Unraveling the Role of the Human Gut Microbiome in Three Conditions: Colorectal Cancer, Group B *Streptococcus* Colonization, and Multiple Sclerosis

By
Elise S. Cowley

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Microbiology)

at the
UNIVERSITY OF WISCONSIN-MADISON
2024

Date of final oral examination: 12/18/2023
The dissertation is approved by the following members of the Final Oral Committee:
Karthik Anantharaman, Associate Professor, Bacteriology
Andrew Hryckowian, Assistant Professor, Medical Microbiology and Immunology and Department of Medicine (Division of Gastroenterology & Hepatology)
Jeniel Nett, Associate Professor, Medicine and Medical Microbiology and Immunology
Nasia Safdar, Professor, Medicine
Laura Jacques, Associate Professor, Obstetrics and Gynecology

Abstract

The human gut microbiome is widely studied primarily through compositional and descriptive studies. Different human health and disease statuses have been linked to both the gut microbiome composition and function. To understand the role of the human microbiome in human health and disease, in this thesis, I have paired compositional characterizations with functional potential via metabolic reconstruction. I utilized human samples paired with participant metadata to understand the role of the gut microbiome in three different health conditions: colorectal cancer (CRC), multiple sclerosis (MS), and group B *Streptococcus* (GBS) colonization. I investigated the role of bacterial sulfur cycle potential in the human gut and its impact on colorectal cancer (CRC) by describing the scope of the bacterial sulfur cycle in the human gut and associations of particular microbial sulfidogenic pathways with colorectal cancer using ~17,000 bacterial genomes from a cohort of 667 individuals **(chapter 2).** We found bacterial sulfur cycle genes are common in the gut and several genes are associated with CRC. Next, I determined the prevalence of group B *Streptococcus* in the human gut in the general adult population using stool samples, and evaluated dietary and health risk factors for GBS colonization using 754 stool samples from adults in Wisconsin **(chapter 3).** In this work, we found GBS is present in 18% of stool samples and an increased abundance of GBS is associated with decreased dental hygiene and increased frequency of iron consumption. Finally, I investigated the role of bacteria and bacteriophages in relapsing remitting multiple sclerosis (RRMS) by evaluating the bacterial and phage composition and functional landscape of the gut microbiome in 75 individuals with RRMS compared to controls from a community in Wisconsin **(chapter 4).** We found the overall bacterial

and phage composition of the gut similar between controls and participants with RRMS. We found the Semi-phosphorylative Entner-Doudoroff pathway, was less likely to be present in in the assembled genomes of participants with MS on disease modifying therapy than those who are not on disease modifying therapy and tryptophan biosynthesis was more likely to be present in bacterial genomes of individuals with RRMS not on disease modifying therapy compared to controls. By understanding the composition and functional capacity of the microbiome in different disease states using human samples, we can better understand how to combine therapeutic approaches to change the composition and function of the microbiome (Figure 1).

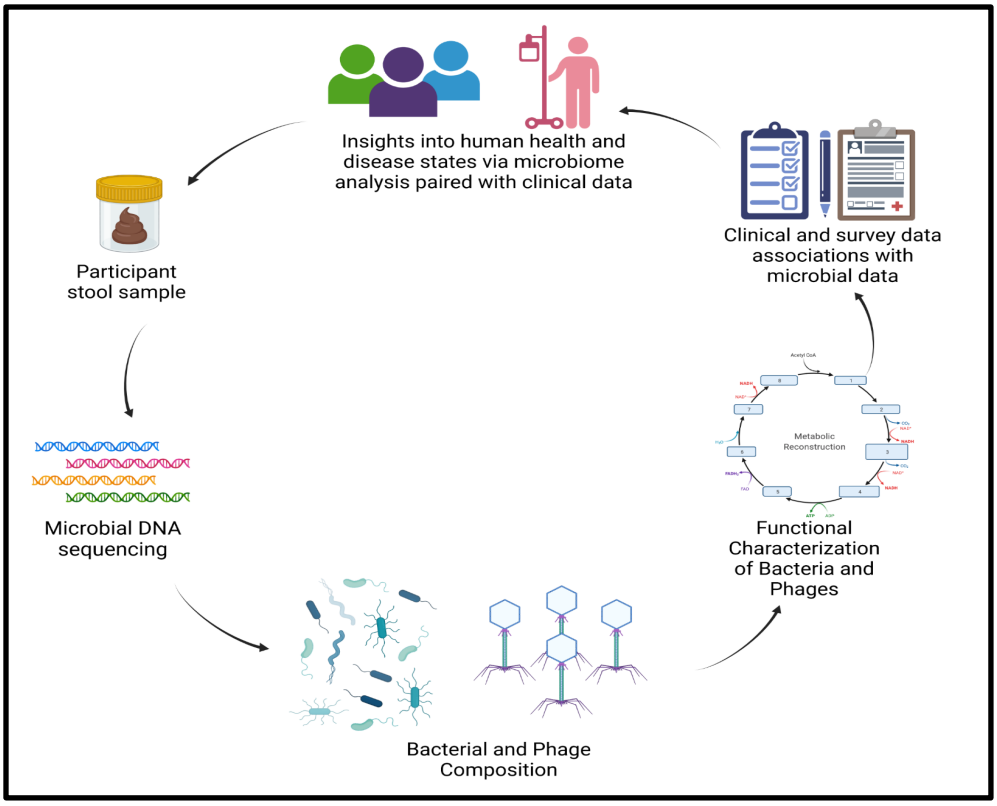


Figure 1. Overview of general approach taken for the projects in this thesis. I will use human stool samples to characterize the gut microbiome composition and functional potential for three different disease states. I paired the microbial analysis with participant metadata in the three disease states to investigate associations between health determinants and the gut microbiome to further our insight into the disease state.

Table of Contents

Acknowledgements

A PhD is a long and difficult journey. I started my PhD in April 2020 so my journey started very isolated and it continued to be that throughout my graduate work, but I was lucky to have many people on my side helping me out and providing bright spots during graduate school.

Thank you to Karthik and Drew for being my PhD advisors, housing me in their labs, and providing me projects to work on. Thank you to all the past and current members of their labs for their helpful discussions and scientific support. In Karthik's lab, I am extremely thankful to Chao, who answered endless bioinformatics questions from me and helped me with debugging code. Patricia Tran provided much needed support during my early days in graduate school and guided me through several bioinformatics obstacles. I am lucky that Patricia also became a friend during graduate school. I am grateful for all our time outside of lab and off campus. In Drew's lab, I want to thank Cate Wilkinson for all the early work she did setting up the Hryckowian lab and the support she provided in getting the Group B *Streptococcus* project off the ground. Thank you to Anna Gregory for her keen scientific mind and being such a source of intentional support and comradery in the Hryckowian lab and for her friendship outside of lab. Anna's presence in the Hryckowian lab is an immensely positive influence and her support during the project I completed in that lab was invaluable. Thank you to Prof. David Hershey who was kind to me in the hallways of the lab building and who I had many enjoyable conversations about gardening and science with. Thank you to Prof.

John Woods for letting me squat in one of his office spaces for my entire graduate school career.

I would not be here without the support staff and administrators in the building, my graduate program, and fellowship program. Thank you to Louise Pape, Terra Theim, Kari Strauss, Kim Besta, Janet Newlands, Chelsea Hanewell, Nichole Monzon, Janna Boehm, Kyle Boldon, and Jacob Schoville.

Thank you to Drs. Nasia Safdar and Jeniel Nett, two of my committee members, who provided clinical perspectives to my research. I am appreciative that Jeniel became an associate director of the MD/PhD program during my time in graduate school and she has been a much-needed voice in that space.

Thank you to Dr. Laura Jacques, another committee member, who I was fortunate to have met during clinical rotations prior to starting my graduate work and she showed me what being an excellent physician meant. I did a whole separate PhD's worth of work on medical education and reproductive healthcare with Dr. Jacques and I am beyond grateful for her steadfast support and mentorship the last four years. She taught me what it meant to be a physician researcher and she always took my ideas seriously. She inspired me with her ability to navigate obstacles and resistance to our research and accomplish a great deal. Besides her unparalleled ability to do excellent research with minimal resources, she was always kind and generous to me. Our weekly meetings were often a highlight of my week. I am grateful to the team that Dr. Jacques built

around her including Amanda Wildenberg, Chelsea Gatica, Cassidy Slinger, Lisa Scott, Amy Godecker, Corinne Hale, Dr. Abby Cutler, Taryn Valley, and all of our collaborators at other institutions. What an extraordinary group. They reminded me how fun research could be and we did work I am truly proud of.

I have been lucky to have many great scientific mentors throughout the years and I would not be here without them. Mrs. Moritz was my first-grade teacher and first got me interested in science. We raised monarch butterflies in our classroom and released them at the river. I still think of how incredible that was. When I went to high school, we were lucky to have an outstanding lineup of science teachers who continued to drive my interest in science, including Mrs. Doyle (biology and anatomy), Mr. Sans (biology II), Mr. Blattman (physics), and Mrs. McGibbon (chemistry). I took three chemistry classes with Mrs. McGibbon during high school and she became a mentor to me, inspiring me to major in chemistry in college. I talked to her about every major life and career decision after high school. Mrs. McGibbon became a very dear friend to me throughout the years and she made time for me when I was back home and we kept in touch via email and texts. She taught me the importance of standing up for what you believe in and speaking your mind and the importance of knowing when to quit. She died this August and it is a tremendous loss for the world. I wish she could have been here to see me get my PhD and I am grateful for the role she played in getting me here. The lessons Mrs. McGibbon taught me will live on and I hope I can instill them in others. At Oregon State University (OSU), Wanda Crannell was my academic advisor and she guided me through my finding my first research experience and getting involved outside of

coursework. I got my first job at OSU as a lab tech in Steve Perakis' forest ecosystems lab. His lab manager and staff scientist, Christina Catricala, hired me, and what an opportunity and stroke of luck that was for me. Chris is one of the best scientist I have ever worked with and the best lab manager I ever worked with. She taught me about the nuts and bolts of lab work and many of the skills she taught me, I have tried to pass down to other labs I worked in. I did undergraduate research with Prof. Kerry McPhail and Chris Thornburg, her graduate student at the time. They both taught me an immense amount about research and guided me through a very fun and interesting undergraduate thesis project. I did a summer internship with Prof. Dianne Newman at Caltech that turned into a job after I graduated from OSU. Dianne is an extraordinary scientist and person. She gave me a lot of independence when I worked for her and a project to work on that changed the trajectory of my career. I worked on a project that put me in direct contact with patients and healthcare workers. Dianne supported my career decisions and helped me get where I am today. She is an excellent mentor and I learned a lot about doing rigorous and interesting science while working for her. Dianne went out of her way to provide opportunities to me such as traveling to Denmark and being a teaching assistant for the Microbial Diversity course at the Marine Biological Laboratories. Through the teaching assistant position at MBL, I met several scientific collaborators, mentors, and friends that I am grateful to for their support especially Srijak Bhatnagar, Emil Ruff, and Kurt Hanselmann.

One of the best decisions I made during my PhD was to volunteer regular and I did so at the Goodman Community Center. I started out driving meals to seniors during the

pandemic and when senior lunch transitioned back in person in 2021, I became a regular Monday first shift volunteer. Besides my meetings with Dr. Jacques, my regular volunteer shift was a weekly highlight. It was fulfilling and engaging. I am appreciative to all the senior participants. I enjoyed getting to know many of them throughout the years and to hear their stories. I worked with extraordinary people including our staff supervisor, Abby Sibilski, and her boss, Gayle Laszewski, both of whom showed me what it means to be excellent at your job and to enjoy it and go the extra mile. I learned so much from the other volunteers that I worked with, Jane Ippolito, Terri Anacker, and Edith Hilliard, and they are all so kind and funny, making for the best senior lunch crew! The staff, volunteers, and seniors made the start of my week fun and meaningful.

I want to thank my Medical Scientist Training Program (MSTP) cohort that I started with in 2017. In particular, I want to thank Patrick Carney for his friendship during the height of the pandemic. I want to thank Josh Choe for asking me to help him with his Eagle Heights Community Garden plot, which led to us getting a second plot. The garden was a source of joy for me and scheming about how many plants we could squish in the plots was a highlight of my graduate school years. Josh also got me very into flower planting and we enabled each other to buy an overwhelming number of flowers to plant each season. Josh taught me a lot over the years while we gardened together, but I am most appreciative that we became better friends through the garden.

Besides the garden, I was fortunate to spend a great deal of time outside during graduate school and I am appreciative for how restorative that time was. I spent a lot of

time at county parks in Madison and state parks around the state. I started bird watching and found great joy in the hawks' nest across from my lab building in Bock and the sandhill cranes in the garden and around town always made me happy. The outdoors and wildlife were very important to me during graduate school and I encourage everyone in graduate school to try to find sources of joy outside of their academic work.

I am grateful to my friends I have made here in Madison and who have helped support me through my PhD. Susan Hromada was the first friend I made in Madison and I am grateful for all of our adventures and misadventures. I enjoyed my many night walks with Sophie Shogren and trying new foods including the infamous key lime pie drink. Anna, Alkas, Ari, and Aslan became my core community in Madison. Thank you for all the dinners, conversations, and welcoming me into your family. Hailey Bussan became one of my best friends during graduate school. I am grateful for all our conversations, errand hangouts, and journeys through construction across the isthmus. Hailey is an outstanding scientist and we had many thought provoking research conversations. Hailey and Anna Gregory were so supportive during graduate school and everyone who goes to graduate school needs to find the friends that will make the journey bearable.

Outside of Madison, I am grateful to my friends who supported me through graduate school. Thank you to Dixie Richards Barry, who has been my best friend since kindergarten and despite not having lived in the same place since high school, we have remained very close. I am grateful for all our conversations and how often Dixie provided a perspective I hadn't thought about. I enjoyed being able to share about

gardening and processing produce with Dixie the last couple of years as we gardened in different places. Thank you to Courtney Van Den Elzen for befriending me all those years ago in Boulder. I am so glad our friendship has weathered being long distance and that we were able to support each other through graduate school. Thank you for the many laughs and thoughtful conversations. I want to thank David Bryan for tolerating me, most of the time, throughout graduate school, and the many years before.

Most importantly, I want to thank my family. Thank you to my extended family who have all supported me from across the country. After years of learning about subjective and objective, I can say for certain that my family is objectively the best. We are loud, argumentative, funny, chaotic yet also very well organized and planned, excellent cooks, hard workers, total shitshows, and I would not want to be a part of any other family. While my years in graduate school were difficult, one good part about graduate school was that I was able to visit my family more than I was able to during my first three years in the MSTP. It was invaluable to get to spend extended periods of time with my family. I was fortunate to be born with two built in best friends, my brother, Mark, and my sister, Anita. They have kept me humble over the years and are a blast to hang out with. My siblings are kind, intelligent, hardworking, and hilarious. All qualities they inherited from our parents, Jeannie and Stan. I am grateful to Lindsey, Zane, and Jory for being such a great addition to our family. There is no one I need to thank more than my parents. It took moving away from home to realize that I have the absolute best parents. Our parents told us we could do anything when we were kids and they supported us through everything we tried to do. Everything I have achieved and

everywhere I have ever been is thanks to my parents and the confidence and work ethic they instilled in us when we were kids. I would not be here, earning a PhD and an MD, without the tremendous support and love of my parents. I have been in academic for many years, and I have been at many institutions around many well-regarded intellectuals and I can say, without a doubt, that my parents are the two most intelligent and hardworking people I have ever met. I am so lucky to have them as parents. Thank you, mom and dad, I love you, and this PhD is for you.

Chapter 1: Introduction - The human gut microbiome and health

Overall Background and Significance

The human body is a complex ecosystem with trillions of microorganisms collectively known as the microbiome both inside and on the body. The human gut, primarily the colon, contains the highest density and diversity of microbes of all the human body associated microbiomes with 100 billion microbial cells per gram of stool [1–4]. The gut microbiome orchestrates a symphony of interactions that profoundly impact our health.

The human gut microbiome is diverse and readily studied via stool samples. Work has been done to characterize the microbial composition of the human gut in different populations and disease states via microbial DNA extracted from stool samples. Microbial communities consist of bacteria, archaea, eukaryotes, and viruses, but the majority of human gut microbiome work has focused on bacterial composition. Bacteria play a substantial role in the human gut including producing certain vitamins, digesting carbohydrates that human cells cannot digest, transforming bile, and they play a role in human immune system development [5].

Diversity and Abundance

The human gut microbiome is diverse, harboring trillions of microorganisms that represent over 1,000 different species. This microbial diversity is essential for overall health, as each species contributes unique functions that collectively support human well-being [6].

Impact on Digestion and Nutrient Absorption

One of the primary roles of the gut microbiome is to aid in digestion and nutrient absorption for the human host. Gut microbes break down complex carbohydrates and fibers that humans cannot digest, providing essential nutrients such as vitamins and short-chain fatty acids (SCFAs). SCFAs serve as energy sources for gut cells and play a vital role in regulating metabolism and inflammation [5,7].

Immune System Modulation

The gut microbiome interacts closely with the immune system, influencing its development and function. The gut microbiome helps train the immune system to distinguish between beneficial and harmful microbes, preventing the body from attacking its own tissues. A diverse and balanced gut microbiome promotes immune tolerance and can reduce the risk of autoimmune diseases [8,9].

The Effects of Diet and Exercise on the Gut Microbiome

Exercise has been shown to impact the gut microbiome by promoting high diversity and decreasing transit time for stool. Some studies have shown exercise can promote so called “beneficial” species in the gut microbiome [10,11]. In addition to exercise, diet has been shown to be a major driver in the composition and functionality of the gut microbiome. A diet rich in fruits, vegetables, and whole grains promotes a diverse and balanced gut microbiome, while a diet high in processed foods, sugary drinks, and artificial sweeteners can lead to an imbalance, potentially increasing the risk of disease. Several studies have also demonstrated that temporary and long term dietary shifts alter the gut microbiome [12–16].

The Gut Microbiome and Human Health

The human gut microbiome is an intricate ecosystem that plays a profound role in human health. Its impact extends beyond digestion, influencing the immune system, mental health, and overall well-being. The gut microbiome has been associated with not only gastrointestinal diseases, but also many diseases that are not traditionally thought of to have an intestinal component. The breadth of diseases, including cardiovascular, neurologic, metabolic, autoimmune, infectious, and gastrointestinal, associated with the gut microbiome demonstrates the effects of the microbial community go beyond the intestinal walls [1,3–5,8,17–20].

The Human Virome

The human virome consists of all viral components of the human microbiome including eukaryotic viruses and bacteriophages (phages). Phages comprise the majority of the human virome [21]. Phages are viruses that infect bacteria and can drive bacterial ecology by conferring functional genes into the bacterial genome in the case of temperate/lysogenic phages or causing cell death in the case of lytic phages. Both lytic and lysogenic phages have been shown to play a role in human health and disease. The overall virome landscape of the human microbiome remains an understudied area, however, it has been shown that the human gut virome is highly stable within individuals, similar to the gut bacteriome and varies between healthy and diseased individuals [22–26].

Gap in Human Gut Microbiome Research

While there are many studies on the human gut microbiome, most have focused on the bacterial composition with little work done on other microbial members and the overall functional capacity of the microbiome. To understand the role of the human microbiome in human health and disease, in this thesis, I have examined understudied drivers of microbiome composition and function. I paired compositional characterizations with functional potential via metabolic reconstruction to study the role of the human microbiome, including bacteria and phages, in human health and disease. I used human stool samples with participant metadata to understand the microbiome in three health conditions: colorectal cancer (CRC), Group B *Streptococcus* (GBS) colonization, and multiple sclerosis. In **chapter 2***,* I performed metabolic reconstructions of the gut microbiomes of an international cohort of participants with and without CRC to describe the scope of the bacterial sulfur cycle in the human gut and associations of particular microbial sulfidogenic pathways with colorectal cancer. In **chapter 3,** I determined the abundance and prevalence of GBS in the gut of the general adult population in Wisconsin and investigated host and microbiome factors that influence GBS gut colonization. Finally in **chapter 4,** I began to compare the gut microbiomes of patients with and without multiple sclerosis in one community in Wisconsin to understand the microbiome composition and functional landscape of the gut microbiome in the relapsing remitting phenotype of multiple sclerosis.

The Human Gut Microbiome and the Wisconsin Idea

In 1904, the current University of Wisconsin-Madison President, Charles Van Hise said that he would, “never be content until the beneficent influence of the university reaches every family in the state", which summarizes the Wisconsin Idea [27,28]. The Wisconsin Idea is that the university should benefit the entire state and people of Wisconsin and reach beyond the bounds of the campus. During my time at UW-Madison, I have come to truly believe in the Wisconsin Idea and having the work done at the university benefit the people across the state and shared with the people in the state. While my thesis contains three very different projects investigating different human diseases, the work embodies the Wisconsin Idea. I took skills that I learned during my dissertation to answer questions and perform research with samples from participants within the state of Wisconsin. In my initial project, **chapter 2,** investigating the role of the bacterial sulfur cycle in the human gut in CRC pathogenesis, I relied on publicly available data from an international cohort of participants. I build a computational skillset using these data. In my next project, **chapter 3,** investigating the prevalence of GBS in the human gut and host and microbial factors that can influence the frequency and abundance of GBS, I used a cohort of participants from across the state of Wisconsin. I built on the skills I learned in **chapter 2** to investigate a question using only individuals from Wisconsin. I took the skills from the previous two chapters to answer a question with a specific group of individuals from one community in Wisconsin in **chapter 3.** My dissertation has extended outside the borders of University of Wisconsin-Madison’s campus.

Chapter 2: Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer

Patricia G. Wolf^[1,2,3,4], **Elise S. Cowley**^[5,6], Adam Breister[5], Sarah Matatov[3], Luke Lucio[3], Paige Polak[3], Jason M. Ridlon[3,4,7,8], H. Rex Gaskins*[3,4,7,8,9,10], Karthik Anantharaman*[5,6]

AFFILIATIONS

[1]Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL, USA.
[2]University of Illinois Cancer Center, University of Illinois at Chicago, Chicago, IL, USA.
[3]Department of Animal Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[4]Division of Nutritional Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[5]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
[6]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA
[7]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[8]Cancer Center at Illinois, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[9]Department of Biomedical and Translational Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA.
[10]Department of Pathobiology, University of Illinois Urbana-Champaign, Urbana, IL, USA.
^These authors contributed equally
*Corresponding author - karthik@bact.wisc.edu; hgaskins@illinois.edu

KEYWORDS – Colorectal cancer, sulfur metabolism, human microbiome, gut, cysteine, taurine, hydrogen sulfide, metagenomics

This chapter has been published in *Microbiome:*

Wolf, P.G.*, Cowley, E.S.*, Breister, A., Matatov, S., Lucio, L., Polak, P., Ridlon, J.M., Gaskins, H. G., Anantharaman, K. Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. *Microbiome* 10, 64 (2022). https://doi.org/10.1186/s40168-022-01242-x.

*these authors contributed equally to this work

Supplementary material for this chapter can be found at the publisher's website.

Abstract

Background: Recent evidence implicates microbial sulfidogenesis as a potential trigger of colorectal cancer (CRC), highlighting the need for comprehensive knowledge of sulfur metabolism within the human gut Microbial sulfidogenesis produces genotoxic hydrogen sulfide (H_2S) in the human colon using inorganic (sulfate) and organic (taurine/cysteine/methionine) substrates, however the majority of studies have focused on sulfate reduction using dissimilatory sulfite reductases (Dsr).

Results: Here we show that genes for microbial sulfur metabolism are more abundant and diverse than previously observed and are statistically associated with CRC. Using ~17,000 bacterial genomes from publicly available stool metagenomes, we studied the diversity of sulfur metabolic genes in 667 participants across different health statuses: healthy, adenoma, and carcinoma. Sulfidogenic genes were harbored by 142 bacterial genera and both organic and inorganic sulfidogenic genes were associated with carcinoma. Significantly, the anaerobic sulfite reductase *(asr)* gene was twice as abundant as *dsr,* demonstrating that Asr is likely a more important contributor to sulfate reduction in the human gut than Dsr. We identified twelve potential pathways for reductive taurine metabolism and discovered novel genera harboring these pathways. Finally, prevalence of metabolic genes for organic sulfur indicate that these understudied substrates may be the most abundant source of microbially derived H_2S.

Conclusions: Our findings significantly expand knowledge of microbial sulfur metabolism in the human gut. We show that genes for microbial sulfur metabolism in

the human gut are more prevalent than previously known, irrespective of health status (i.e., in both healthy and diseased states). Our results significantly increase the diversity of pathways and bacteria that are associated with microbial sulfur metabolism in the human gut. Overall, our results have implications for understanding the role of the human gut microbiome and its potential contributions to the pathogenesis of CRC.

Background

The human gut is a dynamic nutrient rich environment that harbors a diverse metabolically active microbial community. Human health and disease are inextricably linked to microbial composition, however much remains unknown regarding the functional capacity of human gut microbes [29–31]. This has manifested in bacteria or their niches being loosely characterized as “beneficial”, “commensal”, or “deleterious”, which is problematic as microbial functionality is often species-specific and microbes are capable of metabolic shifts based on available substrates [32,33]. Genomic approaches have enabled rapid discovery of novel bacteria whose functional characteristics have yet to be characterized [34]. These discoveries have filled gaps in knowledge regarding the metabolic capacity of human gut microbes and allow design of hypothesis-driven interventions that create beneficial shifts in microbial communities. This approach may have particularly important implications in human diseases for which associations between microbial composition, dietary intake, and disease risks have been established.

For example, there is strong evidence linking a diet high in red and processed meat with colorectal cancer (CRC) [35]. In addition, bacteria capable of producing hydrogen sulfide (H_2S) are associated with a western diet [36,37], colonic inflammation [38], and CRC [39–47]. At μM concentrations [48], endogenously produced H_2S can act as a vasorelaxant [49], reduce endoplasmic reticulum stress [50], and prevent apoptosis [51]. At millimolar concentrations, as commonly found in the colon, H_2S inhibits cytochrome oxidase causing reductive stress and is genotoxic [52–55]. Previous work investigating microbial sulfidogenesis in the

human gut have mostly focused on sulfate reducing bacteria (SRB) that perform inorganic sulfur metabolism [56]. However, recent evidence indicates that organic sulfur metabolism by gut bacteria may be a key mechanism linking diet and CRC [57]. Indeed, CRC-associated bacteria have been shown to produce H_2S via metabolism of sulfur amino acids [34], and the taurine metabolizing *Bilophila wadsworthia* was found previously to be a significant indicator of CRC [39]. Consumption of a diet high in red and processed meat increases colonic concentrations of organic sulfur, which may increase colonic concentrations of microbially derived H_2S to genotoxic levels [39,58,59]. This suggests that sulfur metabolism in the human gut microbiome may be more widespread than originally believed and exposes current gaps in our knowledge of the metabolic functions of CRC-associated bacteria.

Thus, the objective of this study was to use genomic and metagenomic tools to gain a greater understanding of the sulfidogenic capacity of the human gut microbiome. To do so, we investigated the prevalence of sulfidogenic genes in gastrointestinal bacterial genomes, established a network of sulfur metabolic transformations, and identified novel sulfidogenic bacteria. Using newly developed gene databases, 16,936 publicly available bacterial metagenome assembled genomes (MAGs) from human gut microbiomes were mined to compare the relative contribution of inorganic and organic sulfidogenic genes to microbial sulfur metabolism. Gene presence was then compared among disease states in five CRC microbiome studies to evaluate potential contributions of microbial H_2S production to CRC risk. This study provides the most comprehensive analysis of microbial sulfur metabolism in the human gut to date, and

thereby provides a platform for hypothesis-driven experiments characterizing the role of sulfur metabolites in CRC and other inflammatory-associated gut disorders.

Results

Common pathways of microbial sulfur metabolism are prevalent in human gut microbiomes

To understand the diversity, distribution, and ecology of microbial sulfur metabolism in the human gut and its implications in disease, we investigated the complex pathways for sulfur transformations (Fig. 1).

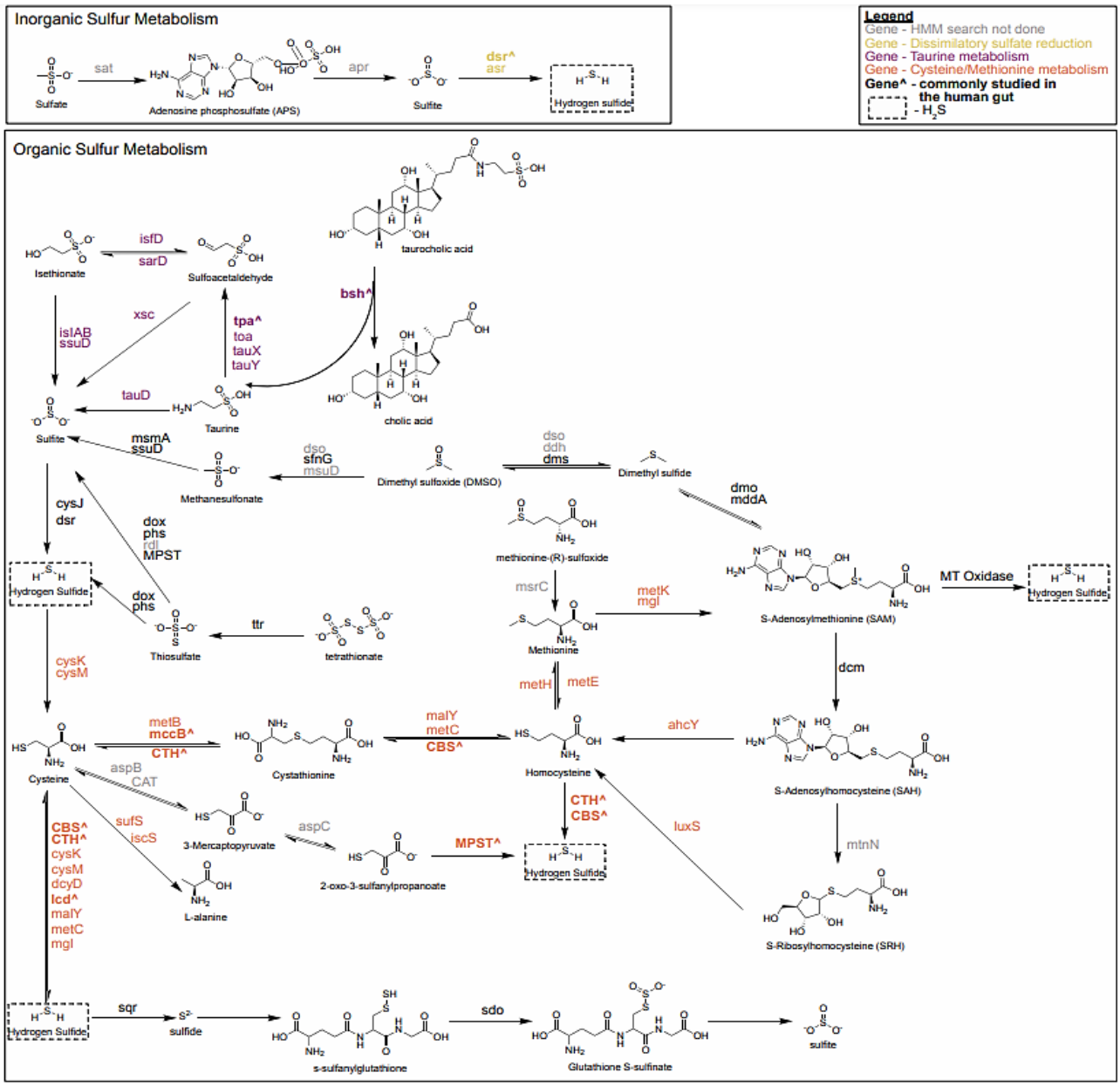


Figure 1. Potential microbial sulfur transformations in the human gut microbiome. Microbial sulfur metabolism results in the production of genotoxic H_2S (dashed box) via metabolism of inorganic sulfate (yellow) or organic sulfur amino acids like cysteine and methionine (maroon), or taurine (orange). Previous studies of microbial sulfidogenesis in the human gut have focused mainly on genes harbored by *Bilophila, Fusobacterium*, and the sulfate reducing bacteria (bolded with a "^"). All genes listed were analyzed in this study except those listed in gray. Reactions are not balanced and only the main sulfur component reactants and products are shown. Some intermediate steps are not shown.

Stool shotgun metagenomic sequence data were used from 5 publicly available studies that investigated the gut microbiome in healthy subjects and patients who had adenoma

or carcinoma of the colon. Collectively, 265 healthy participants, 112 participants with adenoma, and 290 participants with carcinoma were examined. Participant location, associated metadata, and study references are listed in Table 1 [45,60–63]. A total of 16,936 bacterial metagenome assembled genomes (MAGs) were recovered from a previous study that used standardized bioinformatic pipelines for metagenome assembly [64].

Table 1. Overview of original datasets used for this study.

Study	Country of Participant Recruitment	Disease State	Number of Participants	Number of MAGs
Feng [60]	Austria	Control/Healthy	61	1690
		Adenoma	47	1421
		Carcinoma	46	1418
Hannigan [62]	United States, Canada	Control/Healthy	26	101
		Adenoma	23	57
		Carcinoma	26	78
Vogtmann [61]	United States	Control/Healthy	58	1863
		Adenoma	0	0
		Carcinoma	52	1622
Yu [63]	China	Control/Healthy	54	1397
		Adenoma	0	0
		Carcinoma	75	1910
Zeller [45]	France	Control/Healthy	66	1839
		Adenoma	42	956

		Carcinoma	91	2584
Totals		Control/Healthy	265	6890
		Adenoma	112	2434
		Carcinoma	290	7612

A concatenated ribosomal protein tree was created to examine the full diversity of the samples used in this study with the majority of genomes being classified in the phyla Firmicutes, Proteobacteria, and Actinobacteria (Fig. S1). Open reading frames (ORFs) were then predicted followed by homology-based identification of 79 genes associated with microbial sulfur metabolism (Table S1) using available and custom Hidden Markov Models (HMMs). These analyses revealed the breadth of themicrobial sulfur metabolic potential in the human gut and its associations with CRC (Table S2, Fig. S1).

To understand the prevalence of microbial sulfur metabolism in a representative human gut microbiome, 514 fecal microbial genomes were obtained from the Human Microbiome Project (HMP) and surveyed for common metabolic pathways associated with H_2S production from cysteine, taurine, and sulfate/sulfite. Functional genes encoding proteins for cysteine metabolism made up the majority of sequences and included cystathionine-β-lyase (*malY*, *metC*) (11.8%), cystathionine-β-synthase (*CBS*) (27.5%), cysteine desulfhydrase (*lcd*) (2.6%), D-cysteine desulfhydrase (*dcyD*) (17.9%), and methionine-γ-lyase (*mgl*) (14.7%). In contrast, dissimilatory sulfite reductases (*dsrAB*) catalyzing the final step of sulfate and taurine respiration to form H_2S, made up only 18.2% of identified genes (Fig. S2). In total, 313 sulfidogenic genes were identified

from 183 bacterial genomes (35.6% of total genomes), and spanned across six phyla including Proteobacteria (36.1%), Firmicutes (26.2%), Bacteroidetes (20.2%), Fusobacteria (15.8%), Actinobacteria (1.1%), and Synergistetes (0.5%). This key finding demonstrates that pathways for sulfur metabolism were prevalent in human gut microbiome genomes and that cysteine may be an underestimated substrate for microbial sulfur metabolism in the gut.

Genes for anaerobic sulfite reductases (*asrABC*) are more prevalent than dissimilatory sulfite reductases (*dsrAB)* in the human gut

Of the limited literature regarding human colonic sulfidogenic bacteria, the most well studied are the SRB which are capable of reducing inorganic sulfate supplied exogenously by diet or endogenously by degradation of sulfated bile acids and mucins (estimated 1.5-16 and 0.96- 2.6 mmol/day respectively) [56,65–68]. Two enzymes are able to complete the final step of the reaction which catalyzes the six-electron reduction of sulfite to H_2S — dissimilatory sulfite reductase (Dsr) and anaerobic sulfite reductase (Asr) (Fig. 1). It has been proposed that sulfite reduction takes place as a series of two electron transfers to DsrAB from the DsrMKJOP complex via DsrC [69]. Genes for the Dsr pathway, *dsrAB*, are highly conserved among SRB and diversely distributed among phyla in environmental samples [70]. However, culture and PCR based studies of SRB diversity in human stool and colonic mucosa indicate that *dsrAB* is harbored by only five resident genera namely *Bilophila* spp., *Desulfovibrio* spp., *Desulfobulbus* spp., *Desulfobacter* spp., and *Desulfotomaculum* spp. [71–74].

Within our database, *dsrAB* genes were present in 121 MAGs of 16,936 bacterial MAGs (<1% of total MAGs), from human gut samples and in 17.4% of total subjects (Table S3). Taxonomic classification demonstrated genera commonly associated with Dsr activity in the gut microbiota were represented including *Desulfovibrio* spp. and *Bilophila* spp. In addition, six genera were revealed that are not commonly ascribed as human gut SRBs namely *Collinsella* spp., *Eggerthella* spp., *Enterococcus* spp., *Flavinofracter* spp., *Gordonibacter* spp., and *Roseburia* spp (Tables S2, S4). Since previous phylogenetic analyses in environmental samples indicated that *dsrAB* acquisition was often the result of multiple lateral gene transfer events [70], a concatenated gene tree was generated with a reference database of *dsrAB* sequences to observe the consensus phylogeny of *dsrAB* sequences in human gut bacteria. *DsrAB* sequences from sample MAGs separated into three distinct clusters which corresponded with the respective phyla: Actinobacteria, Firmicutes, and Proteobacteria (Fig. 2A). This suggests that lateral gene transfer of *dsrAB* may be less common in SRBs of the human gut than those observed in environmental studies [70].

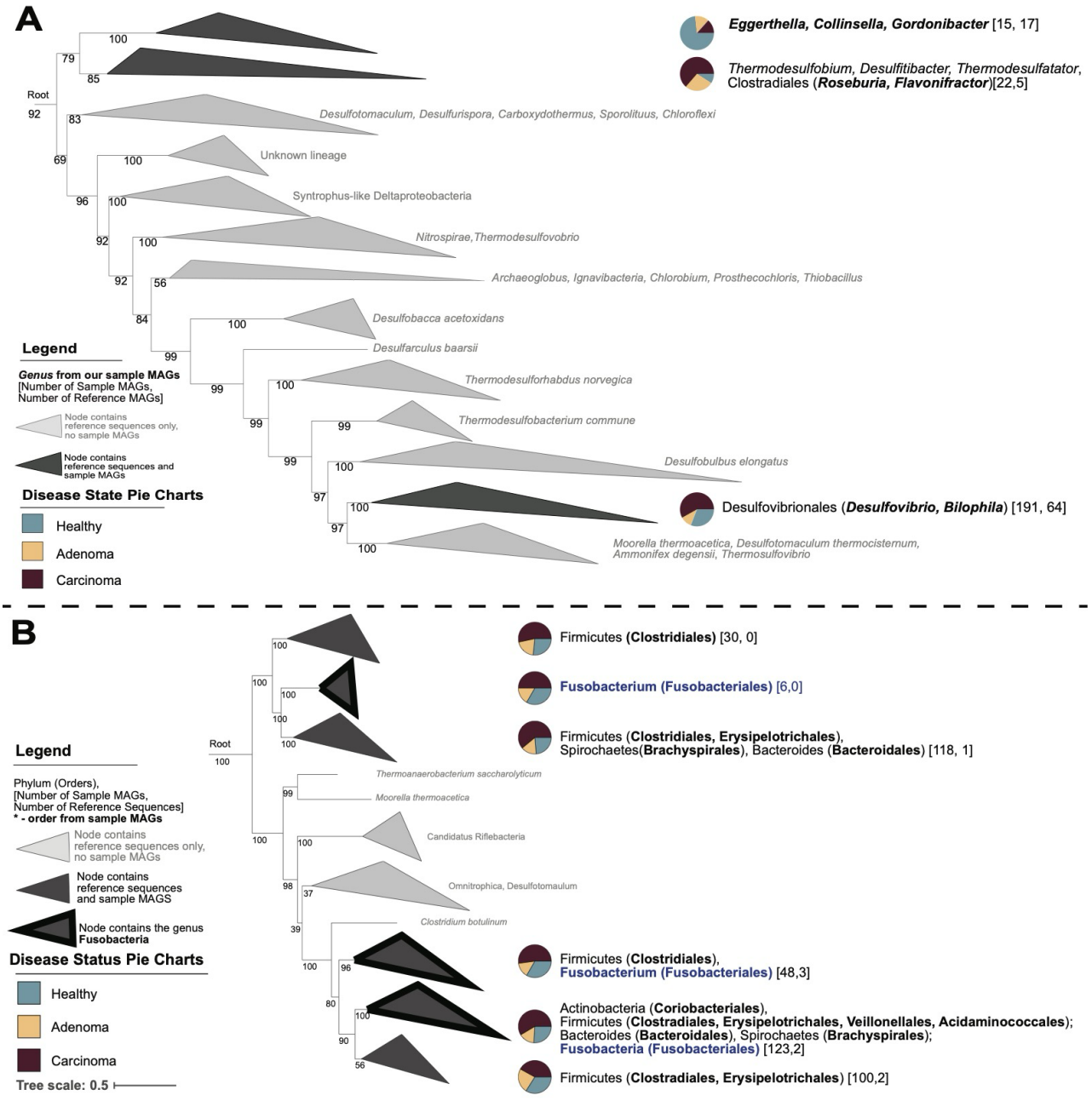


Figure 2. Concatenated protein trees for dissimilatory sulfate reduction pathways. A. Concatenated protein tree showing the diversity of bacteria that possess genes for the final enzyme of the dissimilatory sulfate reduction pathway — dsrAB. B. Concatenated protein tree showing the diversity of bacteria that possess genes for anaerobic sulfite reductase (asrABC), an enzyme also capable of dissimilatory sulfate reduction. Gray clades only contain reference sequences, darker gray clades contain reference sequences and sequences from this study. Bracketed numbers indicate the sequence origin within each clade: [number of sequences from our study, number of sequences from references]. Bacterial genera (dsr) or orders (asr) originating from study samples are bolded. Pie charts indicate the disease state associated with sequences within each clade with blue indicating healthy, yellow adenoma, and maroon carcinoma. Clades outlined in black contain Fusobacterium sequences.

While SRBs may be the most well studied sulfur metabolizing bacteria of the human gut microbiota, studies have focused mainly on bacteria harboring Dsr enzymes. Similar to Dsr enzymes, Asr performs a 6-electron reduction of sulfite to H_2S (Fig. 1). A recent analysis of environmental diversity of the *asrABC* complex revealed its presence in common residents of the gut microbiome including *Fusobacterium nucleatum* and *Clostridium intestinale* [75]. To determine if gut bacteria harbor *asrABC*, we searched the MAG database, revealing that *asrABC* genes were more prevalent in all MAGs and participants than *dsrAB* genes. The *asrABC* genes were present in approximately 2% of total MAGs (388, 390, and 375 MAGs, respectively), and approximately 35% of total subjects (Table S3). Intriguingly, less than a quarter of the subjects that harbored *asrABC* genes also possessed genes for *dsrAB* (53 of 239 subjects). Taxonomic assignments showed thirty-one genera possessed *asrABC,* spread among five phyla including Actinobacteria, Firmicutes, Fusobacteria, Spirochaetes, and a phylum not previously shown to possess *asrABC* — Bacteroidetes [75]. A consensus tree of reference and sample MAG concatenated *asrABC* sequences revealed clustering of *asrABC* genes across six nodes that were distinct from sample MAG phylogeny. Notably, *asrABC* genes possessed by *Fusobacterium* spp. did not cluster together, but were observed in three separate nodes suggesting this pathway was acquired via multiple lateral gene transfers (Fig. 2B). Together, these data indicate that Asr enzymes may be more important contributors to sulfate and sulfite reduction than Dsr in the human gut, and that bacteria acquired these enzymes via a divergent phylogenetic history.

Diverse bacteria harbor pathways for taurine metabolism

Sulfonates are organic sulfur compounds with a SO_3^- moiety that are abundant in marine sediments and detergents, and play an important role in the environmental sulfur cycle [76]. While less is known about the role of sulfonates in the human intestine, there is evidence that metabolism of sulfonates like isethionate and taurine is performed by resident gut microbes [77,78]. Microbial taurine metabolism has gained considerable interest after it was implicated as a potential dietary mechanism of colitis and CRC disparities [38,39]. Representing 18.3% of total free amino acids in the colonic mucosa (13.6 ± 0.5 mmol/kg), taurine is the second most abundant free amino acid in this tissue [79]. Taurine is provided as a substrate to the human gut microbiota either directly through diet or through hydrolysis of taurine conjugated bile acids by the enzyme bile salt hydrolase (BSH). Excess consumption of taurine and cysteine increases tauro-conjugation of secreted bile acids, thus potentially providing additional substrates for bacteria with BSH activity [57]. The *bsh* gene was identified to be present in 15% of total MAGs, and in 89% of total subjects (Table S3). This prevalence was unsurprising, as it has been proposed that hydrolysis of conjugated bile acids may be a detoxification strategy to decrease bile acid toxicity [80], or may serve as a source of nutrients for microbial growth and energy metabolism [56].

Once liberated via BSH or made available through dietary intake, taurine can be metabolized by gut bacteria via oxidative or reductive pathways. Taurine oxidation using the enzyme taurine dioxygenase (TauD) is generally not considered in the literatures a source of H_2S production in the anaerobic environment of the human colon. However,

previous studies have reported an increase of aerotolerant bacteria adherent to the mucosal surface suggesting a luminal gradient of oxygen provided by host tissues [81]. Assessment of *tauD* gene abundance demonstrated that while the enzyme was present in only 1% of total MAGs, these MAGs were present in 23% of subjects (Table S3). As expected, *tauD* genes were present in six facultative genera (*Escherichia* spp., *Enterobacter* spp., *Citrobacter* spp., *Morganella* spp., *Hafnia* spp., and *Raoultella* spp.), however none of these genera possessed genes for anaerobic or dissimilatory sulfite reduction (Table S4). Thus, TauD appears to be primarily used for taurine assimilation and not H_2S production.

The only known bacterium to possess the reductive pathway of taurine metabolism in the human gut is *B. wadsworthia.* However, given the taurine rich environment of the colon, it is likely that other bacteria capable of performing this metabolism remain to be discovered. Thus, to identify candidates that may have the capacity to produce H_2S from taurine in an anaerobic environment, HMM searches of the described cohorts were performed targeting pathways as shown in Fig. 3. Pathway 1 describes the putative taurine reduction pathway previously thought to be possessed by *B. wadsworthia* [56] Our search confirmed that this pathway was not possessed by *Bilophila* spp., as recently described [82]. Instead, analyses revealed two genera that harbor genes for this pathway, namely *Desulfovibrio* spp. and *Flavonifractor* spp. The first step of taurine reduction involves the liberation of the nitrogenous group from taurine producing pyruvate or 2-oxoglutarate via the enzymes taurine-pyruvate aminotransferase (Tpa) or taurine-2-oxoglutarate transaminase (Toa), respectively (Pathway 2). For all MAGs that possess

the latter two genes of this pathway, the genes *tpa* and *toa* co-occur. Notably, *Flavonifractor* spp. also harbors genes for the AsrABC complex, indicating an alternative final step of the taurine reductive pathway (Pathway 2). Additionally, evaluation of three-step pathway combinations revealed 34 genera that possessed the first and final pathway steps, suggesting the pervasiveness of metabolic cooperation in

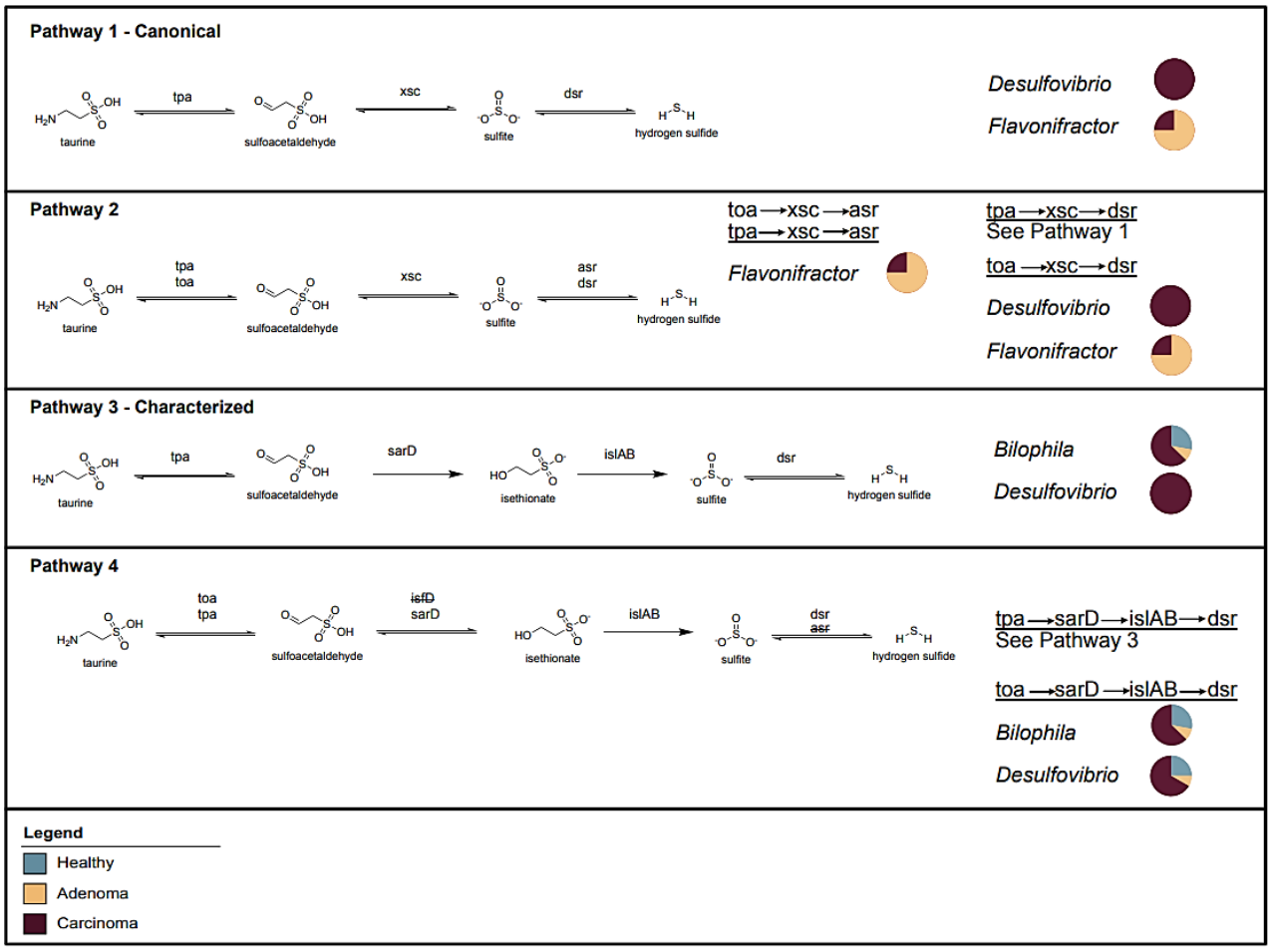


Figure 3. **Characterized and proposed pathways of microbial taurine reduction to H_2S.** Pathway 1 - the canonical pathway of taurine reduction in *Bilophila wadsworthia.* Pathway 2 - putative 3-step reactions for taurine reduction analyzed in this study. Pathway 3 - the recently characterized pathway for taurine reduction in *Bilophila wadsworthia.* Pathway 4 - putative 4-step reactions for taurine reduction analyzed in this study. No complete pathways were found involving genes that are struck through. Genera possessing genes for each complete pathway are listed. Only genera listed were found to have complete pathways. Pie charts indicate the disease state associated with MAGs of each genus with the specified pathway with blue indicating healthy, yellow indicating adenoma, and maroon indicating carcinoma.

the gut and potentially revealing targets for the identification of novel sulfoacetaldehyde acetyltransferases (Xsc) (Fig. 3) (Tables S4, S5).

Pathway 3 represents the recently characterized pathway for taurine reduction in *B. wadsworthia* (Fig. 3) [82]. Gene searches corroborated that *Bilophila* spp. harbored all four genes in this newly defined pathway. However, contrary to what was recently reported, all four genes were also observed in *Desulfovibrio* spp. Since MAGs are unable to allow for granularity at the species level, future work is needed to determine if this pathway is indeed more widespread in resident *Desulfovibrio* species of the human gut. Seven MAGs annotated as *Desulfovibrio* spp. and *Bilophila* spp. also possessed *toa* in the absence of *tpa*, together, this indicates an alternative first step to this pathway (Pathway 4). Notably, unlike the three step Pathway 2, MAGs that possess all genes for the four-step pathway harbor only *dsrAB* and not *asrABC.* In addition, evaluation of four step pathway combinations revealed seven genera missing either the second or third step of the pathway revealing additional targets for gene discovery (Fig. 3) (Tables S4, S5).

Cysteine and methionine are understudied and abundant sources of microbially-derived H_2S in the human gut

Cysteine is a conditionally essential amino acid, which is provided to the intestine directly by diet or the decomposition of methionine (Fig. 1). Methionine restriction alters microbial composition of the gut and down-regulates inflammatory pathways related to

oxidative stress [83,84]. In addition, the production of H_2S via cysteine degradation supports microbial growth and protects from oxidative stress in response to antibiotic treatment [85]. Cysteine metabolizing bacteria are implicated as a source of oral abscess, breath malodor, and delayed wound healing in the oral cavity [86–89], and have been repeatedly associated with CRC [40,42,44,45]. Of late, *F. nucleatum* has been of particular interest in CRC [42,44,90]. Studies have demonstrated association between *F. nucleatum* and the tumor surface in a subset of CRC [40,42,91], *F. nucleatum* DNA in CRC tumors correlate with reduced survival [92,93], and two subspecies of *F. nucleatum* (*vincentii* and *animalis*) have been proposed as part of a microbial signature for fecal-based CRC classification [45]. However, few studies appreciate that *F. nucleatum* is sulfidogenic, and many other bacteria can also produce H_2S from cysteine in the human gut.

To gain an appreciation of the abundance of cysteine metabolism within the human microbiome, MAGs were searched for the following genes associated with cysteine metabolism *dcyD*, *malY*, *metC*, *mgl*, as well as cysteine synthase (*cysK*, *cysM*), cysteine desulfurase *(iscS*, *sufS*), and cystathionine-γ-synthase (*metB*), and the three human orthologs cystathionine-β-synthase (*CBS*), cystathionine-γ-lyase (*CSE/mccB*), and 3-mercaptopyruvate sulfurtransferase (*3MST*). Genes for upstream pathways of methionine and homocysteine metabolism were also analyzed (Fig. 1, Table S1). Searches revealed that all cysteine metabolizing genes were highly present, with *cysK*, *lcd*, *malY*, and *sufS* observed at least once in over 96% of subjects (Table S3). Even those genes that were observed in only 2-5% of total MAGs — namely *dcyD*, *metC*, *cysM*, and *mgl* — were still observed to be present in at least 40% of subjects (Table S3). In accordance with this, genes for microbial pathways for methionine metabolism

were also highly present in subjects, indicating that methionine may be an important source of microbial derived cysteine in the human gut (Table S3). In addition to being abundantly present, genes for cysteine metabolism were also diversely distributed among 13 phyla and 141 genera. Amongst these, 84 genera have not been previously characterized as being sulfidogenic. (Table S4). This may be particularly important in the context of a western diet, as studies using fecal homogenates demonstrate higher production of H_2S from organic sulfur amino acids compared to inorganic sulfate [94], and higher protein intake increases ileal output of protein (2.69 vs. 7.45 g/day) and free amino acids (6.90 vs. 20.48 μmol/mL) [95]. Although additional work is needed to comprehensively resolve cysteine metabolism, together these data indicate that the sulfur amino acids cysteine and methionine may be an understudied and abundant source of microbially-derived H_2S in the human gut.

Microbial sulfur metabolism is statistically associated with colorectal cancer

Sulfate reducing bacteria and H_2S have been implicated in CRC pathogenesis [52,53,58,96,97], however associations of other microbial sulfur metabolism genes with CRC have been less well studied. The proportion of participants in each health status (healthy, adenoma, or carcinoma as evaluated by colonoscopy) category with at least one copy of each gene was determined along with the proportion of the total number of MAGs in each health status with the gene (Fig. 4, Table S3). Genes involved with cysteine and methionine metabolism were generally abundant regardless of health status (Fig. 4A, size of dots) and abundant in many MAGs (Fig. 4A, color of dots). Genes involved in sulfur and taurine metabolism were variable in their distribution among participants in the three health states and in MAGs (Fig. 4).

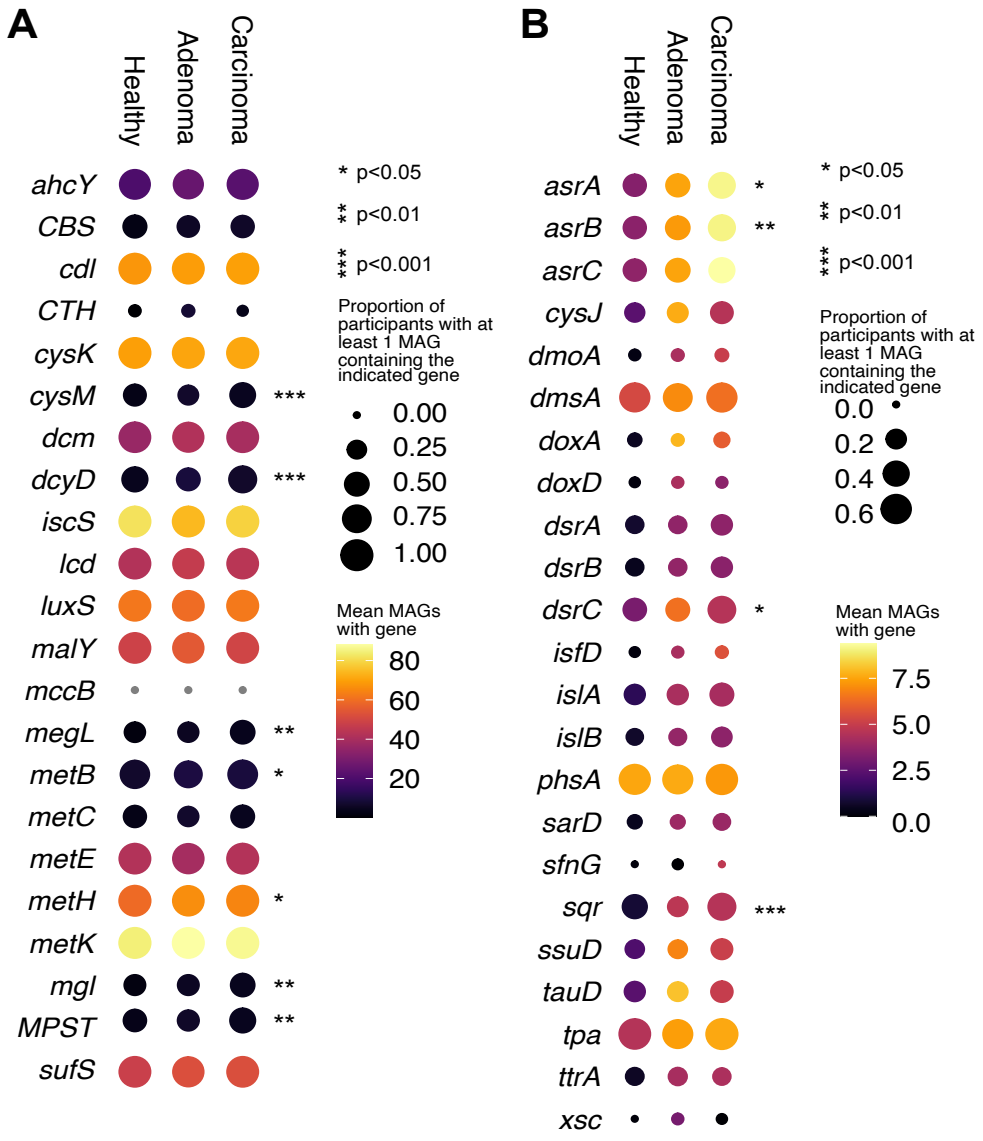


Figure 4. Genes for microbial sulfur metabolism are abundant and significantly associated with colorectal cancer. Dot plots of selected genes related to microbial cysteine and methionine metabolism (A) and taurine and sulfur metabolism (B) across three disease states: healthy, adenoma, and carcinoma. The size of each dot indicates the proportion of participants in each disease state with at least 1 copy of the indicated gene in their bacterial MAGs and the color of each dot indicates the mean number of MAGs with that gene in the subset of participants that have at least 1 copy of the gene. Genes that have a non-random distribution across disease status as analyzed by chi-squared analysis are indicated by asterisks. P-value corrections were done using the Benjamini-Hochberg (BH) Procedure.

To determine if sulfur metabolism was associated with CRC, we conducted statistical tests to study the distribution of sulfur genes in participants across each health status. Genes for tetrathionate metabolism, which was previously implicated as an electron acceptor provided by gut inflammation [98], were present in less than 1% of MAGs and were not significantly different among the three health statuses. Genes involved in cysteine and methionine metabolism including *cysM, dcyD, mgl, metB, metH,* and *sdo*, exhibiteddistributions that were statistically significantly different (at least p-value < 0.05) among participants in the three health states (Fig. 4A, Table S6, Fig. S3, S4), with *sdo, cysM, mgl, metB,* and *metH* being more likely to be found in carcinoma. While 92 genera with cysteine metabolizing genes have been associated with CRC previously, our results indicate that this association may involve sulfidogenesis. Indeed, MAGs for cysteine metabolizing genes were pervasive in genera most commonly associated with CRC, corroborating recent work that observed that genes for cysteine metabolism were significantly more abundant in subjects with CRC [99] (Fig. S5). Sulfur and taurine metabolism genes, *asrA, asrB, dsrC,* and *sqr*, had distributions that were statistically significantly different (at least p-value< 0.05) among the three health states with all being more likely to be found in carcinoma (Fig. 4B, Table S6, Fig. S3, S4).

Intriguingly, for genes related to dissimilatory sulfate reduction, different organisms exhibited different clustering patterns based on both phylogeny and disease state. For the Dsr pathway, the majority of *Desulfovibrio* spp., *Bilophila* spp., *Flavinofracter* spp., and *Roseburia* spp. were harbored by participants with carcinoma while the majority of *Collinsella* spp., *Eggerthella* spp., and *Gordonibacter* spp. originated from healthy

participants (Fig. 2A - pie charts). For the Asr pathway, the majority of sequences were recovered from participants with carcinoma (Fig. 2B - pie charts). These data along with the associations presented in Fig. 4B demonstrate that genes for sulfate reduction are associated with carcinoma.

To gain a deeper understanding of the ecology of microbial sulfur metabolism during colorectal carcinogenesis, we first determined the associations of sulfidogenic genes among different CRC stages from the three datasets that reported staging data, and then investigated growth rates of selected taxa previously considered to be microbial markers of CRC [39,45]. The likelihood of patients having sulfidogenic genes was not significantly different among stages (Supplemental Table 7). However, participants were more likely to have genes for inorganic sulfur metabolism in earlier stages of CRC, while participants having genes for organic sulfur metabolism was uniformly high independent of stage (Fig S6). *Mgl* — an organic sulfur metabolizing gene — was increasingly present across advancing cancer stages. This is intriguing as *mgl* has been shown recently to be the most highly transcribed of cysteine metabolizing genes [99]. Growth rate analysis of bacterial indicator species of CRC did not reveal statistically significant differences among disease states for any species analyzed, however this analysis was restricted by the limited presence of these species within healthy subjects (Fig. S7, Table S8, S9). Future analyses of microbial species that are present across all disease states is needed to determine if other keystone species play a role in CRC pathogenesis than determined previously by taxa abundance.

Discussion

There is compelling evidence implicating microbial H_2S production as an environmental trigger of CRC, however ongoing studies investigating this link have been hampered by the field's incomplete knowledge of sulfur metabolism within the human gut. Here, we performed a comprehensive characterization of the functional capacity of the human gut microbiome to conduct sulfur transformations and produce H_2S. Our investigation into the diversity and ecology of inorganic sulfur metabolism pathways observed that highly conserved functional genes encoding the final step of the sulfate reduction pathway — *dsrAB* — was harbored by six genera not typically targeted as SRB. Additionally, this investigation revealed that *asrABC*, which encodes an enzyme with the same biochemical activity to Dsr, was both twice as abundant in total MAGs and was present in twice as many subjects than *dsrAB.* Together, these data highlight that genes for inorganic sulfur metabolism in the human gut are more widespread than previously established and that *asrABC* may be an important marker to measure the capacity of microbial sulfate reduction within cohorts. Since the diversity of microbial sulfatases have not been characterized, studies that compare substrate specificity and the catalytic efficiencies of these enzymes are needed to truly understand the implications of this expanded view of inorganic sulfur metabolism in the human gut.

While previous investigations of microbial H_2S production and human disease have focused on SRB, bacteria that metabolize sulfur from organic sources have been consistently associated with CRC risk. Prior to this work, *Bilophila wadsworthia* was the only bacterium known to produce H_2S via taurine respiration. However, our analysis

revealed *Desulfovibrio* spp. harbors genes for the characterized 4-step reductive pathway of taurine respiration, and an exploration of twelve taurine reduction pathways revealed two genera with genes for the complete 3-step reduction pathway. Further, 41 unique genera were found to have nearly complete 3- and 4-step pathways, disclosing microbial targets for novel enzyme discovery or investigations of cooperative taurine metabolism. Finally, analyses of diverse pathways for microbial cysteine and methionine degradation revealed these sulfidogenic genes were distributed among diverse bacterial phyla and were abundantly present among subjects. Collectively, these analyses demonstrate that bacteria harboring pathways for organic sulfur metabolism are pervasive in the human gut and likely constitute the most abundant source of microbially derived H_2S.

To determine whether sulfur metabolism was differentially associated along the colorectal carcinoma sequence (healthy → adenoma → carcinoma), presence of sulfidogenic genes was compared between healthy subjects, and patients with adenoma or carcinoma. Genes for both inorganic and organic sulfur metabolism were significantly associated with carcinoma, which is intriguing as both sulfate and taurine metabolism share a final metabolic step. These associations along with the pervasiveness of genes for organic sulfur metabolism in studied MAGs supports the unique hypothesis that organic sulfur metabolism by gut bacteria is a key mechanism linking a western diet and CRC risk (Fig. 5).

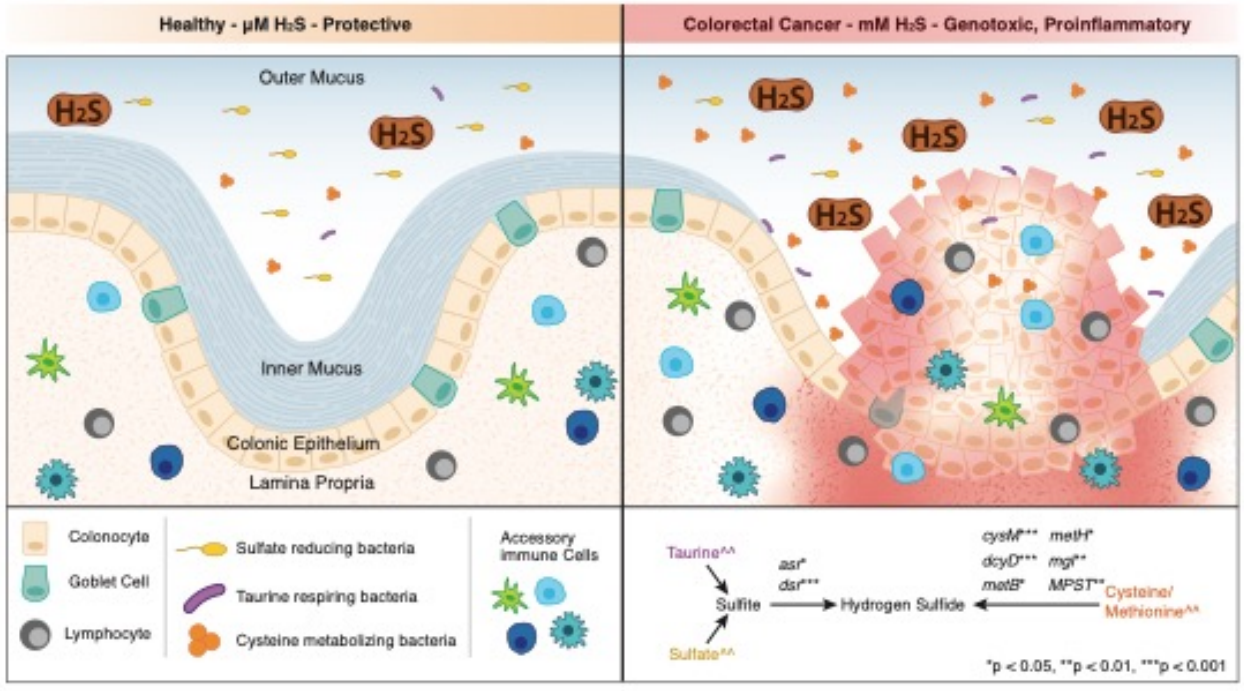


Figure 5. Organic sulfur metabolism by gut bacteria may be a key mechanism linking a western diet and CRC risk. The degradation of sulfomucins by mucolytic bacteria are a key source of inorganic sulfate for sulfate reducing bacteria. At µM concentrations, basal production of H_2S through inorganic sulfate reduction exerts beneficial effects including gut barrier protection and fermentative hydrogen disposal. Intake of a western diet abundant in red and processed meat amplifies the production of taurine conjugated bile acids and increases colonic exposure to dietary sulfur amino acids (taurine, methionine, cysteine). In the context of a western diet, metabolism of organic sulfur amino acids by gut microbes drives the production of H_2S to genotoxic and pro-inflammatory levels (mM concentration). Simplified pathways demonstrate genes for sulfur metabolism that were significantly associated with CRC. ^ indicates inorganic sulfur sources primarily provided by sulfated bile acids and sulfamucins. ^^ indicates organic sulfur sources provided by dietary sulfur amino acids and conjugated bile acids.

A limitation of our study is that our analyses utilized MAGs which do not use all available metagenomic reads as a consequence of assembly and binning. Thus, it is possible that the abundance of sulfur genes present in the human gut is even greater than reported herein. However, the use of MAGs enabled taxonomic information,

reconstruction of full pathways, and calculation of growth rates for selected CRC associated bacteria, thus providing robust new information regarding microbial sulfur metabolism in the human gut. In addition, this work focused on the functional capacity of the gut microbiome for sulfur metabolism based on genetic content. Future work that demonstrates metabolic activity of microbes and their enzymes in vitro, as well as metabolic flux in a complex microbial community, are needed to verify these conclusions.

Conclusions

Overall, the data demonstrate that genes for microbial sulfur metabolism are more diverse than previously recognized, are widely distributed in the human gut microbiome, and are significantly associated with CRC. These data expand what was previously known regarding the diversity of bacteria with the genetic capacity to perform sulfur metabolism, and indicate that genes for organic sulfur metabolism may be the most important contributor of H_2S in the human gut. Our findings provide a foundation for future work characterizing the activity of sulfidogenic enzymes in diverse microbial species, exploring the expression of these genes as affected by health status, and examining microbial H_2S induced tumorigenesis in animal models of CRC and human disease.

Methods

Genomic survey of sulfidogenic genes in Human Microbiome Project genomes

An initial genomic survey was performed using 514 gastrointestinal genomes obtained from the Human Microbiome Project (HMP) in Fall 2018 [100,101]. Reference sequences

were obtained from the National Center for Biotechnology Information (NCBI) using searches for sulfidogenic genes from known residents of the human gut including “cysteine desulfhydrase”, “Cdl”, “Lcd”, “cystathionine-beta-synthase”, “L-methionine–gamma-lyase”, “dissimilatory sulfite reductase”, “dsrA”, “dsrB” and “dsrAB”. Searches of the HMP genes were then performed using BLAST (BLASTv2.8.1+) [102], and alignments with greater than 60% identity and a minimum query coverage of 40 amino acids were retained. To filter non-homologous proteins, gene hits were compared to KEGG.

Sulfur pathway visualization

Sulfur cycle reactions were created in ChemDraw Prime 16.0 and further modified in Affinity Designer.

Downloading MAGs and accessing metadata

The previously reconstructed MAGs from the five cohorts were downloaded from https://opendata.lifebit.ai/table/sgb.The associated file that was downloaded, “download_files.sh” was used to download all 16,936 genomes from the 5 studies [64]. The link to download the metadata (https://www.dropbox.com/s/ht0uyvzzal6exs2/Nine_CRC_cohorts_taxon_profiles.tsv?dl=0) was found at http://segatalab.cibio.unitn.it/data/Thomas_et_al.html from two studies [103,104] and filtered to only include the 5 cohorts used in our study. For samples that had AJCC TNM (Tumor, lymph Node, Metastasis) classification without a stage, the American Cancer Society guidelines to annotate stage based on TNM classification was used. For this study, all “high” (>90% complete, <5% contamination, <0.5% strain

heterogeneity) and “medium” (>50% complete, <5% contamination) quality MAGs were included [64,105].

Gene annotation of MAGs

Prodigal module (prodigal version 2.6.3) of METABOLIC was used to run multiple threads with the -p meta option to annotate open reading frames (ORFs) on all MAGs [106,107].

Sulfur gene identification in MAG database

HMM - Hidden Markov Model (HMM) searches for protein sequences were performed using either hmm profiles from KEGG or custom profiles against a concatenated file of predicted ORFs from all 16,936 MAGs used for this study with HMMSearch version 3.3.0 [108], using trusted cut offs for all sulfur related sequences (Table S1). All HMMs used in this study are available at https://github.com/escowley/HumanGutBacterialSulfurCycle.

Microbial sulfur pathway literature search

To determine the breadth of current knowledge regarding characterized H_2S production in human gut bacteria and their association with CRC, a literature search was performed on genera revealed in our analysis. For MAGs that possessed sulfidogenic genes, the annotations for the “closest genus” were recorded. Web-searches were performed for each genus using the key words “colorectal cancer”, “sulfide”, and “H_2S”.

Genera previously associated with CRC or who have species characterized previously to produce H_2S are noted in Supplemental Table 4.

Taxonomic classifications of MAGs

Taxonomic classifications, "closest" designators, were determined previously by binning the MAGs with a reference set of genomes [64]. To verify taxonomic classifications for MAGs, a concatenated ribosomal protein tree, GTDB-Tk classification, and 16S rRNA alignment were done. Taxonomic classifications from the original study, GTDB-tk, and 16S rRNA genes can be found in Supplemental Table 10. 16S rRNA sequences were extracted from the MAGs using the ssu_finder function of checkM (version 1.0.11) [109]. From the 16,936 MAGs, 2531 16S rRNA sequences were extracted. Extracted 16S rRNA sequences were classified using with SINA aligner (version v1.2.11) (general options - bases remaining unaligned at the end should be attached to the last aligned base, reject sequences below 70% identity) using the search and classify feature (search and classify options - min identity with query sequence 0.95, number of neighbors per query sequence 10, sequence collection used - Ref-NR, search kmer candidates 1000, lca-quorum 0.8, search k-mer len 10, search kmer mm 0, search no fast, taxonomies used for classification SILVA, RDP, GTDB, LTP, EMBL-EBI/ENA), which classifies 16S rRNA sequences based on the least common ancestor and the SILVA reference database (release 138.1) [110–112]. Taxonomy was assigned using GTDB-Tk (version 1.3.0) with database release 95 with the classify_wf function [113–115].

For the concatenated ribosomal protein tree, a curated Hidden Markov Models (HMM) database for single-copy ribosomal proteins (rpL2, rpL3, rpL4, rpL5, rpL6, rpL14, rpL14, rpL15, rpL16, rpL18, rpL22, rpL24, rpS3, rpS3, rpS8, rpS10, rpS17, rpS19) [116] was used to identify these genes in all MAG using hmmsearch (version 3.3.0) using noise cutoffs (--cut_nc) [108]. Once identified, the protein sequences were extracted from the predicted ORFs and imported into Geneious Prime (v 11.1.5). Each sequence was aligned with a reference set using MAFFT (v 7.450, parameters Algorithm: Automatic, Scoring Matrix: BLOSUM62, Gap penalty: 1.53, Offset value: 0.123) [117]. The alignments were manually trimmed and a 95% gap masking threshold was applied to the resulting alignment. The resulting alignments were concatenated for each gene set. The concatenated alignments were exported in fasta format from Geneious and used as the input for making phylogenetic trees with IQTree using version 1.6.9 (-nt AUTO -m MFP -bb 1000 -redo -mset WAG,LG,JTT,Dayhoff -mrate E,I,G,I+G -mfreq FU -wbtl) [118]. The resulting tree file was imported into iToL for visualization and collapsing of nodes followed by final modifications in Affinity Designer [119].

Concatenated protein trees for dissimilatory sulfate reduction genes

To create the concatenated protein trees for *asrABC* and *dsrAB*, reference sequences for each set of genes from a diverse set of environments were utilized [75]. Hmmsearch (version 3.3.0) was used to identify asr and dsr sequences in the database of predicted ORFs [108]. The identified genes were extracted from the predicted ORFs and concatenation and tree building were done according to the protocol previously described for the ribosomal proteins.

Summary calculations and statistical analysis for association of sulfur genes with disease state and stage of CRC

Summary calculations of total number of MAGs with the gene of interest per disease state, mean, median, and standard deviations were performed in R. Summary information can be found in Supplemental Table 3. To identify potential associations between presence of specific microbial sulfur genes and disease state, chi squared tests were performed. First, for each gene, each participant was binarized as either "presence" (at least 1 MAG with at least 1 copy of the gene of interest) or "absence" (no MAGs recovered from the participant's sample had any copies of the gene of interest). For each gene, total presence and absence were tallied for each disease state (healthy, adenoma, carcinoma). Chi square tests were performed for each gene on a 2X3 matrix of presence/absence totals and disease states. P-value corrections were done using the Benjamini-Hochberg (BH) Procedure for final reported p-values. A significant p-value indicates the distribution of the gene among the disease states is not random. Summary values and statistics were all performed in R. Uncorrected and corrected p-values can be found in Supplemental Table 6. Dot plot visualizations with the proportion of participants in each of the three disease states with at least one copy of the gene in their MAGs normalized to total number of participants with that disease state as the size of the dot and mean number of MAGs with a copy of the gene per participant with at least one copy as the color of the dot were made using R with the cowplot package. Data used for these plots can be found in Supplemental Table 3.

Similarly, to identify associations between presence of specific microbial sulfur genes and colorectal cancer stage, the subset of studies with staging data was binarized as described above and binomial logistic regressions were performed. Rao score tests were performed on the resulting binomial logistic regressions to evaluate for significance of each gene having evidence of association with staging of CRC. P-value corrections were done using the Benjamini-Hochberg (BH) Procedure for final reported p-values. Uncorrected and corrected p-values can be found in Supplemental Table 7. Presence/absence values for each gene in each stage were normalized to the total number of participants in that particular stage to generate distribution plots (Fig. S6).

Metabolic reconstruction of MAGs

Metabolic reconstruction of each MAG was accomplished using the METABOLIC-G program of METABOLIC (version 4.0) [107]. Summary information is available at https://github.com/escowley/HumanGutBacterialSulfurCycle.

Determination of growth rates for colorectal cancer indicator bacteria

To determine the growth rates of bacteria previously implicated as indicators of colorectal cancer in the gut community [39,45], the original reads were downloaded and used to generate these genomes from the Hannigan, Yu, Zeller, and Feng studies. For the reads from the Hannigan study, the fasterq-dump (version 2.9.4) module of the SRA toolkit [120]was used, for the reads from the Yu and Zeller studies, the Aspera [121] command line interface (CLI) ascp program (v3.9.1.168954) was used, and for the reads from the Feng study, the Aspera CLI ascp program (v3.9.3.177167) was used,

respectively. For reads from the Zeller study, multiple read sets are deposited for each participant, and the first listed paired-end read set for each genome listed in the European Nucleotide Archive (ENA) metadata was chosen. Reads from the Hannigan study were trimmed and quality filtered with metaWRAP (v1.2.2) using the read_qc[122] module with the option "--skip-bmtagger". Bowtie2 (v2.3.4.1)[123–125] was used with the option "--reorder" to map MAGs classified as those bacteria to the original read sets and shrinksam [126] version 0.9.0 with the "-u" flag to compress mapping files. Growth rates were determined by generating un-filtered indexes of replication with iRep (version 1.10) [127]. Growth rates for the following organisms were determined based on the closest taxonomic designator: *Fusobacterium mortiferum, Fusobacterium ulcerans, Fusobacterium nucleatum, Desulfovibrio piger, Bacteroides fragilis, Escherichia coli, Pyramidobacter piscolens, Clostridium difficile, Clostridium hylemonae, Porphyromonas asaccharolytica, Peptostreptococcus stomatis, Bilophila wadsworthia,* and *Odoribacter splanchnicus* (Table S8). Genomes from the species *F. nucleatum*, *P. piscolens*, and *O. splanchnicus* were not plotted. To determine differences in growth rates between CRC and healthy samples for all 3 studies, a Kruskal Wallis test was performed for each organism. To determine differences in growth rates across all disease states within a study for a particular organism, a Kruskal Wallis test was performed for each organism. P-value corrections were done using the Benjamini-Hochberg (BH) Procedure for final reported p-values. Uncorrected and corrected p-values can be found in Table S9. Code for statistical analysis and generation of plots can be found at https://github.com/escowley/HumanGutBacterialSulfurCycleLink.

DECLARATIONS

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

Not Applicable

CONSENT FOR PUBLICATION

All authors have read and approved the submission of the manuscript and provide consent for publication.

AVAILABILITY OF DATA AND MATERIAL

Raw reads for the original metagenomic studies are deposited at https://www.ebi.ac.uk/ena/browser/view/PRJEB12449, https://www.ebi.ac.uk/ena/browser/view/PRJEB7774, https://www.ebi.ac.uk/ena/browser/view/PRJEB10878, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA389927/, and https://www.ebi.ac.uk/ena/browser/view/PRJEB6070. MAGs were constructed previously[35] and were used for this study (see “downloading bins and accessing metadata” section of the methods). Code and HMM profiles are available at https://github.com/escowley/HumanGutBacterialSulfurCycleLink.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

ESC is an MSTP student and was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by Medical Scientist Training Program grant T32GM008692. PGW was supported at UI-Chicago by the Cancer Education and Career Development Program grant T32CA057699. JMR and HRG were supported by RO1CA204808.

AUTHORS' CONTRIBUTIONS

ESC, PGW, JMR, HRG, and KA conceptualized the project and wrote the manuscript. PGW, SM, LL, PP performed microbial sulfur pathway literature search. AB downloaded the MAGs, ran Prodigal, and generated the growth rates and related plots. ESC and PGW ran HMM searches and generated plots for figures. ESC, PGW, and KA analyzed data. ESC generated phylogenetic trees, generated metabolic reconstructions for MAGs, performed all statistical analyses, and created all final figures. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank the University of Wisconsin—Office of the Vice Chancellor for Research and Graduate Education and the University of Wisconsin-Department of Bacteriology for their support. We (PGW and ESC) are thankful to the Anantharaman, Ridlon, and Gaskins labs for feedback and assistance throughout the process. We thank Madison

Boissiere and Briawna Binion for their help with the literature search. We thank Michael Liou for statistical consulting and coaching throughout this project.

Chapter 3: Determinants of Gastrointestinal Group B *Streptococcus* Carriage in Adults

AUTHORS: Elise S. Cowley[1,2], Ibrahim Zuniga Chaves[1,2], Fauzia Osman[3], Garret Suen[1], Karthik Anantharaman[1], Andrew J. Hryckowian[4,5,^]

AFFILIATIONS:
[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3]Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA
[4]Department of Medical Microbiology & Immunology, University of Wisconsin-Madison, Madison, Wisconsin, USA
[5]Department of Medicine (Division of Gastroenterology & Hepatology), School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA

This chapter has been submitted for publication and is available as a preprint on Biorxiv:

Determinants of Gastrointestinal Group B *Streptococcus* Carriage in Adults
Elise S. Cowley, Ibrahim Zuniga Chaves, Fauzia Osman, Garret Suen, Karthik Anantharaman, Andrew J. Hryckowian bioRxiv 2023.08.17.553755; doi: https://doi.org/10.1101/2023.08.17.553755.

Supplementary material for this chapter can be found on Biorxiv.

Abstract

Background: *Streptococcus agalactiae* (Group B *Streptococcus*, GBS) is a commensal Gram-positive bacterium found in the human gastrointestinal and urogenital tracts. Much of what is known about GBS relates to the diseases it causes in pregnant people and neonates. However, GBS is a common cause of disease in the general population with 90% of GBS mortality occurring in non-pregnant people. There are limited data about the predisposing factors for GBS and the reservoirs in the body. To gain an understanding of the determinants of gastrointestinal GBS carriage, we used stool samples and associated metadata to determine the prevalence and abundance of GBS in the gut microbiome of adults and find risk factors for GBS status.

Methods: We used 754 stool samples collected from adults in Wisconsin from 2016-2017 to test for the prevalence and abundance of GBS using a Taqman probe-based qPCR assay targeting two GBS-specific genes: *cfp* and *sip.* We compared the microbiome compositions of the stool samples by GBS status using 16S rRNA analysis. We compared associations with GBS status and 557 survey variables collected during sample acquisition (demographics, diet, overall health, and reproductive health) using univariate and multivariate analyses.

Results: We found 137/754 (18%) of participants had detectable GBS in their stool samples with a median abundance of 104 copies per nanogram of starting DNA. There was no difference in GBS status or abundance based on gender. Beta-diversity, Bray-Curtis and Unweighted UniFrac, was significantly different based on carrier status of the

participant. Prior to p-value correction, 59/557 (10.6%) survey variables were significantly associated with GBS carrier status and 11/547 (2.0%) variables were significantly associated with abundance (p-value<0.05). After p-value correction, 2/547 (0.4%) variables were associated with GBS abundance: an increased abundance of GBS was associated with a decreased frequency since last dental checkup (p<0.001) and last dental cleaning (p<0.001). Increased GBS abundance was significantly associated with increased frequency of iron consumption (p=0.007) after p-value correction in multivariate models.

Conclusions: GBS is found in stool samples from adults in Wisconsin at similar frequencies as pregnant individuals screened with rectovaginal swabs. We did not find associations between risk factors historically associated with GBS in pregnant people, suggesting that risk factors for GBS carriage in pregnancy may differ from those in the general population. We found that frequency of iron consumption and dental hygiene are risk factors for GBS carriage in Wisconsin adults. Given that these variables were not assayed in previous GBS surveys, it is possible they also influence carriage in pregnant people. Taken together, this work serves as a foundation for future work in developing approaches to decrease GBS abundance in carriers.

Keywords: Human gut microbiome, cross-sectional population-based study, *Streptococcus agalactiae*, Group B *Streptococcus*

Background

Streptococcus agalactiae, also known as Group B Streptococcus (GBS), is a Gram-positive commensal bacterium in the gastrointestinal (GI) and urogenital (UG) tracts of humans [128,129]. While GBS typically colonizes asymptomatically at these body sites, it can cause illnesses such as bacteremia, pneumonia, meningitis, and soft tissue infections especially in adults and children with co-morbidities [130–132].

Much of what is known about GBS is related to diseases in pregnant individual, fetuses, and infants. GBS colonization rates of pregnant people as detected by rectovaginal swabs range from 10-35%, globally, with rates in the United States ranging from 10-30% [133,134]. Based on selective culturing of rectovaginal swabs and urine samples, the risk factors for GBS colonization are history of tobacco use, hypertension, black race, and younger age [135]. This colonization can result in urinary tract infections, chorioamnionitis, post-partum endometritis, and bacteremia in pregnant people [136,137]. Invasive disease has been associated with pregnancy loss, stillbirth, and preterm delivery [138]. GBS can vertically transmit to the neonate during vaginal delivery or infect *in utero*, where it causes early onset GBS disease (0-6 days of life) and is a risk factor for late onset GBS disease (7-89 days of life), which are the leading causes of neonatal sepsis and meningitis [139–143]. In these contexts, GBS causes over 400,000 symptomatic maternal, fetal, and infant cases globally per year [138]. Historically, it has been accepted that the GI tract is the reservoir for GBS leading to vaginal colonization in pregnant

people and work in nonpregnant females has shown that rectal colonization is a strong predictor of vaginal colonization [144,145].

While much of what is known about GBS and human disease is in the context of pregnancy, GBS presents a significant disease burden in non-pregnant adults by causing bacteremia, sepsis, and soft tissue infections [130–132]. GBS disease in nonpregnant adults has steadily increased from 3.6 cases per 100,000 in 1990 to 7.3 in 2007 and 10.9 cases per 100,000 in 2016 with even higher incidences observed among those over the age of 65 years old [130,146]. These clinical data support that the current GBS disease burden is predominantly in non-pregnant people, which is likely due to widely implemented prophylactic strategies to reduce pregnancy-related transmission of GBS to neonates [147–149].

Risk factors for GBS disease in the general population include obesity, diabetes, increased age, and black race [130,150]. Despite this morbidity and mortality, little is known about the factors that dictate GBS colonization. GBS has been documented in the GI and UG tracts of non-pregnant adults, where between 9-32% of healthy adults have detectable GBS via oral swab, urine sample, or anorectal swab [151–155]. Although GBS can be found at these body sites, it is unclear what the risk factors are for colonization at these body sites in the general population. Taken together with the diseases caused by GBS in pregnant people, fetuses, and neonates, there is a need to better understand the reservoirs of GBS and risk factors for its asymptomatic carriage, which can inform our understanding of invasive disease and transmission to at-risk patient populations.

Considering the profound consequences of GBS infections in pregnant patients and neonates, characterizing the presence of GBS in the general population will aid in our understanding of a possible reservoir and mode of transmission of GBS to pregnant patients.

To gain an understanding of the determinants of gastrointestinal GBS carriage in the general adult population, we used stool samples collected by Survey of Health of Wisconsin (SHOW) for the Winning the War on Antibiotic Resistance (WARRIOR) project to determine the frequency and prevalence of GBS in the gut microbiomes of adults in Wisconsin [156]. The samples from the WARRIOR study are from a representative cross section of adults in Wisconsin. The study collected extensive data from the participants on demographics, health, and diet. Understanding the prevalence of GBS in the general population and factors that may be associated with a higher abundance of GBS will further our understanding of the pathophysiology of GBS and inform new therapeutic options. Using samples from the general population will help us identify risk factors for GBS carriage that can inform our understanding of GBS ecology and help guide alternative approaches to coping with this pathogen in a variety of at-risk human populations.

Methods

Stool samples

Human stool samples were previously collected and banked by SHOW from 2016-2017 as part of the WARRIOR project, which was reviewed and approved by the University of

Wisconsin-Madison Institutional Review Board (Protocol #2013-0251) [156,157]. For our study, we used 754 WARRIOR stool samples from participants who agreed to have their samples used in future research. Our study was reviewed and approved by the University of Wisconsin-Madison Institutional Review Board (Protocol #2021-0025).

DNA extractions from stool

Two methods were used to extract DNA from the stool samples. A subset of samples (455/754, 60%) had DNA remaining from a previous extraction and were used in this study [156]. Briefly, DNA was extracted using a bead-beating protocol with additional enzymatic lysis containing mutanolysin, lysostaphin, and lysozyme to help lyse Gram-positive bacterial cell walls. For the remaining 299 stool samples, in order to complete extractions in a high throughput manner, we extracted DNA using the DNeasy PowerSoil Pro kit (Qiagen) following the manufacturer's instructions, starting with 50-250 mg stool and using a TissueLyzer II at 4°C for the homogenization followed by elution of DNA in 75 mL of solution C6.

DNA extractions from bacterial cultures

For positive controls for quantitative polymerase chain reaction (qPCR) assays, we grew three strains of *S. agalactiae* in tryptic soy broth (Neogen NCM0004A) aerobically overnight at 37°C, including: *S.agalactiae* COH1, *S. agalactiae* 10/84, and *S. agalactiae* A909 (obtained from Katy Patras, Baylor College of Medicine). For a negative control, we grew *Clostridioides difficile* 630 in brain heart infusion broth (Neogen NCM0016A) anaerobically overnight at 37°C. One milliliter of each overnight culture was pelleted and

the supernatants were removed. The DNA from the remaining cell pellets were extracted using the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's instructions for Gram-positive bacteria.

DNA quantification

All DNA used in qPCR assays was quantified in duplicate on 96 well plates using the Quant-iT double stranded DNA broad range kit (Invitrogen) with a standard curve from 0 ng/mL – 100 ng/mL DNA. Fluorescent signal was read with Synergy HTX plate reader with excitation of 485/20 and emission of 528/20 and an auto-gain setting. Final DNA quantities in the samples were determined against the standard curve and by averaging duplicate readings.

qPCR to quantify GBS prevalence and abundance in stool samples

We developed a multiplexed Taqman-based qPCR approach targeting two GBS-specific genes, *cfb* and *sip.* These genes were previously used in qPCR protocols to identify the bacterium [158–161] and the assay we developed was based on a protocol previously created for *sip* [162].

For *sip,* we used the primers: 5'-CAG CAA CAA CGA TTG TTT CGC C-3' and 5'-CTT CCT CTT TAG CTG CTG GAA C-3', targeting a 171 base pair region. The Taqman probe for *sip* was: 5'-FAM-AGA CAT ATT - ZEN - CTT CTG CGC CAG CTT TG-3IAkfQ-3'. For *cfb,* we used the primers: 5'-GAA ACA TTG ATT GCC CAG C-3′ and 5′.-AGG AAG ATT TAT CGC ACC TG-3′, targeting a 99 base pair region. The Taqman probe for

cfb was 5'-HEX-CCA TTT GAT AGA CGT TCG TGA AGA G-3BHQ-1 -3'. We ran each reaction in triplicate with both *sip* and *cfb* probes and included positive controls (DNA from three GBS strains), a negative control (DNA from *C. difficile*), and we created a standard curve for each gene from $5x10^{-3}$ ng/µL to $5x10^{-11}$ ng/µL using serial dilutions of synthetic copies of target genes (gBlocks, IDT). All reactions totaled 19.25 µL and included 5 µL of DNA (either extracted from cultures of bacteria, extracted from stool samples, or synthetic DNA for the standard curve), 7.5 µL TaqPath qPCR Master Mix, CG, 0.75 µL of a 10 µM stock each primer (1 reverse and 1 forward for each gene target for 4 total primers), 0.3 µL of a 10 µM stock of each probe, and 3.15 µL water. Samples were held at 96°C for 5 minutes followed by 50 cycles of 96°C for five seconds, 58°C for 10 seconds, and 72°C for 20 seconds [162]. All qPCR was performed on an Applied Biosystems QuantStudio 7 instrument.

We determined GBS status by comparing the threshold cycle of each sample to the negative control. Samples with threshold cycles below the negative control for both *sip* and *cfb* were considered negative for GBS. Samples with threshold cycles above the negative control for both genes were positive for GBS and the abundance was determined by comparing cycle threshold against the standard curve.

16S rRNA marker gene analysis

We used 16S rRNA marker gene data previously generated for the WARRIOR samples [156]. Briefly, the V4 region of the 16S rRNA gene was sequenced on an Illumina MiSeq

using 2 × 250 paired-end reads at the University of Wisconsin Biotechnology Center. Negative controls were included during each step of extraction and amplified and sequenced with the same protocol as described above.

The resulting fastq files were processed with the software QIIME 2 v2021.4 [163]. Demultiplexed raw sequences were imported using the Casava 1.8 format and denoised using DADA2 [164] (via qiime-dada2 plugin) to generate a feature table containing amplicon sequence variants (ASV). ASVs were aligned with MAFFT to construct a phylogenetic tree with fasttree [165]. Taxonomy was assigned using the classify-sklearn naive Bayes taxonomy classifier [166] against the Silva_138 database for 16S rRNA genes [167]. A feature and a taxonomy table, together with the phylogenetic tree were imported in R 4.1.2 as a phyloseq object [168] for further analysis. A total of 756 samples, including 29 negative controls and 727 stool samples, were processed with R. Contaminationwas accounted for by eliminating features based on the prevalence of ASVs in the negative controls using the Decontam package [169], and by removing eukaryotic, chloroplast, mitochondrial or unassigned sequences. Finally, samples without GBS-associated data or with less than 5000 reads were removed. The resulting 693 samples were processed as follows.

For the alpha and beta diversity analysis, samples were rarefied to an even depth of 8396, corresponding to the minimum read count in the dataset. Alpha diversity metrics (Shannon, inverse Simpson, and total observed ASVs) metrics were calculated with Phyloseq. Beta diversity was used to quantify the dissimilarities between the samples.

First, we calculated 3 different types of distance matrices (Bray-Curtis, weighted and unweighted UniFrac and then plotted the ordinations using a principal component analysis (PCoAs). PCoAs were plotted for GBS carrier status, and the logarithmic transformation of the average copy number of GBS. Moreover, to confirm our findings, we extracted the distances for each matrix and plotted the average distance between and within sample types. Average distances were plotted as box plots and significant differences based on the carrier status of GBS were tested using a one-way ANOVA and Tukey's HSD post-hoc test.

Linear correlations with GBS prevalence

To evaluate associations with alpha diversity, simple linear regressions were calculated for prevalence and copy number against the Inverse Simpson's, Observed features, and Shannon's indices. For these analyses, we included only the samples where GBS was detectable via qPCR.

Differential taxonomic abundance

Differential taxonomic abundance was obtained using the ANCOM-BC package in R using default parameters and Benjamini-Hochberg procedure for P-value corrections [170]. Linear discriminant analysis effect size (Lefse) [171] was done using the R package microbiomeMarker [172]. Finally, the QCAT package was used to evaluate microbiome markers using copy numbers as it allows for continuous variables [173].

Random forest classification

To identify taxa that discriminate between the presence and absence of GBS, we used a random forest classifier algorithm from the random forest R package [174]. In summary, we trained and tested our model using the "out of bag" (OOB) error to estimate our model error. The number of trees was set to 1000 and only ASVs with relative abundances ≥ 0.01% were included as input. The classifier was trained on a random selection of 70% of the database composed of 693 samples and 1099 ASVs and validated using the remaining 30%. Finally, prediction performance was measured by the OOB error rate and the mean decrease in Gini coefficient, a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest, for each ASV.

Metadata variable acquisition and selection

We acquired participant metadata from SHOW, which biobanks the samples and manages the data repository associated with the samples. For our analysis of all samples, we selected relevant variables that fit into three categories: demographics, health, and diet. For participants who identified as female, we also analyzed data on reproductive health. Data were collected for some study participants across multiple years, we selected variables relevant for the years the stool samples were collected (2016-2017).

Analysis of WARRIOR study metadata

Differences in GBS outcome by self-identified gender

To determine differences in carriage of GBS by self-identified gender, we used Pearson's Chi-squared test with Yates' continuity correction. To determine differences in abundance of GBS by self-identified gender and for those individuals with detectable GBS, we used Welch's two sample t-test with unequal variances. We used R version 4.2.2 to run these comparisons.

Binary outcome

To determine associations between a binary outcome of GBS (presence or absence of GBS in the stool) and other predictors, we used a logistic regression model and reported odds ratios and confidence intervals. We conducted a univariate analysis first (each predictor in a bivariate analysis), then significant predictors from the independent models were used to construct three multivariate (adjusted) logistic models. Model 1 examined significant predictors from the univariate analysis, model 2 examined predictors commonly associated with GBS carriage or disease risk, as shown in the literature (African American, Type 2 Diabetes Mellitus, and smoking status)[135], and model 3 combined predictors from models 1 and 2. Odds ratio plots were created using log transforms of the odds ratios and 95% confidence intervals. Variables with missing data were addressed using the complete case analysis method. In this instance, any subject with no observations was removed from the specific analysis being conducted. After adjusting for missing data, a total of 555 variables were used for the binary outcome comparisons All analyses were conducted using STATA version 17 (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC).

Continuous outcome

Predictors associated with the abundance of GBS for participants with detectable GBS were examined using a linear regression model with coefficient estimates and 95% confidence intervals. We used a similar approach used above see Binary outcome above. Each predictor was independently tested and determined to either be fit for a multivariate model or not included in the model. We constructed the same three models specified above for the binary outcome. We constructed odds ratio plots as described above. Variables with less than 5% missing data were negligible and used as is, variables with up to 40% missing data were inputted using multiple imputation methods and variables with greater than 50% missing data were excluded from the analysis altogether. After adjusting for missing data, a total of 547 variables were used for the continuous outcome comparisons. All analyses were conducted using STATA version 17 (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC).

Subgroup analysis

A sub-group analysis in females was conducted using the same statistical tests for either the binary outcome or the continuous outcome with the same independent variables for each outcome. All analyses were conducted using STATA version 17

(StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC).

Diagnostics and missing data

We ran collinearity diagnostics to identify independent variables with significant relations as this is disruptive to models. We used a variance inflation factor (VIF) to measure the tolerance of variance. Variables with a VIF value greater than 10 were removed from the models. Secondary pairwise correlations were conducted to supplement our decision to remove collinear variables. We corrected for multiple comparisons effect on the p-value using a false discovery rate (FDR) using the Benjamini-Hochberg (BH) procedure. All p-values were assigned a rank within the outcome group and a critical value using the BH procedure was calculated using false discovery rates of 5% and 10% to discover which worked best for the data. Critical values less than the significant original p-values were also regarded as significant. We graphed the odds from the multivariate analysis using coefficient plots with confidence intervals. All analyses were conducted using STATA version 17 (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC).

Results

Participant Demographics

For this study, we used stool samples from 754 unique individuals across the state of Wisconsin, collected between 2016-2017. Our samples came from 432 (57%) self-

identified females and 322 (43%) self-identified males with ages ranging from 18 to 57 (**Table 1**).

Table 1. Baseline demographics of participants.

Characteristic	n(%) Total participants = 754
Self-Identified Gender	
Female	432 (57%)
Male	322 (43%)
Age (years)	
Minimum	18
Maximum	94
Median	57

GBS is present in gut microbiomes of at widely variable abundances

Using a qPCR-based method, we found that 137/754 (18%) of all participants had detectable GBS in their stool samples with 79/432 (18%) of self-identified females and 58/322 (18%) of self-identified males having detectable GBS (**Figure 1A**). There was no significant difference in proportion of either gender based on GBS status. For samples with detectable GBS, we quantified the amount using a standard curve. The samples contained between 5 to 6,800,532 copies of GBS target DNA per starting nanogram of total DNA (**Figure 1B**). There was no difference in abundances of GBS in the stool based on the gender of the study participants.

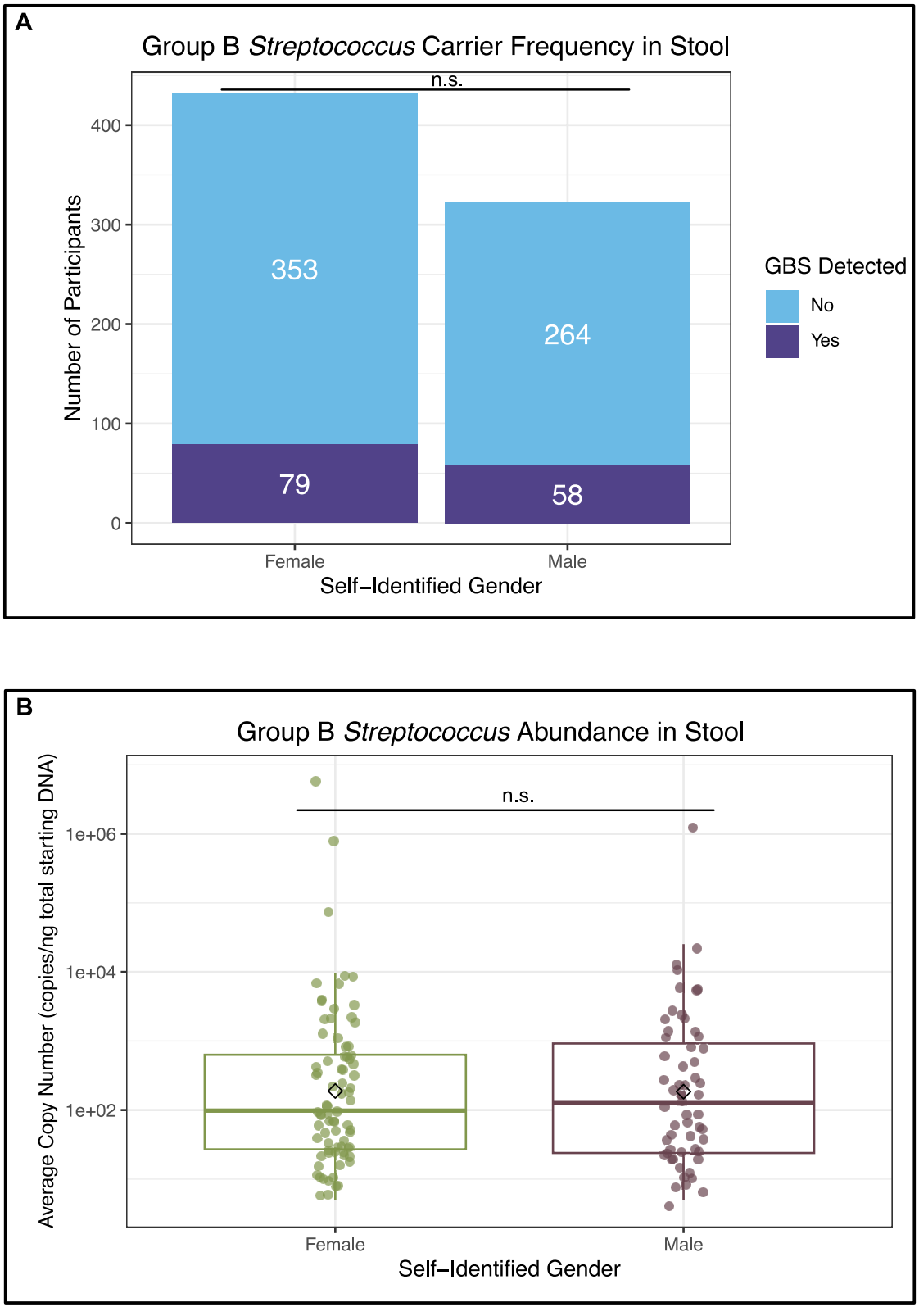


Figure 1. Carriage of GBS and abundance by self-identified gender. A. Presence or absence of GBS in stool samples as determined by qPCR. No significant difference was

found in the prevalence of GBS between genders as indicated by “n.s.” as determined by Chi-squared tests. **B.** Abundance of GBS in stool samples with identifiable GBS as determined by qPCR. Each colored point represents GBS abundance in an individual sample, diamonds represent the mean copies of GBS for each gender and the box plots represent median, interquartile ranges of GBS abundance, and standard deviation for each gender. The median copy number of GBS per starting amount of DNA was 104 and the average was 66,856. The median copy number for self-identified females was 98, with a mean of 97,621, a standard deviation of 768,659 and an interquartile range (IQR) of 605. For males, the median copy number was 126, mean 24,951, standard deviation of 178,687, and an IQR of 898. No significant difference was found in GBS abundance among male and female carriers as indicated by “n.s.” as determined by Welch’s two sample t-test with unequal variances.

Differences in gut microbiome composition of WARRIOR study participants based on GBS carrier status

Of the 754 samples, 693 (92%) had available 16S rRNA marker gene sequencing performed in a previous study [156]. In this subset of samples, no significant correlations were found between GBS presence or abundance and alpha diversity metrics (Inverse Simpson, observed ASV, or Shannon) using linear regression analysis (**Table 2**).

To evaluate differences in bacterial communities we used a PCoA with different distance matrices (Bray-Curtis, Weighted and Unweighted UniFrac) grouped by GBS

carriage and copy number (**Table 2**). Using a PERMANOVA test, we found that the bacterial communities were significantly different according to the presence/absence of GBS in PCoAs with the Bray-Curtis and Unweighted UniFrac distance matrices. Moreover, we found that bacterial communities were significantly different for the abundance of GBS with the log transform for the PCoAs with the Bray-Curtis dissimilarity matrix (**Table 2**).

Additionally, we determined differences in the average beta diversity distance by group based on GBS carrier status. We found statistically significant differences in the average Bray-Curtis, Weighted and Unweighted UniFrac distances across the three comparisons: intragroup distance between all samples with GBS compared to intragroup distance for samples without GBS ($p<0.001$ in all three indices), intergroup distance differences for samples with and without GBS compared to intragroup distances between samples without GBS ($p<0.001$ across all three indices), and finally intergroup distance differences for samples with and without GBS compared to intragroup distances between samples with GBS ($p\leq0.001$ across all three indices) (**Figure 2**).

We compared differential abundances of microbiome members at different taxonomic levels as a function of GBS carrier status. We found no differences at the phylum or ASV level, however at the genus level, we found that the relative abundance of two genera, *Ruminococcus* (p=0.0003 uncorrected, p=0.055 corrected) and *Monoglobus* (p=0.00057 uncorrected, p=0.057 corrected) trended towards being significantly

differentially abundant when GBS is present, specifically both genera trended toward being more abundant when GBS is present.

To gain additional insights into microbes that differentiate GBS carriers from non-carriers, we ran random forest classifiers to identify community signatures predictive of GBS carrier status. This analysis could reliably predict GBS negative individuals, but not GBS positive individuals (error rates = 0.0035 and 0.992, respectively; **Supplemental Table 1**). Several taxa were identified through the random forest classification as good predictors of GBS status, including *Anaerostripes , Ruminococcus ,Collinsella, Dorea, Agathobacter, Blautia, Faecalibacterium, Bacteriodes, and Anaerostripes.* (**Supplemental Table 1**).

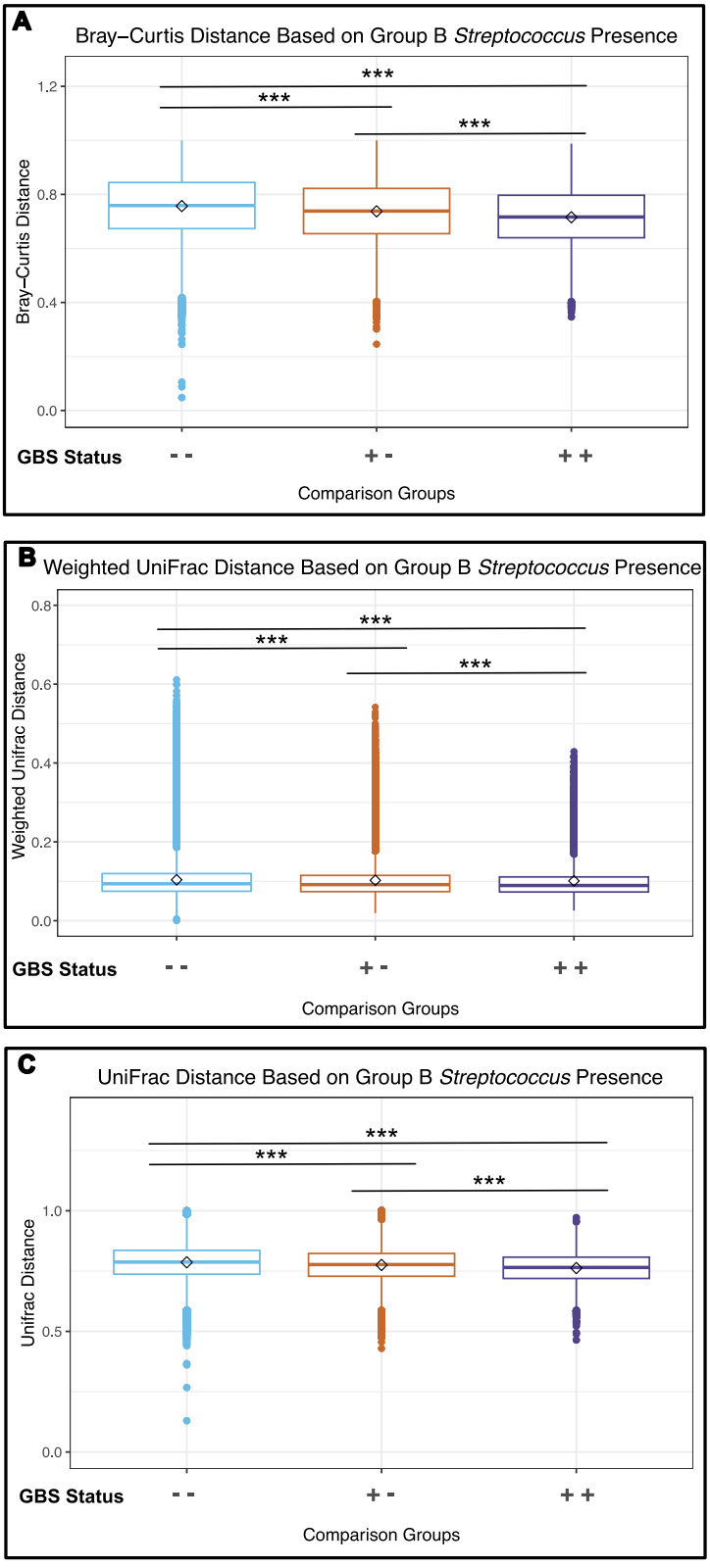


Figure 2. Beta diversity distance differences between GBS groups. Beta diversity distances for inter and intra group comparisons are plotted for three beta diversity indices. Plotted are the intragroup distances for samples without GBS (- -) intragroup distances for samples with GBS (+ +), and the intergroup distances for samples with and without GBS (+ -). The average distance for each of the three groups were compared to test for differences in beta diversity. The three beta diversity indices used were: **A.** Bray-Curtis, **B.** Weighted UniFrac, and **C.** UniFrac. The diamonds indicate the mean for each category. *** indicates a p-value ≤ 0.001.

Table 2. Correlations between alpha and beta diversity metrics and GBS carriage or abundance.

Diversity Metric	Presence or Absence p-value	Abundance p-value
Alpha Diversity		
Inverse Simpson	0.449	0.999
Observed ASV	0.949	0.220
Shannon	0.785	0.636
Beta Diversity		
Bray-Curtis	0.046*	0.176
Log transform		0.013*
Weighted UniFrac	0.205	0.810
Log transform		0.290
Unweighted UniFrac	0.049*	0.183
Log transform		0.064

*indicates p-value<0.05

Differences in characteristics of WARRIOR study participants based on GBS presence and abundance

We completed univariate and multivariate analyses on study participant metadata to understand which host characteristics were associated GBS carriage and which host characteristics were correlated with GBS abundance in carriers. For our analysis, we selected relevant variables that fit into three categories: demographics, health, and diet. For participants who identified as female, we also analyzed data on reproductive health which were not collected for participants who identified as male. The median and interquartile range for all variables considered in our multivariate statistical analysis are listed in **Supplemental Table 2**.

From our univariate analysis for carrier status (GBS present or absent in the stool), we found 59/557 (10.6%) variables were significantly associated with GBS carrier status (p-value < 0.05), but no variables were significantly associated after p-value corrections

with FDR (**Supplemental Table 3**). Similarly, for our three multivariate models, we found no significant associations after p-value correction (**Figure 3, Supplemental Table 4**).

For the univariate analysis of the abundance of GBS from participants with detectable GBS in their stool, we found 11/547 (2.0%) variables were significantly associated with GBS abundance (p-value < 0.05) prior to FDR correction. After FDR correction, we found 2/547 (0.4%) variables associated with GBS abundance: an increased abundance of GBS was associated with an increased time since last dental checkup ($p<0.001$) and last dental cleaning ($p<0.001$) (**Supplemental Table 5**). For the multivariate analysis, we found that in model 1, which included all the variables that were significant from the univariate analysis prior to p-value correction, higher GBS abundance was significantly associated with an increased frequency of iron consumption ($p=0.007$) after p-value correction. No other variables were significantly associated with GBS abundance in any of the other multivariate analyses (**Figure 4, Supplemental Table 6**).

Subgroup Analysis of Associations of GBS for Self-Identified Females

Since GBS has historically been studied in pregnant people, we performed a univariate sub-group analysis of all self-identified females to investigate if there was a difference in the sub-population that were not present in the general population. Similar to the general population, in self-identified females, we tested for variables that differentiate GBS carrier status and for GBS abundance in the subset of female study participants

with detectable GBS. For carrier status, we found 6/556 (1.1%) variables associated with GBS presence or absence prior to p-value correction and we did not find any significant associations after p-value correction with FDR (**Supplemental Table 7**). We found 10/497 (2.0%) variables associated with GBS abundance prior to p-value correction and 2/497 (0.40%) significantly associated after p-value correction (**Supplemental Table 8**). For abundance, similar to the analysis with all participants, an increased abundance of GBS was associated with a decreased frequency since last dental checkup ($p<0.001$) and last dental cleaning ($p<0.001$), after p-value correction (**Supplemental Table 8**). Unlike in the general population, we did not observe a correlation with GBS abundance and frequency of iron consumption.

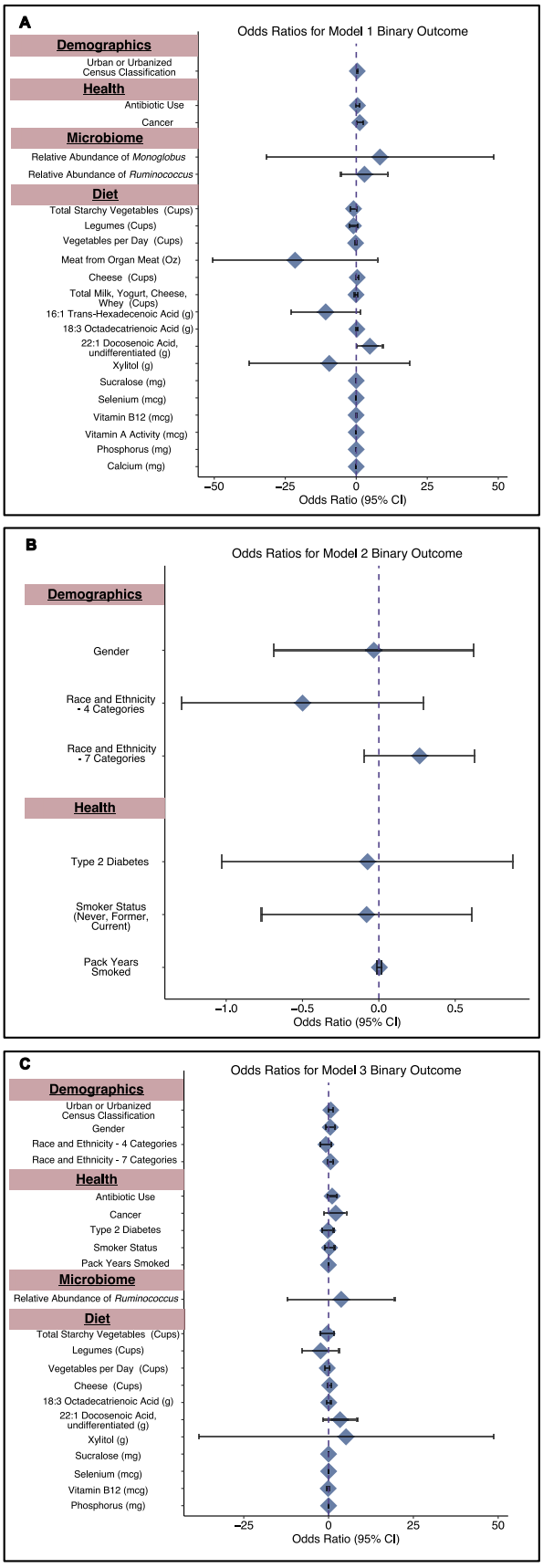


Figure 3. Odds ratios for multivariate models comparing presence or absence of GBS and participant characteristics. The diamonds indicate the odds ratios for each variable with error bars for the 95% confidence interval. **A.** Model 1 compared significant predictors from the univariate analysis prior to FDR correction with GBS presence or absence. **B.** Model 2 compared predictors commonly associated with GBS from previous studies and GBS presence or absence. **C.** Model 3 combined predictors from Models 1 and 2 with GBS carrier status. **indicates p-value ≤ 0.001 after FDR correction

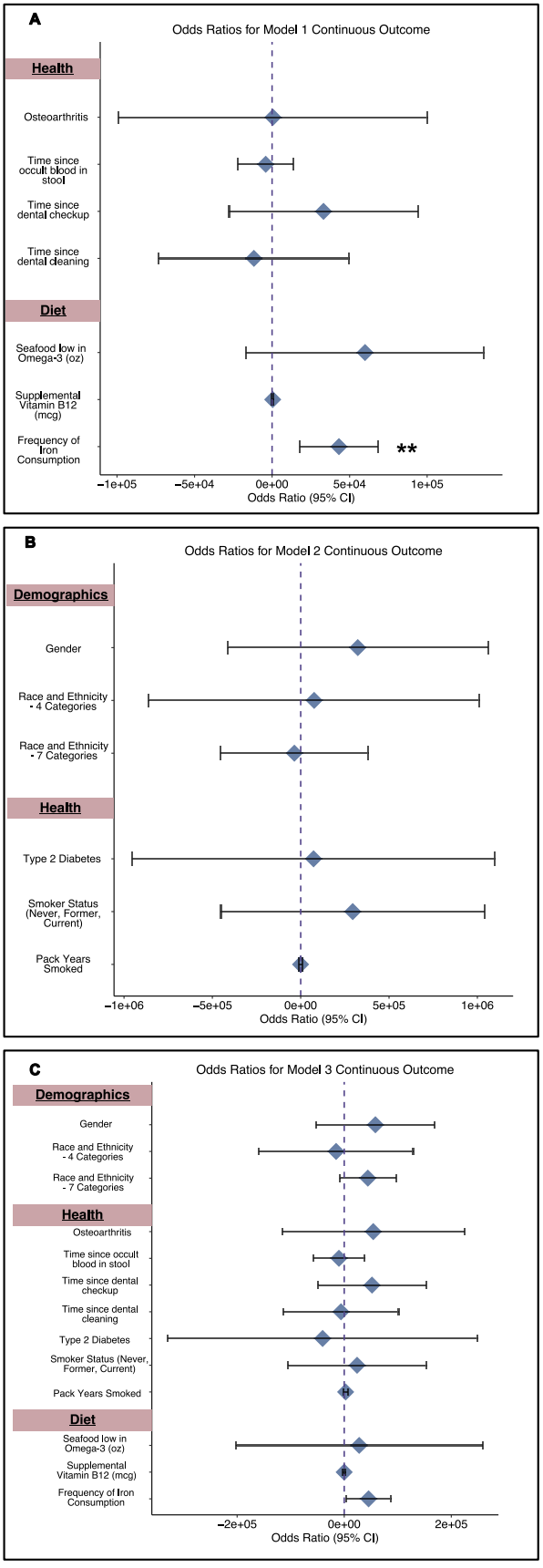


Figure 4. Odds ratios for multivariate models comparing GBS prevalence and participant characteristics. The diamonds indicate the odds ratios for each variable with error bars for the 95% confidence interval. **A.** Model 1 compared significant predictors from the univariate analysis prior to FDR correction with GBS prevalence. **B.** Model 2 compared predictors commonly associated with GBS from previous studies and GBS prevalence. **C.** Model 3 combined predictors from Models 1 and 2 with GBS prevalence. **indicates p-value <0.01 after FDR correction

Discussion

In this study, we used a biobank of 754 stool samples and associated data collected from adults in Wisconsin to better understand the host and microbiome-based factors that influence gastrointestinal GBS carriage.

We found that GBS is present in the stool from a representative cross-sectional sampling of adults in Wisconsin at rates like what is seen in pregnant individuals via rectovaginal swabs. We found that the abundance of GBS varied by orders of magnitude between individuals with GBS and there was no difference in the carrier frequency or abundance based on gender. These data indicate that GBS is common in the distal GI tracts of the general population, providing further evidence that the distal GI tract is an important reservoir for GBS in the human body [144,155].

We carried out 16S rRNA marker gene analysis of the samples in the context of GBS colonization and found that regardless of index used (Bray-Curtis, Weighted UniFrac, Unweighted UniFrac), beta-diversity was significantly different based on GBS carrier status. We did not find robust correlations between alpha-diversity or individual microbes that predicted GBS carriage or GBS abundance in individuals with GBS. It is possible that multiple microbiome configurations or functionally redundant microbiome members influence GBS carriage or that transient GBS carriage, as has been observed for pregnant people, obscures this assessment [175].

In addition, to identify host-centric variables that influence GBS carriage, we leveraged the extensive participant metadata collected as part of the WARRIOR study. While we did not find many statistically significant variables that correlated with GBS carrier frequency or abundance after p-value correction, prior to p-value correction, there were 59/557 (10.6%) variables for GBS carrier status and 11/547 (2.0%) for GBS abundance that were statistically significant ($p<0.05$). The two most robust associations we found after p-value corrections were between lack of dental care and higher iron consumption and increased GBS abundance. *Streptococci* are known inhabitants of the upper GI tract, including the oral cavity [155,176]. Our finding that a lack of dental care is associated with higher GBS burdens in the stool implies that poor dentition leads to an increase in *Streptococci* in the oral cavity, which can subsequently be observed more distal in the GI tract [177]. Iron is an essential nutrient for almost all living organisms, including bacteria, which have evolved strategies to scavenge environmental iron. These strategies enable bacteria to compete with other microbes within microbiomes and have been demonstrated to be important to overcome human immune defenses [178]. In particular, GBS encodes siderophores which aid in environmental iron acquisition [179] and elevated oral iron consumption may favor siderophore-dependent iron acquisition by GBS and increased GBS fitness in the distal GI tract. Alternatively, it is possible that iron indirectly impacts GBS fitness by influencing the abundance of microbes that compete or cooperate with it in the distal GI. Supportively, oral iron consumption has been previously shown to increase the amount of iron available to the gut microbiome and influence its composition [180,181].

Of note, we did not find associations between variables historically associated with GBS carriage in pregnant people in our study population. Specifically, neither GBS carriage nor GBS abundance in carriers was impacted by race, smoking status, age, or diabetes status (Figure 5). This challenges our current understanding of the risk factors for GBS carriage and highlights two possibilities for future study. First, it is possible that risk factors for GBS carriage in pregnancy are different than the risk factors in the general population. Second, it is possible that iron consumption and dental hygiene are risk factors for GBS carriage in pregnant people, but studies have not been done to address these connections. Therefore, further study of both pregnant and non-pregnant adults, informed by this work, will fill these important gaps in understanding relating to the life cycle of GBS. This understanding is a useful starting point developing new interventions, beyond antibiotics, to de-colonize GBS carriers and to prevent GBS infections in at-risk populations. These considerations are relevant given the rising incidence of antibiotic resistance in GBS and our emerging understanding of the collateral damage antibiotics have on human microbiomes [182–187].

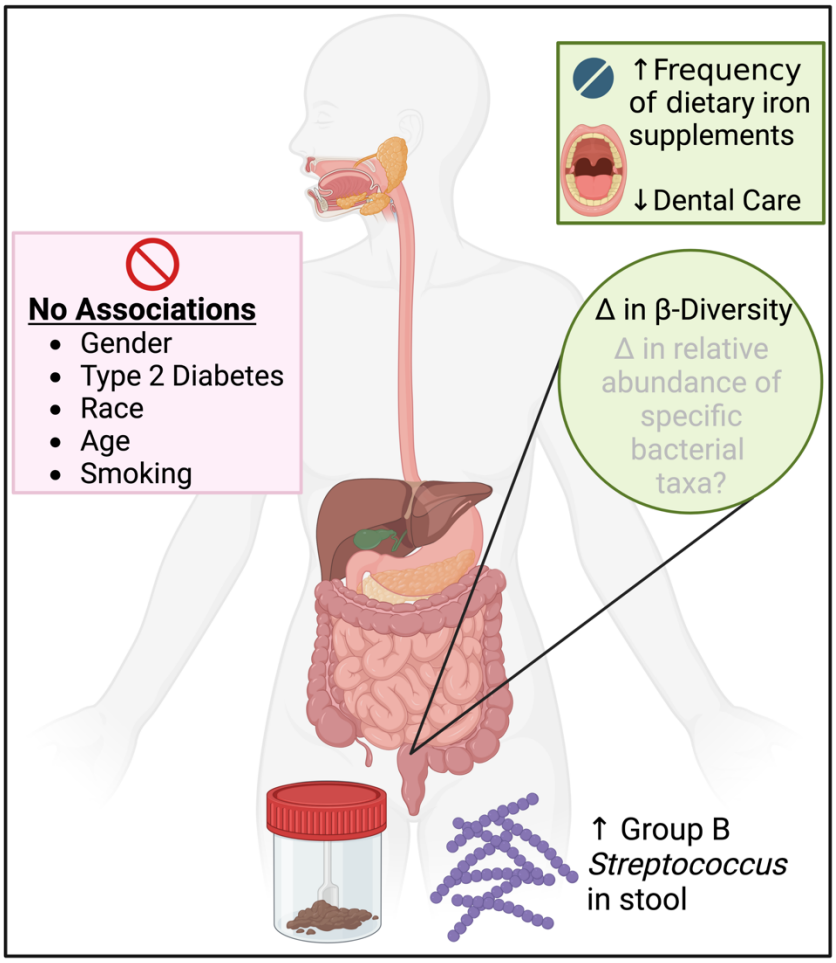


Figure 5. Overview of associations found with GBS in stool from study participants. No associations were found with historically known risk factors for GBS in pregnant individuals including gender, race, Type 2 diabetes, age, or smoking status. A higher burden of GBS in the stool was associated with decreased dental hygiene and an increase in frequency of dietary iron supplements. We found a change in beta-diversity of the gut microbiome was different depending on GBS carrier status. Two genera, *Ruminococcus* (p=0.0003 uncorrected, p=0.055 corrected) and *Monoglobus* (p=0.00057 uncorrected, p=0.057 corrected) trended towards being more abundant when GBS is present. Created with BioRender.com under agreement #DL25UX5VMX.

Limitations of our study include lack of longitudinal samples and using retrospective data. Since we used a biobank, we were limited to the survey questions that were asked with the initial studies and not able to add our own. Our statistical analyses demonstrate

correlations and we are unable to provide causation within this study. Our final p-values were penalized for the high number of metadata variables we investigated. Strengths of our work include the large number of participant samples (754) from a large age range and backgrounds. Another strength is the extensive amount of survey data we were able to compare to GBS carrier status and abundance.

Our work provides a starting point to begin to understand risk factors for GBS carriage in the general population and in pregnant people using basic, translational, and clinical research approaches. Future work could expand on our findings including associations we found significant prior to p-value corrections. In addition, future work could use metabolomics-based approaches to identify microbiome-produced metabolites that correlate with GBS carriage. In addition, longitudinal sampling of study participants would provide important clues into the microbial and metabolic factors that influence transient carriage. Other future directions for this work include the use of experimental models to gain mechanistic insights into the associations we observe and impacts on GBS pathogenesis. This future work could include mouse models to determine the extent to which GBS colonizes distal GI and how dietary iron supplementation influence the abundance of GBS and other interacting microbiome members. Additional work could explore the extent to which transient distal GI colonization is influenced by oral GBS colonization, considering previous observations of high prevalence of oral GBS colonization [155].

Declarations

Abbreviations

Group B *Streptococcus* (GBS), gastrointestinal (GI), urogenital (UG), Survey of Health of Wisconsin (SHOW), Winning the War on Antibiotic Resistance (WARRIOR), quantitative polymerase chain reaction (qPCR).

Ethics approval and consent to participate

Human stool samples were previously collected and banked by the Survey of Health of Wisconsin from 2016-2017, which was reviewed and approved by the University of Wisconsin-Madison Institutional Review Board (Protocol #2013-0251). For this study, we used 754 of those stool samples from participants who agreed to have their samples used in future research. Our study was reviewed and approved by the University of Wisconsin-Madison Institutional Review Board (Protocol #2021-0025).

Availability of data and materials

Raw data and code for qPCR, 16S rRNA amplicon gene, and univariate and multivariate modeling analysis and figure creation along with metadata used in this study are available at: https://github.com/escowley/GroupBStreptococcus_HumanGut. All 16S rRNA sequences are deposited in the NCBI SRA under BioProject ID PRJNA999362. All SHOW and WARRIOR questionnaires and surveys used to obtain metadata from participants are available upon request at https://www.med.wisc.edu/show/data-service-center/.

Competing interests

The authors declare no competing interests.

Funding

ESC is a Medical Scientist Training Program (MSTP) student and was supported in part by MSTP grant T32GM140935 and by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at the University of Wisconsin-Madison. This work was funded by a Centers for Disease Control and Prevention Cooperative Agreement to the University of Wisconsin Prevention Research Center (1U48DP006383 to AJH), a gift from Judy and Sal Troia (AJH), and startup funding from the University of Wisconsin-Madison (AJH).

Authors' contributions

ESC and AJH conceptualized the project. ESC performed all laboratory experiments. ESC analyzed qPCR data. IZC analyzed 16S rRNA data. ESC and FO performed statistical analyses. ESC and IZC created the beta diversity figure. ESC created the remaining figures and all tables. GS, KA, and AJH provided feedback. IZC and FO provided written methods sections. ESC and AJH wrote the manuscript. All authors read, edited, and approved the manuscript prior to submission.

Acknowledgements

We thank Michael Liou and Courtney Van Den Elzen for statistical consulting, guidance, and constant coaching throughout this project. We thank Emmanuel Sampene and Kris

Sankaran for additional statistical discussions. We thank Patrick Carney, Liz Townsend, and Shelby Sandstrom for technical support and advice with qPCR analysis. We thank the staff of the Survey of Health of Wisconsin for assistance with sample and metadata acquisition, especially Maria Nikodemova, Andy Bersch, and Hannah James.

Chapter 4: Differences in Gut Bacterial and Phage Compositions in People with Relapsing Remitting Multiple Sclerosis

TITLE: Differences in Gut Bacterial and Phage Compositions in People with Relapsing Remitting Multiple Sclerosis

AUTHORS: Elise S. Cowley[1,2], Sanjay K. Shukla[3], Karthik Anantharaman[1]

AFFILIATIONS:
[1]Department of Bacteriology, University of Wisconsin-Madison
[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison
[3]Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield Health System

[^]Address correspondence to Karthik Anantharaman, karthik@bact.wisc.edu

ABSTRACT:

Introduction: Multiple sclerosis (MS) is a chronic, autoimmune inflammatory disease of the central nervous system that targets the myelin sheath of neurons, producing symptoms such as numbness, dizziness, vertigo, vision changes, and motor weakness. MS is the most common non-traumatic cause of neurological disability in young people affecting approximately over 2 million people in the world. The most common phenotype of MS is relapsing remitting MS (RRMS), which is characterized by exacerbations and periods of recovery. The etiology of MS is considered multifactorial with genetic and environmental factors influencing MS development. The gut microbiome is thought to play a role in MS. It is unclear if it is a trigger, a driver of disease pathogenesis, or a symptom of the disease. Studies have shown differences in the composition and diversity of people with and without RRMS. We aim to understand the association between the role of the gut microbiome, specifically the bacterial and viral members of

the community, in RRMS compared to healthy controls by characterizing the composition and functional capacity of the microbiome. bacterial and viral

Methods: We used stool samples from 75 individuals with RRMS, 44 on a disease modifying therapy, 31 not on a disease modifying therapy, and 26 controls from a cohort of participants at Marshfield Clinic in Wisconsin collected between 2018-2021. We performed metagenomic sequencing of DNA extracted from stool. We trimmed adapters, quality filtered, assembled, and binned the metagenomes. We taxonomically identified assembled bacterial genomes and built a phylogenetic tree using the genomes. We identified functional capacity of the bacterial genomes using METABOLIC and compared differences in metabolic pathways based on participant type. We identified putative phages in each metagenome using ViWrap and compared viral taxa based on participant type.

Results: We found the bacterial composition in participants with RRMS and controls were similar and largely contained phyla common to the gut environment: *Bacteroidota* and *Firmicutes.* We characterized the phage composition of the stool samples and found the majority of the phages in all groups were from the most abundant class of gut phages: *Caudoviricetes.* We found evidence of 1-3 unique phages for each of the two MS groups, but no phages were unique to the controls. We found the Semi-phosphorylative Entner-Doudoroff pathway, gluconate → glycerate-3-phosphate, was less likely to be present in in the metagenomic assembled genomes of participants with MS on disease modifying therapy than those who are not on disease modifying therapy.

We also found that tryptophan biosynthesis, chorismate → tryptophan, was more likely to be present in bacterial genomes of individuals with RRMS not on disease modifying therapy compared to controls.

Discussion: The few differences we found in bacterial metabolic capabilities in this study provide promising future directions while the phage composition and functional potential remains largely unexamined and an opportunity for future work. Continued work on the gut microbiome and pairing it with clinical metadata is essential to begin to uncover the role of the gut microbiome in MS.

Introduction

Multiple Sclerosis (MS) is a chronic, immune mediated inflammatory disease of the central nervous system which is characterized by demyelination of neurons that results in plaques. Besides traumas, MS is the most common cause of neurological disorders in young people with a mean age of onset between 28-31 years old and it affects more women than men [188–191]. MS is a heterogeneous disorder and can manifest in different phenotypes, one of which is relapsing remitting MS (RRMS) that manifests as exacerbations with neurological symptoms for a defined period followed by recovery. Between flares, patients with RRMS, do not experience signs and symptoms of MS [192,193]. Autoimmune factors and genetic susceptibility are both thought to play a role in the pathogenesis of MS [193,194]. Environmental factors, including viral infections such as EBV, geographic location, and sunlight and vitamin D levels, have been shown to play a role in increased risk for MS [192,195–197].

Recently, studies have demonstrated that the gut microbiome composition differs between patients with MS and healthy controls leading to the hypothesis that the microbiome may play a role in the pathogenesis of MS [198,199]. This is further supported by the fact that most immune cells in the body reside in the gut and MS has known immune system dysregulation. Antibiotic treated and germ free mouse are not susceptible to the animal model of MS, spontaneous experimental autoimmune encephalomyelitis, providing evidence for the importance of the gut microbiome in the development of multiple sclerosis [200–202]. Mouse models of MS have further show that treatment with a probiotic halted the autoimmune progression of MS [198,199,203]. Most gut

microbiome research has focused on bacteria, but bacteriophages or phages have garnered recent attention for their abundance in the gut [204]. Changes in phage functional landscapes have recently been associated with another autoimmune disease, rheumatoid arthritis, and we propose a similar analysis for patients with MS [205].

While there have been several bacterial composition studies of the gut microbiome of individuals with MS in tandem with immunological studies, in this study, we will use metagenomic sequencing of stool samples from individuals with RRMS and controls from Marshfield Clinic in Wisconsin to not only characterize the bacteria and phage composition differences between the groups, but also to begin to examine the potential metabolic differences in the bacterial communities.

Methods

Study design and stool samples

This study obtained approval from institutional research board (IRB) of Marshfield Clinic Health System (MCHS) (IRB protocol SHU10417) and approval by University of Wisconsin-Madison IRB (IRB protocol 2021-0793). All the included subjects signed a written informed consent.

As previously described, 237 patients with MS and 50 controls were recruited at MCHS from 2018-2021 who had been diagnosed with MS in the last two years with exclusion criteria of taking antibiotics, laxatives, or undergoing a colonoscopy in the previous three months [206]. A subset of the 237 patients enrolled in the overall project, including 75

patients with MS, of those 75, 44 patients with RRMS on a disease modifying therapy and 31 patients with RRMS not on a disease modifying therapy, and 26 controls were used for this study.

A fecal sample collection kit was sent to each participant and fecal samples were returned with a frozen cold pack. Stool samples were divided in lab and frozen at -80□C until further processing.

DNA extraction from stool samples and sequencing

Microbial DNA was extracted from stool sample aliquots using a Qiagen DNeasy PowerLyzer PowerSoil Kit and following the kit's protocol.

Library preparation and sequencing of the extracted microbial DNA was performed at UW-Madison Biotechnology Center on an Illumina NovaSeq6000, which produces 2 x 150 base pair reads and samples were sequenced at a depth of 25 million reads per sample.

Metagenomic sequencing analysis

Raw metagenomic reads were trimmed of adapters and quality filtered to remove low quality reads and human DNA using metaWRAP v1.2.2 [207]. The resulting reads were assembled using SPAdes v3.12.0 (parameters: --meta, --k 21,33,55,77,99,127) [208]. We removed scaffolds less than 1500 base pairs and ran VIBRANT v1.2.1 to identify phages [209]. We removed phages identified by VIBRANT as lytic and circular from the

scaffolds and used metaWRAP (parameters: --metabat2 --maxbin2 --metabat1 --concoct) to bin the remaining scaffolds. We performed bin refinement on all the bin outputs from metaWRAP except those from Concoct using DAS Tool v1.1.3 for our final set of metagenomic assembled genomes (MAGs) [210].

Taxonomic Classification of MAGs

We performed taxonomic classification of the MAGs via two methods: 1. GTDB-Tk classification and 2. via a concatenated ribosomal protein tree. We assigned taxonomy using GTDB-Tk v2.3.2 and reference database release 214 with the classify_wf function [211–220]. GTDB-Tk also provided checkM results describing percent completeness and contamination for each MAG.

We created concatenated ribosomal protein trees using a Hidden Markov Model (HMM) database of single copy ribosomal proteins including: rpL2, rpL3, rpL4, rpL5, rpL6, rpL14, rpL15, rpL16, rpL18, rpL22, rpL24, rpS3, rpS8, rpS10, rpS17, and rpS19 [221]. We identified ribosomal proteins in our MAG dataset using the database and hmmsearch (version 3.3.2) with noise cutoffs (--cut_nc) [108]. Using the Prodigal output from checkM, we extracted the identified protein sequences for each ribosomal protein from the ORFs and imported the sequences into Geneious Prime (version 2022.2.2). The extracted protein sequences for each ribosomal protein were aligned with a reference sequence set within Geneious using MAFFT (version 7.490, algorithm: auto, scoring matrix: BLOSUM62, gap open penalty: 1.53, offset value: 0.123). A 95% gap masking threshold was applied to each ribosomal protein alignment and the resulting alignments were

concatenated. The concatenated alignment was manually trimmed and exported in FASTA format. The resulting concatenated alignment was used as the input to make a phylogenetic tree using IQTree (version 2.2.0.3, parameters: -nt AUTO -m MFP -bb 1000 -redo -mset WAG,LG,JTT,Dayhoff -mrate E,I,G,I+G -mfreq FU -wbtl) [222]. The resulting tree was imported into iToL for visualization and manual curation [119].

Metabolic potential characterization of MAGs

We performed metabolic reconstruction of each MAG using the METABOLIC-C function of METABOLIC (version 4.0) [223]. Using the METABOLIC_results excel file output, we summarized the proportion of each participants MAGs that had each KEGG module step present. Using those proportions, we then used DESeq2 (DESeqDataSetFromMatrix function) to determine if there were differences by participant type (RRMS on disease modifying therapy, RRMS not on disease modifying therapy, and control) in the proportion of genomes with or without each of the KEGG steps [224].

Viral characterization of samples

To begin to understand the viral community from the metagenomes, we used ViWrap, a viromics program that combines several programs, allowing the ability to identify viruses from metagenomes, annotate genes, genome bin, assign taxonomy, cluster at the species and genus level, predict the viral host, and analyze the viral genome quality. We analyzed all metagenomes through the full ViWrap v1.2.1 pipeline using the run function. To have the most updated viral taxonomy assignments, we took the vRhyme

output from ViWrap (all_vRhyme_fasta.Nlinked_viral_gn.fasta) for each metagenome and used as input for geNomad v1.7.0 (function annotate) with database 1.5.0, which is based on the International Committee on the Taxonomy of Viruses' Virus Metadata Resource 19 [225]. Using the geNomad output, we summarized any unique viral taxa found only in one participant type (RRMS on disease modifying therapy, RRMS not on disease modifying therapy, and control) and we summarized the total number of unique viral taxa by participant disease type.

Results

Participant Demographics

A subset of the 237 patients enrolled in the overall project, including 75 patients with MS, of those 75, 44 patients with RRMS on a disease modifying therapy and 31 patients with RRMS not on a disease modifying therapy, and 26 controls were used for this study. A summary of the characteristics of the participants are in Table 1. The majority of all participants were female and all participants were white.

Table 1. Demographics of participants in this study. Disease modifying therapies include fingolimod, natalizumab, ocrelizumab, teriflunomide, glatrimer acetate, demthyl fumarate, interferon beta-1a, rituximab, or a combination of the listed therapies.

Characteristic	RRMS on Disease Modifying Therapy (n=44)	RRMS not on Disease Modifying Therapy (n=31)	Control (n=26)
Median Age (at diagnosis for MS and at enrollment for controls in years)	44	58	44

Ethnicity	White	White	White
Gender			
Male n(%)	14 (32%)	7 (23%)	8 (31%)
Female n (%)	30 (68%)	24 (77%)	18 (69%)
Median BMI	27.4	29.7	27.3
Disease Modifying Therapy (n)	44	0	0

Summary of Metagenomic Assembled Genomes and Viral Identification

We reconstructed 3577 MAGs across all participants. Of those 3035 (85%) were considered high or medium quality genomes, >50% complete and <10% contaminated [105,226]. For each participant, we assembled over 30 MAGs (Table 2). We identified 51034 total lytic phages and 10645 lysogenic phages across all the samples with 400-600 lytic phages per sample and ~100 lysogenic phages per sample (Table 2).

Table 2. Summary of metagenomic processing including creation of metagenomic assembled genomes (bins) and viral identification. Reported numbers are per sample in each category.

	RRMS on disease modifying therapy (n=44)	**RRMS not on disease modifying therapy (n=31)**	**Control**
Metagenomic assembling and binning			
Median scaffolds >1500 base pairs	26881.5	25427	25020.5
Median total base pairs in assembly (base pairs)	177082627	170676241	173500063
Median metagenomic bins (Metabat1, Metabat2, Maxbin)	69	65	70
Median refined metagenomic bins	37	33	35
Viral Identification			
Median putative phages	651.5	591	581.5
Median lytic phages	534	482	476.5

Median lysogenic phages	111	101	106

Bacterial MAGs Taxonomic Classification

From taxonomic classification by GTDB-tk and a concatenated ribosomal protein phylogeny, most MAGs were in the phyla Firmicutes or Bacteroidota (Figure 1).

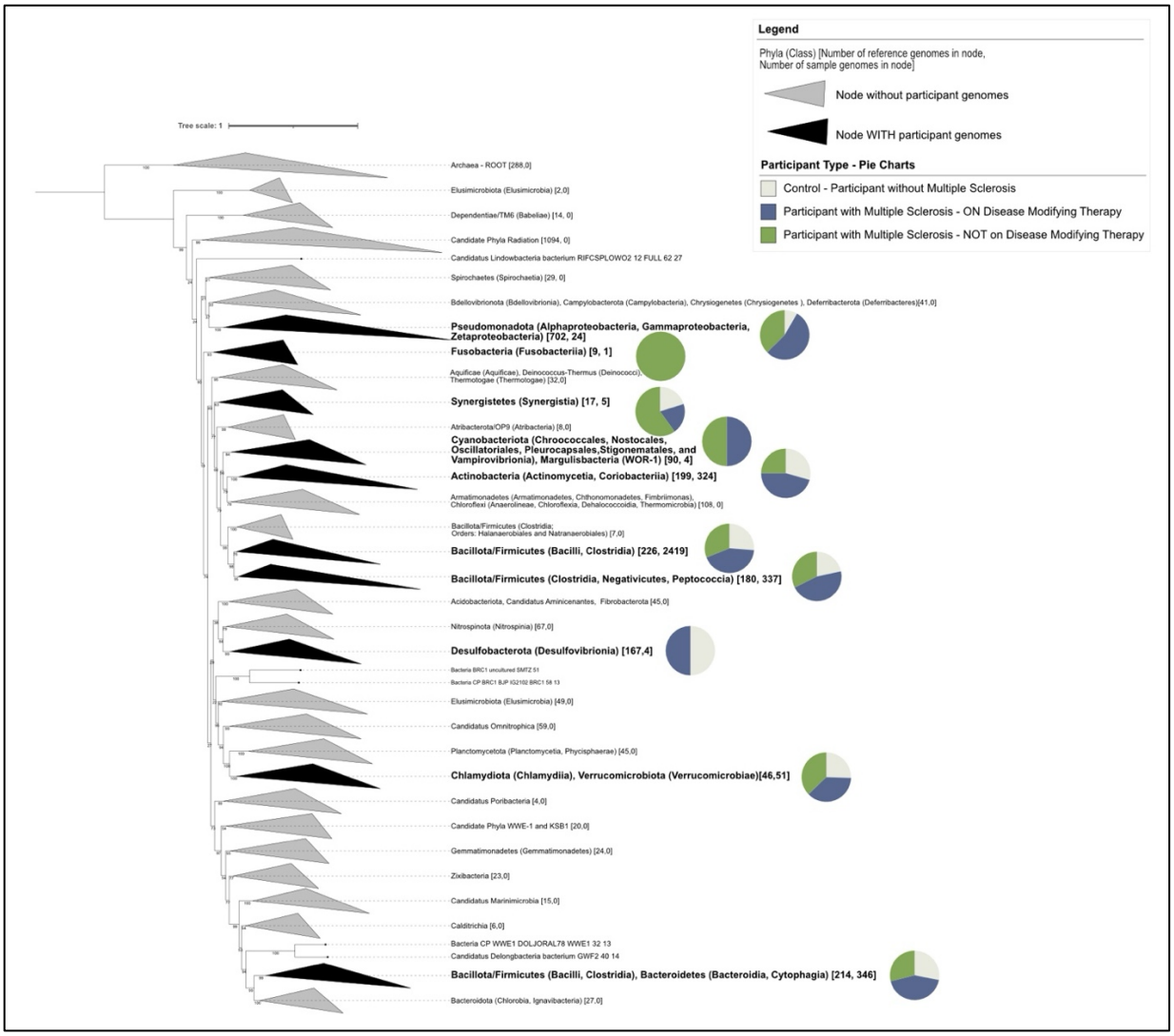


Figure 1. Concatenated ribosomal protein tree to taxonomically classify MAGs. Gray clades only contain reference sequences, darker gray clades contain reference sequences and sequences from participants in this study. Bracketed numbers indicate the sequence origin for each collapsed clade [number of sequences from our study, number of sequences of references]. Pie charts indicate the disease state associated with the sequences from participants in this study in each collapsed clade with white being controls, blue indicating participants with RRMS that are on disease modifying

therapy, and green indicating participants with RRMS that are not on disease modifying therapy.

Bacterial functional differences

From the METABOLIC output, we compared all combinations of participants (RRMS on disease modifying therapy vs RRMS not on disease modifying therapy; RRMS on disease modifying therapy vs controls; RRMS not on disease modifying therapy vs controls; all RRMS vs controls) to determine if there was a difference in the proportion of MAGs with various functional genes from KEGG modules (Tables 3-6). In total, we compared 470 KEGG modules.

After p-value correct, only two pathways were still significant (p-value < 0.05). In the comparison between the two groups of participant with MS, we found that the Semi-phosphorylative Entner-Doudoroff pathway, which takes gluconate and transforms it to glycerate-3-phosphate, was less likely to be present in in the assembled genomes of participants with MS on disease modifying therapy than those who are not on disease modifying therapy (corrected p-value 0.006, Table 3). We also found that tryptophan biosynthesis, chorismate → tryptophan, was more likely to be present in bacterial genomes of individuals with RRMS not on disease modifying therapy compared to controls (corrected p-value 0.02. Table 5). We did not find any pathways differentially abundant when we compared individuals with RRMS on a disease modifying therapy and controls (Table 4) and we did not find differences when we compared all participants with RRMS vs controls (Table 6).

Table 3. Summary of comparison of proportion of genomes with metabolic pathways present for RRMS on disease modifying therapy versus RRMS not on disease modifying therapy. Only uncorrected p-values <0.05 are included and the green shaded row indicates a corrected p-value <0.05.

KEGG Module Step	Log_2 Fold Change	p-value	Corrected p-value
Semi-phosphorylative Entner-Doudoroff pathway, gluconate → glycerate-3P	-0.375022	3.78E-05	0.00579
D-Glucuronate degradation, D-glucuronate → pyruvate + D-glyceraldehyde 3P	-0.336438	0.00088	0.067345
Lysine biosynthesis, DAP aminotransferase pathway, aspartate→lysine	-0.148129	0.001532	0.078124
Galactose degradation, Leloir pathway, galactose → alpha-D-glucose-1P	-0.237705	0.013946	0.485437
Cobalamin biosynthesis, anaerobic, uroporphyrinogen III →sirohydrochlorin → cobyrinate a,c-diamide	-0.154071	0.016215	0.485437
Lysine biosynthesis, DAP dehydrogenase pathway, aspartate →lysine	-0.127705	0.019037	0.485437
Pyridoxal-P biosynthesis, R5P + glyceraldehyde-3P + glutamine → pyridoxal-P	0.204511	0.028379	0.530768
Methionine biosynthesis, aspartate → homoserine → methionine	-0.252162	0.032752	0.530768
Glycolysis (Embden-Meyerhof pathway), glucose → pyruvate	0.051019	0.034142	0.530768
PRPP biosynthesis, ribose 5P → PRPP	0.06684	0.038987	0.530768
NAD biosynthesis, aspartate → quinolinate → NAD	0.106972	0.039379	0.530768
Thiamine salvage pathway, HMP/HET →TMP	0.105753	0.044805	0.530768
Histidine biosynthesis, PRPP → histidine	-0.112134	0.045098	0.530768

Table 4. Summary of comparison of proportion of genomes with metabolic pathways present for RRMS on disease modifying therapy versus controls. Only uncorrected p-values <0.05 are included.

KEGG Module Step	Log_2 Fold Change	p-value	Corrected p-value
Cobalamin biosynthesis, cobyrinate a,c-diamide → cobalamin	0.15553	0.003882	0.50788
Menaquinone biosynthesis, chorismate (+ polyprenyl-PP) → menaquinol	-5.577289	0.007056	0.50788
D-Galacturonate degradation (bacteria), D-galacturonate → pyruvate + D-glyceraldehyde 3P	0.351223	0.012396	0.50788
Glutathione biosynthesis, glutamate → glutathione	-5.093872	0.018534	0.50788
Glycolysis (Embden-Meyerhof pathway), glucose → pyruvate	-0.055686	0.020958	0.50788
Semi-phosphorylative Entner-Doudoroff pathway, gluconate → glycerate-3P	0.202529	0.023956	0.50788
UDP-N-acetyl-D-glucosamine biosynthesis, prokaryotes, glucose → UDP-GlcNAc	-0.096975	0.02963	0.50788
C10-C20 isoprenoid biosynthesis, bacteria	-0.500571	0.0301	0.50788
Cobalamin biosynthesis, anaerobic, uroporphyrinogen III → sirohydrochlorin → cobyrinate a,c-diamide	0.146101	0.033994	0.50788
Reductive pentose phosphate cycle, ribulose-5P → glyceraldehyde-3P	-5.05023	0.038598	0.50788

D-Glucuronate degradation, D-glucuronate → pyruvate + D-glyceraldehyde 3P	0.214934	0.04592	0.50788
Glycogen biosynthesis, glucose-1P → glycogen/starch	-0.097519	0.046911	0.50788
Lysine biosynthesis, DAP aminotransferase pathway, aspartate → lysine	0.107613	0.049444	0.50788
Coenzyme A biosynthesis, pantothenate → CoA	0.073995	0.05262	0.50788
Cytochrome bd ubiquinol oxidase	-4.836799	0.052904	0.50788

Table 5. Summary of comparison of proportion of genomes with metabolic pathways present for RRMS not on disease modifying therapy versus controls. Only uncorrected p-values <0.05 are included and the green shaded row indicates a corrected p-value <0.05.

KEGG Module Step	Log_2 Fold Change	p-value	Corrected p-value
Tryptophan biosynthesis, chorismate → tryptophan	0.254307	0.000152	0.02309
Glutathione biosynthesis, glutamate → glutathione	-5.442974	0.003801	0.139759
Pyridoxal-P biosynthesis, R5P + glyceraldehyde-3P + glutamine → pyridoxal-P	0.295632	0.003923	0.139759
Heme biosynthesis, animals and fungi, glycine → heme	-5.609408	0.004263	0.139759
Menaquinone biosynthesis, chorismate (+ polyprenyl-PP) → menaquinol	-5.581209	0.004597	0.139759
Tyrosine biosynthesis, chorismate → HPP → tyrosine	-5.348714	0.008012	0.202975
Fatty acid biosynthesis, elongation	0.078729	0.013701	0.297502

Cytochrome bd ubiquinol oxidase	-5.170862	0.018407	0.314712
Glyoxylate cycle	-5.164564	0.018634	0.314712
Pyridoxal-P biosynthesis, erythrose-4P → pyridoxal-P	-4.99932	0.03681	0.385214
Reductive pentose phosphate cycle (Calvin cycle)	-4.99932	0.03681	0.385214
Succinate dehydrogenase, prokaryotes	-4.99932	0.03681	0.385214
Galactose degradation, Leloir pathway, galactose → alpha-D-glucose-1P	-0.241369	0.0378	0.385214
ETEC pathogenicity signature, colonization factors	-4.943141	0.040549	0.385214
Fumarate reductase, prokaryotes	-4.943141	0.040549	0.385214
Ubiquinone biosynthesis, prokaryotes, chorismate (+ polyprenyl-PP) → ubiquinol	-4.943141	0.040549	0.385214
Heme biosynthesis, plants and bacteria, glutamate → heme	-1.354657	0.049245	0.387558
Guanine ribonucleotide biosynthesis, IMP → GDP,GTP	-0.067777	0.05128	0.387558

Table 6. Summary of comparison of proportion of genomes with metabolic pathways present for all RRMS participants versus controls. Only uncorrected p-values <0.05 are included.

KEGG Module Step	Log_2 Fold Change	p-value	Corrected p-value
Menaquinone biosynthesis, chorismate (+ polyprenyl-PP) → menaquinol	5.583177	0.000809	0.095986
Glutathione biosynthesis, glutamate → glutathione	5.250936	0.001247	0.095986

Heme biosynthesis, animals and fungi, glycine → heme	5.146387	0.006537	0.226782
Phenylalanine biosynthesis, chorismate → phenylpyruvate → phenylalanine	5.086653	0.007363	0.226782
Tyrosine biosynthesis, chorismate → HPP → tyrosine	5.086653	0.007363	0.226782
Cytochrome bd ubiquinol oxidase	4.988273	0.008893	0.228253
Tryptophan biosynthesis, chorismate → tryptophan	-0.171476	0.011599	0.255181
Cobalamin biosynthesis, cobyrinate a,c-diamide → cobalamin	-0.119973	0.014206	0.273469
Pyridoxal-P biosynthesis, erythrose-4P → pyridoxal-P	4.953137	0.017041	0.291584
Reductive pentose phosphate cycle (Calvin cycle)	4.813106	0.021474	0.330698
Ubiquinone biosynthesis, prokaryotes, chorismate (+ polyprenyl-PP) → ubiquinol	4.84658	0.026506	0.351579
C10-C20 isoprenoid biosynthesis, bacteria	0.469443	0.033008	0.351579
Pyrimidine degradation, uracil → 3-hydroxypropanoate	4.69859	0.033167	0.351579
Undecaprenylphosphate alpha-L-Ara4N biosynthesis, UDP-GlcA → undecaprenyl phosphate alpha-L-Ara4N	4.69859	0.033167	0.351579
Glyoxylate cycle	4.6765	0.034246	0.351579
Glycogen biosynthesis, glucose-1P → glycogen/starch	0.090821	0.036528	0.351579
Formaldehyde assimilation, ribulose monophosphate pathway	0.96168	0.051814	0.443181

Viral Taxonomy

The majority of viruses that were able to be taxonomically classified in all groups were in the class *Caudoviricetes* (Tables 7-9). The phylogenetic level that the viruses identified from the metagenomes can be classified at depends on the sequence length. We found 1-3 classified viruses that were unique to each RRMS participant type, but no viruses unique to the control group (Table 10).

Table 7. Viral taxonomic classifications for participants with RRMS on a disease modifying therapy.

Viral Taxonomic Assignment	Number viruses identified with this taxonomy
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes	15251
Viruses;Monodnaviria;Sangervirae;Phixviricota;Malgrandaviricetes;Petitvirales;Microviridae	259
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Crassvirales	161
Viruses;Monodnaviria;Loebvirae;Hofneiviricota;Faserviricetes;Tubulavirales;Inoviridae	30
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cirlivirales;Circoviridae	27
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Peduoviridae	9
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Herelleviridae	6
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Straboviridae	6
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Vilmaviridae	6
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Autographiviridae	5
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Demerecviridae	3
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Ackermannviridae	2
Viruses	1
Viruses;Bicaudaviridae	1

Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Casjensviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Kyanoviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Mesyanzhinovviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Schitoviridae	1
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cremevirales;Smacoviridae	1
Viruses;Riboviria;Orthornavirae;Duplornaviricota;Vidaverviricetes;Mindivirales;Cystoviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Algavirales;Phycodnaviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Imitervirales;Mimiviridae	1
Viruses;Varidnaviria;Bamfordvirae;Preplasmiviricota;Tectiliviricetes;Kalamavirales;Tectiviridae	1

Table 8. Viral taxonomic classifications for participants with RRMS not on a disease modifying therapy.

Viral Taxonomic Assignment	**Number viruses identified with this taxonomy**
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes	10085
Viruses;Monodnaviria;Sangervirae;Phixviricota;Malgrandaviricetes;Petitvirales;Microviridae	141
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Crassvirales	118
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cirlivirales;Circoviridae	27
Viruses;Monodnaviria;Loebvirae;Hofneiviricota;Faserviricetes;Tubulavirales;Inoviridae	21
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Straboviridae	10
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Herelleviridae	8
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Peduoviridae	6
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Autographiviridae	4

Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Kyanoviridae	3
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Drexlerviridae	2
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Schitoviridae	2
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Ackermannviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Casjensviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Chaseviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Demerecviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Mesyanzhinovviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zierdtviridae	1
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cremevirales;Smacoviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Imitervirales;Mimiviridae	1

Table 9. Viral taxonomic classifications for control participants without RRMS.

Viral Taxonomic Assignment	**Number viruses identified with this taxonomy**
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes	8337
Viruses;Monodnaviria;Sangervirae;Phixviricota;Malgrandaviricetes;Petitvirales;Microviridae	99
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Crassvirales	78
Viruses;Monodnaviria;Loebvirae;Hofneiviricota;Faserviricetes;Tubulavirales;Inoviridae	14
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Herelleviridae	7
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Straboviridae	7
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Peduoviridae	6
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Demerecviridae	5

Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Ackermannviridae	4
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Casjensviridae	3
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Autographiviridae	2
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Schitoviridae	2
Viruses;Bicaudaviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Mesyanzhinovviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Vilmaviridae	1

Table 10. Phages unique to each of the three participant types.

Viral Taxonomic Classification by Disease State	Number of viruses identified with this taxonomy
Unique to RRMS on disease modifying therapy	
Viruses	1
Viruses;Riboviria;Orthornavirae;Duplornaviricota;Vidaverviricetes;Mindivirales;Cystoviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Algavirales;Phycodnaviridae	1
Viruses;Varidnaviria;Bamfordvirae;Preplasmiviricota;Tectiliviricetes;Kalamavirales;Tectiviridae	1
Unique to RRMS not on disease modifying therapy	
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Drexlerviridae	2
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Chaseviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zierdtviridae	1
Unique to only RRMS (on and not on disease modifying therapy)	
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Kyanoviridae	3
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Drexlerviridae	2
Viruses	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Kyanoviridae	1
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cremevirales;Smacoviridae	1
Viruses;Riboviria;Orthornavirae;Duplornaviricota;Vidaverviricetes;Mindivirales;Cystoviridae	1

Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Algavirales;Phycodnaviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Imitervirales;Mimiviridae	1
Viruses;Varidnaviria;Bamfordvirae;Preplasmiviricota;Tectiliviricetes;Kalamavirales;Tectiviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Chaseviridae	1
Viruses;Duplodnaviria;Heunggongvirae;Uroviricota;Caudoviricetes;Zierdtviridae	1
Viruses;Monodnaviria;Shotokuvirae;Cressdnaviricota;Arfiviricetes;Cremevirales;Smacoviridae	1
Viruses;Varidnaviria;Bamfordvirae;Nucleocytoviricota;Megaviricetes;Imitervirales;Mimiviridae	1
Unique to Control	
None	

Discussion

We used metagenomic sequencing of stool samples from 75 individuals with RRMS and 26 controls to characterize the human gut microbiome. We found the bacterial composition in participants with RRMS and controls were similar and largely contained phyla common to the gut environment: *Bacteroidota* and *Firmicutes.* We characterized the phage composition of the stool samples and found the majority of the phages in all groups were from the class *Caudoviricetes,* which are the most abundant class of phages found in the gut [204,227–232]. While we did find evidence of phages unique to the three different participant types (Table 10), we only identified 1-3 phages per unique subtype, which is not robust evidence to base conclusions on. Of note, we did not find any phages unique to the control participants. Future work to determine meaningful differences in the bacterial and phage compositions of the participants will need to include determine abundances of the different taxa.

In addition to taxonomic differences, we explored the potential functional differences of the bacterial taxa and found the Semi-phosphorylative Entner-Doudoroff pathway, which takes gluconate and transforms it to glycerate-3-phosphate, was less likely to be present in in the genomes of participants with MS on disease modifying therapy than those who are not on disease modifying therapy. There is little in the literature about the role of the bacterial Entner-Doudoroff pathway in multiple sclerosis and future work could explore which bacterial taxa have the pathway whether the pathway could have a role in worsening RRMS symptoms or outcomes.

We also found that tryptophan biosynthesis, chorismate → tryptophan, was more likely to be present in bacterial genomes of individuals with RRMS not on disease modifying therapy compared to controls. From previous work on gut microbiome metabolites in patients with MS, we know that bacteria that produce short chain fatty acids (SCFA) have a lower abundance in MS and levels of certain SCFAs such as butyrate are lower in the stool of patients with MS [233–236]. Bile acids have also found to be lower in patients with MS [237]. Tryptophan can be produced both by human cells and by bacteria. Previous work has shown that blood levels of tryptophan metabolism products by human metabolism are reduced in patients with MS [238–240]. Studies have shown mixed results about the impact of microbial tryptophan metabolism on MS with some studies showing a protective effect and others showing microbial tryptophan metabolism enhances MS severity [203,241,242]. Given these contradictory results from previous studies, it is clear that more work is needed to understand the role of bacterial tryptophan metabolism in MS and future work with our data could include incorporating

metabolomics of the stool samples with our genomic data to measure tryptophan and tryptophan derivative levels to compare across participant types.

Limitations of this study include only having one timepoint from each participant and having limited clinical data including not knowing if participants with RRMS are currently having an exacerbation or were in remission at the time of stool collection. For metabolic reconstruction, we only know if genes and pathways are present from DNA sequencing and not activity levels. While the number of controls is fewer than the number of participant samples with RRMS, one strength of the study is the inclusion of control subjects. Other strengths include the inclusion of participants with RRMS both on disease modifying therapy and those not on disease modifying therapy, not only bacterial composition analysis, but also bacterial metabolic reconstruction.

Future work could explore transcriptomics of stool samples from these participants to understand expression levels of the different bacterial pathways and their impacts on disease progression. Additional future work could incorporate clinical data to determine if there are differences in the microbial composition or functionality based on any clinical factors.

Considering the known role of the immune system in MS, future work should incorporate immunological studies alongside microbiological evaluations of participants. Recent seminal work on MS has shown that not only do almost all patients with MS have an increased presence of antibodies to the double stranded DNA Epstein-Barr

Virus (EBV) in their serum, but studies of US military personnel found risk of MS increase 32 fold after individuals were infected with EBV [197,243]. Subsequent work showed that due to molecular mimicry, EBV strongly binds a glial molecule found in the central nervous system and when the immune system attempts to clear the virus, it also inadvertently attacks the nervous system [244]. Future work with the data from this study could identify EBV sequences from the metagenomes while also examining if any phages present in the gut microbiome exhibit molecular mimicry or if there are changes in the phage functional landscape between participants with RRMS and controls.

In this study, we have begun to use extensive sequencing data of 75 participants with RRMS and 26 controls to understand the role bacteria and phages play in RRMS. The few differences we found in bacterial metabolic capabilities provide promising future directions while the phage composition and functional potential remains largely unexamined. Pairing what we have currently discovered and future work with clinical data will provide insights into the role of the gut microbiome in RRMS.

Acknowledgements

We thank the participants of the study. We thank Patricia Tran and Zichao Zhou for bioinformatics guidance. ESC is a Medical Scientist Training Program (MSTP) student and was supported in part by MSTP grant T32GM140935 and by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison. This work was funded by an ICTR MC/MCRI-UW Madison Collaborative Research Pilot Program grant.

Contributions

SS, KA, and ESC conceptualized the project. ESC analyzed metagenomic data. ESC wrote the manuscript.

Declarations

The authors declare no conflict of interest.

Data Availability

Raw sequencing files, metagenomics assembled genomes, and all scripts used for this project are available upon request from Karthik Anantharaman: karthik@bact.wisc.edu.

Chapter 5: Conclusion

In this thesis, I investigated the role of the human gut microbiome in three conditions: colorectal cancer **(chapter 2)**, Group B *Streptococcus* colonization **(chapter 3)**, and multiple sclerosis **(chapter 4)**.

In **chapter 2,** using ~17,000 bacterial genomes from publicly available stool metagenomes, we studied the diversity of sulfur metabolic genes in an international cohort of 667 participants across different health statuses: healthy, adenoma, and carcinoma. We sought to develop a better understanding of the sulfidogenic capacity of the human gut microbiome. We found that genes for microbial sulfur metabolism are more diverse than previously known and are widely distributed across our participant cohort. Bacterial sulfur genes are significantly associated with CRC and our work indicates that genes for organic sulfur metabolism may be the most important contributor of H_2S in the human gut. This work provides a foundation for future work more closely characterizing bacterial sulfidogenic enzymes that may be active in the gut and how the expression of these genes changes with health status of CRC and human disease.

In **chapter 3,** we used a biobank of 754 stool samples and associated data collected from adults in Wisconsin to better understand the host and microbiome-based factors that influence gastrointestinal Group B *Streptococcus* (GBS) carriage. We found the prevalence of GBS in the human gut (18%) of adults in Wisconsin is similar to the rates found by rectovaginal swabs in pregnant individuals. We found the abundance of GBS

in the gut varied widely. We found that an increased GBS abundance was associated with an increased time since last dental checkup and cleaning as well as an increased frequency of exogenous iron consumption. This work provides a starting point to begin to understand risk factors for GBS carriage in the general population and in pregnant people since the risk factors are likely overlapping and by studying GBS in the general population, we can gain better insight into the physiology of GBS in the human body.

In **chapter 4,** we used metagenomic sequencing of 75 participants with relapsing remitting multiple sclerosis (RRMS) and 26 control participants from one community in Marshfield to examine the role of the gut microbiome, particularly bacteria and phages, in RRMS. We found the bacterial composition in participants with RRMS and controls were similar, containing common gut environment phyla: *Bacteroidota* and *Firmicutes.* The majority of the phages in all participants were also from the most abundant class of gut phages: *Caudoviricetes.* The few differences we found in bacterial metabolic capabilities provide promising future directions while the phage composition and functional potential remains largely unexamined. Future work could include pairing more of the clinical data available on participants with the metagenomic data to understand if there are associations between gut microbiome composition or functionality and participant characteristics.

A strength of all my projects is that we were able to leverage metadata on participants for all the projects. A limitation of my work, and many gut microbiome studies, is only having one time point per participant. While cross sectional work is important, having a

time series would be more powerful for all my projects. For the CRC work, we could have seen how the microbiome may have changed as participants with adenoma had polyps removed or how the microbiomes would respond when participants with CRC received treatment. For the GBS work, a longitudinal series would allow us to determine if GBS is a permanent member of the gut microbiome or just transient. Finally for the RRMS work, having a time series would allow us to understand how the gut microbiome changes during MS exacerbations versus recovery and if there are changes when participants change medications.

The three projects that comprise my thesis embody the Wisconsin Idea. I took the skills that I learned from my work on CRC and applied them to people across the state of Wisconsin. This research reaches outside of the bounds of this campus and benefits the people within the state.

References (whole thesis)

1. Vos, W.M. de, Tilg, H., Hul, M.V., and Cani, P.D. (2022). Gut microbiome and health: mechanistic insights. Gut *71*, 1020–1032. 10.1136/gutjnl-2021-326789.

2. Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., McCracken, C., Giglio, M.G., et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. Nature *550*, 61–66. 10.1038/nature23889.

3. Quigley, E.M.M. (2013). Gut Bacteria in Health and Disease. Gastroenterol Hepatol (N Y) *9*, 560–569.

4. Guarner, F., and Malagelada, J.-R. (2003). Gut flora in health and disease. The Lancet *361*, 512–519. 10.1016/S0140-6736(03)12489-0.

5. Bull, M.J., and Plummer, N.T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. Integr Med (Encinitas) *13*, 17–22.

6. Pickard, J.M., Zeng, M.Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunological Reviews *279*, 70–89. 10.1111/imr.12567.

7. Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., and Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr *57*, 1–24. 10.1007/s00394-017-1445-8.

8. Zheng, D., Liwinski, T., and Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. Cell Res *30*, 492–506. 10.1038/s41422-020-0332-7.

9. Wu, H.-J., and Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. Gut Microbes *3*, 4–14. 10.4161/gmic.19320.

10. Monda, V., Villano, I., Messina, A., Valenzano, A., Esposito, T., Moscatelli, F., Viggiano, A., Cibelli, G., Chieffi, S., Monda, M., et al. (2017). Exercise Modifies the Gut Microbiota with Positive Health Effects. Oxid Med Cell Longev *2017*, 3831972. 10.1155/2017/3831972.

11. Clauss, M., Gérard, P., Mosca, A., and Leclerc, M. (2021). Interplay Between Exercise and Gut Microbiome in the Context of Human Health and Performance. Frontiers in Nutrition *8*.

12. Singh, R.K., Chang, H.-W., Yan, D., Lee, K.M., Ucmak, D., Wong, K., Abrouk, M., Farahnik, B., Nakamura, M., Zhu, T.H., et al. (2017). Influence of diet on the gut microbiome and implications for human health. Journal of Translational Medicine *15*, 73. 10.1186/s12967-017-1175-y.

13. Zhang, P. (2022). Influence of Foods and Nutrition on the Gut Microbiome and Implications for Intestinal Health. Int J Mol Sci *23*, 9588. 10.3390/ijms23179588.

14. Leshem, A., Segal, E., and Elinav, E. (2020). The Gut Microbiome and Individual-Specific Responses to Diet. mSystems *5*, 10.1128/msystems.00665-20. 10.1128/msystems.00665-20.

15. Xu, Z., and Knight, R. (2015). Dietary effects on human gut microbiome diversity. Br J Nutr *113*, S1–S5. 10.1017/S0007114514004127.

16. Leeming, E.R., Johnson, A.J., Spector, T.D., and Le Roy, C.I. (2019). Effect of Diet on the Gut Microbiota: Rethinking Intervention Duration. Nutrients *11*, 2862. 10.3390/nu11122862.

17. Fan, Y., and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. Nat Rev Microbiol *19*, 55–71. 10.1038/s41579-020-0433-9.

18. Durack, J., and Lynch, S.V. (2019). The gut microbiome: Relationships with disease and opportunities for therapy. J Exp Med *216*, 20–40. 10.1084/jem.20180448.

19. Chen, Y., Zhou, J., and Wang, L. (2021). Role and Mechanism of Gut Microbiota in Human Disease. Front Cell Infect Microbiol *11*, 625913. 10.3389/fcimb.2021.625913.

20. Manor, O., Dai, C.L., Kornilov, S.A., Smith, B., Price, N.D., Lovejoy, J.C., Gibbons, S.M., and Magis, A.T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun *11*, 5206. 10.1038/s41467-020-18871-1.

21. Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. Cell Host & Microbe *26*, 527-541.e5. 10.1016/j.chom.2019.09.009.

22. Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. Cell Host & Microbe *25*, 195–209. 10.1016/j.chom.2019.01.017.

23. Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res *21*, 1616–1625. 10.1101/gr.122705.111.

24. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563. 10.1038/nature12820.

25. Jalanka-Tuovinen, J., Salonen, A., Nikkilä, J., Immonen, O., Kekkonen, R., Lahti, L., Palva, A., and Vos, W.M. de (2011). Intestinal Microbiota in Healthy Adults: Temporal Analysis Reveals Individual and Common Core and Relation to Intestinal Symptoms. PLOS ONE *6*, e23035. 10.1371/journal.pone.0023035.

26. Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The Long-Term Stability of the Human Gut Microbiota. Science *341*, 1237439. 10.1126/science.1237439.

27. Wisconsin Idea: University of Wisconsin–Madison https://www.wisc.edu/wisconsin-idea/.

28. Charles Richard Van Hise (President: 1903-1918) https://www.library.wisc.edu/archives/exhibits/campus-history-projects/chancellors-and-presidents-of-the-university-of-wisconsin-madison/charles-richard-van-hise-president-1903-1918/.

29. Tremaroli, V., and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. Nature *489*, 242–249. 10.1038/nature11552.

30. Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The Impact of the Gut Microbiota on Human Health: An Integrative View. Cell *148*, 1258–1270. 10.1016/j.cell.2012.01.035.

31. Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214. 10.1038/nature11234.

32. Raibaud, O., and Goldberg, M.E. (1973). The tryptophanase from Escherichia coli K-12. II. Comparison of the thermal stabilities of apo-, holo-, and hybrid enzymes. J Biol Chem *248*, 3451–3455.

33. Warren, Y.A., Citron, D.M., Merriam, C.V., and Goldstein, E.J.C. (2005). Biochemical Differentiation and Comparison of Desulfovibrio Species and Other Phenotypically Similar Genera. Journal of Clinical Microbiology *43*, 4041–4045. 10.1128/JCM.43.8.4041-4045.2005.

34. Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., McCracken, C., Giglio, M.G., et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. Nature *550*, 61–66. 10.1038/nature23889.

35. Islami, F., Goding Sauer, A., Miller, K.D., Siegel, R.L., Fedewa, S.A., Jacobs, E.J., McCullough, M.L., Patel, A.V., Ma, J., Soerjomataram, I., et al. (2018). Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA Cancer J Clin *68*, 31–54. 10.3322/caac.21440.

36. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563. 10.1038/nature12820.

37. O'Keefe, S.J.D., Li, J.V., Lahti, L., Ou, J., Carbonero, F., Mohammed, K., Posma, J.M., Kinross, J., Wahl, E., Ruder, E., et al. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun *6*, 6342. 10.1038/ncomms7342.

38. Devkota, S., Wang, Y., Musch, M.W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., Antonopoulos, D.A., Jabri, B., and Chang, E.B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10 −/− mice. Nature *487*, 104–108. 10.1038/nature11225.

39. Yazici, C., Wolf, P.G., Kim, H., Cross, T.-W.L., Vermillion, K., Carroll, T., Augustus, G.J., Mutlu, E., Tussing-Humphreys, L., Braunschweig, C., et al. (2017). Race-dependent association of sulfidogenic bacteria with colorectal cancer. Gut *66*, 1983–1994. 10.1136/gutjnl-2016-313321.

40. Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., Hold, G.L., et al. (2013). Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. Cell Host & Microbe *14*, 207–215. 10.1016/j.chom.2013.07.007.

41. Gao, Z., Guo, B., Gao, R., Zhu, Q., and Qin, H. (2015). Microbiota disbiosis is associated with colorectal cancer. Front. Microbiol. *6*. 10.3389/fmicb.2015.00020.

42. Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res *22*, 292–298. 10.1101/gr.126573.111.

43. Chen, W., Liu, F., Ling, Z., Tong, X., and Xiang, C. (2012). Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer. PLOS ONE *7*, e39743. 10.1371/journal.pone.0039743.

44. Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R.A., et al. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res *22*, 299–306. 10.1101/gr.126516.111.

45. Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol *10*, 766. 10.15252/msb.20145645.

46. Wolf, P.G. (2018). Microbial pathways of sulfur metabolism and colorectal cancer risk.

47. Huycke, M.M., and Gaskins, H.R. (2004). Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. Exp Biol Med (Maywood) *229*, 586–597. 10.1177/153537020422900702.

48. Furne, J., Saeed, A., and Levitt, M.D. (2008). Whole tissue hydrogen sulfide concentrations are orders of magnitude lower than presently accepted values. Am J Physiol Regul Integr Comp Physiol *295*, R1479-1485. 10.1152/ajpregu.90566.2008.

49. Yang, G., Wu, L., Jiang, B., Yang, W., Qi, J., Cao, K., Meng, Q., Mustafa, A.K., Mu, W., Zhang, S., et al. (2008). H2S as a Physiologic Vasorelaxant: Hypertension in Mice with Deletion of Cystathionine -Lyase. Science *322*, 587–590. 10.1126/science.1162667.

50. Zheng, J., Zhao, T., Yuan, Y., Hu, N., and Tang, X. (2015). Hydrogen sulfide (H2S) attenuates uranium-induced acute nephrotoxicity through oxidative stress and inflammatory response via Nrf2-NF-κB pathways. Chemico-Biological Interactions *242*, 353–362. 10.1016/j.cbi.2015.10.021.

51. Guo, C., Liang, F., Shah Masood, W., and Yan, X. (2014). Hydrogen sulfide protected gastric epithelial cell from ischemia/reperfusion injury by Keap1 s-sulfhydration, MAPK dependent anti-apoptosis and NF-κB dependent anti-inflammation pathway. European Journal of Pharmacology *725*, 70–78. 10.1016/j.ejphar.2014.01.009.

52. Attene-Ramos, M.S., Wagner, E.D., Plewa, M.J., and Gaskins, H.R. (2006). Evidence That Hydrogen Sulfide Is a Genotoxic Agent. Mol Cancer Res *4*, 9–14. 10.1158/1541-7786.MCR-05-0126.

53. Attene-Ramos, M.S., Wagner, E.D., Gaskins, H.R., and Plewa, M.J. (2007). Hydrogen sulfide induces direct radical-associated DNA damage. Mol Cancer Res *5*, 455–459. 10.1158/1541-7786.MCR-06-0439.

54. Attene-Ramos, M.S., Nava, G.M., Muellner, M.G., Wagner, E.D., Plewa, M.J., and Gaskins, H.R. (2010). DNA damage and toxicogenomic analyses of hydrogen sulfide in human intestinal epithelial FHs 74 Int cells. Environ. Mol. Mutagen., NA-NA. 10.1002/em.20546.

55. Deplancke, B., and Gaskins, H.R. (2003). Hydrogen sulfide induces serum-independent cell cycle entry in nontransformed rat intestinal epithelial cells. FASEB j. *17*, 1310–1312. 10.1096/fj.02-0883fje.

56. Carbonero, F., Benefiel, A.C., Alizadeh-Ghamsari, A.H., and Gaskins, H.R. (2012). Microbial pathways in colonic sulfur metabolism and links with health and disease. Frontiers in Physiology *3 NOV*, 1–11. 10.3389/fphys.2012.00448.

57. Ridlon, J.M., Wolf, P.G., and Gaskins, H.R. (2016). Taurocholic acid metabolism by gut microbes and colon cancer. Gut Microbes *7*, 201–215. 10.1080/19490976.2016.1150414.

58. Nguyen, L.H., Ma, W., Wang, D.D., Cao, Y., Mallick, H., Gerbaba, T.K., Lloyd-Price, J., Abu-Ali, G., Hall, A.B., Sikavi, D., et al. (2020). Association Between Sulfur-Metabolizing Bacterial Communities in Stool and Risk of Distal Colorectal Cancer in Men. Gastroenterology *158*, 1313–1325. 10.1053/j.gastro.2019.12.029.

59. Dordević, D., Jančíková, S., Vítězová, M., and Kushkevych, I. (2021). Hydrogen sulfide toxicity in the gut environment: Meta-analysis of sulfate-reducing and lactic acid bacteria in inflammatory processes. Journal of Advanced Research *27*, 55–69. 10.1016/j.jare.2020.03.003.

60. Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, X., Jie, Z., et al. (2015). Gut microbiome development along the colorectal adenoma–carcinoma sequence. Nature Communications *6*, 6528. 10.1038/ncomms7528.

61. Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A.Y., Hercog, R., Goedert, J.J., Shi, J., Bork, P., and Sinha, R. (2016). Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. PLOS ONE *11*, e0155362. 10.1371/journal.pone.0155362.

62. Hannigan, G.D., Duhaime, M.B., Ruffin, M.T., Koumpouras, C.C., and Schloss, P.D. (2018). Diagnostic Potential and Interactive Dynamics of the Colorectal Cancer Virome. mBio *9*. 10.1128/mBio.02248-18.

63. Yu, J., Feng, Q., Wong, S.H., Zhang, D., Liang, Q. yi, Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut *66*, 70–78. 10.1136/gutjnl-2015-309800.

64. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell *176*, 649-662.e20. 10.1016/j.cell.2019.01.001.

65. Rabus, R., Hansen, T.A., and Widdel, F. (2013). Dissimilatory Sulfate- and Sulfur-Reducing Prokaryotes. The Prokaryotes, 309–404. 10.1007/978-3-642-30141-4_70.

66. Florin, T., Neale, G., Gibson, G.R., Christl, S.U., and Cummings, J.H. (1991). Metabolism of dietary sulphate: absorption and excretion in humans. Gut *32*, 766–773. 10.1136/gut.32.7.766.

67. Alnouti, Y. (2009). Bile Acid Sulfation: A Pathway of Bile Acid Elimination and Detoxification. Toxicological Sciences *108*, 225–246. 10.1093/toxsci/kfn268.

68. Corfield, A.P., Wagner, S.A., Clamp, J.R., Kriaris, M.S., and Hoskins, L.C. (1992). Mucin degradation in the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. Infect Immun *60*, 3971–3978. 10.1128/IAI.60.10.3971-3978.1992.

69. Joseph W. Lengeler, Gerhart Drews, and Hans G. Schlegel (1999). Biology of the Prokaryotes (Wiley-Blackwell).

70. Müller, A.L., Kjeldsen, K.U., Rattei, T., Pester, M., and Loy, A. (2015). Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. The ISME Journal *9*, 1152–1165. 10.1038/ismej.2014.208.

71. Gibson, G.R., Macfarlane, G.T., and Cummings, J.H. (1988). Occurrence of sulphate-reducing bacteria in human faeces and the relationship of dissimilatory sulphate reduction to methanogenesis in the large gut. Journal of Applied Bacteriology *65*, 103–111. https://doi.org/10.1111/j.1365-2672.1988.tb01498.x.

72. Nava, G.M., Carbonero, F., Croix, J.A., Greenberg, E., and Gaskins, H.R. (2012). Abundance and diversity of mucosa-associated hydrogenotrophic microbes in the healthy human colon. The ISME Journal *6*, 57–70. 10.1038/ismej.2011.90.

73. Zinkevich, V., and Beech, I.B. (2000). Screening of sulfate-reducing bacteria in colonoscopy samples from healthy and colitic human gut mucosa. FEMS Microbiology Ecology *34*, 147–155. 10.1016/S0168-6496(00)00086-6.

74. Fite, A., Macfarlane, G.T., Cummings, J.H., Hopkins, M.J., Kong, S.C., Furrie, E., and Macfarlane, S. (2004). Identification and quantitation of mucosal and faecal desulfovibrios using real time polymerase chain reaction. Gut *53*, 523–529. 10.1136/gut.2003.031245.

75. Anantharaman, K., Hausmann, B., Jungbluth, S.P., Kantor, R.S., Lavy, A., Warren, L.A., Rappé, M.S., Pester, M., Loy, A., Thomas, B.C., et al. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. The ISME Journal *12*, 1715–1728. 10.1038/s41396-018-0078-0.

76. Lie, T.J., Pitta, T., Leadbetter, E.R., Godchaux III., W., and Leadbetter, J.R. (1996). Sulfonates: novel electron acceptors in anaerobic respiration. Arch Microbiol *166*, 204–210. 10.1007/s002030050376.

77. Lie, T.J., Godchaux, W., and Leadbetter, E.R. (1999). Sulfonates as Terminal Electron Acceptors for Growth of Sulfite-Reducing Bacteria (Desulfitobacterium spp.) and Sulfate-Reducing Bacteria: Effects of Inhibitors of Sulfidogenesis. Appl. Environ. Microbiol. *65*, 4611–4617. 10.1128/AEM.65.10.4611-4617.1999.

78. Laue, H., Denger, K., and Cook, A.M. (1997). Taurine reduction in anaerobic respiration of Bilophila wadsworthia RZATAU. Appl Environ Microbiol *63*, 2016–2021. 10.1128/AEM.63.5.2016-2021.1997.

79. Ahlman, B., Leijonmarck, C.-E., and Wernerman, J. (1993). The content of free amino acids in the human duodenal mucosa. Clinical Nutrition *12*, 266–271. 10.1016/0261-5614(93)90044-5.

80. Islam, K.B.M.S., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T., and Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. Gastroenterology *141*, 1773–1781. 10.1053/j.gastro.2011.07.046.

81. Albenberg, L., Esipova, T.V., Judge, C.P., Bittinger, K., Chen, J., Laughlin, A., Grunberg, S., Baldassano, R.N., Lewis, J.D., Li, H., et al. (2014). Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. Gastroenterology *147*, 1055-1063.e8. 10.1053/j.gastro.2014.07.020.

82. Peck, S.C., Denger, K., Burrichter, A., Irwin, S.M., Balskus, E.P., and Schleheck, D. (2019). A glycyl radical enzyme enables hydrogen sulfide production by the human intestinal bacterium Bilophila wadsworthia. Proc Natl Acad Sci U S A *116*, 3171–3176. 10.1073/pnas.1815661116.

83. Wu, G., Shi, Y., Han, L., Feng, C., Ge, Y., Yu, Y., Tang, X., Cheng, X., Sun, J., and Le, G.-W. (2020). Dietary Methionine Restriction Ameliorated Fat Accumulation, Systemic Inflammation, and Increased Energy Metabolism by Altering Gut Microbiota in Middle-Aged Mice Administered Different Fat Diets. J. Agric. Food Chem. *68*, 7745–7756. 10.1021/acs.jafc.0c02965.

84. Liu, G., Yu, L., Fang, J., Hu, C.-A.A., Yin, J., Ni, H., Ren, W., Duraipandiyan, V., Chen, S., Al-Dhabi, N.A., et al. (2017). Methionine restriction on oxidative stress and immune response in dss-induced colitis mice. Oncotarget *8*, 44511–44520. 10.18632/oncotarget.17812.

85. Shatalin, K., Shatalina, E., Mironov, A., and Nudler, E. (2011). H2S: a universal defense against antibiotics in bacteria. Science *334*, 986–990. 10.1126/science.1209855.

86. Rizzo, A.A. (1967). The possible role of hydrogen sulfide in human periodontal disease. I. Hydrogen sulfide production in periodontal pockets. Periodontics *5*, 233–236.

87. Horowitz, A., and Folke, L.E. (1973). Hydrogen sulfide production in the periodontal environment. J Periodontol *44*, 390–395. 10.1902/jop.1973.44.7.390.

88. Ng, W., and Tonzetich, J. (1984). Effect of hydrogen sulfide and methyl mercaptan on the permeability of oral mucosa. J Dent Res *63*, 994–997. 10.1177/00220345840630071701.

89. Yaegaki, K., Qian, W., Murata, T., Imai, T., Sato, T., Tanaka, T., and Kamoda, T. (2008). Oral malodorous compound causes apoptosis and genomic DNA damage in human gingival fibroblasts. J Periodontal Res *43*, 391–399. 10.1111/j.1600-0765.2007.01052.x.

90. Marchesi, J.R., Dutilh, B.E., Hall, N., Peters, W.H.M., Roelofs, R., Boleij, A., and Tjalsma, H. (2011). Towards the Human Colorectal Cancer Microbiome. PLOS ONE *6*, e20447. 10.1371/journal.pone.0020447.

91. McCoy, A.N., Araújo-Pérez, F., Azcárate-Peril, A., Yeh, J.J., Sandler, R.S., and Keku, T.O. (2013). Fusobacterium Is Associated with Colorectal Adenomas. PLOS ONE *8*, e53653. 10.1371/journal.pone.0053653.

92. Mima, K., Sukawa, Y., Nishihara, R., Qian, Z.R., Yamauchi, M., Inamura, K., Kim, S.A., Masuda, A., Nowak, J.A., Nosho, K., et al. (2015). Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. JAMA Oncol *1*, 653–661. 10.1001/jamaoncol.2015.1377.

93. Mima, K., Nishihara, R., Qian, Z.R., Cao, Y., Sukawa, Y., Nowak, J.A., Yang, J., Dou, R., Masugi, Y., Song, M., et al. (2016). Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut *65*, 1973–1980. 10.1136/gutjnl-2015-310101.

94. Florin, T.H.J. (1991). Hydrogen sulphide and total acid-volatile sulphide in faeces, determined with a direct spectrophotometric method. Clinica Chimica Acta *196*, 127–134. 10.1016/0009-8981(91)90065-K.

95. Chacko, A., and Cummings, J.H. (1988). Nitrogen losses from the human small bowel: obligatory losses and the effect of physical form of food. Gut *29*, 809–815.

96. Scanlan, P.D., Shanahan, F., and Marchesi, J.R. (2009). Culture-independent analysis of desulfovibrios in the human distal colon of healthy, colorectal cancer and polypectomized individuals. FEMS Microbiology Ecology *69*, 213–221. 10.1111/j.1574-6941.2009.00709.x.

97. Cai, W.-J., Wang, M.-J., Ju, L.-H., Wang, C., and Zhu, Y.-C. (2010). Hydrogen sulfide induces human colon cancer cell proliferation: Role of Akt, ERK and p21. Cell Biology International *34*, 565–572. https://doi.org/10.1042/CBI20090368.

98. Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsolis, R.M., et al. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. Nature *467*, 426–429. 10.1038/nature09415.

99. Braccia, D.J., Jiang, X., Pop, M., and Hall, A.B. (2021). The capacity to produce hydrogen sulfide (H2S) via cysteine degradation is ubiquitous in the human gut microbiome. bioRxiv, 2021.03.24.436781. 10.1101/2021.03.24.436781.

100. Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Jacob, B., Ratner, A., Liolios, K., Pagani, I., Huntemann, M., et al. (2012). IMG/M-HMP: A Metagenome Comparative Analysis System for the Human Microbiome Project. PLOS ONE *7*, e40151. 10.1371/journal.pone.0040151.

101. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The Human Microbiome Project. Nature *449*, 804–810. 10.1038/nature06244.

102. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410. 10.1016/S0022-2836(05)80360-2.

103. Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., et al. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nature Medicine *25*, 667–678. 10.1038/s41591-019-0405-7.

104. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nature Medicine *25*, 679–689. 10.1038/s41591-019-0406-6.

105. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology *35*, 725–731. 10.1038/nbt.3893.

106. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119. 10.1186/1471-2105-11-119.

107. Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., and Anantharaman, K. (2020). METABOLIC: High-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. bioRxiv, 761643. 10.1101/761643.

108. HMMER http://hmmer.org/.

109. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. 10.1101/gr.186072.114.

110. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project:

improved data processing and web-based tools. Nucleic Acids Research *41*, D590–D596. 10.1093/nar/gks1219.

111. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glöckner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research *42*, D643–D648. 10.1093/nar/gkt1209.

112. Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics *28*, 1823–1829. 10.1093/bioinformatics/bts252.

113. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol *36*, 996–1004. 10.1038/nbt.4229.

114. Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol *38*, 1079–1086. 10.1038/s41587-020-0501-8.

115. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics *36*, 1925–1927. 10.1093/bioinformatics/btz848.

116. Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature Communications *7*, 13219. 10.1038/ncomms13219.

117. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol *30*, 772–780. 10.1093/molbev/mst010.

118. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution *32*, 268–274. 10.1093/molbev/msu300.

119. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research *47*, W256–W259. 10.1093/nar/gkz239.

120. SRA-Tools - NCBI http://ncbi.github.io/sra-tools/.

121. Aspera - Connect (2019). https://www.ibm.com/aspera/connect.

122. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis | Microbiome | Full Text https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0541-1.

123. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. 10.1038/nmeth.1923.

124. Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics *35*, 421–432. 10.1093/bioinformatics/bty648.

125. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25. 10.1186/gb-2009-10-3-r25.

126. Thomas, B. (2013). bcthomas/shrinksam.

127. Brown, C.T., Olm, M.R., Thomas, B.C., and Banfield, J.F. (2016). Measurement of bacterial replication rates in microbial communities. Nature Biotechnology *34*, 1256–1263. 10.1038/nbt.3704.

128. Schuchat, A. (1998). Epidemiology of Group B Streptococcal Disease in the United States: Shifting Paradigms. Clin Microbiol Rev *11*, 497–513. 10.1128/CMR.11.3.497.

129. Armistead, B., Oler, E., Adams Waldorf, K., and Rajagopal, L. (2019). The Double Life of Group B Streptococcus: Asymptomatic Colonizer and Potent Pathogen. Journal of Molecular Biology *431*, 2914–2931. 10.1016/j.jmb.2019.01.035.

130. Skoff, T.H., Farley, M.M., Petit, S., Craig, A.S., Schaffner, W., Gershman, K., Harrison, L.H., Lynfield, R., Mohle-Boetani, J., Zansky, S., et al. (2009). Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990-2007. Clin Infect Dis *49*, 85–92. 10.1086/599369.

131. Schwartz, B., Schuchat, A., Oxtoby, M.J., Cochi, S.L., Hightower, A., and Broome, C.V. (1991). Invasive group B streptococcal disease in adults. A population-based study in metropolitan Atlanta. JAMA *266*, 1112–1114.

132. Farley, M.M., Harvey, C., Stull, T., Smith, J.D., Schuchat, A., Wenger, J.D., and Stephens, D.S. (1993). A Population-Based Assessment of Invasive Disease Due to Group B Streptococcus in Nonpregnant Adults. N Engl J Med *328*, 1807–1811. 10.1056/NEJM199306243282503.

133. Russell, N.J., Seale, A.C., O'Driscoll, M., O'Sullivan, C., Bianchi-Jassir, F., Gonzalez-Guarin, J., Lawn, J.E., Baker, C.J., Bartlett, L., Cutland, C., et al. (2017). Maternal Colonization With Group B Streptococcus and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. Clinical Infectious Diseases *65*, S100–S111. 10.1093/cid/cix658.

134. Verani, J.R., McGee, L., Schrag, S.J., and Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC) (2010). Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010. MMWR Recomm Rep *59*, 1–36.

135. Edwards, J.M., Watson, N., Focht, C., Wynn, C., Todd, C.A., Walter, E.B., Heine, R.P., and Swamy, G.K. (2019). Group B Streptococcus (GBS) Colonization and Disease among Pregnant Women: A Historical Cohort Study. Infectious Diseases in Obstetrics and Gynecology *2019*, e5430493. https://doi.org/10.1155/2019/5430493.

136. Krohn, M.A., Hillier, S.L., and Baker, C.J. (1999). Maternal Peripartum Complications Associated with Vaginal Group B Streptococci Colonization. The Journal of Infectious Diseases *179*, 1410–1415. 10.1086/314756.

137. Deutscher, M., Lewis, M., Zell, E.R., Taylor, T.H., Jr., Van Beneden, C., Schrag, S., and for the Active Bacterial Core Surveillance Team (2011). Incidence and Severity of Invasive Streptococcus pneumoniae, Group A Streptococcus, and Group B Streptococcus Infections Among Pregnant and Postpartum Women. Clinical Infectious Diseases *53*, 114–123. 10.1093/cid/cir325.

138. Seale, A.C., Bianchi-Jassir, F., Russell, N.J., Kohli-Lynch, M., Tann, C.J., Hall, J., Madrid, L., Blencowe, H., Cousens, S., Baker, C.J., et al. (2017). Estimates of the Burden of Group B Streptococcal Disease Worldwide for Pregnant Women, Stillbirths, and Children. Clinical Infectious Diseases *65*, S200–S219. 10.1093/cid/cix664.

139. Wood, E.G., and Dillon, H.C. (1981). A prospective study of group B streptococcal bacteriuria in pregnancy. Am J Obstet Gynecol *140*, 515–520. 10.1016/0002-9378(81)90226-x.

140. Persson, K., Bjerre, B., Elfström, L., Polberger, S., and Forsgren, A. (1986). Group B streptococci at delivery: high count in urine increases risk for neonatal colonization. Scand J Infect Dis *18*, 525–531. 10.3109/00365548609021657.

141. Melin, P. (2011). Neonatal group B streptococcal disease: from pathogenesis to preventive strategies. Clin Microbiol Infect *17*, 1294–1303. 10.1111/j.1469-0691.2011.03576.x.

142. Madrid, L., Seale, A.C., Kohli-Lynch, M., Edmond, K.M., Lawn, J.E., Heath, P.T., Madhi, S.A., Baker, C.J., Bartlett, L., Cutland, C., et al. (2017). Infant Group B Streptococcal Disease Incidence and Serotypes Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis *65*, S160–S172. 10.1093/cid/cix656.

143. Edmond, K.M., Kortsalioudaki, C., Scott, S., Schrag, S.J., Zaidi, A.K.M., Cousens, S., and Heath, P.T. (2012). Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet *379*, 547–556. 10.1016/S0140-6736(11)61651-6.

144. Badri, M.S., Zawaneh, S., Cruz, A.C., Mantilla, G., Baer, H., Spellacy, W.N., and Ayoub, E.M. (1977). Rectal Colonization with Group B Streptococcus: Relation to Vaginal Colonization of Pregnant Women. The Journal of Infectious Diseases *135*, 308–312. 10.1093/infdis/135.2.308.

145. MEYN, L.A., KROHN, M.A., and HILLIER, S.L. (2009). Rectal Colonization by Group B Streptococcus as a Predictor of Vaginal Colonization. Am J Obstet Gynecol *201*, 76.e1-76.e7. 10.1016/j.ajog.2009.02.011.

146. Francois Watkins, L.K., McGee, L., Schrag, S.J., Beall, B., Jain, J.H., Pondo, T., Farley, M.M., Harrison, L.H., Zansky, S.M., Baumbach, J., et al. (2019). Epidemiology of Invasive Group B Streptococcal Infections Among Nonpregnant Adults in the United States, 2008-2016. JAMA Intern Med *179*, 479. 10.1001/jamainternmed.2018.7269.

147. Langley, G., Schaffner, W., Farley, M.M., Lynfield, R., Bennett, N.M., Reingold, A., Thomas, A., Harrison, L.H., Nichols, M., Petit, S., et al. (2015). Twenty Years of Active Bacterial Core Surveillance. Emerg Infect Dis *21*, 1520–1528. 10.3201/eid2109.141333.

148. Phares, C.R. (2008). Epidemiology of Invasive Group B Streptococcal Disease in the United States, 1999-2005. JAMA *299*, 2056. 10.1001/jama.299.17.2056.

149. Prevention of Group B Streptococcal Early-Onset Disease in Newborns: ACOG Committee Opinion Summary, Number 797 (2020). Obstetrics & Gynecology *135*, 489–492. 10.1097/AOG.0000000000003669.

150. Pitts, S.I., Maruthur, N.M., Langley, G.E., Pondo, T., Shutt, K.A., Hollick, R., Schrag, S.J., Thomas, A., Nichols, M., Farley, M., et al. (2018). Obesity, Diabetes, and the Risk of Invasive Group B Streptococcal Disease in Nonpregnant Adults in the United States. Open Forum Infectious Diseases *5*, ofy030. 10.1093/ofid/ofy030.

151. Manning, S.D., Neighbors, K., Tallman, P.A., Gillespie, B., Marrs, C.F., Borchardt, S.M., Baker, C.J., Pearlman, M.D., and Foxman, B. (2004). Prevalence of Group B Streptococcus Colonization and Potential for Transmission by Casual Contact in Healthy Young Men and Women. Clinical Infectious Diseases *39*, 380–388. 10.1086/422321.

152. Bliss, S.J., Manning, S.D., Tallman, P., Baker, C.J., Pearlman, M.D., Marrs, C.F., and Foxman, B. (2002). Group B Streptococcus Colonization in Male and Nonpregnant Female University Students: A Cross-Sectional Prevalence Study. Clinical Infectious Diseases *34*, 184–190. 10.1086/338258.

153. Edwards, M.S., Rench, M.A., Palazzi, D.L., and Baker, C.J. (2005). Group B Streptococcal Colonization and Serotype-Specific Immunity in Healthy Elderly Persons. Clinical Infectious Diseases *40*, 352–357. 10.1086/426820.

154. van Kassel, M.N., Janssen, S.W.C.M., Kofman, S., Brouwer, M.C., van de Beek, D., and Bijlsma, M.W. (2021). Prevalence of group B streptococcal colonization in the healthy non-pregnant population: a systematic review and meta-analysis. Clinical Microbiology and Infection *27*, 968–980. 10.1016/j.cmi.2021.03.024.

155. Martins, E.R., Nascimento do Ó, D., Marques Costa, A.L., Melo-Cristino, J., and Ramirez, M. (2022). Characteristics of Streptococcus agalactiae Colonizing Nonpregnant Adults Support the Opportunistic Nature of Invasive Infections. Microbiology Spectrum *10*, e01082-22. 10.1128/spectrum.01082-22.

156. Eggers, S., Malecki, K.M., Peppard, P., Mares, J., Shirley, D., Shukla, S.K., Poulsen, K., Gangnon, R., Duster, M., Kates, A., et al. (2018). Wisconsin microbiome study, a cross-sectional investigation of dietary fibre, microbiome composition and antibiotic-resistant organisms: rationale and methods. BMJ Open *8*, e019450. 10.1136/bmjopen-2017-019450.

157. Nieto, F.J., Peppard, P.E., Engelman, C.D., McElroy, J.A., Galvao, L.W., Friedman, E.M., Bersch, A.J., and Malecki, K.C. (2010). The Survey of the Health of Wisconsin (SHOW), a novel infrastructure for population health research: rationale and methods. BMC Public Health *10*, 785. 10.1186/1471-2458-10-785.

158. Bergh, K., Stoelhaug, A., Loeseth, K., and Bevanger, L. (2004). Detection of group B streptococci (GBS) in vaginal swabs using real-time PCR with TaqMan probe hybridization. Indian J Med Res *119*, 221–223.

159. Bergseng, H., Bevanger, L., Rygg, M., and Bergh, K. (2007). Real-time PCR targeting the sip gene for detection of group B streptococcus colonization in pregnant women at delivery. Journal of Medical Microbiology *56*, 223–228. 10.1099/jmm.0.46731-0.

160. Escobar, D.F., Diaz-Dinamarca, D.A., Hernández, C.F., Soto, D.A., Manzo, R.A., Alarcón, P.I., Pinto, C.H., Bastias, D.N., Oberg-Bravo, C.N., Rojas, R., et al. (2020). Development and analytical validation of real-time PCR for the detection of Streptococcus agalactiae in pregnant women. BMC Pregnancy Childbirth *20*. 10.1186/s12884-020-03038-z.

161. Carrillo-Ávila, J.A., Gutiérrez-Fernández, J., González-Espín, A.I., García-Triviño, E., and Giménez-Lirola, L.G. (2018). Comparison of qPCR and culture methods for group B Streptococcus colonization detection in pregnant women: evaluation of a new qPCR assay. BMC Infectious Diseases *18*, 305. 10.1186/s12879-018-3208-4.

162. Khatami, A., Randis, T.M., Chamby, A., Hooven, T.A., Gegick, M., Suzman, E., A'Hearn-Thomas, B., Steenhoff, A.P., and Ratner, A.J. (2018). Improving the Sensitivity of Real-time PCR Detection of Group B Streptococcus Using Consensus Sequence-Derived Oligonucleotides. Open Forum Infect Dis *5*. 10.1093/ofid/ofy164.

163. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol *37*, 852–857. 10.1038/s41587-019-0209-9.

164. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods *13*, 581–583. 10.1038/nmeth.3869.

165. Katoh, K., Misawa, K., Kuma, K.I., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research *30*, 3059–3066. 10.1093/nar/gkf436.

166. Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 1–17. 10.1186/s40168-018-0470-z.

167. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research *41*, 590–596. 10.1093/nar/gks1219.

168. McMurdie, P.J., Holmes, S., Kindt, R., Legendre, P., and O'Hara, R. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE *8*, e61217. 10.1371/journal.pone.0061217.

169. Davis, N.M., Proctor, Di.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 1–14. 10.1186/s40168-018-0605-2.

170. Lin, H., and Peddada, S. Das (2020). Analysis of compositions of microbiomes with bias correction. Nature Communications *11*, 1–11. 10.1038/s41467-020-17041-7.

171. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., Tabbaa, D., Highlander, S., Sodergren, E., et al. (2011). Metagenomic biomarker discovery and explanation. Genome Biology *12*, R60. 10.1186/gb-2011-12-6-r60.

172. Cao, Y., Dong, Q., Wang, D., Zhang, P., Liu, Y., and Niu, C. (2022). microbiomeMarker: an R/Bioconductor package for microbiome marker identification and visualization. Bioinformatics *38*, 4027–4029. 10.1093/bioinformatics/btac438.

173. Tang, Z.Z., Chen, G., Alekseyenko, A. V., and Li, H. (2017). A general framework for association analysis of microbial communities on a taxonomic tree. Bioinformatics *33*, 1278–1285. 10.1093/bioinformatics/btw804.

174. Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. R News *2*, 18–22.

175. Kwatra, G., Adrian, P.V., Shiri, T., Buchmann, E.J., Cutland, C.L., and Madhi, S.A. (2014). Serotype-Specific Acquisition and Loss of Group B Streptococcus Recto-Vaginal Colonization in Late Pregnancy. PLOS ONE *9*, e98778. 10.1371/journal.pone.0098778.

176. Abranches, J., Zeng, L., Kajfasz, J.K., Palmer, S.R., Chakraborty, B., Wen, Z.T., Richards, V.P., Brady, L.J., and Lemos, J.A. (2018). Biology of Oral Streptococci. Microbiology Spectrum *6*, 10.1128/microbiolspec.gpp3-0042–2018. 10.1128/microbiolspec.gpp3-0042-2018.

177. Hillman, E.T., Lu, H., Yao, T., and Nakatsu, C.H. (2017). Microbial Ecology along the Gastrointestinal Tract. Microbes and environments *32*, 300–313. 10.1264/jsme2.ME17017.

178. Cassat, J.E., and Skaar, E.P. (2013). Iron in Infection and Immunity. Cell Host & Microbe *13*, 509–519. 10.1016/j.chom.2013.04.010.

179. Clancy, A., Loar, J.W., Speziali, C.D., Oberg, M., Heinrichs, D.E., and Rubens, C.E. (2006). Evidence for siderophore-dependent iron acquisition in group B streptococcus. Molecular Microbiology *59*, 707–721. 10.1111/j.1365-2958.2005.04974.x.

180. Yilmaz, B., and Li, H. (2018). Gut Microbiota and Iron: The Crucial Actors in Health and Disease. Pharmaceuticals (Basel) *11*, 98. 10.3390/ph11040098.

181. Rusu, I.G., Suharoschi, R., Vodnar, D.C., Pop, C.R., Socaci, S.A., Vulturar, R., Istrati, M., Moroșan, I., Fărcaș, A.C., Kerezsi, A.D., et al. (2020). Iron Supplementation Influence on the Gut Microbiota and Probiotic Intake Effect in Iron Deficiency—A Literature-Based Review. Nutrients *12*, 1993. 10.3390/nu12071993.

182. Dahesh, S., Hensler, M.E., Van Sorge, N.M., Gertz, R.E., Schrag, S., Nizet, V., and Beall, B.W. (2008). Point Mutation in the Group B Streptococcal pbp2x Gene Conferring Decreased Susceptibility to β-Lactam Antibiotics. Antimicrobial Agents and Chemotherapy *52*, 2915–2918. 10.1128/aac.00461-08.

183. Longtin, J., Vermeiren, C., Shahinas, D., Tamber, G.S., McGeer, A., Low, D.E., Katz, K., and Pillai, D.R. (2011). Novel Mutations in a Patient Isolate of Streptococcus agalactiae with Reduced Penicillin Susceptibility Emerging after Long-Term Oral Suppressive Therapy. Antimicrobial Agents and Chemotherapy *55*, 2983–2985. 10.1128/aac.01243-10.

184. Kimura, K., Suzuki, S., Wachino, J., Kurokawa, H., Yamane, K., Shibata, N., Nagano, N., Kato, H., Shibayama, K., and Arakawa, Y. (2008). First molecular characterization of group B streptococci with reduced penicillin susceptibility. Antimicrob Agents Chemother *52*, 2890–2897. 10.1128/aac.00185-08.

185. Persson, E., Berg, S., Bergseng, H., Bergh, K., Valsö-lyng, R., and Trollfors, B. (2008). Antimicrobial susceptibility of invasive group B streptococcal isolates from south-west Sweden 1988–2001. Scandinavian Journal of Infectious Diseases *40*, 308–313. 10.1080/00365540701678702.

186. Simoes, J.A., Aroutcheva, A.A., Heimler, I., and Faro, S. (NaN/NaN/NaN). Antibiotic Resistance Patterns of Group B Streptococcal Clinical Isolates. Infectious Diseases in Obstetrics and Gynecology *12*, 1–8. 10.1080/10647440410001722269.

187. Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A. (2008). The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. PLOS Biology *6*, e280. 10.1371/journal.pbio.0060280.

188. Ramagopalan, S.V., and Sadovnick, A.D. (2011). Epidemiology of multiple sclerosis. Neurol Clin *29*, 207–217. 10.1016/j.ncl.2010.12.010.

189. Noseworthy, J.H., Lucchinetti, C., Rodriguez, M., and Weinshenker, B.G. (2000). Multiple sclerosis. N Engl J Med *343*, 938–952. 10.1056/NEJM200009283431307.

190. Alonso, A., and Hernán, M.A. (2008). Temporal trends in the incidence of multiple sclerosis: a systematic review. Neurology *71*, 129–135. 10.1212/01.wnl.0000316802.35974.34.

191. Goodin, D.S. (2014). The epidemiology of multiple sclerosis: insights to disease pathogenesis. Handb Clin Neurol *122*, 231–266. 10.1016/B978-0-444-52001-2.00010-8.

192. Weiner, H.L. (2004). Multiple sclerosis is an inflammatory T-cell-mediated autoimmune disease. Arch Neurol *61*, 1613–1615. 10.1001/archneur.61.10.1613.

193. Weinshenker, B.G. (1994). Natural history of multiple sclerosis. Ann Neurol *36 Suppl*, S6-11. 10.1002/ana.410360704.

194. International Multiple Sclerosis Genetics Consortium (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science *365*. 10.1126/science.aav7188.

195. Hernán, M.A., Zhang, S.M., Lipworth, L., Olek, M.J., and Ascherio, A. (2001). Multiple sclerosis and age at infection with common viruses. Epidemiology *12*, 301–306. 10.1097/00001648-200105000-00009.

196. GBD 2016 Multiple Sclerosis Collaborators (2019). Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol *18*, 269–285. 10.1016/S1474-4422(18)30443-5.

197. Bjornevik, K., Cortese, M., Healy, B.C., Kuhle, J., Mina, M.J., Leng, Y., Elledge, S.J., Niebuhr, D.W., Scher, A.I., Munger, K.L., et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science *375*, 296–301. 10.1126/science.abj8222.

198. Mielcarz, D.W., and Kasper, L.H. (2015). The gut microbiome in multiple sclerosis. Curr Treat Options Neurol *17*, 344. 10.1007/s11940-015-0344-7.

199. Bhargava, P., and Mowry, E.M. (2014). Gut microbiome and multiple sclerosis. Curr Neurol Neurosci Rep *14*, 492. 10.1007/s11910-014-0492-2.

200. Berer, K., Mues, M., Koutrolos, M., Rasbi, Z.A., Boziki, M., Johner, C., Wekerle, H., and Krishnamoorthy, G. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. Nature *479*, 538–541. 10.1038/nature10554.

201. Ochoa-Repáraz, J., Mielcarz, D.W., Ditrio, L.E., Burroughs, A.R., Foureau, D.M., Haque-Begum, S., and Kasper, L.H. (2009). Role of Gut Commensal Microflora in the Development of Experimental Autoimmune Encephalomyelitis1. The Journal of Immunology *183*, 6041–6050. 10.4049/jimmunol.0900747.

202. Lee, Y.K., Menezes, J.S., Umesaki, Y., and Mazmanian, S.K. (2011). Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. Proceedings of the National Academy of Sciences *108*, 4615–4622. 10.1073/pnas.1000082107.

203. He, B., Hoang, T.K., Tian, X., Taylor, C.M., Blanchard, E., Luo, M., Bhattacharjee, M.B., Freeborn, J., Park, S., Couturier, J., et al. (2019). Lactobacillus reuteri Reduces the Severity of Experimental Autoimmune Encephalomyelitis in Mice by Modulating Gut Microbiota. Front. Immunol. *10*. 10.3389/fimmu.2019.00385.

204. Camarillo-Guerrero, L.F., Almeida, A., Rangel-Pineros, G., Finn, R.D., and Lawley, T.D. (2020). Massive expansion of human gut bacteriophage diversity. bioRxiv, 2020.09.03.280214. 10.1101/2020.09.03.280214.

205. Mangalea, M.R., Paez-Espino, D., Kieft, K., Chatterjee, A., Chriswell, M.E., Seifert, J.A., Feser, M.L., Demoruelle, M.K., Sakatos, A., Anantharaman, K., et al. (2021). Individuals at risk for rheumatoid arthritis harbor differential intestinal bacteriophage communities with distinct metabolic potential. Cell Host & Microbe *0*. 10.1016/j.chom.2021.03.020.

206. Elsayed, N.S., Valenzuela, R.K., Kitchner, T., Le, T., Mayer, J., Tang, Z.-Z., Bayanagari, V.R., Lu, Q., Aston, P., Anantharaman, K., et al. (2023). Genetic risk score in multiple sclerosis is associated with unique gut microbiome. Sci Rep *13*, 16269. 10.1038/s41598-023-43217-4.

207. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome *6*, 158. 10.1186/s40168-018-0541-1.

208. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. CP in Bioinformatics *70*, e102. 10.1002/cpbi.102.

209. Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome *8*, 90. 10.1186/s40168-020-00867-0.

210. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol *3*, 836–843. 10.1038/s41564-018-0171-1.

211. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research *50*, D785–D794. 10.1093/nar/gkab776.

212. Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.-A., Davín, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., and Hugenholtz, P. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. Nat Microbiol *6*, 946–959. 10.1038/s41564-021-00918-8.

213. Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. Nature Biotechnology, 1–8. 10.1038/s41587-020-0501-8.

214. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol *36*, 996–1004. 10.1038/nbt.4229.

215. Matsen, F.A., Kodner, R.B., and Armbrust, E.V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics *11*, 538. 10.1186/1471-2105-11-538.

216. Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun *9*, 5114. 10.1038/s41467-018-07641-9.

217. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119. 10.1186/1471-2105-11-119.

218. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE *5*, e9490. 10.1371/journal.pone.0009490.

219. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLOS Computational Biology *7*, e1002195. 10.1371/journal.pcbi.1002195.

220. Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology *17*, 132. 10.1186/s13059-016-0997-x.

221. Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun *7*, 13219. 10.1038/ncomms13219.

222. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534. 10.1093/molbev/msaa015.

223. Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., and Anantharaman, K. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome *10*, 33. 10.1186/s40168-021-01213-8.

224. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550. 10.1186/s13059-014-0550-8.

225. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2023). Identification of mobile genetic elements with geNomad. Nat Biotechnol, 1–10. 10.1038/s41587-023-01953-y.

226. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell *176*, 649-662.e20. 10.1016/j.cell.2019.01.001.

227. Manrique, P., Dills, M., and Young, M.J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. Viruses *9*, 141. 10.3390/v9060141.

228. Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature *466*, 334–338. 10.1038/nature09199.

229. Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. Proceedings of the National Academy of Sciences *113*, 10400–10405. 10.1073/pnas.1601060113.

230. Castro-Mejía, J.L., Muhammed, M.K., Kot, W., Neve, H., Franz, C.M.A.P., Hansen, L.H., Vogensen, F.K., and Nielsen, D.S. (2015). Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. Microbiome *3*, 64. 10.1186/s40168-015-0131-4.

231. Smith, L., Goldobina, E., Govi, B., and Shkoporov, A.N. (2023). Bacteriophages of the Order Crassvirales: What Do We Currently Know about This Keystone Component of the Human Gut Virome? Biomolecules *13*. 10.3390/biom13040584.

232. Zuppi, M., Hendrickson, H.L., O'Sullivan, J.M., and Vatanen, T. (2022). Phages in the Gut Ecosystem. Front Cell Infect Microbiol *11*, 822562. 10.3389/fcimb.2021.822562.

233. Duscha, A., Gisevius, B., Hirschberg, S., Yissachar, N., Stangl, G.I., Dawin, E., Bader, V., Haase, S., Kaisler, J., David, C., et al. (2020). Propionic Acid Shapes the Multiple Sclerosis Disease Course by an Immunomodulatory Mechanism. Cell *180*, 1067-1080.e16. 10.1016/j.cell.2020.02.035.

234. Takewaki, D., Suda, W., Sato, W., Takayasu, L., Kumar, N., Kimura, K., Kaga, N., Mizuno, T., Miyake, S., Hattori, M., et al. (2020). Alterations of the gut ecological and functional microenvironment in different stages of multiple sclerosis. Proceedings of the National Academy of Sciences *117*, 22402–22412. 10.1073/pnas.2011703117.

235. Park, J., Wang, Q., Wu, Q., Mao-Draayer, Y., and Kim, C.H. (2019). Bidirectional regulatory potentials of short-chain fatty acids and their G-protein-coupled receptors in autoimmune neuroinflammation. Sci Rep *9*, 8837. 10.1038/s41598-019-45311-y.

236. Saresella, M., Marventano, I., Barone, M., La Rosa, F., Piancone, F., Mendozzi, L., d'Arma, A., Rossi, V., Pugnetti, L., Roda, G., et al. (2020). Alterations in Circulating Fatty Acid Are Associated With Gut Microbiota Dysbiosis and Inflammation in Multiple Sclerosis. Frontiers in Immunology *11*.

237. Bhargava, P., Smith, M.D., Mische, L., Harrington, E., Fitzgerald, K.C., Martin, K., Kim, S., Reyes, A.A., Gonzalez-Cardona, J., Volsko, C., et al. (2020). Bile acid metabolism is altered in multiple sclerosis and supplementation ameliorates neuroinflammation. J Clin Invest *130*, 3467–3482. 10.1172/JCI129401.

238. Baidina, T.V., Trushnikova, T.N., and Danilova, M.A. (2018). [Interferon-induced depression and peripheral blood serotonin in patients with multiple sclerosis]. Zh Nevrol Psikhiatr Im S S Korsakova *118*, 77–81. 10.17116/jnevro201811808277.

239. Nourbakhsh, B., Bhargava, P., Tremlett, H., Hart, J., Graves, J., and Waubant, E. (2018). Altered tryptophan metabolism is associated with pediatric multiple sclerosis risk and course. Annals of Clinical and Translational Neurology *5*, 1211–1221. 10.1002/acn3.637.

240. Lim, C.K., Bilgin, A., Lovejoy, D.B., Tan, V., Bustamante, S., Taylor, B.V., Bessede, A., Brew, B.J., and Guillemin, G.J. (2017). Kynurenine pathway metabolomics predicts and provides mechanistic insight into multiple sclerosis progression. Sci Rep *7*, 41473. 10.1038/srep41473.

241. Fettig, N.M., and Osborne, L.C. (2021). Direct and indirect effects of microbiota-derived metabolites on neuroinflammation in multiple sclerosis. Microbes and Infection *23*, 104814. 10.1016/j.micinf.2021.104814.

242. Montgomery, T.L., Eckstrom, K., Lile, K.H., Caldwell, S., Heney, E.R., Lahue, K.G., D'Alessandro, A., Wargo, M.J., and Krementsov, D.N. (2022). Lactobacillus reuteri tryptophan metabolism promotes host susceptibility to CNS autoimmunity. Microbiome *10*, 198. 10.1186/s40168-022-01408-7.

243. Ascherio, A., and Munger, K.L. (2007). Environmental risk factors for multiple sclerosis. Part I: The role of infection. Annals of Neurology *61*, 288–299. 10.1002/ana.21117.

244. Lanz, T.V., Brewer, R.C., Ho, P.P., Moon, J.-S., Jude, K.M., Fernandez, D., Fernandes, R.A., Gomez, A.M., Nadj, G.-S., Bartley, C.M., et al. (2022). Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. Nature *603*, 321–327. 10.1038/s41586-022-04432-7.

Appendix A – Characterization of bacterial composition of surgical site infections after gynecological surgery

All appendices represent work outside of my main dissertation that I worked on that has either been published, has been accepted, or has received a revise and resubmit decision.

During my PhD, I was fortunate to work with Dr. Laura Jacques on several projects separate from my main dissertation research. The following appendices A-F were projects completed under her mentorship.

This appendix has been published in *America Journal of Obstetrics and Gynecology:* Cowley, E.S., Jacques, L., Powell, A.M., Al-Niaimi, A., Pop-Vicas, A. Characterization of bacterial composition of surgical site infections after gynecological surgery. *AJOG.* 227, 2 (2022). https://doi.org/10.1016/j.ajog.2022.02.033.

Title: Characterization of bacterial composition of surgical site infections after gynecological surgery

Authors: **Ms Elise S. Cowley**[1,2] BS, Dr Laura Jacques[3] MD, Dr Anna M. Powell[4] MD, Dr Ahmed Al-Niaimi[3,5] MD, Dr Aurora Pop-Vicas* MD

*Corresponding author: popvicas@medicine.wisc.edu

Affiliations

[1]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, USA

[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA

[3]Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI USA

[4]Department of Gynecology and Obstetrics, Johns Hopkins School of Medicine, Baltimore, Maryland

[5]Carbone Cancer Center, University of Wisconsin-Madison, Madison, WI USA

[6]Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI USA

Keywords: Anaerobic bacteria, Commensal bacteria, Deep incisional surgical site infection, Deep/organ surgical site infection, Gynecological surgery, Intestinal bacteria, Microbiological culturing, Pre-operative antibiotics, Skin bacteria, Superficial surgical site infection, Surgical site infections, Vaginal bacteria

Objective: Surgical site infections (SSIs) occur within 30 days of an operation, superficially at the incision site, deeply at the incision site, or in organ and deep spaces accessed during the operation[1]. SSIs are the most common complication of gynecological surgeries and can result in significant morbidity[2,3]. Specifically, hysterectomies have an SSI complication rate of 1-2%[4], with most pathogens arising from the endogenous

microbes of the skin, vagina, and gastrointestinal tract. Here, we characterize the bacterial composition of gynecological SSIs to help inform prophylactic antibiotic choices.

Study Design We retrospectively reviewed all microbiological cultures from patients with SSIs, as part of a quality improvement departmental initiative using an institutional infection control database that includes all gynecological surgeries for benign and malignant indications at an academic tertiary care center during 2010-2020.

Results: Among 10,495 gynecologic surgeries performed during the ten-year span, 192 (1.8%) SSIs were diagnosed. 142/192 (74%) occurred in cases that involved hysterectomies, (8 vaginal, 134 abdominal) and 157/192 (82%) of SSIs occurred among patients with suspected preoperative gynecologic malignancies. Of all the gynecological surgeries, 4,818 (46%) involved hysterectomies with 141 (2.9%) complicated by SSIs. For all gynecological cases, the median surgery duration was 2.15 hours [IQR: 2.43 hours], while for cases involving hysterectomies it was 3.15 hours [IQR: 1.7 hours]. For cases with SSIs, the median surgery duration was 4.1 hours [IQR: 2.27] with a median time to infection of 13 days [IQR: 8 days].

Pre-operative antibiotics were administered to 186 (97%) patients who developed SSIs, including cefoxitin (44%), cefazolin (19%), cefazolin and metronidazole (10%), and non-beta-lactam regimens (22%). Pre-operative antibiotics were administered to 78% of patients who did not developed SSIs, including cefoxitin (48%), cefazolin (23%), cefazolin and metronidazole (12%), and non-beta-lactam regimens (14%). All patients

who developed an SSI received preoperative surgical site skin preparation with either chlorohexidine (76%) or iodine-based preparations (24%) in the operating room.

Table 1 shows the classification of microbiological results by SSI location type. Strict anaerobic species were isolated from a total of 52 SSIs (27.0%), primarily from organ/space SSIs (67%). These included *Anaerococcus, Arcanobacterium pyogenes, Bacteroides fragilis* and other *Bacteroides sp*., *Clostridium sp., Eggerthella lenta, Fusobacterium sp., Peptostreptococcus sp*., *Peptoniphilus sp., Prevotella sp., Porphyromonas sp., Propionebacterium avidum, Ruminococcus gnavus*, and *Veillonella sp.*

Table 1. SSI characteristics by type of infection and type of bacteria.

	Organ/space (intra-abdominal) SSI N = 84 n (%)	**Deep incisional SSI N = 18** n (%)	**Superficial SSI N = 90** n (%)
Skin bacteria[a]	5 (6.0)	4 (22.2)	26 (28.9)
Skin and intestinal or vaginal bacteria	13 (15.4)	3 (16.7)	12 (13.3)
Intestinal or vaginal bacteria[b]	51 (60.7)	10 (55.5)	28 (31.1)
Other* and intestinal or vaginal bacteria[b]	4 (4.8)	0 (0)	1 (1.1)
Other*	2 (2.4)	0 (0)	3 (3.3)
Not cultured	9 (10.7)	1 (5.6)	20 (22.3)

[a]**Skin bacteria**: *Corynebacterium sp., Propionibacterium sp*., *Staphylococcus aureus, Staphylococcus coagulase negative sp.,* gram-positive rods
[b]**Intestinal or vaginal bacteria**: *Actinomyces, Anaerococcus sp., Atopobium vaginae, Bacteroides fragilis,* other *Bacteroides sp., Candida albicans, Candida dubiliniensis, Citrobacter sp, Clostridium sp., Eggerthella lenta, Enterobacter sp., Enterococcus*

faecalis, Enterococcus faecium, Escherichia coli, Fusobacterium sp., Gardnerella vaginalis, Gemella morbillorum, Klebsiella sp., Lactobacillus, Morganella morganii, Peptostreptococcus sp., Peptoniphilus sp., Pseudomonas aeruginosa, *Prevotella sp., Porphyromonas sp., Ruminococcus gnavus, Salmonella sp*.**,** *Streptococcus agalactiae, Streptococcus anginosus, Streptococcus viridans* group, *Veilllonella sp.,* anaerobic gram-positive cocci, anaerobic gram-positive rods
*__Other__: *Acinetobacter sp., Arcanobacterium pyogenes,* Methicillin-Resistant *Staphylococcus aureus, Pasteurella multocida.*

Conclusions: Superficial and deep incisional SSIs are predominantly associated with bacteria typically found in the intestine and vagina, although skin bacteria are also commonly present. Organ/space SSIs are predominantly associated with intestinal bacteria, with anaerobes being recovered in more than three-quarters of these infections. Previous work has shown that cefazolin and metronidazole are superior at preventing SSIs in patients undergoing hysterectomies, particularly anaerobic SSIs, than with only a cephalosporin[5]. Given the prevalence of anaerobic organisms that we found in all types of SSIs, further research should compare the effectiveness of preoperative antimicrobial prophylactic agents in preventing anaerobic related SSIs after gynecological surgery.

Disclosure statement of conflicts

The authors report no conflict of interest.

Disclosure of Funding

ESC is an MSTP student and was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by Medical Scientist Training Program grant T32GM008692.

Acknowledgements

We thank our individual funding sources.

References

1. Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *Infect Control Hosp Epidemiol*. 1992;13(10):606-608.

2. Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. *Infect Control Hosp Epidemiol*. 1999;20(11):725-730. doi:10.1086/501572

3. Prevention of Infection After Gynecologic Procedures: ACOG Practice Bulletin, Number 195. *Obstetrics & Gynecology*. 2018;131(6):e172. doi:10.1097/AOG.0000000000002670

4. Lake AG, McPencow AM, Dick-Biascoechea MA, Martin DK, Erekson EA. Surgical site infection after hysterectomy. *American Journal of Obstetrics and Gynecology*. 2013;209(5):490.e1-490.e9. doi:10.1016/j.ajog.2013.06.018

5. Till SR, Morgan DM, Bazzi AA, et al. Reducing surgical site infections after hysterectomy: metronidazole plus cefazolin compared with cephalosporin alone. *American Journal of Obstetrics and Gynecology*. 2017;217(2):187.e1-187.e11. doi:10.1016/j.ajog.2017.03.019

Appendix B – Peripartum uterine clostridial myonecrosis: A report of two fatal cases

This appendix has received a decision of revise and resubmit from the *Wisconsin Medical Journal.* All supplemental materials are available from the corresponding author.

Title: Peripartum uterine clostridial myonecrosis: A report of two fatal cases

Authors: Laura Jacques, MD[1]^, Bridget Kelly[1], MD, John Soehl, MD[2], Matthew Wagar[1], Janine Rhoades, MD[1], **Elise S. Cowley**, BS[3,4], Peter G. Pryde, MD[5], Abigail Cutler, MD[1], David Eschenbach, MD[6],

^ Correspondence should be addressed to Laura Jacques, Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-Madison, McConnell Hall, 4th Floor, 1010 Mound St., Madison, WI, 53715, laura.jacques@wisc.edu

Affiliations

[1] Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-Madison, 1010 Mound St., Madison, WI 53715 USA

[2] Department of Maternal Fetal Medicine, Women & Infants Hospital, 101 Dudley Street. Providence RI, 02905. USA.

[3] Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA

[4] Microbiology Doctoral Training Program, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA

[5] Madison Anesthesiology Consultants, 202 South Park Street, Madison, WI 53715

[6] Emeritus, Department of Obstetrics and Gynecology, University of Washington, 1959 NE Pacific St. Box 356460, Seattle, WA 98195.

Keywords: Peripartum Uterine Clostridial Myonecrosis; Clostridium Septicum; Clostridium Sordellii; Gas Gangrene; Intra-Amniotic Infection; Maternal Mortality;

Obstetric Infection; Hysterectomy; Point-of-Care Ultrasound; Necrotizing Soft Tissue Infection

Abstract

Introduction

Uterine clostridial myonecrosis is a rare infection associated with a high mortality rate. This report presents two cases of maternal mortality resulting from peripartum clostridial myonecrosis of the uterus.

Case Presentation

Case one is a 30-year-old nullipara who presented in labor at term with an intra-amniotic infection and fetal demise. She rapidly developed septic shock and cesarean hysterectomy was performed for a suspected necrotizing uterine infection later identified to be *Clostridium septicum*. Case two is an adolescent who presented in septic shock following first trimester medication abortion and died during emergent exploratory laparotomy; cultures grew *Clostridium sordellii.* Both patients expired within 18 hours of hospital admission.

Discussion

Given the rapidly progressive course of clostridial infections, maintaining a high index of suspicion is imperative for ensuring timely diagnosis and effective treatment. Prompt recognition of clinical features associated with clostridial myonecrosis– abdominal pain, tachycardia, leukocytosis and hyponatremia– is essential in preventing mortality. The utilization of point-of-care ultrasound may expedite the diagnosis of uterine myonecrosis. When uterine myonecrosis is suspected, immediate initiation of penicillin-based antibiotics, alongside clindamycin, and aggressive surgical intervention

including hysterectomy, are essential for ensuring survival. Although the decision to perform a hysterectomy can be challenging, especially in cases involving child-bearing aged patients, it is a vital step to avert a fatal outcome.

Conclusion

By presenting these cases, we aim to raise awareness of this uncommon, but highly lethal infection to expedite diagnosis and treatment to improve patient outcomes.

Introduction

Severe sepsis from intra-amniotic infection (IAI) is a rare but devastating complication of both childbirth and abortion.[1–4] Intra-amniotic infection is typically polymicrobial and isolation of a single causative bacterial species is uncommon.[5] The exception is IAI due to toxin-producing bacteria like *Clostridium* and Group A *Streptococcus* (GAS) which, even in isolation, can cause lethal necrotizing infections. Nearly all maternal infectious deaths today in developed countries are caused by these organisms.[6–8] Though uncommon, clostridial myonecrosis (gas gangrene) of the uterus can result from prolonged rupture of membranes, retained products of conception, and post-abortive sequelae.[9–11] We describe two cases of maternal death due to peripartum clostridial myonecrosis of the uterus.

Cases

Case 1-Clostridium septicum

A 30-year-old G2P0010 without any prenatal care at approximately 42-weeks gestation presented with ruptured membranes and an intrauterine fetal demise in breech presentation. Vital signs were remarkable for fever of 38.2 C, pulse of 127 beats per minute (bpm) and hypertension with a blood pressure of 167/81. Laboratory evaluation including a complete blood count (CBC), comprehensive metabolic panel (CMP), and urine protein to creatinine ratio (P:C) showed leukocytosis of 19.6 10^3/ µL and hyponatremia of 126 mmol/L and urine P:C was 0.57 while creatinine, AST, ALT, lactate and platelet counts were normal. She had no known history of hypertension and was treated for pre-eclampsia with severe features with a 4-gram bolus and then 2-gram per hour infusion of magnesium sulfate and one dose of 20 mg of intravenous Labetalol. Her blood pressures responded initially to 134/77 and she never required another dose of Labetalol. She received an epidural for pain control but initially declined labor augmentation, antibiotic treatment for presumed IAI or imaging. Two hours after admission she accepted and received ampicillin and gentamicin; her fever defervesced, and her tachycardia resolved. Vitals signs were recorded every 15 minutes and she remained normotensive with adequate urine output of 80 mL/hour over the next four hours. Labor progressed without augmentation and contractions were detected every three minutes on tocometry. Seven hours after admission, she reported increasing abdominal pain despite the epidural. Her condition acutely deteriorated, with vital signs 15 minutes before decompensation showing a blood pressure of 101/65, pulse of 88 bpm, oxygen saturation 96%. Then, 15 minutes later, her blood pressure and oxygen saturation dropped to 68/37 and 91% while her pulse remained 85 bpm. On physical examination, she was found to be newly lethargic and her fundal height, measured at 37

cm on admission, was palpated at the level of her diaphragm. Due to abdominal pain and the rapid increase in her fundal height, a bedside transabdominal ultrasound was attempted to assess the uterus and abdomen for signs of abruption and blood accumulation, but the presence of diffuse high-amplitude echoes and myometrial shadowing made visualization challenging. Chest X-ray and bedside echocardiogram were without acute findings. Laboratory evaluation was now significant for a white blood cell (WBC) count of 24.6 10^3/ µL, hemoglobin 11.9 g/dL, creatinine of 1.34 mg/dL, sodium 128 mmol/L and magnesium of 7.4 mg/dL. Her antibiotics were broadened to piperacillin/tazobactam and vancomycin and she initially responded to IV blood pressure support. She was now completely dilated and plus two station, and began pushing for an attempted breech vaginal delivery. Her condition worsened during 15 minutes of pushing with minimal descent of the fetal breech, so a vertical midline cesarean delivery under general anesthesia was performed. Upon entry into the abdomen, the uterus was markedly enlarged and edematous with diffuse mottling. A classical hysterotomy, that notably did not bleed, caused the eruption of large volume foul smelling gas. A demised fetus, which weighed 6800 grams due to remarkable edema and gaseous distension, was delivered. Hysterectomy was completed to obtain source control of the infection. Disseminated intravascular coagulation was diagnosed intraoperatively with an INR of 4.0 and an undetectable fibrinogen level. She was transferred to the intensive care unit postoperatively and two hours later experienced a cardiac arrest, was unable to be resuscitated and was pronounced dead. Uterine and placental cultures grew *Escherichia coli* and *Clostridium septicum*, two sets of peripheral blood cultures collected on admission had no growth. Pathology exam of the

uterus demonstrated extensive decidual and myometrial necrosis (Figure 1). Placental pathology revealed acute necrotizing chorioamnionitis with Gram variable rods in the intermembranous space (Figure 2). Her family declined autopsy of the patient and fetus.

Case 2-Clostridium sordellii

The second case is an adolescent primigravida who presented to the emergency department with worsening lower abdominal pain eight days after initiating medication abortion with mifepristone and misoprostol at 8-weeks of gestation. She had normal vital signs and laboratory evaluation, except a WBC count of 23.9 10^3/uL. Ultrasound showed vascular, echogenic material in the endometrial cavity presumed to be retained products of conception, so she underwent an uncomplicated manual vacuum aspiration in the emergency department and was discharged home.

She returned to the emergency department the next day with worsening abdominal pain and altered mental status. She was hypotensive (blood pressure of 30/15 mmHg) and tachycardic (heart rate 110 bpm) but afebrile and oxygenating normally. Her abdominal exam was notable for distention, guarding and rebound tenderness to palpation. A bedside ultrasound demonstrated large volume abdominal free fluid. The patient was taken to the operating room for exploratory laparotomy due to concern for intra-abdominal hemorrhage.

Exploration of the abdominal cavity revealed large volume ascites, edematous bowel and normal appearing appendix and adnexa. The uterus was small, firm and normal in appearance. Her laboratory analysis resulted intraoperatively and were notable for hemoglobin of 19.8 g/dL, WBC of 135.1 10^3/μL, sodium of 129 mmol/L,

creatinine of 2.25 mg/dL and a lactate of 8.5 mmol/L. Blood cultures and an endometrial biopsy for culture were sent when the elevated WBC count resulted and vancomycin, piperacillin-tazobactam and clindamycin were started. The patient's clinical status deteriorated further and a hysterectomy was performed for source control. During the hysterectomy, the patient experienced cardiac arrest that did not respond to resuscitation, and she expired on the operating table. The blood cultures never grew an organism but the endometrial culture revealed *Clostridium sordellii* and final uterine pathology demonstrated extensive necrosis of the endometrial lining extending to myometrium with acute inflammation and hemorrhagic necrosis of the bilateral adnexa. Autopsy findings were consistent with death due to toxic shock.

Discussion

Clostridia are anaerobic, spore-forming, gram-positive rods that are a commensal in the vaginal secretions of 5-20% of asymptomatic people and are a rare cause of obstetric and gynecologic infections.[2,9–16] In this report, we present two cases of uterine clostridial myonecrosis (gas gangrene), a rare and deadly manifestation of clostridial infection with mortality rates between 50-100% (9,10). In cases of clostridial myonecrosis, *C. perfringens* is most commonly isolated (found in 80% of cases), followed by *C. novyi* and *C. septicum*. Other rarer clostridial species causing myonecrosis include *C. histolyticum*, *C. sordellii* and *C. fallax.*[13,15,17] Most fatal cases of postpartum infection have involved one of these Clostridial species or other toxin producing bacteria, such as Group A Strep.[4]

Pathogenesis

Clostridial uterine myonecrosis occurs when clostridial organisms within the normal intestinal or vaginal flora spread into new spaces through a breach in anatomical defense barriers.[10] Severity of infection results from bacterial production of toxins.[13]

C. septicum

The bacterium encountered in the first case is *C. septicum,* an organism commonly identified in gas gangrene, but a rare cause of obstetric and gynecologic infection. We identified seven other cases in the literature – three with gynecologic malignancy and four peripartum patients.[12,18–23] *C. septicum* is an aerotolerant organism 300 times more virulent than *C. perfringens.*[15,21] The virulence of *C. septicum* stems from its four toxins, the most lethal of which is the alpha toxin, which produces lecithinase, a phospholipase that causes ischemic tissue necrosis and hemolysis. The creation of devitalized tissue by toxin-generated ischemia produces an anaerobic environment that results in rapid disease spread. Antibiotics cannot penetrate this devitalized tissue, making rapid source control via surgical debridement imperative.

C. sordellii

C. sordellii is an uncommon and transient isolate of vaginal and rectal flora, found in 3.4% of asymptomatic reproductive-aged women in the United States.[24] Uterine myonecrosis from *C. sordellii* constitutes a rapidly progressing and nearly always fatal infection. Though rare, it is a known complication of medication abortion, and has also been seen in peripartum infections and after gynecologic surgery.[25–29] Similar to *C. septicum,* virulence results from the production of seven known toxins, of which the hemorrhagic and lethal toxins are considered primarily responsible for pathogenic infection.

Diagnosis

While clinical manifestations of *C. sordellii* and *C. septicum* share some features, each presents differently (Table 1). Both infections are associated with tachycardia and pain out of proportion to examination findings. Patients with *C. septicum* typically have fever, sometimes hemorrhagic bullae or soft tissue crepitations and – once shock develops -- hypotension, multiple organ failure and disseminated intravascular coagulation.[21] Patients with *C. sordellii* infection are typically afebrile, but have refractory hypotension, tachycardia and a profound capillary leak syndrome causing large volume ascites.[25,27]

Laboratory findings are similar in both infections and early on are remarkable for hyponatremia and marked leukocytosis. Hyponatremia is caused by both sepsis physiology and the extensive third spacing of fluids seen in necrotizing soft tissue infections. Serum sodium levels less than 135 mEq/L and WBC count of greater 15.4 10^3/ μL are useful thresholds to distinguish necrotizing soft tissue infections from non-necrotizing infections.[24,25,30] Leukocytosis, while present in both clostridial infections, is markedly elevated (50-200 10^3/μL) in cases of *C. sordellii*.[25,27,31] Later laboratory findings for both infections include elevated lactate and creatine kinase levels as well as signs of multisystem organ failure such as elevated creatinine and liver function tests. *C. septicum* often causes coagulopathy–including thrombocytopenia, fibrinogenemia, anemia and elevated INR – where hemoconcentration is seen with *C. sordellii.* Blood and all tissue specimens suspected to be infected should be sent for Gram stain and culture. Gram stain will show gram-positive or gram-variable rods and often a paucity of neutrophils, thought due to toxin mediated destruction of these cells.[15] Culture is less

useful in the initial diagnostic phase as clostridial species can take several days to speciate.

Imaging, typically CT scan, may demonstrate myometrial gas, which can expedite confirmation of a diagnosis. In our first case, we saw diffuse echogenic foci with “dirty posterior shadowing” on bedside ultrasound of the uterus. In retrospect, this finding was likely due to the presence of extensive gas in the myometrium; similar ultrasonographic descriptions of intramuscular gas exist in the literature in cases of emphysematous cholecystitis.[30] Our review of the literature revealed no other reports of utilizing point-of-care ultrasound to evaluate for the presence of myometrial gas. We propose this as another diagnostic tool to aid in more rapid diagnosis of uterine myonecrosis.

Treatment

The principles of treatment for clostridial myonecrosis are the same regardless of causative species: broad-spectrum antibiotics; early, aggressive surgical debridement of infected tissues; and expeditious cardiovascular and blood product support. Clostridial species are nearly universally susceptible to penicillin-based antibiotics. Clindamycin is recommended due to its inhibition of bacterial protein synthesis and mitigating effects of circulating clostridial toxins, however an improved clinical benefit has not been documented.[32,33] While antibiotic therapy is necessary, the rapidly expanding volume of devitalized tissue in clostridial necrotizing infections means antibiotic penetrance is minimal. Early surgical exploration and aggressive resection of all necrotic tissue is required to gain source control and allow for increased antibiotic efficacy.

Conclusion

Though rare, uterine clostridial myonecrosis can have tragic consequences, as demonstrated by these two cases. Early recognition of the clinical features associated with clostridium infections -- pain out of proportion to exam, tachycardia, leukocytosis, hyponatremia and myometrial gas on imaging – may allow for prompt diagnosis and life-saving treatment. Most clostridium infections occur in patients of child-bearing age, rendering the decision to perform hysterectomy difficult. However, swift surgical action for effective infection source control is imperative for patient survival. Clinicians must utilize a high index of suspicion to promptly diagnose and treat these rare but devastating infections.

Figure 1. Gross uterine pathology. Gross uterine specimen for Case 1 demonstrating diffuse decidual and myometrial necrosis with complete cervical necrosis.

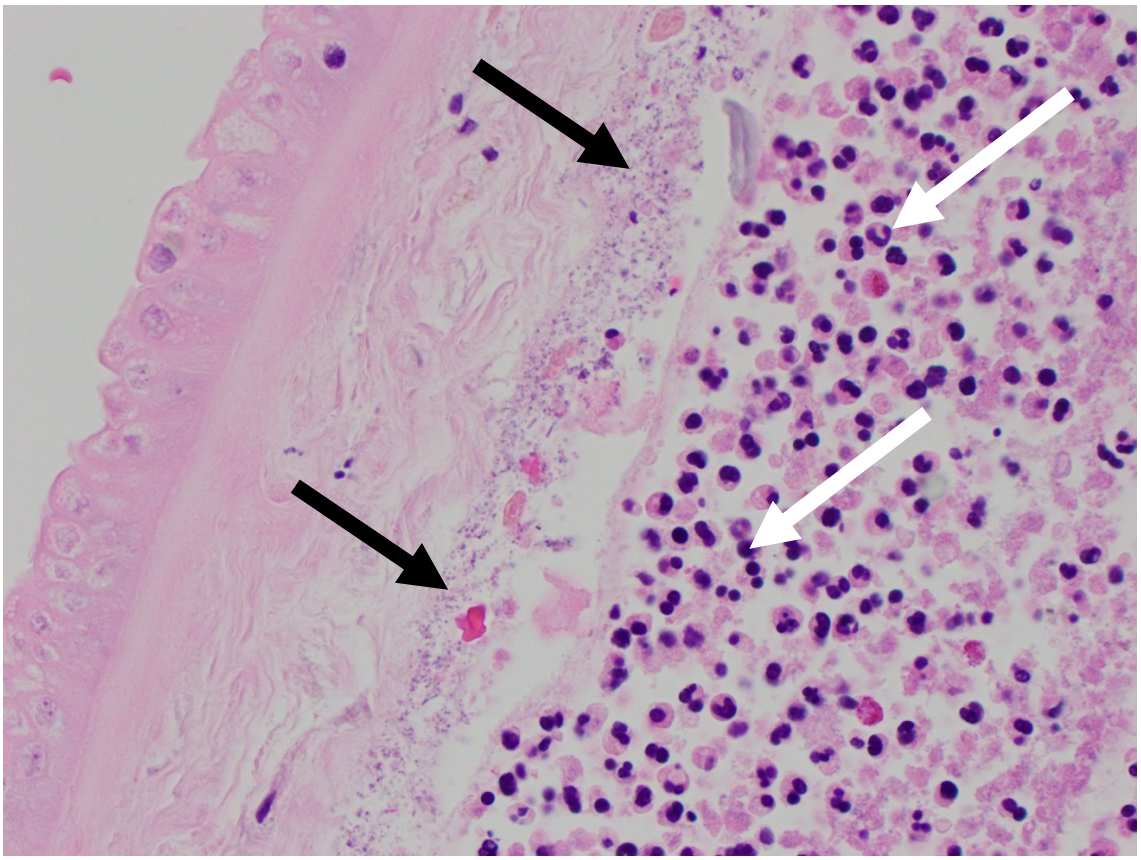


Figure 2. Placental pathology. Chorioamnionitis associated with Case 1 with gram variable rods present within intermembranous space (black arrows). Polymorphonuclear granulocytes present diffusely within chorion (white arrows).

CLINICAL CONSIDERATIONS					
ORGANISM	PATHOGENESIS	CLINICAL EXAM	LABORATORY	IMAGING	MANAGEMENT
C. septicum	4 unique toxins	**Fever** Hemorrhagic bullae Soft tissue crepitations Right heart	**↑WBC** 15-30 $10^3/\mu L$ Coagulopathy ↓ Hemoglobin ↓ Platelets ↓ Fibrinogen ↑ INR	Echocardiogram Right heart failure	**Broad spectrum antibiotics** • Clindamycin AND • Piperacillin-tazobactam
Both	**Toxin production** Ascending infection from vaginal or rectal flora	Pain out of proportion to exam Shock Refractory hypotension Tachycardia Oliguria	↑WBC ↑CK ↑Lactate ↑Creatinine ↑LFTs ↓Na Gram stain Gram+/Gram variable rods ↓ Neutrophils	CT scan/ultrasound **Myometrial gas**	**Hemodynamic stabilization** • Multidisciplinary care team • Adequate intravenous and central line access • Aggressive fluid and pressor support
C. sordellii	7 unique toxins	**No fever** Capillary leak syndrome Large volume ascites Edema	**↑↑↑ WBC** 50-200 $10^3/\mu L$ Hemoconcentration	Fast scan/CT scan Ascites	**Early and aggressive source control** • Consider hysterectomy • Consider unilateral or bilateral salping-oophorectomy

Table 1. Unique (white) and shared (gray) features of *Clostridium septicum* and *Clostridium sordellii.* This table highlights the key similarities and differences in the pathogenesis, clinical presentation and management for the two *Clostridial* species.

Acknowledgements and Disclosures

The authors thank Amanda Wildenberg and Cassidy Slinger for administrative support and Dr. Michael Fritsch, Department of Pathology, University of Wisconsin, Madison, for the photos provided as figures.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. ESC is an MSTP student and was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by MSTP grant T32GM140935.

None of the authors have a financial or other conflict of interest.

The authors have no disclaimers.

References

1. Gibbs, R.S., and Duff, P. (1991). Progress in pathogenesis and management of clinical intraamniotic infection. Am J Obstet Gynecol *164*, 1317–1326. 10.1016/0002-9378(91)90707-x.

2. Al-Ostad, G., Kezouh, A., Spence, A.R., and Abenhaim, H.A. (2015). Incidence and risk factors of sepsis mortality in labor, delivery and after birth: Population-based study in the USA. Journal of Obstetrics and Gynaecology Research *41*, 1201–1206. 10.1111/jog.12710.

3. Edwards, S.E., Grobman, W.A., Lappen, J.R., Winter, C., Fox, R., Lenguerrand, E., and Draycott, T. (2015). Modified obstetric early warning scoring systems (MOEWS): validating the diagnostic performance for severe sepsis in women with chorioamnionitis. American Journal of Obstetrics and Gynecology *212*, 536.e1-536.e8. 10.1016/j.ajog.2014.11.007.

4. Dempsey, A. (2012). Serious Infection Associated With Induced Abortion in the United States. Clinical Obstetrics and Gynecology *55*, 888–892. 10.1097/GRF.0b013e31826fd8f8.

5. Romero, R., Miranda, J., Kusanovic, J.P., Chaiworapongsa, T., Chaemsaithong, P., Martinez, A., Gotsch, F., Dong, Z., Ahmed, A.I., Shaman, M., et al. (2015). Clinical chorioamnionitis at term I: microbiology of the amniotic cavity using cultivation and molecular techniques. J Perinat Med *43*, 19–36. 10.1515/jpm-2014-0249.

6. Tanaka, H., Katsuragi, S., Hasegawa, J., Tanaka, K., Osato, K., Nakata, M., Murakoshi, T., Sekizawa, A., Kanayama, N., Ishiwata, I., et al. (2019). The most common causative bacteria in maternal sepsis-related deaths in Japan were group A Streptococcus: A nationwide survey. Journal of Infection and Chemotherapy *25*, 41–44. 10.1016/j.jiac.2018.10.004.

7. van Dillen, J., Zwart, J., Schutte, J., and van Roosmalen, J. (2010). Maternal sepsis: epidemiology, etiology and outcome. Current Opinion in Infectious Diseases *23*, 249–254. 10.1097/QCO.0b013e328339257c.

8. Carapetis, J.R., Steer, A.C., Mulholland, E.K., and Weber, M. (2005). The global burden of group A streptococcal diseases. The Lancet Infectious Diseases *5*, 685–694. 10.1016/S1473-3099(05)70267-X.

9. Eschenbach, D.A. (1983). Vaginal infection. Clin Obstet Gynecol *26*, 186–202. 10.1097/00003081-198303000-00023.

10. Decker, W.H., and Hall, W. (1966). Treatment of abortions infected with clostridium welchii. American Journal of Obstetrics & Gynecology *95*, 394–399. 10.1016/0002-9378(66)90124-4.

11. Kurashina, R., Shimada, H., Matsushima, T., Doi, D., Asakura, H., and Takeshita, T. (2010). Spontaneous Uterine Perforation Due to Clostridial Gas Gangrene Associated with Endometrial Carcinoma. Journal of Nippon Medical School *77*, 166–169. 10.1272/jnms.77.166.

12. Moore, D.H., Ghosh, K., and Sutton, G.P. (1994). Spontaneous Occurrence of Gangrene Due to Clostridium septicum in a Patient With AdvancedEndometrial Carcinoma. Infectious Diseases in Obstetrics and Gynecology *2*, 34–37. 10.1155/S1064744994000372.

13. Halpin, T.F., and Molinari, J.A. (2002). Diagnosis and Management of Clostridium Perfringens Sepsis and Uterine Gas Gangrene: Obstetrical and Gynecological Survey *57*, 53–57. 10.1097/00006254-200201000-00024.

14. Kocarev, M., Girn, Z., Collyer, T.C., and Swindells, S. (2012). Uterine Clostridial Myonecrosis After Thermal Balloon Endometrial Ablation. Obstetrics & Gynecology *120*, 483–485. 10.1097/AOG.0b013e31825a725d.

15. Cohen-Poradosu, R., and Kasper, D.L. (2020). Anaerobic Infections: General Concepts. In Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Elsevier), pp. 2923–2932.

16. Onderdonk, A.B., and Garrett, W. (2020). Diseases Caused by Clostridium. In Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Elsevier), pp. 2960–2968.

17. Present, D.A., Meislin, R., and Shaffer, B. (1990). Gas gangrene. A review. Orthop Rev *19*, 333–341.

18. Haas, L.E.M., Tjan, D.H.T., and van Zanten, A.R.H. (2006). Fatal Clostridium septicum infection in a young pregnant woman. Neth J Med *64*, 254–255.

19. Khoo, C.L., Meskhi, A., and Harris, C.P. (2013). Fatal *Clostridium septicum* following medical termination of pregnancy. Journal of Obstetrics and Gynaecology *33*, 530–530. 10.3109/01443615.2013.788138.

20. McDonald, R.E., and Moola, S. (2012). Clostridium septicum infection in a young pregnant patient. Case Reports *2012*, bcr2012006254–bcr2012006254. 10.1136/bcr-2012-006254.

21. Prinssen, H.M., Hoekman, K., and Burger, C.W. (1999). Clostridium septicumMyonecrosis and Ovarian Cancer: A Case Report and Review of Literature. Gynecologic Oncology *72*, 116–119. 10.1006/gyno.1998.5209.

22. Zurmeyer, S., Fotopoulou, C., Braicu, E., Schlichting, U., and Sehouli, J. (2013). Clostridium septicum can cause distant myonecrosis in patients with ovarian cancer. Anticancer Res *33*, 1585–1589.

23. Rimawi, B.H., Graybill, W., Pierce, J.Y., Kohler, M., Eriksson, E.A., Shary, M.T., Crookes, B., and Soper, D.E. (2014). Necrotizing Fasciitis and Toxic Shock Syndrome from *Clostridium septicum* following a Term Cesarean Delivery. Case Reports in Obstetrics and Gynecology *2014*, 1–5. 10.1155/2014/724302.

24. Chong, E., Winikoff, B., Charles, D., Agnew, K., Prentice, J.L., Limbago, B.M., Platais, I., Louie, K., Jones, H.E., and Shannon, C. (2016). Vaginal and Rectal Clostridium sordellii and Clostridium perfringens Presence Among Women in the United States: Obstetrics & Gynecology *127*, 360–368. 10.1097/AOG.0000000000001239.

25. Aldape, M.J., Bayer, C.R., Bryant, A.E., and Stevens, D.L. (2016). A novel murine model of Clostridium sordellii myonecrosis: Insights into the pathogenesis of disease. Anaerobe *38*, 103–110. 10.1016/j.anaerobe.2016.01.004.

26. Centers for Disease Control and Prevention (CDC) (2005). Clostridium sordellii toxic shock syndrome after medical abortion with mifepristone and intravaginal misoprostol--United States and Canada, 2001-2005. MMWR Morb Mortal Wkly Rep *54*, 724.

27. Fischer, M., Bhatnagar, J., Guarner, J., Reagan, S., Hacker, J.K., Van Meter, S.H., Poukens, V., Whiteman, D.B., Iton, A., Cheung, M., et al. (2005). Fatal toxic shock syndrome associated with Clostridium sordellii after medical abortion. N Engl J Med *353*, 2352–2360. 10.1056/NEJMoa051620.

28. Soper, D.E. (2007). Abortion and Clostridial Toxic Shock Syndrome. Obstetrics & Gynecology *110*, 2. 10.1097/01.AOG.0000288224.47104.7b.

29. Fjerstad, M., Trussell, J., Sivin, I., Lichtenberg, E.S., and Cullins, V. (2009). Rates of Serious Infection after Changes in Regimens for Medical Abortion. New England Journal of Medicine *361*, 145–151. 10.1056/NEJMoa0809146.

30. Ballard, D.H., Mazaheri, P., Raptis, C.A., Lubner, M.G., Menias, C.O., Pickhardt, P.J., and Mellnick, V.M. (2020). Fournier Gangrene in Men and Women: Appearance on CT, Ultrasound, and MRI and What the Surgeon Wants to Know. Can Assoc Radiol J *71*, 30–39. 10.1177/0846537119888396.

31. Aldape, M.J., Bryant, A.E., and Stevens, D.L. (2006). Clostridium sordellii Infection: Epidemiology, Clinical Findings, and Current Perspectives on Diagnosis and Treatment. Clinical Infectious Diseases *43*, 1436–1446. 10.1086/508866.

32. Stevens, D.L., Bryant, A.E., and Hackett, S.P. (1995). Antibiotic effects on bacterial viability, toxin production, and host response. Clin Infect Dis *20 Suppl 2*, S154-157. 10.1093/clinids/20.supplement_2.s154.

33. Hamada, S., Nakajima, M., Kaszynski, R.H., Otaka, S., Goto, H., Matsui, H., Fushimi, K., Yamaguchi, Y., and Yasunaga, H. (2022). Association between adjunct clindamycin and in-hospital mortality in patients with necrotizing soft tissue infection due to group A Streptococcus: a nationwide cohort study. Eur J Clin Microbiol Infect Dis *41*, 263–270. 10.1007/s10096-021-04376-2.

Appendix C – Qualitative pilot study: Longitudinal perspectives from people who had second-trimester abortions for fetal anomaly

This appendix has received a decision of revise and resubmit from the *Wisconsin Medical Journal.* All supplemental materials are available from the corresponding author.

TITLE Qualitative pilot study: Longitudinal perspectives from people who had second-trimester abortions for fetal anomaly

AUTHORS

Elise S. Cowley, BS[1,2 ^], Taryn M. Valley, MA[3,4 ^], Kelsey Christianson, BA[5], Suzanne Walczak, MD[6], Jessica A. Greenblatt, MD[6], Laura Jacques, MD[3*]

^ these authors contributed equally to this work

* Corresponding Author:

Laura Jacques, MD

University of Wisconsin-Madison, School of Medicine and Public Health

Department of Obstetrics and Gynecology

McConnell Hall, 4th Floor

1010 Mound St

Madison, WI

53715

608-890-2053

laura.jacques@wisc.edu

AFFILIATIONS

[1]Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA

[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA

[3]Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-Madison, 1010 Mound St., Madison, WI 53715 USA
[4]Department of Anthropology, University of Wisconsin-Madison, 1180 Observatory St., Madison, WI, 53706, USA
[5]School of Medicine and Public Health, University of Wisconsin-Madison, 750 Highland Ave, Madison, WI, 53726 USA
[6]Department of Obstetrics and Gynecology, Medical College of Wisconsin, 8701 Watertown
Plank Road, Milwaukee, WI, 53226, USA

Abstract

Background: We investigated patient experience with abortion for fetal anomaly, about which little is known.

Methods: This qualitative, longitudinal pilot study surveyed seven patients after abortion for fetal anomaly at one Wisconsin hospital, from 2012-2014.

Results: Patients: 1) felt having a choice to have an abortion and choose the modality is imperative, and they remained certain in their decision-making over time; 2) described initially strong, then lacking, social support; 3) processed grief; and 4) identified resource constraints.

Discussion: Patients emphasized the importance of having the choice to choose abortion and the abortion modality, remaining confident in their decision-making over

time. This qualitative pilot study provides areas for future intervention to improve care for people undergoing abortion for fetal anomaly.

Introduction

Nearly 1% of pregnant people have an abortion for fetal anomaly.[1,2] Yet little literature examines patient experiences with abortion for fetal anomalies or how patients' needs change over time.[2–4] We studied experiences with second-trimester abortion for fetal anomalies in Wisconsin, where patients increasingly encounter barriers accessing abortion.[5] In this qualitative pilot study, we examined patients' longitudinal 1) decision-making certainty, 2) emotional experiences, 3) perceived support, and 4) resource needs, after abortion for fetal anomaly.

Materials and Methods

From July 2012 until February 2014, we recruited patients at Froedtert Hospital, an academic and community-partnered hospital in Milwaukee, Wisconsin. At the time, barriers to abortion access in Wisconsin included a mandated 24-hour waiting period and consent certificate. Froedtert Hospital only provided abortion services given maternal health risk or life-threatening fetal anomalies. Patients qualified for the study if they were over 18 years old, English-speaking, and 14-24 weeks gestation with a pregnancy complicated by a lethal fetal anomaly.

We emailed consented participants two REDCap surveys, the first survey four to five days after their abortion and the second three months later (Appendices A and B). Surveys contained multiple-choice, short answer, and free text questions. We analyzed qualitative data using an inductive-deductive approach, establishing codes then themes.[6] Four researchers synchronously coded data in NVivo12 (QSR International, Australia), reaching consensus on themes.

Results

Of 36 eligible patients, 12 (33%) underwent dilation and evacuation (D&E) and 24 (67%) labor-induction. Ten (28%) enrolled; three dropped out without completing either survey, and seven patients completed both surveys (paired survey response rate = 70%; follow-up survey response rate = 100%). Four surveyed chose D&E and three labor-induction.

We identified four themes in both immediate and follow-up surveys: patients 1) felt certain in their *decision-making;* 2) emphasized need for *emotional support*; 3) shared

processing of grief; and 4) noted *resource constraints*. Below, we discuss these themes using representative patient quotations.

Certainty in Decision-making

Certainty about having an abortion

All patients considered the capacity to choose abortion in the setting of fetal anomaly essential, even patients who disapproved of abortion in other circumstances. Patients imagined feeling upset or helpless if they had needed to continue the pregnancy. One patient described, *"I would have felt trapped in my own body [without the abortion]. It was hard enough for me to have to carry my baby as long as I did"* (Table 1:1). One patient summarized how choice was essential: *"No one should be able to tell you that you don't have an option*" (Table 1:2); another patient said it would be difficult to *"carry my child full term knowing it would not survive"* (Table 1:3). Most patients shared that, though difficult, they felt certain their decision to terminate the pregnancy was right for them and their families (Table 1:4). On follow-up, participants continued to describe certainty in their decision to have an abortion (Table 2:1-3). One patient stated, *"I would not have changed my decision. I feel like it was the best thing to do for our family"*(Table 2:4).

Certainty about chosen abortion method

Participants universally felt that having a choice between abortion modalities, D&E or labor-induction, was critical for them and their families (Table 2:5-8). One patient reported:

> *"[T]he entire event would have been much worse if D&E was not available. If I would have had to have gone through L&D, that would have been much, much worse. It would have taken me much longer to heal, both emotionally and physically."* (Table 2:9)

Nearly all patients reported feeling that they had chosen the correct abortion method on three-month follow-up and wanted to reassure others faced with similar choices to trust they would choose correctly.

> *"Keep you [sic] head up ... trust yourself that you are doing the right thing. Regardless of what anyone says, you made the right decision if you are putting the needs of your family and that little baby first."* (Table 2:10)

Only one patient wished they had chosen a different method, labor-induction instead of D&E:

> *"I think I wish I would have labored to have him so I could have held him. I really wish I could have cuddled him. At the time I made the decision because I didn't want to be with other women delivering healthy babies."* (Table 2:11)

Emotional support

Healthcare staff

All patients reported strong support from non-physician healthcare staff, but varying experiences with physician provided support (Table 1:5-6). One patient described:

> *"The staff was great. From the nurse who checked us in at the day surgery center to the anesthesia team. Everyone was very compassionate and made sure my husband and I both had everything we needed."* (Table 1:7)

However, one patient felt that while the diagnosis of fetal anomaly was given professionally, they wished for more emotional connection with their physician (Table 1:8).

Family and Friends

Immediately after their abortion, patients expressed feeling well supported emotionally by family, friends, and colleagues (Table 1:9-11). Over time, some people felt that others wished they would move on faster than they were able and couldn't fully understand their experience (Table 2:12). One patient explained:

> *"I think just having people, especially our parents, understand that what we went through was hard and that we are not completely over it and when we are ready we will try again. It is not something to just jump back into....there are a lot of emotions still involved."* (Table 2:13)

On follow-up, people reflected on the limits of support (Table 2:14). One patient explained:

> *"I don't think anyone can really understand what you are going through unless they went through the exact same thing. There are days when my husband does not even totally understand and he will admit that. I think people try to understand and try to empathize, but they do not really understand."* (Table 2:15)

Seeking shared experiences

Relatedly, patients reported seeking people with shared experiences, through structured programs or family connections. When people shared experiences, participants on follow-up reported feeling more supported (Table 2:16-17).

Processing of grief

Participants uniformly demonstrated attachment to their pregnancy both immediately and upon follow-up. Directly after the procedure, participants reported significant grief: *"I was so sad. I hadn't realized how much wanted him until the diagnosis"* (Table 1:12). All patients reported calling the "entity that was lost" by name or "baby." Alongside sadness, sometimes participants expressed anger or numbness (Table 1:13-14).

As time passed, people experienced grief differently than directly after their abortions (Table 2:18-20). *"I still hurt and cry when I think about it, as my baby would be due to be born this month. I have good days but sometimes have days I cry through when I think of him"* (Table 2:21). Over time people felt increasingly "at peace" with their abortion decisions, process, and outcome (Table 2:22). One participant echoed many shared sentiments on follow-up:

> *"It was hard at first when my pregnancy suddenly ended and there was no baby to show for it. However, everyone has been really supportive and I am very happy and at peace now with the whole situation."* (Table 2:23)

Resource constraints

In our final theme, some respondents struggled with logistics surrounding their abortion. People were concerned about finances, including procedure expenses, potential lack of insurance coverage, time off work for recovery, and burial expenses (Table 1:15-16). People also mentioned travel distance for desired modality or accessing support services as obstacles (Table 1:17). One patient explained: *"I'd like to attend a support group at Froedtert, but it's about 45 min. away. Wish it was closer,"* (Table 2:24).

Table 1. Themes regarding experiences with a pregnancy complicated by fetal anomaly and representative quotations from **initial surveys** related to themes and sub-themes.

Sub-theme	Quote #	**Themes**	
		Representative quotations	Patient, Modality, Immediate or Follow-up survey
		Certainty in Decision-Making	
Importance of the ability to choose abortion and certainty around this choice	1	I would have felt trapped in my own body [without the abortion]. It was hard enough for me to have to carry my baby as long as I did after finding out. I think that emotionally I would have been a basket case. I don't quite know how I would have been able to handle it. My husband would have been the same. I don't think that he would have been able to look at me. I had a hard time looking at myself.	1236, D&E, I
	2	I would be upset [if I had not had the option to have an abortion]. It is an option for people to make when getting the upsetting news that your child will not survive. No one should be able to tell you that you don't have an option.	1231, D&E, I
	3	It would have been even harder to have to carry my child full term knowing it would not survive. Not	1233, IOL, I

		having the option for labor induction would have been very upsetting.	
	4	It is a decision we made together and the best decision for us as a couple.	1237, D&E, I
		Emotional Support	
Physician and healthcare staff support	5	[Doctors] gave us the space that we needed, but were still available for questions. They did not rush us to any decisions.	1237, D&E, I
	6	The doctors presented the information very well and were all very empathetic. We could not have asked for better people to give us the news.	1231, D&E, I
	7	The staff was great. From the nurse who checked us in at the day surgery center to the anesthesia team. Everyone was very compassionate and made sure my husband and I both had everything we needed.	1237, D&E, I
	8	The diagnosis was presented in a medical and somewhat 'sterile' way, not very emotional.	1234, IOL, I
Family and friend support	9	It was a very difficult decision but my family as well as the staff we interacted with made the situation go smoothly. Everyone really helped my husband and I to let our feelings be hear [sic] and really honor the memory of our daughter.	1233, IOL, I
	10	With the support of my husband as well as our families and friends it was as easy as it could be. They were there to bring food and do things around the house. Getting back to work was just exhausting, but my staff helped a great deal with easing me back in.	1231, D&E, I
	11	We received great support from family and friends. They provided their own experiences as well as people to talk to when we needed to vent or cry.	1233, IOL, I
		Processing of Grief	
	12	I was so sad. I hadn't realized how much wanted him until the diagnosis.	1238, D&E, I
	13	[I felt n]umbness at times. Anger and disgust at the baby.	1234, IOL, I

	14	Of course I felt sadness, but I was surprised at how angry I was. Why was this happening to us?	1231, D&E, I
		Resource Constraints	
	15	All I could think about was the time to recover and the expense with labor. My insurance covered almost all of the surgery.	1237, D&E, I
	16	If finance was not an issue, I would have given my precious a service and burial.	1239, IOL, I
	17	We chose labor induction. It seemed to be the safer option at that point in the pregnancy. I was also told there might not be a place closer than Chicago that offers D&E at the stage of my pregnancy, and I liked the idea of being closer to my home.	1234, IOL, I

*(D&E) indicates a quote from a patient who terminated their pregnancy by dilation and evacuation
*(IOL) indicates a quote from a patient who terminated their pregnancy via induction of labor.
*(I) Immediate follow-up survey response
*(F) Follow-up survey response

Table 2. Themes regarding experiences with a pregnancy complicated by fetal anomaly and representative quotations from **follow up surveys** related to themes and sub-themes.

Sub-theme	Quote #	**Themes**	
		Representative quotations	Patient, Modality, Immediate or Follow-up survey
		Certainty in Decision-Making	
Importance of the ability to choose abortion and	1	It was hard at first when my pregnancy suddenly ended and there was no baby to show for it. However, everyone has been really supportive and I am very	1233, IOL, F

certainty around this choice		happy and at peace now with the whole situation.	
	2	I still feel that I made the correct decision [to have an abortion].	1239, IOL, F
	3	I would still be pregnant if he had not been born on his own. I do enjoy being pregnant but it would be so difficult being pregnant knowing the baby would die when he was born. I felt I needed to end the pregnancy to start my grieving process sooner and not prolong	1238, D&E, F
	4	I would not have changed my decision. I feel like it was the best thing to do for our family.	1231, D&E, F
Importance of choice in abortion **modality** and certainty about chosen abortion method, D&E or IOL	5	A surgical procedure I think would have been a little more stressful.	IOL, 1233, F
	6	I would not like it [D&E] because I want to see my baby regardless of the deformity.	1239, IOL, F
	7	Labor induction was the right decision for us. D&E makes me uncomfortable.	1234, IOL, F
	8	I feel good about my decision to have a D&E. It was the best decision for me and my husband.	1236, D&E, F
	9	I think the entire event would have been much worse if D&E was not available. If I would have had to have gone through L&D, that would have been much, much worse. It would have taken me much longer to heal, both emotionally and physically.	1237, D&E, F
	10	Keep you [sic] head up and trust yourself. If your choices are the best for that little baby, then trust yourself that you are doing the right thing. Regardless of what anyone says, you made the right decision if you are putting the needs of your family and that little baby first.	1237, D&E, F
	11	I think I wish I would have labored to have him so I could have held him. I really wish I could have cuddled him. At the time I made the decision because I didn't want to be with other women delivering healthy babies.	1237, D&E, F
		Emotional Support	

Family and friend support	12	The number of people that ignore or forget what you've gone thru. I think it's easier for them to not mention it, since they don't want to bring it up/make me feel sad, etc. By them not saying anything, it makes me think they don't care.	1234, IOL, F
	13	I think just having people, especially our parents, understand that what we went through was hard and that we are not completely over it and when we are ready we will try again. It is not something to just jump back into....there are a lot of emotions still involved.	1231, D&E, F
	14	I don't think anyone can truly understand besides me. My husband and parents are dealing with the loss of their first grandchild, and my husband with his first child. I was the one carrying her, and no one else can completely understand that. But, I also don't expect them too [sic].	1237, D&E, F
	15	I don't think anyone can really understand what you are going through unless they went through the exact same thing. There are days when my husband does not even totally understand and he will admit that. I think people try to understand and try to empathize, but they do not really understand.	1231, D&E, F
Seeking shared experiences	16	Everyone in the support group has gone through a similar experience so they understand.	1238, D&E, F
	17	My mother lost a child during pregnancy so she was able to share her feelings about that with me.	1233, IOL, F
		Processing of Grief	
	18	I cried a lot at first and didn't have any interest in life. As time has gone by I had to go back to work and take care of my children. Life has gone on and I have good days and bad days.	1238, D&E, F
	19	Mentally, at times I find myself thinking about the baby and the what ifs. Physically, I am back to my normal routines.	1239, IOL, F

	20	I have my good days and my bad days. I just take what I can handle and if I need to step away from an [sic] situation I do and come back when my emotions calm down. now I have some bad days and then [sic] come unexpectedly. I cry more that I'm not pregnant again then I do about what I had to go threw [sic].	1234, IOL, F
	21	I still hurt and cry when I think about it, as my baby would be due to be born this month. I have good days but sometimes have days I cry through when I think of him.	1238, D&E, F
	22	I truly believe that I made the right decisions for my family and my little girl. I am at peace with my decisions.	1237, D&E, F
	23	It was hard at first when my pregnancy suddenly ended and there was no baby to show for it. However, everyone has been really supportive and I am very happy and at peace now with the whole situation.	1233, IOL, F
		Resource Constraints	
	24	I'd like to attend a support group at Froedtert, but it's about 45 min. away. Wish it was closer.	1234, IOL, F

*(D&E) indicates a quote from a patient who terminated their pregnancy by dilation and evacuation
*(IOL) indicates a quote from a patient who terminated their pregnancy via induction of labor.
*(I) Immediate follow-up survey response *(F) Follow-up survey response
*(F) Follow-up survey response

Discussion

Our longitudinal findings (**Figure 1**) reinforce previous work documenting how peoples' needs evolve over time after abortion for fetal anomaly.[2,7,8] Patients require immediate and ongoing emotional, logistical, and financial support.[9,10] Our findings suggest patients could benefit from accessible support groups connecting individuals with shared experiences. In our study, few respondents reported regret and most remained

certain in their decision-making around whether and how to terminate their pregnancies, a novel finding. These findings could be shared with patients during pre-procedure counseling, advising they are likely to later feel they've decided correctly.

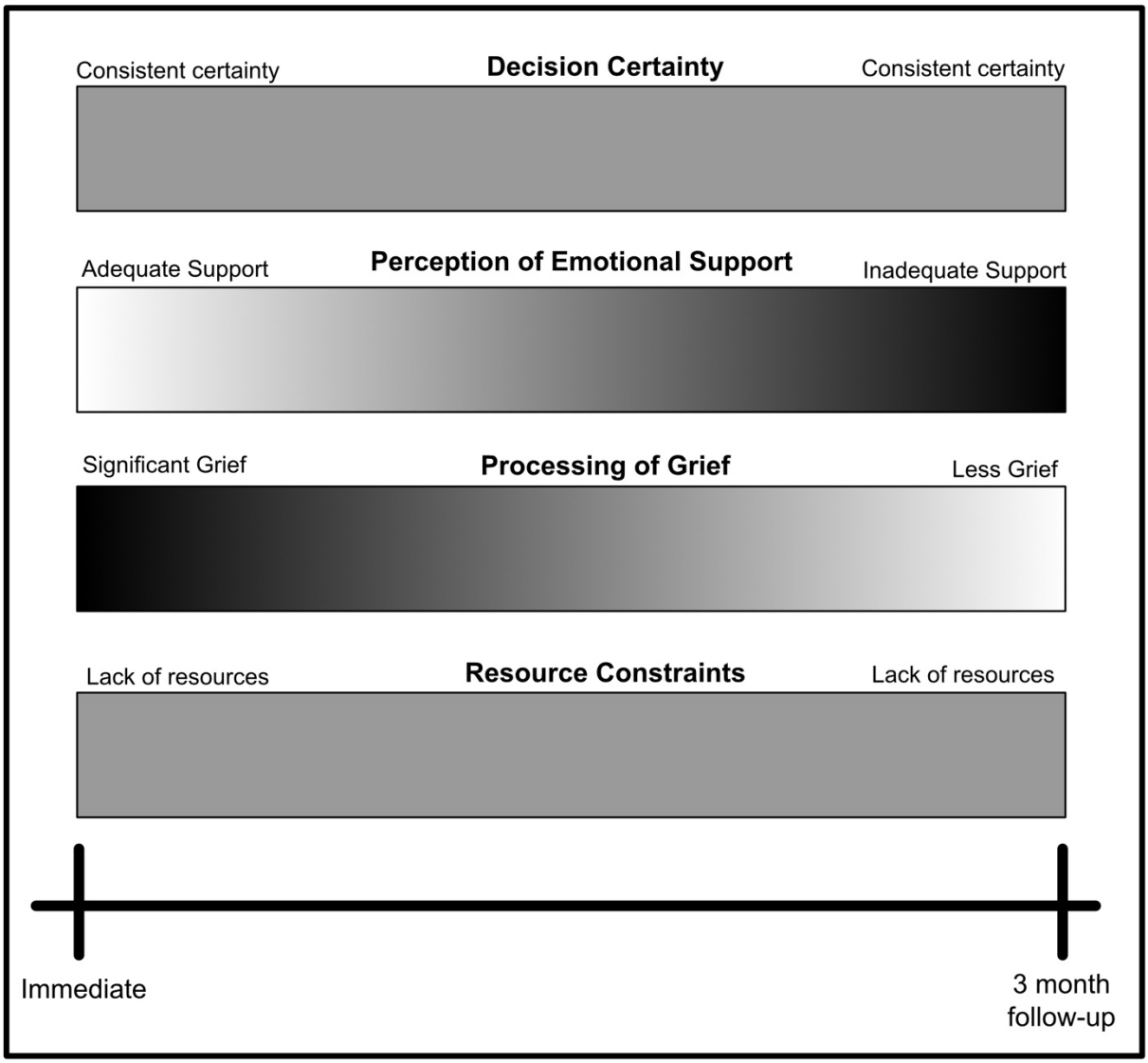


Figure 1. Shifting perspectives of people immediately and longitudinally post-abortion for second trimester for fetal anomaly. The majority of people reported feeling consistently certain in their decision to have an abortion and their choice of modality. Immediately after their abortion, patients reported feeling well supported, but later wanted more support. Immediately, people described significant grief. They reported continuing, but waning grief on follow-up. Though the types of resources

people needed changed over time, patients consistently detailed how lack of resources limited their decision making.

Since the Supreme Court decision in June 2022 overturned *Roe vs Wade* and federal protection for abortion, abortion in the case of fetal anomaly has been banned in Wisconsin. Patients in our study described the stress and harm they would have experienced without the choice to have an abortion or select the modality (D&E versus labor-induction). Findings from these few patients can be used to design further research on effects of Wisconsin's abortion ban on patients diagnosed with fetal anomaly.

Despite our small sample size, our response rate, thematic consistency, and response breadth strengthen analysis. Research involving stigmatized behavior and rare conditions draws meaningful conclusions from small samples. This study is unique, as abortion is increasingly inaccessible in post-*Roe* Wisconsin.[5] This qualitative pilot study provides groundwork for future research with people who have second-trimester abortion for fetal anomaly.

Ethics Approval and Consent to Participate

The project has been approved by the Medical College of Wisconsin Institutional Review Board (ID# IRB PRO00016460). Written consent was obtained from the original Primary Investigators, Drs. Jessica Greenblatt and Suzanne Walczak, stating that only de-identified data is being provided for analysis by current investigators.

Acknowledgements

We thank the patients for making this study possible and our individual funding sources. We thank Andrea Zorbas and Amanda Wildenberg for administrative support.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. ESC is a Medical Scientist Training Program (MSTP) student and was supported by a National Library of Medicine training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by MSTP grant T32GM140935. TMV is an MSTP student who, during the work on this paper, was funded in part by the CORE Lab, funded by a large, anonymous family foundation, and in part by MSTP grant T32GM140935. KC was supported by the Herman and Gwendolyn Shapiro Foundation through a summer research award and UW-Madison Department of Obstetrics and Gynecology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of Data and Materials

The datasets generated and analyzed during the current study are not publicly available due to the specific nature of the content discussed and possible identifiability of participants from their data, but are available from the corresponding author on reasonable request.

References

1. Byrne, J.J., Morgan, J.L., Twickler, D.M., McIntire, D.D., and Dashe, J.S. (2020). Utility of follow-up standard sonography for fetal anomaly detection. Am J Obstet Gynecol *222*, 615.e1-615.e9. 10.1016/j.ajog.2020.01.001.

2. Maguire, M., Light, A., Kuppermann, M., Dalton, V.K., Steinauer, J.E., and Kerns, J.L. (2015). Grief after second-trimester termination for fetal anomaly: a qualitative study. Contraception *91*, 234–239. 10.1016/j.contraception.2014.11.015.

3. Kerns, J., Vanjani, R., Freedman, L., Meckstroth, K., Drey, E.A., and Steinauer, J. (2012). Women's decision making regarding choice of second trimester termination method for pregnancy complications. International Journal of Gynecology & Obstetrics *116*, 244–248. 10.1016/j.ijgo.2011.10.016.

4. Kerns, J.L., Light, A., Dalton, V., McNamara, B., Steinauer, J., and Kuppermann, M. (2018). Decision satisfaction among women choosing a method of pregnancy termination in the setting of fetal anomalies and other pregnancy complications: A qualitative study. Patient Education and Counseling *101*, 1859–1864. 10.1016/j.pec.2018.06.012.

5. Post-Dobbs changes in abortion access in Wisconsin and nationally (2023). (UW Collaborative for Reproductive Equity).

6. Bernard, H.R. (2006). Qualitative Data Analysis 1: Text Analysis. In Research Methods in Anthropology: Qualitative and Quantitative Approaches (AltaMira Press), pp. 492–521.

7. Gammeltoft, T., Minh Hăng, T., Thỉ Hiêp, N., and Thỉ Thúy Hảnh, N. (2008). Late-Term Abortion for Fetal Anomaly: Vietnamese Women's Experiences. Reproductive Health Matters *16*, 46–56. 10.1016/S0968-8080(08)31373-1.

8. Korenromp, M.J., Page-Christiaens, G.C.M.L., van den Bout, J., Mulder, E.J.H., and Visser, G.H.A. (2009). Adjustment to termination of pregnancy for fetal anomaly: a longitudinal study in women at 4, 8, and 16 months. American Journal of Obstetrics and Gynecology *201*, 160.e1-160.e7. 10.1016/j.ajog.2009.04.007.

9. Ramdaney, A., Hashmi, S.S., Monga, M., Carter, R., and Czerwinski, J. (2015). Support Desired by Women Following Termination of Pregnancy for a Fetal Anomaly. J Genet Counsel *24*, 952–960. 10.1007/s10897-015-9832-x.

10. Asplin, N., Wessel, H., Marions, L., and Georgsson Öhman, S. (2014). Pregnancy termination due to fetal anomaly: Women's reactions, satisfaction and experiences of care. Midwifery *30*, 620–627. 10.1016/j.midw.2013.10.013.

Appendix D – Obstetrics and Gynecology Applicant Perceptions of Residency Program Culture with Virtual Interviews: A Qualitative Analysis of Social Media Posts

This appendix has received a decision of revise and resubmit from *BMC Medical Education.* All supplemental materials are available from the corresponding author.

Title: Obstetrics and Gynecology Applicant Perceptions of Residency Program Culture with Virtual Interviews: A Qualitative Analysis of Social Media Posts

Authors: Laura H Jacques[1], **Elise S. Cowley**[2,3], Shanaya M. Hebgen[4], Ryan J. Spencer[1], Corinne M. Hale[5]

L. Jacques is an associate professor and medical student clerkship director, Department of Obstetrics and Gynecology, University of Wisconsin, Madison; ORCID: 0000-0003-2343-8358
E.S. Cowley is an MD/PhD candidate, Department of Bacteriology, University of Wisconsin, Madison; ORCID: 0000-0002-0862-7321.
S. Hebgen is an MD candidate, University of Wisconsin School of Medicine and Public Health, Madison.
Ryan Spencer is an associate professor and the director of the Obstetrics and Gynecology Residency Program, Department of Obstetrics and Gynecology, University of Wisconsin, Madison.
C.M. Hale is a PhD/MPH candidate, Department of Anthropology, University of Wisconsin, Madison; ORCID: 0000-0002-9805-2102.
Affiliations:

[1]Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-Madison, 1010 Mound St., Madison, WI 53715 USA
[2]Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA
[3]Microbiology Doctoral Training Program, University of Wisconsin-Madison, 1550 Linden Dr., Madison, WI, 53706 USA
[4]Doctor of Medicine program, University of Wisconsin School of Medicine and Public Health, 750 Highland Ave, Madison, WI, 53726 USA
[5]Department of Anthropology, University of Wisconsin-Madison, 1180 Observatory Dr., Madison, WI, 53706, USA
^ Correspondence should be addressed to Laura Jacques, Department of Obstetrics and Gynecology, School of Medicine and Public Health, University of Wisconsin-

Madison, McConnell Hall, 4th Floor, 1010 Mound St., Madison, WI, 53715,
laura.jacques@wisc.edu

Abstract

Background

In the United States, Obstetrics and Gynecology residency interviews are instrumental in assessing the compatibility between medical student applicants and residency programs during the match process. Applicant perceptions of Obstetrics and Gynecology residency culture are a key component in determining how they rank residency programs. In 2020, residency interviews transitioned to a virtual format, and little is known about how applicants evaluated program culture during this first round of universal virtual interviews.

Medical students in the United States commonly use Reddit, a popular social media platform, to discuss residency programs and share interview experiences. We explored Obstetrics and Gynecology applicants' considerations regarding residency program culture during the first universal virtual interview season in 2020-2021 by analyzing posts on a Google spreadsheet accessed through Reddit.

Methods

In 2022, we imported 731 posts from the "2020-21 OB GYN Residency Applicant Spreadsheet" Google spreadsheet posted to the 2020-2021 Residency Interview Spreadsheet megathread on the r/medicalschool subreddit to NVivo 12(QSR

International, Burlington, MA), a qualitative analysis software program. Three investigators used qualitative inductive techniques to code and identify themes.

Results

Applicants used visual, verbal and behavioral cues during virtual Obstetrics and Gynecology residency interviews to understand three components of the workplace culture: prioritization of diversity, equity and inclusion, social environment, and resident workload.

Conclusions

Obstetrics and Gynecology residency programs convey information about their culture during virtual interviews through the behavior, appearances and responses of residents and interviewers to applicant questions. To ensure they accurately represent their culture to applicants, programs should consider educating residents and faculty around the implications of interview-day conduct.

Keywords: Residency Interviews, Obstetrics and Gynecology, Virtual Interviews, Program Culture, Applicant Perceptions, Residency Program Fit

Introduction

When medical students in the United States (US) apply for Obstetrics and Gynecology (ObGyn) residencies, they take into account various factors, including location and institutional reputation. However, consistently ranking as a top priority for both applicants and programs is the interview day experience and “goodness of fit”.[1-5] While there is no universally agreed-upon definition for “goodness of fit”, it is generally understood to refer to the alignment between a program’s culture and an applicant’s

priorities.[6] To ensure clarity in our discussions, we will consistently use the term 'fit' to signify the alignment or compatibility between ObGyn residency programs and applicants throughout this manuscript.

Residency interviews serve as a critical means to assess the fit between applicants and programs, influencing how students ultimately rank programs during the Match® process.[2,3] The onset of the COVID-19 pandemic in the US forced the 2020-2021 interview season to transition online to combat the spread of SARS-CoV2. Ultimately, even after the threat from COVID-19 diminished, the Association of American Medical Colleges (AAMC) and the National Resident Matching Program (NRMP) recommended the adoption of virtual interviewing for all residency programs to address socioeconomic disparities, reduce environmental impact, and enhance applicant satisfaction.[7,8] Though virtual interviewing conveys many advantages, both applicants and program directors have found it difficult to assess “fit”, especially when compared to traditional in-person interviewing.[1,5,9]

If programs want to accurately convey their culture to applicants, they must understand which aspects of culture applicants value and how they gather this information virtually. There is little data on how applicants evaluate program culture and fit during virtual interviews. To address this gap, we explored US ObGyn applicants’ considerations regarding residency program culture and applicant fit during virtual interviews by analyzing a Google spreadsheet posted to Reddit, a popular social media platform.

Materials and Methods

Data source

Reddit, a popular social media platform in the United States, counts nearly 40% of people 18-29 years of age as users, and functions as an interactive online bulletin board.[10,11] Users post content and engage through narrative comments within topic-specific communities called “subreddits” which contain discussion boards known as threads. Reddit is increasingly used as a qualitative data source due to its ability to capture real-time, often unfiltered impressions and discussions of events.[12-17]

Every year the subreddit /r/medicalschool creates a publicly available “megathread”, a consolidated thread used for major events or popular topics within a subreddit, where medical students discuss residency programs. Specialty-specific spreadsheets are posted within the megathread. On May 27, 2022, we accessed the “2020-21 OB GYN Residency Applicant Spreadsheet” Google spreadsheet posted to the 2020-2021 “Residency Interview Spreadsheet” megathread on the r/medicalschool subreddit. We chose this time frame because it represented the first interview season where there was widespread use of virtual interviewing. We analyzed every comment from the three tabs most pertinent to our research question - “Name & Shame”, “Student Reviews,” and “PM_Pearls”. The 'Name and Shame' tab features student reviews detailing their interview experiences with specific residency programs. In this context, 'Name' refers to students disclosing the program they interviewed with, while 'Shame' pertains to any issues, comments, or activities by the programs or their representatives that caused applicants distress or frustration. "Student reviews" is a tab where students evaluate their home institution and respond to questions about the experience at their program from other students. “PM_Pearls” is a section where

applicants and program managers can engage in discussions and address applicant questions and concerns.

A single researcher removed duplicate posts and identifying information like usernames and location. We imported 731 unique anonymous posts to NVivo 12(QSR International, Burlington, MA), a qualitative analysis software.

Analysis

We analyzed 731 poster comments using thematic qualitative analysis, as described by Kim and colleagues.[18] Three investigators, C.H., E.S.C. and S.H., used an inductive approach to generate codes based on the data and create a codebook. The final codebook contained 11 codes and 38 subcodes. The first 25 posts were collectively coded by C.H., E.S.C., and S.H. to reach a consensus, after which the investigators individually coded all remaining posts. All five authors independently reviewed the coding reports, met together as a group to discuss and generate the final three themes presented below. This study is considered program evaluation and was exempted from review by the UW-Madison Institutional Review Board.

Results

We identified three themes that relate to what applicants value in a residency program's culture: 1. Prioritization of diversity, equity, and inclusion (DEI), 2. Social environment and resident (mis)treatment, and 3. Resident workload. Our analysis also revealed three coherent subthemes for each main theme regarding *how* applicants were trying to determine these three aspects of residency culture during virtual interviews; they used a. *visual, b. verbal* and c. *behavioral* clues to understand program culture (Fig. 1).

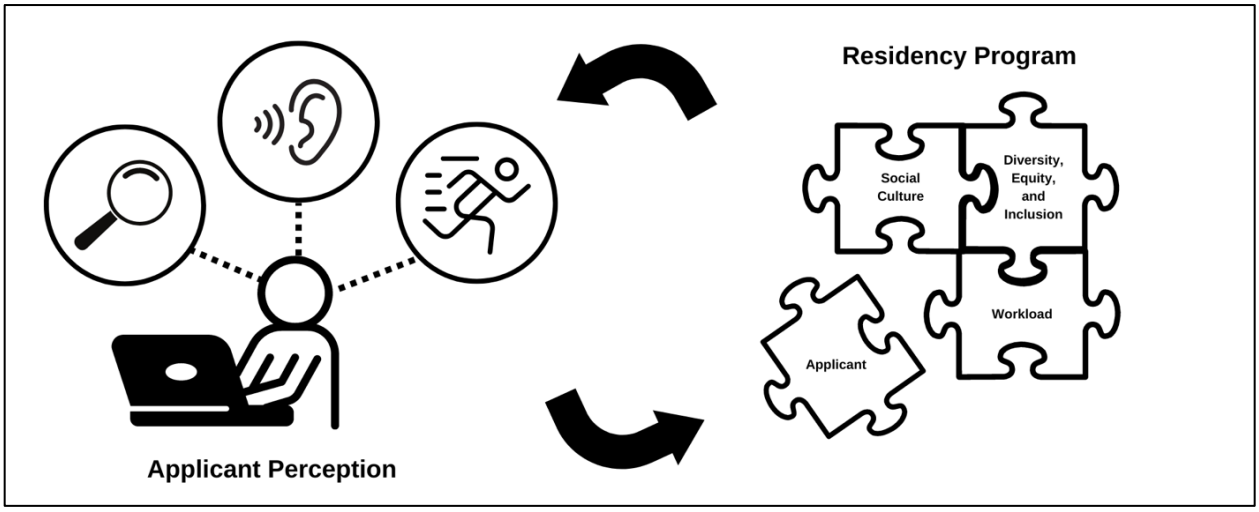


Figure 1. **How Residency Applicants Determine Program Fit** Obstetrics and Gynecology residency applicants use *visual, verbal*, and *behavioral* cues to determine residency program culture and their fit within that culture during virtual interviews.

1. Prioritization of Diversity, Equity and Inclusion

Applicants were keenly interested in how programs were addressing issues around diversity, equity and inclusion (DEI) during the 2020-2021 application cycle[19,20]. This time period in the US was marked by significant cultural events like the tragic death of George Floyd and COVID-19 driven health disparities that thrust matters of social justice and racial inequality into the spotlight.[20,21] Posters commented on the varying sources, visual, verbal and behavioral, they used to try to determine how residency programs were handling issues of race, equity and inclusion. This theme was so prevalent in our data, that we dedicated a separate manuscript to exploring this theme

and the impact of DEI on the residency match process[19]. We also deemed it essential to include a more concise exploration of the DEI theme in this paper, as it emerged as a key factor in applicants' assessments of program culture and played a significant role in shaping their perspectives during the residency selection process.

a. *Visual cues:* Applicants considered the visual diversity among interviewers, current residents

and applicants as an indicator of inclusivity. For instance, one post mentioned, *"Was disappointed by lack of diversity in applicant pool during my IV [interview] day".* Several other posters agreed with one replying, *"yup especially given they have ZERO black residents right now"*. One applicant expressed disappointment with the lack of perceived diversity at an event specifically designed for under-represented in medicine applicants, "*Had us attend a 6-hour diversity 2nd look with ZERO black or Latinx (from appearances and names) applicants present. Idk [I don't know] what they're doing but that was my sign."*

a. *Verbal cues*: Verbal cues about the inclusivity of program culture included how interviewers

asked and responded to diversity, equity and inclusion (DEI) focused questions, microaggressions, and culturally insensitive or racist comments made by members of the department. One student of color described being asked an inappropriate question by an associate program director, *"I was asked what I do to be anti-racist by the APD [Associate Program Director]. I'm Black so that was super off putting and it was a very poor question to ask a Black woman."* Others empathized with that student's experience and reaffirmed their assertion that it was a racially motivated question, *"YIKES. I was*

also not asked this (and I'm white!)". Applicants also posted about experiencing microaggressions during interviews such as "*Question on diversity efforts was met with how they look for "competent" applicants, WTF"* and "*Of course got the "Your English (sic) so good" thrown in by half my interviewers. Just a ton of microaggressions I have neither the energy or interest in repeating"*. Several posters described racist comments made by interviewers such as a program chair who referred to undocumented people as "*illegal immigrants*" or residents characterizing the community where they live as "*the hood*", which they found "*absolutely appalling"* and *"blatantly racist."*

b. *Behavioral cues:* Applicants also evaluated program culture based on the behavior of programs

around DEI-related issues, such as what faculty posted on their private social media accounts and whether institutions have faced discrimination-related litigation. Posts range from critical, *"Associate Dean of GME ([Institution name]) has been sued THREE TIMES for discrimination, yet still retains his position… definitely sensing a systemic problem here."* to noting positive actions like hiring more female faculty and increasingly diversifying their resident classes.

2. Social Environment and Resident (Mis)treatment

Applicants actively discussed their perceptions of the social dynamics within residency programs, emphasizing their interest in residents' interactions and how they were treated by faculty and the institution. Visual, verbal and behavioral cues played a crucial role in applicants' assessments of the social environment within the residency community and between faculty and residents.

a. *Visual cues:* Applicants noted that the presence or absence of residents from interview days

visually signaled to them whether the program valued its residents. One poster registered "*No residents present during IV [interview] day*" as a "Shame" and another poster agreed replying, *"I think it is a huge red flag to not have residents present at an interview day and understand why op [original post] would state this*."

b. *Verbal cues:* Verbal cues included the interviewing style, degree of interviewer interest in

applicants and types of questions asked by interviewers to determine the social environment. Interrogative or behavioral interviewing were viewed the most negatively as they seemed too intense or didn't facilitate programs and applicants to get to know each other. An exemplar post regarding an applicant's perspective on a more interrogatory interview: *The interviews were like oral exams where we were asked obgyn questions- if you have a 15 min interview you should use that time to get to know us personally, not test us on info that we clearly were tested on in the usmle exams*." Behavioral interviewing was also widely seen as a negative experience as exemplified by the following highly-engaged post (+4 – indicating four additional posters agreed with the preceding comment) and additional comments (<) expressed agreement:

> *"Wanted to love this program. Seems like a great place and the residents were really nice, but my interview consisted of purely behavioral questions asking me to describe times where I messed up. Spent the whole day talking about every bad experience I've ever had and at the end of the day I just felt like shit. Can't rank them high cuz (sic) I just had such a terrible interview experience +4 <*

Agree, I felt like they did not get to know me at all. They also didn't respond after I did, so I told them a horror story and then they just stared at me and asked the next question... over and over."

The types of questions interviewers asked applicants also influenced their perception of program culture with many posters "shaming" programs for asking "illegal questions" about marital status, relationships and what other programs applicants were applying to. As one poster states, *"Illegal questions all over. PD asked me "so now that we are talking, tell me the real reason why you applied to this program" weird way to ask that question."*

c. *Behavioral cues:* Applicants gauged a program's attitude toward its residents and what the social

environment was like by observing behavioral cues, including how residents were prioritized for COVID-19 vaccines, interviewer conduct and the behavior of residents during pre-interview social events. The interview cycle included in our analysis was the first during the COVID-19 pandemic, and prioritization (or lack thereof) of residents for the initial COVID-19 vaccinations was seen as an indicator of how the program valued residents. A representative post states, *"Sacrificing their residents to COVID by not prioritizing them for vaccination. Institutions treats their residents poorly, clearly. Top program so clearly think they can get away with this crap. Don't rank friends."* A popular post with applicants discussed the positive and negative ways residency programs were handling COVID-19 vaccine distribution in the early stages of the pandemic and what this meant about the way the program and the hospital system valued residents:

> *"Saw on Twitter that the PD supported all the obgyn residents protesting... they arranged for L&D coverage from attendings/APPs so that residents could protest and it wouldn't affect patient care >> yeah, but where were the admin in advocating for the residents prior to that? why were they not in the relevant rooms, or if they were, why were their voices not prioritized? >> yep exactly, love that the program leadership is supportive but it's going to be a long 4 years if your INSTITUTION leadership treats residents like cattle. << Agreed. I know people want to simp [1] because it's[Institution name] but this is a huge red flag as to the way institutions treat residents and I'm glad that at least covid is highlighting this at multiple places."*

Disrespectful interviewer conduct was also commonly discussed by applicants. Students found interviewers arriving late or leaving early, not reading application materials, having webcams turned off or multi-tasking during the interview with the sense that faculty do not value trainees. A representative post reads:

> *"The residents were really kind and I enjoyed my interview experience other than that but the faculty seemed SO uninterested in interviewing me. I interviewed with the interim PD, APD and 2 other faculty members. Some of the faculty had their cameras turned off and the APD answered a phone call when I was mid-sentence and left and never came back."*

Another poster describes a negative interaction with a program director,

---

[1] "Simp" is a derogatory slang term to describe someone was overly self-sacrificing.

> *“PD was chomping away on baby carrots during my interview, unmuted. And she decided to try to fix her internet connection issues during my interview (about halfway through the day) instead of during the break time. Overall got the sense she was uninterested.”*

One post more explicitly states how the interviewer’s behavior was viewed as a surrogate for what it would be like to work with them, *“One of my interviewers clearly didn't read my application. I would not want to work with that faculty member.”*

Applicants commonly used pre-interview social events to observe the social culture of a program focusing on residents’ enthusiasm and friendliness. Notably, impressions of socials were often mixed, even from applicants who attended the same events. A representative post states, *“The residents seemed really hostile towards each other, especially the upper years. Did anyone else experience this?”* and another replied, *“not at all! I had the opposite impression. I loved the interactions between the residents, thought they got along incredibly well. It's one of the biggest things that stood out to me about the program”*.

3. Resident Workload

Applicants wanted to find programs that offered high-quality training while also providing sufficient support for maintaining a work-life balance. They assessed the workload of residents at each program, seeking to strike a balance between rigorous training and personal well-being. Visual, verbal, and behavioral cues observed during virtual interviews and related social events were leveraged by applicants to gauge these aspects of residency program culture.

a. *Visual cues:* Applicants assumed residents were overworked if they appeared tired or continually

mentioned how hard they worked, as seen in this post, *“Kept repeating how busy of a program they are. Residents seemed tired at meet and greet. Overall sense that they are overworked here.”* Another post describes the appearance of the interns at a pre-interview social event and then a medical student from that institution confirms the impression the residents are overworked *“Interns looked a little rough at the meet and greet lol, they all seemed so exhausted :/ << As somebody that goes to [Med School Name], they definitely work a lot. Its 12+ hours in all services except outpatient.”*

b. *Verbal cues:* Verbal cues included questions from interviewers about how applicants intended to

balance the responsibilities of family and pet-ownership while being a resident, implying a lack of program support for residents’ personal lives. A representative post states, *“I also had some very uncomfortable questions like "well what are you going to do with your dog all day while you're at work?" And "does your husband understand how much you're going to be working?" (in a condescending, unfriendly tone).”*

c. *Behavioral cues:* Criticism of interview interruptions by pets or children were seen as evidence

that the program may not understand the needs of its members as described in this post, *“During the social the residents multiple times made a very big deal about how unprofessional it was to have a dog or a partner around while you are interviewing. Like multiple times they were like "we take note if this happens and it looks bad." Like yall*

chill it's a damn pandemic and some of us live in studio apartments and can't pay for pet care etc. Just rude."

Discussion

We found that in 2020-2021, US applicants to ObGyn residency programs prioritized diversity and a supportive social environment within a residency program and were wary of signs of resident mistreatment and burnout when considering program culture and their fit within that culture. Consistent with our findings, the 2022 NRMP applicant survey revealed that applicants rated "goodness of fit" and the "interview day experience" as top priorities when ranking programs. The survey also noted that the majority of applicants found that the virtual environment presented challenges to determining their "fit" within a program and to ascertaining programs' commitment to DEI. Applicants seemed to adapt to the virtual environment over time however and reported increased comfort assessing residency culture and fit in 2022 versus 2021.[1] Our data provides additional context to the NRMP data. The qualitative nature of our study allows for a more detailed description of the specific components of residency program culture that applicants value, as well as how applicants are adapting to the virtual environment and accessing this information online.

Similar to applicants, the corresponding 2022 program director NRMP survey found that the majority of US program directors felt that the virtual platform disadvantaged them in showcasing their programs and finding aligned applicants.[5] Our data challenges this assumption and suggests that programs *are* conveying their program culture virtually, but perhaps in ways program directors have not considered. Programs can leverage our findings to inform improvements to interview processes so

that both programs and applicants are able to effectively assess compatibility. Our results indicate programs should be mindful of DEI tenets when selecting interviewers and interviewees. Additionally, programs should ensure resident involvement in the interview process and make sure their residents appear, and are, well-rested prior to interview days and social events. Programs could consider faculty development around respectful interview behavior and reflect on how interview styles, such as behavioral or interrogative interviewing, convey their residency culture. Over time, as both applicants and programs become more adept at projecting themselves virtually, we anticipate applicants and program directors will gain greater confidence in assessing compatibility through virtual means.

Perceived fit continues to be a priority for both applicants and programs, but it is important to remember that there is no evidence that a good fit results in better outcomes and that a focus on fit may have negative consequences. When fit is defined as similarity, it can be a vehicle for unconscious bias perpetuating inequities and limiting diversity within programs.[6] Further research should evaluate whether achieving program and applicant alignment results in positive outcomes like improved resident workplace satisfaction or lower rates of resident burnout or generates negative outcomes such as reduced racial, ethnic and socioeconomic diversity within residency programs.

The strengths of our research are the qualitative nature, which allows for a more in-depth exploration of applicant considerations around program culture. Additionally, the anonymous nature of the posts gives a different, and perhaps less filtered, perspective of the applicant experience than traditional interview or survey data. As people are typically posting in real time, shortly after their interview experiences, recall

bias is mitigated. There are however limitations to our work. There is selection bias, as the data collected from Reddit represents the perspectives of applicants who chose to post their experiences on social media, which may not be representative of all applicants. Additionally, there is no way to verify the veracity of posts or an accurate number of posters.

Conclusions

Our work offers valuable insights into the considerations of US ObGyn applicants when evaluating program culture and fit during virtual interviews. We found that applicants were interested in the overall social environment of the program and how programs and departments prioritized DEI and resident well-being. Residency programs can benefit from understanding how their selection of interviewers, interview styles, and resident and faculty conduct contribute to conveying these aspects of program culture to applicants. While our research specifically focuses on applicants to ObGyn residencies, the findings likely hold relevance for applicants across specialties.

Acknowledgements and Funding Sources:
We wish to thank Andrea Zorbas, Sharon Blohowiak, Lisa Scott and Cassidy Slinger for administrative support. The project was supported by institutional funding from the University of Wisconsin Department of Obstetrics and Gynecology. ESC is a Medical Scientist Training Program (MSTP) student and was supported in part by MSTP grant T32GM140935 and by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison. ESC

and CMH were partially funded by a grant from an anonymous family foundation.

References

1. National Resident Matching Program, Data Release and Research Committee. Results of the 2022 NRMP Applicant Survey by Preferred Specialty and Applicant Type. National Resident Matching Program, Washington, DC. 2022. URL: https://www.nrmp.org/wp-content/uploads/2022/09/NRMP-2022-Applicant-Survey-Report-Final.pdf. Accessed September 1, 2023.
2. George KE, Gressel GM, Ogburn T, Woodland MB, Banks E. Surveying Obstetrics and Gynecology Residents About Their Residency Applications, Interviews, and Ranking. J Grad Med Educ. 2021 Apr;13(2):257-265. doi: 10.4300/JGME-D-20-00939.1. Epub 2021 Apr 16. PMID: 33897960; PMCID: PMC8054595.
3. Domingo A, Rdesinski RE, Stenson A, et al. Virtual Residency Interviews: Applicant Perceptions Regarding Virtual Interview Effectiveness, Advantages, and Barriers. J Grad Med Educ. 2022;14(2):224-228. doi:10.4300/JGME-D-21-00675.1.
4. Tout W, Oyola S, Sharif Z, VanGompel EW. Applicant Evaluation of Residency Programs in a Virtual Format: A Mixed-Methods Study. Fam Med. 2022 Nov;54(10):804-813. doi: 10.22454/FamMed.2022.148473. Epub 2022 Sep 13. PMID: 36350745.
5. NRMP PD survey. URL: https://www.nrmp.org/wp-content/uploads/2022/09/PD-Survey-Report-2022_FINALrev.pdf. Accessed September 1, 2023.

6. Shappell E, Schnapp B. The F Word: How "Fit" Threatens the Validity of Resident Recruitment. J Grad Med Educ. 2019 Dec;11(6):635-636. doi: 10.4300/JGME-D-19-00400.1. PMID: 31871561; PMCID: PMC6919185.
7. Interviews in GME: Where Do We Go From Here? AAMC. URL: https://www.aamc.org/about-us/mission-areas/medical-education/interviews-gme-where-do-we-go-here. Accessed August 25, 2023.
8. NRMP supports AAMC interview guidance for 2022-23 residency selection cycle. URL: https://www.nrmp.org/about/news/2022/05/nrmp-supports-aamc-interview-guidance-for-2022-23-residency-selection-cycle/. Accessed September 25, 2023.
9. Virtual Experience and holistic REVIEW in the transition to residency: an examination of the 2021 and 2022 main residency matches: research brief. URL: https://www.nrmp.org/wp-content/uploads/2022/07/2022-Virtual-Exper-Research-Brief_Final.pdf. Accessed September 25, 2023.
10. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. Social Media + Society. 2021;7(2):205630512110190. doi:10.1177/20563051211019004.
11. Auxier B, Anderson M. Social Media Use in 2021. Pew Research Center: Internet, Science & Tech. Published April 7, 2021. Accessed September 25, 2023. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/
12. Arshonsky J, Krawczyk N, Bunting A, Frank D, Friedman S, Bragg M. Informal Coping Strategies Among People Who Use Opioids During COVID-19: Thematic Analysis of Reddit Forums. JMIR Form Res2022; 6(3):e32871. URL: https://formative.jmir.org/2022/3/e32871 DOI: 10.2196/32871.

13. Auxier B, Anderson M. Social Media Use in 2021. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/. Published April 7, 2021. Accessed September 25, 2023.
14. Higgins JA, Lands M, Valley TM, Carpenter E, Jacques L. Real-Time Effects of Payer Restrictions on Reproductive Healthcare: A Qualitative Analysis of Cost-Related Barriers and Their Consequences among U.S. Abortion Seekers on Reddit. International Journal of Environmental Research and Public Health. 2021; 18(17):9013. URL: https://doi.org/10.3390/ijerph18179013
15. Nutley SK, Falise AM, Henderson R, Apostolou V, Mathews CA, Striley CW. Impact of the COVID-19 Pandemic on Disordered Eating Behavior: Qualitative Analysis of Social Media Posts. JMIR Ment Health. 2021 Jan 27;8(1):e26011. doi: 10.2196/26011. PMID: 33465035; PMCID: PMC7842857.
16. Belcher RE, Sim D, Meykler M, Owens-Walton J, Hassan N, Rubin RS, Malik RD. A qualitative analysis of female Reddit users' experiences with low libido: how do women perceive their changes in sexual desire? J Sex Med. 2023 Feb 27;20(3):287-297. doi: 10.1093/jsxmed/qdac045. PMID: 36763942.
17. Harrison LN, Neiterman E, MacEachen E, Gibson MF. Navigating return to sex: A qualitative Reddit analysis of parents' perceptions about the timing of resuming sex after a birth. Sex Reprod Healthc. 2022 Dec;34:100782. doi: 10.1016/j.srhc.2022.100782. Epub 2022 Sep 16. PMID: 36179395.
18. Kim H, Sefcik JS, Bradway C. Characteristics of Qualitative Descriptive Studies: A Systematic Review. Res Nurs Health. 2017;40(1):23-42. doi:10.1002/nur.21768.

19. Hale C, Cowley E, Hebgen S, Spencer R, Jacques L. Diversity, Equity, and Inclusion or Bigotry, Inequity, and Exclusion? Applicant Perceptions of Obstetrics and Gynecology Residency Programs. Preprint at: https://www.medrxiv.org/content/10.1101/2023.10.13.23297010v1. (2023)
20. Academic medical leaders and learners reflect on police brutality, racism, and the path forward. URL: https://www.aamc.org/news/academic-medical-leaders-and-learners-reflect-police-brutality-racism-and-path-forward. Published 06 May, 2020. Accessed 25 September, 2023.
21. Lopez L, Hart LH, Katz MH. Racial and Ethnic Health Disparities Related to COVID-19. JAMA. 2021;325(8):719–720. doi:10.1001/jama.2020.26443

Appendix E – An online alternative: A qualitative study of virtual abortion values clarification workshops

This appendix has been published in *Medical Education Online:*

Williams, M., Cowley, E.S., Valley, T.M. Farooque, A., Shultz, Z., Godecker, A., Askins, J., Jacques, L. An online alternative: A qualitative study of virtual abortion values clarification workshops. *Med. Ed. Online.* 28, 1 (2023).

https://doi.org/10.1080/10872981.2023.2258004.

Supplemental materials are available at the publisher's website.

Title: An online alternative: A qualitative study of virtual abortion values clarification workshops

Authors: Margaret Williams, MD[1], **Elise S. Cowley**, BS[2,3], Taryn M. Valley, MA[4,5], Alma Farooque, MD[5], Zoey Shultz, MD[5], Amy Godecker, PhD[1], Jacquelyn Askins, BA[1], Laura Jacques, MD[1]

1. Dept of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI USA

2. Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI USA

3. Dept of Bacteriology, University of Wisconsin-Madison, Madison, WI USA

4. Dept of Anthropology, University of Wisconsin-Madison, Madison, WI USA

5. School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI USA

Corresponding Author

Margaret Williams

mwilliams26@uwhealth.org

202 S Park St. 5E-UPH

Madison, WI 53715

Phone: 205-910-4415

Fax: 608-417-4270

Abstract:

Background: Following the U.S. Supreme Court *Dobbs* decision, access to abortion education is increasingly regionally dependent. Participation in values clarification workshops on abortion can improve abortion knowledge and reduce stigma. Traditionally, values clarification workshops occur in person, yet medical education increasingly utilizes online learning. We sought to understand how a virtual platform impacted medical students and Obstetrics and Gynecology (ObGyn) residents' experience with a values clarification workshop on abortion.

Methods: We conducted values clarification workshops over Zoom with medical students and ObGyn residents at four midwestern teaching hospitals from January 2021-December 2021 during the COVID-19 pandemic. We held semi-structured interviews with participants and facilitators to learn about how the virtual format

impacted their experience with the workshop. Four researchers analyzed transcripts using an inductive approach to generate codes then themes.

Results: We interviewed 24 medical students, 13 ObGyn residents, and five workshop facilitators. Participants and facilitators found the virtual platform to have both unique advantages and disadvantages. Four central themes were identified: 1) Screen as a barrier: participants noted obstacles to conversation and intimacy. 2) Emotional safety: participants felt comfortable discussing sensitive topics. 3) Ease of access: participants could access virtual workshops regardless of location. 4) Technology-specific features: Zoom features streamlined aspects of the workshop and allowed for anonymous contributions to discussion.

Conclusions: Our findings suggest that a virtual platform can be a convenient and effective way to deliver values clarification workshops on abortion, and this technology could be leveraged to expand access to this training in areas without trained facilitators.

Key words: abortion, medical education, virtual, online, values clarification

Introduction

Teaching on abortion is limited in medical education and after the U.S. Supreme Court decision in *Dobbs v. Jackson Women's Health Organization* overturned the federal right to abortion in June 2022, increasingly regionally dependent [1–4]. At the time of writing, 15 states have total abortion bans with very limited exceptions, severely restricting training for at least 230 ObGyn residents and an even greater number of medical students [5,6]. Medical schools and residency programs are now working to fill in the gaps in abortion training for medical trainees [7].

One intervention shown to increase knowledge and support around abortion among participants are Values Clarification and Attitude Transformation (VCAT) Workshops. VCAT workshops have been conducted in a variety of settings around the world reaching broad audiences including clinicians, international development workers, and policy makers [8,9]. In these workshops, participants engage in discussion on abortion-related scenarios, guided by a trained facilitator, and explore their personal and professional beliefs about abortion in an open-minded space [10]. VCAT workshops have traditionally been conducted in-person; however due to COVID-19 in-person learning restrictions and to meet the need for expanded access to abortion education, we adapted a VCAT workshop to a virtual format. Virtual education has been an increasingly utilized and effective method of curriculum delivery [11,12]. The aim of this study was to understand participants' and facilitators' experiences participating in the workshop virtually, assessing both strengths and limitations of the virtual format.

Materials and Methods

Study population

Five facilitators led a total of 29 workshops with medical trainees (medical students and ObGyn residents), of which 26 workshops were with medical students and four with ObGyn residents. We conducted the "Four Corners" portion of a values clarification workshop on abortion at four midwestern teaching hospitals over Zoom [10]. All medical students on their core ObGyn clinical clerkship at three medical schools and all ObGyn residents not on post-call or vacation at four residency programs participated in the workshops virtually from December to January 2021.

Surveys

Prior to the workshop, participants received an email inviting them to complete a voluntary 23-item Qualtrics survey assessing their attitudes (17 statements) and behavioral intentions (six statements) surrounding abortion and demographic information (Appendix 1). This survey was adapted from a previously published survey evaluating the impact of in-person VCAT workshops with international healthcare workers [8]. Participants received a ten-dollar Amazon gift card for completing the survey. Using the 17 attitudes statements on the survey, we created a summative abortion attitude score ranging from 0 (most negative) to 100 (most positive) for each responder [8].

Workshop

Prior to the start of each workshop, participants were required to fill out an online Google form containing 12 statements about abortion and select whether they strongly agreed, agreed, strongly disagreed or disagreed with each statement (Appendix 2). Their responses were de-identified, and each participant was emailed a set of responses from one of their colleagues and was asked to participate with those responses during the workshop.

The workshops were facilitated on Zoom by ObGyn faculty at each institution who received a standardized facilitation guide. The facilitator read each of the 12 Google form statements aloud and then opened Zoom polls for each statement, asking participants to reply with their colleague's de-identified responses. In the traditional Four Corners exercise, each corner of a room is labeled with the four possible responses (Strongly Agree, Agree, Strongly Disagree, Disagree) and participants move to the corner of the room that corresponds with their colleague's response. Participants are asked to reflect on the visual representation of the variety of beliefs held by their colleagues. Two to three of the statements are used as small group discussions, and people in each corner discuss why someone might hold the belief that is presented on their paper. They then share their thoughts with the large group in a facilitated discussion. In the virtual adaptation moving to the four corners of the room is simulated by responding to a zoom poll and the visual representation of the group's values is achieved by displaying poll responses to participants. Zoom breakout rooms are used for small group discussions. In both the in-person and virtual workshops, participants

work from the beliefs presented on the anonymous form of responses that they receive and not their own beliefs. This encourages empathy building through cognitive flexibility.

Interviews

After the workshop, a subset of trainees were invited to participate in a semi-structured interview about their experience. To minimize the effect of volunteer bias, the baseline survey abortion attitude scores for participants were sorted from most negative to most positive and grouped into quartiles. We sought to interview 3 medical students from the lowest and highest attitude quartiles from each school. Of medical students, 72 with attitude scores randomly selected from the highest and lowest quartiles were invited until we reached our target (Figure 1). We also sought to interview 3 non-responders to the initial survey from each institution. Resident participants were also invited to be interviewed about their experience with a goal of 36 interviews. We emailed 4 rounds of invitations, or fewer if we reached the goal number of interviewees prior to that. All residents were eventually invited. All facilitators were invited to be interviewed as well. Interview participants received a one-hundred dollar Amazon gift card.

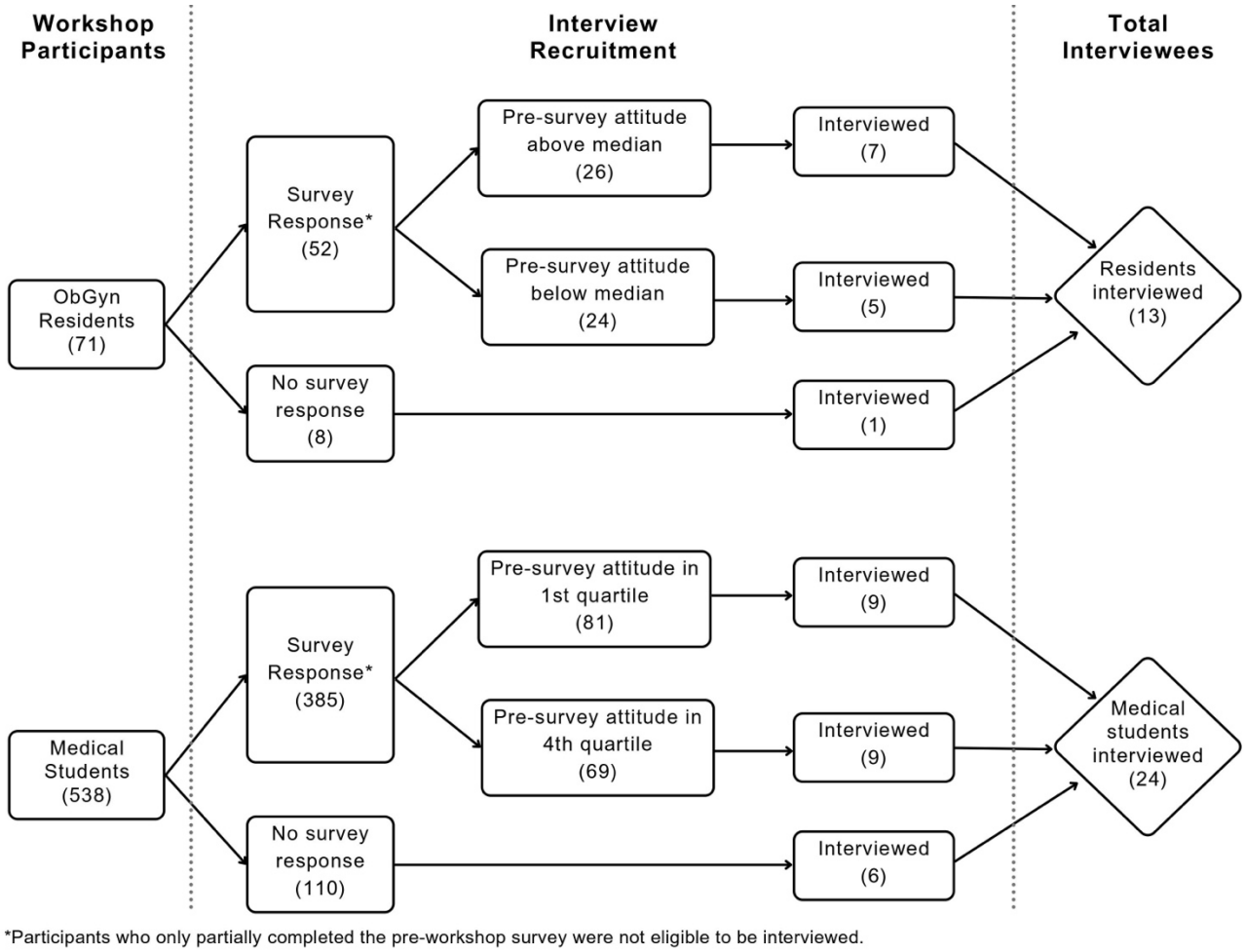


Figure 1. Interview participant recruitment

All interviews were conducted virtually on Zoom by one trained interviewer affiliated with the project, but not involved in student or resident evaluation, using a standardized interview guide (Appendix 3), which was created with input from four authors (TMV, ESC, JA, LJ). Interview questions focused on the experience of participating in the Four Corners activity, students' impressions of their colleagues' abortion beliefs, implications for future practice, as well as likes, dislikes and surprises about the workshop.

Interviews were recorded with participants' consent and transcribed using Zoom autotranscription followed by final transcription by two authors (TMV, AF).

Data analysis

We used an inductive qualitative approach to identify common themes in the data [13]. Four qualitatively-trained researchers (TMV, ESC, AF, ZBS) analyzed interview transcripts. We coded five initial transcripts synchronously to establish consensus and generate a codebook using NVivo software. We then coded remaining transcripts in pairs to ensure consensus throughout. The pair TMV and ESC coded 17 transcripts; AF and ZBS coded 18 transcripts as a pair. One transcript was coded by TMV alone and one was omitted from the coding process because the interviewee did not participate in the workshop. Individual team members (TMV, ESC, AF, ZBS, MW, LJ) identified themes from the codes and then met to establish consensus themes.

Results

In total, 182 medical students and 70 ObGyn residents were invited to participate, of whom 24 medical students and 13 residents were ultimately interviewed (Figure 1). All five workshop facilitators were interviewed. Four central themes about the virtual experience were identified during data analysis: *the screen as a barrier*, *emotional safety, ease of access, and technology-specific features*. These themes are described further with supplemental exemplary quotes in Table 1.

Table 1. Supplemental Quotes

	Supplemental Quotes
Screen as a barrier	R1. It does kind of make it a little bit difficult to have like a free flowing conversation because there are a lot of times where

Participants noted obstacles to conversation and intimacy.	people will all try to kind of speak up at once, and then there's a little bit of a delay for some people, and then you realize you're accidentally talking or something when you didn't even mean to, and then you both stop, and no one talks for like 30 seconds.
	R2. Different approaches to something that is polarizing like abortion can be seen through body language and like unspoken language and things like that which you kind of lose in a virtual sense. Like right now, because I'm trying to be poised I'm like playing with my sock, which is something that like I might do when I'm like nervous or thinking too much and you might not necessarily get that because I'm off screen.
	M1. I guess you can add that to the list of like bonuses for Zoom is that you really get to like watch all of these people and also watch people react, a little bit differently, because you have the checkerboard of like you know 40 faces or whatever it might be. And I thought that was a unique feature because I could also kind of prepare myself in case it was going to get more charged.
	R3. You can't really have side conversations, it's really just you know one person is talking and everyone else is listening, and I think it might have been interesting if it were in person to kind of have had the opportunity to talk with another co resident kind of one-on-one about different ideas.
	M2. I think it's a lot easier to judge people when you're not with them in the flesh...like when people are in person, it often humanizes other people's responses more. [The computer screen is] like this shield and protection, and I think when we take that barrier down, it makes a lot easier for people to humanize each other and to be more real with those responses.
Emotional safety *Participants felt comfortable discussing sensitive topics.*	R4. Trying to put a little distance between you and ideas, to see what other people think about them, I think a virtual platform is actually really good for that...For some reason I felt like using zoom for that activity kind of diffused some of like the tension that could be there in my mind at least.
	F1. I think the virtual platform, maybe added a safety element, because you kind of feel like you're in your own safe space where you're not necessarily in the same space with other people that may have different opinions than you.
	M3. Especially with a sensitive topic, I think that there is a lot of power to having a safe space and to being in that safe space. ... like you could always, you know, like turn your video off and take a breather, take a break, grab a sip of water or something. So it's kind of nice to be able to have that opt out for a few minutes and then jump in if you did need it.
	M4. There's some degree of protection that you feel when you're just sitting in front of your computer screen rather than in front of a group of people...You wonder if your viewpoints will be

	incorrectly perceived by the group. So it feels a little bit safer to be behind your computer, especially in conversation setting that many of us aren't as used to.
Ease of Access *The virtual format increased participants' and facilitators' access to the workshop.*	M5. If it was voluntary, and I had a choice between doing it over Zoom and having to go to a lecture hall, I may opt to do the one over Zoom and not actually walk to the one that's going to physically take some time. I think it brings down some of those access barriers for people.
	F2. We had actually planned to drive to each other's institutions so facilitate each other's workshops, which now seems so quaint, like the concept of driving two hours to give one lecture, seems insane, but at the time we were like that's normal, it's not that far, and now the beauty of Zoom which weirdly I didn't immediately recognize, is that we can facilitate each other's sessions from home, like it's super easy.
Technology-specific features *Zoom features streamlined aspects of the workshop and allowed for anonymous contributions to discussion.*	M6. We've all gotten so used to zoom especially you know as medical students, it's something we do a lot. So that's kind of nice to like just able to like instantly break out and talk and like come back and, like the way people can share their screens and stuff you know to like look at the survey things and talk at the same time that can be useful.
	F3. I do think, specifically for Four Corners, I think this works way better. Because in person, when you have people having to physically move around the room, it wastes a lot of time. Or even some of the other values clarification is where people stand on a spectrum. All of those where people are physically having to move, even though it's more engaging, I find that it takes a lot longer. So I think in general, we are able to get through a lot more in a single hour virtually, than we are in person.
Legend: M indicates a medical student quote, R indicates a resident, and F indicates a facilitator. Each number indicates a different interviewee.	

The screen as a barrier

One disadvantage of the virtual format was that the computer screen created a barrier between participants, which was noted by a majority of the trainees (13 out of 24 medical students and 8 out of 13 residents) and all facilitators. Some participants noted that others kept their cameras off the entire session or stayed muted during the

discussion, which they felt hindered conversation. Some felt that the flow of conversation was impeded over Zoom. They recounted instances when participants would accidentally interrupt one another due to delays in internet connection or because it was difficult to detect cues that another participant was about to speak (Table 1: R1), which might have been easier to distinguish in person.

In addition to maintaining the flow of conversation, participants felt that nonverbal cues were important for other reasons. Many trainees wished they could have seen other participants' body language, such as signs of discomfort after the facilitator read a question aloud or during another participant's response (Table 1: R2). On the other hand, one medical student did report feeling that some cues such as facial expressions were *easier* to detect over Zoom due to the "*checkerboard*" of faces visible on the screen, making it possible to observe several individuals' reactions at once (Table 1: M1).

One facilitator who also had experience facilitating in-person VCAT workshops described challenges due to the lack of nonverbal cues. They recounted difficulty "*read[ing] the room*," or gauging participants' level of understanding and engagement, through the virtual format. They explained:

> *"I had become pretty in depth with having a sense of the vibe of my room when I was doing this in person and making sure who I was losing because I could see their face really clearly. So if I could tell that one table just was not engaging, when we broke into small groups, I'd make sure that I went to that table and tried to*

> *engage them a little bit more... whereas here I can't, I can't see them, and I can't try to bring them back in, so I think that's probably my biggest [issue with the virtual format].*"

In addition to Zoom's impact on group conversation, some participants also noticed the lack of individual side conversations over Zoom. They described wishing that they could process a thought with a peer sitting near them, rather than sharing all thoughts with the group (Table 1: R3). One medical student recalled texting a friend after the workshop to debrief, but expressed that they would have preferred to be able to discuss their thoughts in person "*like us standing in the hallway after*."

Participants also felt that the workshop's virtual format lacked intimacy. They felt their colleagues were more distant, less vulnerable, and hidden behind their screens. One resident reflected, "*[These are] impactful and important discussions to have, and I feel like if you just have it with a black screen it's not as meaningful.*" A few participants feared that this lack of intimacy may have prevented participants from humanizing one another's responses to challenging and controversial questions (Table 1: M2). As one medical student explained:

> *"I think that there's some sense of anonymity in it being virtual even though you can see other people's faces, you're still behind a screen, and so I think that it can be harder to see other people as people rather than just their opinions."*

While participants acknowledged the benefits of increased anonymity during a sensitive conversation, some trainees felt that they should be challenged to have uncomfortable

conversations. Two students felt that the workshop may have been more valuable for medical trainees' professional development if it had required in-person conversation, in order to better prepare them for future discussions or conflicts with healthcare colleagues and patients, "*propelling them into getting used to that scenario.*"

Emotional safety

Although many felt that the virtual platform created barriers, approximately half of the participants (11 out of 24 medical students, 4 out of 13 residents, and 4 out of 5 facilitators) also expressed that the virtual platform created a sense of emotional safety (Table 1: R4, F1). Compared to an in-person classroom, trainees found that the anonymity provided by Zoom helped them feel more comfortable sharing their thoughts, at times referring to the virtual format as a *"safe space*" where they felt "*protect[ed]*" (Table 1: M3, M4). One medical student explained,

> *"It's a little bit less personal, and I think maybe for some individuals with such a charged topic or such a, such a sensitive topic, maybe [virtual VCAT workshops] would be a little bit better. You just feel like it's a little bit less putting yourself out there, less, less chance to feel embarrassed."*

Another medical student felt that "*starting these conversations can be a little bit less intimidating over Zoom."*

People also utilized the functionality of an online platform to increase their feelings of security. Some described how they hid uncontrolled facial expressions by turning off their camera or simply walked away from the computer when they needed an emotional

break (Table 1: M3). One medical student appreciated being able to use the camera functionality to monitor and adjust their facial expressions:

> *"You could have like a safe space, like you could turn your camera off if you want or just like not even be in view of it and be able to read these things and kind of formulate your thoughts, or you don't have to see the reactions of your classmates as you're reading stuff too. I think I feel like it kind of gave some safety to it."*

Virtual classrooms also allow people to participate from a location of their choosing, often at home. Multiple facilitators felt that the comfort of participants' home environments allowed them to talk more openly than they might in a different setting. As one medical student stated, "*People are just kind of inherently more comfortable when they're in their own homes."*

Ease of Access

The virtual format increased access to the workshop for participants and facilitators. Approximately one-quarter of those interviewed (5 out of 24 medical students, 2 out of 13 residents, and 4 out of 5 facilitators) mentioned that the virtual workshop was easy to access. Medical students liked being able to join the workshop from home or wherever was convenient (Table 1: M5). A medical student who was working nights during the workshop explained "*it would have been even harder*" to participate if the workshop had been held in person. Trainees also described the ability to participate during rural or away rotations at a significant distance from their main campuses, when they otherwise might have been unable to participate. One resident explained:

> "F*rom a logistics perspective, like residents at different sites can participate. Sometimes, with the didactics it's hard to get everyone together, and so having people be able to kind of tune in from wherever they were I thought was helpfu*l."

Facilitators also appreciated easier access to the virtual workshop. Facilitators discussed the benefits of being able to leverage the virtual platform to lead sessions at remote locations, either because their learners were at outlying sites or because they were asked to facilitate sessions at different institutions (Table 1: F2). Thinking of students rotating in remote locations, a facilitator reflected:

> *"They don't go to a lot of the teaching...because it was held in person, and they are hours away from here; and now, we can have them all attend to the same workshop, at the same time."*

One facilitator felt the virtual option helped them secure an outside facilitator for their residents, which they thought might help their residents feel more comfortable:

> "*Residents report to me, and I always felt like there was a power differential in terms of them expressing their opinions or ideas or beliefs, even if you attempt to be super supportive... And so I thought that that was a really interesting and innovative thing, that we could facilitate sessions for residents and medical students at remote locations. And then it just makes it so much easier*."

Some facilitators proposed that this method of remote facilitation could allow the workshop to be expanded to other institutions where a family planning clinician may not be locally available or comfortable leading the workshop.

Technology-specific features

Several participants, approximately one-fifth, described that the virtual workshop was efficient due to technological features (2 out of 24 medical students, 4 out of 13 residents, and 3 out of 5 facilitators). Participants and facilitators enjoyed quickly transitioning in and out of breakout groups and the ability to complete surveys ahead of time to maximize discussion time during the workshop (Table 1: M6). One resident described:

> "*I liked that we were able to, one, sort of streamline breaking up into groups. I liked that, you know, all of the pre-work was completed online before, there wasn't like a lot of clutter and I feel like it was something that was relatively compact in terms of its timing, but we accomplished a lot*."

However, some people also described delays with the virtual format, such as slower internet speeds, which negatively affected the workshop's efficiency.

Zoom-specific features also affected the virtual experience. Many participants liked the polling feature, as it allowed facilitators to easily present survey responses. Some appreciated the polls' objectivity, displaying their colleagues' anonymous survey responses, which many like this medical student saw as a true representation of their cohort's beliefs:

> *"A good amount of people believe certain things that you wouldn't have expected, so I liked being able to see those numbers and put it into perspective."*

A few participants also commented on the Zoom chat feature. One resident liked that participants could send a message to the facilitator if they didn't want to share a thought

or question aloud, "*if someone didn't feel comfortable in that environment*." However, one facilitator described an experience during which they became the mediator between two anonymous participants expressing their thoughts in the chat:

> *"It became sort of a back and forth of anonymous discussion that I was then reading to the group so, whereas in person, hopefully, those two people would have had the discussion with each other. Instead here, it became, you know, the anonymity allowed people to say things that they might not have been as comfortable saying in person."*

Breakout rooms were another commonly utilized feature during the workshop. Some participants felt that the transitions in and out of breakout rooms were smooth and more efficient than if the transition to small groups had occurred in person. Those with previous in-person workshop experience also felt that the virtual platform saved time by decreasing the amount of time participants spent physically moving around the room (Table 1: F3). Some also felt that the virtual breakout rooms created more privacy and less "*chaos*" compared to multiple small group discussions occurring at once, "*in a big room with a lot of noise it's actually not a very fruitful conversation*."

However, others felt that the transitions between online breakout rooms inhibited discussion among participants. A few participants expressed that more facilitators would be needed to effectively mediate conversations taking place in multiple breakout rooms. One medical student described being in a breakout room without a facilitator as "*awkward*." Facilitators echoed these sentiments. One facilitator felt that their sudden

appearance in a virtual breakout room altered the course of conversation, and worried that if they closed the breakout rooms, they might unknowingly bring a robust discussion in a different room to an end prematurely; they summarized, *"the virtual way of doing it really cuts off discussion."*

Discussion

Values clarification workshops on abortion have proven to be an effective tool for educating international healthcare workers about abortion and reducing abortion stigma [8,9]. They have also been shown to be an effective tool for ObGyn residents training at religiously-affiliated hospitals in the United States. ObGyn residents in this setting who participated in VCAT workshops showed increased acceptance of abortion care post-workshop [9]. Previous research has only evaluated the impact of in-person workshops, but there has not yet been an evaluation of a virtual adaptation of these workshops.

Many studies have evaluated the efficacy of online modalities for medical education and demonstrated similar efficacy to in-person learning, including students' ability to retain knowledge and develop communication skills [11,12,14,15]. A separate qualitative analysis of this study's data supported this hypothesis for VCAT in particular: participation in the virtual VCAT workshop helped trainees understand their own and others' views on abortion and practice professional communication, similar to outcomes for in-person VCAT workshops [16]. Our study adds to the existing literature by describing how the virtual format affected participants' experiences engaging in VCAT workshops.

Our findings suggest that using an online platform to deliver values clarification workshops on abortion provides both unique advantages and disadvantages to in-person instruction. The original workshop takes steps to create a safe environment for participants by having them participate using an anonymous colleague's survey responses, rather than their own. Themes from the interviews we conducted with medical trainees highlight how the virtual platform additionally allowed for more comfortable discussion around a stigmatized topic. Participants attributed their feelings of emotional safety during these discussions to the distance between participants created by the virtual format. Additionally, the online format lowered barriers, primarily travel time and effort, to accessing the workshop for both trainees and facilitators.

One commonly cited drawback to online education is a lack of social connection among learners and educators. In medical education, learners have reported feeling less connected and described challenges to communicating virtually in online courses [17,18]. This finding is supported by data in our study. A majority of participants and facilitators in the virtual VCAT workshop described the virtual platform as a barrier to connecting with others, citing a lack of intimacy and difficulty detecting non-verbal communication cues. Educators should weigh how the virtual format may both contribute to emotional safety and simultaneously reduce intimacy among learners when determining whether a virtual format best fits the specific needs of their learners and learning environment.

Despite some limitations of online education, virtual VCAT workshops may be a timely intervention after the Supreme Court *Dobbs* decision. Recent data shows that 56

ObGyn residency programs, approximately one-fifth of all programs, are in states with the most restrictive abortion bans [6]. We cannot know how medical trainees' attitudes and behavioral intentions towards abortion will be affected over time in the aftermath of *Dobbs*. In this context, virtual values clarification workshops may become increasingly useful for combatting abortion stigma, as remote facilitators of the online workshop can reach medical trainees in more restrictive states where trained facilitators may be unavailable or less comfortable leading workshops. The findings of our study may assist medical educators by helping them weigh the advantages and disadvantages of a virtual VCAT workshop in their particular legal and cultural context, providing evidence to inform whether this format may benefit their trainees.

Our study design is a strength of this research. This was a multi-institution study, and we recruited participants with a range of baseline attitudes towards abortion prior to the workshop as well as non-responders. Additionally, conducting semi-structured interviews allowed us to elicit participants' nuanced descriptions of their experiences with the workshop, and produced a rich dataset. Despite efforts to recruit interviewees with a range of opinions about abortion, selection bias remains a limitation of our study. Additionally, we anticipate limitations to generalizability of our data, given that our cohort consisted of Midwestern medical trainees and faculty. As the workshop is expanded to other regions, future research should assess trainees' experiences in other geographic areas.

Using a virtual platform to deliver values clarification workshops on abortion is feasible and provides specific advantages of anonymity, safety, and accessibility, although at the potential cost of reduced vulnerability among participants.

Acknowledgements

The authors thank Andrea Zorbas, Sharon Blohowiak, Amanda Wildenberg, and Kelly Winum for administrative support. The authors also thank Nathan Jones and the UW-Madison Survey Center.

This project was partially funded through a departmental grant. TMV is a Medical Scientist Training Program (MSTP) student funded in part by the CORE Lab, funded by a large, anonymous family foundation, and in part by Medical Scientist Training Program grant T32GM140935. ESC is an MSTP student and was supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by MSTP grant T32GM140935.

Declaration of Interest Statement

Ethical approval has been waived for this study by the University of Wisconsin-Madison Minimal Risk IRB in August 2020, reference 2020-0803. The authors have no conflicts of interest to report.

References

1. Landy, U., Darney, P.D., and Steinauer, J. eds. (2021). Advancing Women's Health Through Medical Education: A Systems Approach in Family Planning and Abortion (Cambridge University Press) 10.1017/9781108884709.

2. Rennison, C., Woodhead, E.J., Horan, C., Lohr, P.A., and Kavanagh, J. (2022). Abortion education in UK medical schools: a survey of medical educators. BMJ Sex Reprod Health *48*, 210–216. 10.1136/bmjsrh-2021-201387.

3. Espey, E., Ogburn, T., Chavez, A., Qualls, C., and Leyba, M. (2005). Abortion education in medical schools: a national survey. Am J Obstet Gynecol *192*, 640–643. 10.1016/j.ajog.2004.09.013.

4. Vinekar, K., Karlapudi, A., Nathan, L., Turk, J.K., Rible, R., and Steinauer, J. (2022). Projected Implications of Overturning Roe v Wade on Abortion Training in U.S. Obstetrics and Gynecology Residency Programs. Obstetrics & Gynecology *140*, 146. 10.1097/AOG.0000000000004832.

5. Institute, G. Interactive Map: US Abortion Policies and Access After Roe. https://states.guttmacher.org/policies/.

6. Ryan Program Locations Ryan Program. https://ryanprogram.org/home/overview/ryan-program-locations/.

7. Mengesha, B., Zite, N., and Steinauer, J. (2022). Implications of the Dobbs Decision for Medical Education: Inadequate Training and Moral Distress. JAMA *328*, 1697–1698. 10.1001/jama.2022.19544.

8. Turner, K.L., Pearson, E., George, A., and Andersen, K.L. (2018). Values clarification workshops to improve abortion knowledge, attitudes and intentions: a pre-post assessment in 12 countries. Reprod Health *15*, 40. 10.1186/s12978-018-0480-0.

9. Guiahi, M., Wilson, C., Claymore, E., Simonson, K., and Steinauer, J. (2021). Influence of a values clarification workshop on residents training at Catholic Hospital programs. Contraception: X *3*, 100054. 10.1016/j.conx.2021.100054.

10. Turner, K., L., and Chapman Page, K. (2008). Abortion attitude transformation: A values clarification toolkit for global audiences.

11. Pei, L., and Wu, H. (2019). Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. Medical Education Online *24*, 1666538. 10.1080/10872981.2019.1666538.

12. Kyaw, B.M., Posadzki, P., Paddock, S., Car, J., Campbell, J., and Tudor Car, L. (2019). Effectiveness of Digital Education on Communication Skills Among Medical

Students: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res *21*, e12967. 10.2196/12967.

13. Thomas, D.R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation *27*, 237–246. 10.1177/1098214005283748.

14. Co, M., Chung, P.H.-Y., and Chu, K.-M. (2021). Online teaching of basic surgical skills to medical students during the COVID-19 pandemic: a case–control study. Surg Today *51*, 1404–1409. 10.1007/s00595-021-02229-1.

15. Almhdawi, K.A. (2022). Influencing Medical Students' Knowledge and Attitudes Related to Disability: A Randomized Controlled Trial. Am J Phys Med Rehabil *101*, 113–118. 10.1097/PHM.0000000000001740.

16. Valley, T.M., Cowley, E.S., Farooque, A., Shultz, Z.B., Williams, M., Askins, J., Godecker, A., and Jacques, L. (2023). "We had to put ourselves in their shoes": Experiences of Medical Students and ObGyn Residents with a Values Clarification Workshop on Abortion. Preprint at medRxiv, 10.1101/2023.02.16.23286043 10.1101/2023.02.16.23286043.

17. Evans, A.Z., Adhaduk, M., Jabri, A.R., and Ashwath, M.L. (2023). Is Virtual Learning Here to Stay? A Multispecialty Survey of Residents, Fellows, and Faculty. Current Problems in Cardiology *48*, 101641. 10.1016/j.cpcardiol.2023.101641.

18. Reinhart, A., Malzkorn, B., Döing, C., Beyer, I., Jünger, J., and Bosse, H.M. (2021). Undergraduate medical education amid COVID-19: a qualitative analysis of enablers and barriers to acquiring competencies in distant learning using focus groups. Medical Education Online *26*, 1940765. 10.1080/10872981.2021.1940765.

Appendix F – Abortion Attitudes and Behavioral Intentions of Obstetrics and Gynecology Residents at Four Midwestern Residency Programs in 2021

This appendix has been accepted at the *Wisconsin Medical Journal*, but is not in press as of the date of this defense. Supplemental materials are available from the corresponding author.

Title: Abortion Attitudes and Behavioral Intentions of Obstetrics and Gynecology Residents at Four Midwestern Residency Programs in 2021

Authors: Abigail S. Cutler[1], **Elise S. Cowley**[2,3], Jessika A. Ralph[4], Jessie Chen[4], Amy Godecker[1], Jordan Ward[1], Sarah Hutto[4], Laura Jacques[1]

Affiliations:

[1]Dept of Obstetrics and Gynecology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America

[2]Dept of Bacteriology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America

[3]Microbiology Doctoral Training Program, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, United States of America

[4] Dept of Obstetrics and Gynecology, University of Minnesota, Minneapolis, Minnesota, United States of America

Corresponding Author:

Abigail S. Cutler, MD, MPH

451 Junction Road

Madison, WI 53717

608.265.7601

ascutler@wisc.edu

Abstract

Background: In June 2022, the U.S. Supreme Court announced its decision in *Dobbs V. Jackson Women's Health Organization* to overturn *Roe v Wade*. As a result, half of U.S. states now face proposed or in-effect abortion bans, which affect the ability of obstetrics and gynecology (ObGyn) residency programs to provide abortion training.

Objective: To establish ObGyn residents' pre-*Dobbs* attitudes toward abortion, desire to learn about abortion, and intentions about providing abortion care in their future practice.

Methods: From January through December 2021, we surveyed 70 ObGyn residents at four programs in Wisconsin and Minnesota to assess their attitudes toward abortion, desire to learn about abortion, and intentions about providing abortion care in their future practice.

Results: A total of 55 out of 70 (79%) ObGyn residents completed the survey. Most reported highly favorable attitudes toward abortion, nearly all found the issue of abortion to be important, and the majority planned to incorporate abortion care into their future

work. There were no differences in median attitude scores or behavioral intentions among institutions.

Conclusions: Prior to the reversal of *Roe v Wade,* ObGyn residents in Minnesota and Wisconsin viewed abortion as important health care and intended to provide this care after graduation.

Introduction

The reversal of *Roe V. Wade* by the United States Supreme Court on June 24, 2022 has resulted in proposed or in-effect abortion bans spanning half the country.[1] Despite that nearly a quarter of women will have an abortion in their lifetimes and that abortion training is a required component of Obstetrics and Gynecology (ObGyn) resident education by the Accreditation Council of Graduate Medical Education, nearly half of all ObGyn residency programs now struggle to provide clinical training in this common health care service.[2–6] The media has also raised concerns that wide geographic variations in abortion legality will adversely shape where physicians choose to train and ultimately practice, which could further exacerbate existing ObGyn shortages.[7–9] To better understand how the fall of *Roe* may affect U.S. ObGyn residents' career decisions, it is crucial to understand what their baseline attitudes were towards abortion prior to the fall of *Roe*, what their desire was to learn about abortion, and the importance they placed on being able to provide abortion care in their future work. Current literature on these topics is sparse.[10] Prior to the fall of *Roe,* we assessed attitudes and career intentions toward abortion among ObGyn residents in Minnesota and Wisconsin where

19.5% and 15.3% of counties, respectively, already qualify as maternity care deserts and access to abortion, post-*Dobbs,* significantly differs. [11,12]

Methods

We included all ObGyn residents who were scheduled to participate in a workshop at the University of Minnesota Twin Cities (UMN), University of Wisconsin—Madison (UW), Medical College of Wisconsin (MCW) and Aurora-Sinai Milwaukee (Aurora) between January and December 2021. A required component of their residency didactics curriculum, the workshop, adapted from the Values Clarification and Attitudes Transformation (VCAT) workshop published by Turner et al, was designed to help participants explore their attitudes toward abortion.[13,14] All residents were required to attend this workshop unless they had an approved absence (e.g. post-call, vacation). We emailed all residents a link to a confidential, voluntary survey using a web-based platform (Qualtrics). The first screen of the survey informed potential participants that completion of the survey would be considered consent to participate in our research study. Survey participants received a ten-dollar Amazon gift-card link.

We gathered demographic information and used a previously published questionnaire (adapted from workshop materials published Turner et al; see Supplemental Materials) to assess attitudes toward abortion care and behavioral intentions for future practice.[13,14] To assess attitudes, we asked the degree to which participants agreed with 17 statements about abortion using a five-point Likert scale. To assess behavioral intentions, we posed six yes/no questions regarding intent to learn about, advocate for,

refer patients to, and provide abortion care. To compare attitudes among participants overall rather than item-by-item, we followed Turner et al.'s analytic methodology and created summative attitude scores which ranged from zero (most negative toward abortion) to 100 (most positive toward abortion). Summative attitude scores were calculated by summing the five-point responses, dividing by five times the total number of items (17) and multiplying by 100. Summative behavioral intention scores were calculated similarly, dividing the sum of positive responses by the total number of items (6), and multiplying by 100. [14]

To compare attitude scores by demographic characteristics, we used the Wilcoxon rank sum test and the Kruskal-Wallis equality of distributions rank test, as appropriate. Given the observed distribution of answers to the behavioral intention questions (80% of respondents answered "Yes" to all six items), we dichotomized responses into "all yes" vs "any no" and used Fisher's exact test to test for significant differences by demographic characteristics. P-values <0.05 were considered statistically significant. The study was reviewed and considered exempt by the University of Wisconsin-Madison Minimal Risk IRB and reviewed and approved by the University of Minnesota IRB.

Results

A total of 55 out of 70 (79%) ObGyn residents completed the survey, 17 of 21 (81%) from UMN, 14 of 20 (70%) from UW, 14 of 16 (88%) from MCW, and 10 of 13 (77%) from Aurora. Of residents who completed the survey, a majority (n=46, 84%) identified

as women, were born in the Midwest (n=28, 51%), and did not identify with a particular religion (n=36, 65%) (Table 1).

Table 1. Median attitude and behavioral intention scores of survey participants, by selected demographic characteristics.

Resident characteristics	N (%)	Attitude Score, Median (Interquartile Range)	P value	Behavioral Intention Score, n (%) saying "Yes" to all 6 items	P value
Total	**55 (100)**	**95 (87-98)**		**44 (80)**	
Institution			0.436		0.298
University of Minnesota - Twin Cities	17 (31)	95 (93-99)		16 (94)	
Medical College of Wisconsin	14 (25)	92 (85-96)		10 (71)	
Aurora-Sinai Milwaukee	10 (18)	96 (91-99)		8 (80)	
University of Wisconsin - Madison	14 (25)	96 (87-98)		10 (71)	
Gender			0.413		0.179
Man	9 (16)	95 (95-99)		9 (100)	
Woman	46 (84)	95 (87-98)		35 (76)	
Birthplace			0.819		0.545
Outside the United States	5 (9)	91 (88-92)		4 (80)	
United States, not Midwest	22 (40)	95 (89-99)		16 (73)	
United States, Midwest	28 (51)	96 (86-98)		24 (86)	
Religiosity			0.046		
Do not identify with a religion	36 (65)	95 (90-99)		30 (83)	
Identify with religion; incorporate into daily life none/little	6 (11)	96 (92-99)		6 (100)	
Identify with religion;	13 (24)	87 (84-96)		8 (62)	

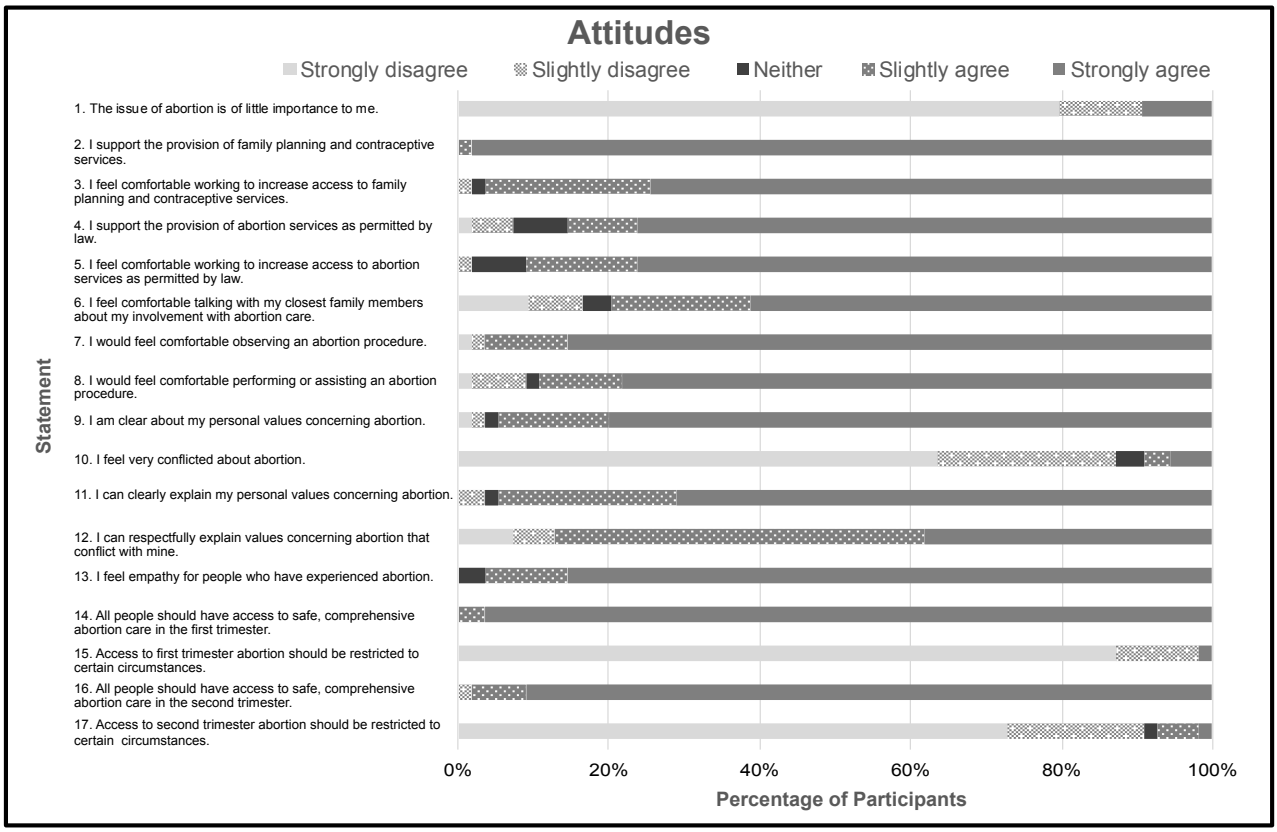
incorporate into daily life some/quite a bit/great deal					
Post-Graduate Year			0.773		0.831
1	17 (31)	95 (87-98)		14 (82)	
2	14 (25)	96 (85-99)		10 (71)	
3	12 (22)	96 (93-99)		10 (83)	
4	12 (22)	95 (89-97)		10 (83)	
Interested in pursuing a fellowship			0.533		0.712
No	39 (71)	95 (88-99)		32 (82)	
Yes	16 (29)	95 (86-97)		12 (75)	

Wilcoxon rank sum or Kruskal-Wallis rank tests for attitude scores; Fisher's exact test for behavioral intentions; P-values <0.05 considered statistically significant

The median attitude score for ObGyn residents was 95 (Interquartile Range: 87-98). Differences in the distributions of attitude and behavioral intention scores among institutions were not statistically significant (Table 1). Behavioral intention scores did not significantly differ across participant characteristics. Attitude scores significantly differed only by religiosity: residents who did not identify with a religion held more positive overall attitudes than residents for whom religion impacted their daily lives (Table 1). Nearly all 55 residents (n=49, 89%) disagreed with the statement "the issue of abortion has little importance to me". Almost all residents agreed that "All people should have access to safe, comprehensive abortion care in the first (n=55, 100%) and second (n=54, 98%) trimester. Nearly all residents (n=52, 95%) wanted to learn more about the need for safe, comprehensive abortion care, 54 (98%) planned to refer people seeking

abortion to safe services, and 46 (84%) planned to provide abortion care in their future careers (Figure 1).

A



B

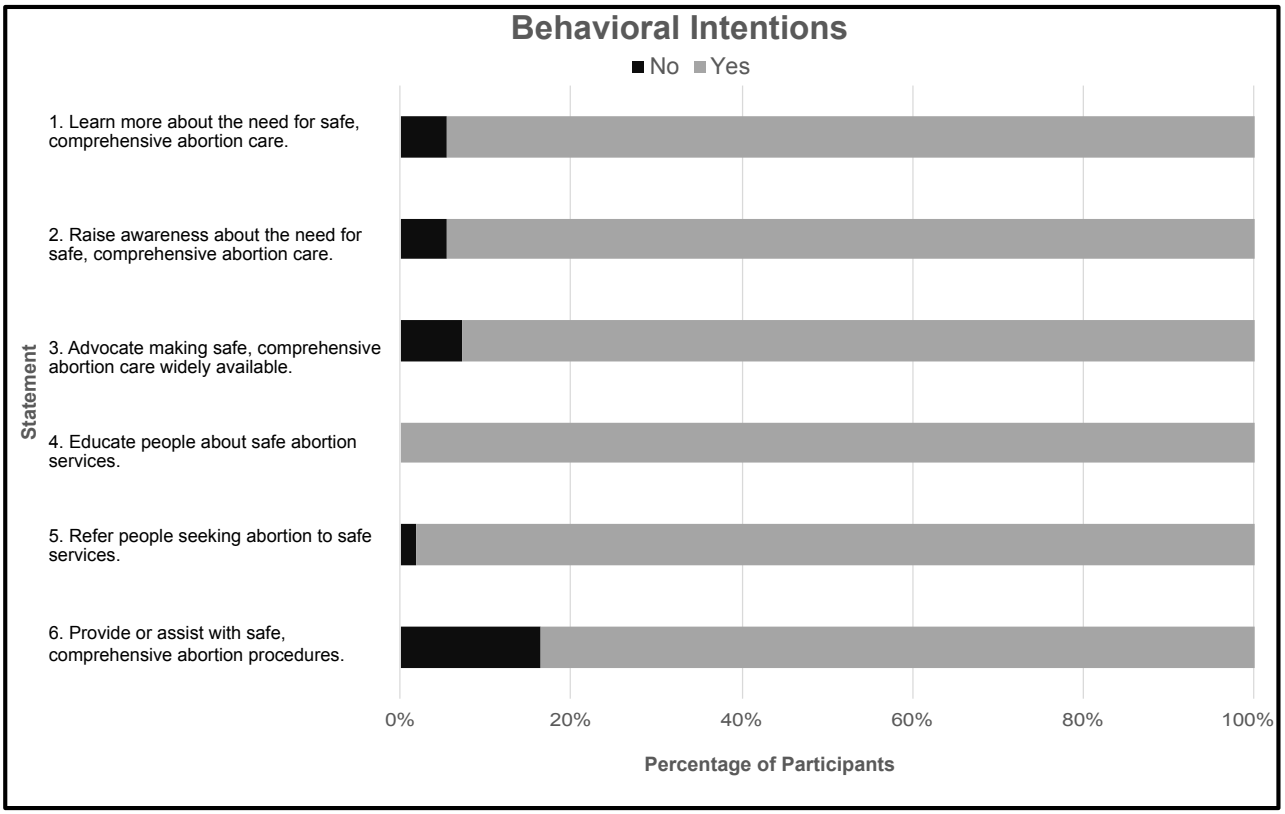


Figure 1. Summary of attitudes and behavioral intentions on abortion for all participants. Percentage of participants for each response option for A. attitudes and B. behavioral intentions on abortion.

Discussion

Our research demonstrates that prior to the fall of *Roe*, ObGyn residents in our sample from Wisconsin and Minnesota held highly favorable attitudes toward abortion, believed abortion should be available to patients, desired education and training in abortion care, and planned to directly provide or refer patients for abortion care in their future practice. Notably, our study was conducted in two states that faced different abortion access

restrictions even prior to the *Dobbs* decision. Unlike Wisconsin, where restrictions to abortion proliferated prior to the fall of *Roe*, Minnesota had fewer barriers to access in place.[15,16] Despite these differences in legal landscapes, we found no significant difference in attitudes toward abortion or plans to incorporate abortion into their future work between state populations.

Following *Dobbs,* the two states' legal landscapes have diverged even farther from each other: Minnesota has codified abortion rights into its state constitution and Wisconsin has reverted to an 1849 state law that criminalizes the provision of abortion in nearly all circumstances.[17,18] Although our data do not offer insight into the extent to which post-*Dobbs* restrictions are shaping resident recruitment and decision-making, this study is strengthened by a high survey response rate and provides critical baseline data to understand how the *Dobbs* decision will affect ObGyn residents and future career plans going forward. Understanding the forces that shape the future ObGyn workforce is key, particularly in light of concerns pre-dating *Dobbs* about impending ObGyn shortages in certain areas of the country.[19] As other researchers have noted, physician attitudes not only have the capacity to guide important stakeholders such as media, policymakers and voters, but they also can carry weight in their institutions, whose responsibilities include recruiting and retaining a robust and willing labor force.[10]

It is worth noting that even when abortion was a constitutional federally protected right, only 60% of ObGyn residents reported "routine" access to abortion training; satisfaction with abortion training was positively and independently correlated with the routine

availability of this training.[20] As graduate medical education transitions to a post-*Roe* world, residency programs should evaluate their current recruitment and educational strategies to maximally ensure that all ObGyn residents who are legally able to receive adequate abortion training do so. This may necessitate establishing out-of-state training partnerships for programs in restricted states. Education leaders should also work within academic training sites to reduce or extinguish wherever possible all other institutional-level barriers to the provision of abortion care in the inpatient and outpatient settings. Finally, new initiatives may include implementing and studying interventions designed to improve attitudes and behavioral intentions toward abortion care – such as mandated values clarification workshops – so that demand for training among residents remains high.[21]

Conclusion

Prior to the fall of *Roe,* most ObGyn residents in two Midwestern states with significantly different abortion access held highly favorable attitudes toward abortion and planned to provide abortion care in their future practice. How and to what extent recent seismic changes in the legal landscape will shape the future, post-*Roe* ObGyn workforce remains unknown. While some graduating residents who value abortion training and provision may avoid practicing in states where abortion is restricted, others may be drawn to practice where the need for advocacy is high. Future research should directly evaluate how post-*Roe* state-level abortion restrictions, such as those in Wisconsin, impact both recruitment into ObGyn residency programs, career decision-making

among graduating residents, as well as the availability, accessibility and quality of pregnancy-related health care for patients.

Acknowledgements:
We wish to thank Andrea Zorbas, Sharon Blohowiak, Amanda Wildenberg, and Kelly Winum for administrative support. We thank Allison Linton and Kristina Kaljo for their contributions to study design and recruitment. We thank Nathan Jones and the UW-Madison Survey Center.

References

1. 26 States Are Certain or Likely to Ban Abortion Without Roe: Here's Which Ones and Why (2021). Guttmacher Institute. https://www.guttmacher.org/article/2021/10/26-states-are-certain-or-likely-ban-abortion-without-roe-heres-which-ones-and-why.

2. Kortsmit, K. (2021). Abortion Surveillance — United States, 2019. MMWR Surveill Summ *70*. 10.15585/mmwr.ss7009a1.

3. Jones, R.K., and Jerman, J. (2017). Population Group Abortion Rates and Lifetime Incidence of Abortion: United States, 2008–2014. Am J Public Health *107*, 1904–1909. 10.2105/AJPH.2017.304042.

4. ACGME (2021). Obstetrics and Gynecology Milestones - The Accreditation Council for Graduate Medical Education.

5. Vinekar, K., Karlapudi, A., Nathan, L., Turk, J.K., Rible, R., and Steinauer, J. (2022). Projected Implications of Overturning Roe v Wade on Abortion Training in U.S. Obstetrics and Gynecology Residency Programs. *00*, 4.

6. Murphy, B. (2023). After Dobbs, M4s face stark reality when applying for residency. American Medical Association. https://www.ama-assn.org/medical-students/preparing-residency/after-dobbs-m4s-face-stark-reality-when-applying-residency.

7. Meyerhofer, K. (2022). In a post-Roe world, some medical students rethink plans to practice in Wisconsin. Journal Sentinel. https://www.jsonline.com/story/news/education/2022/06/29/roe-ruling-abortion-ban-affects-wisconsin-doctor-recruitment-uw-madison-medical-college-mcw-ob-gyn/7747097001/.

8. Patel, K. (2022). The consequences of abortion bans for people who don't want one are growing. MSNBC.com. https://www.msnbc.com/opinion/msnbc-opinion/abortion-bans-are-dangerous-roadblock-america-s-future-doctors-n1297263.

9. Weiner, S. (2022). How the repeal of Roe v. Wade will affect training in abortion and reproductive health. AAMC. https://www.aamc.org/news/how-repeal-roe-v-wade-will-affect-training-abortion-and-reproductive-health.

10. Schmuhl, N.B., Rice, L.W., Wautlet, C.K., and Higgins, J.A. (2021). Physician attitudes about abortion and their willingness to consult in abortion care at a Midwestern academic medical center. Contraception *104*, 278–283. 10.1016/j.contraception.2021.04.030.

11. Fontenot, J., Lucas, R., Stoneburner, A., Brigance, C., Hubbard, K., Jones, E., and Mishkin, K. (2023). Where You Live Matters: Maternity Care Deserts and the Crisis of Access and Equity in Minnesota.

12. Fontenot, J., Lucas, R., Stoneburner, A., Brigance, C., Hubbard, K., Jones, E., and Mishkin, K. (2023). Where You Live Matters: Maternity Care Deserts and the Crisis of Access and Equity in Wisconsin.

13. Turner, K., L., and Chapman Page, K. (2008). Abortion attitude transformation: A values clarification toolkit for global audiences.

14. Turner, K.L., Pearson, E., George, A., and Andersen, K.L. (2018). Values clarification workshops to improve abortion knowledge, attitudes and intentions: a pre-post assessment in 12 countries. Reproductive Health *15*, 40. 10.1186/s12978-018-0480-0.

15. Higgins JA, Ufot MI, Williamson A. (2022). Wisconsin state laws impacting abortion access (University of Wisconsin Collaborative for Reproductive Equity.).

16. Nelson, E. Minnesota is bracing for an increase in abortion demand. https://www.startribune.com/minnesota-braces-for-influx-of-out-of-state-abortion-patients/600093418/.

17. Petrovic, P. (2022). Wisconsin's 173-year-old ban allows only life-saving 'therapeutic abortions.' No one knows what that means. Wisconsin Watch. http://wisconsinwatch.org/2022/08/wisconsins-173-year-old-ban-allows-only-life-saving-therapeutic-abortions-no-one-knows-what-that-means/.

18. HF 1 as introduced - 93rd Legislature (2023 - 2024) https://www.revisor.mn.gov/bills/text.php?number=HF1&type=bill&version=0&session=ls93&session_year=2023&session_number=0.

19. Orvos, J.M. (2017). ACOG releases new study on ob/gyn workforce: trends similar to those seen in previous studies expected to continue. Contemporary OB/GYN *62*, 50–53.

20. Horvath, S., Zite, N., Turk, J., Ogburn, T., and Steinauer, J. (2021). Resident Abortion Care Training and Satisfaction: Results from the 2020 Council on Resident Education in Obstetrics and Gynecology In-Training Examination Survey. Obstetrics & Gynecology *138*, 472–474. 10.1097/AOG.0000000000004512.

21. Guiahi, M., Wilson, C., Claymore, E., Simonson, K., and Steinauer, J. (2021). Influence of a values clarification workshop on residents training at Catholic Hospital programs. Contraception: X *3*, 100054. 10.1016/j.conx.2021.100054.

Appendix G – An in-room observation study of hand hygiene and contact precaution compliance for *Clostridioides difficile* patients

This appendix has been published in *American Journal of Infection Control:*

Barker, A.K., Cowley, E.S., McKinley, L. Wright, M.O., Shultz, Z., Safdar, N. An in-room observation study of hand hygiene and contact precaution compliance for Clostridioides difficile patients. *Am J Infect Control.* 47, 10 (2019). https://doi.org/10.1016/j.ajic.2019.03.031.

An in-room observation study of hand hygiene and contact precaution compliance for *Clostridioides difficile* patients

Anna K. Barker, PhD[#a], **Elise S. Cowley**, HBS[#b], Linda McKinley, RN, MPH, CIC, FAPIC[c], Marc-Oliver Wright, MS, CIC, FAPIC, MT (ASCP)[d], Nasia Safdar, MD, PhD[c,d,e,*]

[a]Department of Population Health Sciences, University of Wisconsin-Madison, School of Medicine and Public Health, Madison, WI

[b]Medical Scientist Training Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI

[c]William S. Middleton Memorial Veterans Hospital, Madison, WI

[d]Infection Control, University of Wisconsin Hospitals and Clinics, Madison, WI

[e]Division of Infectious Diseases, Department of Medicine, University of Wisconsin-Madison,
School of Medicine and Public Health, Madison, WI

[#] These authors contributed equally to this work.

Abstract

Using an innovative, covert, in-room observer method to evaluate infection control practices for patients with Clostridioides difficile infection, we found no difference between physician and nursing hand hygiene compliance and contact precaution usage. There was also no diurnal variation in hand hygiene practices, but decreased contact precaution usage at night. Conversely, hospital-wide data from overt observations collected over the same time period showed significantly higher hand hygiene compliance among nurses than physicians.

Keywords

Clostridioides difficile infection; Hand hygiene compliance; Diurnal effects; Covert observations

Introduction

Clostridioides difficile is a major cause of hospital-acquired infection in the United States.[1] Its prevention requires compliance with hand hygiene, particularly soap and water, and contact precautions. Despite numerous prior studies of hand hygiene and contact precautions, few focus on patients with *C difficile* infection (CDI). Furthermore, most rely on overt observations or self-reported data, as opposed to covert observations.[2] Overt reporters consistently overestimate covertly measured hand hygiene rates by 20%−60%, making it difficult to accurately track compliance and intervention effectiveness.[3-5]

Given *C difficile's* unique infection control considerations and the urgency of reducing infection, novel approaches to measuring compliance are needed. Electronic monitoring systems recently became available to automate tracking. However, their use is limited by high upfront cost and unclear accuracy.[6,7] Covert observation remains the gold-standard method, yet is typically also costly and labor intensive.[6] Therefore, we conducted an innovative, in-room observation study using volunteers and light-duty staff to assess hand hygiene and contact precaution practices of hospital employees and visitors interacting with CDI patients.

Methods

The study was conducted from December 2015 to June 2018, at a 565-bed tertiary-care hospital in Madison, Wisconsin. We employed an in-room methodology to evaluate hand hygiene and contact precaution compliance using volunteer and light-duty staff as covert observers. Observers spent at least 15 minutes inside each CDI patient room, collecting data on entering hospital employees and visitors. Hand hygiene compliance at entry was defined as using alcohol handrub or soap and water. Rubbing hands together on room approach was allowed, as people were presumed to have applied handrub immediately prior. At study inception, our facility had sustained hyperendemic rates of CDI. Therefore, compliance at exit was defined as soap and water use at an inside or outside room sink. Contact precaution compliance was defined as wearing gown and gloves at room entry. Data were categorized into day (7 AM to 3 PM) and night (3 PM to 7 AM) shifts.

All observers completed one-on-one training with a single infection preventionist prior to data collection. They wore identification badges and red polo volunteer shirts, neither of which were visible through the opaque isolation gowns donned on room entry. In an effort to maintain covert observations, we trained and used 195 different observers, to date. Each typically conducted observations once weekly over the course of a college semester.

Observations were recorded on OpenText Teleform (OpenText Corp, Waterloo, ON) scannable forms and scanned into a Microsoft Access (Microsoft Corp, Redmond, WA) database. If questioned by staff, covert observers used a standard script to discuss the project (Fig 1).

Script for introducing self to patient:

Good morning/afternoon/evening Mr. / Mrs. _____. My name is _____. As part of our regular work here at UW Hospital, we like to make sure we are providing the safest care possible. We make sure we are doing this by watching how we care for our patients. If it is alright with you, I would like to sit here in your room for a few minutes and observe anyone who might come in to visit or take care of you. Would that be ok with you? […] Thank you.

When asked by staff what you are doing there:

My name is _____, and I am working with the department of Infection Control and Nursing Quality looking at patients on enhanced contact isolation precautions. I'm looking at both individual practice and room setup. If you have any questions, you can contact ***** in Infectious Diseases (*****) or ****** in Infection Control (*****).

Once we start. ***After an observation of non-compliance:***

For staff: "I am conducting observations for enhanced contact precautions and observed that you did not follow the guidelines (before you entered the patient's room / during patient care / after leaving the patient's room). Please be sure to do this to prevent the spread of infections."

When told "but I didn't touch the patient:"

"Often times, we go into a room without the intention of touching anything, but end up touching a countertop, a monitor, a remote control, a pump, something…it is UW policy and best to adhere to isolation precautions to ensure that we are not putting our patients or ourselves at risk--regardless of whether or not we plan to touch anything in the room."

If you receive a negative response from someone who was non-compliant please contact: ***, MD (for physicians), ****** (for nursing and nursing assistants) or ****** (for any others).***

Fig 1. Script provided to observers to address issues related to patient involvement and staff concerns.

For comparison, monthly historical hand hygiene compliance data were also obtained from the hospital's infection control database for this time period. These data are routinely collected by trained overt observers on all patients, not just those with CDI. Hand hygiene at entry and exit were reported as a composite measure. Compliance

was defined as described earlier, with exit practices liberalized to also include alcohol handrub.

Analyses were performed in R software (R Foundation for Statistical Computing, Vienna, Austria), using the $\chi 2$ test or the t test. This was a quality-improvement study, deemed exempt from review by the University of Wisconsin Health Sciences institutional review board.

Results

A total of 2,889 covert observations were collected from CDI rooms. Among these, visitors had the lowest rates of compliance for all 3 measures: hand hygiene at entry, hand hygiene at exit, and contact precautions (Table 1). Overall hand hygiene compliance among health care workers was 71.6% and 73.7% at entry and exit, respectively. Hand hygiene compliance was comparable between nurses and physicians at entry (70.9% vs 75.0%; $P = .37$) and exit (75.7% vs 71.4%; $P = .17$). Contact precautions compliance was also the same at 80.4% for nurses and 83.7% for physicians ($P = .21$). Overall health care worker compliance with contact precautions was 80.6%.

Table 1. Hand hygiene and contact precautions practices, evaluated by population type and time of observation.

	Hand hygiene at entry; n (%)				Hand hygiene at exit; n (%)					Contact precautions; n (%)						
	Compliant		Non-compliant		Compliant		Non-compliant			Compliant		Non-compliant				
	Alcohol rub	S/W	Neither	*P*-value*	S/W in room	S/W outside room	Alcohol rub	Neither	*P*-value*	Gown tied, gloves	Gown untied, gloves	Gown only, tied	Gown only, untied	Gloves only	None	*P*-value*
Population type				<.001					<.001							<.001
Physician	108 (73.0)	3 (2.0)	37 (25.0)		115 (43.9)	72 (27.5)	24 (9.2)	51 (19.5)		170 (55.6)	86 (28.1)	10 (3.3)	11 (3.6)	11 (3.6)	18 (5.9)	
Nurse	497 (67.8)	23 (3.1)	213 (29.1)		665 (52.5)	293 (23.1)	97 (7.7)	211 (16.7)		1061 (68.6)	182 (11.8)	75 (4.9)	29 (1.9)	93 (6.0)	106 (6.9)	
Visitor	24 (27.9)	10 (11.6)	52 (60.5)		39 (32.0)	8 (6.6)	9 (7.4)	66 (54.1)		143 (21.8)	32 (6.3)	111 (21.8)	40 (7.9)	2 (0.4)	181 (35.6)	
Other healthcare†	94 (69.6)	2 (1.5)	39 (28.9)		108 (46.4)	45 (19.3)	19 (8.2)	61 (26.2)		209 (66.1)	39 (12.3)	16 (5.1)	13 (4.1)	16 (5.1)	23 (7.3)	
Other‡	44 (59.5)	3 (4.1)	27 (36.5)		41 (30.6	17 (12.7)	13 (9.7)	63 (47.0)		102 (57.3)	46 (25.8)	5 (2.8)	9 (5.1)	8 (4.5)	8 (4.5)	
Time of observation				.17					.87							<.001
Day	295 (57.5)	11 (2.1)	207 (40.4)		601 (50.7)	221 (18.6)	60 (5.1)	303 (25.6)		941 (60.2)	242 (15.5)	80 (5.1)	50 (3.2)	67 (4.3)	183 (11.7)	
Night	85 (60.7)	8 (5.7)	47 (33.6)		113 (49.6)	47 (20.6)	24 (10.5)	44 (19.3)		187 (56.3)	32 (10.6)	33 (9.6)	14 (4.2)	20 (6.0)	46 (13.9)	

S/W, soap and water.

*Compliant versus non-compliant across all 5 population types.

†Includes pharmacy, physical therapy, occupational therapy, respiratory therapy, phlebotomy, radiology technicians, and other health care workers.

‡Includes environmental cleaning, transport, food services, and other non-health care workers.

There was no difference in hand hygiene compliance for CDI patients at entry or exit between day and night shifts. However, there was a decrease in contact precaution use at night that remained significant among employee-only data, when visitors were excluded (83.3% vs 74.6%; *P* = .001).

A total of 101,833 hand hygiene observations were obtained by infection control. Compliance from these hospital-wide, overt observations was significantly higher for

nurses than physicians (97.2% vs 90.5%; P < .001). Similar rates were reported for other health care providers (97.1%) and non-health care provider employees (93.0%).

Discussion

There was no difference in hand hygiene or contact precaution compliance between covertly observed physicians and nurses working with CDI patients. However, hospital-wide overt observation data over the same time period showed significantly higher nursing compliance. Although hospital-wide and CDI-specific rates are not directly comparable, hand hygiene compliance at exit is typically higher for patients under contact precautions than the general hospital population.[8,9] Therefore, our CDI-specific hand hygiene rates likely overestimate hospital-wide compliance and underestimate differences between overt and covertly observed measurements at our institution.

The conflicting nature of our in-room and overt observation findings are directly in line with recent systematic reviews of hand hygiene practices, which illustrate a lack of clarity regarding the effect of provider type on compliance.[2] Studies reporting higher nursing compliance typically used overt observations or self-report methodology. However, overt observers overestimate compliance twice as much when evaluating nurses as physicians.[3,4] Therefore, it is not surprising that our institution's nursing and physician hand hygiene rates were more similar among covert than overt observations.

In addition to using distinct CDI-specific and hospital-wide populations for in-room and overt observations, we were also limited by the inability of in-room observers to record outside-room hand hygiene that was not visible from inside. Observers could not doff gown and gloves and exit the patient's room between observations, which resulted in a disproportionate amount of missing data regarding hand hygiene at entry. Considering

that all subjects had access to the same outside-room sink and alcohol dispensers, differential rates of visibility are not expected to have biased this assessment of hand hygiene compliance.

Finally, to our knowledge, this is the first study to evaluate diurnal effects of contact precaution usage. Notably, the decrease in compliance overnight cannot be fully explained by low adherence among visitors, as the association remained significant when visitors were excluded. It is possible that fewer providers working overnight results in increased clinical burdens and perceived lack of time, both of which contribute to low compliance with contact precautions.[10]

Conclusions

Covert observations do not simply replicate the findings of routine, overt observations, but provide a more realistic estimate of compliance with infection control practices. The added burden of covert observation is warranted for CDI patients, given the importance of accurate hand hygiene measurements for this population.[1] The cost of conducting covert observations can be minimized by relying on hospital volunteers and light-duty staff trained in infection control monitoring.

Acknowledgements

The authors acknowledge the work of UW Health volunteers who conducted observations that made this contribution possible.

Funding/support: AKB and ESC are supported by NIH grant MSTP T32 GM008692. AKB was also supported as a pre-doctoral trainee under NIH awards UL1TR000427 and TL1TR000429.

References

1. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). Clin Infect Dis 2018;66:e1–48. [PubMed: 29462280]
2. Erasmus V, Daha TJ, Brug H, Richardus JH, Behrendt MD, Vos MC, et al. Systematic review of studies on compliance with hand hygiene guidelines in hospital care. Infect Control Hosp Epidemiol 2010;31:283–94. [PubMed: 20088678]
3. Wu K-S, Lee SS-J, Chen J-K, Chen YS, Tsai HC, Chen YJ, et al. Identifying heterogeneity in the Hawthorne effect on hand hygiene observation: a cohort study of overtly and covertly observed results. BMC Infect Dis 2018;18:369. [PubMed: 30081843]
4. Kovacs-Litman A, Wong K, Shojania KG, Callery S, Vearncombe M, Leis JA. Do physicians clean their hands? Insights from a covert observational study. J Hosp Med 2016;11:862–4. [PubMed: 27378510]
5. Kwok YLA, Juergens CP, McLaws M-L. Automated hand hygiene auditing with and

without an intervention. Am J Infect Control 2016;44:1475–80. [PubMed: 27776819]

6. Ward MA, Schweizer ML, Polgreen PM, Gupta K, Reisinger HS, Perencevich EN. Automated and electronically assisted hand hygiene monitoring systems: a systematic review. Am J Infect Control 2014;42:472–8. [PubMed: 24773785]
7. Boyce JM. Hand hygiene compliance monitoring: current perspectives from the USA. J Hosp Infect 2008;70:S2–7.
8. Morgan DJ, Pineles L, Shardell M, Graham MM, Mohammadi S, Forrest GN, et al. The effect of contact precautions on healthcare worker activity in acute care hospitals. Infect Control Hosp Epidemiol 2013;34:69–73. [PubMed: 23221195]
9. Swoboda SM, Earsing K, Strauss K, Lane S, Lipsett PA. Isolation status and voice prompts improve hand hygiene. Am J Infect Control 2007;35: 470–6. [PubMed: 17765560]
10. Efstathiou G, Papastavrou E, Raftopoulos V, Merkouris A. Factors influencing nurses' compliance with standard precautions in order to avoid occupational exposure to microorganisms: a focus group study. BMC Nursing 2011;10:1. [PubMed: 21255419]

Appendix H - METABOLIC: High-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks

This appendix has been published in *Microbiome:*

Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., Anantharaman, K. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome.* 10, 33 (2022). https://doi.org/10.1186/s40168-021-01213-8.

All supplemental Materials are available on the publisher's website.

METABOLIC: High-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks

Zhichao Zhou[1], Patricia Q. Tran[1,2], Adam M. Breister[1], Yang Liu[3], Kristopher Kieft[1,4], **Elise S. Cowley**[1,4], Ulas Karaoz[5], Karthik Anantharaman[1]*

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, 53706, USA
[2]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, 53706, USA
[3]Institute for Advanced Study, Shenzhen University, Shenzhen, Guangdong Province, 518060, China
[4]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, 53706, USA
[5]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

*Correspondence to Karthik Anantharaman, karthik@bact.wisc.edu

ABSTRACT

Background: Advances in microbiome science are being driven in large part due to our ability to study and infer microbial ecology from genomes reconstructed from mixed

microbial communities using metagenomics and single-cell genomics. Such omics-based techniques allow us to read genomic blueprints of microorganisms, decipher their functional capacities and activities, and reconstruct their roles in biogeochemical processes. Currently available tools for analyses of genomic data can annotate and depict metabolic functions to some extent, however, no standardized approaches are currently available for the comprehensive characterization of metabolic predictions, metabolite exchanges, microbial interactions, and microbial contributions to biogeochemical cycling.

Results: We present METABOLIC (METabolic And BiogeOchemistry anaLyses In miCrobes), a scalable software to advance microbial ecology and biogeochemistry studies using genomes at the resolution of individual organisms and/or microbial communities. The genome-scale workflow includes annotation of microbial genomes, motif validation of biochemically validated conserved protein residues, metabolic pathway analyses, and calculation of contributions to individual biogeochemical transformations and cycles. The community-scale workflow supplements genome-scale analyses with determination of genome abundance in the microbiome, potential microbial metabolic handoffs and metabolite exchange, reconstruction of functional networks, and determination of microbial contributions to biogeochemical cycles. METABOLIC can take input genomes from isolates, metagenome-assembled genomes, or single-cell genomes. Results are presented in the form of tables for metabolism and a variety of visualizations including biogeochemical cycling potential, representation of sequential metabolic transformations, community-scale microbial functional networks

using a newly defined metric ‘MW-score’ (metabolic weight score), and metabolic Sankey diagrams. METABOLIC takes ~3 hours with 40 CPU threads to process ~100 genomes and corresponding metagenomic reads within which the most compute-demanding part of hmmsearch takes ~45 mins, while it takes ~5 hours to complete hmmsearch for ~3600 genomes. Tests of accuracy, robustness, and consistency suggest METABOLIC provides better performance compared to other software and online servers. To highlight the utility and versatility of METABOLIC, we demonstrate its capabilities on diverse metagenomic datasets from the marine subsurface, terrestrial subsurface, meadow soil, deep sea, freshwater lakes, wastewater, and the human gut.

Conclusion: METABOLIC enables the consistent and reproducible study of microbial community ecology and biogeochemistry using a foundation of genome-informed microbial metabolism, and will advance the integration of uncultivated organisms into metabolic and biogeochemical models. METABOLIC is written in Perl and R and is freely available at https://github.com/AnantharamanLab/METABOLIC under GPLv3.

Keywords: functional traits, metagenome-assembled genomes, microbiome, biogeochemistry, metabolic potential, microbial functional networks.

INTRODUCTION

Metagenomics and single-cell genomics have transformed the field of microbial ecology by revealing a rich diversity of microorganisms from diverse settings, including terrestrial

[1-3] and marine environments [4, 5] and the human body [6]. These approaches can provide an unbiased and insightful view into microorganisms mediating and contributing to biogeochemical activities at a number of scales ranging from individual organisms to communities [7-9]. Recent studies have also enabled the recovery of hundreds to thousands of genomes from a single sample or environment [8, 10, 11]. However, analyses of ever-increasing datasets remain a challenge. For example, there is a lack of scalable and reproducible bioinformatic approaches for characterizing metabolism and biogeochemistry, as well as standardizing their analyses and representation for large datasets.

Microbially-mediated biogeochemical processes serve as important driving forces for the transformation and cycling of elements, energy, and matter among the lithosphere, atmosphere, hydrosphere, and biosphere [12]. Microbial communities in natural environmental settings exist in the form of complex and highly connected networks that share and compete for metabolites [13-15]. The interdependent and cross-linked metabolic and biogeochemical interactions within a community can provide a relatively high level of plasticity and flexibility [16]. For instance, multiple metabolic steps within a specific pathway are often separately distributed in a number of microorganisms and they are interdependent on utilizing the substrates from the previous step [2, 17, 18]. This scenario, referred to as ‘metabolic handoffs’, is based on sequential metabolic transformations, and provides the benefit of high resilience of metabolic activities which make both the community and function stable in the face of perturbations [17, 18]. It is therefore highly

valuable to obtain the information of microbial metabolic function from the perspective of individual genomes as well as the entire microbial community.

Currently, there are many quantitative software and platforms for reconstructing species and community-level metabolic networks [19-25]. They are largely based on building microbial metabolic models containing reactions for substrate utilization and product generation 1[5, 19]. Based on individual microbial models, metabolic phenotypes for the whole community can be further predicted [15]. These approaches allow providing mechanistic bases for predicting and thus operating community metabolisms based on the given environmental conditions and predicted microbial phenotypes [26]. Thus they are more focused on illustrating the operating principles of community metabolisms and the underlying metabolic networks of connected reactions to achieve better outcomes for metabolite production [21, 22], industrial applications [19], drug discovery [19], etc.

Yet, seldom have approaches been developed to study the functional role of microorganisms in the context of biogeochemistry and community-level functional networks [27, 28]. Such tools are based on the principles of facilitating the understanding of microbially-mediated biogeochemical activities. The tools ask for identifying and providing metabolic predictions on the functional details, transformations of nutrients and energy, and functional connections for microorganisms within the community [29]. The resulting genome-informed microbial metabolisms are important for understanding the microbial roles within a whole community in mediating the biogeochemical processes. Currently, such quantitative approaches to interpret functional details,

reconstruct metabolic relationships, and visualize microbial functional networks are still limited [27, 28].

Prediction of microbial metabolism relies on the annotation of protein function for microorganisms using a number of established databases, e.g., KEGG [30], MetaCyc [31], Pfam [32], TIGRfam [33], SEED/RAST [34], and eggNOG [35]. However, these results are often highly detailed, and therefore can be overwhelming to users. Obtaining a functional profile and identifying metabolic pathways in a microbial genome can involve manual inspection of thousands of genes [36]. Organizing, interpreting, and visualizing such datasets remains a challenge and is often untenable especially with datasets larger than one microbial genome. There is a critical need for approaches and tools to identify and validate the presence of metabolic pathways, biogeochemical function, and connections in microbial communities in a user-friendly manner. Such tools addressing this gap would also allow standardization of methods and easier integration of genome-informed metabolism into biogeochemical models, which currently rely primarily on physicochemical data and treat microorganisms as black boxes [37]. A recent statistical study indicates that incorporating microbial community structure in biogeochemical modeling could significantly increase model accuracy of processes that are mediated by narrow phylogenetic guilds via functional gene data, and processes that are mediated by facultative microorganisms via community diversity metrics [38]. This highlights the importance of integrating microbial community and genomic information into the prediction and modeling of biogeochemical processes.

Here we present the software METABOLIC (METabolic And BiogeOchemistry anaLyses In miCrobes), a toolkit to profile metabolic and biogeochemical traits, and functional networks in microbial communities based on microbial genomes. METABOLIC integrates annotation of proteins using KEGG [30], TIGRfam [33], Pfam [32], custom hidden Markov model (HMM) databases [2], dbCAN2 [39], and MEROPS [40], incorporates a protein motif validation step to accurately identify proteins based on prior biochemical validation, and determines presence or absence of metabolic pathways based on KEGG modules. METABOLIC also produces user-friendly outputs in the form of tables and figures including a summary of microbial functional profiles, biogeochemically-relevant pathways, functional networks at the scale of individual genomes and community levels, and microbial contribution to the biogeochemical processes.

METHODS

HMM databases used by METABOLIC

To generate a broad range of metabolic gene HMM profiles, we integrated three sets of HMM-based databases, which are KOfam [41] (July 2019 release, containing HMM profiles for KEGG/KO with predefined score thresholds), TIGRfam [33] (Release 15.0), Pfam [32] (Release 32.0), and custom metabolic HMM profiles [2]. In order to achieve a better HMM search result excluding non-specific hits, we have tested and manually curated cutoffs for those HMM databases listed above into the resulting HMMs: KOfam database - KOfam suggested values; TIGRfam/Pfam/Custom databases - manually curated by adjusting noise cutoffs (NC) or trusted cutoffs (TC) to avoid potential false positive hits. For the KOfam suggested cutoffs, we considered both the score type (full

length or domain) and the score value to assign whether an individual protein hit is significant or not. HMM databases were used as the reference for hmmsearch [42] to find protein hits of input genomes. Prodigal [43] was used to annotate genomic sequences (the method used to find ORFs by Prodigal can be set by METABOLIC as “meta” or “single”), or a user can provide self-annotated proteins (with extensions of “.faa”) to facilitate incorporation into existing pipelines. Methods on the manual curation of these HMM databases are described in the next section.

Curation of cutoff scores for metabolic HMMs

Two curation methods for adjusting NC or TC of TIGRfam/Pfam/Custom databases were used for a specific HMM profile. First, we parsed and downloaded representative protein sequences according to either the corresponding KEGG identifier or UniProt identifier [44]. We then randomly subsampled a small portion of the sequences (10% of the whole collection if this was more than 10 sequences, or at least 10 sequences) as the query to search against the representative protein collections [42]. Subsequently, we obtained a collection of hmmsearch scores by pairwise sequence comparisons. We plotted scores against hmmsearch hits and selected the mean value of the sharpest decreasing interval as the adjusted cutoff (approximately the F1 score). Second, we downloaded a collection of proteins that belong to a specific HMM profile and pre-checked the quality and phylogeny of these proteins by reconstructing and manually inspecting phylogenetic trees. We applied pre-checked protein sequences as the query search against a set of training metagenomes (data not shown). We then obtained a collection of hmmsearch scores of resulting hits from the training metagenomes. By

using a similar method as described above, the cutoff was selected as the mean value of the sharpest decreasing interval.

The following example demonstrates how the method above was used to curate the cutoffs for hydrogenase enzymes. We then expanded this method to all genes using a similar method. We downloaded the individual protein collections for each hydrogenase functional group from the HydDB [45], which included [FeFe] Group A-C series, [Fe] Group, and [NiFe] Group 1-4 series. The individual hydrogenase functional groups were further categorized based on the catalyzing directions, which included H2-evolution, H2-uptake, H2-sensing, electron-bifurcation, and bidirection. To define the NC cutoff ('--cut_nc' in hmmsearch) for individual hydrogenase groups, we used the protein sequences from each hydrogenase group as the query to hmmsearch against the overall hydrogenase collections. By plotting the resulting hmmsearch hit scores against individual hmmsearch hits, we selected the mean value of the sharpest decreasing interval as the cutoff value.

Motif validation

To automatically validate protein hits and avoid false positives, we introduced a motif validation step by comparing protein motifs against a manually curated set of highly conserved residues in important proteins. This manually curated set of highly conserved residues is derived from either reported works or protein alignments from this study. We chose 20 proteins associated with important metabolisms (with a focus on important biogeochemical cycling steps) that are prone to be misannotated into proteins within the

same protein family. Details of these proteins are provided in Additional file 8: Dataset S1. For example, DsrC (sulfite reductase subunit C) and TusE (tRNA 2-thiouridine synthesizing protein E) are similar proteins that are commonly misannotated. Both of them are assigned to the family KO:K11179 in the KEGG database. To avoid assigning TusE as a sulfite reductase, we identified a specific motif for DsrC but not TusE (GPXKXXCXXXGXPXPXXCX", where "X" stands for any amino acid) [46]. We used these specific motifs to filter out proteins that have high sequence similarity but functionally divergent homologs.

Annotation of carbohydrate-active enzymes and peptidases

For carbohydrate-active enzymes (CAZymes), dbCAN2 [39] was used to annotate proteins with default settings. The hmmscan parser and HMM database (2019-09-05 release) were downloaded from the dbCAN2 online repository (http://bcb.unl.edu/dbCAN2/download/) [39]. The non-redundant library of protein sequences which contains all the peptidase/inhibitor units from the peptidase (inhibitor) database MEROPS [40] (known as the 'MEROPS pepunit' database) was used as the reference database to search against putative peptidases and inhibitors using DIAMOND. The settings used for the DIAMOND BLASTP search were "-k 1 -e 1e-10 --query-cover 80 --id 50" [47]. We used the 'MEROPS pepunit' database due to the fact that it only includes the functional unit of peptidases/inhibitors [40] which can effectively avoid potential non-specific hits.

Implementation of METABOLIC-G and METABOLIC-C

To target specific applications in processing omics datasets, we have implemented two versions of METABOLIC: METABOLIC-G (genome version) and METABOLIC-C (community version). METABOLIC-G intakes only genome files and provides analyses for individual genome sequences (including three kinds of genomes, e.g., single-cell genomes, isolate genomes, and metagenome-assembled genomes). The analyzing procedures of METABOLIC-G for all these three kinds of genomes are the same.

METABOLIC-C includes an option for users to include metagenomic reads for mapping to metagenome-assembled genomes (MAGs). Using Bowtie 2 (version ≥ v2.3.4.1) [48], metagenomic BAM files were generated by mapping all input metagenomic reads to gene collections from input genomes. Subsequently, SAMtools (version ≥ v0.1.19) [49], BAMtools (version ≥ v2.4.0) [50], and CoverM (https://github.com/wwood/CoverM) were used to convert BAM files to sorted BAM files and to calculate the gene coverage. To calculate the relative abundance of a specific biogeochemical cycling step, all the coverage of genes that are responsible for this step were summed up and normalized by overall gene coverage. Reads from single-cell and isolate genomes can also be mapped in an identical manner to metagenomes. The gene coverage result generated by metagenomic read mapping was further used in downstream processing steps to conduct community-scale interaction and network analyses.

Classifying microbial genomes into taxonomic groups

To study community-scale interactions and networks of each microbial group within the whole community, we classified microbial genomes into individual taxonomic groups.

GTDB-Tk v0.1.3 [51] was used to assign taxonomy of input genomes with default settings. GTDB-Tk can provide automated and objective taxonomic classification based on the rank-normalized Genome Taxonomy Database (GTDB) taxonomy within which the taxonomy ranks were established by a sophisticated criterion counting the relative evolutionary divergence (RED) and average nucleotide identity (ANI) [51, 52]. Subsequently, genomes were clustered into microbial groups at the phylum level, except for Proteobacteria which were replaced by its subordinate classes due to its wide coverage. Taxonomic assignment information for each genome was used in the downstream community analyses.

Analyses and visualization of metabolic outputs, biogeochemical cycles, MW-scores, functional networks, and metabolic Sankey diagrams

To visualize the outputted metabolic results, the R script "draw_biogeochemical_cycles.R" was used to draw the corresponding metabolic pathways for individual genomes. We integrated HMM profiles that are related to biogeochemical activities and assigned HMM profiles to 31 distinct biogeochemical cycling steps (See details in "METABOLIC_template_and_database" folder on the GitHub page). The script can generate figures showing biogeochemical cycles for individual genomes and the summarized biogeochemical cycle for the whole community. By using the results of metabolic profiling generated from hmmsearch and gene coverage from the mapping of metagenomic reads, we can depict metabolic capacities of both individual genomes and all genomes within a community as a whole. The community-level diagrams, including sequential transformation diagrams, functional

network diagrams, and metabolic Sankey diagrams, were generated using both metabolic profiling and gene coverage results. The diagrams are made by the scripts “draw_sequential_reaction_diagram.R”, “draw_metabolic_Sankey_diagram.R”, and “draw_functional_network_diagram.R”, respectively (For details, refer to GitHub wiki pages).

MW-score (metabolic weight score) is a metric reflecting the functional capacity and abundance of a microbial community in co-sharing functional networks. It was calculated at the community-scale level based on results of metabolic profiling and gene coverage from metagenomic read mapping as described above. We divided metabolic/biogeochemical cycling steps (31 in total) into a finer level – function (51 functions in total) – for better resolution on reflecting functional networks. By using similar methods for determining metabolic interactions (as described above), we selected functions that are shared among genomes. MW-score for each function was calculated by summing up all the coverage values of each function (calculated by summing up all coverage values of genomes that contain this function) and subsequently normalizing it by the overall function coverage. For each function, the contribution percentage of each microbial phylum (the default taxonomic level setting) was also calculated accordingly. One can also change the taxonomic level setting to the resolution of “class”, “order”, “family”, or “genus” to calculate the corresponding contribution percentage of each microbial group. Two equations are provided as follows to calculate each function’s MW-score (1) and the percentage of contribution of each microbial group to the MW-score (2):

$$\mathrm{MW}_{f_\mathrm{i}} = \frac{\sum_{g=g_1}^{g_\mathrm{n}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{i}}}{\sum_{g=g_1, f=f_1}^{g_\mathrm{n}, f_\mathrm{n}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{n}}} \quad (1)$$

$$\mathrm{Cperc}_{f_\mathrm{i} p_\mathrm{j}} = \left(\frac{\sum_{g=g_k}^{g_\mathrm{l}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{i}}}{\sum_{g=g_1, f=f_1}^{g_\mathrm{n}, f_\mathrm{n}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{n}}} \Big/ \frac{\sum_{g=g_1}^{g_\mathrm{n}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{i}}}{\sum_{g=g_1, f=f_1}^{g_\mathrm{n}, f_\mathrm{n}} \mathrm{C}_{g_\mathrm{n}} \cdot \mathrm{S}_{f_\mathrm{n}}} \right) \times 100\% \quad (2)$$

within which $g_k \ldots g_\mathrm{l} \in p_\mathrm{j}$

In equation (1), MW refers to MW-score. fi refers to the studied function (f) which ranks in the (i) position amongst all functions. g1 and gn indicate the first and the last genome amongst all genomes. f1 and fn indicate the first and the last function amongst all functions. Cg means the coverage of a genome and Sf means the presence (denoted as 1) or absence (denoted as 0) state of a function within that genome. In equation (2), Cprec refers to the contribution percentage of a microbial group to the MW-score. pj means the studied group (p) which ranks in the (j) position amongst all groups. gk and gl indicate the genomes which rank in the (k) position and the (l) position amongst all genomes; the additional note g_k…g_l ∈p_j indicates all the genomes between these two belong to the studied group pj.

Example of METABOLIC analysis

An example of community-scale analyses including element biogeochemical cycling and sequential reaction analyses, functional network and metabolic Sankey visualization, and MW-score calculation were conducted using a metagenomic dataset of microbial community inhabiting deep-sea hydrothermal vent environment of Guaymas Basin in the Pacific Ocean [53]. It contains 98 MAGs and 1 set of metagenomic reads (genomes were available at NCBI BioProject PRJNA522654 and metagenomic reads were deposited to NCBI SRA with accession as SRR3577362).

A metagenomic-based study of the microbial community from an aquifer adjacent to Colorado River, located near Rifle, has provided an accurate reconstruction of the metabolism and ecological roles of the microbial majority [2]. From underground water and sediments of the terrestrial subsurface at Rifle, 2545 reconstructed MAGs were obtained (genomes are under NCBI BioProject PRJNA288027). They were used as the in silico dataset to test METABOLIC's performance. First, all the microbial genomes were dereplicated by dRep v2.0.5 [54] to pick the representative genomes for downstream analysis using the setting of '-comp 85'. Then, METABOLIC-G was applied to profile the functional traits of these representative genomes using default settings. Finally, the metabolic profile chart was depicted by assigning functional traits to GTDB taxonomy-clustered genome groups.

Test on software performance for different environments

To benchmark and test the performance of METABOLIC in different environments, eight datasets of metagenomes and metagenomic reads from marine, terrestrial, and human environments were used. These included marine subsurface sediments [55] (Deep biosphere beneath Hydrate Ridge offshore Oregon), freshwater lake [56] (Lake Tanganyika, eastern Africa), colorectal cancer (CRC) patient gut [57], healthy human gut [57], deep-sea hydrothermal vent [53] (Guaymas Basin, Gulf of California), terrestrial subsurface sediments and water [2] (Rifle, CO, USA), meadow soils [58] (Angelo Coastal Range Reserve, CA, USA), and advanced water treatment facility [59] (Groundwater Replenishment System, Orange County, CA, USA). Default settings were used for running METABOLIC-C.

Comparison of community-scale metabolism

To compare the metabolic profile of two environments at the community scale, MW-score was used as the benchmarker. Two sets of environment pairs were compared, including the pair of marine subsurface sediments [55] and terrestrial subsurface sediments [2] and the pair of freshwater lake [56] and deep-sea hydrothermal vent [53]. To demonstrate differences between these environments in specific biogeochemical processes, we focused on the biogeochemical cycling of sulfur. The sulfur biogeochemical cycling diagrams were depicted with the annotation of the number and the coverage of genomes that contain each biogeochemical cycling step.

Metabolism in human microbiomes

To inspect the metabolism of microorganisms in the human microbiome (associated with skin, oral mucosa, conjunctiva, gastrointestinal tracts, etc.), a subset of KOfam HMMs (139 HMM profiles) were used as markers to depict the human microbiome metabolism (parsed by HuMiChip targeted functional gene families [60]). They included 10 function categories as follows: amino acid metabolism, carbohydrate metabolism, energy metabolism, glycan biosynthesis and metabolism, lipid metabolism, metabolism of cofactors and vitamins, metabolism of other amino acids, metabolism of terpenoids and polyketides, nucleotide metabolism, and translation. The CRC and healthy human gut (healthy control) sample datasets were used as the input (Accession IDs: BioProject PRJEB7774 Sample 31874 and Sample 532796). Heatmap of presence/absence of these functions were depicted by R package “pheatmap” [61] with 189 horizontal entries

(there are duplications of HMM profiles among function categories; for detailed human microbiome metabolism markers, refer to Additional file 9: Dataset S2).

Representation of microbial cell metabolism

To provide a schematic representation of the metabolism of microbial cells, two microbial genomes were used as examples, Hadesarchaea archaeon 1244-C3-H4-B1 and Nitrospirae bacteria M_DeepCast_50m_m2_151. METABOLIC-G results of these two genomes, including functional traits and KEGG modules, were used to draw the cell metabolism diagrams.

Metatranscriptome analysis by METABOLIC

METABOLIC-C can take metatranscriptomic reads as input into transcript coverage calculation and integrate the result into downstream community analyses. METABOLIC-C uses a similar method to that of gene coverage calculation, including mapping transcriptomic reads to the gene collection from input genomes, converting BAM files to sorted BAM files, and calculating the transcript coverage. The raw transcript coverage was further normalized by the gene length and metatranscriptomic read number in Reads Per Kilobase of transcript, per Million mapped reads (RPKM). Hydrothermal vent and background seawater transcriptomic reads from Guaymas Basin (NCBI SRA accessions: SRR452448 and SRR453184) were used to test the outcome of metatranscriptome analysis.

RESULTS

Given the ever-increasing number of microbial genomes from microbiome studies, we developed METABOLIC to enable the metabolic pathway analysis and the visualization of biogeochemical cycles and community-scale functional networks. METABOLIC has an improved methodology to get fast, accurate, and robust annotation results, and it integrates a variety of visualization functions for better interpreting the community-level functional interactions and microbial contributions. While METABOLIC relies on microbial genomes and metagenomic reads for underpinning its analyses for community-level functional interactions, it can easily integrate transcriptomic datasets to provide an activity-based measure of community networks. The scalable capacity, wide utility, and compatibility for analyzing datasets from various environments make it a well-tailored tool for metabolic profiling of large sets of genomes. In the following sections, the microbial community consisting of 98 MAGs from a deep-sea hydrothermal vent was used as the input dataset if not mentioned otherwise.

Workflow to determine the presence of metabolic pathways

METABOLIC is written in Perl and R and is expected to run on Unix, Linux, or macOS. The prerequisites are described on METABOLIC's GitHub wiki pages (https://github.com/AnantharamanLab/METABOLIC/wiki). The input folder requires microbial genome sequences in FASTA format and an optional set of genomic/metagenomic reads which were used to reconstruct those genomes (Figure 1). The annotated proteins from input genomic sequences are queried against HMM databases (KEGG KOfam, Pfam, TIGRfam, and custom HMMs) using hmmsearch implemented within HMMER [42] which applies methods to detect remote homologs as

sensitively and efficiently as possible. After the hmmsearch step, METABOLIC subsequently validates the primary outputs by a motif-checking step for a subset of protein families; only those protein hits which successfully pass this step are regarded as positive hits.

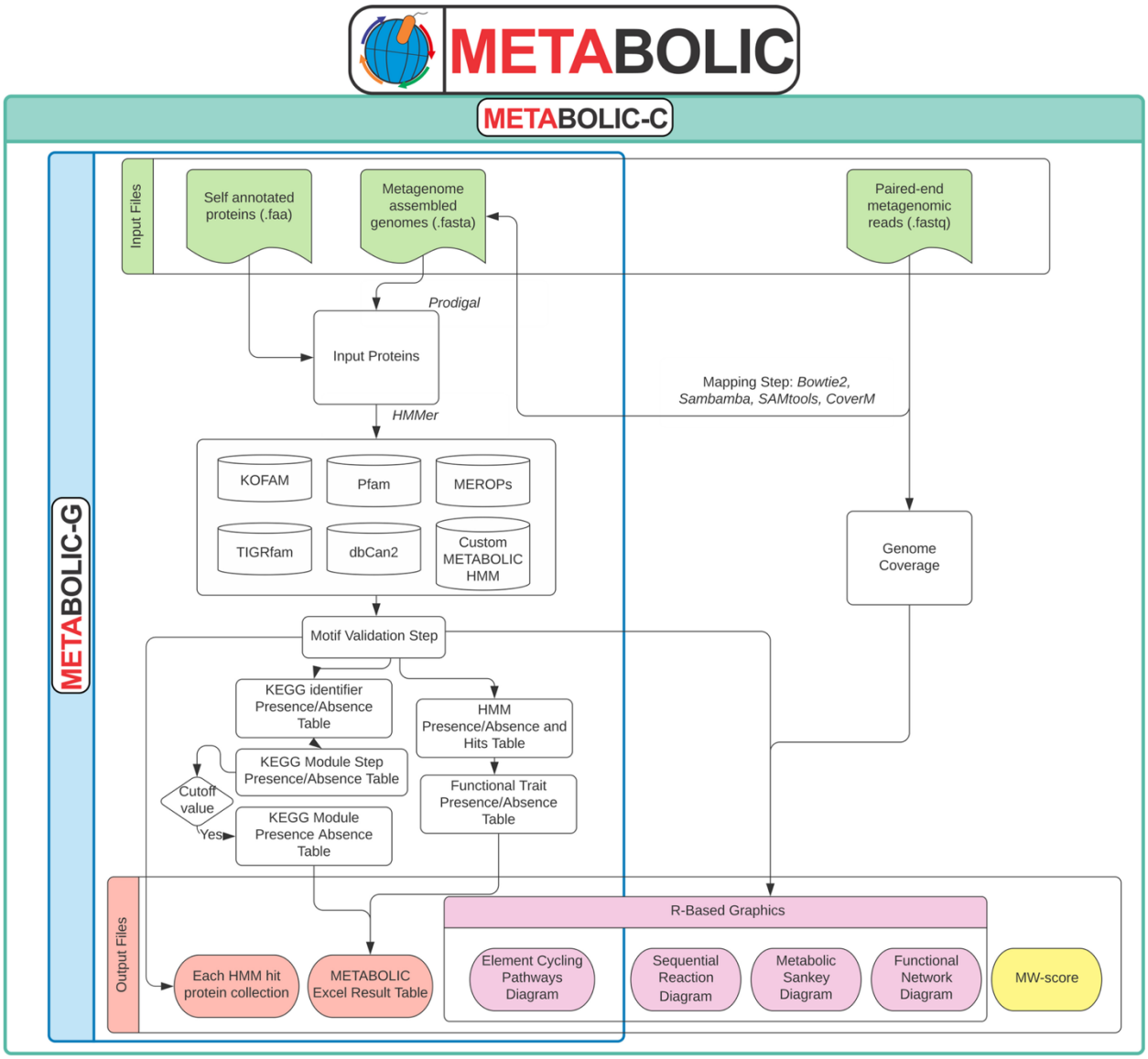


Figure 1. An outline of the workflow of METABOLIC. Detailed instructions are available at https://github.com/AnantharamanLab/METABOLIC/wiki. METABOLIC-G workflow is specifically shown in the blue box and METABOLC-C workflow is shown in the green square.

METABOLIC relies on matches to the above databases to infer the presence of specific metabolic pathways in microbial genomes. Individual KEGG annotations are inferred in the context of KEGG modules for a better interpretation of metabolic pathways. A KEGG module is comprised of multiple steps with each step representing a distinct metabolic function. We parsed the KEGG module database [62] to link the existing relationship of KO identifiers to KEGG module identifiers to project our KEGG annotation result into the interactive network which was constructed by individual building blocks – modules – for better representation of metabolic blueprints of input genomes. In most cases, we used KOfam HMM profiles for KEGG module assignments. For a specific set of important metabolic marker proteins and commonly misannotated proteins, we also applied the TIGRfam/Pfam/custom HMM profiles and motif-validation steps. The software has customizable settings for increasing or decreasing the priority of specific databases, primarily meant to increase annotation confidence by preferentially using custom HMM databases over KEGG KOfam when both targeting the same set of proteins.

Since individual genomes from metagenomes and single-cell genomes can often have incomplete metabolic pathways due to their low completeness compared to isolate genomes, we provide an option to determine the completeness of a metabolic pathway (or a module here). A user-defined cutoff is used to set the threshold of completeness for a given module to be assigned as present (the default cutoff is the presence of 75% of metabolic steps/genes within a given module), which is then used to produce a KEGG module presence/absence table. All modules exceeding the cutoff value are

determined to be present. Meanwhile, the presence/absence information for each module step is also summarized in an overall output table to facilitate further detailed investigations.

Outputs consist of six different results that are reported in an Excel spreadsheet (Additional file 1: Figure S1). These contain details of protein hits (Additional file 1: Figure S1A) which include both presence/absence and protein names, presence/absence of functional traits (Additional file 1: Figure S1B), presence/absence of KEGG modules (Additional file 1: Figure S1C), presence/absence of KEGG module steps (Additional file 1: Figure S1D), carbohydrate-active enzyme (CAZyme) hits (Additional file 1: Figure S1E) and peptidase/inhibitor hits (Additional file 1: Figure S1F). For each HMM profile, the protein hits from all input genomes can be used to construct phylogenetic trees or further be combined with reference protein collections for detailed evolutionary analyses.

Quantitative visualization of biogeochemical cycles and sequential reactions

After METABOLIC generates protein and pathway annotation results, the software further identifies and highlights specific pathways of importance in microbiomes associated with energy metabolism and biogeochemistry. To visualize pathways of biogeochemical importance, it generates schematic profiles for nitrogen, carbon, sulfur, and other elemental cycles for each genome. The set of genomes used as input is considered the “community”, and each genome within is considered an “organism”. A summary schematic diagram at the community level integrates results from all individual

genomes within a given dataset (Figure 2) and includes computed abundances for each step in a biogeochemical cycle if the genomic/metagenomic read datasets are provided. The genome number labeled in the figure indicates the number/quantity of genomes that contain the specific gene components of a biogeochemical cycling step (Figure 2) [2]. In other words, it represents the number of organisms within a given community inferred to be able to perform a given metabolic or biogeochemical transformation. The abundance percentage indicates the relative abundance of microbial genomes that contain the specific gene components of a biogeochemical cycling step among all microbial genomes in a given community (Figure 2) [2].

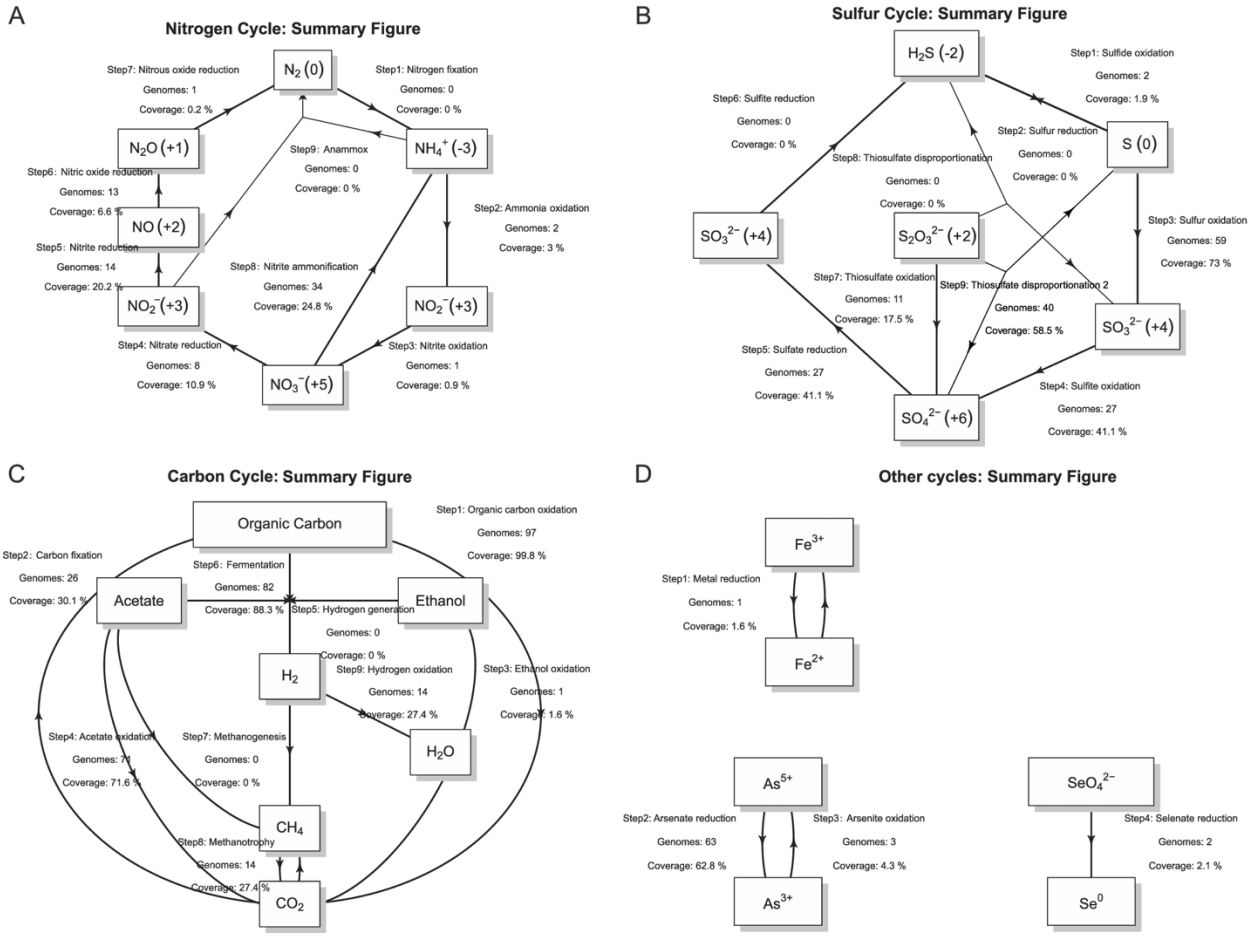


Figure 2. Summary scheme of biogeochemical cycling processes at the community scale. Each arrow represents a single transformation/step within a cycle.

Labels above each arrow are (from top to bottom): step number and reaction, number of genomes that can conduct these reactions, metagenomic coverage of genomes (represented as a percentage within the community) that can conduct these reactions. The numbers in brackets next to the nitrogen or sulfur-containing compounds are chemical states of the nitrogen or sulfur atoms in these compounds.

Microorganisms in nature often do not encode pathways for the complete transformation of compounds. For example, microorganisms possess partial pathways for denitrification that can release intermediate compounds like nitrite, nitric oxide, and nitrous oxide in lieu of nitrogen gas which is produced by complete denitrification [63]. A greater energy yield could be achieved if one microorganism conducts all steps associated with a pathway (such as denitrification) [2] since it could fully use all available energy from the reaction. However, in reality, few organisms in microbial communities carry out multiple steps in complex pathways; organisms commonly rely on other members of microbial communities to conduct sequential reactions in pathways [2, 64, 65]. Thus, to study this metabolic scenario in microbial communities, METABOLIC summarizes and enables visualization of the genome number and coverage (relative abundance) of microorganisms that are putatively involved in the sequential transformation of both important inorganic and organic compounds (Figure 3). This provides a quantitative calculation of microbial interactions and connections using shared metabolites associated with inorganic and organic transformations. Additionally, it shows the intuitive pattern of quantity and abundance of microorganisms that are able to conduct partial or all steps for a given pathway, which potentially reflects the degree of resilience of a microbial community.

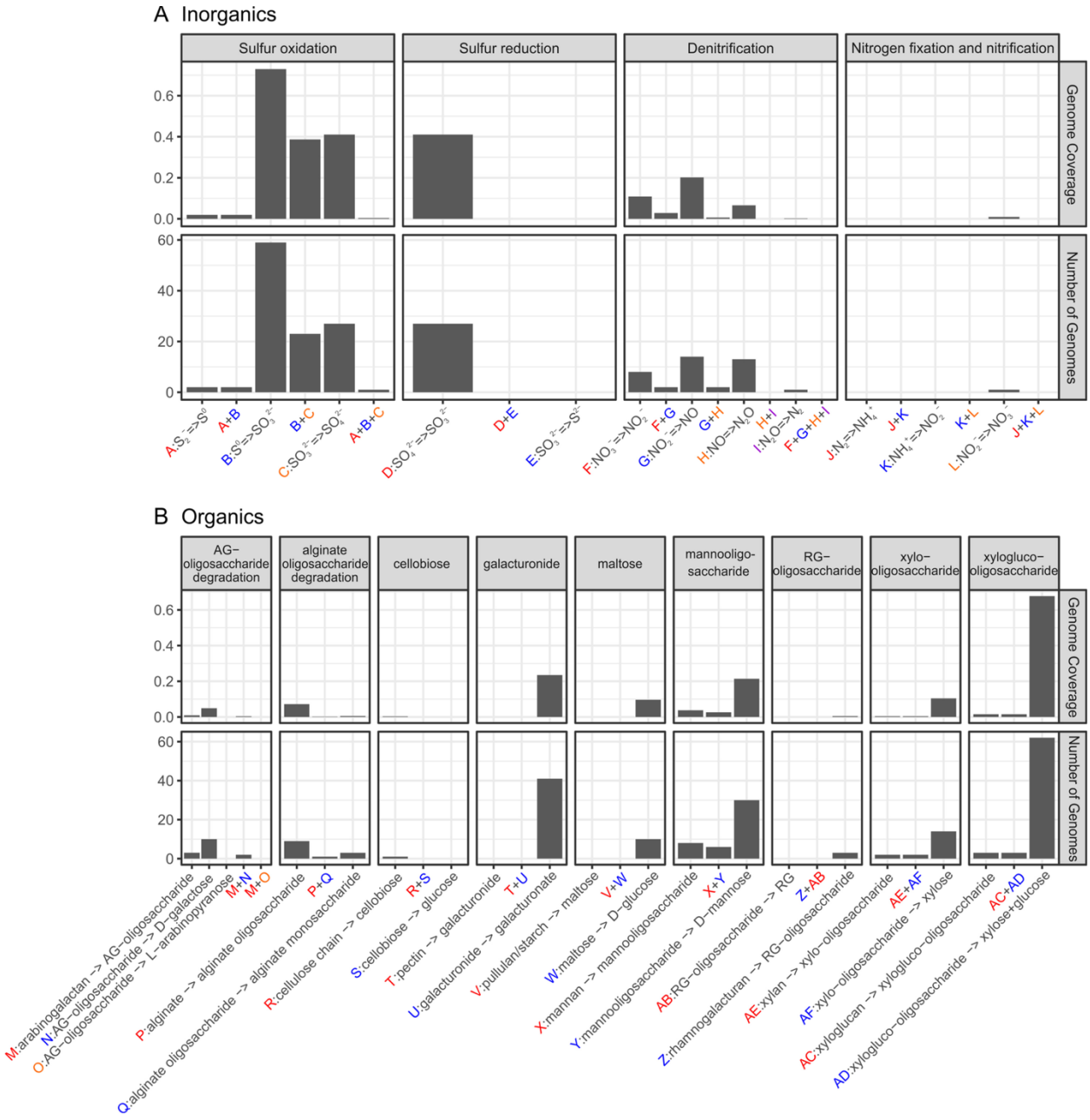


Figure 3. Schematic figure of sequential metabolic transformations. (A) the sequential transformation of inorganic compounds; (B) the sequential transformation of organic compounds. X-axes describe individual sequential transformations indicated by letters. The two panels describe the number of genomes and genome coverage (represented as a percentage within the community) of organisms that are involved in certain sequential metabolic transformations. The deep-sea hydrothermal vent dataset was used for these analyses.

Calculation and visualization of functional networks, metabolic weight scores (MW-scores), and microbial contribution to metabolic reactions

Given the microbial pathway abundance information generated by METABOLIC, we identified co-existing metabolisms in microbial genomes as a measure of connections between different metabolic functions and biogeochemical steps. In the context of biogeochemistry, this approach allows the evaluation of relatedness among biogeochemical steps and the connection contribution by microorganisms. This is enabled at the resolution of individual microbial groups based on the phylogenetic classification (Figure 4) assigned by GTDB-Tk [51]. As an example, we have demonstrated this approach on a microbial community inhabiting deep-sea hydrothermal vents. We divided the microbial community of deep-sea hydrothermal vents into 18 phylum-level groups (except for Proteobacteria which were divided into their subordinate classes). The functional network diagrams were depicted at the resolution of both individual phyla and the entire community level (Additional file 10: Dataset S3). Figure 4 demonstrates metabolic connections that were represented with individual metabolic/biogeochemical cycling steps depicted as nodes, and the connections between two given nodes depicted as edges. The size of a given node is proportional to the degree (number of connections to each node). The thickness of a given edge was depicted based on the average of gene coverage values of two biogeochemical cycling steps (the connected nodes). More edges connecting two nodes represent more connections between these two steps. The color of the edge corresponds to the taxonomic group. At the whole community level, more abundant microbial groups were more represented in the diagram (Figure 4). Overall,

METABOLIC provides a comprehensive approach to construct and visualize functional networks associated with important pathways of energy metabolism and biogeochemical cycles in microbial communities and ecosystems.

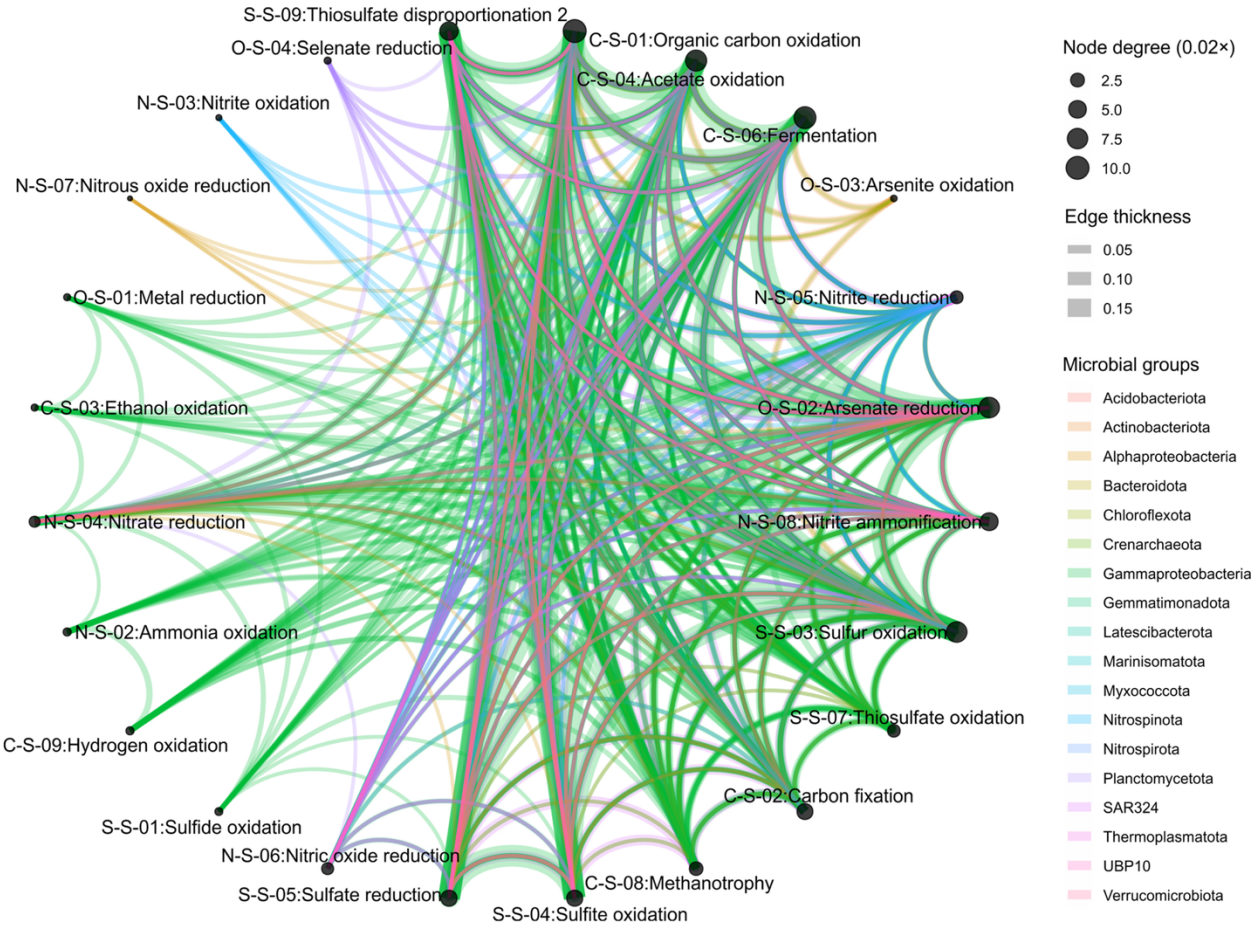


Figure 4. Functional network showing connections between different functions in the microbial community. Nodes represent individual steps in biogeochemical cycles; edges connecting two given nodes represent the functional connections between nodes, which are enabled by organisms that can conduct both biogeochemical processes/steps. The size of the node was depicted according to the degree (number of connections to each node). The thickness of the edge was depicted according to the average gene coverage values of the two connected biogeochemical cycling steps – for example, thiosulfate oxidation and organic carbon oxidation. The color of the edges was assigned based on the taxonomy of the represented genome. The deep-sea hydrothermal vent dataset was used for these analyses.

To address the lack of quantitative and reproducible measures to represent potential metabolic interactions in microbial communities, we developed a new metric that we termed MW-score (metabolic weight scores) (Equations 1 and 2). MW-scores quantitatively measure “function weights” within a microbial community as reflected by the metabolic profile and gene coverage. As metabolic potential for the whole community was profiled into individual functions that either mediated specific pathways or transformed certain substrates into products, a function weight that reflects the abundance fraction for each function can be used to represent the overall metabolic potential of the community. MW-scores resolved the functional capacity and abundance in the co-sharing functional networks as studied and visualized in the above section. More frequently shared functions and their higher abundances lead to higher MW-scores, which quantitatively reflects the function weights in functional networks (Figure 5). MW-score reflects the same functional networking pattern as the above description on the edges (networking lines) connecting the nodes (metabolic steps) that – more edges connecting two nodes indicates two steps are more co-shared, thicker edges indicate higher gene abundance for the metabolic steps. The MW-scores can integratively represent these two networking patterns and serve as metrics to measure these function weights. At the same time, we also calculated each microbial group’s (phylum in this case) contribution to the MW-score of a specific function within the community (Figure 5). A higher microbial group contribution percentage value indicates that one function is more represented by the microbial group (for both gene presence and abundance) in the functional networks. MW-scores provide a quantitative measure

of comparing function weights and microbial group contributions within functional networks.

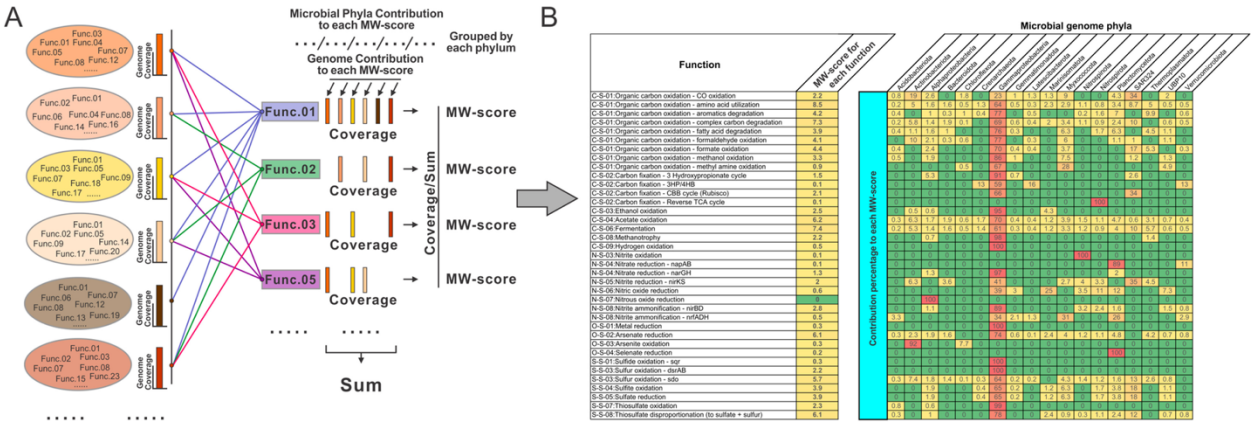


Figure 5. Description, calculation, and result table of MW-scores. (A) The calculation method for the MW-score within a community based on a given metagenomic dataset. Each circle stands for a genome within the community, and the adjacent bar stands for its genome coverage within the community. The coverage values of encoded genes for all functions were summed up as the denominator, and the coverage value of encoded genes for each function was used as the numerator, and the MW-score was calculated accordingly for each function. (B) The resulting table of MW-score for the deep-sea hydrothermal vent metagenomic dataset. MW-score for each function was given in a separated column, and the rest of the table indicates the contribution percentage to each MW-score of the genomes grouped in each phylum. The MW-score of "N-S-07:Nitrous oxide reduction" was not exactly 0 but rounded to 0 due to the original number being less than 0.05. Additionally, contribution percentages were also rounded to only retain one digit after the decimal points; consequently, the sum contribution percentages for some functions slightly deviate from 100%.

To understand the contributions of microbial groups associated with specific metabolic and biogeochemical transformations, we developed an approach to visualize the connections among specific taxonomic groups, metabolic reactions, and entire biogeochemical cycles such as carbon, nitrogen, and sulfur cycles. Our approach involves the use of Sankey diagrams (also called 'Alluvial' plots) to represent the fractions of metabolic functions that are contributed by various microbial groups in a given community (Figure 6). It allows visualization of metabolic reactions as the link

between microbial contributors clustered as taxonomic groups and biogeochemical cycles at a community level (Figure 6 and Additional file 10: Dataset S3). The function fraction was calculated by accumulating the genome coverage values of genomes from a specific microbial group that possesses a given functional trait. The width of curved lines from a specific microbial group to a given functional trait indicates their corresponding proportional contribution to a specific metabolism (Figure 6). Alternatively, the genomic/metagenomic datasets which are used in constructing the above two diagrams: functional network diagram (Figure 4) and metabolic Sankey diagram (Figure 6), can be replaced by transcriptomic/metatranscriptomic datasets, and correspondingly, the gene coverage values will be replaced by gene expression values, and therefore, diagrams will represent the transcriptional activity patterns of functional network and microbial contribution to metabolic reactions (Additional file 2, 3, 4, and 5: Figure S2, S3, S4, and S5).

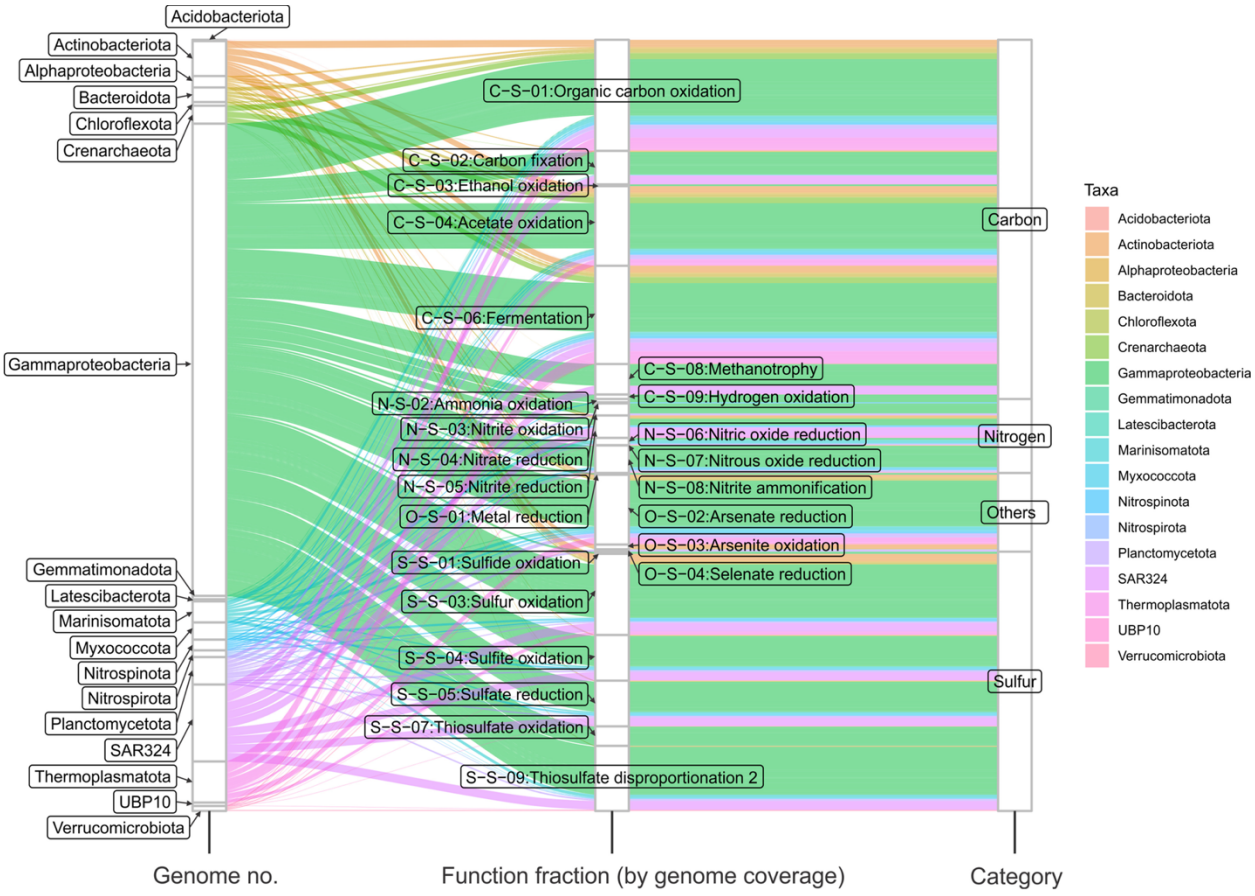


Figure 6. Metabolic Sankey diagram representing the contributions of microbial genomes to individual metabolic and biogeochemical processes and entire elemental cycles. Microbial genomes are represented at the phylum-level resolution. The three columns from left to right represent taxonomic groups scaled by the number of genomes, the contribution to each metabolic function by microbial groups calculated based on genome coverage, and the contribution to each functional category/biogeochemical cycle. The colors were assigned based on the taxonomy of the microbial groups. The deep-sea hydrothermal vent dataset was used for these analyses.

To demonstrate this part of the workflow in reality, the microbial community consisting of 98 MAGs from a deep-sea hydrothermal vent was used as a test dataset. After running the bioinformatic analyses described above, resulting tables and diagrams were compiled and visualized accordingly (Figure 4, 5, 6 and Additional file 10: Dataset S3). Results for functional networks and MW-scores of the deep-sea hydrothermal vent

environment indicate that the microbial community depends on mixotrophy and sulfur oxidation for energy conservation and involves arsenate reduction potentially responsible for detoxification/arsenate resistance [66]. MW-scores indicate that amino acid utilization, complex carbon degradation, acetate oxidation, and fermentation are the major heterotrophic metabolisms for this environment; CO2-fixation and sulfur oxidation also occupy a considerable functional fraction, which indicates heterotrophy and autotrophy both contribute to energy conservation (Figure 5). As represented by both MW-scores and metabolic Sankey diagram, Gammaproteobacteria are the most numerically abundant group in the community and they occupy significant functional fractions amongst both heterotrophic and autotrophic metabolisms (MW-score contribution ranging from 59-100%) (Figure 5, 6), which is consistent with previous findings in the Guaymas Basin hydrothermal environment [53, 67]. Meanwhile, MW-scores also explicitly reflect the involvement of other minor electron donors in energy conservation which are mainly contributed by Gammaproteobacteria, such as hydrogen and methane (Figure 5). This is also consistent with previous findings [53, 67] and indicates the accuracy and sensitivity of MW-scores to reflect metabolic potentials.

METABOLIC performance demonstration

To test METABOLIC's performance on speed, we applied the software (METABOLIC-C mode) to analyze the metagenomic dataset which includes 98 MAGs from a deep-sea hydrothermal vent, and two sets of metagenomic reads (that are subsets of original reads with 10 million reads for each pair comprising ~10% of the total reads). The total running time was ~3 hours using 40 CPU threads in a Linux version 4.15.0-48-generic

server (Ubuntu v5.4.0). The most compute-demanding step is hmmsearch, which took ~45 mins. When tested on another dataset comprising ~3600 microbial genomes (data not shown), METABOLIC could complete hmmsearch in ~5 hours by using 40 CPU threads, indicating its scalable capability on analyzing thousands of genomes.

In order to test the accuracy of the results predicted by METABOLIC, we picked 15 bacterial and archaeal genomes from Chloroflexi, Thaumarchaeota, and Crenarchaeota which are reported to have 3 hydroxypropionate cycle (3HP) and/or 3-hydroxypropionate/4-hydroxybutyrate cycle (3HP/4HB) for carbon fixation. METABOLIC predicted results in line with annotations from the KEGG genome database which can be visualized in KEGG Mapper (Table 1). Our predictions are also in accord with biochemical evidence of the existence of corresponding carbon fixation pathways in each microbial group: 1) 3 out of 5 Chloroflexi genomes are predicted by both METABOLIC and KEGG to possess the 3HP pathway and none of all these Chloroflexi genomes are predicted to possess the 3HP/4HB pathway. This is consistent with current reports based on biochemical and molecular experiments that only organisms from the phylum Chloroflexi are known to possess the 3HP pathway [68] (Table 1). 2) All 5 Thaumarchaeota genomes and 2 out of 5 Crenarchaeota genomes are predicted by both METABOLIC and KEGG to possess the 3HP/4HB pathway and none of these Thaumarchaeota and Crenarchaeota genomes are predicted to possess the 3HP pathway. This is consistent with current reports that only the 3HP/4HB pathway could be detected in Crenarchaeota and Thaumarchaeota[69, 70] (Table 1). We have also applied METABOLIC on a large well-studied dataset comprising 2545 metagenome-

assembled genomes from terrestrial subsurface sediments and groundwater [2]. The annotation results of METABOLIC are consistent with previously described reports (Additional file 6, 10: Figure S6, Dataset S3). These results suggest that METABOLIC can provide accurate annotations and perform well as a functional predictor for microbial genomes and communities.

Table 1. The carbon fixation metabolic traits of 15 tested bacterial and archaeal genomes predicted by both METABOLIC and KEGG genome database.

				METABOLIC result		KEGG genome pathway	
				Carbon fixation		Carbon fixation	
Accession ID	Organism	KEGG Organism Code	Group	3HP cycle	3HP/4HB cycle	3HP cycle	3HP/4HB cycle
GCA_000011905.1	*Dehalococcoides mccartyi* 195	det	Chloroflexi	Absent	Absent	Absent	Absent
GCA_000017805.1	*Roseiflexus castenholzii* DSM 13941	rca	Chloroflexi	Present	Absent	Present	Absent
GCA_000018865.1	*Chloroflexus aurantiacus* J-10-fl	cau	Chloroflexi	Present	Absent	Present	Absent
GCA_000021685.1	*Thermomicrobium roseum* DSM 5159	tro	Chloroflexi	Absent	Absent	Absent	Absent
GCA_000021945.1	*Chloroflexus aggregans* DSM 9485	cag	Chloroflexi	Present	Absent	Present	Absent
GCA_000299395.1	*Nitrosopumilus sediminis* AR2	nir	Thaumarchaeota	Absent	Present	Absent	Present
GCA_000698785.1	*Nitrososphaera viennensis* EN76	nvn	Thaumarchaeota	Absent	Present	Absent	Present
GCA_000875775.1	*Nitrosopumilus piranensis* D3C	nid	Thaumarchaeota	Absent	Present	Absent	Present

GCA_000812185.1	*Nitrosopelagicus brevis* CN25	nbv	Thaumarchaeota	Absent	Present	Absent	Present
GCA_900696045.1	*Nitrosocosmicus franklandus* NFRAN1	nfn	Thaumarchaeota	Absent	Present	Absent	Present
GCA_000015145.1	*Hyperthermus butylicus* DSM 5456	hbu	Crenarchaeota	Absent	Absent	Absent	Absent
GCA_000017945.1	*Caldisphaera lagunensis* DSM 15908	clg	Crenarchaeota	Absent	Present	Absent	Present
GCA_000148385.1	*Vulcanisaeta distributa* DSM 14429	vdi	Crenarchaeota	Absent	Absent	Absent	Absent
GCA_000193375.1	*Thermoproteus uzoniensis* 768-20	tuz	Crenarchaeota	Absent	Present	Absent	Present
GCA_003431325.1	*Acidilobus* sp. 7A	acia	Crenarchaeota	Absent	Absent	Absent	Absent

Currently, several software packages and online servers are available for genome annotation and metabolic profiling. Comparing to other software/online servers including GhostKOALA [71], BlastKOALA [71], KAAS [72], RAST/SEED [34], and eggNOG-mapper [73], METABOLIC is unique in its ability to integrate multi-omic information towards elucidating and visualizing community-level functional connections and the contribution of microorganisms to biogeochemical cycles (Figure 7A). Additionally, in order to compare the prediction performance of METABOLIC to others, we conducted parallel in silico experiments (Figure 7B). We used two representative bacterial genomes as the test datasets. We randomly picked 100 protein sequences from individual genomes and submitted them to annotation by these six software/online servers. Predicted protein

annotations by individual software and online servers were compared to their original annotations that were provided by the NCBI database (Additional file 11, 12: Dataset S4, S5). According to statistical methods of evaluating binary classification [74], the following parameters were used to make the comparison: 1) recall (also referred to as the sensitivity) as the true positive rate, 2) precision (also referred to as the positive predictive value) which indicates the reproducibility and repeatability of a measurement system, 3) accuracy which indicates the closeness of measurements to their true values, and 4) F1 value which is the harmonic mean of precision and recall, and reflects both these two parameters. Among the tested software/online servers, the performance parameters of METABOLIC consistently placed it as the top 3 and top 2 software for recall and F1 and the top 1 and top 2 software for precision and accuracy. These results demonstrate that METABOLIC (Figure 7B) provides robust performance and consistent metabolic prediction that facilitate accurate and reliable applicability for downstream data visualization and community-level analyses.

A

Software/Tool	METABOLIC	KAAS	GhostKOALA	BlastKOALA	RAST	eggNOG-mapper
Online/Standalone	Standalone	Online	Standalone	Standalone	Both	Both
Genome sequence input	●	●			●	●
Protein input	●	●	●	●		●
Genome/Metagenome input	●	●	●	●	●	●
Genome metabolic profile	●	●	●	●	●	●
Genome metabolic profile visualization	●	●	●	●		
Sequential reaction analysis	●					
Community-level analysis	●					
Metagenome coverage analysis	●					

B

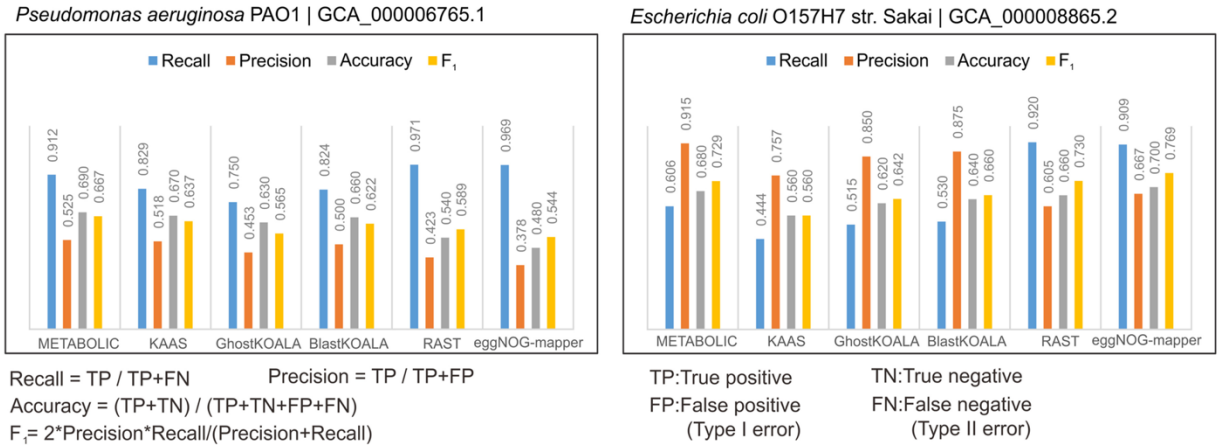


Figure 7. Comparison of METABOLIC with other software packages and online servers. (A) Comparison of the workflows and services, (B) Comparison of performance of protein prediction for two representative genomes, Pseudomonas aeruginosa PAO1, and Escherichia coli O157H7 str. sakai.

To demonstrate the application and performance of METABOLIC in different samples, we tested eight distinct environments (marine subsurface, terrestrial subsurface, deep-sea hydrothermal vent, freshwater lake, gut microbiome from patients with colorectal cancer, gut microbiome from healthy control, meadow soil, wastewater treatment facility). Overall, we found METABOLIC to perform well across all the environments to profile microbial genomes with functional traits and biogeochemical cycles (Additional file 10: Dataset S3). Among these tested environments, we also performed community-scale metabolic comparisons based on the MW-score (Figure 8). MW-score fraction at the community scale reflects the overall metabolic profile distribution pattern.

Specifically, we compared samples from terrestrial and marine subsurface and samples from hydrothermal vent and freshwater lake. We observed that terrestrial subsurface contains more abundant metabolic functions related to nitrogen cycling compared to the marine subsurface (Figure 8A), consistent with the previous characterization of these two environments [2, 75]. Deep-sea hydrothermal vent samples had a considerably high concentration of methane and hydrogen [53] as compared to Lake Tanganyika (freshwater lake). Consistent with this phenomenon, the deep-sea hydrothermal vent microbial community has more abundant metabolic functions associated with methanotrophy and hydrogen oxidation (Figure 8B). In order to focus on a specific biogeochemical cycle, we applied METABOLIC to compare sulfur-related metabolisms at the community scale for these two environment pairs (Additional file 7: Figure S7). Terrestrial subsurface contains genomes covering more sulfur cycling steps compared to marine subsurface (7 steps vs 3 steps) (Additional file 7: Figure S7A). Freshwater lake contains genomes involving almost all the sulfur cycling steps except for sulfur reduction, while deep-sea hydrothermal vent contains less sulfur cycling steps (8 steps vs 6 steps) (Additional file 7: Figure S7B). Nevertheless, deep-sea hydrothermal vent has a higher fraction of genomes (59/98) and a higher relative abundance (73%) of these genomes involving sulfur oxidation compared to the freshwater lake (Additional file 7: Figure S7B). This indicates that the deep-sea hydrothermal vent microbial community has a more biased sulfur metabolism towards sulfur oxidation, which is consistent with previous metabolic characterization on the dependency of elemental sulfur in this environment [53, 76-78]. Collectively, by characterizing community-scale

metabolism, METABOLIC can facilitate the comparison of overall functional profiles as well as for a particular elemental cycle.

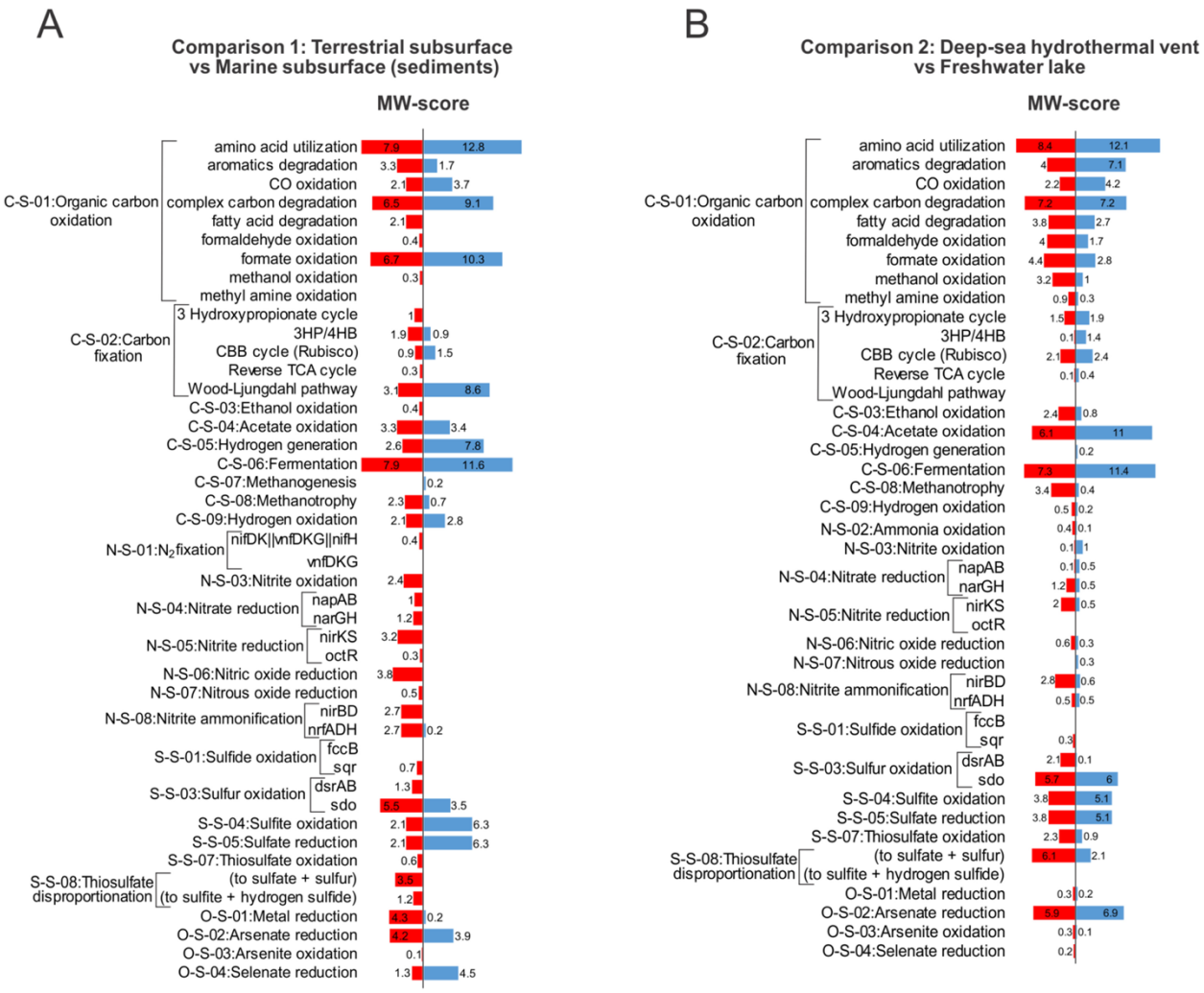


Figure 8. Community metabolism comparison based on MW-scores. (A) Comparison between terrestrial subsurface (left red bars) and marine subsurface (right blue bars); (B) Comparison between deep-sea hydrothermal vent (left red bars) and freshwater lake (right blue bars). MW-scores were calculated as gene coverage fractions for individual metabolic functions. Functions with MW-scores in both environments as zero were removed from each panel, e.g., N-S-02:Ammonia oxidation, N-S-09:Anammox, S-S-02:Sulfur reduction, and S-S-06:Sulfite reduction in Panel (A), and C-S-07:Methanogenesis, N-S-01:N2 fixation, N-S-09:Anammox, S-S-02:Sulfur reduction, and S-S-06:Sulfite reduction in Panel (B). Details for MW-score and each microbial group contribution refer to Supplementary Dataset S3.

METABOLIC enables accurate reconstruction of cell metabolism

To demonstrate applications of reconstructing and depicting cell metabolism based on METABOLIC results, two microbial genomes were used as an example (Figure 9). As illustrated in Figure 9A, Hadesarchaea archaeon 1244-C3-H4-B1 has no TCA cycling gene components, which is consistent with previous findings in archaea within this class [79]. Gluconeogenesis/glycolysis pathways are also lacking in the genome; since gluconeogenesis is the central carbon metabolism responsible for generating sugar monomers which will be further biosynthesized to polysaccharides as important cell structural components [80], the lack of this pathway could be due to genome incompleteness. As an enigmatic archaeal class newly discovered in the recent decade, Hadesarchaea have distinctive metabolisms that separate them from conventional euryarchaeotal groups. They almost lost all TCA cycle gene components for the production of acetyl-CoA; while they could metabolize amino acids in a heterotrophic lifestyle [79]. It is posited that the Hadesarchaea genome has been subjected to a streamlining process possibly due to nutrient limitations in their surrounding environments [79]. Due to their metabolic novelty and limited available genomes at the current time, there are still uncertainties on unknown/hypothetical genes and pathways and unclassified metabolic potential across the whole class. The previous metabolic characterization on four Hadesarchaea genomes indicates Hadesarchaea members could anaerobically oxidize CO, and H2 was produced as the side product [79]. In the Hadesarchaea archaeon 1244-C3-H4-B1 genome, METABOLIC results indicate the loss of all anaerobic carbon-monoxide dehydrogenase gene components, which suggests the distinctive metabolism of this Hadesarchaea archaeon from others and highlights the accuracy of METABOLIC in reflecting functional details.

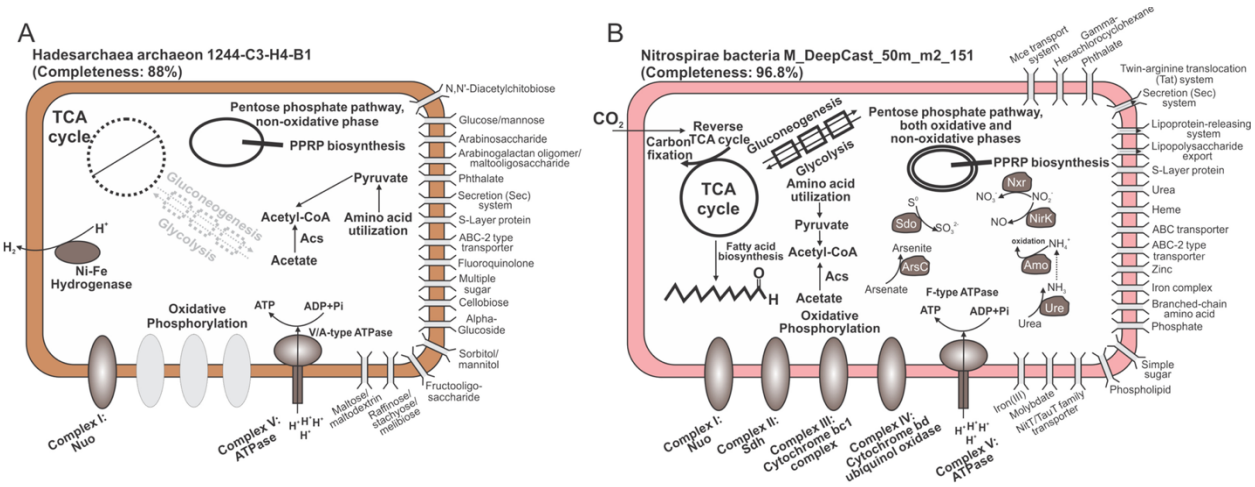


Figure 9. Cell metabolism diagrams of two microbial genomes. (A) cell metabolism diagram of Hadesarchaea archaeon 1244-C3-H4-B1 (B) cell metabolism diagram of Nitrospirae bacteria M_DeepCast_50m_m2_151. The absent functional pathways/complexes were labeled with dash lines.

We also reconstructed the metabolism for Nitrospirae bacteria M_DeepCast_50m_m2_151, a member of the Nitrospirae phylum reconstructed from Lake Tanganyika [56] (Figure 9B). It contains the full pathway for the TCA cycle and gluconeogenesis/glycolysis. Furthermore, it also has the full set of oxidative phosphorylation complexes for energy conservation and functional genes for nitrite oxidation to nitrate. Other nitrogen cycling metabolisms identified in this genome include ammonium oxidation, urea utilization, and nitrite reduction to nitric oxide. The reverse TCA cycle pathway was identified for carbon fixation. The metabolic profiling result is in accord with the fact that Nitrospirae is a well-known nitrifying bacterial class capable of nitrite oxidation and living an autotrophic lifestyle [80]. Additionally, their more abundant distribution in nature compared to other nitrite-oxidizing bacteria such as Nitrobacter indicates their significant contribution to nitrogen cycling in the environment [80]. This highlights the ability of METABOLIC in reflecting functional details of more common and prevalent microorganisms compared to the Hadesarchaea archaeon. Notably as

discovered from METABOLIC analyses, this bacterial genome also contains a wide range of transporter enzymes on the cell membrane, including mineral and organic ion transporters, sugar and lipid transporters, phosphate and amino acid transporters, heme and urea transporters, lipopolysaccharide and lipoprotein releasing system, bacterial secretion system, etc., which indicates its metabolic versatility and potential interactive activities with other organisms and the ambient environment. Collectively, METABOLIC result of functional profiling provides an intuitively-represented summary of a single microbial genome which enables depicting cell metabolism for better visualizing the functional capacity.

METABOLIC accurately represents metabolism in the human microbiome

In addition to resolving microbial metabolism and biogeochemistry in environmental microbiomes, METABOLIC also accurately identifies metabolic traits associated with human microbiomes. The implications of microbial metabolism on human health largely remain a black box, much like microbial contributions to biogeochemical cycling. We demonstrate the utility of METABOLIC in human microbiomes using publicly available data from stool samples collected from patients with colorectal cancer and healthy individuals. From this study, we selected stool metagenomes from one colorectal cancer (CRC) and an age and sex-matched healthy control to conduct the comparison. The heatmap indicates the human microbiome functional profiles of both samples based on the marker gene presence/absence patterns (Figure 10). As an example of METABOLIC's application, we demonstrate that there were 28 markers with variations > 10% in terms of the marker-containing genome fractions between these two samples

(Figure 10, Additional file 13: Dataset S6). These 28 markers involved all the ten metabolic categories except for lipid metabolism and translation, suggesting the broad functional differences between these two samples. In addition to analyzing human microbiome specific functional markers, METABOLIC can be used to visualize elemental nutrient cycling and analyze metabolic interactions in human microbiomes. Overall it enables systematic characterization of the composition, structure, function, and interaction of microbial metabolisms in the human microbiome and facilitates omics-based studies of microbial community on human health [60].

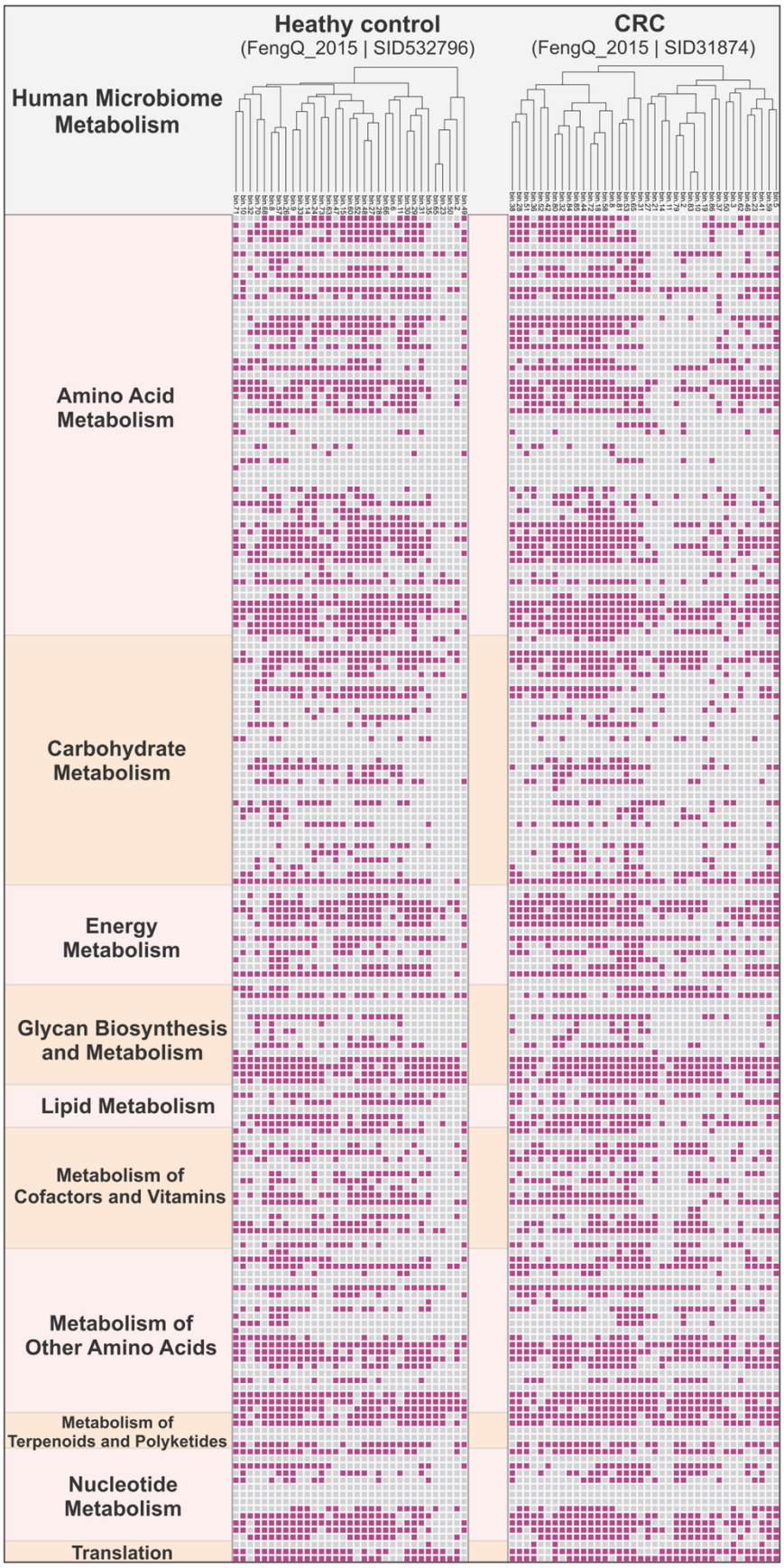

Heathy control
(FengQ_2015 | SID532796)
CRC
(FengQ_2015 | SID31874)
Human Microbiome Metabolism
Amino Acid Metabolism
Carbohydrate Metabolism
Energy Metabolism
Glycan Biosynthesis and Metabolism
Lipid Metabolism
Metabolism of Cofactors and Vitamins
Metabolism of Other Amino Acids
Metabolism of Terpenoids and Polyketides
Nucleotide Metabolism
Translation

Figure 10. Presence/Absence map of human microbiome metabolisms of a colorectal cancer (CRC) patient and a healthy control gut sample. The heatmap has summarized 189 horizontal entries (189 lines) based on 139 key functional gene families that covered 10 function categories. Purple cells indicate presence and gray cells indicate absence. Detailed KEGG KO identifier IDs and protein information for each function category were described in Supplementary Dataset S2.

DISCUSSION

The rapid increase in the availability of sequenced microbial genomes, metagenome-assembled genomes, and single-cell genomes has significantly benefited ecogenomic research on unraveling microbial functional roles and their metabolic contribution to biogeochemical cycles. Tools that enable to conduct accurate and reproducible functional profiling on genomic blueprints at the scale of both individual microorganisms and the whole microbial community offered significant applications and advances. They are fundamental to facilitate understanding of community-level functions, activities, interactions, and functional contributions in the era of multi-omics. An ideal tool for microbial biogeochemical profiling needs consideration on better organizing, interpreting, and visualizing the functional profile information; this is especially important for dealing with thousands of genomes reconstructed from metagenomes and studying the community-scale interactive metabolisms. Meanwhile, fast, accurate, robust performance and wide usage of the tool will allow for providing reliability and efficiency.

Here we developed METABOLIC for profiling metabolisms, biogeochemical pathways, and community-scale functional networks. Instead of solely depending on widely adopted protein annotation databases, in METABOLIC two additional steps were added in order to accurately predict protein functions and reconstruct metabolic pathways.

First, for TIGRfam/Pfam/Custom HMM profile databases, default NC/TC thresholds are often set too low to avoid noisy signals especially for annotating proteins from large sets of metagenomes wherein similar protein families often co-exist. This frequently leads to misannotations. To avoid this, we collected hmmsearch scores of previous annotation results and plotted these scores as a function of all annotations, and manually curated NC/TC by specifically picking the sharpest decreasing interval as the adjusted cutoff. Second, the motif validation step involves comparing potential hits to a set of manually curated highly conserved amino acid residues. This helps to distinguish two protein families with high sequence identity but different functions which are often difficult to separate by HMM profile-based annotations. These two steps help to filter out non-specific and cross-talking hits of important functional proteins for downstream bioinformatic analyses. After obtaining predicted metabolic pathways, many other software/online servers mostly provide raw annotation results with overwhelming yet unorganized details on characterizing protein functions. For microbial ecologists it is fundamental to provide organized and intuitive results to facilitate understanding on the whole landscape of biogeochemical cycling capacities. In METABOLIC, such a function was developed to enable visualizing the presence/absence state of each step of biogeochemical cycles for individual genomes and the whole microbial community. Combined with gene abundance information calculated by metagenomic read mapping, we can identify the relative abundance for each step of biogeochemical cycles. Furthermore, METABOLIC can also visualize sequential reaction patterns for important organic and inorganic compound transformations. This visualization function of METABOLIC is practical for representing the “metabolic handoff” scenario of within-

community interactions [2]. METABOLIC can be implemented in human microbiome with the same performance. Recently, METABOLIC was applied to stool metagenomic samples from 667 individuals who either were healthy or had adenomas or carcinomas of the colon, to profile organic/inorganic sulfate reduction and sulfide production [81]. This has considerably enlarged the utility of METABOLIC in community-scale investigation on human microbiomes for purposes of systematic microbiota-disease studies.

Previously, the community networks reflected by microbial genomes mostly focused on modeling reactions that are linked by metabolizing substrates and generating products [15, 19, 26]. On the contrary, METABOLIC was developed for a different purpose to study microbially-mediated biogeochemical processes. In METABOLIC the community-scale functional network provides an intuitive perspective on the metabolic connectivity among biogeochemical/metabolic steps and microbial contributions to these functions. MW-score, a metric that was built based on the same notion and methodology, offers quantitative measurement for these connected functions. Combined together they represent which functions are more centralized (connected with others) and important (weighted with higher relative abundance) in the co-sharing functional networks and which groups of microbial players contribute to these functions. Additionally, metabolic Sankey diagrams can be drawn to further visualize the microbial group contributions to different functions and biogeochemical cycles. As gene coverages generated by metagenomic read mapping can be replaced by transcript coverages generated by transcriptomic reads mapping, we broaden the usage in reflecting active function connections and weights. In practical applications, functional networks and MW-scores can be made in a standardized, reproducible, and normalized manner, so parallel

comparisons between communities (or samples) are applicable. The visualized network and Sankey diagram can also offer intuitive representations of functional connections and microbial contribution at both individual function and community-scale levels by using customized color schemes. There are other read-based metagenomic profiling tools, e.g., MetaPhlAn [28] and MEGAN [82], that can study the taxonomical and functional composition of microbiome at the community-scale level. Compared to read-based approaches which largely depend on the comprehensiveness of reference databases to capture microbial organisms, METABOLIC depends on the annotation of MAGs that is free from the limitation of reference databases on novel and rare organism characterization. METABOLIC specifically provides additional functionalities on annotation validation, result organization, and visualization which are meaningful to give reliable and easily accessible functional profiling results for microbial ecologists and biogeochemists to have a comprehensive understanding on the whole landscape of biogeochemical cycling capacities.

CONCLUSIONS

Metabolic functional profile of microbial genomes at the scale of individual organisms and communities is essential to have a comprehensive understanding of ecosystem processes, and as a conduit for enabling functional trait-based modeling of biogeochemistry. We have developed METABOLIC as a metabolic functional profiler that goes above and beyond current frameworks of genome/protein annotation platforms in providing protein annotations and metabolic pathway analyses that are used for inferring the contribution of microorganisms, metabolism, interactions, activity, and biogeochemistry at the community-scale. METABOLIC facilitates standardization

and integration of genome-informed metabolism into metabolic and biogeochemical models. We anticipate that METABOLIC will enable easier interpretation of microbial metabolism and biogeochemistry from metagenomes and genomes and enable microbiome research in diverse fields.

Declarations

Acknowledgments

We thank the comments and suggestions from the users of METABOLIC, which helped to improve and expand the functions of this software.

Funding

We thank the University of Wisconsin - Office of the Vice-Chancellor for Research and Graduate Education, University of Wisconsin – Department of Bacteriology, Wisconsin Alumni Research Foundation, and the University of Wisconsin – College of Agriculture and Life Sciences for their support. KA acknowledges support from the National Science Foundation under Grant No. NSF-DBI 2047598. PQT was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). KK is supported by a Wisconsin Distinguished Graduate Fellowship Award from the University of Wisconsin-Madison, and a William H. Peterson Fellowship Award from the Department of Bacteriology, University of Wisconsin-Madison. ESC is supported by an NLM training grant to the Computation and Informatics in Biology and in part by the Medicine Scientist Training Program (NLM 5T15LM007359).

Authors' contributions

ZZ and KA conceptualized and designed the study. ZZ and PQT wrote the Perl and R scripts. ZZ ran the test data and improved the software. YL provided a part of the databases. PQT, AMB, KK, ESC, and UK provided ideas and comments, helped to set up the GitHub page, and contributed to improving the overall performance of the software. ZZ and KA wrote the manuscript, and all authors contributed and approved the final edition of the manuscript.

Corresponding author

Correspondence to Karthik Anantharaman.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Wu X, Holmfeldt K, Hubalek V, Lundin D, Astrom M, Bertilsson S, Dopson M: Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep

biosphere reveal metabolic partitioning among populations. ISME J 2016, 10:1192-1203.
2. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, et al: Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 2016, 7:13219.
3. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, et al: Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol 2018, 3:328-336.
4. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T, Hentschel U: Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. ISME J 2011, 5:61-70.
5. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV: Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. Science 2012, 335:587-590.
6. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al: Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell 2019, 176:649-662 e620.
7. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA: Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 2017, 35:725.
8. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW: Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2017, 2:1533-1542.
9. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K: A new view of the tree of life. Nat Microbiol 2016, 1:16048.
10. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA: Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. ISME J 2020, 14:79-90.
11. Ruuskanen MO, Colby G, St Pierre KA, St Louis VL, Aris-Brosou S, Poulain AJ: Microbial genomes retrieved from High Arctic lake sediments encode for adaptation to cold and oligotrophic environments. Limnol Oceanogr 2020, 65:S233-S247.
12. Madsen EL: Microorganisms and their roles in fundamental biogeochemical cycles. Curr Opin Biotechnol 2011, 22:456-464.
13. Abreu NA, Taga ME: Decoding molecular interactions in microbial communities. FEMS Microbiol Rev 2016, 40:648-663.
14. Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C: Microbial syntrophy: interaction for the common good. FEMS Microbiol Rev 2013, 37:384-406.
15. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR: Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci U S A 2015, 112:6449.

16. Baker BJ, Lazar CS, Teske AP, Dick GJ: Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. Microbiome 2015, 3.
17. Morris BE, Henneberger R, Huber H, Moissl-Eichinger C: Microbial syntrophy: interaction for the common good. FEMS Microbiol Rev 2013, 37:384-406.
18. Graf DR, Jones CM, Hallin S: Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N2O emissions. PLoS One 2014, 9:e114118.
19. Machado D, Andrejev S, Tramontano M, Patil KR: Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res 2018, 46:7542-7553.
20. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al: The BioCyc collection of microbial genomes and metabolic pathways. Briefings in Bioinformatics 2019, 20:1085-1093.
21. Diener C, Gibbons SM, Resendis-Antonio O: MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. mSystems 2020, 5:e00606-00619.
22. Zimmermann J, Kaleta C, Waschina S: gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. Genome Biol 2021, 22:81-81.
23. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, Ong WK, Subhraveti P, Caspi R, Fulcher C, et al: Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics 2021, 22:109-126.
24. Zorrilla F, Buric F, Patil KR, Zelezniak A: metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. Nucleic Acids Res 2021, 49:e126-e126.
25. Belcour A, Frioux C, Aite M, Bretaudeau A, Hildebrand F, Siegel A: Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. eLife 2020, 9:e61968.
26. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY: Current status and applications of genome-scale metabolic models. Genome Biol 2019, 20:121.
27. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, et al: DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 2020, 48:8883–8900.
28. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, et al: Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife 2021, 10:e65088.
29. Hug Laura A, Co R: It Takes a Village: Microbial Communities Thrive through Interactions and Metabolic Handoffs. mSystems, 3:e00152-00117.
30. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000, 28:27-30.
31. Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, et al: MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 2006, 34:D511-516.

32. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al: Pfam: the protein families database. Nucleic Acids Res 2014, 42:D222-230.
33. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O: TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res 2007, 35:D260-264.
34. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M: The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 2013, 42:D206-D214.
35. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al: eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 2016, 44:D286-D293.
36. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y: KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008, 36:D480-D484.
37. Schimel J: 1.13 - Biogeochemical Models: Implicit versus Explicit Microbiology. In Global Biogeochemical Cycles in the Climate System. Edited by Schulze E-D, Heimann M, Harrison S, Holland E, Lloyd J, Prentice IC, Schimel D. San Diego: Academic Press; 2001: 177-183
38. Graham EB, Knelman JE, Schindlbacher A, Siciliano S, Breulmann M, Yannarell A, Beman JM, Abell G, Philippot L, Prosser J, et al: Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? Front Microbio 2016, 7.
39. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y: dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 2018, 46:W95-W101.
40. Rawlings ND, Barrett AJ, Finn R: Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2016, 44:D343-D350.
41. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H: KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 2019, 36:2251–2252.
42. Finn RD, Clements J, Eddy SR: HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 2011, 39:W29-37.
43. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ: Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010, 11:119.
44. UniProt C: UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 2019, 47:D506-D515.
45. Sondergaard D, Pedersen CN, Greening C: HydDB: A web tool for hydrogenase classification and analysis. Sci Rep 2016, 6:34212.
46. Venceslau SS, Stockdreher Y, Dahl C, Pereira IA: The "bacterial heterodisulfide" DsrC is a key protein in dissimilatory sulfur metabolism. Biochim Biophys Acta 2014, 1837:1148-1164.

47. Buchfink B, Xie C, Huson DH: Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015, 12:59-60.
48. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nat Methods 2012, 9:357.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25:2078-2079.
50. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT: BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011, 27:1691-1692.
51. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P: A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 2018, 36:996-1004.
52. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH: GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 2020, 36:1925-1927.
53. Anantharaman K, Breier JA, Sheik CS, Dick GJ: Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proc Natl Acad Sci U S A 2013, 110:330.
54. Olm MR, Brown CT, Brooks B, Banfield JF: dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 2017, 11:2864.
55. Glass JB, Ranjan P, Kretz CB, Nunn BL, Johnson AM, Xu M, McManus J, Stewart FJ: Microbial metabolism and adaptations in Atribacteria-dominated methane hydrate sediments. Environ Microbiol 2021, 23:4646-4660.
56. Tran PQ, Bachand SC, McIntyre PB, Kraemer BM, Vadeboncoeur Y, Kimirei IA, Tamatamah R, McMahon KD, Anantharaman K: Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. ISME J 2019, 15:1971-1986.
57. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, et al: Gut microbiome development along the colorectal adenoma–carcinoma sequence. Nat Commun 2015, 6:6528.
58. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, Lane KR, Thomas BC, Pan C, Northen TR, Banfield JF: Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nat Microbiol 2019, 4:1356-1367.
59. Stamps BW, Leddy MB, Plumlee MH, Hasan NA, Colwell RR, Spear JR: Characterization of the Microbiome at the World's Largest Potable Water Reuse Facility. Front Microbio 2018, 9.
60. Tu Q, He Z, Li Y, Chen Y, Deng Y, Lin L, Hemme CL, Yuan T, Van Nostrand JD, Wu L, et al: Development of HuMiChip for Functional Profiling of Human Microbiomes. PLoS One 2014, 9:e90546.
61. Kolde R, Kolde MR: Package 'pheatmap'. R Package 2015, 1:790.
62. Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, Kanehisa M: Modular architecture of metabolic pathways revealed by conserved sequences of reactions. Journal of Chemical Information and Modeling 2013, 53:613-622.

63. Kuypers MMM, Marchant HK, Kartal B: The microbial nitrogen-cycling network. Nat Rev Microbiol 2018, 16:263-276.
64. Hug LA, Co R: It Takes a Village: Microbial Communities Thrive through Interactions and Metabolic Handoffs. mSystems 2018, 3:e00152-00117.
65. Graf DRH, Jones CM, Hallin S: Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community Structure for N2O Emissions. PLoS One 2014, 9:e114118.
66. Mukhopadhyay R, Rosen BP, Phung LT, Silver S: Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol Rev 2002, 26:311-325.
67. Zhou Z, Liu Y, Pan J, Cron BR, Toner BM, Anantharaman K, Breier JA, Dick GJ, Li M: Gammaproteobacteria mediating utilization of methyl-, sulfur- and petroleum organic compounds in deep ocean hydrothermal plumes. ISME J 2020, 14:3136-3148.
68. Shih PM, Ward LM, Fischer WW: Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. Proc Natl Acad Sci U S A 2017, 114:10749-10754.
69. Berg IA, Kockelkorn D, Buckel W, Fuchs G: A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science 2007, 318:1782-1786.
70. Pester M, Schleper C, Wagner M: The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr Opin Microbiol 2011, 14:300-306.
71. Kanehisa M, Sato Y, Morishima K: BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 2016, 428:726-731.
72. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 2007, 35:W182-W185.
73. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al: eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 2019, 47:D309-D314.
74. Olson DL, Delen D: Advanced data mining techniques. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg 2008.
75. Glass JB, Ranjan P, Kretz CB, Nunn BL, Johnson AM, McManus J, Stewart FJ: Adaptations of Atribacteria to life in methane hydrates: hot traits for cold life. bioRxiv 2019, 1:536078.
76. Anantharaman K, Breier JA, Dick GJ: Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J 2015, 10:225.
77. Anantharaman K, Duhaime MB, Breier JA, Wendt K, Toner BM, Dick GJ: Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. Science 2014, 344:757-760.
78. Zhou Z, Tran PQ, Kieft K, Anantharaman K: Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. ISME J 2020, 14:2060-2077.
79. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs K-U, Teske AP, Ettema TJG: Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. Nat Microbiol 2016, 1:16002.

80. Madigan MT, John M. Martinko, Kelly S. Bender, Daniel H. Buckley, and David Allan Stahl: Brock Biology of Microorganisms. Fourteenth edition edn. Boston: Pearson; 2015.
81. Wolf PG, Cowley ES, Breister A, Matatov S, Lucio L, Polak P, Ridlon JM, Gaskins HR, Anantharaman K: Diversity and distribution of sulfur metabolism in the human gut microbiome and its association with colorectal cancer. bioRxiv 2021:2021.2007.2001.450790.
82. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R: MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol 2016, 12:e1004957.

Appendix I – Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes

This appendix has been published in *Gut Microbes:*

Lee J.W., Cowley E.S., Wolf P.G., Doden H.L., Murai T., Caicedo K.Y.O, Ly L.K., Sun F., Takei H., Nittono H., Daniel S.L., Cann I., Gaskins H.R., Anantharaman K., Alves J.M.P, Ridlon J.M. Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes. *Gut Microbes.* 14,1 (2022). https://doi.org/10.1080/19490976.2022.2132903.

All supplemental Materials are available on the publisher's website.

Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes

Jae Won Lee, [a,b] **Elise S. Cowley**, [c,d] Patricia G. Wolf, [a,b,e,f,*] Heidi L. Doden, [a,b] Tsuyoshi Murai, [g] Kelly Yovani Olivos Caicedo, [h] Lindsey K. Ly, [a,i] Furong Sun, [j] Hajime Takei, [k] Hiroshi Nittono, [k] Steven L. Daniel, [l] Isaac Cann, [a,b,i,m] H. Rex Gaskins, [a,b,i,n] Karthik Anantharaman, [c] João M. P. Alves, [h] and Jason M. Ridlon [a,b,i,n,o,p]

[a]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA

[b]Department of Animal Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA

[c]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

[d]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA

[e]Institute for Health Research and Policy, University of Illinois Chicago, Chicago, IL, USA

[f]University of Illinois Cancer Center, University of Illinois Chicago, Chicago, IL, USA

[g]School of Pharmaceutical Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan

[h]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

[i]Division of Nutritional Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA

[j]Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois Urbana-Champaign, IL, USA

[k]Junshin Clinic Bile Acid Institute, Tokyo, Japan

[l]Department of Biological Sciences, Eastern Illinois University, Charleston, IL, USA

[m]Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, IL, USA

[n]Cancer Center at Illinois, Urbana, IL, USA

[o]Center for Advanced Study, University of Illinois Urbana-Champaign, Urbana, IL, USA

[p]Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, VA, USA

CONTACT Jason M. Ridlon ude.sionilli@noldirmj Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA

*Present affiliation: Department of Nutrition Science, Purdue University, West Lafayette, IL, USA

Abstract

The gut microbiome of vertebrates is capable of numerous biotransformations of bile acids, which are responsible for intestinal lipid digestion and function as key nutrient-signaling molecules. The human liver produces bile acids from cholesterol predominantly in the A/B-*cis* orientation in which the sterol rings are “kinked”, as well as small quantities of A/B-*trans* oriented “flat” stereoisomers known as “primary allo-bile acids”. While the complex multi-step bile acid 7α-dehydroxylation pathway has been well-studied for conversion of “kinked” primary bile acids such as cholic acid (CA) and chenodeoxycholic acid (CDCA) to deoxycholic acid (DCA) and lithocholic acid (LCA), respectively, the enzymatic basis for the formation of “flat” stereoisomers allo-deoxycholic acid (allo-DCA) and allo-lithocholic acid (allo-LCA) by Firmicutes has remained unsolved for three decades. Here, we present a novel mechanism by which Firmicutes generate the ”flat” bile acids allo-DCA and allo-LCA. The BaiA1 was shown to catalyze the final reduction from 3-oxo-allo-DCA to allo-DCA and 3-oxo-allo-LCA to allo-LCA. Phylogenetic and metagenomic analyses of human stool samples indicate that BaiP and BaiJ are encoded only in Firmicutes and differ from membrane-associated bile acid 5α-reductases recently reported in Bacteroidetes that indirectly generate allo-LCA from 3-oxo-Δ^4-LCA. We further map the distribution of *baiP* and *baiJ*

among Firmicutes in human metagenomes, demonstrating an increased abundance of the two genes in colorectal cancer (CRC) patients relative to healthy individuals.

KEYWORDS: Secondary allo-bile acids, bile acid dehydroxylation, bile acid 5α-reductases, Firmicutes, colorectal cancer

Introduction

Bile acid synthesis in the liver represents a major route for removal of cholesterol from the body and bile acids function as an emulsifying agent for the digestion of lipid-soluble dietary components in the aqueous lumen of the small bowel.[1] In humans, the liver synthesizes two abundant primary bile acids, cholic acid (CA; 3α-,7α-,12α-trihydroxy-5β-cholan-24-oic acid) and chenodeoxycholic acid (CDCA; 3α-,7α-dihydroxy-5β-cholan-24-oic acid) from cholesterol. Before active secretion from the liver, bile acids are conjugated to either taurine or glycine at the C-24 carboxyl group.[1] When bile acids reach the terminal ileum, they are actively transported across the epithelium into portal blood and returned to the liver in a process known as enterohepatic circulation (EHC). Daily, several hundred milligrams of bile acids escape EHC and enter the large intestine. Colonic bacteria are capable of carrying out numerous biotransformations of primary bile acids to diverse secondary bile acids in the large intestine. The composition of intestinal and fecal bile acids in germ-free animals reflects the biliary composition.[2–5] Meanwhile, in conventional animals with a normal gut microbiota, fecal bile acid composition is diversified from only a few primary bile acids synthesized by the host to

an estimated ~400 secondary bile acid products.[6,7] Bacterial modifications to bile acids provide a form of interdomain communication given that beyond mere lipid-digesting detergents, bile acids are important nutrient-signaling molecules.[8] Indeed, microbial metabolism of bile acids is widely recognized to contribute to numerous human disorders including, but not limited to, cancers of the liver[9,10] and colon,[11] obesity, type 2 diabetes, nonalcoholic fatty liver disease (NAFLD),[12,13] cholesterol gallstone disease,[14,15] Alzheimer's disease,[16,17] and cardiovascular disease.[18]

A myriad of microbial bile acid biotransformations occur in the large intestine and include two key transformations. First, the conjugated bile acids are hydrolyzed to unconjugated bile acids and glycine or taurine by bile salt hydrolase (BSH).[19] Second, the unconjugated primary bile acids CA and CDCA are converted to deoxycholic acid (DCA; 3α-,12α-dihydroxy-5β-cholan-24-oic acid) and lithocholic acid (LCA; 3α-hydroxy-5β-cholan-24-oic acid)[20] via 7α-dehydroxylation, respectively. BSH (EC 3.5.1.24) enzymes are widely distributed among predominant microbial phyla within the domains Bacteria and Archaea inhabiting the human GI tract and catalyze the substrate-limiting deconjugation of bile acid amides.[19] The resulting major secondary bile acids routinely measured in human fecal samples are unconjugated derivatives of DCA and LCA.[20] A bile acid inducible (*bai*) regulon encoding enzymes involved in the conversion of CA to DCA (Figure 1), and CDCA and ursodeoxycholic acid (UDCA; 3α-,7β-dihydroxy-5β-cholan-24-oic acid) to LCA has been elucidated over the past three decades in strains of *Lachnoclostridium scindens* (formerly *Clostridium scindens), Peptacetobacter hiranonis* (formerly *Clostridium hiranonis*), and *Lachnoclostridum hylemonae* (formerly *Clostridium hylemonae*).[20] Discovery and characterization of *bai* genes have allowed

recent studies to extend the species distribution of 7-dehydroxylating bacteria into new families within the Firmicutes through bioinformatics-based searches of metagenomic sequence databases.[21,22] Similarly, comparison of the distribution of *bai* genes between fecal metagenomes obtained from healthy and disease cohorts has also enabled the association of the abundance of *bai* genes with risk for adenomatous polyps[23] or colorectal cancer.[24] This agrees with bile acid metabolomic studies that demonstrate increased fecal and serum DCA and LCA derivatives in subjects at high risk for CRC.[25–30] Conversely, lower abundance of *bai* genes is associated with bile acid dysbiosis characterized by increased fecal conjugated primary bile acids in inflammatory bowel diseases.[31,32]

There are additional *bai* genes yet to be accounted for in strains of *L. scindens* that result in the formation of stereoisomers of DCA and LCA known as “secondary allo-bile acids”. In 1991, Hylemon et al.[33] reported that allo-deoxycholic acid (allo-DCA; 3α-,12α-dihydroxy-5α-cholen-24-oic acid) formation is a CA-inducible side-product of bile acid 7-dehydroxylation by *L. scindens*. During the conversion of cholesterol to the primary bile acids CA and CDCA, the liver enzyme Δ^4-3-ketosteroid-5β-reductase (3-oxo-Δ^4-steroid-5β-reductase; AKR1D1) saturates the Δ^4-bond generating steroid A/B rings in the *cis*-orientation which appear “kinked” (Figure 1). When CA is transported into bacteria expressing *bai* genes, the first oxidative steps of bile 7-dehydroxylation, catalyzed by BaiA and BaiCD, “resetting” A/B ring stereochemistry through formation of the 3-keto-Δ^4 structure.[20] This is followed by the rate-limiting 7α-dehydration (BaiE).[34] The BaiCD was shown to then re-establish stereochemistry by catalyzing the conversion of 3-oxo-Δ^4-

DCA (12α-hydroxy-3-oxo-5β-chol-4-en-24-oic acid) to 3-oxo-DCA (12α-hydroxy-3-oxo-5β-cholan-24-oic acid), which is further reduced by BaiA1 and BaiA2 to DCA.[35] The current model of bile acid 7α-dehydroxylation suggests that another enzyme, currently unknown, acts on 3-oxo-Δ^4-DCA to form the alternative stereoisomer, 3-oxo-allo-DCA (12α-hydroxy-3-oxo-5α-cholan-24-oic acid), which is reduced by another unknown reductase to allo-DCA. Secondary allo-bile acids have a “flat” shape owing to hydrogenation that results in an A/B-*trans* orientation (Figure 1). While few studies have reported measurement of allo-DCA and allo-LCA (3-oxo-5α-cholan-24-oic acid), two studies have shown these bile acids are enriched in the feces of patients with CRC.[36,37] Derivatives of allo-LCA are also reported to be enriched in Japanese centenarians,[38] although there is a paucity of measurement of secondary allo-bile acids across populations and disease states. Thus, determining the gene(s) encoding reductases in *L. scindens* and other gut microbes responsible for the formation of allo-DCA and allo-LCA is of biomedical importance.

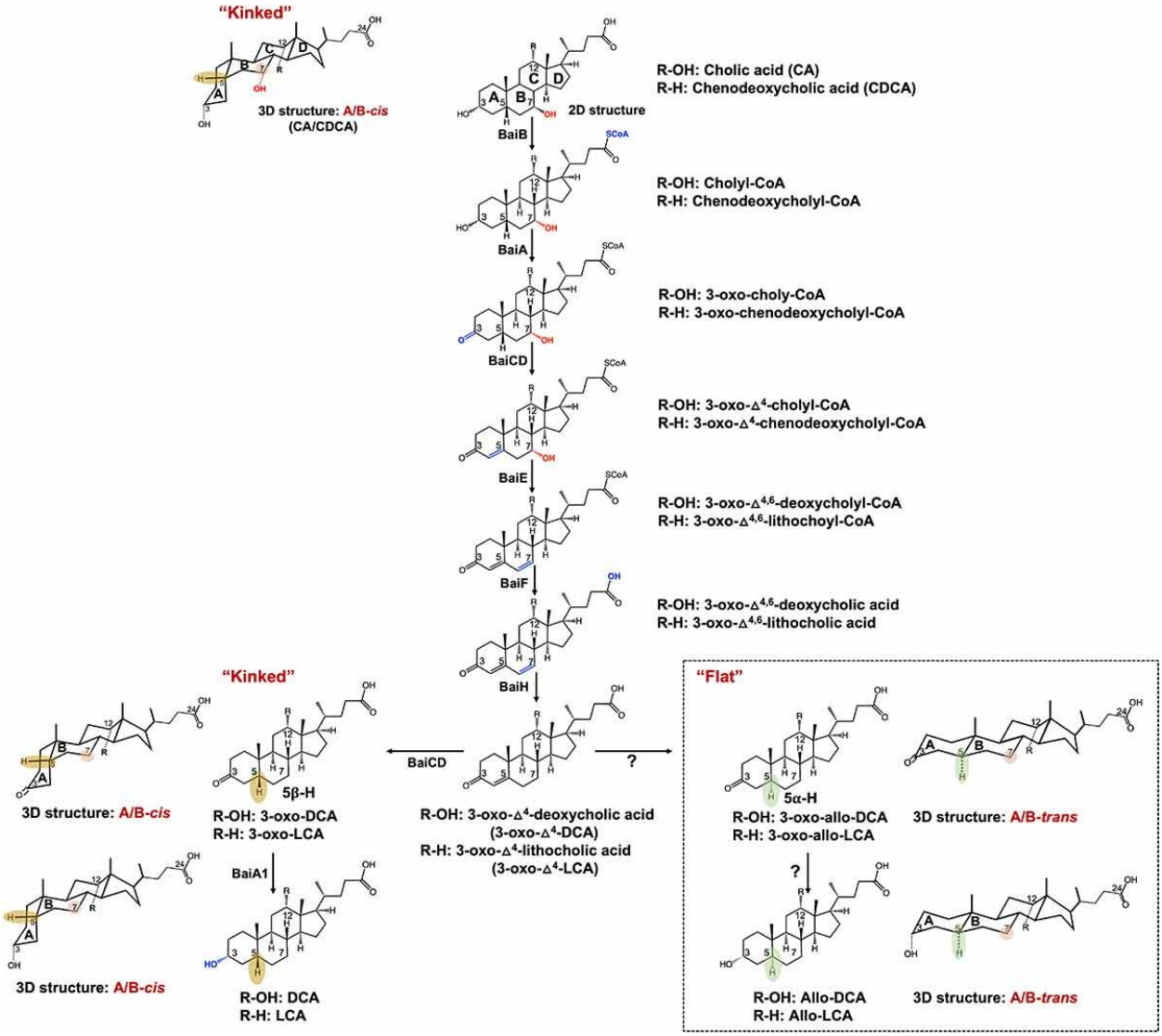


Figure 1. A proposed pathway for the 7α-dehydroxylation of cholic acid (CA) and chenodeoxycholic acid (CDCA) to deoxycholic acid (DCA) and allo-deoxycholic acid (allo-DCA), and lithocholic acid (LCA) and allo-lithocholic acid (allo-LCA).

BaiB, Bile acid CoA ligase; BaiA, 3α-hydroxysteroid dehydrogenase; BaiCD, 3-dehydro-$\Delta^{Citation4}$-7α-oxidoreductase; BaiE, 7α-dehydratase; BaiF, CoA transferase; BaiH, 3-dehydro-$\Delta^{Citation4}$-7β-oxidoreductase. The enzymes involved in the sequential reduction of 3-oxo-Δ^4-DCA and allo-DCA are currently unknown.

We recently reported genome-wide transcriptome profiling of *L. scindens* ATCC 35704 in the presence of CA and DCA and identified a potential candidate bile acid-inducible 3-oxo-Δ^4-5α-reductase.[39] Here, we confirm that this candidate bile acid-inducible gene encodes a novel bile acid 3-oxo-Δ^4-5α-reductase responsible for secondary allo-bile acids formation. We have named this gene in *L. scindens* ATCC 35704 the *baiP* gene. We previously reported identification of the *baiJ* gene as part of a polycistronic operon in *L. scindens* VPI 12708 and *L. hylemonae* DSM 15053, whose function remained unknown.[40] Our current study reports that the *baiJ* gene also encodes a bile acid 3-oxo-Δ^4-5α-reductase. The *baiP* and *baiJ* genes are distributed solely among the Firmicutes. Identification of these *bai* genes may provide the ability to predict and potentiate the formation of alternative forms of secondary bile acids whose ring structures are “flat” rather than the “kinked” form produced by the host. Indeed, we developed Hidden Markov Models (HMMs) of *bai* proteins and determined the distribution of *baiP* and *baiJ* in human metagenomes, demonstrating increased abundance in colorectal cancer (CRC) patients relative to healthy individuals.

Results

The HDCHBGLK_03451 gene from *L. scindens* ATCC 35704 encodes a bile acid 5α-reductase, yielding secondary allo-bile acids

Prior work established that allo-DCA is a CA-induced side-product of CA metabolism in cell-extracts of *L. scindens* VPI 12708[33] (Figure 1). We previously identified *L. scindens* ATCC 35704 gene HDCHBGLK_03451 as CA-inducible and suggested this is a likely

candidate for bile acid 5α-reductase[39] (Figure 2a). The gene HDCHBGLK_03451 encodes a 563 amino acid protein comprising FMN (flavin mononucleotide) and FAD (flavin adenine dinucleotide)-binding domains (Figure 2b). The HDCHBGLK_03451 gene from *L. scindens* ATCC 35704 was codon-optimized for *E. coli* and overexpressed in *E. coli* (Figure 2c) for resting cell assays with bile acid intermediates (Figure 2d). The stereochemistry of the A/B ring junction is lost during the steps leading up to and following 7α-dehydration of CA (BaiE),[41] resulting in formation of a 7α-deoxy-3-oxo-Δ^4-intermediates of DCA or LCA, respectively, which are reduced by the BaiH yielding 3-oxo-Δ^4-intermediates.[35] The 3-oxo-Δ-^{4}intermediate is then predicted to yield either 3-oxo-DCA (BaiCD) or 3-oxo-allo-DCA (BaiP). The same enzymatic steps are involved in the conversion of CDCA to 3-oxo-Δ^4-LCA followed by conversion to 3-oxo-LCA (3-oxo-5β-cholan-24-oic acid) or 3-oxo-alloLCA (3-oxo-5α-cholan-24-oic acid) by BaiCD or BaiP, respectively (Figure 1).

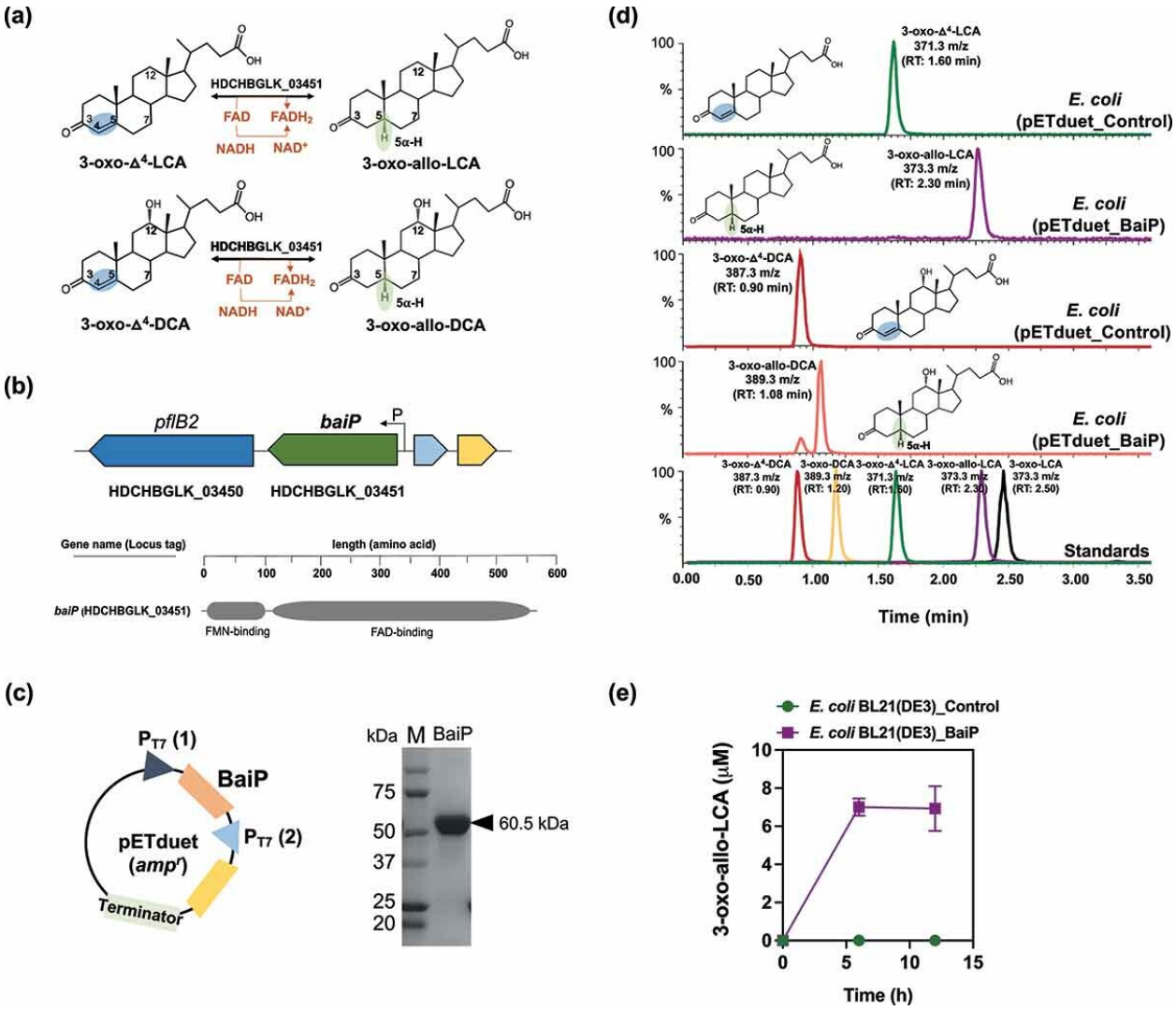


Figure 2. The baiP gene from L. scindens ATCC 35704 encodes a bile acid 5α-reductase. (a) Formation of bile acid stereoisomers after reduction of 3-oxo-Δ^4-LCA and 3-oxo-Δ^4-DCA by 5α-reductase. (b) Gene organization of *baiP* with genomic context and domain structure of BaiP. (c) Cloning strategy for heterologous expression of N-terminal his-tagged recombinant BaiP in *E. coli* BL21(DE3). SDS-PAGE confirms expression of 60.5 kDa recombinant BaiP. (d) Representative LC/MS chromatograms after resting cell assay with *E. coli* BL21(DE3) pETduet_Control or pETduet_BaiP incubated in anaerobic PBS containing 50 μM 3-oxo-Δ^4-LCA (Top panels 1 & 2) or

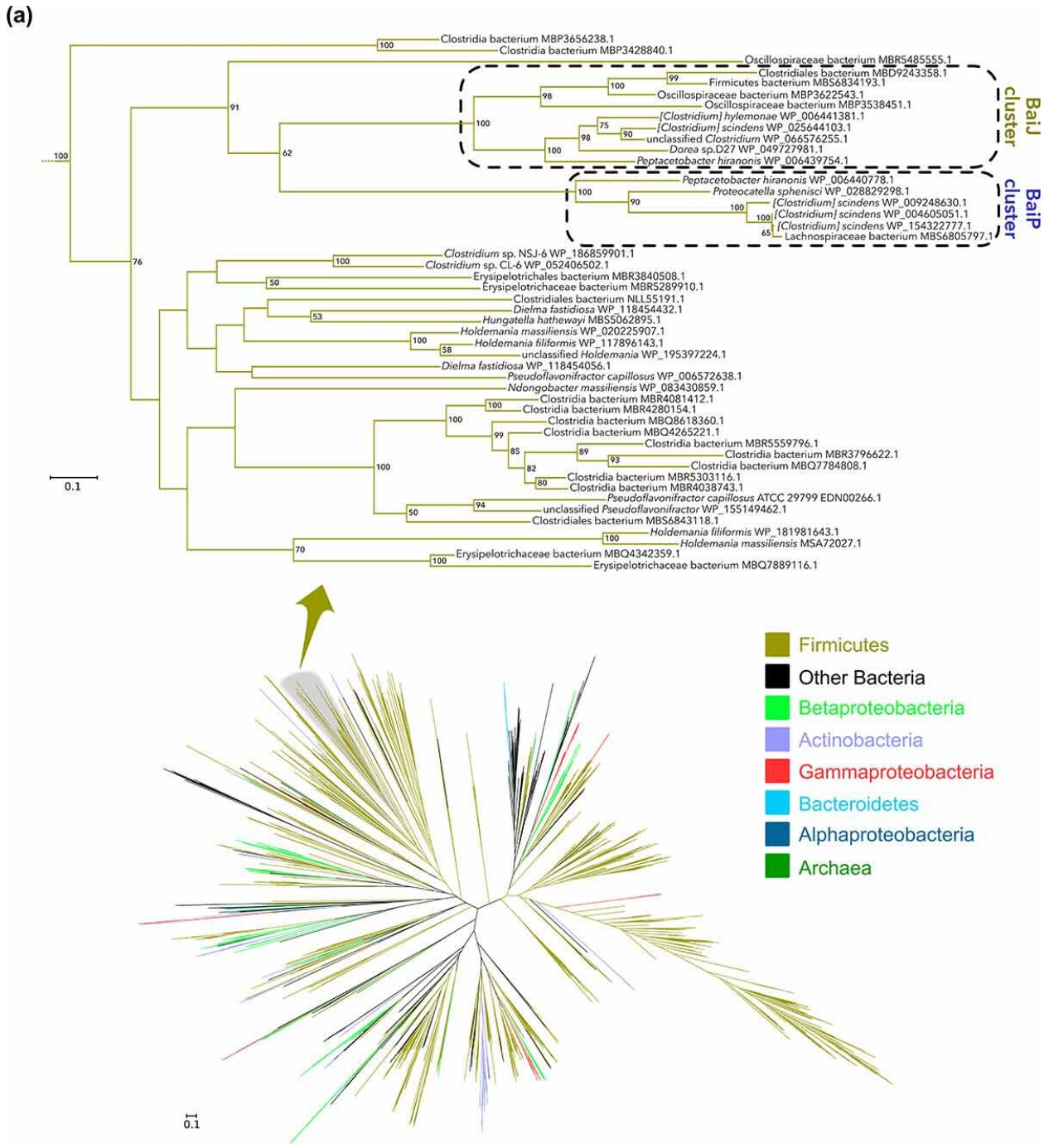
50 μM 3-oxo-Δ^4-DCA (Bottom panels 3 & 4). Standards are shown in Panel 5 (bottom). (e) Time course of 3-oxo-allo-LCA production by the *E. coli* BL21(DE3) pETduet_BaiP strain. Data points indicate the mean concentration of 3-oxo-allo-LCA ± SD (three biological replicates).

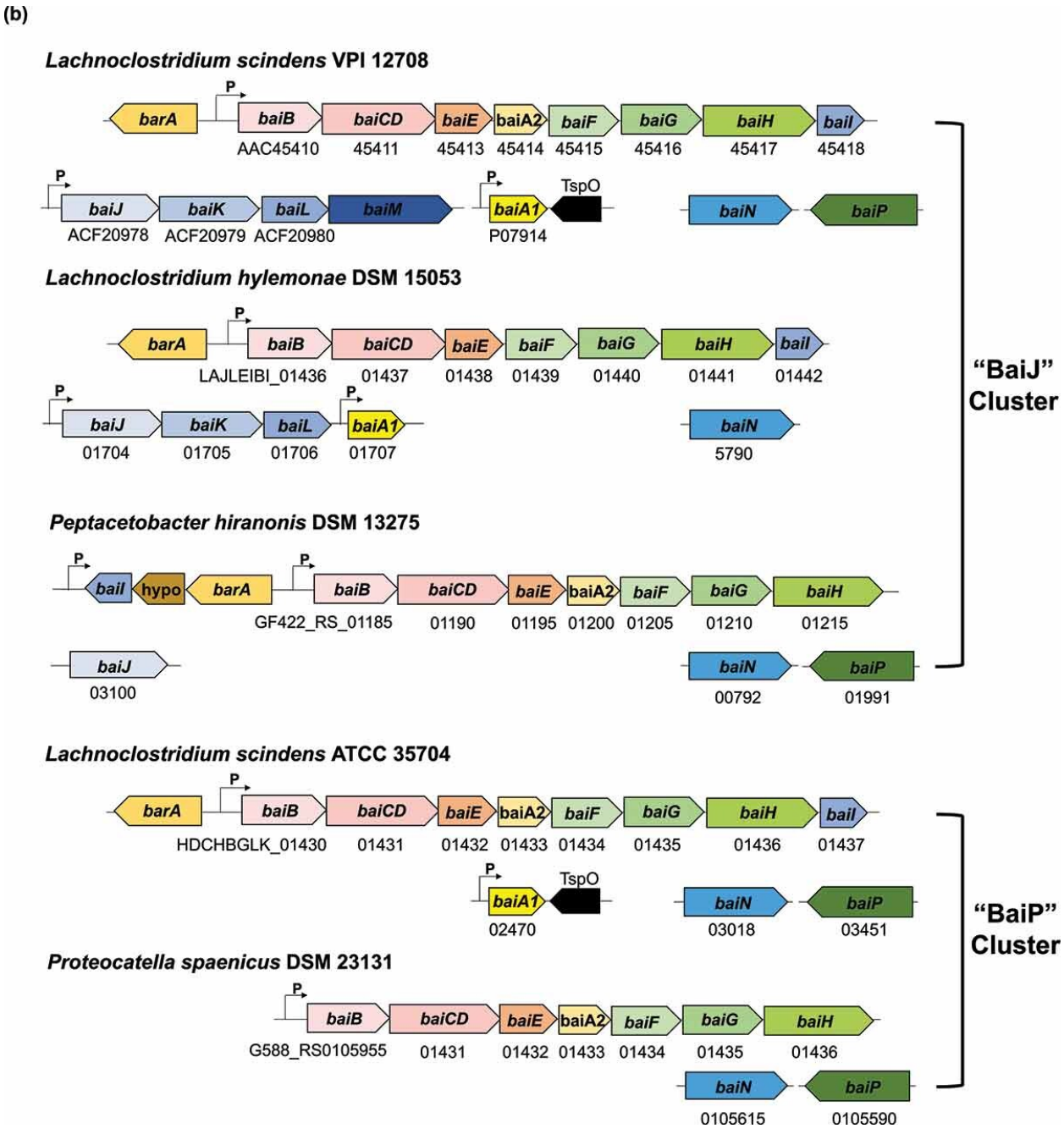
We therefore chemically synthesized 3-oxo-Δ^4-DCA and 3-oxo-Δ^4-LCA and incubated these substrates (50 μM) with *E. coli* expressing HDCHBGLK_03451 under anaerobic conditions in PBS. When 3-oxo-Δ^4-LCA was present as the substrate, 3-oxo-allo-LCA (RT = 2.30 min; *m/z* = 373.3) was synthesized, but not 3-oxo-LCA (RT = 2.50 min; *m/z* = 373.3) (Figure 2d). The 6 h reaction yielded 7.00 ± 0.46 μM 3-oxo-allo-LCA (Figure 2e). Similarly, incubation of resting cells with 3-oxo-Δ^4-DCA yielded a product (RT = 1.08 min; *m/z* = 389.26) consistent with 3-oxo-allo-DCA (RT = 1.08 min; *m/z* = 389.26), but not 3-oxo-DCA (RT = 1.20 min; *m/z* = 389.26) (Figure 2d). These data confirm that HDCHBGLK_03451 encodes a novel bile acid 5α-reductase, and we propose the name *baiP* for this gene (See **Supplementary material, Figure S1**).
We previously reported a cortisol-inducible operon (*desABCD*) in *L. scindens* ATCC 35704 encoding steroid-17,20-desmolase (DesAB) and NADH-dependent steroid 20α-hydroxysteroid dehydrogenase (DesC).[42] DesC reversibly forms cortisol and 20α-dihydrocortisol,[42] and DesAB catalyzes the side-chain cleavage of cortisol yielding 11β-hydroxyandrostenedione (11β-OHAD).[43] Because substrates and products in the desmolase pathway have 3-oxo-Δ^4-structures analogous to 3-oxo-Δ^4-DCA and 3-oxo-Δ^4-LCA, we next performed resting cell assays with *E. coli* strain expressing the BaiP

enzyme. LC/MS analysis of reaction products indicates that cortisol and 11β-OHAD were not substrates for BaiP (**Figure S2**).

Phylogenetic analysis of BaiP followed by functional assay reveals the baiJ gene also encodes bile acid 5α-reductase

Having provided experimental evidence that *baiP* encodes an enzyme with bile acid 5α-reductase activity, we wanted to determine the phylogeny of the BaiP from *L. scindens* ATCC 35704. A subtree of the >1,400 sequences representing close relatives of the BaiP from *L. scindens* ATCC 35704 was generated (Figure 3a). The proteins most closely related to BaiP from *L. scindens* ATCC 35704 in the "BaiP Cluster" were from *Lachnoclostridium* strains MSK.5.24, GGCC_0168, and Lachnospiraceae bacterium 5_1_57FAA. Additional FAD-dependent oxidoreductase BaiP candidates from a penguin isolate, *Proteocatella sphenisci* DSM 23131 (76% sequence identity), and *P. hiranonis*[15,44] (72% sequence identity) were also identified at high bootstrap values (90–100%). Previous work established *bai* genes in *P. hiranonis*,[45] although the present data provide first indication that *P. hiranonis* has the potential to form secondary allo-bile acids (Figure 3a, 3b). *P. sphenisci* has also been reported to encode the *bai* polycistronic operon,[21,22] and our demonstration that *P. sphenisci* harbors *baiP* indicate that secondary allo-bile acids may constitute part of the bile acid metabolome of penguin guano (Figure 3b).


(a)
Clostridia bacterium MBP3656238.1
Clostridia bacterium MBP3428840.1
Oscillospiraceae bacterium MBR5485555.1
Clostridiales bacterium MBD9243358.1
Firmicutes bacterium MBS6834193.1
Oscillospiraceae bacterium MBP3622543.1
Oscillospiraceae bacterium MBP3538451.1
[Clostridium] hylemonae WP_006441381.1
[Clostridium] scindens WP_025644103.1
unclassified Clostridium WP_066576255.1
Dorea sp.D27 WP_049727981.1
Peptacetobacter hiranonis WP_006439754.1
BaiJ cluster
Peptacetobacter hiranonis WP_006440778.1
Proteocatella sphenisci WP_028829298.1
[Clostridium] scindens WP_009248630.1
[Clostridium] scindens WP_004605051.1
[Clostridium] scindens WP_154322777.1
Lachnospiraceae bacterium MBS6805797.1
BaiP cluster
Clostridium sp. NSJ-6 WP_186859901.1
Clostridium sp. CL-6 WP_052406502.1
Erysipelotrichales bacterium MBR3840508.1
Erysipelotrichaceae bacterium MBR5289910.1
Clostridiales bacterium NLL55191.1
Dielma fastidiosa WP_118454432.1
Hungatella hatheway MBS5062895.1
Holdemania massiliensis WP_020225907.1
Holdemania filiformis WP_117896143.1
unclassified Holdemania WP_195397224.1
Dielma fastidiosa WP_118454056.1
Pseudoflavonifractor capillosus WP_006572638.1
Ndongobacter massiliensis WP_083430859.1
Clostridia bacterium MBR4081412.1
Clostridia bacterium MBR4280154.1
Clostridia bacterium MBQ8618360.1
Clostridia bacterium MBQ4265221.1
Clostridia bacterium MBR5559796.1
Clostridia bacterium MBR3796622.1
Clostridia bacterium MBQ7784808.1
Clostridia bacterium MBR5303116.1
Clostridia bacterium MBR4038743.1
Pseudoflavonifractor capillosus ATCC 29799 EDN00266.1
unclassified Pseudoflavonifractor WP_155149462.1
Clostridiales bacterium MBS6843118.1
Holdemania filiformis WP_181981643.1
Holdemania massiliensis MSA72027.1
Erysipelotrichaceae bacterium MBQ4342359.1
Erysipelotrichaceae bacterium MBQ7889116.1
0.1
Firmicutes
Other Bacteria
Betaproteobacteria
Actinobacteria
Gammaproteobacteria
Bacteroidetes
Alphaproteobacteria
Archaea
0.1


(b)
Lachnoclostridium scindens VPI 12708
barA
baiB
baiCD
baiE
baiA2
baiF
baiG
baiH
baiI
AAC45410
45411
45413
45414
45415
45416
45417
45418
baiJ
baiK
baiL
baiM
ACF20978
ACF20979
ACF20980
baiA1
TspO
P07914
baiN
baiP
Lachnoclostridium hylemonae DSM 15053
LAJLEIBI_01436
01437
01438
01439
01440
01441
01442
01704
01705
01706
01707
5790
Peptacetobacter hiranonis DSM 13275
baiI
hypo
GF422_RS_01185
01190
01195
01200
01205
01210
01215
03100
00792
01991
"BaiJ" Cluster
Lachnoclostridium scindens ATCC 35704
HDCHBGLK_01430
01431
01432
01433
01434
01435
01436
01437
02470
03018
03451
Proteocatella spaenicus DSM 23131
G588_RS0105955
0105615
0105590
"BaiP" Cluster

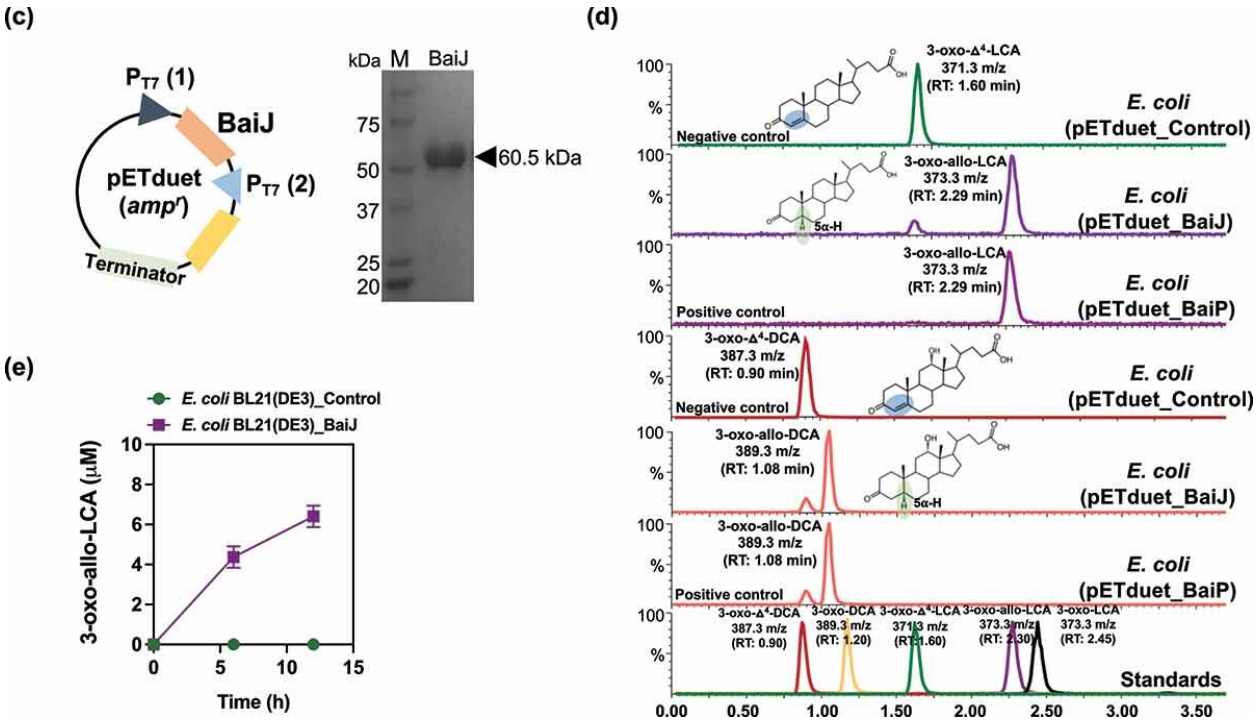


Figure 3. Large scale phylogenetic analysis of BaiP from L. scindens ATCC 35704 reveals baiJ gene from C. scindens VPI 12708 encodes a bile acid 5α-reductase. (a) Maximum-likelihood tree of >2,300 protein sequences from NCBI's non-redundant database that were similar to BaiP from *L. scindens*. The subtree containing BaiP from *L. scindens* formed two clusters containing BaiP sequences (Purple) from other Firmicutes known to convert CA to DCA. The second cluster contains BaiJ proteins, representing several strains known to convert CA to DCA. (b) Arrangement of genes in the bile acid inducible (*bai*) operon in various species of bile acid 7α-dehydroxylating gut bacteria. The gene encoding enzymes carrying out bile acid metabolism in gut bacteria capable of producing secondary allo-bile acids. Biochemical pathway leading to secondary allo-bile acid formation is shown in Figure 1. (c) Cloning strategy for *baiJ* gene from *L. scindens* VPI 12708 and SDS-PAGE after purification of recombinant His-tagged BaiJ. (d) Representative LC/MS chromatographs after resting cell assay with *E.*

coli BL21(DE3) pETduet_Control or pETduet_BaiJ incubated in anaerobic PBS containing 50 μM 3-oxo-Δ^4-LCA (Top panels 1 & 2) compared to pETduet_BaiP (Panel 3). Panels 4 & 5 display chromatograms of reaction products formed after incubation of *E. coli* BL21(DE3) pETduet_Control or pETduet_BaiJ incubated in anaerobic PBS containing 50 μM 3-oxo-Δ^4-DCA compared to pETduet_BaiP (Panel 6). Standards are shown in Panel 7 (bottom). (e) Time course of 3-oxo-allo-LCA production by the *E. coli* BL21(DE3) pETduet_BaiJ strain. Data points indicate the mean concentration of 3-oxo-allo-LCA ± SD (two biological replicates)

A second closest FAD-dependent oxidoreductase cluster (~45% ID) to BaiP from *L. scindens* ATCC 35704 was composed of the previously named BaiJ proteins from *L. scindens* VPI 12708, *L. hylemonae* DSM 15053, and *P. hiranonis* DSM13275, as well as *Dorea* sp. D27, and an unclassified *Clostridium* sp. ("BaiJ Cluster"). Prior work established a novel *bai* operon in which the *baiJ* gene is adjacent to the *baiK* gene on a polycistronic operon in *L. scindens* VPI 12708 and *L. hylemonae* DSM 15053.[40] Evidence was also presented that *L. scindens* VPI 12708 and *L. hylemonae* DSM 15053 formed allo-DCA.[46] It was then reported that the BaiK is a paralog of BaiF in *L. scindens* VPI 12708, and both proteins catalyze bile acid coenzyme A transferase from the end-product secondary bile acids, DCA~SCoA and allo-DCA~SCoA, to primary bile acids including CA, CDCA, allo-CA, and UDCA.[40] The *baiJ* gene has been shown previously to be enriched in the gut microbiome in mouse models of liver cancer and CRC,[9,24] diseases reported to be enriched in secondary allo-bile acids in the biliary pool in the few studies that have measured them.[47] Taken together, the close phylogenetic

clustering of BaiJ with BaiP indicates that the *baiJ* gene may also encode a bile acid 5α-reductase isoform (Figure 3a, 3b).[21,44,45]

To test this hypothesis, we cloned and overexpressed the *baiJ* gene from *L. scindens* VPI 12708 (accession number: ACF20978) in *E. coli* BL21(DE3) (Figure 3c), and measured conversion of 3-oxo-Δ^4-LCA and 3-oxo-Δ^4-DCA in resting cell assays (Figure 3d). When 3-oxo-Δ^4-LCA (RT = 1.60; *m/z* = 371.25) was the substrate, a product eluting at the same position as 3-oxo-allo-LCA (RT = 2.29; *m/z* = 373.27), but not as 3-oxo-LCA (RT = 2.45; *m/z* = 373.26), was observed. An anaerobic resting cell assay (6 h) resulted in the formation of 4.4 ± 0.54 μM 3-oxo-allo-LCA (Figure 3e). Similarly, when 3-oxo-Δ^4-DCA (RT = 0.90; *m/z* = 387.25) was the substrate, a product that eluted at the same position as 3-oxo-allo-DCA (RT = 1.08; *m/z* = 389.26), and different from 3-oxo-DCA (RT = 1.20; *m/z* = 389.27), was observed (Figure 3d). These results establish a function for the *baiJ* gene product and indicate that strains of *L. scindens* and other bile acid 7α-dehydroxylating bacteria encode distinct bile acid 5α-reductase isoforms.

BaiP and BaiA1 catalyze consecutive final reductive steps in the formation of allo-DCA and allo-LCA

Having established that BaiP converts 3-oxo-Δ^4-LCA to 3-oxo-allo-LCA, we next sought to identify an enzyme from *L. scindens* ATCC 35704 catalyzing the final reductive step from 3-oxo-allo-LCA to allo-LCA. There is compelling evidence that BaiA1 and BaiA2 enzymes catalyze the first oxidative and last reductive steps in the pathway.[35,48,49] This comes from substrate-specificity and kinetic analyses of BaiA1 and BaiA2 showing that

3-oxo-DCA and 3-oxo-LCA are substrates[48] and by the observation that BaiA is sufficient for the final reductive step yielding DCA.[35] Prior work established that the *baiA* genes encode bile acid 3α-hydroxysteroid dehydrogenase (3α-HSDH) that catalyze the first oxidation step, formation of 3-oxo-7α-hydroxy-5β-bile acids, and the final reductive step generating 7-deoxy-3α-hydroxy-5β-bile acids.[49] However, the ability of BaiA enzymes to recognize allo-bile acids has not been established (Figure 4a). The *baiA1* gene from *L. scindens* ATCC 35704 was codon-optimized for *E. coli* and overexpressed in *E. coli* alone or in combination with *baiP* (Figure 4b). Whole cell *E. coli* assays with overexpressed BaiA1 converted 3-oxo-allo-LCA (RT = 2.30 min; *m/z* = 373.2) to a product consistent with allo-LCA (RT = 2.74 min; *m/z* = 375.3), but not LCA (RT = 2.68 min; *m/z* = 375.3). *E. coli* expressing both BaiP and BaiA1 converted 3-oxo-Δ^4-LCA (RT = 1.65 min; *m/z* = 371.3) to allo-LCA (RT = 2.74 min; *m/z* = 375.3) and 3-oxo-Δ^4-DCA (RT = 0.75 min; *m/z* = 387.3) to allo-DCA (RT = 1.51 min; *m/z* = 391.3) confirming the role of BaiP and BaiA1 in the cooperative catalysis of the two final steps in formation of secondary allo-bile acids (Figure 4c).

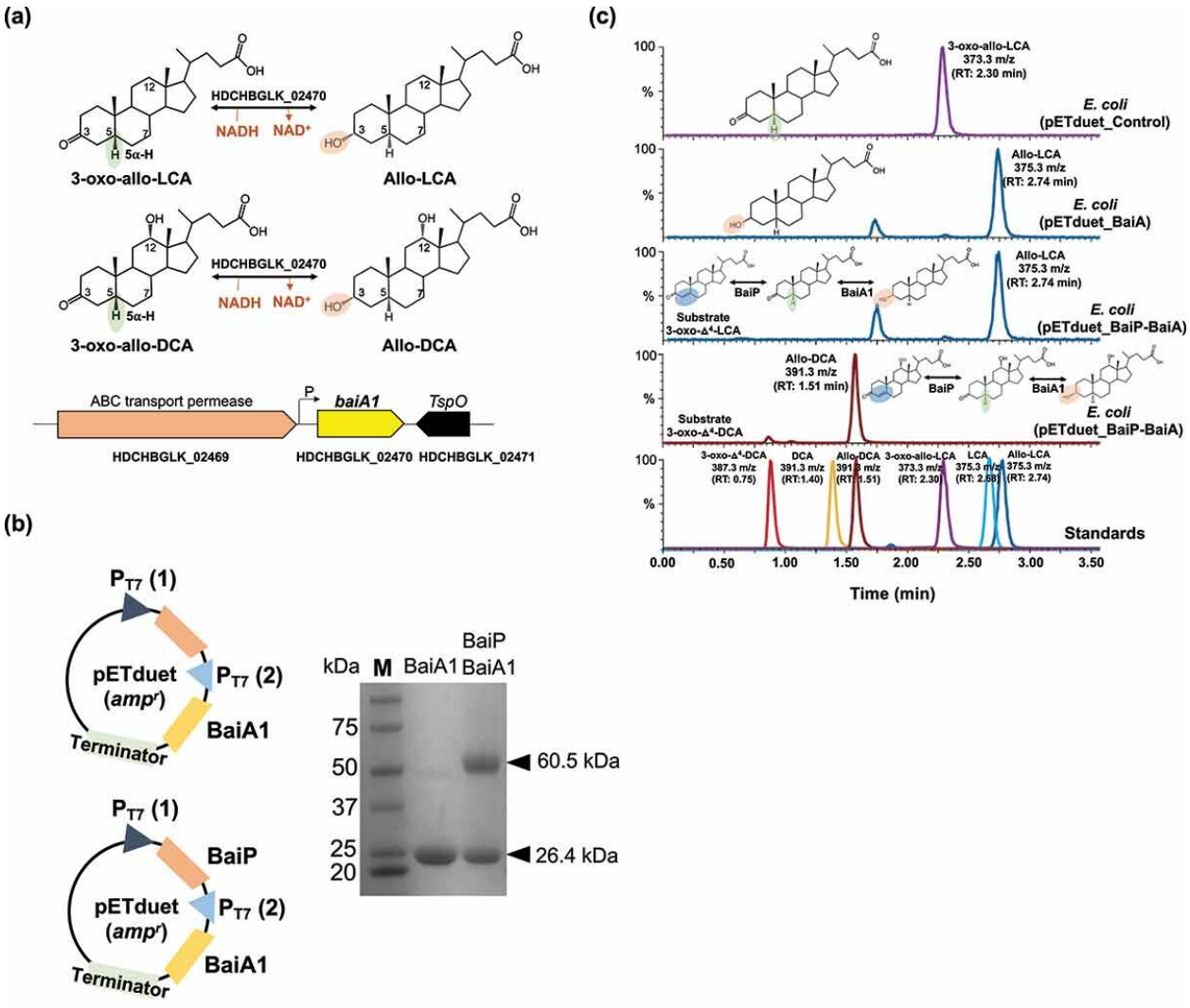


Figure 4. Recombinant BaiA1 from L. scindens ATCC 35704 catalyzes the final reductive step in the formation of allo-DCA and allo-LCA. (a) Formation of bile acid stereoisomers after reduction of 3-oxo-allo-LCA and 3-oxo-allo-DCA by 3α-HSDH and gene organization of *baiA1* in *L. scindens* ATCC 35704. (b) Cloning strategy of *baiA1* and *baiA1* + *baiP* in pETduet. SDS-PAGE of His-tagged purified recombinant BaiA1 and BaiA1 + BaiP expressed in *E. coli* BL21(DE3). (c) Representative LC/MS chromatograms after resting cell assay with *E. coli* BL21(DE3) pETduet_Control or pETduet_BaiA1 incubated in anaerobic PBS containing 50 μM 3-oxo-allo-LCA (Top

panels 1 & 2), *E. coli* BL21(DE3) pETduet_BaiP-BaiA1 incubated with 50 μM 3-oxo-Δ^4-LCA (Panel 3) and *E. coli* BL21(DE3) pETduet_BaiP-BaiA1 incubated with 50 μM 3-oxo-Δ^4-DCA (Panel 4). Standards are shown in Panel 5 (bottom). The overall two-step reaction is shown on the panels.

A previous bioinformatics study hypothesized based on gene context and annotation that *CLOSCI_00522*, a gene directly downstream from *baiN* (*CLOSCI_00523*), encodes a predicted NAD(FAD)-utilizing dehydrogenase involved in the final reductive step[31,] (**Figure S1**). This gene was named "*baiO"*.[31] An organism may encode several proteins from different lineages that have similar catalytic activity. Indeed, the BaiN[50] is predicted to catalyze similar sequential reactions to BaiH and BaiCD.[35] We therefore tested the hypothesis that the previously annotated *baiO* encodes either a bile acid 3-oxo-Δ^4-reductase and/or bile acid 3α-HSDH. We cloned the *baiO* in pETduet and verified the expression after His-tag purification and SDS-PAGE (**Figure S1a, S1b**). Analysis of bile acid products after 24 h incubation of *E. coli* expressing BaiO enzyme in a resting cell assay with either 3-oxo-LCA, 3-oxo-DCA (**Figure S1c, S1d)**, 3-oxo-Δ^4-LCA, or 3-oxo-Δ^4-DCA (**Figure S1e, S1f)**, did not yield a detectable product by LC/MS. While this does not disprove that *CLOSCI_00522* is involved in bile acid metabolism, we were not able to confirm its function.

The distribution of *baiP* and *baiJ* genes in public human metagenome datasets

Having shown that BaiP clusters with the previously identified BaiJ from *L. hylemonae* DSM 15053, the next objective was to determine the presence of *bai* genes involved in

bile acid 7-dehydroxylation among bacterial genomes from human stool samples. We utilized reference sequences of BaiP and BaiJ as well as BaiE and BaiCD (Figure 5a) to generate HMMs in order to search public human metagenomic databases. We expected that the occurrence of BaiE and BaiCD which are co-transcribed on the multi-gene *bai* operon will coincide with the relative abundances of BaiP and BaiJ. As expected, genes for BaiE and BaiCD as well as BaiP and BaiJ were observed to have similar relative frequency (1% and 0.9% of total metagenome assembled genomes (MAGs), respectively). All genes were largely represented by unclassified Firmicutes and *Lachnospiraceae*. (Figure 5a). Representative genera were analyzed to identify candidates which possess multiple genes of the Bai operon which revealed that unclassified Firmicutes, unclassified *Lachnospiraceae*, and *Flavonifractor* harbored all four genes analyzed. This pathway analysis also revealed the novel finding that *Flavonifractor* and *Pseudoflavonifractor* harbor genes for bile acid 7-dehydroxylation. Intriguingly, while *bai* genes represented approximately 1% of total MAGs, genes were detected in approximately one third of subjects (BaiCD 35%, BaiE 35%, BaiJ 30%, and BaiP 28%). An analysis of differences in gene presence among healthy subjects and those with adenoma and carcinoma revealed that the genes had the greatest abundance in patients with carcinoma, and that the genes *baiCD, baiE*, and *baiJ* were significantly associated with carcinoma (Figure 5b**, Table S4, S5**)

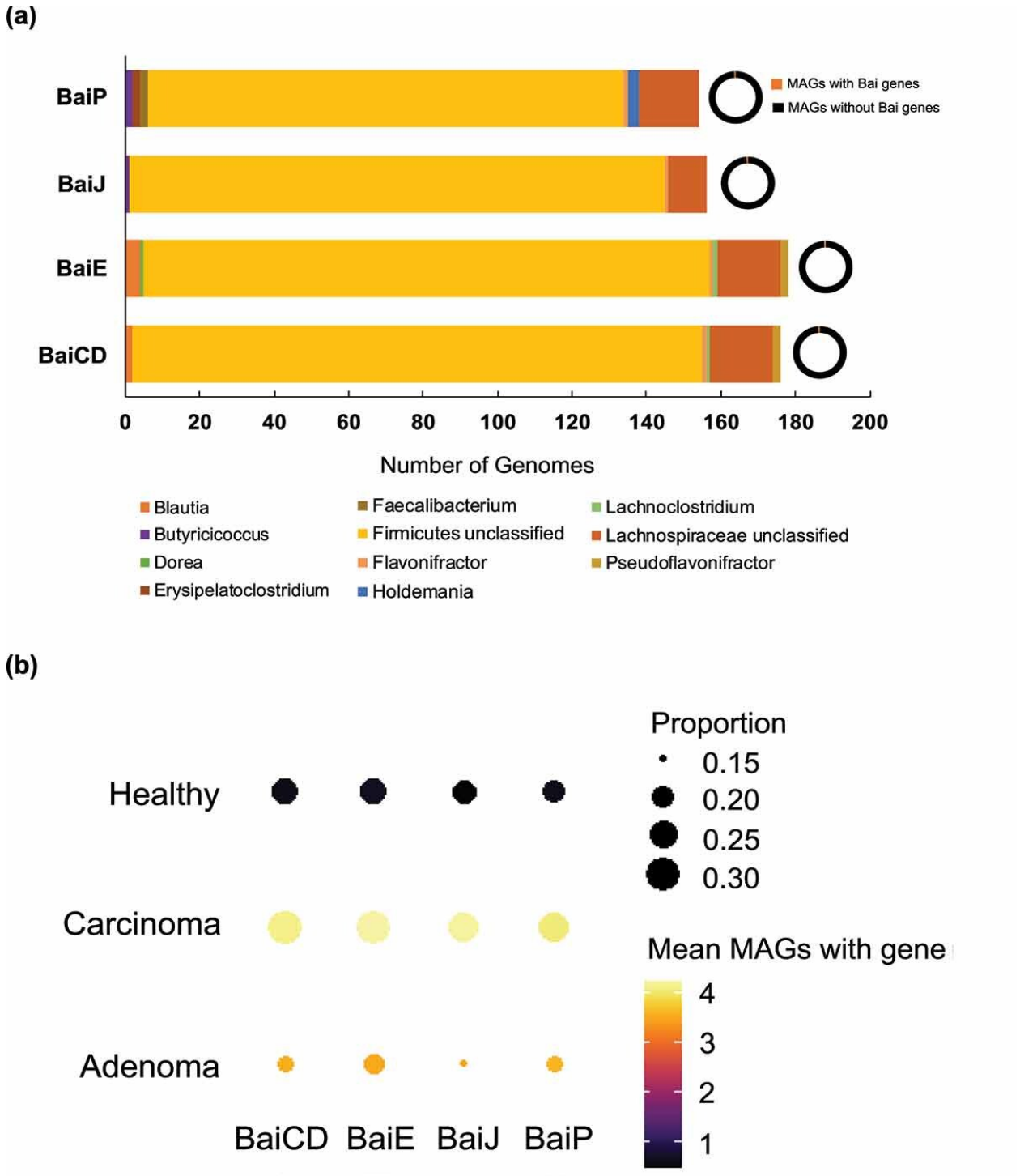

(a)
BaiP
BaiJ
BaiE
BaiCD
MAGs with Bai genes
MAGs without Bai genes
0
20
40
60
80
100
120
140
160
180
200
Number of Genomes
Blautia
Butyricicoccus
Dorea
Erysipelatoclostridium
Faecalibacterium
Firmicutes unclassified
Flavonifractor
Holdemania
Lachnoclostridium
Lachnospiraceae unclassified
Pseudoflavonifractor
(b)
Healthy
Carcinoma
Adenoma
BaiCD
BaiE
BaiJ
BaiP
Proportion
0.15
0.20
0.25
0.30
Mean MAGs with gene
4
3
2
1

Figure 5. Hidden-Markov Model search reveals enrichment of bai genes in colorectal carcinoma. (a) Distribution of microbial genomes with putative 5α-reductase genes (*baiP* and *baiJ*) present across the five metagenomic studies. (b) Dot plots of selected genes related to allo-bile acids production across three disease states: carcinoma, adenoma, and healthy. The size of each dot indicates the proportion of participants with at least one copy of the gene in their bacterial metagenomic assembled genomes (MAGs) and the color of each dot indicates the mean number of MAGs with that gene in the subset of participants that have at least one copy of the gene.

Discussion

The results of the current study add to a growing literature demonstrating that the colonic microbes are capable of "resetting" stereochemistry of sterols undergoing enterohepatic circulation through expression of 5α-reductase and 5β-reductase enzymes. So far, two mechanisms have been identified: (1) A direct mechanism whereby bacteria encoding the multi-step bile acid 7α-dehydroxylation pathway convert primary bile acids to either secondary bile acids via BaiCD/BaiN or as shown herein secondary allo-bile acids via BaiP/BaiJ activities; and (2) an indirect mechanism in which certain species of Bacteroidetes convert 5β-secondary bile acids DCA and LCA to 3-oxo-Δ^4-intermediates, followed by reduction to secondary allo-bile acids.[38] The current work is thus a significant advance toward determining the enzymatic basis for the formation of secondary allo-bile acids by the gut microbiome (Figure 6).

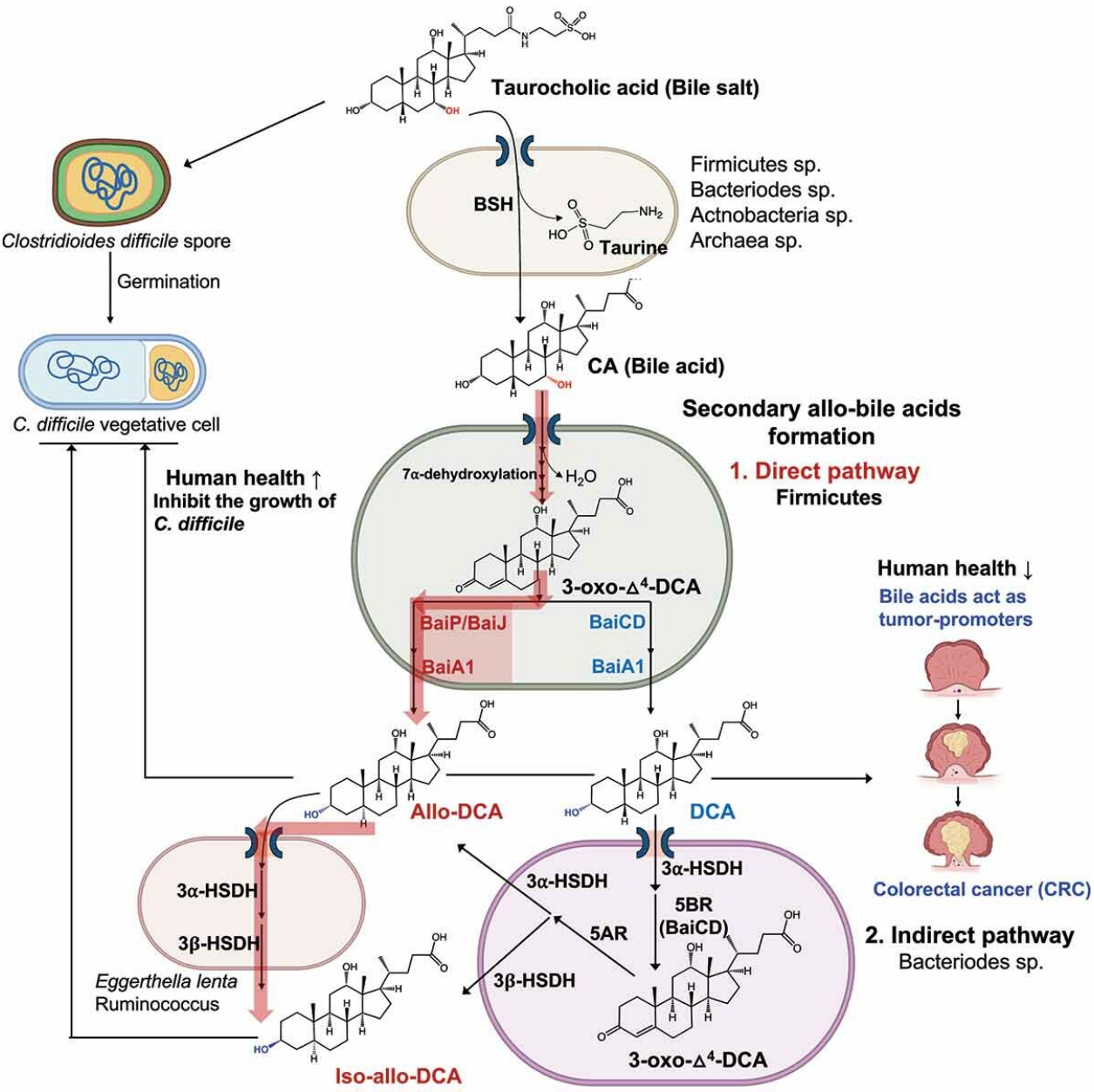


Figure 6. Direct and indirect formation of secondary allo-bile acids, and their potential consequences. Taurocholic acid is deconjugated, mainly in the large intestine, by diverse gut microbial taxa. Free cholic acid is imported into a few species of Firmicutes that harbor the *bai* regulon. **Direct Pathway**: After several oxidative steps, and rate-limiting 7α-dehydration, 3-oxo-Δ^4-DCA becomes a substrate for BaiCD forming

DCA or BaiP/BaiJ forming alloDCA. **Indirect Pathway**: DCA is imported into Bacteroidetes strains that express 3α-HSDH and 5β-reductase (5BR) which converts DCA to 3-oxo-Δ^4-DCA. Expression of 5α-reductase (5AR) and 3β-HSDH sequentially reduce 3-oxo-Δ^4-DCA to iso-allo-DCA. Alternatively, allo-DCA generated by Firmicutes can be isomerized to iso-allo-DCA by species expressing 3α-HSDH and 3β-HSDH such as *Eggerthella lenta*. While taurocholic acid is a germination factor for *C. difficile, s*econdary bile acids such as DCA and secondary allo-bile acids are inhibitory toward *C. difficile* vegetative cells in the GI tract. Secondary bile acids, including DCA and allo-DCA, are associated with increased risk of colorectal cancer (CRC).

Bile acid intermediates in the 7α-dehydroxylation pathway have been determined previously. Björkhem et al.[51] utilized [3β-^{3}H] [24–^{14}C] and [5β-^{3}H] [24–^{14}C] labeled cholic acid in whole cells and cell extracts of *L. scindens* VPI 12708, observing loss of both 3β- and 5β-hydrogens during conversion of CA to DCA.[33] Administration of [3β–^{3}H] [24–C^{14}C] CA and [5β–^{3}H] [24–C^{14}C] CA to volunteers followed by analysis of tritium loss after extraction from duodenal aspirates confirmed that 3–oxo–Δ^4–bile acid intermediates were formed during conversion of CA to DCA.[33] Subsequent work incubating [24–^{14}C] CA with cell extracts of *L. scindens* VPI 12708 revealed a multi-enzyme pathway necessary to convert CA to DCA (and CDCA to LCA).[52] Hylemon and Björkhem (1991) isolated nine [24–^{14}C] CA intermediates after incubation with cell-free extracts of CA-induced whole cells of *L. scindens* VPI 12708 providing the biochemical framework to search for enzymes involved in bile acid 7α-dehydroxylation.[33] Subsequent work determined that bile acid 7α-dehydroxylation proceeds by two

oxidation steps yielding a 7α-hydroxy-3-oxo-Δ^4-intermediate, the substrate for the rate-limiting enzyme, bile acid 7α-dehydratase (BaiE).[34,41,53] Removal of the C7-hydroxyl yields a 7-deoxy-3-oxo-Δ^4-intermediate which is then reduced by flavoproteins BaiN[50] or BaiH[35] to a 7-deoxy-3-oxo-Δ^4-intermediate. The BaiCD and BaiA isoforms then convert 7-deoxy-3-oxo-Δ^4-intermediates to DCA or LCA.[35,53] One of the bile acid-inducible [24–^{14}C] CA metabolites identified was [24–^{14}C] allo-DCA, indicating that *L. scindens* possesses an enzyme with bile acid 5α-reductase distinct from BaiCD (bile acid 5β-reductase).[33] The current results establish conclusively that the *baiP* and *baiJ* genes encode bile acid 5α-reductases in different strains of *L. scindens* and related Firmicutes that catalyze the formation of allo-DCA and allo-LCA.

Previous work also demonstrated that BaiA1 and BaiA2 catalyze both the initial oxidation and final reduction in the formation of DCA and LCA.[35,48] However, a recent report named a gene (*CLOSCI_00522*) adjacent to *baiN*, the "*baiO*" that encodes a predicted 61 kDa flavin-dependent dehydrogenase proposed to catalyze the final reductive step in the pathway.[31] We tested both BaiA1 and BaiO for reduction of allo-DCA and allo-LCA. While the function of *CLOSCI_00522* in bile acid metabolism remains unclear, our results have extended the functional role of the BaiA1. We determined for the first time that this enzyme converts 3-oxo-allo-DCA and 3-oxo-allo-LCA to allo-DCA and allo-LCA, respectively.

The functional role of the previously reported *baiJKL* operon in *L. scindens* VPI 12708 and *L. hylemonae* DSM 15053 has also been extended by the current study.[40] Ridlon

and Hylemon (2012) reported that BaiK and BaiF catalyze bile acid~CoA transferase from secondary bile acids, including allodeoxycholyl~SCoA, to primary bile acids.[40] The *baiJ* gene was annotated as “flavin-dependent fumarate reductase” and “3-ketosteroid-Δ^1-dehydrogenase”, and is co-expressed with *baiKL* under the control of the conserved *bai* promoter.[40] We previously observed bile acid induction of *baiJKL* genes by RT-PCR[40] and RNA-Seq[54] in *L. hylemonae* DSM 15053. Also, the *baiJ* gene was reported to be enriched in the gut microbiome in mouse models of liver cancer and CRC.[9,24] Fecal secondary allo-bile acids have also been reported to be enriched in GI cancers.[47] Phylogenetic analysis of BaiP from *L. scindens* ATCC 35704 revealed two clusters harboring Firmicutes encoding the *bai* pathway, many of which, such as *P. hiranonis, L. hylemonae*, and strains of *L. scindens*, are known to convert CA and CDCA to DCA and LCA, respectively. These clusters are also represented by taxa such as *Dorea* sp. D27, *P. sphenisci*, and *Oscillospiraceae* MAGs whose genome sequences contain *bai* operons.[21,22] Clusters with more distant homologs of BaiP are also worth examining in future studies for novel bile acid 3-oxo-Δ^4-reductases. Mining human metagenomic datasets for “core” Bai proteins (BaiCD, BaiE) as well as BaiP and BaiJ sequences confirmed that these enzymes are only encoded in Firmicutes. Roughly a third of healthy, adenoma, and carcinoma subjects had detectable BaiE enzymes representing ~1% of MAGs. A combination of low abundance bile acid 7-dehydroxylating Firmicutes and stringency of the HMM search likely explains the low representation of subjects with detectable Bai enzymes. Intriguingly, and in line with previous reports,[24] Bai enzymes are enriched in CRC subjects relative to healthy subjects.

There is a paucity of studies on secondary allo-bile acids, and the literature which exists is conflicting as to whether to regard these hydrophobic “flat” bile acids as beneficial, disease promoting, or contextually important.[36–38,47] Recent work measured the secondary allo-bile acid iso-allo-LCA in fecal samples at an average concentration of ~20 μM, and that low micromolar levels, such as those achieved in our resting cell assays, inhibit the growth of gram-positive pathogens including *Clostridioides difficile*[38] (Figure 6). There is a recent growing interest in the immune mechanisms of action of secondary bile acid derivatives and isomers in the colon. Secondary bile acid derivatives, including 3-oxo-DCA, 3-oxo-LCA, iso-DCA (3β, 12α-dihydroxy-5β-cholan-24-oic acid), iso-LCA (3β-hydroxy-5β-cholan-24-oic acid), and certain secondary allo-bile acids (e.g. iso-allo-LCA: 3β-hydroxy-5α-cholan-24-oic acid), regulate the balance of regulatory T cells (Treg) and pro-inflammatory T_H17 cells by promoting expansion of Tregs.[55–57] The current work is thus an important contribution in a rapidly evolving area of the role of diverse bile acid metabolites generated by the gut microbiome on mechanisms underlying host health and disease.

Materials and methods

Bacterial strains and chemicals

E. coli Top10 [F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G] competent cells from Invitrogen (Carlsbad, CA, USA) were used for manipulation of plasmids, and *E. coli* BL21(DE3) [F−, *ompT, hsdSB*(rB− mB−), *gal, dcm, rne131* (DE3)] was also purchased from

Invitrogen and used for protein expression. 3-oxo-Δ^4-LCA, 3-oxo-allo-LCA, 3-oxo-LCA, allo-LCA, LCA, and 3-oxo-DCA were purchased from Steraloids (Newport, RI, USA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Gold Biotechnology (St. Louis, MO, USA). All other reagents were of the highest possible purity and purchased from Fisher Scientific (Pittsburgh, PA, USA).

Bile acid synthesis

Authentic 3-oxo-Δ^4-DCA and allo-DCA were synthesized as previously described[58] and confirmed by nuclear magnetic resonance (NMR) spectroscopy (**Fig. S3, S4**).

Cloning of *bai* operon genes from *L. scindens* strains

The strains/plasmids, primers, and synthetic DNA sequences used in this study are listed in **Table S1, S2**, and **S3**, respectively. First, *baiP* gene encoding FAD-dependent oxidoreductase and *baiA1* gene encoding 3α-HSDH from *L. scindens* ATCC 35704, *baiJ* gene encoding FAD-dependent oxidoreductase from *L. scindens* VPI 12708, and *baiO* encoding a predicted 61 kDa flavin-dependent dehydrogenase were codon-optimized for *E. coli* and synthesized using gBlocks service from Integrated DNA Technologies (IDT, IA, USA). To construct a BaiP, BaiJ, BaiO or BaiA1 expression plasmid (pBaiP, pBaiJ, pBaiO or pBaiA1), a DNA fragment (vector fraction) was amplified from the pETduet plasmid using a primer pair of V1-F and V1-R, V1-F and V1-R, V1-F and V1-R, or V2-F and V2-R, respectively. Another DNA fragment (insert fraction) was amplified from the synthetic oligomers of BaiP, BaiJ, BaiO or BaiA1 using a primer pair of BaiP-F and BaiP-R, BaiJ-F and BaiJ-R, BaiO-F and BaiO-R or BaiA1-F

and BaiA1-R, respectively. The two pairs of PCR products were ligated together by *in vitro* homologous recombination using a Gibson assembly cloning kit (NEB, Boston, MA, USA), respectively. For construction of a BaiP and BaiA1 co-expression plasmid (pBaiP-A1), a DNA fragment (vector fraction) was amplified from the pBaiP plasmid using a pair of the primers V2-F and V2-R, and another DNA fragment (insert fraction) was amplified from the synthetic oligomer of BaiA1 using a pair of the primers BaiA1-F and BaiA1-R. The two PCR products were ligated together by the Gibson assembly cloning kit (NEB)

Recombinant plasmids (**Table S1**) were transformed into chemically competent *E. coli* Top10 cells via heat-shock method, respectively, plated, and grown for overnight at 37°C on lysogeny broth (LB) agar plates supplemented with appropriate antibiotics (Ampicillin: 100 μg/ml). A single colony from each transformation was inoculated into LB medium (5 ml) containing the corresponding antibiotic. The cells were subsequently centrifuged (3,220 × g, 10 min, 4°C) and plasmids were extracted from the cell pallets using QIAprep Spin Miniprep kit (Qiagen, CA, USA). The sequences of the inserts were confirmed by Sanger sequencing (ACGT Inc, Wheeling, IL, USA).

Heterologous expression and purification of Bai enzymes in *E. coli*

For protein expression, the extracted recombinant plasmids were transformed into *E. coli* BL21(DE3) cells by use of electroporation method, respectively, and cultured overnight at 37°C on LB agar plates supplementary with appropriate antibiotics. Selected colonies were inoculated into 10 mL of LB medium containing the corresponding antibiotic and grown at 37°C for 6 h with vigorous aeration. The pre-

cultures were added to fresh LB medium (1 L), supplemented with appropriate antibiotics, and aerated at 37°C until reaching an OD_{600} (optical density of a sample measured at a wavelength of 600 nm) of 0.3. IPTG was added to each culture at a final concentration of 0.1 mM to induce and the temperature was decreased to 16°C. Following 16 h of culturing, cells were pelleted by centrifugation (4000 × g, 30 min, 4°C) and resuspended in 30 ml of binding buffer (20 mM Tris-HCl, 300 mM NaCl, 10 mM 2-mercaptoethanol, pH 7.9). The cell suspension was subjected to an ultra sonicator (Fisher Scientific) and the cell debris was separated by centrifugation (20,000 × g, 40 min, 4°C).

The recombinant protein in the soluble fraction was then purified using TALON Metal Affinity Resin (Clontech Laboratories, CA, USA) per manufacturer's protocol. The recombinant protein was eluted using an elution buffer composed of 20 mM Tris-HCl, 300 mM NaCl, 10 mM 2-mercaptoethanol, and 250 mM imidazole at pH 7.9. The resulting purified protein was analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Whole cell bile acid conversion assay

E. coli BL21(DE3) strains harboring the constructed plasmids were cultured aerobically at 25°C on LB medium (10 mL) supplementary with appropriate antibiotics and expressed the corresponding proteins by IPTG induction at 25°C. Following 16 h of culturing, the strains were pelleted by centrifugation (3,220 × g, 10 min) and washed twice with anaerobic PBS solution. The washed *E. coli* strains were inoculated along with 50 μM bile acid substrates (3-oxo-Δ^4-LCA, 3-oxo-Δ^4-DCA, or 3-oxo-allo-LCA) into

10 mL of PBS and incubated anaerobically at room temperature for 12 h. The whole cell reaction cultures were centrifuged at 3,220 × g for 10 min to remove bacterial cells and adjusted the pH of the supernatant to pH 3.0 by adding 25 µL of 2 N HCl. Bile acid metabolites were extracted by vortexing with two volumes of ethyl acetate for 1 to 2 min. The organic layer was recovered and evaporated under nitrogen gas. The products were dissolved in 200 µL methanol and analyzed by liquid chromatography-mass spectrometry (LC-MS).

Liquid chromatography-mass spectrometry

LC-MS analysis for all samples was performed using a Waters Acquity UPLC system coupled to a Waters SYNAPT G2-Si ESI mass spectrometer (Milford, MA, USA). For the bile acids as substrates and products of whole cell bioconversion assay by the *E. coli* strains expressing BaiP, BaiJ, or BaiP-A1 enzymes (3-oxo-Δ^4-LCA, 3-oxo-Δ^4-DCA, 3-oxo-LCA, 3-oxo-allo-LCA, 3-oxo-DCA, LCA, allo-LCA, DCA, and allo-DCA) analysis, LC was performed with a Waters Acquity UPLC HSS T3 C18 column (1.8 µm particle size, 2.1 mm × 100 mm) at a column temperature of 40°C. Samples were injected at 0.2 µL. Mobile phase A was a mixture of acetonitrile and methanol (50/50, v/v), and B was 10 mM ammonium acetate. The mobile phase composition was 75% of mobile phase A and 25% of mobile phase B and ran an isocratic mode. The flow rate of the mobile phase was 0.5 mL/min. MS was carried out in negative ion mode with a desolvation temperature of 400°C and desolvation gas flow of 800 L/hr. The capillary voltage was 2,000 V. Source temperature was 120°C, and the cone voltage was 30 V. Chromatographs and mass spectrometry data were analyzed using Waters MassLynx

software. Analytes were identified according to their mass and retention time. For quantification of 3-oxo-allo-LCA produced by the *E. coli* BL21(DE3) expressing BaiP/BaiJ strains, a standard curve was obtained, and then 3-oxo-allo-LCA was quantified based on the standard curve (**Figure S5**). The limit of detection (LOD) for 3-oxo-Δ^4-LCA, 3-oxo-allo-LCA, and allo-LCA was 0.1 μmol/L.

For the cortisol and 11β-OHAD as substrates and products of whole cell bioconversion assay by the *E. coli* strain expressing BaiP enzyme analysis, LC was performed with a Waters Acquity UPLC BEH C18 column (1.7 μm particle size, 2.1 mm × 50 mm) at a column temperature of 40°C. Samples were injected at 0.2 μL. Mobile phase A was a mixture of 95% water, 5% acetonitrile, and 0.1% formic acid, and B was a mixture of 95% acetonitrile, 5% water, and 0.1% formic acid. The mobile phase gradient was as follows: 0 min 100% mobile phase A, 0.5 min 100% A, 6.0 min 30% A, 7.0 min 0% A, 8.1 min 100% A, and 10.0 min 100% A. The flow rate of the mobile phase was 0.5 mL/min. MS was carried out in positive ion mode with a desolvation temperature of 450°C and desolvation gas flow of 800 L/hr. The capillary voltage was 3,000 V. Source temperature was 120°C, and the cone voltage was 30 V.

NMR spectroscopy

To determine the molecular structure of the chemically synthesized 3-oxo-Δ^4-DCA and allo-DCA at the atomic level, NMR spectroscopy was performed.^{1}H-NMR spectra were recorded on a JNM-ECA-500 spectrometer (JEOL Co., Tokyo, Japan) at 500 MHz, with pyridine-D_5 as the solvent. Chemical shifts are given as the δ-value with

tetramethylsilane (TMS) as an internal standard. The abbreviation used here: s, singlet; d, doublet; bs, broad singlet.

Phylogenetic analysis

Sequences for phylogenetic analyses were retrieved from NCBI's NR protein database using the sequence of HDCHBGLK_03451 as the query and limiting the number of resulting database matches to five thousand and allowing a maximum alignment E-value of 1E-10 for BLASTP v. 2.12.0 +.[59] The retrieved alignments showed high sequence conservation, therefore the worst E-value seen in the alignments was about 3E-37.

Given the high sequence similarities observed in the search step, sequences were clustered with USEARCH v. 11.0.667[60] to remove redundancy from the dataset. The cluster_fast command was used with an identity threshold of at least 95% to cluster sequences. Each cluster was represented in the phylogenetic analysis by one representative, the centroid sequence. The only exception was the sequences in the same cluster as the query sequence used above, in which case all sequences from the cluster were used in the analysis, instead of just the centroid. Clustering resulted in 1,603 sequences included in the downstream analyses.

Centroids 25% shorter or longer than the average sequence length calculated for the whole dataset (596 amino acids) were removed from the dataset, thus keeping in the analysis only sequences with at least 446 and at most 744 amino acids in length. The 1,460 protein sequences remaining in the dataset were aligned by MUSCLE v. 3.8.1551[61] and the best-fitting sequence substitution model was identified using

ModelTest-NG v. 0.1.7.[62] Phylogenetic tree inference was performed using the maximum likelihood criterion as implemented by RAxML v. 8.2.12,[63] using the WAG sequence substitution model with empirical residue frequencies, gamma-distributed substitution rates, and bootstrap pseudoreplicates (whose number, 250, was determined automatically by the program at run-time). The resulting phylogenetic tree was edited with TreeGraph2 v. 2.15.0–887[64] and Dendroscope v. 3.7.6[65] and further cosmetic adjustments were performed with the Inkscape vector editor (https://inkscape.org/ last accessed on January, 20th, 2022).

Bai gene identification in MAG database

A database of publicly available MAGs from five cohorts varying in CRC status was previously annotated for open reading frames and used for this study.[66,67] Custom Hidden Markov Model (HMM) profiles were created for each of the 4 genes of interest (*baiCD, baiE, baiP*, and *baiJ*) by creating an alignment of reference protein sequences in this study and blastp results with 60% identity to those reference sequences and then passing the alignments to hmmbuild to create an HMM profile. Initial HMM cutoffs were generated by querying protein sequences from the Human Microbiome Project.[66] To further refine HMM profile cutoffs, blast databases were made of each alignment and a concatenated file of predicted open reading frames from the 16,936 MAGs described earlier were queried against the alignment databases. The MAG database was searched using the HMM profiles with finalized cutoffs and hmmsearch within HMMER 3.1b2. All custom HMM profiles used for these searches can be found at: https://github.com/escowley/BileAcid_LeeJ.

Summary calculations and statistical analysis for association of Bai genes with disease state from MAG database

Summary calculations of number of gene hits in the MAG database, number of participants with the gene of interest, and disease information were performed in R and can be found in **Table S4**. Methods for determining associations between Bai genes and disease state were previously described.[66] Briefly, chi squared tests were performed on a dataset of binarized participants that were designated as "presence" if any of their MAGs contained a copy of the gene of interest or "absence" if none of their recovered MAGs contained a copy of the gene of interest. P-values less than 0.05 are designated as significant (**Table S5**).

Acknowledgments

E.S.C. is a Medical Scientist Training Program (MSTP) student and was supported by a National Library of Medicine training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison, and in part by MSTP grant T32 GM140935. P.G.W. was supported at UI-Chicago by the Cancer Education and Career Development Program grant T32CA057699. L.K.L. was supported by NSF GRFP 2017224867. H.L.D. was supported by the David H. and Norraine A. Baker Graduate Fellowship in Animal Sciences at Illinois.

Funding Statement

This work was supported by National Institutes of Health grants (1RO1 CA204808-01 [J.M.R., H.R.G.], R01 GM134423-01A1 [J.M.R.], R03 AI147127-01A1 [J.M.P.A, J.M.R.]) as well as UIUC Department of Animal Sciences Matchstick grant, and Hatch ILLU-538-916.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The code and HMM profiles used for this study are openly available at: https://github.com/escowley/BileAcid_LeeJ.

Supplementary material

Supplemental data for this article can be accessed online at https://doi.org/10.1080/19490976.2022.2132903

References

1. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. Arch Intern Med. 1999;159(22):2647–19. doi: 10.1001/archinte.159.22.2647. [PubMed] [CrossRef] [Google Scholar]

2. Midtvedt T. Microbial bile acid transformation. Am J Clin Nutr. 1974;27(11):1341–1347. doi: 10.1093/ajcn/27.11.1341. [PubMed] [CrossRef] [Google Scholar]

3. Swann JR, Want EJ, Geier FM, Spagou K, Wilson ID, Sidaway JE, Nicholson JK, Holmes E. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. Proc Natl Acad Sci. 2011;108(supplement_1):4523–4530. doi: 10.1073/pnas.1006734107. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

4. Narushima S, Itoha K, Miyamoto Y, Park SH, Nagata K, Kuruma K, Uchida K. Deoxycholic acid formation in gnotobiotic mice associated with human intestinal bacteria. Lipids. 2006;41(9):835–843. doi: 10.1007/s11745-006-5038-1. [PubMed] [CrossRef] [Google Scholar]

5. Gustafsson BE, Midtvedt T, Norman A. Metabolism of cholic acid in germfree animals after the establishment in the intestinal tract of deconjugating and 7 alpha-dehydroxylating bacteria. Acta Pathol Microbiol Scand. 1968;72(3):433–443. doi: 10.1111/j.1699-0463.1968.tb00457.x. [PubMed] [CrossRef] [Google Scholar]

6. Lan K, Su M, Xie G, Ferslew B C, Brouwer K L, Rajani C, Liu C, Jia W. Key role for the 12-hydroxy group in the negative ion fragmentation of unconjugated C24 bile acids. Anal Chem. 2016;88(14):7041–7048. doi: 10.1021/acs.analchem.6b00573 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Quinn, RA, Melnik, AV, Vrbanac, A, Fu, T, Patras, KA, Christy, LK, Bodai, Z, Belda-Ferre, Pedro, Tripathi, Anupriya, Chung, L K, et al. Global chemical effects of the microbiome include new bile-acid conjugations. Nature. 2020;579(7797):123–129 doi: 10.1038/s41586-020-2047-9 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

8. Zhou H, Hylemon P B. Bile acids are nutrient signaling hormones. Steroids. 2014;86:62–68. doi: 10.1016/j.steroids.2014.04.016 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

9. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature. 2013;499(7456):97–101. doi: 10.1038/nature12347. [PubMed] [CrossRef] [Google Scholar]

10. Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, Agdashian D, Terabe M, Berzofsky J A, Fako V, et al. Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells. Science. 2018;360(6391):eaan5931. doi: 10.1126/science.aan5931. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

11. Ocvirk S, O'Keefe SJ. Influence of bile acids on colorectal cancer risk: potential mechanisms mediated by diet gut microbiota interactions. Current Nutrition Reports. 2017;6(4):315–322. doi: 10.1007/s13668-017-0219-5. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

12. Chaudhari SN, Harris DA, Aliakbarian H, Luo JN, Henke MT, Subramaniam R, Vernon AH, Tavakkoli A, Sheu EG, Devlin AS, et al. Bariatric surgery reveals a gut-restricted TGR5 agonist with anti-diabetic effects. Nat Chem Biol. 2021;17(1):20–29. doi: 10.1038/s41589-020-0604-z. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

13. Chiang JYL, Ferrell JM. Discovery of farnesoid X receptor and its role in bile acid metabolism. Mol Cell Endocrinol. 2022;548:111618. doi: 10.1016/j.mce.2022.111618. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

14. Berr F, Kullak-Ublick GA, Paumgartner G, Münzing W, Hylemon PB. 7 alpha-dehydroxylating bacteria enhance deoxycholic acid input and cholesterol saturation of

bile in patients with gallstones. Gastroenterology. 1996;111(6):1611–1620. doi: 10.1016/S0016-5085(96)70024-0. [PubMed] [CrossRef] [Google Scholar]

15. Wells JE, Berr F, Thomas LA, Dowling RH, Hylemon PB. Isolation and characterization of cholic acid 7alpha-dehydroxylating fecal bacteria from cholesterol gallstone patients. Journal of Hepatology. 2000;32(1):4–10. doi: 10.1016/S0168-8278(00)80183-X. [PubMed] [CrossRef] [Google Scholar]

16. Marksteiner J, Blasko I, Kemmler G, Koal T, Humpel C. Bile acid quantification of 20 plasma metabolites identifies lithocholic acid as a putative biomarker in Alzheimer's disease. Metabolomics. 2018;14(1):1. doi: 10.1007/s11306-017-1297-5. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

17. MahmoudianDehkordi S, Arnold M, Nho K, Ahmad S, Jia W, Xie G, Louie G, Kueider-Paisley A, Moseley MA, Thompson JW, et al. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease - an emerging role for gut microbiome. Alzheimers Dement. 2019;15(1):76–92. doi: 10.1016/j.jalz.2018.07.217. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

18. Porez G, Prawitt J, Gross B, Staels B. Bile acid receptors as targets for the treatment of dyslipidemia and cardiovascular disease. J Lipid Res. 2012;53(9):1723–1737. doi: 10.1194/jlr.R024794. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

19. Jones, BV, Begley, M, Hill, C, Gahan, CGM, Marchesi, JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc Natl Acad Sci. 2008;105(36):13580–13585 doi: 10.1073/pnas.0804437105 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

20. Ridlon, JM, Harris, SC, Bhowmik, S, Kang, D-J, Hylemon, PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes. 2016;7(1): 22–39. doi: 10.1080/19490976.2015.1127483 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

21. Kim KH, Park D, Jia B, Baek JH, Hahn Y, Jeon CO. Identification and characterization of major bile acid 7α-dehydroxylating bacteria in the human gut. mSystems 2022:e0045522. [PMC free article] [PubMed]

22. Vital M, Rud T, Rath S, Pieper DH, Schlüter D. Diversity of bacteria exhibiting bile acid-inducible 7α-dehydroxylation genes in the human gut. Comput Struct Biotechnol J. 2019;17:1016–1019. doi: 10.1016/j.csbj.2019.07.012. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

23. Hale VL, Chen J, Johnson S, Harrington SC, Yab TC, Smyrk TC, Nelson H, Boardman LA, Druliner BR, Levin TR, et al. Shifts in the fecal microbiota associated with adenomatous polyps. Cancer Epidemiol Biomarkers Prev. 2017;26(1):85–94. doi: 10.1158/1055-9965.EPI-16-0337. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

24. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25(4):679–689. doi: 10.1038/s41591-019-0406-6. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

25. Ochsenkühn T, Bayerdörffer E, Meining A, Schinkel M, Thiede C, Nüssler V, Sackmann M, Hatz R, Neubauer A, Paumgartner G, et al. Colonic mucosal proliferation

is related to serum deoxycholic acid levels. Cancer. 1999;85(8):1664–1669. doi: 10.1002/(SICI)1097-0142(19990415)85:8<1664::AID-CNCR4>3.0.CO;2-O. [PubMed] [CrossRef] [Google Scholar]

26. van Faassen A, Ochsenkühn T, Houterman S, van der Ploeg EM, Bueno-de-Mesquita BH, Ocké MC, Bayerdörffer E, Janknegt RA. Plasma deoxycholic acid is related to deoxycholic acid in faecal water. Cancer Lett. 1997;114(1–2):293–294. doi: 10.1016/S0304-3835(97)04683-1. [PubMed] [CrossRef] [Google Scholar]

27. Bayerdörffer E, Mannes GA, Ochsenkühn T, Dirschedl P, Wiebecke B, Paumgartner G. Unconjugated secondary bile acids in the serum of patients with colorectal adenomas. Gut. 1995;36(2):268–273. doi: 10.1136/gut.36.2.268. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

28. Bayerdörffer E, Mannes GA, Richter WO, Ochsenkühn T, Wiebecke B, Köpcke W, Paumgartner G. Increased serum deoxycholic acid levels in men with colorectal adenomas. Gastroenterology. 1993;104(1):145–151. doi: 10.1016/0016-5085(93)90846-5. [PubMed] [CrossRef] [Google Scholar]

29. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K, Gaskins HR, O'Keefe SJ. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. Am J Clin Nutr. 2013;98(1):111–120. doi: 10.3945/ajcn.112.056689. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

30. Ocvirk S, Wilson AS, Posma JM, Li JV, Koller KR, Day GM, Flanagan CA, Otto JE, Sacco PE, Sacco FD; Ocvirk S, Wilson AS, Posma JM, Li JV, Koller KR, Day GM, et al . A prospective cohort analysis of gut microbial co-metabolism in Alaska native and rural African people at high and low risk of colorectal cancer. Am J Clin Nutr.

2020;111(2):406–419. doi: 10.1093/ajcn/nqz301. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

31. Heinken A, Ravcheev D A, Baldini F, Heirendt L, Fleming R M, Thiele I. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome. 2019;7(1):75. doi: 10.1186/s40168-019-0689-3. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

32. Fiorucci, S, Carino, A, Baldoni, M, Santucci, L, Costanzi, E, Graziosi, L, Distrutti, E, Biagioli, M. Bile acid signaling in inflammatory bowel diseases. Dig Dis Sci. 2021;66(3):674–693. doi: 10.1007/s10620-020-06715-3. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

33. Hylemon PB, Melone PD, Franklund CV, Lund E, Björkhem I. Mechanism of intestinal 7 alpha-dehydroxylation of cholic acid: evidence that allo-deoxycholic acid is an inducible side-product. J Lipid Res. 1991;32(1):89–96. doi: 10.1016/S0022-2275(20)42247-3. [PubMed] [CrossRef] [Google Scholar]

34. Bhowmik S, Chiu H-P, Jones DH, Chiu H-J, Miller MD, Xu Q, Farr, C L, Ridlon, J M, Wells, J E, Elsliger, M-A, et al. Structure and functional characterization of a bile acid 7α dehydratase BaiE in secondary bile acid synthesis. Proteins. 2016;84(3):316–331. doi: 10.1002/prot.24971. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

35. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA, et al. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. Nature. 2020;582(7813):566–570. doi: 10.1038/s41586-020-2396-4. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

36. Tadano T, Kanoh M, Matsumoto M, Sakamoto K, Kamano T. Studies of serum and feces bile acids determination by gas chromatography-mass spectrometry. Rinsho Byori. 2006;54:103–110. [PubMed] [Google Scholar]

37. Tadano T, Kanoh M, Kondoh H, Matsumoto M, Mimura K, Kanoh Y, Sakamoto K, Kamano T. Kinetic analysis of bile acids in the feces of colorectal cancer patients by gas chromatography-mass spectrometry (GC-MS). Rinsho Byori. The Japanese Journal of Clinical Pathology. 2007;55(5):417–427. [PubMed] [Google Scholar]

38. Sato Y, Atarashi K, Plichta DR, Arai Y, Sasajima S, Kearney SM, Suda W, Takeshita K, Sasaki T, Okamoto S, et al. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. Nature. 2021;599(7885):458–464. doi: 10.1038/s41586-021-03832-5. [PubMed] [CrossRef] [Google Scholar]

39. Devendran S, Shrestha R, Alves JMP, Wolf PG, Ly L, Hernandez AG, Méndez-García C, Inboden A, Wiley J, Paul O, et al. *Clostridium scindens* ATCC 35704: integration of nutritional requirements, the complete genome sequence, and global transcriptional responses to bile acids. Appl Environ Microbiol. 2019;85(7):e00052–19. doi: 10.1128/AEM.00052-19. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

40. Ridlon, JM, Hylemon, PB. Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens*, a bile acid 7α-dehydroxylating intestinal bacterium. J Lipid Res. 2012;53(1): 66–76. doi: 10.1194/jlr.M020313 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

41. Dawson JA, Mallonee DH, Björkhem I, Hylemon PB. Expression and characterization of a C24 bile acid 7 alpha-dehydratase from *Eubacterium* spstrain VPI

12708 in *Escherichia coli*. J Lipid Res. 1996;37(6):1258–1267. doi: 10.1016/S0022-2275(20)39155-0. [PubMed] [CrossRef] [Google Scholar]

42. Ridlon JM, Ikegawa S, Alves JM, Zhou B, Kobayashi A, Iida T, Mitamura K, Tanabe G, Serrano M, De Guzman A, et al. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. J Lipid Res. 2013;54(9):2437–2449. doi: 10.1194/jlr.M038869. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

43. Devendran, Saravanan, Mythen, SM, Ridlon, JM. The *desA* and *desB* genes from *Clostridium scindens* ATCC 35704 encode steroid-17,20-desmolase. J Lipid Res. 2018;59(6):1005–1014. doi: 10.1194/jlr.M083949. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

44. Chen X-J, Wang Z-Q, Zhou Z-Y, Zeng N-Y, Huang Q-F, Wang Z-W, Tang W-L, Zhou H-W. Characterization of *Peptacetobacter hominis* gen. nov., sp. nov., isolated from human faeces, and proposal for the reclassification of *Clostridium hiranonis* within the genus *Peptacetobacter*. Int J Syst Evol Microbiol. 2020;70:2988–2997. [PubMed] [Google Scholar]

45. Wells JE, Hylemon PB. Identification and characterization of a Bile Acid 7α-Dehydroxylation operon in *clostridium* sp. Strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces. Appl Environ Microbiol. 2000;66(3):1107–1113. doi: 10.1128/AEM.66.3.1107-1113.2000. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

46. Ridlon, JM, Kang, D-J, Hylemon, PB. Isolation and characterization of a bile acid inducible 7alpha-dehydroxylating operon in *Clostridium hylemonae* TN271. Anerobe.

2010;16(2):137–146. doi: 10.1016/j.anaerobe.2009.05.004. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

47. Shiffka SJ, Kane MA, Swaan PW. Planar bile acids in health and disease. Biochim Biophys Acta Biomembr. 2017;1859(11):2269–2276. doi: 10.1016/j.bbamem.2017.08.019. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

48. Bhowmik S, Jones DH, Chiu H-P, Park I-H, Chiu H-J, Axelrod HL, Farr CL, Tien HJ, Agarwalla S, Lesley SA. Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. Proteins. 2014;82(2):216–229. doi: 10.1002/prot.24353. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

49. Mallonee DH, Lijewski MA, Hylemon PB . Expression in *Escherichia coli* and characterization of a bile acid-inducible 3 alpha-hydroxysteroid dehydrogenase from *Eubacterium sp*. strain VPI 12708. Curr Microbiol. 1995;30(5):259–263. doi: 10.1007/BF00295498. [PubMed] [CrossRef] [Google Scholar]

50. Harris SC, Devendran S, Alves JMP, Mythen SM, Hylemon PB, Ridlon JM. Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium *Clostridium scindens* ATCC 35704. Biochim Biophys Acta Mol Cell Biol Lipids. 2018;1863(3):276–283. doi: 10.1016/j.bbalip.2017.12.001. [PubMed] [CrossRef] [Google Scholar]

51. Björkhem, I, Einarsson, K, Melone, P, Hylemon, PB. Mechanism of intestinal formation of deoxycholic acid from cholic acid in humans: evidence for a 3-oxo-delta 4-

steroid intermediate. J Lipid Res. 1989;30(7):1033–1039. doi: 10.1016/S0022-2275(20)38290-0. [PubMed] [CrossRef] [Google Scholar]

52. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res. 2006;47(2):241–259. doi: 10.1194/jlr.R500013-JLR200. [PubMed] [CrossRef] [Google Scholar]

53. Kang DJ, Ridlon JM, Moore DR 2nd, Barnes S, Hylemon PB. Clostridium scindens baiCD and baiH genes encode stereo–specific 7alpha/7beta–hydroxy–3–oxo–delta4–cholenoic acid oxidoreductases. Biochim Biophys Acta. 2008. Jan-Feb;1781(1-2):16-25. doi: 10.1016/j.bbalip.2007.10.008. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

54. Ridlon JM, Devendran S, Alves JM, Doden H, Wolf PG, Pereira GV, Ly L, Volland A, Takei H, Nittono H, et al. The '*in vivo* lifestyle' of bile acid 7α-dehydroxylating bacteria: comparative genomics, metatranscriptomic, and bile acid metabolomics analysis of a defined microbial community in gnotobiotic mice. Gut Microbes. 2020;11(3):381–404. doi: 10.1080/19490976.2019.1618173. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

55. Song X, Sun X, Oh SF, Wu M, Zhang Y, Zheng W, Geva-Zatorsky N, Jupp R, Mathis D, Benoist C, et al. Microbial bile acid metabolites modulate gut RORγ^{+} regulatory T cell homeostasis. Nature. 2020;577(7790):410–415. doi: 10.1038/s41586-019-1865-0. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

56. Li W, Hang S, Fang Y, Bae S, Zhang Y, Zhang M, Wang G, McCurry MD, Bae M, Paik D, et al. A bacterial bile acid metabolite modulates Treg activity through the nuclear

hormone receptor NR4A1. Cell Host Microbe. 2021;29(1366–77.e9):1366–1377.e9. doi: 10.1016/j.chom.2021.07.013. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

57. Hang S, Paik D, Yao L, Kim E, Trinath J, Lu J, Ha S, Nelson BN, Kelly SP, Wu L, et al. Bile acid metabolites control T_H17 and T_{reg} cell differentiation. Nature. 2019;576(7785):143–148. doi: 10.1038/s41586-019-1785-z. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

58. Leppik RA. Improved synthesis of 3-keto, 4-ene-3-keto, and 4,6-diene-3-keto bile acids. Steroids. 1983;41(4):475–484. doi: 10.1016/0039-128X(83)90087-9. [PubMed] [CrossRef] [Google Scholar]

59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T L. BLAST+: architecture and applications. BMC Bioinform. 2009;10(1):421. doi: 10.1186/1471-2105-10-421. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

60. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–2461. doi: 10.1093/bioinformatics/btq461. [PubMed] [CrossRef] [Google Scholar]

61. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 2004;5(1):113. doi: 10.1186/1471-2105-5-113. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

62. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2019;37:291–294. doi: 10.1093/molbev/msz189. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313. doi: 10.1093/bioinformatics/btu033. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

64. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinform. 2010;11:7. [PMC free article] [PubMed] [Google Scholar]

65. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012;61(6):1061–1067. doi: 10.1093/sysbio/sys062. [PubMed] [CrossRef] [Google Scholar]

66. Wolf PG, Cowley ES, Breister A, Matatov S, Lucio L, Polak P, Ridlon JM, Gaskins HR, Anantharaman K. Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. Microbiome. 2022;10(1):64. doi: 10.1186/s40168-022-01242-x. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

67. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell. 2019;176(3):649–62.e20. doi: 10.1016/j.cell.2019.01.001. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Appendix J – Diversity and ecology of microbial sulfur metabolism

This work has been accepted for publication in *Nature Reviews Microbiology*

Zhichao Zhou[1+], Patricia Q. Tran[1+], **Elise S. Cowley[1,2]**, Elizabeth Trembath-Reichert[3], Karthik Anantharaman[1,4*]

[1] Department of Bacteriology, University of Wisconsin-Madison, USA

[2] Microbiology Doctoral Training Program, University of Wisconsin-Madison, USA

[3] School of Earth and Space Exploration, Arizona State University, USA

[4] Department of Integrative Biology, University of Wisconsin-Madison, USA

Supplemental materials are available from the corresponding author: Karthik Anantharaman (karthik@bact.wisc.edu).

[+]Co-first authors

*Corresponding author: Karthik Anantharaman (karthik@bact.wisc.edu)

Abstract

Sulfur plays a pivotal role in interactions among the atmosphere, lithosphere, pedosphere, hydrosphere, and biosphere, and the functioning of living organisms. Within Earth's crust, mantle, and atmosphere, sulfur undergoes diverse geochemical transformations due to natural and anthropogenic factors. In the biosphere, sulfur is indispensable, participating in the formation of amino acids, proteins, coenzymes, and vitamins. Microorganisms in the biosphere play a key role in the cycling of sulfur compounds through oxidation, reduction, and/or disproportionation reactions for generation of energy and bioassimilation of sulfur compounds. Microbial sulfur metabolism is abundant in both aerobic and anaerobic environments and interconnected with the biogeochemical cycles of important elements such as carbon, nitrogen, and iron. Through metabolism,

competition, or cooperation, microorganisms metabolizing sulfur can drive the consumption of organic carbon, loss of fixed nitrogen, and production of greenhouse gases. Given the increasing significance sulfur metabolism in environmental alteration and the intricate involvement of microorganisms in sulfur dynamics, a timely reevaluation of the sulfur cycle is imperative. This discusses advances in our understanding of microbial sulfur metabolism, provides holistic context for the sulfur cycle in the face of rapidly changing ecosystems on Earth, and examines the significance of microbially-mediated sulfur transformation reactions in different environments, ecosystems, and microbiomes.

Introduction

Sulfur is the fifth most abundant element on Earth. Sulfur is ubiquitous in diverse environments, including marine, terrestrial, soil, freshwater, and the human body. Historically, geological activities constituted the primary sources of sulfur[2], but in recent centuries, anthropogenic activities have exerted a substantial influence on the sulfur cycle[4]. Instances such as acid rain caused by the release of sulfates into aquatic ecosystems[6] and combustion of fossil fuels leading to the release of sulfur dioxide into the atmosphere[1] highlight the impact of human activities on sulfur dynamics. The sulfur cycle encompasses effects on climate change and environmental conservation, and involves economic benefits for human beings, including sulfur utilization[7] and agricultural activities[4]. Comprehending biotic and abiotictransformations of sulfur compounds is crucial for understanding their implications in the overall sulfur cycle, as well as their interactions with other macro- and microorganisms and elemental cycles.

Sulfur can exist in organic and inorganic forms, and in various oxidation states, ranging from highly reduced hydrogen sulfide to highly oxidized sulfate. Inorganic forms of sulfur comprise various sulfides, elemental sulfur, thiosulfate, tetrathionate, sulfite, and sulfate. Hydrogen sulfide is prevalent in anoxic environments, functioning as a reduced compound for microbial energy metabolism. Elemental sulfur commonly occurs in volcanic and hydrothermal settings, and is most commonly sourced from incomplete abiotic or biotic oxidation of hydrogen sulfide. In contrast, sulfate dominates oxic settings like soils and oceans, supporting plant and microbial growth through assimilatory transformations to organic sulfur, and microbial metabolism through dissimilatory transformations to hydrogen sulfide. Organic forms of sulfur encompass sulfur-containing amino acids such as cysteine and methionine, and other important biotic compounds

including amino acid derivatives such as taurine, thioethers such as dimethyl sulfide (DMS), dimethylsulfoniopropionate (DMSP) which serves as a common biological osmolyte, and sulfonates such as methanesulfonate.

The conversions between oxidation states and the various forms of sulfur are often mediated by microorganisms. These microbes make vital contributions to energy flow in the environment and make essential sulfur compounds (including organosulfur compounds) available to higher organisms, such as animals and plants[8]. For example, sulfate-reducing bacteria are proficient in in anoxic environments converting sulfate to hydrogen sulfide[9], whereas sulfur-oxidizing bacteria catalyze the reverse reaction in both anoxic and oxic environments, transforming hydrogen sulfide into sulfate through oxidation[10]. Archaea, frequently inhabiting extreme environments, participate in diverse metabolic pathways contributing to sulfur cycling including sulfate reduction, sulfur oxidation, and sulfur reduction altering the oxidation state of sulfur compounds based on nutrient availability[11,12]. Viruses can infect these microorganisms, controlling their populations and providing auxiliary metabolic functions by manipulating their microbial hosts, and thereby influencing the transformations of sulfur compounds[13,14]. These microbial and viral interactions and processes play an important role in the accessibility of crucial sulfur compounds for organisms spanning various trophic levels.

While elemental cycles of other important elements such as carbon and nitrogen cycles have received extensive attention, the sulfur cycle remains relatively less understood[15]. The sulfur cycle plays an important role in ongoing climate change and shapes interactions between the atmosphere, lithosphere, pedosphere, hydrosphere, and biosphere. Given the specific importance of the microbial sulfur cycle which plays a

central role in these processes on Earth, we aim to enhance our understanding of microbial sulfur metabolism, and examine the microbial diversity and functional roles within sulfur-associated microbiomes and ecosystems.

Global sulfur cycling

The global cycling of sulfur encompasses the dynamic movement of sulfur in various forms across diverse ecosystems and environments on Earth (Fig. 1). Sulfur has its largest reservoirs in the Earth's lithosphere, amounting to $12\pm6\times10^9$ Teragrams (Tg) (mostly as minerals in the crust) and in Earth's hydrosphere to the order of $1.3\pm0.1\times10^9$ Tg (mostly as sulfate in the oceans)[15]. In concert with its abundance, sulfur transformations are among key biological processes, as it is incorporated into living organisms, contributing to the formation of amino acids, vitamins, cofactors, hormones, and other compounds essential for maintaining biological functions[16]. Sulfur is important for agricultural activities and serves as a key fertilizer for plant growth[17]. Additionally, the cycling of sulfur plays a crucial role in maintaining the balance of ecosystems and supporting various essential biological processes.

Sulfur can exist in organic and inorganic forms, each with its unique sources and processes[16]. Inorganic sulfur compounds originate from three major processes – biotic processes, abiotic processes, and human activities. Biotic processes in particular play crucial roles in the formation of inorganic sulfur. In the oceans, algae and phytoplankton produce a secondary metabolite called dimethylsulfoniopropionate (DMSP), which serves as a significant source of dimethyl sulfide (DMS), the most abundant biotic sulfur compound emitted into the atmosphere which is finally oxidized into inorganic sulfur

compounds (i.e., SO_2 and SO_4^{2-})[16]. Additionally, microbial decomposers in soils and during waste processing contribute to the inorganic sulfur supply[18,19], and sulfide minerals can undergo microbial leaching to generate solubilized metals and sulfur compounds[20]. Abiotic processes also contribute significantly to the sulfur cycle. Volcanic activity releases sulfur dioxide and hydrogen sulfide into the atmosphere[1,3]. Geothermally active fields are natural sources of atmospheric hydrogen sulfide[3]. Additionally, oxidative weathering of sulfide minerals buried in sediments can regenerate sulfate and release it into the terrestrial hydrosphere[21]. Human activities are another significant factor in the sulfur budget. The combustion of fossil fuels containing sulfur results in the generation of sulfur dioxide[1]. In industry, high-temperature reactions between sulfur or sulfur-containing compounds and organic materials, such as during petrochemical processing, tanning, and coke production, produce a significant amount of hydrogen sulfide[22].

Sulfur circulates readily and widely among terrestrial, atmospheric, and oceanic environments. Sulfate is transferred to terrestrial soils, freshwater bodies, and oceans through wet/dry atmospheric deposition[5]. Over time, inorganic sulfur is transported to the ocean via runoff, which serves as the main reservoir of inorganic sulfur in the form of sulfate[15]. In soils, sulfate can be absorbed by plants and assimilated into organic sulfur. Within the terrestrial hydrosphere, sulfur mobility varies depending on its forms[23]. Dissolved sulfate in the aqueous phase moves rapidly, while organic sulfur bound in organisms and sulfate bound to particles are less mobile. Microbially-mediated processes in sediments enable the interchangeability of organic sulfur and inorganic sulfur through mineralization and immobilization[24]. Sulfide generated primarily through microbially-mediated sulfate reduction can be deposited in sediments as sulfide minerals[25]. The

deposition of organic sulfur and sulfide minerals are two major sulfur sinks in terrestrial environments.

The dynamic interplay of abiotic and biotic processes collectively contributes to the complexity of the sulfur cycle (Fig. 1). Throughout the subsequent sections, we focus on examining the biochemical basis of sulfur transformations, microbial diversity of sulfur-associated microbiomes, and unraveling the ecological roles of microorganisms in various environments driving the sulfur cycle.

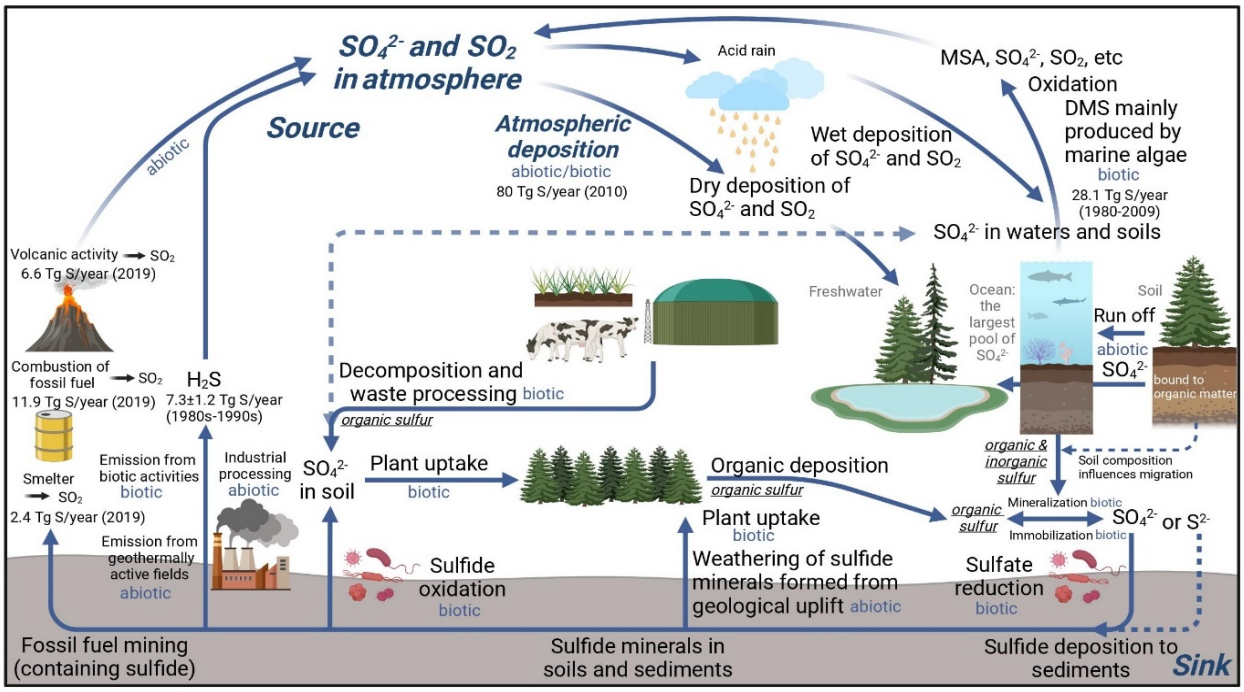


Fig 1. Overview of the global sulfur cycle. This figure illustrates the processes and interactions spanning sulfur transformations in terrestrial, atmospheric, and aquatic environments. Sulfur sources encompass biotic contributions such as decomposition of organisms, uptake by plants, waste processing, and microbial sulfur metabolism (such as sulfide oxidation , sulfate reduction, and generation of DMS in the oceans). Abiotic sources comprise volcanic activity, emissions from geothermally active fields, and weathering of sulfide minerals resulting from geological uplift. Anthropogenic sources involve fossil fuel combustion and various industrial processes. The biotic/abiotic state for each sulfur flux and transformation are labeled accordingly. Sulfur sinks include marine environments hosting the largest pool of SO_4^{2-}, terrestrial environments like soil and freshwater housing the SO_4^{2-} pool and the deposition of organic sulfur and sulfide to terrestrial sediments, and integration into the biosphere through assimilation into proteins and essential biological components. The annual budgets for global sulfur sources and sinks deposited in the atmosphere are labeled accordingly. Numbers for different sulfur species indicate teragrams of sulfur per year. The original dataset source years or decades are labeled in brackets. The SO_2 emission budgets from three major sources were based on Dahiya et al (2020)[1], the total H_2S emission budget was based on Watts (2000)[3], and the total sulfur budget for atmospheric deposition was based on Rubin et al (2022)[5].

Biochemical basis of microbially-mediated sulfur cycling

The microbially-mediated sulfur cycle encompasses oxidation, reduction, and disproportionation reactions (Fig. 2). Within the framework of sulfur cycling, oxidation refers to the elevation of the oxidation state of sulfur-containing substrates (such as in transformation of sulfide (oxidation state of -2) to sulfate (oxidation state of +6)), whereas reduction refers to a reduction in their oxidation state. Disproportionation is characterized

by substrates undergoing simultaneous oxidation and reduction reactions, resulting in products with one of higher oxidation state and the other one of lower oxidation state. Sulfur cycling microorganisms commonly rely on other members of microbiomes to conduct sequential multi-step sulfur transformations[26,27]. The interconnected sulfur metabolism pathways mediated by different members of microbiomes facilitate a synergistic transformation of sulfur compounds for substrate and energy utilization. In this section, we will delve into the enzymatic and chemical nature of sulfur-associated metabolic reactions and describe the common taxa conducting them (Caspi, R. et al.[28] and references therein for the subsections).

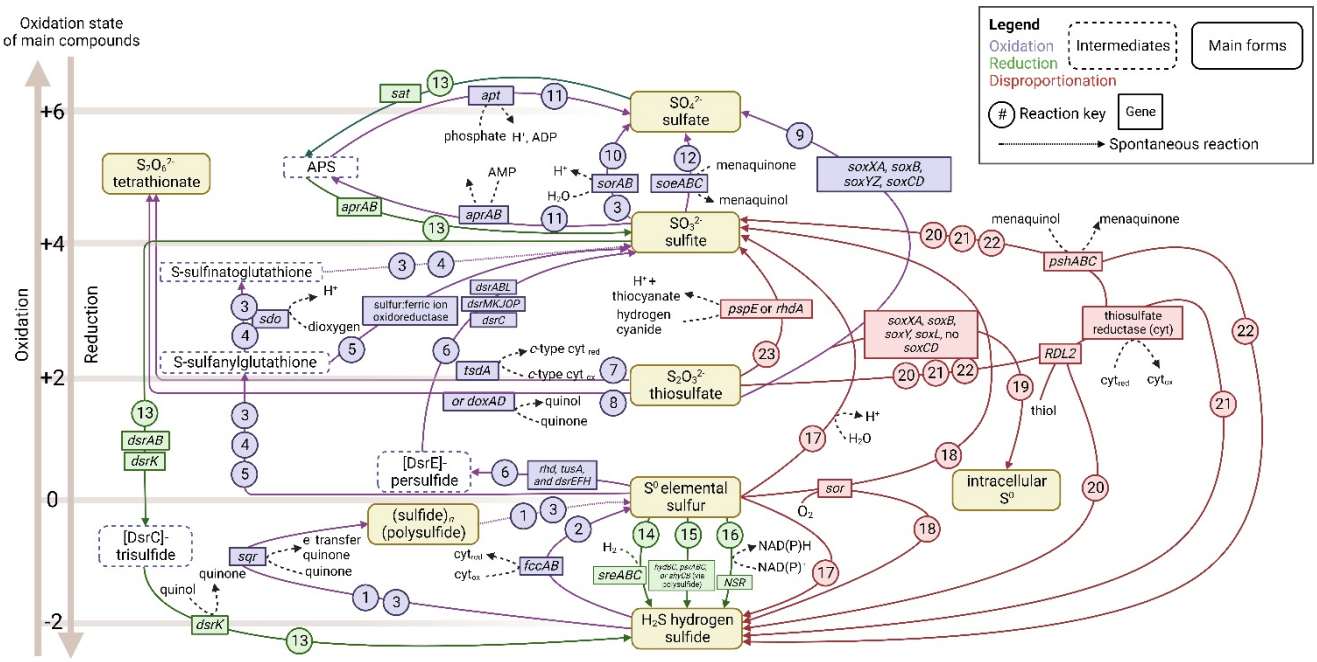


Fig. 2 | Summary figure of sulfur oxidation (purple), reduction (green), and disproportionation (red) reactions. Genes responsible for each reaction are shown in rectangular boxes. The most important oxidation states of sulfur are illustrated on the left. Spanning from the lowest to the highest oxidation states, major oxidation processes encompass sulfide (S^{2-}) oxidation (reactions 1, 2, and 3), elemental sulfur (S^0) oxidation (reactions 4, 5, and 6), thiosulfate ($S_2O_3^{2-}$) oxidation (reactions 7, 8, and 9), and sulfite (SO_3^{2-}) oxidation (reactions 10, 11, and 12). Conversely, major reduction processes include dissimilatory sulfate (SO_4^{2-}) and sulfite (SO_3^{2-}) reduction (reaction 13), and elemental sulfur (S^0) reduction (reactions 14, 15, and 16). Additionally, intermediate oxidation states entail the disproportionation processes, encompassing elemental sulfur disproportionation (reactions 17 and 18) and thiosulfate disproportionation (reactions 17, 19, 20, 21, 22, and 23). Other sulfur compounds represent important intermediates such as polysulfide $(S^{2-})_n$ which participates in the oxidation of hydrogen sulfide to elemental sulfur, tetrathionate ($S_4O_6^{2-}$) which is the product of two thiosulfate oxidation pathways, adenosine phosphosulfate (APS) which is an intermediate in sulfate reduction and sulfur oxidation, and S-sulfinatoglutathione and S-sulfanylglutathione which participate in the oxidation of elemental sulfur. Sulfite is a critical intermediate in sulfur metabolism which participates in oxidation, reduction, and disproportionation pathways. Dashed boxes represent intermediates, and solid boxes represent the main forms of sulfur compounds. Reaction numbers and corresponding details are described in Table S1.

Table 1. Reactions, Genes, and Gibbs free energy of all reactions associated with sulfur metabolism. The full names of genes are listed in Table S1. The term "N/A" indicates that the corresponding $\Delta_rG'°$ value is not available in the literature at present.

Reaction no.	Sulfur metabolism	Reaction	$\Delta_rG'°$ (kcal/mol)	Genes
Reactions 1 and 3	Sulfide oxidation ($S^{2-} \rightarrow S^0$)	n hydrogen sulfide + n electron-transfer quinone → $(sulfide)_n$	16.89	*sqr*, *sdo*, *sorA*, *sorB*

		+ n electron-transfer quinol (main reaction)		
Reaction 2	Sulfide oxidation ($S^{2-} \rightarrow S^0$)	hydrogen sulfide + 2 oxidized *c*-type cytochrome → sulfur(0) + 2 reduced *c*-type cytochrome + 2 H^+	-6.13	*fccAB*
Reaction 4	*Elemental Sulfur oxidation ($S^0 \rightarrow SO_3^{2-}$)*	sulfur(0) + glutathione → *S*-sulfanylglutathione	-2.34	spontaneous
		S-sulfanylglutathione + dioxygen + H_2O → sulfite + glutathione + 2 H^+	-100.75	*sdo*
Reaction 5	*Elemental Sulfur oxidation ($S^0 \rightarrow SO_3^{2-}$)*	sulfur(0) + glutathione → *S*-sulfanylglutathione	-2.34	spontaneous
		S-sulfanylglutathione + 4 Fe^{3+} + 3 H_2O → sulfite + 4 Fe^{2+} + glutathione + 6 H^+	-87.58	sulfur:ferric ion oxidoreductase
Reaction 6	*Elemental Sulfur oxidation ($S^0 \rightarrow SO_3^{2-}$)*	intracellular S^0_n (in sulfur globules) + a thiol → intracellular S^0_{n-1} + a perthiol	N/A	spontaneous
		a perthiol + a [DsrE protein]-L-cysteine →a [DsrE protein] persulfide + a thiol (overall reaction for sulfur transfers)	N/A	*rhd*, *tusA*, *dsrEFH*,
		a [DsrE protein] persulfide + a [DsrC protein]-dithiol + an oxidized electron carrier → a [DsrE protein]-L-cysteine + a [DsrC]-trisulfide + a reduced two electron	N/A	*dsrC*, *dsrMKJOP*

		carrier (overall reaction for [DsrC]-trisulfide recycling)		
		[DsrC]-trisulfide + NAD^+ + 3 H_2O ↔ sulfite + [DsrC protein]-dithiol + NADH + 3 H^+	N/A	*dsrABL*
Reaction 7	*Thiosulfate oxidation ($S_2O_3^{2-}$ → SO_4^{2-})*	2 oxidized *c*-type cytochrome$_{[periplasm]}$ + 2 thiosulfate$_{[periplasm]}$ ↔ tetrathionate$_{[periplasm]}$ + 2 reduced c-type cytochrome$_{[periplasm]}$ + 2 $H^+_{[periplasm]}$	-21.08	*tsdA*
Reaction 8	*Thiosulfate oxidation ($S_2O_3^{2-}$ → SO_4^{2-})*	2 thiosulfate + a quinone → tetrathionate + quinol	1.93	*doxAD*
Reaction 9	*Thiosulfate oxidation ($S_2O_3^{2-}$ → SO_4^{2-})*	thiosulfate + 5 H_2O → 2 sulfate + 8e^- + 11 H^+ (overall reaction)	N/A	*soxXABYZ(CD)$_2$*
Reaction 10	*Sulfite oxidation (SO_3^{2-} → SO_4^{2-})*	sulfite + 2 oxidized *c*-type cytochrome + H_2O → sulfate + 2 reduced *c*-type cytochrome + 2 H^+	-16.72	*sorAB*
Reaction 11	*Sulfite oxidation (SO_3^{2-} → SO_4^{2-})*	sulfite + AMP + oxidized AprM electron-transfer protein + 2 H^+ ↔ adenosine 5'-phosphosulfate + reduced AprM electron-transfer protein	N/A	*aprAB*
		adenosine 5'-phosphosulfate + phosphate → sulfate + ADP + H^+	-8.71	*apt, sat*

Reaction 12	*Sulfite oxidation* *($SO_3^{2-} \rightarrow SO_4^{2-}$)*	sulfite + menaquinone + H_2O → sulfate + menaquinol	-21.51	*soeABC*
Reaction 13	*Dissimilatory sulfate reduction* *($SO_4^{2-} \rightarrow S^{2-}$)*	sulfate + ATP + H^+ ↔ adenosine 5'-phosphosulfate + diphosphate	-20.45	*sat*
		adenosine 5'-phosphosulfate + reduced QmoABC electron-transfer protein complex ↔ sulfite + AMP + oxidized QmoABC electron-transfer protein complex + 2 H^+	N/A	*aprAB*
		sulfite + [DsrC protein]-dithiol + NADH + 3 H^+ ↔ [DsrC]-trisulfide + NAD^+ + 3 H_2O	N/A	*dsrAB (reductive)*
		[DsrC]-trisulfide + 2 electron-transfer quinol → hydrogen sulfide + [DsrC protein]-dithiol + 2 electron-transfer quinone	-9.54	*dsrK*
Reaction 14	*Sulfur reduction* *($S^0 \rightarrow S^{2-}$)*	sulfur(0) + H_2 → hydrogen sulfide	-19.64	*sreABC*
Reaction 15	*Sulfur reduction* *($S^0 \rightarrow S^{2-}$)*	n sulfur(0) + hydrosulfide → $(sulfide)_{n+1}$ + H^+	N/A	spontaneous
		$(sulfide)_n$ + H_2 → hydrogen sulfide + $(sulfide)_{n-1}$	N/A	*hydBC, psrABC, or shyCB*
Reaction 16	*Sulfur reduction* *($S^0 \rightarrow S^{2-}$)*	sulfur(0) + NAD(P)H + H^+ → hydrogen sulfide + $NAD(P)^+$	-1.78	NAD(P)H sulfur oxidoreductase (NSR)

Reaction 17	*Disproportionation of sulfur into sulfite and hydrogen sulfide (anaerobic and aerobic)*	3 sulfur(0) + 3 H_2O → sulfite + 2 hydrogen sulfide + 2 H^+	-16.65	not characterized
Reaction 18	*Disproportionation of sulfur into sulfite and hydrogen sulfide (anaerobic and aerobic)*	4 sulfur(0) + 4 H_2O + dioxygen → 2 hydrogen sulfide + 2 sulfite + 4 H^+	-119.74	*sor (sulfur oxygenase reductase)*
Reaction 19	*Disproportionation of thiosulfate into sulfate and elemental sulfur (Sox pathway)*	thiosulfate + intracellular $S^0_{(n)}$ + H_2O → sulfate + intracellular $S^0_{(n+1)}$ + 2e^- + 3 H^+ (overall reaction)	N/A	*soxXABYL*
Reaction 20	*Disproportionation of thiosulfate into sulfite and hydrogen sulfide (thiol)*	thiosulfate + 2 thiol → sulfite + hydrogen sulfide + disulfide + H^+	6.82	*RDL2*
Reaction 21	*Disproportionation of thiosulfate into sulfite and hydrogen sulfide (cytochrome)*	thiosulfate + 2 reduced type I cytochrome c_3 + H^+ → sulfite + hydrogen sulfide + 2 oxidized type I cytochrome c_3	-0.67	thiosulfate reductase (cytochrome)
Reaction 22	*Disproportionation of thiosulfate into sulfite and hydrogen sulfide (quinone)*	thiosulfate + menaquinol → sulfite + hydrogen sulfide + menaquinone + H^+	-5.45	*phsABC*
Reaction 23	*Disproportionation of thiosulfate into sulfite and thiocyanate (rhodanese)*	thiosulfate + hydrogen cyanide → sulfite + thiocyanate + 2 H^+	-30.83	*pspE or rhdA*

Oxidation reactions

Sulfide oxidation (S^{2-}→ S^0)

Sulfide oxidation to elemental sulfur can be conducted using two different metabolic pathways[10]. First, sulfide:quinone oxidoreductase (Sqr) transforms hydrogen sulfide to $(sulfide)_n$ using electron-transfer quinones, with the $(sulfide)_n$ then being spontaneously converted to sulfur, and stored in sulfur globules intracellularly in some bacteria, such as the model purple sulfur bacterium *Allochromatium vinosum*. Second, flavocytochrome *c* sulfide dehydrogenase (Fcc) can oxidize hydrogen sulfide using *c*-type cytochromes. Microbial sulfide oxidation is common in both oxic and anoxic environments and can be coupled to diverse electron acceptors including oxygen, nitrate, iron, and manganese. For example, sulfide oxidation is an important reaction driving anoxygenic photosynthesis by organisms from diverse lineages of bacteria including Chlorobiota, Firmicutes, Acidobacteriota, Chloroflexota, Proteobacteria, and Gemmatimonadota.

Elemental sulfur oxidation ($S^0 \rightarrow SO_3^{2-}$)

Elemental sulfur is typically oxidized different by autotrophic and heterotrophic organisms. In heterotrophs, the initial step of aerobic sulfur oxidation involves the spontaneous combination of glutathione with sulfur to form S-sulfanylglutathione. Subsequently, S-sulfanylglutathione dioxygenase (Sdo) oxidizes S-sulfanylglutathione to S-sulfinatoglutathione, and then transformed into sulfite[29]. Alternatively, sulfur oxidizing autotrophs, such as organisms from Gammaproteobacteria (like *Thioglobus*/SUP05) use the reverse-type dissimilatory sulfite reductase (rDsr, oxidative type) to oxidize sulfur to sulfite[30]. In this pathway, elemental sulfur stored within sulfur globules can undergo spontaneous reduction to persulfide before entering the cytoplasm. It is then transferred to DsrC by DsrEFH complex. DsrMKJOP complex facilitates the oxidation of the combined molecule, producing a [DsrC]-trisulfide. Finally, DsrABL catalyzes its oxidation

to sulfite[31]. This reaction direction is the opposite of canonical dissimilatory sulfite reduction.

Thiosulfate oxidation ($S_2O_3^{2-} \rightarrow SO_4^{2-}$)

Thiosulfate is an inorganic sulfur anion ($S_2O_3^{2-}$) that can be both oxidized and reduced, and tends to form bonds with metals. Three types of thiosulfate oxidation reactions exist, namely, type I, II, and III. Type I and Type II generate tetrathionate, whereas Type III uses the *sox* pathway and generate sulfate. The typical Sox multi-enzyme complex used in Type III thiosulfate oxidation comprises subunits of SoxXABYZ$(CD)_2$[32]. SoxXA is heterodimeric *c*-type cytochrome, SoxYZ is a sulfur carrier, Sox$(CD)_2$ is sulfur dehydrogenase, and SoxB acts as the sulfate thiol esterase, releasing sulfate from the cysteine S-thiosulfonate[32]. The *sox* pathway is encoded by organisms in various bacteria genera, such as *Aquifex*, *Paracoccus*, and *Rhodovulum* within the bacterial phyla Aquificota and Proteobacteria.

Sulfite oxidation ($SO_3^{2-} \rightarrow SO_4^{2-}$)

Sulfite precedes the final oxidized form of sulfur species, sulfate. For prokaryotes, there are three major pathways of oxidizing sulfite and generating sulfate[33,34]. First, Sor, the sulfite dehydrogenase (cytochrome), can directly oxidize sulfite, and the electrons generated during catalysis can be passed to the oxidized form of cytochrome *c* for making protons. Second, adenosine 5'-phosphosulfate (APS) is the intermediate within the oxidation pathway; it can either be oxidized by adenylylsulfate:phosphate adenylyltransferase (Apr) or sulfate adenylyltransferase (Sat) to sulfate. Third, Soe, the sulfite dehydrogenase (quinone), can directly oxidize sulfite with menaquinone to generate sulfate and menaquinol. The microbial taxa ranges for these three major pathways are listed in Table S1.

Reduction reactions

Assimilatory sulfate reduction

Sulfur reduction processes involve the reduction of oxidized forms of sulfur. Important reactions such as sulfate reduction, can be assimilatory or dissimilatory. Assimilatory sulfate reduction involves the reduction of sulfate to primarily produce organosulfur compounds for biosynthesis such as in amino acids, proteins, and other cellular components. The four assimilatory sulfate reduction pathways and corresponding microbial taxa ranges are summarized in Table S1. In contrast, dissimilatory sulfate reduction involves the reduction of sulfate (as an electron acceptor) to either generate energy and/or link to other metabolic processes (such as carbon fixation or heterotrophy). Overall, sulfate reduction to either sulfite or hydrogen sulfide, is a multi-step process.

Dissimilatory sulfate and sulfite reduction (SO_4^{2-} or $SO_3^{2-} \rightarrow S^{2-}$)

Microorganisms that have the capacity for complete reduction of sulfate to sulfide, are referred to as SRM (sulfate-reducing microorganisms)[9]. In this process, sulfate is first reduced to APS by the activity of Sat, which uses ATP and releases diphosphate[9]. Then, APS reductase (Apr) converts APS to sulfite[9]. Finally, sulfite is converted to H_2S by dissimilatory sulfite reductases (Dsr)[9]. During dissimilatory sulfate reduction, sulfate is used as the terminal electron acceptor (instead of what would normally be oxygen), and organisms must expel the sulfide from the cell because H_2S can be toxic. A small proportion of microorganisms lack the ability to reduce sulfate to sulfite, but can reduce sulfite to sulfide and are termed sulfite-reducing microorganisms. Diverse microorganisms in thermophilic, mesophilic, and psychrophilic environments encode these metabolic pathways including archaea from four phyla including Halobacterota,

Thermoproteota, Thermoplasmatota, and Hydrothermarchaeota, and bacteria from twenty seven distinct phyla (Table S1).

Sulfur reduction ($S \rightarrow S^{2-}$)

Although many sulfate- or sulfite-reducing microorganisms can carry out sulfur reduction (sulfur reduction is a "step" in both oxidative and reductive pathways), few organisms are true sulfur reducers (those that can not use sulfate or sulfite but require elemental sulfur as a starting point). Elemental sulfur can be directly reduced to hydrogen sulfide by a sulfur reductase (Sre)[35] or NAD(P)H sulfur oxidoreductase (Nsr)[36], or elemental sulfur can be first transformed to polysulfide spontaneously and then reduced to hydrogen sulfide[37]. Hydrogen serves as the terminal electron donor for both pathways. The Sre enzyme is widely distributed across numerous bacterial and archaeal species (Table S1). Conversely, Nrs is only distributed in a few archaeal thermophiles, exemplified by *Pyrococcus*. Similarly, several microbial species employ specific enzymes encoding for the polysulfide pathway. Examples include sulfur reductase (HydBC) in *Pyrococcus furiosus*, polysulfide reductase (PsrABC) in *Wolinella succinogenes*, and sulfide dehydrogenase (ShyCB) in *Pyrococcus furiosus*.

Disproportionation

Disproportionation of sulfur into sulfite and hydrogen sulfide (anaerobic and aerobic)

In this process, sulfur serves as both the electron donor and acceptor, with oxygen acting as an additional electron acceptor. No additional electron donors, acceptors, or cofactors are involved in the reaction. The reaction yields sulfite and hydrogen sulfide as products, while thiosulfate can also be formed through a non-enzymatic reaction between sulfite and sulfur. The enzyme responsible for this reaction is the soluble cytoplasmic sulfur

oxygenase reductase (Sor)[38], frequently found in thermophilic organisms from the archaeal phyla Thermoplasmatota and Thermoproteota, and the bacterial phylum Aquificota.

Disproportionation of thiosulfate into sulfate and elemental sulfur (Sox pathway)
The Sox pathway facilitates the oxidation of thiosulfate to sulfate when all the components SoxXABYZ$(CD)_2$ are present. However, in the absence of SoxCD, thiosulfate undergoes disproportionation into sulfate and elemental sulfur. When SoxCD is present, both the sulfone-sulfur (sulfur atom attached to oxygen) and sulfane-sulfur (sulfur atom without oxygen attached) within a thiosulfate molecule are oxidized to form sulfate; conversely, in the absence of SoxCD, the sulfane moiety is transferred to generate sulfur globules, which are then either deposited in the periplasmic space (e.g., in Beggiatoa, Chromatiaceae) or extracellularly (e.g., in Chlorobiaceae, Ectothiorhodospiraceae) [39,40].

Disproportionation of thiosulfate into sulfite and hydrogen sulfide (thiol)
Within this pathway, a thiol-dependent thiosulfate-thiol sulfurtransferase (RDL2) facilitates the transfer of sulfane sulfur to a thiol compound, resulting in the formation of sulfite and a persulfide[41]. When an excess of thiol is present, the persulfide can spontaneously release free hydrogen sulfide as the final product[41]. This pathway is frequently found in several bacterial genera including acidophilic organisms such as *Acidithiobacillus* and *Thiobacillus*, non-acidophilic organisms such as *Acinetobacter*, and fungi such as *Saccharomyces*.

Disproportionation of thiosulfate into sulfite and hydrogen sulfide (cytochrome)
In this pathway, the conversion of thiosulfate to sulfite and sulfide is facilitated by thiosulfate reductase and hydrogenase, consuming hydrogen in the process. An essential component in this process is the presence of cytochrome c_3, which serves as an electron

shuttle between hydrogenase and thiosulfate reductase[42]. This pathway is frequently found in several bacterial genera from the phylum Proteobacteria including *Desulfocapsa*, *Desulfotomaculum*, and *Desulfovibrio*.

Disproportionation of thiosulfate into sulfite and hydrogen sulfide (quinone)

This pathway is facilitated by the membrane-bound quinone-dependent thiosulfate reductase (Phs), which functions as a molybdopterin-containing oxidoreductase[43]. It is an energy-dependent process that can be powered by NADH and/or other fermentation products[44]. This pathway is frequently found in several Enterobacteriaceae genera that are common commensal and pathogenic members of the human and animal gut microbiomes including *Salmonella*, *Proteus*, *Citrobacter*, *Budvicia*, and *Edwardsiella*.

Disproportionation of thiosulfate into sulfite and thiocyanate (rhodanese/thiosulfate sulfurtransferase)

Thiosulfate and cyanide can be disproportionated by rhodanese (thiosulfate sulfurtransferase) into sulfite and thiocyanate[45]. The enzyme rhodanese has been discovered in both prokaryotes and eukaryotes, and is widely distributed in biology ranging from heterotrophic bacteria such *E. Coli*, autotrophic sulfur oxidizing Proteobacteria, to Mammalia. It has been suggested that rhodanese is involved in the detoxification of cyanide in both mammals[46] and bacteria[47].

Organic Sulfur transformations

A comprehensive understanding of organic sulfur transformation has been discussed in previous works[48], while for the purposes of this review, we emphasize specific reactions closely related to inorganic forms[14]. Notably, assimilatory sulfate reduction plays a crucial role in incorporating sulfide into cysteine or homocysteine for biosynthesis, contributing

to the formation of essential cellular components. Subsequently, homocysteine/cysteine can be converted into methionine, serving as the primary source of sulfur-containing proteinogenic amino acids in cellular organisms. Organic sulfur compounds such as taurine, alkylsulfonate/isethionate, and methanesulfonate can undergo catalysis, leading to the production of sulfite. Likewise, other organic sulfur compounds, including 3-mercaptopyruvate and cystathionine, can be catalyzed to generate sulfide through homocysteine/cysteine. The inorganic sulfur compounds produced from these processes, such as sulfide can enter both assimilatory and dissimilatory sulfur metabolism, completing the sulfur transformation loop. Sulfide, as an important product of organic sulfur transformations, plays significant roles in both human and environmental systems. For instance, sulfide is genotoxic and has proinflammatory effects on the gut epithelium[49], and contributes to "cryptic sulfur cycling"[50] through organosulfur degradation in oxic conditions[51]. Organic sulfur transformation can interact with inorganic sulfur metabolisms and we describe instances of these in the following section on illustrating sulfur cycling across different ecosystems.

Microbial diversity of sulfur-cycling organisms

Dissimilatory sulfur-metabolizing lineages are widely distributed across the tree of life

Microbes across the entire spectrum of life are involved in sulfur cycling: archaea, bacteria, eukaryotes, and viruses. Here, we identify the most abundant and important microorganisms and metabolisms in sulfur-cycling microbiomes. We focus on six key functions associated with oxidation, reduction, and disproportionation metabolisms: sulfide oxidation, elemental sulfur oxidation, thiosulfate oxidation, sulfur reduction, sulfite reduction, and thiosulfate disproportionation (Fig. 3). We searched GTDB species representatives (release 202) to specifically summarize the distribution of these functions

across all prokaryotic phyla/classes. The distribution of functions was guided by the following rules for presence of proteins in microbial genomes, *fccB* and *sqr* for sulfide oxidation, *dsr* (oxidative), *sdo*, and *sor* for sulfur oxidation, *SoxXYZA* with *soxCD* for thiosulfate oxidation, *sreABC* and *sor* for sulfur reduction, *dsr* (reductive) and *asrABC* for sulfite reduction, and *phsA* and *SoxXYZA* without *soxCD* for thiosulfate disproportionation. Across 127 bacterial phyla and 47 archaeal classes, 102 and 26 were identified to contain at least one of the six key functions, respectively, suggesting that sulfur-metabolizing lineages are common (Fig. 3). It should be noted that these two trees only show a subset of dissimilatory sulfur metabolizing functions that are common and widely distributed (Table S1). Sulfate reduction (to sulfite) and sulfite oxidation reactions are not shown as they are not exclusive to dissimilatory metabolism and are widely distributed across all domains of life.

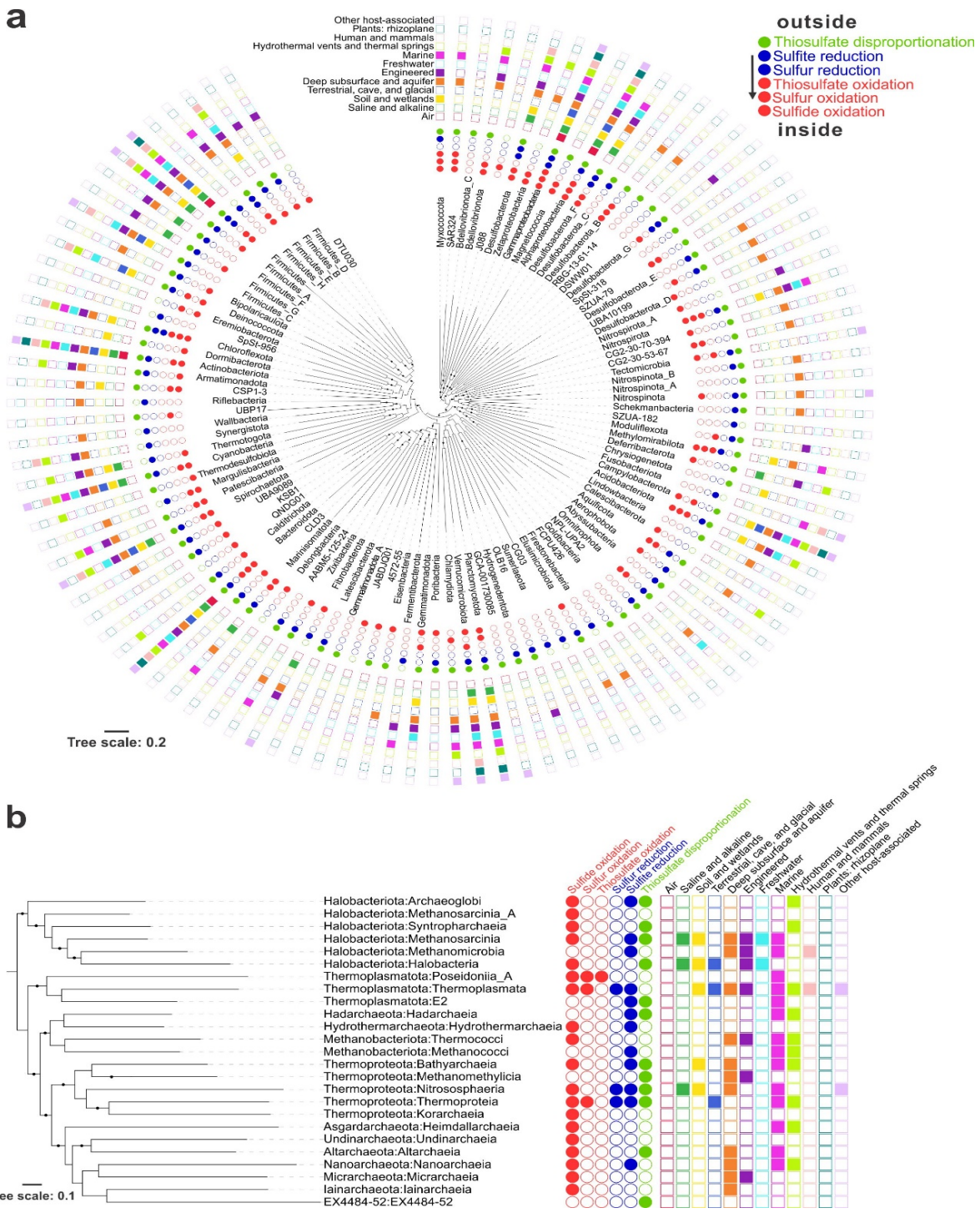


Fig. 3 | Sulfur metabolism across the tree of life. Bacterial (**a**) and archaeal (**b**) tree of life based on GTDB species representatives with annotations for key functions associated with dissimilatory sulfur metabolism. The bacterial tree was classified to the phylum level, except for the Proteobacteria which were classified to the class level. The archaeal tree was classified to the class level. Filled circles indicate the presence of a specific function (i.e., if any species representative within the class/phylum possessed the function). Similarly, a filled square represents identification of a class/phylum in the environmental category (i.e., if any species representative from that class/phylum exists within that category). Bootstrap (UFBoot) support values (≥80%) are labeled as black dots. One representative genome was randomly picked from each class/phylum to build the phylogenetic tree.

Viruses cannot conduct metabolic reactions independently but are obligate intracellular parasites relying on hosts for energy and metabolic precursors to propagate. Viral-encoded auxiliary metabolic genes (AMGs) can provide viruses with fitness advantages for propagation by maintaining, driving, or short-circuiting important step(s) of the metabolic pathway(s) of host cells[52,53]. AMGs have been found to significantly contribute to the biogeochemical cycling of essential elements such as carbon[54-58], nitrogen[59,60], and sulfur[61,62] in diverse environments, including oceans, freshwater lakes, and terrestrial soils, ultimately shaping the biogeochemical landscape of their ecosystems. Viruses can affect sulfur metabolism in their hosts by encoding genes involved in both dissimilatory and assimilatory sulfur metabolism[13,14] (Box 1).

Environmental distribution of sulfur-metabolizing lineages

To study the distribution of sulfur-metabolizing prokaryotes, we classified them into 12 environmental categories encompassing natural environments, engineered environments, and human systems. Various bacterial lineages, including Actinobacteriota, Bacteroidota, Firmicutes, Alphaproteobacteria, and Gammaproteobacteria, exhibited a widespread distribution across all environmental categories (Fig. 3a). In general, archaea were not as widely distributed as bacteria. Among archaeal classes, Thermoplasmata demonstrated the broadest distribution, extending to 8 out of the 12 environmental categories (Fig. 3b). With rapidly increasing numbers of genomes in public databases from diverse environments, future studies will enable an increasingly representative portrait of the taxonomic diversity of sulfur metabolism, and their distribution in nature.

Sulfur cycling across ecosystems

Abiotic and biotic reactions associated with the sulfur cycle take place in diverse environments encompassing the atmosphere, soils, aquatic systems, and the human body. Importantly, these environments serve as habitats for various microorganisms involved in sulfur cycling. Specifically, six representative ecosystems were chosen based on the significance or considerable global impact of microbial sulfur transformations occurring within them. Marine oxygen minimum zones (OMZs) are crucial for the rapid cycling of sulfur compounds and the maintenance of low sulfide concentrations, resulting in “cryptic sulfur cycling”. Marine hydrothermal vents are vital locations for sulfur-metabolizing chemosynthetic communities that dominate primary production through H_2S and S^0 oxidation. Freshwater lakes exhibit diverse sulfur metabolisms based on stratification and oxygen levels. The atmosphere plays a crucial role in the global sulfur cycle, with marine bacteria contributing approximately 10% to the annual global DMS production. Soils and the human gut are key ecosystems that have a significant influence on the balance of inorganic and organic sulfur budgets, thereby impacting sustainable soil ecosystem and plant growth as well as human health, respectively. The subsequent sections elaborate on the influence of environmental factors and chemicals on the composition of microbial communities and their interactions, ultimately shaping the overall sulfur cycling process.

Marine oxygen minimum zones

Marine oxygen minimum zones (OMZs) are geochemical regimes in the world’s oceans located at intermediate depths (~100-1000 m) that are characterized by depleted oxygen levels (below 20 μmol/kg)[63]. Within OMZs, microorganisms rapidly cycle sulfur compounds by reducing sulfate to sulfide and oxidizing sulfide to sulfate, but sulfide concentrations remain at sub-micromolar levels. Consequently, it is difficult to observe

the sulfur transformations using traditional biogeochemical measurements[64]. The close proximity of sulfate reduction and sulfide oxidation, for example, in marine particle aggregates, facilitates a simultaneous coupled activity that maintains low sulfide concentrations, leading to what is known as “cryptic sulfur cycling”[50]. In some cases, anoxic waters can accumulate sulfide, which are then termed euxinic.

Sulfur metabolism in OMZs is closely linked to carbon and nitrogen cycling by autotrophic and heterotrophic microorganisms. The remineralization of organic matter by heterotrophs is mediated by electron acceptors, starting with oxygen in the upper OMZ, transitioning to nitrate/nitrite, followed by sulfate/sulfite in the core OMZ[50,65] (Fig. 4). Meanwhile, groups of chemosynthetic microorganisms dominated by the gammaproteobacterial “SUP05” clade play a crucial role in the rapid removal of hydrogen sulfide by coupling sulfide oxidation to dissimilatory reduction of nitrate/nitrite[50,64] (Fig. 4).

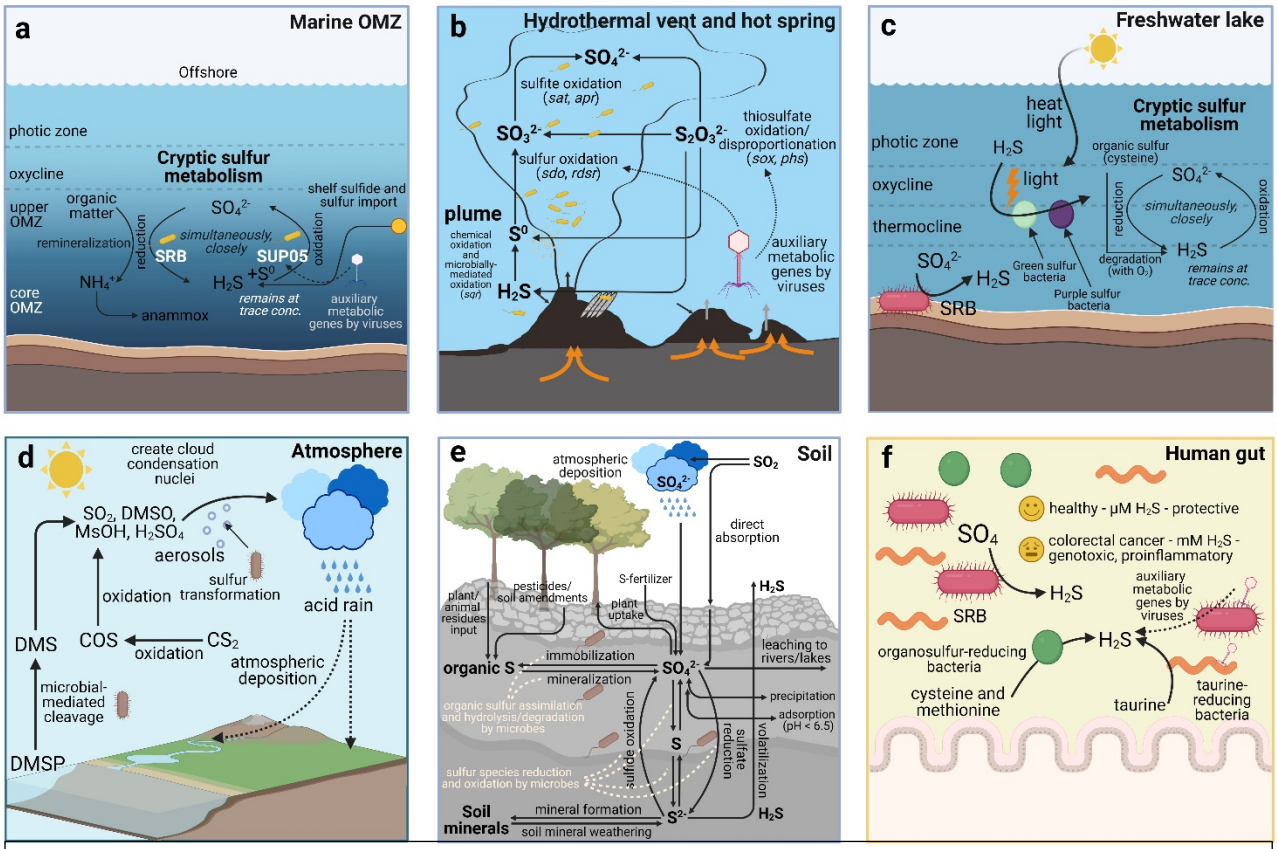


Fig. 4 | Microbial interactions of the sulfur cycle in marine, freshwater, atmosphere, soil, and human gut systems. a, The close interplay between sulfate reduction and sulfide oxidation processes in marine OMZs leads to the rapid turnover of sulfide, which maintains it at low concentrations, leading to "cryptic sulfur cycling". **b,** Sulfur oxidation-dominated energy landscape determines the structure and function of microbiomes in hydrothermal vent and related hot spring ecosystems. **c**, In freshwater lakes, light, temperature, and oxygen availability influence the distribution and biogeochemical processes of microorganisms. A "cryptic sulfur cycle" is driven by the reduction of sulfate or organosulfur compounds to form hydrogen sulfide that rapidly recycled. **d**, In the atmosphere, microbially-driven DMS promotes aerosol formation that creates clouds for microbial activities and transports sulfur compounds across the globe. **e**, In soils, microbially-driven reactions contribute to the oxidation/reduction and immobilization/mineralization processes of inorganic and organic sulfur compounds. **f**, In human gut, both microbially-mediated inorganic and organic sulfur metabolisms, and auxiliary metabolism mediated by viruses contribute to the production of H_2S, which conveys either protective or pro-inflammatory effects at different concentrations.

Hydrothermal vents and hot springs

Hydrothermal vents are globally distributed along mid-ocean ridges on the ocean floor[66]. Sulfur metabolizing chemosynthetic communities are common in three distinct vent habitats: hydrothermal vent chimneys dominated by thermophilic and mesophilic communities, hydrothermal vent plumes dominated by psychrophilic communities, and

symbionts associated with vent-associated animals (such as tubeworms, clams, mussels, and snails)[66]. Hydrothermal vent chimneys are characterized by steep geochemical gradients created by the mixing of hydrothermally-derived fluids rich in reduced substrates and cold oxidized seawater[66]. This mixing of fluids forms plumes that rise hundreds of meters above the seafloor and spread over thousands of kilometers[67]. H_2S and S^0 are the dominant energy sources for primary production in most hydrothermal microbiomes across the globe[68,69]. Sulfur metabolism in hydrothermal vent chimneys is dominated by members of Gammaproteobacteria and Campylobacteria[69], but the distribution of these two bacterial lineages is governed by sulfide/sulfur concentrations with Campylobacteria favored at higher and Gammaproteobacteria at lower concentrations[70]. Similarly, a core hydrothermal vent plume microbiome was recently identified identified to contain at least 14 microbial genera dominated by sulfur-metabolizing microorganisms. These include microorganisms from diverse lineages including Actinobacteriota, Campilobacterota, Chloroflexi, Planctomycetota, Alphaproteobacteria, Gammaproteobacteria, and SAR324[68]. Similarily, terrestrial hydrothermal ecosystems, commonly referred to as hot springs, encompass environments rich in sulfate, sulfide, and methane, with distinct micro-niches characterized by varying oxygen levels[71]. Organisms in hot springs can conduct a variety of sulfur metabolisms, including aerobic sulfide oxidation[72], anaerobic sulfate/sulfite reduction[73], anoxygenic photosynthetic-dependent sulfur oxidation[73], and coupled anaerobic methane and dissimilatory sulfur metabolisms[71]. Microbial sulfur metabolism in terrestrial hot springs plays a pivotal role in global sulfur cycling and holds significant implications for understanding the evolution of ancient sulfur and carbon metabolisms[71,74].

Freshwater lakes

Freshwater lakes are primarily characterized by differences in location (temperate vs. tropical) and stratification (holomictic vs. meromictic), which are then associated with differences in sulfur metabolism and biogeochemistry. In holomictic lakes, such as Lake Mendota (located in Wisconsin, USA)[75], differences in surface and bottom temperatures promote stratification in certain seasons; when seasons change, turnover and mixing of lake water occurs. Contrastingly, in meromictic lakes, such as Lake Tanganyika[76], layers do not intermix, the bottom layer does not turn over for years, decades, or even hundreds of years, and is generally hypoxic. The gradients of oxygen, temperature, and light availability generally control the distribution and biogeochemical processes of microorganisms[77]. Green and purple sulfur bacteria which can survive in well-lit but oxygen-deficient areas, oxidize sulfide or elemental sulfur, use light energy for autotrophic metabolisms, and tend to congregate in the oxycline and photic zone[78]. In the lower layers and sediments, where oxygen is limited, sulfate reduction is abundant and drives the transformations of organic matter[76,79].

Based on data collected from 308 lakes spanning diverse landscapes and trophic states, including urbanized, agricultural, and intermediate conditions, microbially-mediated sulfur metabolisms exhibit certain commonalities[80]. In the majority of surface waters, Gammaproteobacteria were a prominent group and associated with the oxidation and disproportionation of thiosulfate. Dissimilatory sulfate/sulfite reduction, reliant on lower oxygen levels, was less prevalent in surface waters, which aligned with their higher oxygenation levels[80]. All lakes generally contained lower sulfate levels than oceans (1-50 mg/L vs. ~2,700 mg/L)[81,82]. Similar to OMZs, a “cryptic sulfur cycle” also occurs in freshwater, but differs in the source of sulfide which can be from sulfate reduction or from

organosulfur degradation in oxic conditions[51]. Commonly, sulfide oxidation in lakes may be driven by the reduction of Fe(III)[83], phototrophy[84], or denitrification[85] (Box 2).

Atmosphere

Microbial cell densities in the atmosphere are significantly lower compared to Earth's surface biomes, typically measured in cells per cubic meter instead of grams or milliliters. Bacterial densities in near-surface air are generally estimated to be between 10^4 and 10^5 cells per cubic meter[86], with the highest concentration in clouds ranging from 10^3 to 10^6 cells per milliliter[87,88]. The major source of sulfur transferred from ocean to atmosphere is through DMS. Marine phytoplankton and algae produce DMSP, which is enzymatically broken down by bacteria into DMS[89]. Subsequently, DMS is oxidized in the atmosphere, leading to the formation of various sulfur-containing compounds[90]. These compounds serve as aerosols, acting as cloud condensation nuclei and influencing global climate by promoting cloud formation[90]. Annually, approximately 10% of global DMS production (around 300 Tg/year) is contributed by bacteria entering the atmosphere[91]. Microbially-driven sulfur transformations can happen in the atmosphere because microbes reside on many primary aerosols, such as dust[92], DMS[90], and sea spray[93]. These bioaerosols contribute to the formation of cloud droplets, which contain water and provide shade against direct UV radiation, serving as "hotspots" for microbial activity[92]. In fact, recent metagenomic studies have revealed that bioaerosol microbes mediate sulfur transformations including the formation of sulfate[94]. Additionally, carbonyl sulfide (COS)[95] and carbon disulfide (CS_2)[96] are important sulfur gases that contribute to the sulfate aerosol layer in the atmosphere, with both natural and industrial origins[3]. Furthermore, microbes can consume COS as a trace gas in the atmosphere at ambient

concentrations[97]. Through atmospheric deposition, aerosols containing microbially-driven sulfur compounds also play a significant role in the global transportation of sulfur.

Soil

Terrestrial soils are characterized by both organic and inorganic forms of sulfur. Inorganic sulfur is subjected to the influence of precipitation, adsorption, oxidation, and reduction[24]; the latter two are mainly mediated by microorganisms. In contrast, organic sulfur is subjected to immobilization and mineralization[24]. About 5% of sulfur in soils is in the form of inorganic sulfur[98]. Microbial oxidation of reduced sulfur species, i.e., S^{2-}, S^0, and $S_2O_3^{2-}$, can be performed by both autotrophic and heterotrophic microorganisms. For example, *Thiobacillus* sp. uses elementary sulfur, thiosulfate, or polythionates as energy sources and fixes CO_2[99]. *Pseudomonas aeruginosa* PAO1 can consume various sulfur species (H_2S, S^{2-}, and HS^-) when growing on organic carbon under aerobic conditions[100].

Organic sulfur represents 95% of soil sulfur and comprises two primary forms: ester sulfates (C–O–S) and carbon-bonded S (C–S)[101]. Particularly, ester sulfates comprise 30-70% of the organic sulfur pool[102]. The biochemical mineralization of ester sulfates involves microbially-mediated enzymatic hydrolysis by sulfatases[24]. The release of SO_4^{2-} by sulfatase is controlled by the sulfur supply to soil microorganisms in the environment. Meanwhile, the biological mineralization of carbon-bonded S is mediated by microorganisms for use as a carbon source and for energy demands, with SO_4^{2-} being released as a by-product[24]. In soil, the sulfatase-dependent mineralization process is much faster and thus more important for short-term sulfur cycling[103]. While microorganisms control the metabolism and composition of soil organic sulfur compounds, these processes control sulfur supply for plant growth since plants depend

on sulfate released by organic sulfur mineralization[103]. Overall, soil microbiomes contribute significantly to the balance of inorganic/organic sulfur budgets that are critical for healthy and sustainable soil ecosystems.

Human gut

The focus of sulfur compounds in human health has largely been on H_2S. In the human body, H_2S can have both protective and anti-inflammatory effects, as well as pro-inflammatory effects, mediating various disease processes[104,105]. Studies on sulfur microbiomes in the human gut have explored dissimilatory sulfate reduction[106], organosulfur metabolism[107], and assimilatory sulfur metabolism[104], all of which are primarily mediated by bacteria. However, recent evidence suggests that viruses can also influence organic sulfur metabolism in the human gut[14].

Understanding the role of H_2S in the human gut is crucial for comprehending the potential impact of bacteria-produced H_2S on disease development and considering strategies to mitigate its effects (Fig. 4). At low concentrations (μM level), H_2S can promote vasorelaxation and prevent cell death[108]. However, in the gut, it is commonly found at higher concentrations (mM level), leading to cellular stress and genotoxicity[109,110]. Bacteria that produce H_2S have been associated with increased risks of colorectal cancer (CRC) due to the inflammation they cause in the gut epithelium[49]. Furthermore, bacterial-produced H_2S has been implicated in inflammatory bowel disease (IBD)[49,111]. Research on bacterial-produced H_2S and its correlation with human gut pathogenesis has predominantly focused on sulfate reduction mediated by dissimilatory sulfite reductases; nevertheless, recent studies have shed light on organic sulfur metabolism, indicating its potential connection to CRC[112]. The diversity of bacterial sulfur cycling in the human gut

is more taxonomically diverse and prevalent than previously recognized, with organic sulfur metabolisms, such as cysteine/methionine reduction and taurine reduction, playing a more significant role and contributing to a greater extent to H_2S production[107].

The microbial sulfur cycle and climate change

The sulfur cycle on Earth plays key roles in both impacting, and being impacted by climate change on Earth. For example, the microbal sulfur cycle plays key roles in the loss of nitrogen from ecosystems, the anaerobic oxidation of carbon globally (by SRMs), influences climate patterns thorugh production of climate active compounds such as dimethyl sulfide, and drives the expansion of globally distributed OMZs.

The Earth's climate system is highly responsive to the release of dimethyl sulfide (DMS) from the oceans[113]. The influx of microbially produced DMS into the atmosphere leads to alterations in aerosol formation and cloud nucleation, subsequently affecting solar radiation reaching the ocean surface. The cooling effect resulting from DMS-induced reduction in radiation is of a comparable magnitude to the warming effect attributed to anthropogenic CO_2 emissions[114]. In the future, climate variations such as increased global mean temperatures and elevated CO_2 will exert reciprocal feedback on oceanic DMS production. Currently, only a minor portion (~10%) of phytoplanktonic DMSP undergoes conversion via microbially-mediated processes, resulting in the release of DMS into the atmosphere[115]. Various other processes, including microbial assimilation (e.g., demethylation pathways) and photolysis, contribute to the overall DMSP sink budget[115]. It is posited that even a slight alteration in the microbially-mediated DMSP cycling budget can lead to substantial changes in DMS production and, subsequently, its emission into the atmosphere. Ocean warming influences growth rates and modifies the

functional trait composition of marine phytoplankton. Simultaneously, temperature regulates the metabolic rates of marine heterotrophs. Consequently, the delicate balance between the autotrophic activity of sulfate assimilation and DMSP production, and the heterotrophic activity of DMS production profoundly influences the oceanic DMS budget and sulfur cycling interactions between the ocean and atmosphere[115].

As global temperatures increase, the microbial sulfur cycle is predicted to play important roles in aquatic systems worldwide. Higher temperatures will cause OMZs in the world's oceans to rapidly expand in volume, and freshwater lakes to become more stratified and anoxic. The sulfur cycle within marine OMZs and freshwater lakes has multifaceted effects on the global climate, influencing primary production, nitrogen loss, and greenhouse gas emissions[116]. Higher temperatures will lead to increased anoxia, thereby leading to enchanced benthic phosphorus release during sulfidic events attributed to the activities of sulfate-reducing and sulfide-oxidizing bacteria[116]. Subsequencly, phosphorus, as a limiting nutrient in aquatic systems, can be released to the surface layer and thereby stimulate primary production[116]. Increased anoxia will also lead to enhanced denitrification activity associated with sulfide/sulfur oxidation, increased production of the green house gas N_2O (from incomplete denitrification), and thereby contribute to substantially increased nitrogen loss in oceans and lakes[116].

Both natural and anthropogenic factors can influence the global sulfur cycle. Natural factors including ocean surface temperature, surface winds, currents, and mixing of the water column play a pivotal role in influencing DMS seawater concentrations and,

consequently, the sea-to-air sulfur flux[115]. In contrast, anthropogenic factors include the utilization of fossil fuels that significantly elevates atmospheric sulfur deposition and results in the generation of acid rain which poses substantial environmental and health hazards, and agricultural additions of sulfur-containing fertilizers and pesticides that have negative impacts on soil and water quality, and human health[4].

In North America, a shift in the negative impacts of the sulfur cycle has been documented since the 1970s[4] with a switch from atmospheric emissions to agricultural additions of sulfur. Given ongoing climate change, intensified crop production, and shifts of sulfur additions, it is imperative to reevaluate the environmental consequences of the microbial sulfur cycle.

Conclusions

Microbial sulfur metabolism plays a fundamental role in energy conversion and facilitates the dynamic movement of sulfur in various forms across diverse ecosystems and environments on Earth. Microbial sulfur metabolism has been a bedrock of energy metabolism on Earth since the evolution of life[117]. While we have come a long way into uncovering the role of microorganisms in the sulfur cycle, much remains to be learned about these processes. Sulfur transformations can be difficult to track because of the rapid turnover of interconnected redox reactions in nature, thus leading to cryptic sulfur cycling. Despite the challenges in studying this, diverse approaches such as omics techniques, stable isotope labeling, ecophysiology, modeling, and microscopy are being employed to overcome these difficulties. These approaches enable the investigation of sulfur flux, reaction rates, and microbial energetics *in situ* and in model systems.

Sulfur cycling occurs in varied environmental settings. Under different conditions, the reaction types and rates are significantly influenced by the distribution of available

electron donors and acceptors and environmental factors. Sulfur exists in multiple valence states in its inorganic forms, but many of these states, including polythionates ($S_2O_3^{2-}$ and $S_4O_6^{2-}$)[118,119], have seldom been studied due to technical difficulties in detecting them. Exploring new pathways mediating the transformations of inorganic sulfur compounds can provide new insights for global sulfur cycling. Additionally, further investigations on the coupling of sulfur transformations with other elemental cycles can identify novel biogeochemical pathways and illuminate their overlooked activities. Studying these patterns can help us understand, predict, and manipulate the direction and fate of sulfur reactions in different environmental contexts, including natural environments and man-made biological systems.

We are increasingly recognizing that microbial sulfur metabolism plays a critical role in natural and man-made environments, however, future research should examine its implications in biomedicine. Investigating the role of microbial sulfur metabolism in the gut microbiome dysbiosis and the progression and development of associated diseases, such as IBD, Crohn's disease, and CRC, can help elucidate the underlying mechanisms of pathogenesis and develop new approaches for treatment and prevention. To conclude, it is imperative that future research continues to explore the multifaceted roles of microbial sulfur metabolism across diverse scientific disciplines, ranging from biogeochemistry to medicine. These insights will have profound implications for environmental sustainability and biogeochemistry, improve our understanding of human health, and develop the necessary knowledge and tools to foster a healthier planet.

Box 1. Viruses and sulfur metabolism

Viruses encode a variety of sulfur AMGs facilitating the dissimilatory sulfur metabolism for enhancement of viral fitness. Currently, AMGs involving sulfur oxidation and thiosulfate oxidation/disproportionation have been discovered in numerous environments, ranging from oceans, freshwater, human host-associated, and engineered ecosystems[13] (Table 1). The process of virus-augmented elemental sulfur oxidation and thiosulfate oxidation generates additional electrons through exergonic reactions, which can be utilized by the electron transport chain for ATP production. This process provides fitness advantages to viruses by promoting virion propagation.

Viral AMGs also encode five categories of genes involving the transformation of organosulfur and assimilatory sulfur metabolism[14]. They augment the degradation of organosulfur compounds to sulfite or sulfide, manipulate different organosulfur compound forms, and accelerate sulfur fixation. Organosulfur AMGs viruses infect all three domains of life and widely influence the environmental sulfur cycle and human health by interacting with hosts and releasing sulfide and microbial lysate into environments or human systems, such as gastrointestinal tracts[14]. Sulfide produced by virus auxiliary metabolisms indirectly or directly conveys important influences on viral fitness, host physiology, surrounding microbiomes and environments (*ref.* 45[14] and references therein). For example, viruses can stimulate the growth of their host organisms and aid in maintaining a balanced redox state, scavenging free radicals, and promoting the biosynthesis of proteins and amino acids. Assimilated sulfur also contributes to constructing viral structural proteins for virion production. Additionally, thiol modification of nucleic acids can contribute to stress responses and regulate gene expression for both viruses and hosts; thiol components of enzymes are essential for viral functions such as

dsDNA recombination, integration, and dsDNA repair. Finally, the disequilibrium caused by elevated virus-driven sulfide production in conjunction with pathogenic bacteria can accelerate the inflammation in the gastrointestinal tract and further develop into severe diseases, such as colorectal cancer.

The details of dissimilatory and assimilatory sulfur AMGs. The full gene name for each gene abbreviation is listed in Table S2.

AMG genes	**Environments**	**Function**
dsrA and *dsrC* (sulfur oxidation)	Hydrothermal vents (*dsrA* and *dsrC*)[62], Oxygen minimum zones[61] (*dsrC*), Epipelagic ocean (*dsrC*)[120]; Marine (*dsrA*)[13]; marine/saline, human host-associated, and engineered[13] (*dsrC*)	Enzyme complex DsrAB can oxidize elemental sulfur carried by sulfur carrier protein DsrC to sulfite. This is the rate-limiting step for microbial hosts and yields the most electrons. Viruses augment these metabolisms[13], and potential microbial hosts are the SUP05 group of Gammaproteobacteria.
soxYZ (thiosulfate oxidation /disproportionation)	Hydrothermal vents (*soxYZ*)[62], Epipelagic ocean (*soxYZ*)[120]	Sulfur carrier SoxYZ mediates thiosulfate oxidation/disproportionation continuously. Viral SoxYZ probably alleviates the worn-out state of host SoxYZ. Potential hosts include different lineages of Proteobacteria.
soxC and *soxD* (thiosulfate oxidation)	Oil seeps (*soxC* and *soxD*), Freshwater sediment (*soxD*)[13]	SoxCD oxidizes the disulfanyl group to sulfosulfanyl group during thiosulfate oxidation. This reaction yields six electrons. Viruses augment these activities[13].
cysC, cysN, cysD, cysH, cysNC, cysJ (assimilatory sulfate reduction)	Marine (*cysJ*, *cysC*, *cysD*, *cysH*), Freshwater (*cysC*, *cysD*, *cysH*), Engineered (*cysC*, *cysN*, *cysD*, *cysH*), Soil	They mediate the incorporation of sulfide into cysteine for biosynthesis. Viral AMGs likely short-circuit the assimilatory

	(*cysNC*, *cysD*, *cysH*), Hydrothermal vents (*cysC*, *cysH*), Saline and alkaline (*cysC*, *cysH*), Wetlands (*cysH*), Thermal springs (*cysH*), Mammals: gut (*cysD*, *cysH*), other host-associated (*cysH*), Plants: rhizoplane (*cysH*), Human: oral (*cysD*, *cysH*), Human: gut (*cysD*, *cysH*)[14]	sulfur pathway, thus impacting cysteine metabolism and potentially drive sulfide production[14].
cysK, *cysM*, *malY*, *dcyD*, *metC*, *metY* (direct sulfide production)	Marine (*cysK*, *cysM*, *metC*), Freshwater (*cysK*, *metC*), Engineered (*cysK*, *cysM*), Hydrothermal vents (*cysK*), other host-associated (*cysK*, *dcyD*, *metY*), Human: oral (*dcyD*), Human: gut (*cysK*, *cysM*, *malY*, *metY*)[14]	They directly produce sulfide through degradation of organosulfur compounds, i.e., cysteine and homocysteine. The produced sulfide can provide fitness advantages to viruses.
tauD, *ssuD*, *msmA* (direct sulfite production)	Marine (*tauD*, *ssuD*, *msmA*), Freshwater (*tauD*, *ssuD*), Engineered (*tauD*), Saline and alkaline (*tauD*)[14]	They directly produce sulfite through organosulfur compounds, i.e., taurine, alkylsulfonate/isethionate, and methanesulfonate. Sulfite can further be fed into dissimilatory and assimilatory sulfate reduction.
metB, *metH*, *metE*, *metK*, *mtnN*, *dcm*, *ahcY*, *luxS*, *msrC*, *megL*, *aspB* (indirect sulfide production)	Marine (*metK*, *mtnN*, *dcm*, *ahcY*, *megL*, *aspB*), Freshwater (*metB*, *metK*, *mtnN*, *dcm*, *ahcY*, *aspB*), Engineered (*metH*, *metE*, *metK*, *dcm*, *ahcY*, *megL*, *aspB*), Soil (*metH*, *metE*, *dcm*, *aspB*), Hydrothermal vents (*metH*, *metK*, *mtnN*, *dcm*), Saline and alkaline (*metK*, *dcm*), Deep subsurface (*dcm*), Wetlands (*metK*, *mtnN*, *dcm*), Thermal springs (*dcm*), Mammals: gut (*metK*, *mtnN*, *dcm*, *luxS*), other host-associated (*metK*, *dcm*), Plants: rhizoplane (*metH*, *metE*, *dcm*, *ahcY*), Human: oral (*metH*, *metE*, *metK*, *mtnN*, *dcm*, *luxS*, *aspB*), Human: gut (*metE*, *metK*, *mtnN*, *dcm*, *luxS*,	They indirectly produce sulfide by converting organosulfur compounds, such as methionine and cystathionine, to cysteine or homocysteine. The produced sulfide can provide fitness advantages to viruses.

	aspB), Human: vagina (*metK*, *dcm*)[14]	
cysE, *nrnA*, *speE*, *metA*, *mtnK*, *mtnA*, *mtnD*, *lysC*, *thrA*, *asd*, *hom*, *mdh*, *cysQ* (indirect sulfur metabolism)	Marine (*nrnA*, *speE*, *lysC*, *thrA*, *mdh*, *cysQ*), Freshwater (*speE*, *mtnD*, *cysQ*), Engineered (*cysE*, *lysC*, *mdh*), Soil (*cysE*), Hydrothermal vents (*speE*), Deep subsurface (*metA*), Mammals: gut (*lysC*), other host-associated (*cysE*, *nrnA*, *mtnK*, *mtnA*, *lysC*, *asd*), Plants: rhizoplane (*metA*), Human: oral (*nrnA*, *asd*, *mdh*), Human: gut (*metA*, *asd*, *hom*, *mdh*), Human: vagina (*nrnA*, *lysC*, *hom*)[14]	Their Influence extends to the synthesis of organosulfur compounds, which are major precursors for sulfide production.

Box 2. Microbially-mediated interaction of sulfur with other biogeochemical cycles, such as nitrogen, iron, manganese, and phosphorus.

Sulfur is a biogeochemical cycle tightly linked with nitrogen[121] and iron[122]. In marine OMZs, the cryptic sulfur cycle involves denitrification-dependent sulfide oxidation for rapid sulfide removal[50]. This coupling of sulfide oxidation and nitrate reduction to nitrite, N_2, or N_2O has also been observed at the sulfide/nitrate interface of other redox-stratified marine systems[123,124]. It has been reported that sulfide/sulfur-oxidizing microbes (*Thioglobus perditus*) and elemental sulfur can be transported from sulfidic continental shelf waters to offshore OMZs. *Thioglobus perditus* can perform complete denitrification coupled with sulfur and sulfide oxidation[125]. Additionally, ammonium liberated from organic matter remineralization by sulfate reduction contributes to a large fraction of ammonium load for anerobic ammonium oxidation (anammox) in marine OMZs[50]. While not observed to be abundant in the environment, in industrial proceses, sulfate reduction can be coupled with ammonium oxidation under anaerobic conditions, termed “sulfammox” as an alternative for nitrite-dependent anammox[126]. Two bacterial species, *Bacillus Benzoevorans*[127] and *Brocadia Anammoxoglobus sulfate*[128], can perform the sulfammox process. The

combination of anammox, sulfammox, and autotrophic denitrification processes can achieve high efficiency in treating high contents of NH_4^+ and SO_4^{2-} in wastewater. All of these suggest a tight interplay of sulfur and nitrogen cycles[126].

Recent reports indicate that *Shewanella oneidensis* can reduce sulfur to sulfide under alkaline conditions[129]. The resulting hydrogen sulfide can abiotically reduce goethite (α-FeO(OH)), thus overcoming their energetic challenges when performing enzymatic Fe (III) reduction in alkaline environments. This discovery indicates some dissimilatory metal-reducing bacteria like *Shewanella oneidensis* possess multiple enzymatic machineries to reduce Fe (III) and sulfur in various environmental conditions. There are further examples of microorganisms that function as both sulfur and iron metabolizers. Many purple sulfur bacteria (Gammaproteobacteria) and green sulfur bacteria (Chlorobiota) are phototrophic ferrotrophs[130] that oxidize both Fe^{2+} and sulfide. The sulfur reducer *Desulfuromonas acetoxidans* could also conduct dissimilatory Fe(III) reduction when coupled with organic compounds[131]. From a microbial ecology perspective, sulfur and iron cycling can occur simultaneously[83] and the presence of iron-sulfur microbial communities has been documented in various environments, including fjords[132], water columns[83], sediments[133], coal mines[134], wetlands[135], and coastal sediments[122]. The coupling of iron and sulfur cycles is globally widespread and plays a crucial role in biogeochemical processes.

Sulfurimonas marisnigri, a bacterial isolate from the Black Sea, exhibits autotrophic growth and can reduce MnO_2 to Mn_2^+ through the full oxidation of hydrogen sulfide and thiosulfate to sulfate[136]. The genus *Sulfurimonas* is widely distributed in redox-transitional environments, including hydrothermal vents, marine sediments and water columns, and terrestrial habitats[137]. Consistent with the metabolic traits of *Sulfurimonas marisnigri*[136],

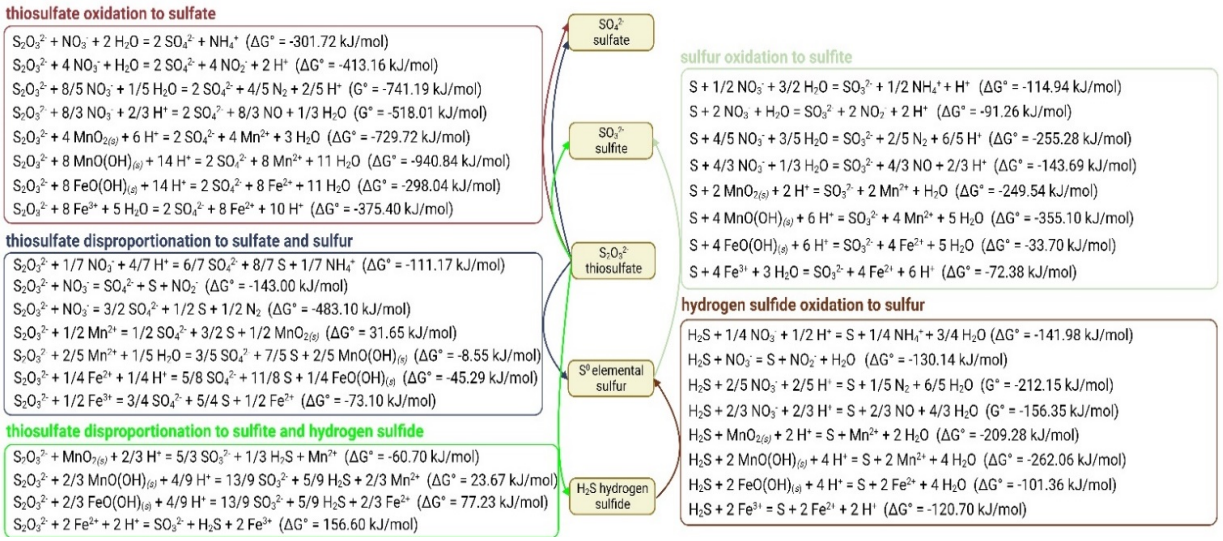
the addition of MnO_2, but not FeOOH, has been found to enhance sulfate production during incubation experiments involving marine sediments[138]. This highlights the potentially extensive biogeochemical consequences of MnO_2-mediated sulfide oxidation in global oceans. Three groups of bacteria, including Methylococcales, Pseudomonadales, and Marinimicrobia, have been found to contain the potential for sulfur and manganese oxidation in their genomes, facilitating energy conservation using hydrothermally-derived reduced substrates[68].

Low-oxygen conditions can favor the release of phosphorus from the sediments to the water column, thus, enhancing the primary production and driving further anoxia[116]. High rates of surface exported organic matter can greatly fuel sulfate reduction in continental shelf sediments coupled to organic matter mineralization in anoxic conditions, thus releasing phosphorus to the water column[139]. Furthermore, some giant sulfur-oxidizing bacteria (such as *Thioploca* and *Thiomargarita*) can store polyphosphate reserves[140]. In low sulfide conditions, these organisms uptake phosphate from the environment and store them as polyphosphate; when sulfide concentrate increases, they oxidize polyphosphate and release dissolved phosphate[140]. The accumulation of phosphate can, in turn, enhance primary production and sustain sulfidic conditions in marine oxygen minimum zones. The feedback between sulfur and phosphorus cycles is crucial for unraveling ecosystem dynamics and comprehending nutrient transformations.

Multiple lines of evidence support the identification of chemolithoautotrophic microorganisms that drive the conversion of methane, nitrogen, and iron, as predicted by thermodynamics[141]. Notable examples include the discovery of anaerobic ammonium oxidation (anammox) bacteria[142] and nitrate/nitrite-dependent anaerobic methane

oxidation (n-DAMO) bacteria[143]. This highlights the notion that the thermodynamically feasible process can be attributed to particular microorganisms or simple combinations thereof, which are so-called “impossible” microorganisms due to highly strict and specific requirements for growth. Sulfur species demonstrate a lower redox potential than nitrogen, oxygen, iron, and manganese, thereby exhibiting a greater thermodynamic tendency for oxidation driven by electron acceptors such as NO_3^-, MnO_2/MnO(OH), and Fe^{3+}/FeO(OH), suggesting that microorganisms can harness these exergonic reactions under anaerobic conditions.

Here, we provide a limited collection of predicted reactions to show the possibility of anaerobic metabolisms mediated by “impossible” microorganisms. Denitrification-dependent sulfide oxidation and Mn (IV) reduction-dependent sulfide/thiosulfate oxidation have been discovered in bacteria belonging to the SUP05/Thioglobaceae clade of Gammaproteobacteria[50,64] and *Sulfurimonas marisnigri*[136], respectively. However, the microorganisms responsible for Fe (III) and Mn (IV/III)-dependent sulfur oxidation and disproportionation have yet-to-be identified. To potentially uncover these elusive microorganisms, investigation of iron and manganese hydroxide-rich seabed or freshwater sediments is crucial, as they provide a promising avenue for discovering these previously unidentified microorganisms.



Highlighted references

Ref 2: Roerdink, D. Redrawing the early sulfur cycle. Nat. Geosci. 13, 526-527, doi:10.1038/s41561-020-0608-z (2020).

Significance: This paper indicates the potential for the Archaean atmosphere to have been adequately oxygenated, as inferred from early-stage sulfur metabolic processes on Earth.

Ref 4: Hinckley, E.-L. S., Crawford, J. T., Fakhraei, H. & Driscoll, C. T. A shift in sulfur-cycle manipulation from atmospheric emissions to agricultural additions. Nat. Geosci. 13, 597-604, doi:10.1038/s41561-020-0620-3 (2020).

Significance: This paper describes the role of sulfur processes in agricultural contexts and their impacts and consequences on human and environmental health.

Ref 13: Kieft, K. et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. Nat. Commun. 12, 3503, doi:10.1038/s41467-021-23698-5 (2021).

Significance: This paper studies the distribution, diversity, and ecology of bacteriophage auxiliary metabolism associated with inorganic sulfur tranformations.

Ref 30: Anantharaman, K. et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 12, 1715-1728, doi:10.1038/s41396-018-0078-0 (2018).

Significance: This paper reveals a substantial degree of previously undiscovered microbial diversity associated with dissimilatory sulfur cycling.

Ref 48: Moran, M. A. & Durham, B. P. Sulfur metabolites in the pelagic ocean. Nat. Rev. Microbiol. 17, 665-678, doi:10.1038/s41579-019-0250-1 (2019).

Significance: This paper provides a summary of the release and uptake of organic sulfur metabolites between marine phytoplanktons and microorganisms.

Ref 50: Canfield, D. E. et al. A cryptic sulfur cycle in oxygen-minimum–zone waters off the Chilean coast. Science 330, 1375-1378, doi:10.1126/science.1196889 (2010).

Significance: This paper unveils a “cryptic sulfur cycle” in the oceans, a process in which sulfide produced by sulfate reduction is immediately reoxidized back to sulfate by microorganisms.

Ref 62: Anantharaman, K. et al. Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. Science 344, 757-760, doi:10.1126/science.1252229 (2014).

Significance: This paper describes the discovery of auxiliary metabolic genes involving sulfur oxidation and chemolithotrophy in viruses.

Ref 66: Dick, G. J. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat. Rev. Microbiol. 17, 271-283, doi:10.1038/s41579-019-0160-2 (2019).

Significance: This review provides a comprehensive summary of the microbiomes and their associated biogeochemical activities in deep-sea hydrothermal vents.

ref 68: Zhou, Z. et al. Sulfur cycling connects microbiomes and biogeochemistry in deep-sea hydrothermal plumes. ISME J., doi:10.1038/s41396-023-01421-0 (2023).

Significance: This paper describes the ecological and evolutionary basis of microbiome adaptations to the sulfur cycling-dominated geochemistry of hydrothermal vents.

Ref 91: Lovelock, J. E., Maggs, R. J. & Rasmussen, R. A. Atmospheric Dimethyl Sulphide and the Natural Sulphur Cycle. Nature 237, 452-453, doi:10.1038/237452a0 (1972).

Significance: This paper describes the role of atmospheric dimethyl sulfide in mediating the transfer of sulfur from the sea to land.

Ref 107: Wolf, P. G. et al. Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. Microbiome 10, 64, doi:10.1186/s40168-022-01242-x (2022).

Significance: This paper uncovers extensive diversity of sulfur metabolic genes in the human gut and highlights the substantial contribution of organic sulfur metabolism to H_2S production and its association with disease.

Ref 117: Mateos, K. et al. The evolution and spread of sulfur cycling enzymes reflect the redox state of the early Earth. Science Advances 9, eade4847, doi:10.1126/sciadv.ade4847 (2023).

Significance: This paper reveals evolutionary paths of early sulfur metablizing enzymes and their association with the redox state of the early Earth.

Competing interests
The authors declare no competing interests.

Acknowledgments
PT was supported by a Natural Science and Engineering Research Council of Canada Doctoral Fellowship and a Wisconsin Distinguished Graduate Award Fellowship from the University of Wisconsin-Madison. ESC is an MSTP student and was supported in part by Medical Scientist Training Program grant T32GM008692 and by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359) at UW-Madison. KA acknowledges support from the US National Science Foundation under grant number OCE 2049478.

References

1 Dahiya, S. *et al.* Ranking the World's Sulfur Dioxide (SO_2) Hotspots: 2019–2020. 48 (Center for Research on Energy and Clean Air & Greenpeace India, Chennai, India, 2020).

2 Roerdink, D. Redrawing the early sulfur cycle. *Nat. Geosci.* **13**, 526-527, doi:10.1038/s41561-020-0608-z (2020).

3 Watts, S. F. The mass budgets of carbonyl sulfide, dimethyl sulfide, carbon disulfide and hydrogen sulfide. *Atmos. Environ.* **34**, 761-779, doi:10.1016/S1352-2310(99)00342-8 (2000).

4 Hinckley, E.-L. S., Crawford, J. T., Fakhraei, H. & Driscoll, C. T. A shift in sulfur-cycle manipulation from atmospheric emissions to agricultural additions. *Nat. Geosci.* **13**, 597-604, doi:10.1038/s41561-020-0620-3 (2020).

5 Rubin, H. J. *et al.* Global Nitrogen and Sulfur Budgets Using a Measurement-Model Fusion Approach. *EGUsphere* **2022**, 1-22, doi:10.5194/egusphere-2022-873 (2022).

6 Lens, P. N. L. & Kuenen, J. G. The biological sulfur cycle: novel opportunities for environmental biotechnology. *Water Sci. Technol.* **44**, 57-66, doi:10.2166/wst.2001.0464 (2001).

7 Nehb, W. & Vydra, K. in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH, 2000).

8 Wu, B. *et al.* Microbial sulfur metabolism and environmental implications. *Sci. Total Environ.* **778**, 146085, doi:10.1016/j.scitotenv.2021.146085 (2021).

9 Simon, J. & Kroneck, P. M. H. in *Adv. Microb. Physiol.* Vol. 62 (ed Robert K. Poole) 45-117 (Academic Press, 2013).

10 Sousa, F. M., Pereira, J. G., Marreiros, B. C. & Pereira, M. M. Taxonomic distribution, structure/function relationship and metabolic context of the two families of sulfide dehydrogenases: SQR and FCSD. *Biochim. Biophys. Acta* **1859**, 742-753, doi:10.1016/j.bbabio.2018.04.004 (2018).

11 Baker, B. J. *et al.* Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* **5**, 887-900, doi:10.1038/s41564-020-0715-z (2020).

12 Offre, P., Spang, A. & Schleper, C. Archaea in Biogeochemical Cycles. *Annu. Rev. Microbiol.* **67**, 437-457, doi:10.1146/annurev-micro-092412-155614 (2013).

13 Kieft, K. *et al.* Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* **12**, 3503, doi:10.1038/s41467-021-23698-5 (2021).

14 Kieft, K. *et al.* Virus-associated organosulfur metabolism in human and environmental systems. *Cell Reports* **36**, 109471, doi:10.1016/j.celrep.2021.109471 (2021).

15 Friend, J. P. in *Chemistry of the Lower Atmosphere* (ed S. I. Rasool) 177-201 (Springer US, 1973).

16 Madigan, M., Bender, K., Buckley, D., Sattley, W. & Stahl, D. *Brock Biology of Microorganisms. 15th Global Edition*. (Pearson Education, 2018).

17 Eriksen, J. in *Advances in Agronomy* Vol. 102 55-89 (Academic Press, 2009).

18 Hopkins, D. W. in *Soil analysis in forensic taphonomy* 65-78 (CRC Press, 2008).

19 Song, L., Wang, Y., Zhang, R. & Yang, S. Microbial Mediation of Carbon, Nitrogen, and Sulfur Cycles During Solid Waste Decomposition. *Microb. Ecol.*, doi:10.1007/s00248-022-02056-y (2022).
20 Suzuki, I. Microbial leaching of metals from sulfide minerals. *Biotechnol. Adv.* **19**, 119-132, doi:10.1016/S0734-9750(01)00053-2 (2001).
21 Bottrell, S. H. & Newton, R. J. Reconstruction of changes in global sulfur cycling from marine sulfate isotopes. *Earth-Sci. Rev.* **75**, 59-83, doi:10.1016/j.earscirev.2005.10.004 (2006).
22 *Hydrogen sulfide*. (World Health Organization, 1981).
23 Edwards, P. J. Sulfur cycling, retention, and mobility in soils: A review. (U.S. Department of Agriculture, Forest Service, Northeastern Research Station, 1998).
24 Wilhelm Scherer, H. Sulfur in soils. *J. Plant Nutr. Soil Sci.* **172**, 326-335, doi:10.1002/jpln.200900037 (2009).
25 Picard, A., Gartman, A. & Girguis, P. R. What Do We Really Know about the Role of Microorganisms in Iron Sulfide Mineral Formation? *Front. Earth Sci.* **4**, 68, doi:10.3389/feart.2016.00068 (2016).
26 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219, doi:10.1038/ncomms13219 (2016).
27 Hug, L. A. & Co, R. It Takes a Village: Microbial Communities Thrive through Interactions and Metabolic Handoffs. *mSystems* **3**, e00152-00117, doi:10.1128/mSystems.00152-17 (2018).
28 Caspi, R. *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34**, D511-516, doi:10.1093/nar/gkj128 (2006).
29 Rohwerder, T. & Sand, W. The sulfane sulfur of persulfides is the actual substrate of the sulfur-oxidizing enzymes from *Acidithiobacillus* and *Acidiphilium* spp. *Microbiology* **149**, 1699-1710, doi:10.1099/mic.0.26212-0 (2003).
30 Anantharaman, K. *et al.* Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* **12**, 1715-1728, doi:10.1038/s41396-018-0078-0 (2018).
31 Pott, A. S. & Dahl, C. Sirohaem sulfite reductase and other proteins encoded by genes at the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology* **144**, 1881-1894, doi:10.1099/00221287-144-7-1881 (1998).
32 Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A. & Fischer, J. Prokaryotic sulfur oxidation. *Curr. Opin. Microbiol.* **8**, 253-259, doi:10.1016/j.mib.2005.04.005 (2005).
33 Kappler, U. & Dahl, C. Enzymology and molecular biology of prokaryotic sulfite oxidation1. *FEMS Microbiol. Lett.* **203**, 1-9, doi:10.1111/j.1574-6968.2001.tb10813.x (2001).
34 Boughanemi, S., Infossi, P., Giudici-Orticoni, M.-T., Schoepp-Cothenet, B. & Guiral, M. Sulfite oxidation by the quinone-reducing molybdenum sulfite dehydrogenase SoeABC from the bacterium Aquifex aeolicus. *Biochim. Biophys. Acta* **1861**, 148279, doi:10.1016/j.bbabio.2020.148279 (2020).

35 Laska, S., Lottspeich, F. & Kletzin, A. Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiology* **149**, 2357-2371, doi:10.1099/mic.0.26455-0 (2003).

36 Schut, G. J., Bridger, S. L. & Adams, M. W. W. Insights into the Metabolism of Elemental Sulfur by the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Characterization of a Coenzyme A- Dependent NAD(P)H Sulfur Oxidoreductase. *J. Bacteriol.* **189**, 4431-4441, doi:10.1128/jb.00031-07 (2007).

37 Klimmek, O., Kröger, A., Steudel, R. & Holdt, G. Growth of *Wolinella succinogenes* with polysulphide as terminal acceptor of phosphorylative electron transport. *Arch. Microbiol.* **155**, 177-182, doi:10.1007/BF00248614 (1991).

38 Kletzin, A. Coupled enzymatic production of sulfite, thiosulfate, and hydrogen sulfide from sulfur: purification and properties of a sulfur oxygenase reductase from the facultatively anaerobic archaebacterium Desulfurolobus ambivalens. *J. Bacteriol.* **171**, 1638-1643, doi:10.1128/jb.171.3.1638-1643.1989 (1989).

39 Hensen, D., Sperling, D., Trüper, H. G., Brune, D. C. & Dahl, C. Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Mol. Microbiol.* **62**, 794-810, doi:10.1111/j.1365-2958.2006.05408.x (2006).

40 Welte, C. *et al.* Interaction between Sox proteins of two physiologically distinct bacteria and a new protein involved in thiosulfate oxidation. *FEBS Lett.* **583**, 1281-1286, doi:10.1016/j.febslet.2009.03.020 (2009).

41 Chen, Z. *et al.* The Complete Pathway for Thiosulfate Utilization in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **84**, e01241-01218, doi:10.1128/AEM.01241-18 (2018).

42 Aketagawa, J., Kobayashi, K. & Ishimoto, M. Purification and Properties of Thiosulfate Reductase from *Desulfovibrio vulgaris*, Miyazaki F1. *The Journal of Biochemistry* **97**, 1025-1032, doi:10.1093/oxfordjournals.jbchem.a135144 (1985).

43 Alami, N. & Hallenbeck, P. C. Cloning and characterization of a gene cluster, *phsBCDEF*, necessary for the production of hydrogen sulfide from thiosulfate by *Salmonella typhimurium*. *Gene* **156**, 53-57, doi:10.1016/0378-1119(94)00930-Q (1995).

44 Stoffels, L., Krehenbrink, M., Berks Ben, C. & Unden, G. Thiosulfate Reduction in *Salmonella enterica* Is Driven by the Proton Motive Force. *J. Bacteriol.* **194**, 475-485, doi:10.1128/jb.06014-11 (2012).

45 Ray, W. K., Zeng, G., Potters, M. B., Mansuri, A. M. & Larson, T. J. Characterization of a 12-Kilodalton Rhodanese Encoded by *glpE* of *Escherichia coli* and Its Interaction with Thioredoxin. *J. Bacteriol.* **182**, 2277-2284, doi:10.1128/jb.182.8.2277-2284.2000 (2000).

46 Nandi, D. L., Horowitz, P. M. & Westley, J. Rhodanese as a thioredoxin oxidase. *Int. J. Biochem. Cell Biol.* **32**, 465-473, doi:10.1016/S1357-2725(99)00035-7 (2000).

47 Cipollone, R., Ascenzi, P., Tomao, P., Imperi, F. & Visca, P. Enzymatic Detoxification of Cyanide: Clues from *Pseudomonas aeruginosa* Rhodanese. *J. Mol. Microbiol. Biotechnol.* **15**, 199-211, doi:10.1159/000121331 (2008).

48 Moran, M. A. & Durham, B. P. Sulfur metabolites in the pelagic ocean. *Nat. Rev. Microbiol.* **17**, 665-678, doi:10.1038/s41579-019-0250-1 (2019).

49 Carbonero, F., Benefiel, A., Alizadeh-Ghamsari, A. & Gaskins, H. R. Microbial pathways in colonic sulfur metabolism and links with health and disease. *Front. Physiol.* **3**, 448, doi:10.3389/fphys.2012.00448 (2012).

50 Canfield, D. E. *et al.* A cryptic sulfur cycle in oxygen-minimum–zone waters off the Chilean coast. *Science* **330**, 1375-1378, doi:10.1126/science.1196889 (2010).

51 Patricia, Q. T. *et al.* Physiological and genomic evidence of cysteine degradation and aerobic hydrogen sulfide production in freshwater bacteria. *mSystems* **8**, e0020123, doi:10.1128/msystems.00201-23 (2021).

52 Sullivan, M. B. *et al.* Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts. *PLoS Biol.* **4**, e234, doi:10.1371/journal.pbio.0040234 (2006).

53 Hurwitz, B. L. & U'Ren, J. M. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin. Microbiol.* **31**, 161-168, doi:10.1016/j.mib.2016.04.002 (2016).

54 Chen, L.-X. *et al.* Large freshwater phages with the potential to augment aerobic methane oxidation. *Nat. Microbiol.* **5**, 1504-1515, doi:10.1038/s41564-020-0779-9 (2020).

55 Bragg, J. G. & Chisholm, S. W. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS One* **3**, e3550, doi:10.1371/journal.pone.0003550 (2008).

56 Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11013-11018, doi:10.1073/pnas.0401526101 (2004).

57 Emerson, J. B. *et al.* Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870-880, doi:10.1038/s41564-018-0190-y (2018).

58 Trubl, G. *et al.* Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems* **3**, e00076-00018, doi:10.1128/mSystems.00076-18 (2018).

59 Ahlgren, N. A., Fuchsman, C. A., Rocap, G. & Fuhrman, J. A. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode *amoC* nitrification genes. *ISME J.* **13**, 618-631, doi:10.1038/s41396-018-0289-4 (2019).

60 Cassman, N. *et al.* Oxygen minimum zones harbour novel viral communities with low diversity. *Environ. Microbiol.* **14**, 3043-3065, doi:10.1111/j.1462-2920.2012.02891.x (2012).

61 Roux, S. *et al.* Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Elife* **3**, e03125, doi:10.7554/eLife.03125 (2014).

62 Anantharaman, K. *et al.* Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science* **344**, 757-760, doi:10.1126/science.1252229 (2014).

63 Lam, P. & Kuypers, M. M. Microbial nitrogen cycling processes in oxygen minimum zones. *Ann. Rev. Mar. Sci* **3**, 317-345, doi:10.1146/annurev-marine-120709-142814 (2011).

64 Stewart, F. J., Ulloa, O. & DeLong, E. F. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ. Microbiol.* **14**, 23-40, doi:10.1111/j.1462-2920.2010.02400.x (2012).

65 Carolan, M. T., Smith, J. M. & Beman, J. M. Transcriptomic evidence for microbial sulfur cycling in the eastern tropical North Pacific oxygen minimum zone. *Front. Microbio.* **6**, 334, doi:10.3389/fmicb.2015.00334 (2015).

66 Dick, G. J. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* **17**, 271-283, doi:10.1038/s41579-019-0160-2 (2019).

67 German, C. R. & Von Damm, K. L. in *Treatise on Geochemistry* (eds Heinrich D. Holland & Karl K. Turekian) 181-222 (Pergamon, 2003).

68 Zhou, Z. *et al.* Sulfur cycling connects microbiomes and biogeochemistry in deep-sea hydrothermal plumes. *ISME J.*, doi:10.1038/s41396-023-01421-0 (2023).

69 Zhou, Z., Tran, P. Q., Kieft, K. & Anantharaman, K. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *ISME J.* **14**, 2060-2077, doi:10.1038/s41396-020-0669-4 (2020).

70 Zhou, Z., St. John, E., Anantharaman, K. & Reysenbach, A.-L. Global patterns of diversity and metabolism of microbial communities in deep-sea hydrothermal vent deposits. *Microbiome* **10**, 241, doi:10.1186/s40168-022-01424-7 (2022).

71 McKay, L. J. *et al.* Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* **4**, 614-622, doi:10.1038/s41564-019-0362-4 (2019).

72 Tamazawa, S., Takasaki, K., Tamaki, H., Kamagata, Y. & Hanada, S. Metagenomic and Biochemical Characterizations of Sulfur Oxidation Metabolism in Uncultured Large Sausage-Shaped Bacterium in Hot Spring Microbial Mats. *PLoS One* **7**, e49793, doi:10.1371/journal.pone.0049793 (2012).

73 Kubo, K., Knittel, K., Amann, R., Fukui, M. & Matsuura, K. Sulfur-metabolizing bacterial populations in microbial mats of the Nakabusa hot spring, Japan. *Syst. Appl. Microbiol.* **34**, 293-302, doi:10.1016/j.syapm.2010.12.002 (2011).

74 Colman, D. R. *et al.* Geobiological feedbacks and the evolution of thermoacidophiles. *ISME J.* **12**, 225-236, doi:10.1038/ismej.2017.162 (2018).

75 Linz, A. M. *et al.* Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* **6**, e6075, doi:10.7717/peerj.6075 (2018).

76 Tran, P. Q. *et al.* Depth-discrete metagenomics reveals the roles of microbes in biogeochemical cycling in the tropical freshwater Lake Tanganyika. *ISME J.* **15**, 1971-1986, doi:10.1038/s41396-021-00898-x (2019).

77 Sigee, D. C. *Freshwater microbiology: biodiversity and dynamic interactions of microorganisms in the aquatic environment*. (John Wiley & Sons, Ltd, 2005).

78 Kushkevych, I., Procházka, J., Gajdács, M., Rittmann, S. K.-M. R. & Vítězová, M. Molecular Physiology of Anaerobic Phototrophic Purple and Green Sulfur Bacteria. *Int. J. Mol. Sci.* **22**, 6398, doi:10.3390/ijms22126398 (2021).

79 Lovley, D. R. & Klug, M. J. Sulfate Reducers Can Outcompete Methanogens at Freshwater Sulfate Concentrations. *Appl. Environ. Microbiol.* **45**, 187-192, doi:10.1128/aem.45.1.187-192.1983 (1983).

80 Garner, R. E. *et al.* A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nat. Microbiol.*, doi:10.1038/s41564-023-01435-6 (2023).

81 Halevy, I., Peters, S. E. & Fischer, W. W. Sulfate Burial Constraints on the Phanerozoic Sulfur Cycle. *Science* **337**, 331-334, doi:10.1126/science.1220224 (2012).
82 Jamieson, J. W., Wing, B. A., Farquhar, J. & Hannington, M. D. Neoarchaean seawater sulphate concentrations from sulphur isotopes in massive sulphide ore. *Nat. Geosci.* **6**, 61-64, doi:10.1038/ngeo1647 (2013).
83 Berg, J. S. *et al.* Microbial diversity involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of Lake Pavin. *PLoS One* **14**, e0212787, doi:10.1371/journal.pone.0212787 (2019).
84 Tsuji, J. M. *et al.* Anoxygenic photosynthesis and iron–sulfur metabolic potential of Chlorobia populations from seasonally anoxic Boreal Shield lakes. *ISME J.* **14**, 2732-2747, doi:10.1038/s41396-020-0725-0 (2020).
85 İnceoğlu, Ö. *et al.* Vertical Distribution of Functional Potential and Active Microbial Communities in Meromictic Lake Kivu. *Microb. Ecol.* **70**, 596-611, doi:10.1007/s00248-015-0612-9 (2015).
86 Burrows, S. M., Elbert, W., Lawrence, M. G. & Pöschl, U. Bacteria in the global atmosphere – Part 1: Review and synthesis of literature data for different ecosystems. *Atmospheric Chem. Phys.* **9**, 9263-9280, doi:10.5194/acp-9-9263-2009 (2009).
87 Sattler, B., Puxbaum, H. & Psenner, R. Bacterial growth in supercooled cloud droplets. *Geophys. Res. Lett.* **28**, 239-242, doi:10.1029/2000GL011684 (2001).
88 Hill, K. A. *et al.* Processing of atmospheric nitrogen by clouds above a forest environment. *J. Geophys. Res. Atmos.* **112**, doi:10.1029/2006JD008002 (2007).
89 Todd, J. D. *et al.* Structural and Regulatory Genes Required to Make the Gas Dimethyl Sulfide in Bacteria. *Science* **315**, 666-669, doi:10.1126/science.1135370 (2007).
90 Curson, A. R. J., Todd, J. D., Sullivan, M. J. & Johnston, A. W. B. Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nat. Rev. Microbiol.* **9**, 849-859, doi:10.1038/nrmicro2653 (2011).
91 Lovelock, J. E., Maggs, R. J. & Rasmussen, R. A. Atmospheric Dimethyl Sulphide and the Natural Sulphur Cycle. *Nature* **237**, 452-453, doi:10.1038/237452a0 (1972).
92 Šantl-Temkiv, T., Amato, P., Casamayor, E. O., Lee, P. K. H. & Pointing, S. B. Microbial ecology of the atmosphere. *FEMS Microbiol. Rev.* **46**, fuac009, doi:10.1093/femsre/fuac009 (2022).
93 Patterson, J. P. *et al.* Sea Spray Aerosol Structure and Composition Using Cryogenic Transmission Electron Microscopy. *ACS Central Science* **2**, 40-47, doi:10.1021/acscentsci.5b00344 (2016).
94 Liu, H. *et al.* Airborne microorganisms exacerbate the formation of atmospheric ammonium and sulfate. *Environ. Pollut.* **263**, 114293, doi:10.1016/j.envpol.2020.114293 (2020).
95 Chin, M. & Davis, D. D. A reanalysis of carbonyl sulfide as a source of stratospheric background sulfur aerosol. *J. Geophys. Res. Atmos.* **100**, 8993-9005, doi:10.1029/95JD00275 (1995).

96 Lennartz, S. T. *et al.* Marine carbonyl sulfide (OCS) and carbon disulfide (CS_2): a compilation of measurements in seawater and the marine boundary layer. *Earth Syst. Sci. Data* **12**, 591-609, doi:10.5194/essd-12-591-2020 (2020).
97 Greening, C. & Grinter, R. Microbial oxidation of atmospheric trace gases. *Nat. Rev. Microbiol.* **20**, 513-528, doi:10.1038/s41579-022-00724-x (2022).
98 Barber, S. A. *Soil nutrient bioavailability: a mechanistic approach*. (John Wiley & Sons, 1995).
99 Beller, H. R. *et al.* The Genome Sequence of the Obligately Chemolithoautotrophic, Facultatively Anaerobic Bacterium *Thiobacillus denitrificans*. *J. Bacteriol.* **188**, 1473-1488, doi:10.1128/jb.188.4.1473-1488.2006 (2006).
100 Xia, Y. *et al.* Sulfide production and oxidation by heterotrophic bacteria under aerobic conditions. *ISME J.* **11**, 2754-2766, doi:10.1038/ismej.2017.125 (2017).
101 Kertesz, M. A. & Mirleau, P. The role of soil microbes in plant sulphur nutrition. *J. Exp. Bot.* **55**, 1939-1945, doi:10.1093/jxb/erh176 (2004).
102 Schoenau, J. J. & Malhi, S. S. in *Sulfur: A Missing Link between Soils, Crops, and Nutrition* 1-10 (American Society of Agronomy, 2008).
103 Kovar, J. L. & Grant, C. A. in *Soil Management: Building a Stable Base for Agriculture* 103-115 (American Society of Agronomy, 2011).
104 Barton, L. L., Ritz, N. L., Fauque, G. D. & Lin, H. C. Sulfur Cycling and the Intestinal Microbiome. *Dig. Dis. Sci.* **62**, 2241-2257, doi:10.1007/s10620-017-4689-5 (2017).
105 Winyard, Paul G. *et al.* Measurement and meaning of markers of reactive species of oxygen, nitrogen and sulfur in healthy human subjects and patients with inflammatory joint disease. *Biochem. Soc. Trans.* **39**, 1226-1232, doi:10.1042/BST0391226 (2011).
106 Kushkevych, I. *et al.* Recent Advances in Metabolic Pathways of Sulfate Reduction in Intestinal Bacteria. *Cells* **9**, 698, doi:10.3390/cells9030698 (2020).
107 Wolf, P. G. *et al.* Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. *Microbiome* **10**, 64, doi:10.1186/s40168-022-01242-x (2022).
108 Yang, G. *et al.* H_2S as a Physiologic Vasorelaxant: Hypertension in Mice with Deletion of Cystathionine γ-Lyase. *Science* **322**, 587-590, doi:10.1126/science.1162667 (2008).
109 Attene-Ramos, M. S., Wagner, E. D., Plewa, M. J. & Gaskins, H. R. Evidence That Hydrogen Sulfide Is a Genotoxic Agent. *Mol. Cancer Res.* **4**, 9-14, doi:10.1158/1541-7786.MCR-05-0126 (2006).
110 Deplancke, B. & Gaskins, H. R. Hydrogen sulfide induces serum-independent cell cycle entry in nontransformed rat intestinal epithelial cells. *FASEB J.* **17**, 1310-1312, doi:10.1096/fj.02-0883fje (2003).
111 Sartor, R. B. Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology* **134**, 577-594, doi:10.1053/j.gastro.2007.11.059 (2008).
112 Ridlon, J. M., Wolf, P. G. & Gaskins, H. R. Taurocholic acid metabolism by gut microbes and colon cancer. *Gut Microbes* **7**, 201-215, doi:10.1080/19490976.2016.1150414 (2016).
113 Gunson, J. R. *et al.* Climate sensitivity to ocean dimethylsulphide emissions. *Geophys. Res. Lett.* **33**, L07701, doi:10.1029/2005GL024982 (2006).

114 Hopkins, F. E., Archer, S. D., Bell, T. G., Suntharalingam, P. & Todd, J. D. The biogeochemistry of marine dimethylsulfide. *Nature Reviews Earth & Environment* **4**, 361-376, doi:10.1038/s43017-023-00428-7 (2023).

115 Jackson, R. & Gabric, A. Climate Change Impacts on the Marine Cycling of Biogenic Sulfur: A Review. *Microorganisms* **10**, 1581, doi:10.3390/microorganisms10081581 (2022).

116 Callbeck, C. M. *et al.* Sulfur cycling in oceanic oxygen minimum zones. *Limnol. Oceanogr.* **66**, 2360-2392, doi:10.1002/lno.11759 (2021).

117 Mateos, K. *et al.* The evolution and spread of sulfur cycling enzymes reflect the redox state of the early Earth. *Science Advances* **9**, eade4847, doi:10.1126/sciadv.ade4847 (2023).

118 Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A. & Fischer, J. Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? *Appl. Environ. Microbiol.* **67**, 2873-2882, doi:10.1128/AEM.67.7.2873-2882.2001 (2001).

119 Druschel, G. K., Hamers, R. J. & Banfield, J. F. Kinetics and mechanism of polythionate oxidation to sulfate at low pH by O_2 and Fe_3^+. *Geochim. Cosmochim. Acta* **67**, 4457-4469, doi:10.1016/S0016-7037(03)00388-0 (2003).

120 Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689, doi:10.1038/nature19366 (2016).

121 Madsen, E. L. Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* **22**, 456-464, doi:10.1016/j.copbio.2011.01.008 (2011).

122 Thamdrup, B., Fossing, H. & Jørgensen, B. B. Manganese, iron and sulfur cycling in a coastal marine sediment, Aarhus bay, Denmark. *Geochim. Cosmochim. Acta* **58**, 5115-5129, doi:10.1016/0016-7037(94)90298-4 (1994).

123 Lavik, G. *et al.* Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**, 581-584, doi:10.1038/nature07588 (2009).

124 Walsh, D. A. *et al.* Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**, 578-582, doi:10.1126/science.1175309 (2009).

125 Callbeck, C. M. *et al.* Oxygen minimum zone cryptic sulfur cycling sustained by offshore transport of key sulfur oxidizing bacteria. *Nat. Commun.* **9**, 1729, doi:10.1038/s41467-018-04041-x (2018).

126 Dominika, G., Joanna, M. & Jacek, M. Sulfate reducing ammonium oxidation (SULFAMMOX) process under anaerobic conditions. *Environ. Technol. Innov.* **22**, 101416, doi:10.1016/j.eti.2021.101416 (2021).

127 Cai, J., Jiang, J. & Zheng, P. Isolation and identification of bacteria responsible for simultaneous anaerobic ammonium and sulfate removal. *Sci. China Chem.* **53**, 645-650, doi:10.1007/s11426-010-0053-8 (2010).

128 Liu, S. *et al.* Application of anaerobic ammonium-oxidizing consortium to achieve completely autotrophic ammonium and sulfate removal. *Bioresour. Technol.* **99**, 6817-6825, doi:10.1016/j.biortech.2008.01.054 (2008).

129 Flynn, T. M., O'Loughlin, E. J., Mishra, B., DiChristina, T. J. & Kemner, K. M. Sulfur-mediated electron shuttling during bacterial iron reduction. *Science* **344**, 1039-1042, doi:10.1126/science.1252066 (2014).

130 Kappler, A. *et al.* An evolving view on biogeochemical cycling of iron. *Nat. Rev. Microbiol.* **19**, 360-374, doi:10.1038/s41579-020-00502-7 (2021).
131 Roden Eric, E. & Lovley Derek, R. Dissimilatory Fe(III) Reduction by the Marine Microorganism *Desulfuromonas acetoxidans*. *Appl. Environ. Microbiol.* **59**, 734-742, doi:10.1128/aem.59.3.734-742.1993 (1993).
132 Buongiorno, J. *et al.* Complex Microbial Communities Drive Iron and Sulfur Cycling in Arctic Fjord Sediments. *Appl. Environ. Microbiol.* **85**, e00949-00919, doi:10.1128/AEM.00949-19 (2019).
133 Jin, L. *et al.* Abundant iron and sulfur oxidizers in the stratified sediment of a eutrophic freshwater reservoir with annual cyanobacterial blooms. *Sci. Rep.* **7**, 43814, doi:10.1038/srep43814 (2017).
134 Küsel, K. Microbial cycling of iron and sulfur in acidic coal mining lake sediments. *Water Air Soil Pollut. Focus* **3**, 67-90, doi:10.1023/A:1022103419928 (2003).
135 Karimian, N., Johnston, S. G. & Burton, E. D. Iron and sulfur cycling in acid sulfate soil wetlands under dynamic redox conditions: A review. *Chemosphere* **197**, 803-816, doi:10.1016/j.chemosphere.2018.01.096 (2018).
136 Henkel, J. V. *et al.* A bacterial isolate from the Black Sea oxidizes sulfide with manganese(IV) oxide. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12153-12155, doi:10.1073/pnas.1906000116 (2019).
137 Han, Y. & Perner, M. The globally widespread genus *Sulfurimonas*: versatile energy metabolisms and adaptations to redox clines. *Front. Microbio.* **6**, 989, doi:10.3389/fmicb.2015.00989 (2015).
138 Aller, R. C. & Rude, P. D. Complete oxidation of solid phase sulfides by manganese and bacteria in anoxic marine sediments. *Geochim. Cosmochim. Acta* **52**, 751-765, doi:10.1016/0016-7037(88)90335-3 (1988).
139 Lomnitz, U. *et al.* Benthic phosphorus cycling in the Peruvian oxygen minimum zone. *Biogeosciences* **13**, 1367-1386, doi:10.5194/bg-13-1367-2016 (2016).
140 Brock, J. & Schulz-Vogt, H. N. Sulfide induces phosphate release from polyphosphate in cultures of a marine Beggiatoa strain. *ISME J.* **5**, 497-506, doi:10.1038/ismej.2010.135 (2011).
141 in 't Zandt, M. H., de Jong, A. E. E., Slomp, C. P. & Jetten, M. S. M. The hunt for the most-wanted chemolithoautotrophic spookmicrobes. *FEMS Microbiol. Ecol.* **94**, fiy064, doi:10.1093/femsec/fiy064 (2018).
142 Kuenen, J. G. Anammox bacteria: from discovery to application. *Nat. Rev. Microbiol.* **6**, 320-326, doi:10.1038/nrmicro1857 (2008).
143 Raghoebarsing, A. A. *et al.* A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* **440**, 918-921, doi:10.1038/nature04617 (2006).

Glossary
Sulfide minerals

A class of minerals containing sulfide or disulfide as the major anion. The metal cation usually comes from Fe^{2+}, Cu^{2+}, Pb^{2+}, Zn^{2+}, Ag^{+}, and Hg^{2+}. Sulfide minerals deposited in

sediments, such as pyrite (FeS_2), can be leached by microorganisms or oxidatively weathered to sulfate and enter the terrestrial hydrosphere for sulfur cycling.

Atmospheric deposition

A transport process for gases and particles from the atmosphere deposited onto the surface of terrestrial and aquatic systems. Elements, nutrients, and pollutants are transported in the process, for example, nitrogen, sulfur, and heavy metals. This involves contains wet and dry depositions. The former indicates deposition by means of rain, sleet, snow, or fog; the latter indicates deposition without precipitation processes.

Mineralization

The conversion of chemical compounds in organic matter to inorganic forms. The soluble chemicals are then available to plants.

Immobilization

The opposite process of mineralization that converts inorganic compounds to organic compounds by plants or microorganisms. The compounds are inaccessible to plants.

Disproportionation

A redox reaction where the substrate is in the intermediate oxidation state, and the products are two compounds with one of a higher oxidation state and the other of a lower oxidation state.

Cryptic sulfur cycling

A tight link between sulfate reduction and sulfide oxidation enables rapid sulfide turnover, ensuring sulfide concentration remains below detectable levels. Traditional

biogeochemical measurements are unable to directly detect the sulfur transformation process. Instead, omics-based profiling of activities or stable isotope labeling of reactions are relied upon to investigate these process.

Auxiliary metabolic genes

Host-derived metabolic genes encoded by viruses to influence and manipulate their hosts. They either augment, redirect, or short-circuit host metabolisms by encoding corresponding enzymes.

Chemosynthetic

Chemosynthetic processes involve the biological conversion of inorganic carbon into organic matter for cellular biosynthesis using energy sources other than sunlight.

Holomictic lake

Lakes typically experience an annual occurrence where the temperature and density of the surface and bottom layers are the same, facilitating the mixing of water from distinct layers.

Meromictic lake

Meromictic lake is the opposite of holomictic lake in that its layers never intermix.

Dissimilatory sulfate/sulfite reduction

A form of anaerobic respiration where sulfate/sulfite is used as the electron acceptor by microorganisms to produce hydrogen sulfide.

Assimilatory sulfate/sulfite reduction

A process where sulfate/sulfite is reduced to cysteine for biosynthesis in living organisms. Sulfate/sulfite is assimilated to organic sulfur compounds eventually though hydrogen sulfide can be an intermediate during the process.