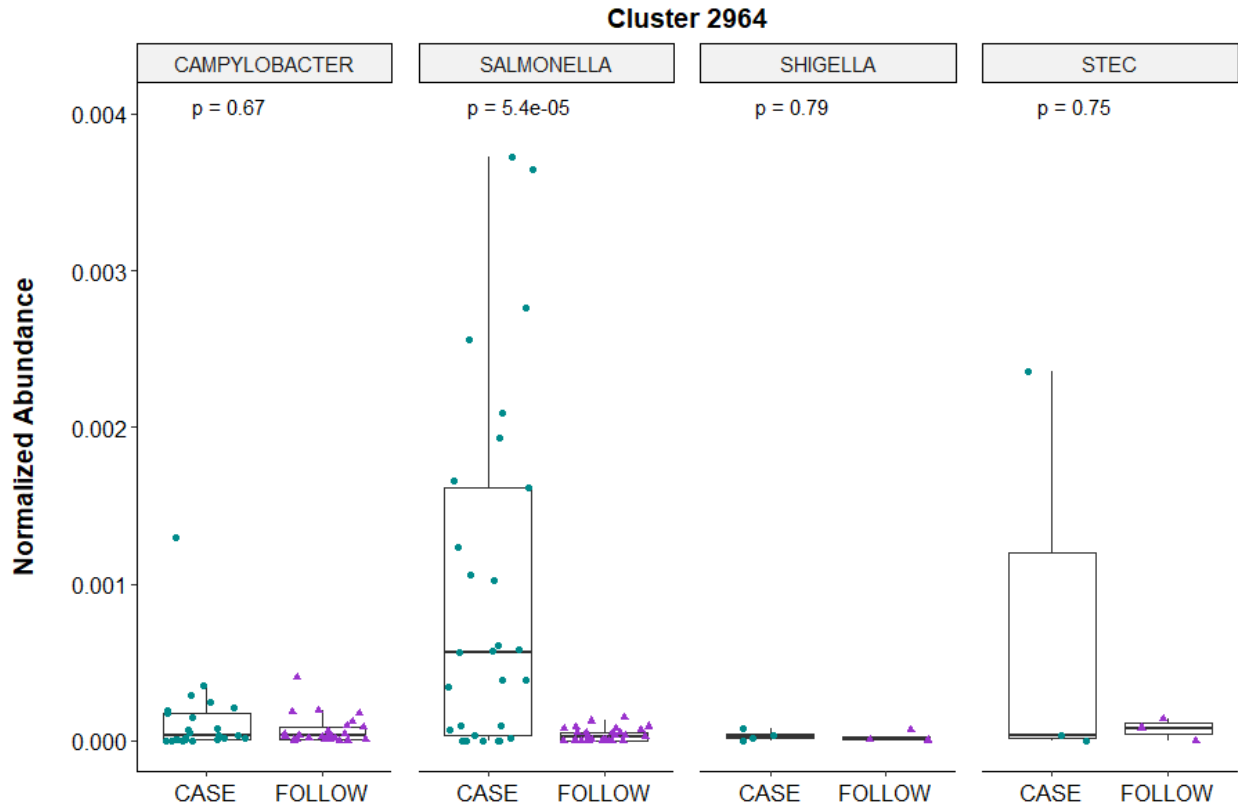
**Figure C.31. Normalized abundance of Cluster 299 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 299 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



**Figure C.32. Mean decrease in accuracy plot from random forest analysis of nonpolar metabolites stratified by infecting pathogen.**

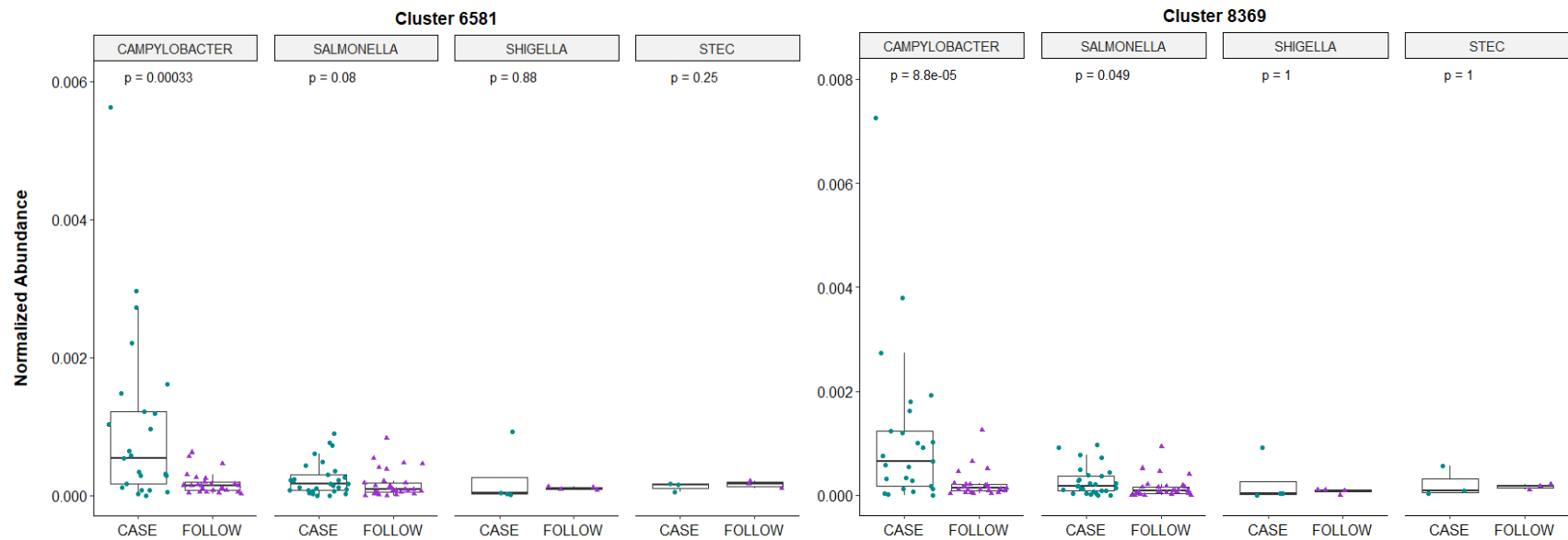
Random forest was run on nonpolar metabolite intensities and set to classify samples by infecting pathogen. The dot plot displays the top-30 clusters (metabolites) which were found to be of highest importance when distinguishing infections among different pathogens. Cluster index is noted on the y-axis, while mean decrease in accuracy is plotted on the x-axis. A higher mean decrease in accuracy metric indicates greater importance in the classifying algorithm.



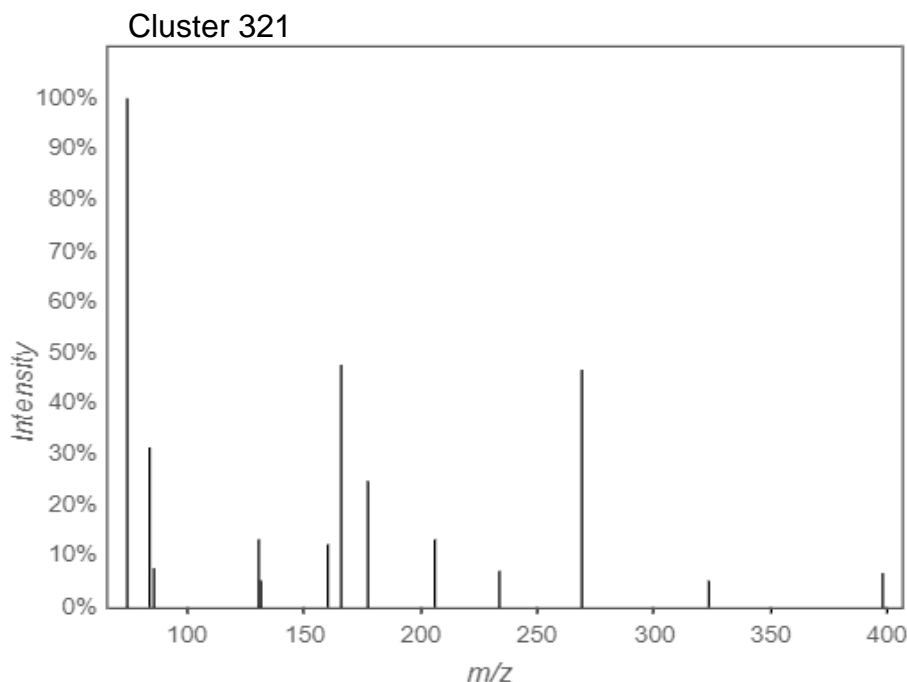
**Figure C.33. Normalized abundance of Cluster 2964 among cases and follow-ups separated by infecting pathogen.**

Normalized abundances of Cluster 2964 are displayed. The box-plot is faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



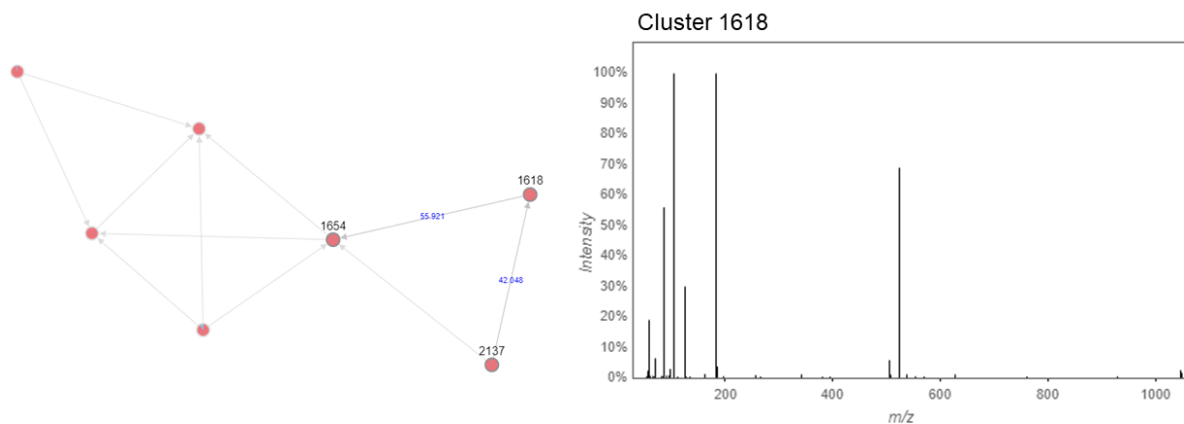


**Figure C.34. Normalized abundance of Clusters 6581 and 8369 among cases and follow-ups separated by infecting pathogen.** Normalized abundances of each cluster are displayed. The box-plots are faceted by infecting pathogen and stratified by health status, with samples represented by circles (cases, green) or triangles (follow-ups, purple). Data points are offset from the vertical to allow for clear interpretation of all samples. Within each box, the median is displayed as the thick black bar; the first and third quartiles are shown by the bottom and the top of each box, respectively. P-values displayed on the plot were calculated using the Wilcoxon signed-rank test for paired samples. P-values are indicated at the top of the plot for case-follow-up comparisons within each infecting pathogen group.



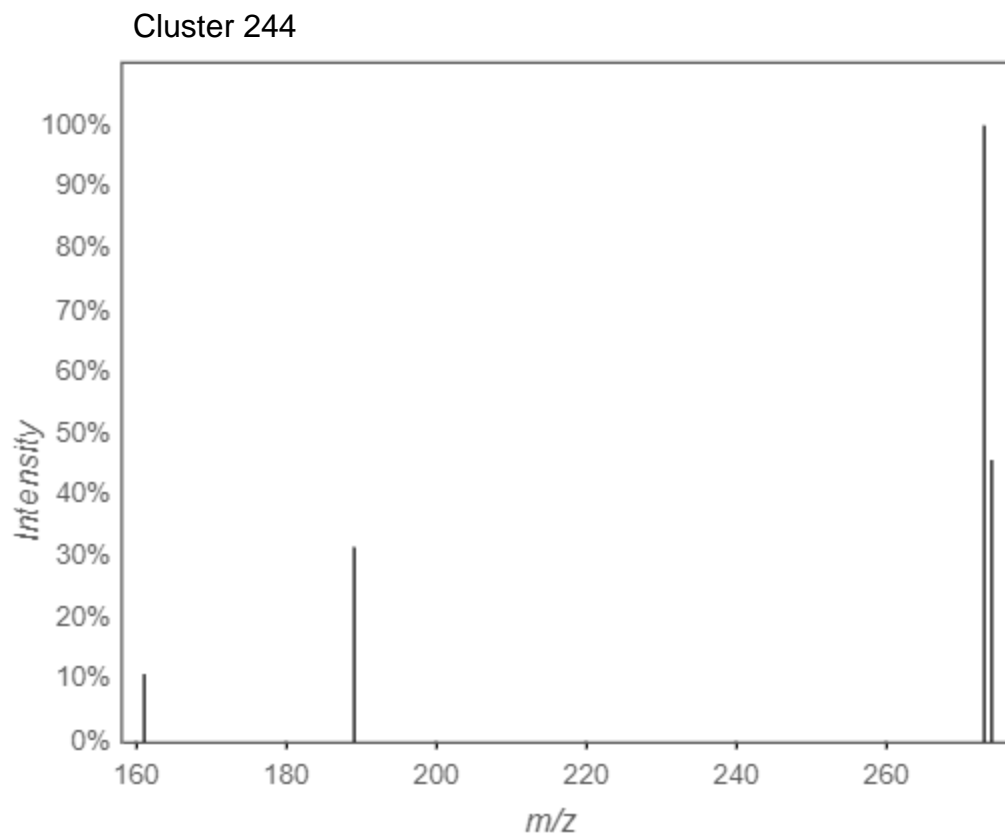
**Figure C.35. MS<sup>2</sup> spectra for Cluster 321, which was a prevalent nonpolar metabolite in cases.**

The MS<sup>2</sup> spectra Cluster 321 is shown. This cluster was a singleton (i.e., not affiliated with a molecular network) and was not assigned an annotation in GNPS.

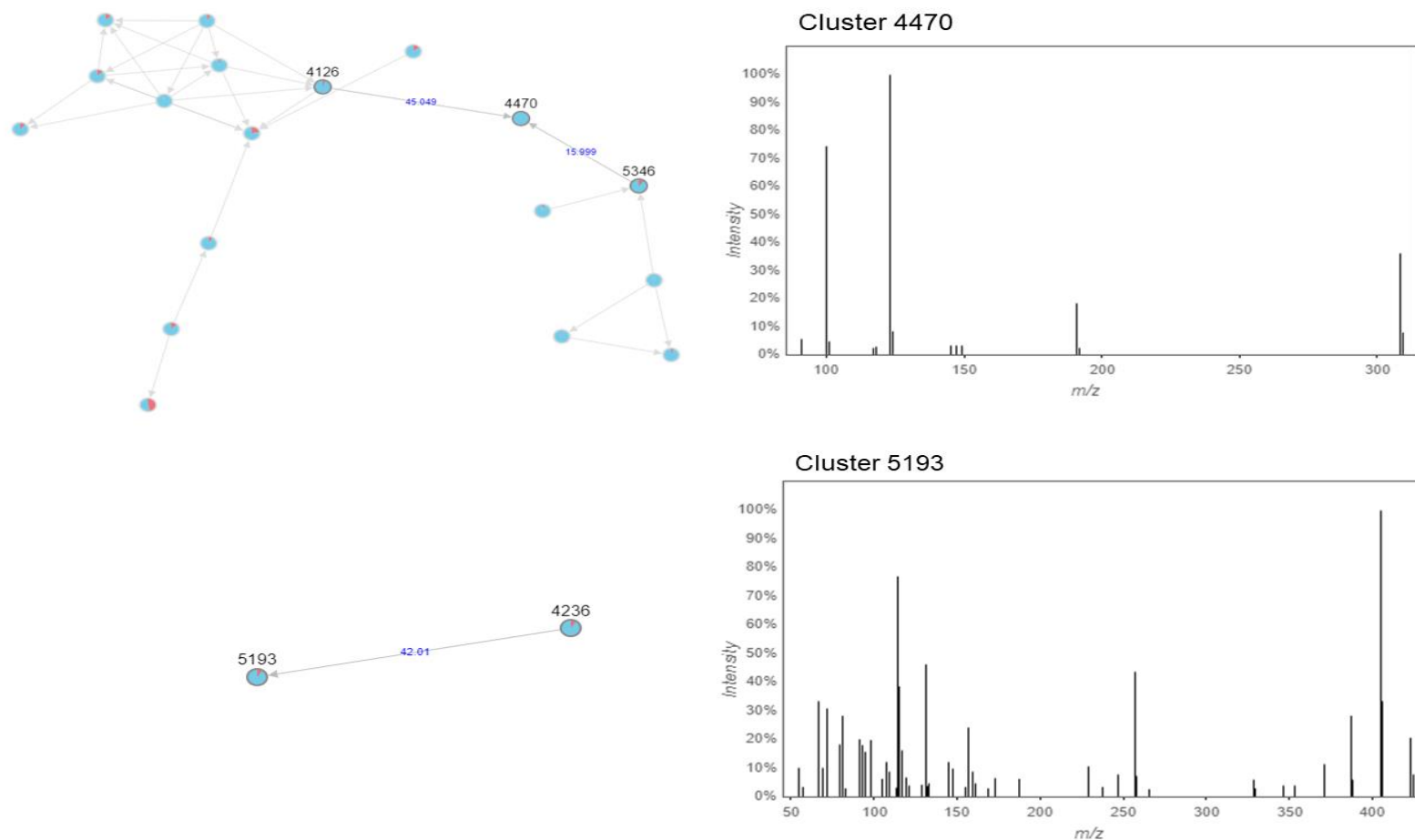


**Figure C.36. Molecular network and MS<sup>2</sup> spectra for Cluster 1618, which was prevalent among cases.**

A molecular network constructed in GNPS (left) shows the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra Cluster 1618 is shown (right).

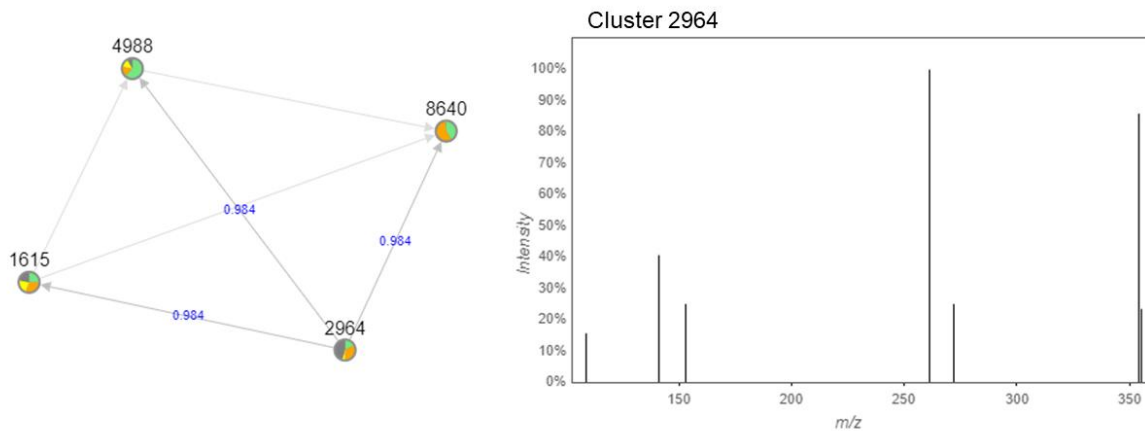


**Figure C.37. MS<sup>2</sup> spectra for Cluster 244, which was prevalent among cases.**  
The MS<sup>2</sup> spectra Cluster 244 is shown. This cluster was a singleton (i.e., not affiliated with a molecular network) and was not assigned an annotation in GNPS.



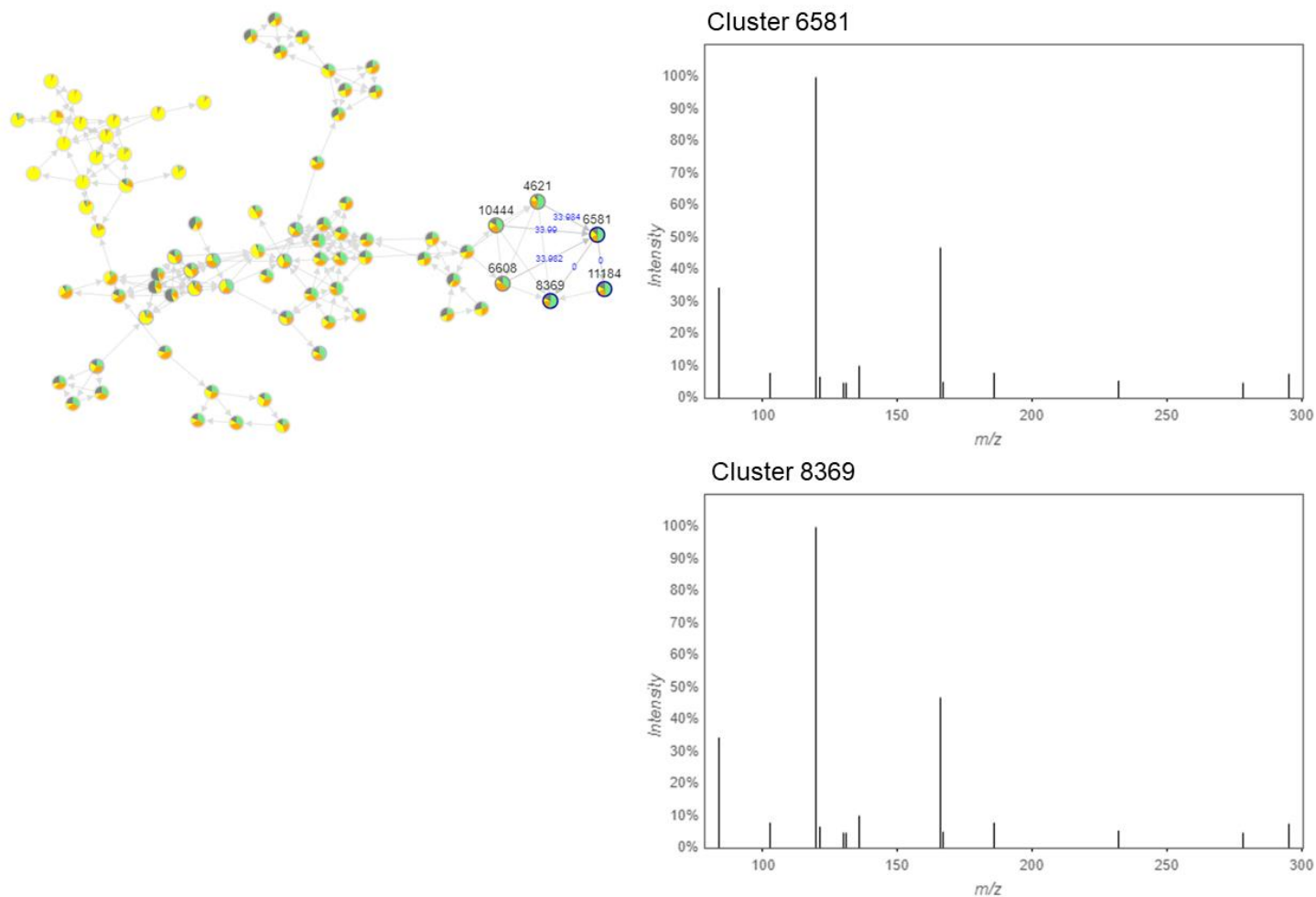
**Figure C.38. Molecular networks and MS<sup>2</sup> spectra for Clusters 4470 and 5193, metabolites found more consistently among follow-ups.**

Molecular networks constructed in GNPS (left) show the interrelatedness of multiple metabolite clusters for Cluster 4470 and 5193. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in cases (red) and follow-ups (blue). The MS<sup>2</sup> spectra for these clusters are shown (right).



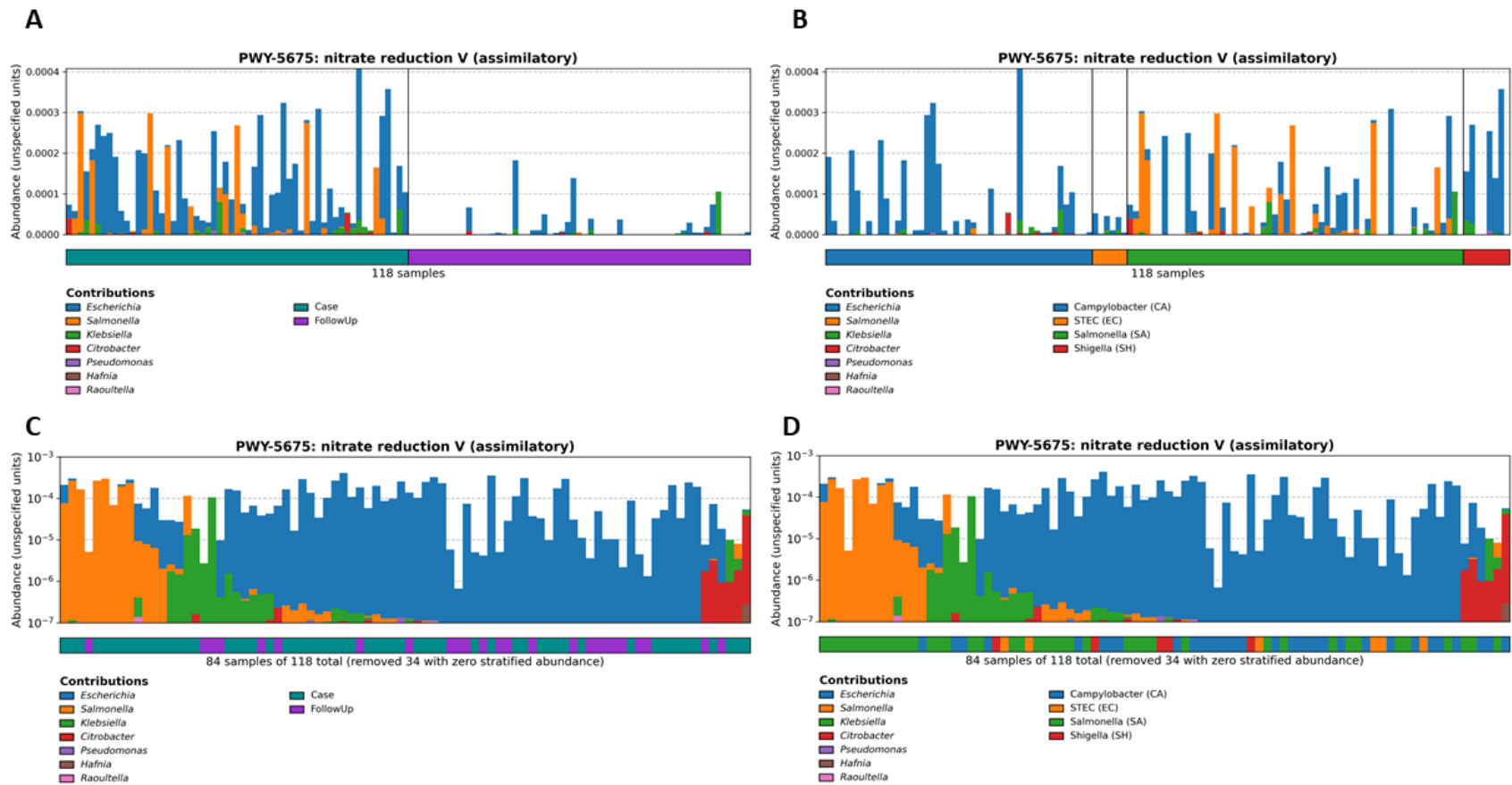
**Figure C.39. The molecular network and MS<sup>2</sup> spectra for Cluster 2964, which was indicated to be more abundant among people with *Salmonella* infection.**

Molecular networks constructed in GNPS (left) show the interrelatedness of multiple metabolite clusters for Cluster 2964. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in patients with different types of infection; *Campylobacter* (green), *Salmonella* (orange), *Shigella* (yellow), or STEC (gray). The MS<sup>2</sup> spectra for Cluster 2964 is also shown (right).



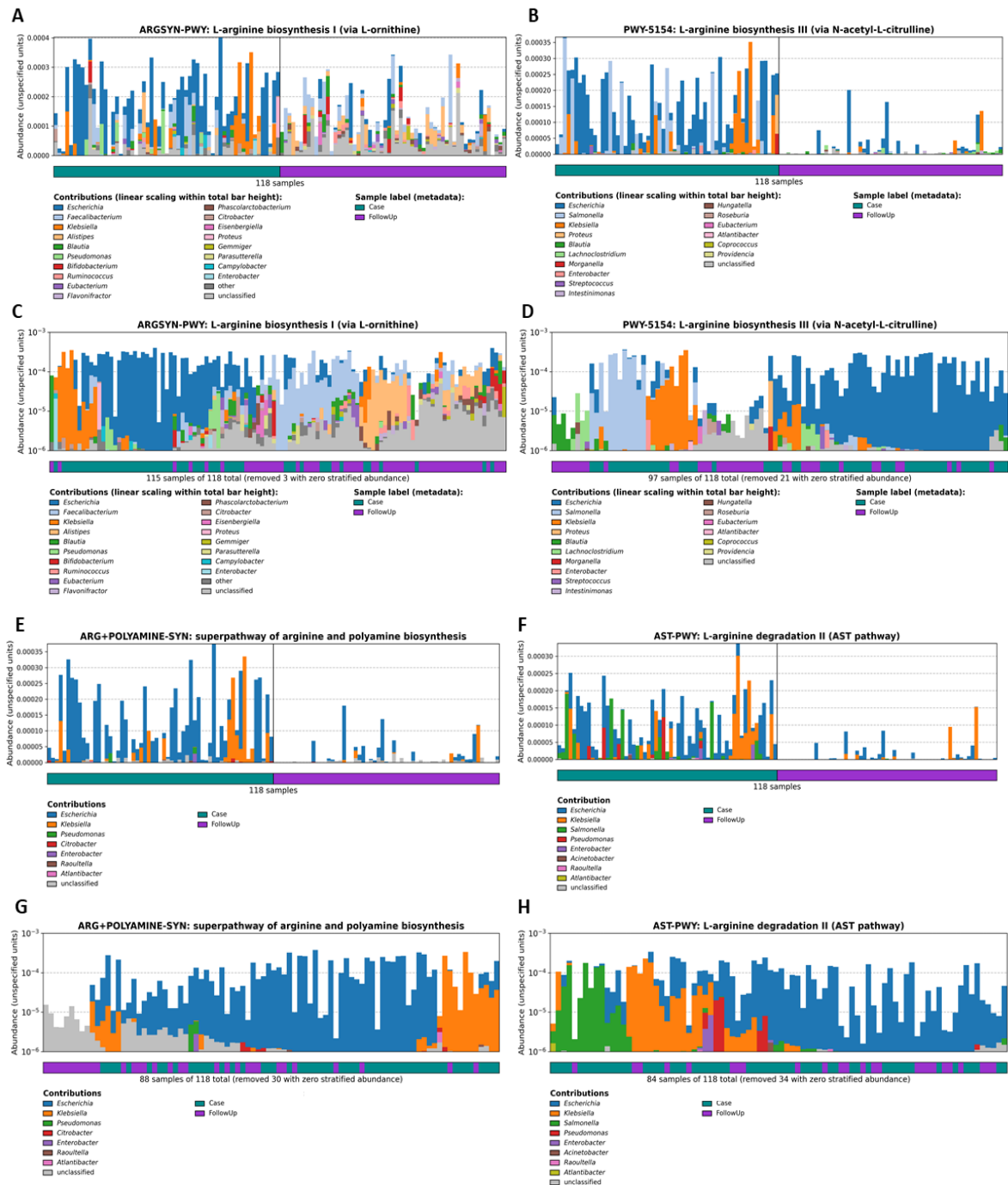
**Figure C.40. The molecular network and MS<sup>2</sup> spectra for Clusters 6581 and 8369, which were indicated to be more abundant among people with *Campylobacter* infection.**

Molecular networks constructed in GNPS (left) show the interrelatedness of multiple metabolite clusters. Nodes are labeled with their cluster index (black) and edges are labeled with the associated mass difference between two connected nodes (blue). Pie-charts on each node indicate the proportion of that node that was found in patients with different types of infection; *Campylobacter* (green), *Salmonella* (orange), *Shigella* (yellow), or STEC (gray). The MS<sup>2</sup> spectra for Clusters 6581 and 8369 are also shown (right).



**Figure C.41. Relative abundances of PWY-5675: nitrate reduction V among infected and recovered patients.**

Barplots show the relative abundance of PWY-5675 calculated by HUMAnN 3.0 stratified by health status (A, C) and infecting pathogen (B, D). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples when stratified by health status; for infecting pathogen, samples are grouped by *Campylobacter* (blue), *Salmonella* (green), *Shigella* (red), or STEC (orange). The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for each pathway attributed to that specific genus. Sample relative abundances were first plotted (A, B) and also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C, D).



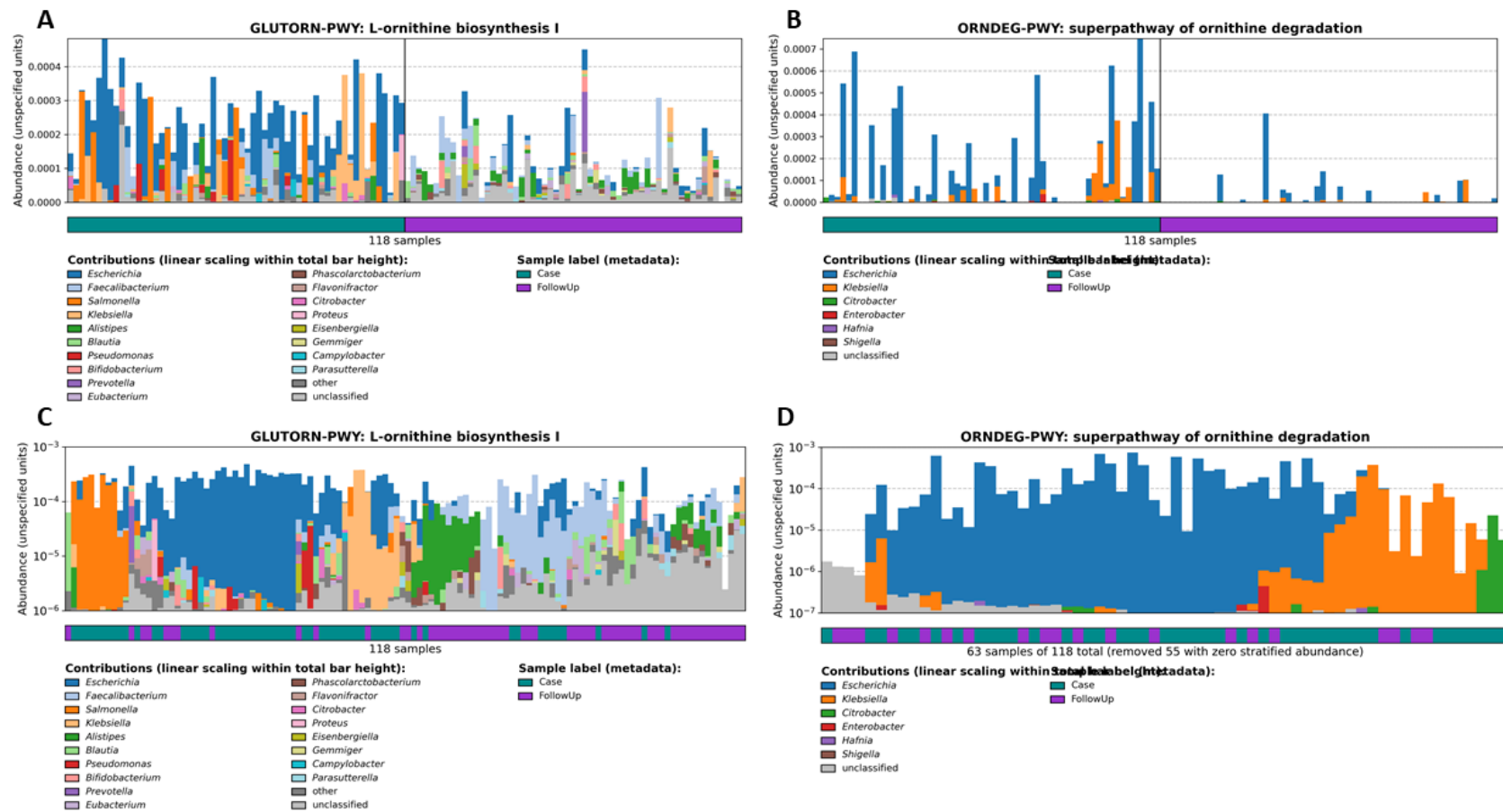
**Figure C.42. Relative abundances of various arginine metabolism pathways among infected and recovered patients.**

Barplots show the relative abundance of four relevant arginine biosynthesis or degradation pathways calculated by HUMAnN 3.0 stratified by health status. The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for each pathway attributed to that specific genus.



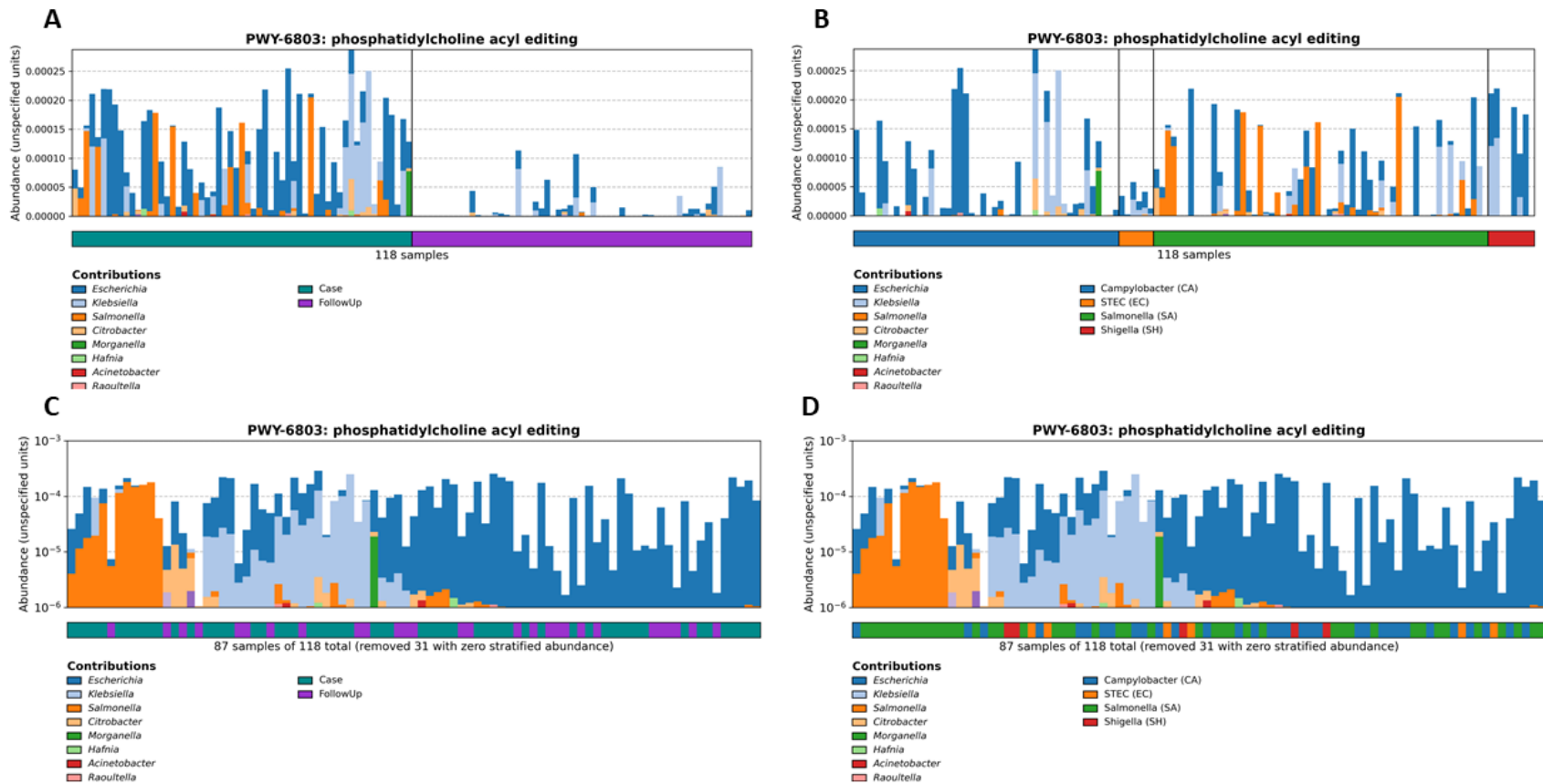
**Figure C.42 (cont'd)**

Sample relative abundances were first plotted (A, B, E, F) and also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C, D, G, H).



**Figure C.43. Relative abundances of various ornithine metabolism pathways among infected and recovered patients.**

Barplots show the relative abundance of two relevant ornithine biosynthesis or degradation pathways calculated by HUMAnN 3.0 stratified by health status. The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples. The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for each pathway attributed to that specific genus. Sample relative abundances were first plotted (A, B) and also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C, D).



**Figure C.44. Relative abundances of PWY-6803: phosphatidylcholine acyl editing among infected and recovered patients.** Barplots show the relative abundance of PWY-6803 calculated by HUMAnN 3.0 stratified by health status (A, C) and infecting pathogen (B, D). The horizontal color bar on the bottom designates case (green) vs. follow-up (purple) samples when stratified by health status; for infecting pathogen, samples are grouped by *Campylobacter* (blue), *Salmonella* (green), *Shigella* (red), or STEC (orange). The ‘Contributions’ section displays genera found to be associated with the pathway of interest; colors in the stacked barplots show the proportion of relative abundances for each pathway attributed to that specific genus. Sample relative abundances were first plotted (A, B) and also clustered by Bray-Curtis dissimilarity to explore clustering relevant to specific genera associations and abundance (C, D).

## **REFERENCES**

## REFERENCES

1. Louis P, Flint HJ. 2017. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology* 19:29-41.
2. Richards JL, Yap YA, Mcleod KH, Mackay CR, Mariño E. 2016. Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases. *Clinical & Translational Immunology* 5:e82.
3. Duncan SH, Holtrop G, Lobley GE, Calder AG, Stewart CS, Flint HJ. 2004. Contribution of acetate to butyrate formation by human faecal bacteria. *British Journal of Nutrition* 91:915-923.
4. Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, Petersen KF, Kibbey RG, Goodman AL, Shulman GI. 2016. Acetate mediates a microbiome–brain– $\beta$ -cell axis to promote metabolic syndrome. *Nature* 534:213-217.
5. Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, Flint HJ, Louis P. 2014. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME Journal* 8:1323-1335.
6. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. 2002. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 52:2141-2146.
7. Wolters M, Ahrens J, Romani-Pérez M, Watkins C, Sanz Y, Benítez-Páez A, Stanton C, Günther K. 2019. Dietary fat, the gut microbiota, and metabolic health – A systematic review conducted within the MyNewGut project. *Clinical Nutrition* 38:2504-2520.
8. Fan Y, Pedersen O. 2021. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology* 19:55-71.
9. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. 2018. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210-215.
10. Jumpertz R, Le DS, Turnbaugh PJ, Trinidad C, Bogardus C, Gordon JI, Krakoff J. 2011. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *The American Journal of Clinical Nutrition* 94:58-65.
11. Roager HM, Hansen LBS, Bahl MI, Frandsen HL, Carvalho V, Gøbel RJ, Dalgaard MD, Plichta DR, Sparholt MH, Vestergaard H, Hansen T, Sicheritz-Pontén T, Nielsen HB, Pedersen O, Lauritzen L, Kristensen M, Gupta R, Licht TR. 2016. Colonic transit time is

- related to bacterial metabolism and mucosal turnover in the gut. *Nature Microbiology* 1:16093.
12. Janssen AWF, Kersten S. 2015. The role of the gut microbiota in metabolic health. *The FASEB Journal* 29:3111-3123.
  13. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto J-M, Kennedy S, Leonard P, Li J, Burgdorf K, Grarup N, Jørgensen T, Brandslund I, Nielsen HB, Juncker AS, Bertalan M, Levenez F, Pons N, Rasmussen S, Sunagawa S, Tap J, Tims S, Zoetendal EG, Brunak S, Clément K, Doré J, Kleerebezem M, Kristiansen K, Renault P, Sicheritz-Ponten T, De Vos WM, Zucker J-D, Raes J, Hansen T, Bork P, Wang J, Ehrlich SD, Pedersen O. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500:541-546.
  14. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, Segal E. 2019. Structural variation in the gut microbiome associates with host health. *Nature* 568:43-48.
  15. Gomaa EZ. 2020. Human gut microbiota/microbiome in health and diseases: a review. *Antonie van Leeuwenhoek* 113:2019-2040.
  16. Lin L, Zhang J. 2017. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC Immunology* 18.
  17. Schippa S, Conte M. 2014. Dysbiotic Events in Gut Microbiota: Impact on Human Health. *Nutrients* 6:5786-5805.
  18. Cândido FG, Valente FX, Grześkowiak ŁM, Moreira APB, Rocha DMUP, Alfenas RDCG. 2018. Impact of dietary fat on gut microbiota and low-grade systemic inflammation: mechanisms and clinical implications on obesity. *International Journal of Food Sciences and Nutrition* 69:125-143.
  19. Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. 2016. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529:212-215.
  20. Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget AL, DelmÉ E, Cousin BA, Sulpice T, Chamontin B, Ferrière J, Tanti J-FO, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin RM. 2007. Metabolic Endotoxemia Initiates Obesity and Insulin Resistance. *Diabetes* 56:1761-1772.
  21. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Doré J. 2009. Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflammatory Bowel Diseases* 15:1183-1189.
  22. Duboc H, Rainteau D, Rajca S, Humbert L, Farabos D, Maubert M, Grondin V, Jouet P, Bouhassira D, Seksik P, Sokol H, Coffin B, Sabate JM. 2012. Increase in fecal primary

- bile acids and dysbiosis in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterology & Motility* 24:513-e247.
23. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne Illness Acquired in the United States -- Major Pathogens. *Emerging Infectious Disease* 17:7-15.
  24. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, Jervis R, Lathrop S, Muse A, Duwell M, Smith K, Tobin-D'Angelo M, Vugia DJ, Zablotsky Kufel J, Wolpert BJ, Tauxe R, Payne DC. 2020. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food -- Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016-2019. *MMWR Morbidity and Mortality Weekly Report* 69:509-514.
  25. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* 3:45-45.
  26. Hansen ZA, Cha W, Nohomovich B, Newton DW, Lephart P, Salimnia H, Khalife W, Shade A, Rudrik JT, Manning SD. 2021. Comparing gut resistome composition among patients with acute *Campylobacter* infections and healthy family members. *Scientific Reports* 11.
  27. Ingram DD, Franco SJ. 2014. 2013 NCHS Urban-rural Classification Scheme for Counties, 166 ed, vol Stat 2. National Center for Health Statistics.
  28. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* 6:3-6.
  29. Nayfach S, Pollard KS. 2015. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biology* 16:51-51.
  30. Beghini F, Mciver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10.
  31. Team RC. 2017. R: A language and environment for statistical computing., *on R Foundation for Statistical Computing*. <https://www.R-project.org/>. Accessed
  32. Ma S. 2021. MMUPHin: Meta-analysis Methods with Uniform Pipeline for Heterogeneity in Microbiome Studies, vol R package version 1.8.0.

33. 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, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, Van Der Hoof JJJ, et al. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nature Methods* 17:905-908.
34. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapon CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, 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 C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* 34:828-837.
35. 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 Bioinformatics* 11:395.
36. Katajamaa M, Miettinen J, Oresic M. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22:634-636.
37. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T. 2010. MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* 45:703-714.
38. Mohimani H, Gurevich A, Shlemov A, Mikheenko A, Korobeynikov A, Cao L, Shcherbin E, Nothias L-F, Dorrestein PC, Pevzner PA. 2018. Dereplication of microbial metabolites through database search of mass spectra. *Nature Communications* 9.
39. Ono K, Demchak B, Ideker T. 2014. Cytoscape tools for the web age: D3.js and Cytoscape.js exporters. *F1000Research* 3:143.
40. 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 Research* 13:2498-2504.
41. Breiman L. 2001. *Machine Learning* 45:5-32.



42. Liaw A, Wiener M. 2002. Classification and Regression by randomForest. *R News* 2:18-22.
43. Pang Z, Chong J, Zhou G, Morais L, De, David Anderson, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J. 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* 49:W388-W396.
44. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K. 2018. Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition* 57:1-24.
45. Kriss M, Hazleton KZ, Nusbacher NM, Martin CG, Lozupone CA. 2018. Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Current Opinion in Microbiology* 44:34-40.
46. Dash NR, Al Bataineh MT. 2021. Metagenomic Analysis of the Gut Microbiome Reveals Enrichment of Menaquinones (Vitamin K2) Pathway in Diabetes Mellitus. *Diabetes & Metabolism Journal* 45:77-85.
47. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host & Microbe* 2:119-129.
48. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra AM, Laughlin RC, Gomez G, Wu J, Lawhon SD, Popova IE, Parikh SJ, Adams LG, Tsois RM, Stewart VJ, Bäumlér AJ. 2013. Host-Derived Nitrate Boosts Growth of *E. coli* in the Inflamed Gut. *Science* 339:708-711.
49. Cynober L. 1994. Can arginine and ornithine support gut functions? *Gut* 35:S42-S45.
50. Leclerc M, Bedu-Ferrari C, Etienne-Mesmin L, Mariadassou M, Lebreuilly L, Tran S-L, Brazeau L, Mayeur C, Delmas J, Rue O, Denis S, Blanquet-Diot S, Ramarao N. 2021. Nitric Oxide Impacts Human Gut Microbiota Diversity and Functionalities. *mSystems* 6.
51. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research* 48:D445-D453.
52. Karl JP, Fu X, Wang X, Zhao Y, Shen J, Zhang C, Wolfe BE, Saltzman E, Zhao L, Booth SL. 2015. Fecal menaquinone profiles of overweight adults are associated with gut microbiota composition during a gut microbiota-targeted dietary intervention. *The American Journal of Clinical Nutrition* 102:84-93.
53. Karl JP, Meydani M, Barnett JB, Vanegas SM, Barger K, Fu X, Goldin B, Kane A, Rasmussen H, Vangay P, Knights D, Jonnalagadda SS, Saltzman E, Roberts SB, Meydani SN, Booth SL. 2017. Fecal concentrations of bacterially derived vitamin K forms are associated with gut microbiota composition but not plasma or fecal cytokine

- concentrations in healthy adults. *The American Journal of Clinical Nutrition* 106:1052-1061.
54. Hassouneh SA-D, Loftus M, Yooseph S. 2021. Linking Inflammatory Bowel Disease Symptoms to Changes in the Gut Microbiome Structure and Function. 12.
  55. Nowicka B, Kruk J. 2010. Occurrence, biosynthesis and function of isoprenoid quinones. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1797:1587-1605.
  56. Zeng MY, Inohara N, Nuñez G. 2017. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunology* 10:18-26.
  57. Pérez-Cobas AE, Artacho A, Knecht H, Ferrús ML, Friedrichs A, Ott SJ, Moya A, Latorre A, Gosalbes MJ. 2013. Differential Effects of Antibiotic Therapy on the Structure and Function of Human Gut Microbiota. *PLoS ONE* 8:e80201.
  58. van der Veen JN, Kennelly JP, Wan S, Vance JE, Vance DE, Jacobs RL. 2017. The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1859:1558-1572.
  59. Goh YQ, Cheam G, Wang Y. 2021. Understanding Choline Bioavailability and Utilization: First Step Toward Personalizing Choline Nutrition. *Journal of Agricultural and Food Chemistry* 69:10774-10789.
  60. Wright AT. 2019. Gut commensals make choline too. *Nature Microbiology* 4:4-5.
  61. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, Didonato JA, Lusis AJ, Hazen SL. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472:57-63.
  62. Dyle MC, Ebert SM, Cook DP, Kunkel SD, Fox DK, Bongers KS, Bullard SA, Dierdorff JM, Adams CM. 2014. Systems-based Discovery of Tomatidine as a Natural Small Molecule Inhibitor of Skeletal Muscle Atrophy. *Journal of Biological Chemistry* 289:14913-14924.
  63. Adams CM, Ebert SM, Dyle MC. 2015. Use of mRNA expression signatures to discover small molecule inhibitors of skeletal muscle atrophy. *Current Opinion in Clinical Nutrition and Metabolic Care* 18:263-268.
  64. Lamontagne Boulet M, Isabelle C, Guay I, Brouillette E, Langlois J-P, Jacques P-É, Rodrigue S, Brzezinski R, Beauregard PB, Bouarab K, Boyapelly K, Boudreault P-L, Marsault É, Malouin F. 2018. Tomatidine Is a Lead Antibiotic Molecule That Targets *Staphylococcus aureus* ATP Synthase Subunit C. *Antimicrobial Agents and Chemotherapy* 62:AAC.02197-17.
  65. Boulanger S, Mitchell G, Bouarab K, Marsault E, Cantin A, Frost EH, Deziel E, Malouin F. 2015. Bactericidal Effect of Tomatidine-Tobramycin Combination against Methicillin-

- Resistant *Staphylococcus aureus* and *Pseudomonas aeruginosa* Is Enhanced by Interspecific Small-Molecular Interactions. *Antimicrobial Agents and Chemotherapy* 59.
66. Barbieri R, Coppo E, Marchese A, Daglia M, Sobarzo-Sanchez E, Fazel Nabavi S, Mohammad Nabavi S. 2017. Phytochemicals for human disease: An update on plant-derived compounds antibacterial activity. *Microbiological Research* 196:44-68.
  67. Guthrie L, Wolfson S, Kelly L. 2019. The human gut chemical landscape predicts microbe-mediated biotransformation of foods and drugs. *eLife* 8.
  68. Lee H-S, Shin HJ, Jang KH, Kim TS, Oh K-B, Shin J. 2005. Cyclic Peptides of the Nocardamine Class from a Marine-Derived Bacterium of the Genus *Streptomyces*. *Journal of Natural Products* 68:623-625.
  69. Shaaban KA, Singh S, Elshahawi SI, Wang X, Ponomareva LV, Sunkara M, Copley GC, Hower JC, Morris AJ, Kharel MK, Thorson JS. 2014. Venturicidin C, a new 20-membered macrolide produced by *Streptomyces* sp. TS-2-2. *The Journal of Antibiotics* 67:223-230.
  70. Bolourian A, Mojtahedi Z. 2018. *Streptomyces*, shared microbiome member of soil and gut, as 'old friends' against colon cancer. *FEMS Microbiology Ecology* 94.
  71. Iser JH, Sali A. 1981. Chenodeoxycholic Acid. *Drugs* 21:90-119.
  72. Ridlon JM, Kang D-J, Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research* 47:241-259.
  73. Ridlon JM, Kang D-J, Hylemon PB. 2010. Isolation and characterization of a bile acid inducible  $7\alpha$ -dehydroxylating operon in *Clostridium hylemonae* TN271. *Anaerobe* 16:137-146.
  74. Johnson CH, Gonzalez FJ. 2012. Challenges and opportunities of metabolomics. *Journal of Cellular Physiology* 227:2975-2981.
  75. Consortium THMP. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207-214.
  76. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohny RP, Li W, de Rinaldis E, Bell JT, Venter JC, Nelson KE, Spector TD, Falchi M. 2019. Interplay between the human gut microbiome and host metabolism. *Nature Communications* 10.

**CHAPTER 5**  
Conclusions and Future Directions

The human gut microbiome is important for shaping human health and is involved in immune system homeostasis, protection against foreign microbes, and production and modification of key metabolites (1-3). However, even healthy gut microbial communities undergo periods of ecological change such as disturbance or invasion (4). In the context of human hosts, microbial invasion in the gut environment is typically associated with entry of foreign pathogens, specifically those that cause enteric disease. Enteric pathogens are responsible for greater than 9.4 million foodborne infections every year (5). Specifically, the CDC has reported consistently high incidence of infections caused by *Campylobacter*, *Salmonella*, *Shigella*, and STEC (6). In addition to their virulence, these pathogens have also been implicated for their ability to harbor and disseminate antimicrobial resistance (7, 8). Indeed, each of the four enteric pathogens included in this study is listed as a serious threat for causing antibiotic resistant infection (9). Considering that the human gut microbiome is a notable reservoir for AMR (10) and these pathogens interact with the resident microbiota during infection, an interesting ecology of AMR spread emerges which warrants further characterization (11). Additionally, it has been demonstrated that host- and microbe-mediated metabolism differs between healthy and diseased states (12, 13). Although these metabolic discrepancies have been widely explored for multiple chronic diseases, our understanding of metabolic fluctuations related to acute enteric infection is scant. Therefore, these studies were undertaken to elucidate community changes in the human gut microbiome, resistome, and metabolome associated with enteric infection.

Overall, the findings presented in this dissertation describe important changes within human gut microbial communities during and after infection. Herein, we examined stools from patients infected with an enteric pathogen (cases) and a subset of follow-up samples from the same patients submitted after they recovered from the initial infection. Additionally, members of

these patients' households also supplied stool samples for comparison. In the first analysis (Chapter 2), which was limited to patients with *Campylobacter* infections, we documented a distinct change in the diversity and composition of ARGs among infected cases compared to healthy controls. Importantly, these cases displayed increased abundance of MDR genes and were dominated by members of *Proteobacteria*, a phylum whose enrichment has previously been connected to gut disruption and inflammation (14, 15). Controls, on the other hand, displayed high relative abundance of ARGs related to tetracyclines and MLS, signatures that are in line with other studies that have explored resistance among healthy individuals (16, 17). Similar findings were reflected in Chapter 3 in which we sought to characterize both the microbiome and resistome composition in infected and recovered gut communities. Infected communities displayed higher resistome diversity; however, taxonomic diversity was significantly lower during infection than after recovery, and cases had an expansion of *Proteobacteria*, specifically among members of *Enterobacteriaceae*. This expansion was consistently observed regardless of the pathogen causing each infection, which is likely due to enhanced inflammation in the gut during infection. Indeed, the abundance decreased in the recovered samples to levels that were similar to those observed in healthy controls. Support for these findings comes from a prior study showing that expansion of *Enterobacteriaceae* is linked to host-mediated inflammatory responses (15). Developing new treatments that limit the overgrowth of specific microbial populations during infection should be explored to decrease the burden of disease, particularly in the most susceptible populations.

Chapter 3 also involved exploration of specific associations between ARGs and relevant taxa among cases and follow-ups. Correlation network analyses revealed notable connections between *Escherichia* and *Shigella* and multiple *mdt* genes, which are relevant to the MdtABC-

ToIC MDR efflux system (18). *Salmonella* registered significant connections to many metal and biocide resistance genes that were only detected in individuals with *Salmonella* infection, suggesting the pathogen was directly associated with expansion of these genes. Host-tracking analysis, which was performed to examine these network associations, demonstrated similar results. Specifically, *Escherichia* comprised the largest portion of ARG-carrying contigs among cases, with nearly 30% of its ARGs being relevant to MDR. Other *Enterobacteriaceae*, such as *Salmonella* and *Klebsiella*, were also primary hosts to ARGs in infected guts. Importantly, the distribution of ACCs among genera did not appear to shift substantially upon recovery from infection, though there were overall fewer ACCs attributed to the top genera compared to cases. These findings suggest that even as the gut microbiome recovers and demonstrates taxonomic shifts, key members of the community maintain their resistance capacity establishing the human gut as a critical reservoir of ARGs. Indeed, the observed abundance of ARGs attributed to members of *Enterobacteriaceae* during and even after infection is alarming, as these pathobionts are capable of transferring resistance to other pathogens in the gut (11). The mobility of resistance genes among *Enterobacteriaceae* is incredibly high (19), and horizontal gene transfer rates are enhanced in the gut during inflammation (20).

While these findings shed some light on compositional trends associated with enteric infection, further characterization of the mechanisms behind these shifts is needed. One method which could help address this need is consideration of different sample types. For example, due to their varied binding affinities and niche habitation, microbes present in the mucosal layer of gastrointestinal tract have been found to differ from those isolated from stool (21) and these microbes may differentially interact with the host and other microbes (22). Additionally, consideration of stool alone does not allow direct investigation of host-mediated changes such as

inflammation during infection. Rather, collection of serum or blood samples enables direct measurement of immune-modulated features such as inflammatory cytokines, which have been found to vary in response to different microbial stimuli (23). Indeed, inclusion of multiple sample types would provide a much more comprehensive view of the ecological changes we have documented during enteric infection.

Another consideration that should be taken into account in future work is the implementation of longitudinal sampling. Although this dissertation has captured relevant information for microbial gut communities during and after infection, the minutiae of recovery cannot be captured. Additionally, approaches that further explore the mobility of ARGs among members of gut microbial communities should be applied. Previous work has demonstrated that certain groups of bacteria, primarily belonging to *Enterobacteriaceae*, are capable of disseminating resistance among different genera through shared plasmids (19, 24). As members of this family are overrepresented during infection, the potential repercussions of increased ARG transmission are concerning. Therefore, employing computational methods that capture plasmids and other relevant MGEs within gut communities is of great importance. Future studies should also consider exploring plasmids and their microbial hosts through laboratory techniques such as the Hi-C method (25) or SMRT sequencing (26).

Following characterization of the resistome and microbiome of paired cases and follow-ups in Chapter 3, we sought to define the functional capacity of these samples during and after infection. Therefore, Chapter 4 included computational prediction of metabolic pathways as well as metabolite analysis via untargeted metabolomics. This investigation revealed that patient stools during enteric infection not only display different functional potential via pathway prediction, but also contain markedly different metabolites than stools collected from the same



patients upon recovery. Enteric infection was associated with more diverse metabolic pathways, including those relevant to enhanced nitrogen and amino acid metabolism. Because infection was associated with an increased abundance of *Escherichia* and other members of *Enterobacteriaceae* (Chapter 3), the observed increase in these metabolic pathways is consistent with previous studies showing a link between inflammation-associated nitrate production and overgrowth of *Enterobacteriaceae* (27). Examination of gut metabolites indicated that recovered patients contained greater metabolic diversity for both polar and nonpolar metabolites. Since computational pathway prediction only considers microbially-derived metabolic capacity, this discrepancy is not unexpected. In fact, the incongruence between these methods suggests that host-derived metabolites are also critically important for human health.

Our exploration of metabolic shifts associated with enteric infection furthers our understanding of these nuanced gut community changes. While a majority of metabolites quantified via untargeted metabolomics are currently unidentifiable due to limited compound annotation, descriptions concerning patterns of intensity, presence/absence, and diversity associated with infection still hold significant meaning. To assist in addressing some of the limitations of this work, future studies should consider integrating multiple ‘omics techniques to more comprehensively characterize the gut environment. Previous studies have combined methods such as metagenomics, metatranscriptomics, and metametabolomics to thoroughly characterize changes in gut microbiota relevant to antibiotic treatment (28) and fatty liver disease (29). Indeed, establishing connections between specific microbial features and metabolite composition may help explain pieces of this complex ecology. Another relevant approach could apply targeted metabolomics of SCFAs to our samples. These compounds are known to benefit the human host through maintenance of metabolic homeostasis (30); additionally, their absence

has been implicated in multiple states of dysbiosis (31). Examining levels of SCFA production, particularly through use of multiple ‘omics techniques, would enhance our understanding of more detailed metabolic alterations taking place during enteric infection.

Altogether, the findings described in this dissertation capture key changes in the human gut resistome, microbiome, and metabolome related to enteric infection. Indeed, these documented shifts substantiate the need for further characterization of microbial responses to perturbations such as those caused by invading pathogens. Given the increased prevalence of antimicrobial resistance genes observed during infection in our study, particular attention should be paid to the ubiquity and transmission of resistance during periods of ecological change. Additionally, greater work is required to define the ecological mechanisms at play during these periods of flux and future studies should incorporate methods that comprehensively measure both microbial- and host-related responses. To be sure, the ecology of enteric infection and its impacts on the human gut microbiome warrant continued study.

## **REFERENCES**

## REFERENCES

1. Wu H-J, Wu E. 2012. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes* 3:4-14.
2. Kamada N, Chen GY, Inohara N, Núñez G. 2013. Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology* 14:685-690.
3. Louis P, Flint HJ. 2017. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology* 19:29-41.
4. Costello EK, Stagaman K, Dethlefsen L, Bohannon BJM, Relman DA. 2012. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science* 336:1255-1262.
5. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, Griffin PM. 2011. Foodborne Illness Acquired in the United States -- Major Pathogens. *Emerging Infectious Disease* 17:7-15.
6. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, Jervis R, Lathrop S, Muse A, Duwell M, Smith K, Tobin-D'Angelo M, Vugia DJ, Zablotsky Kufel J, Wolpert BJ, Tauxe R, Payne DC. 2020. Preliminary Incidence and Trends of Infections with Pathogens Transmitted Commonly Through Food -- Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2016-2019. *MMWR Morbidity and Mortality Weekly Report* 69:509-514.
7. Pickering LK. 2004. Antimicrobial resistance among enteric pathogens. *Seminars in Pediatric Infectious Diseases* 15:71-77.
8. Ballal M. 2016. Chapter 4 - Trends in Antimicrobial Resistance Among Enteric Pathogens: A Global Concern, p 63-92. *In* Kon K, Rai M (ed), *Antibiotic Resistance*. Academic Press.
9. CDC. 2019. *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA.
10. Kim D-W, Cha C-J. 2021. Antibiotic resistome from the One-Health perspective: understanding and controlling antimicrobial resistance transmission. *Experimental & Molecular Medicine* 53:301-309.
11. Wallace MJ, Fishbein SRS, Dantas G. 2020. Antimicrobial resistance in enteric bacteria: current state and next-generation solutions. *Gut Microbes* 12:e1799654.
12. Gomaa EZ. 2020. Human gut microbiota/microbiome in health and diseases: a review. *Antonie van Leeuwenhoek* 113:2019-2040.
13. Lin L, Zhang J. 2017. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC Immunology* 18.

14. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, Zell A, Newton DW, Salimnia H, Lephart P, Sundin D, Khalife W, Britton RA, Rudrik JT, Manning SD. 2015. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* 3:45-45.
15. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host & Microbe* 2:119-129.
16. Feng J, Li B, Jiang X, Yang Y, Wells GF, Zhang T, Li X. 2018. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environmental Microbiology* 20:355-368.
17. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, Pan Y, Li J, Zhu L, Wang X, Meng Z, Zhao F, Liu D, Ma J, Qin N, Xiang C, Xiao Y, Li L, Yang H, Wang J, Yang R, Gao GF, Wang J, Zhu B. 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications* 4.
18. Reygaert WC. 2018. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiology* 4:482-501.
19. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy* 73:1121-1137.
20. Stecher B, Denzler R, Maier L, Bernet F, Sanders MJ, Pickard DJ, Barthel M, Westendorf AM, Krogfelt KA, Walker AW, Ackermann M, Dobrindt U, Thomson NR, Hardt W-D. 2012. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences* 109:1269-1274.
21. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans ADL, de Vos WM. 2002. Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Applied and Environmental Microbiology* 68:3401-3407.
22. Carroll IM, Chang Y-H, Park J, Sartor RB, Ringel Y. 2010. Luminal and mucosal-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Gut Pathogens* 2:19.
23. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, Ter Horst R, Jansen T, Jacobs L, Bonder MJ, Kurilshikov A, Fu J, Joosten LAB, Zhernakova A, Huttenhower C, Wijmenga C, Netea MG, Xavier RJ. 2016. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* 167:1125-1136.e8.
24. Brolund A, Rajer F, Giske CG, Melefors Ö, Titelman E, Sandegren L. 2019. Dynamics of Resistance Plasmids in Extended-Spectrum- $\beta$ -Lactamase-Producing Enterobacteriaceae during Postinfection Colonization. *Antimicrobial Agents and Chemotherapy* 63.

25. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragooczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326:289-293.
26. Beaulaurier J, Zhu S, Deikus G, Mogno I, Zhang XS, Davis-Richardson A, Canepa R, Triplett EW, Faith JJ, Sebra R, Schadt EE, Fang G. 2018. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nature Biotechnology* 36.
27. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra AM, Laughlin RC, Gomez G, Wu J, Lawhon SD, Popova IE, Parikh SJ, Adams LG, Tsohis RM, Stewart VJ, Bäumlér AJ. 2013. Host-Derived Nitrate Boosts Growth of *E. coli* in the Inflamed Gut. *Science* 339:708-711.
28. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, Otto W, Rojo D, Bargiela R, Von Bergen M, Neulinger SC, Däumer C, Heinsen F-A, Latorre A, Barbas C, Seifert J, Dos Santos VM, Ott SJ, Ferrer M, Moya A. 2013. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* 62:1591-1601.
29. Del Chierico F, Nobili V, Vernocchi P, Russo A, De Stefanis C, Gnani D, Furlanello C, Zandonà A, Paci P, Capuani G, Dallapiccola B, Miccheli A, Alisi A, Putignani L. 2017. Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* 65:451-464.
30. den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. 2013. The role of short-chain fatty acids in the interplay between diet, gut, microbiota, and host energy metabolism. *Journal of Lipid Research* 54:2325-2340.
31. Parada Venegas D, De la Fuente MK, Landskron G, Gonzalez MJ, Quera R, Dijkstra G, Harmsen HJM, Faber KN, Hermoso MA. 2019. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Frontiers in Microbiology* 10.