Systems approaches to infer microbial community interactions and their impacts on ecosystem function

Alex Carr

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Nitin Baliga, Chair

Sean Gibbons

Dave Stahl

Amy Willis

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

Systems approaches to infer microbial community interactions and their impacts on ecosystem function

Alex Carr

Chair of the Supervisory Committee:
Nitin Baliga
Institute for Systems Biology
Molecular Engineering and Sciences

Microbes and the communities they form play critical roles in ecosystem function, from facilitating the biogeochemical cycling of essential nutrients, such as carbon, nitrogen, and sulfur, to acting as the foundation of complex food webs. Microbes also play important roles as eukaryotic symbionts, where they can have profound effects on host fitness. Thus, understanding how microbial community interactions and environmental context shape the functional capabilities of microbiomes is of vital importance if we want to engineer these systems to address challenges in human health and the health of natural ecosystems. Here, I show how systems approaches can be leveraged to overcome the inherent limitations of inferring microbial community interactions directly from correlation structure, which has been the standard approach in the microbiome field. Specifically, I highlight how multi-omic characterization, MCMMs, and synthetic communities (SynComs) can be leveraged to better understand niche competition between nitrate-reducing bacteria and sulfate-reducing bacteria in oxygen-depleted ecosystems, the importance of pathway partitioning in nitrate-reducing communities, and the role of nitrate-reducing communities in nitrous oxide emissions. I also highlight how microbial community-scale metabolic models (MCMMs) can be leveraged to predict *Clostridioides difficile* (*C. difficile*) colonization in the human gut microbiome, provide mechanistic insights into the niche of *C. difficile* across different community contexts, and assess probiotic interventions designed to inhibit *C. difficile* growth.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This thesis acknowledgment is a tribute to all the people who aided me in my academic journey. First of all, I would like to thank my supervisors, Nitin Baliga and Sean Gibbons, for pushing me to become a better scientist and whose unwavering support has been instrumental in the completion of this thesis. I would also like to acknowledge the support and contributions of the many individuals that I have worked with during my time at the Institute for Systems Biology and the University of Washington. In particular, I would like to thank Christian Diener, Anne Otwell, Kris Hunt, James Wilson, Jake Valenzuela, Serdar Turkarslan, and Dave Stahl. Additionally, I would like to thank my Molecular Engineering and Sciences program advisors past and present Doug Ballard and Paul Neubert for their support in helping me navigate the various academic hurdles that presented themselves while I was a graduate student. Finally, I would like to thank my friends and family for always being there for me through the good times and the bad.

Chapter 1. INTRODUCTION

1.1 THE IMPORTANCE OF MICROBIAL ECOLOGY

Microbes and the communities they form encompass a tremendous amount of biodiversity, are ubiquitous across ecosystems, and provide critical support to all life forms, including humans [1]. Most ecosystem processes essential to life on Earth are mediated by microbes [2]. Among these are the cycling of nitrogen, where microbes transform atmospheric nitrogen into bioavailable forms consumed by plants and animals, the breakdown of complex carbon largely inaccessible to other forms of life, and the production of atmospheric oxygen [3, 4]. In recent years, thanks to large-scale initiatives such as the Earth Microbiome Project and the Human Microbiome Project, significant progress has been made in cataloging the diversity of microorganisms in the natural environment and our bodies [5, 6]. These efforts have led to significant developments in our understanding of how microbial biodiversity associates with host phenotypes and environmental processes. We have learned, for instance, that our intestinal microbiota plays key roles in our nutrition, reproduction, behavior, and overall health and, when perturbed, can lead to the development of a variety of disorders [7–9]. We have also gained deeper insight into the ways abiotic environmental factors like pH, nutrient availability, and temperature shape observed biodiversity. As a consequence, recent work has shown that natural bacterial communities can serve as quantitative geochemical biosensors [10]. However, mechanistically linking observed biodiversity to community function, host phenotype, and ecosystem processes is still a major challenge with many open questions yet to be answered [11, 12].

1.2 CHALLENGES IN MICROBIAL ECOLOGY: PREDICTIVE UNDERSTANDING OF COMMUNITY FUNCTION AND DYNAMICS

Building predictive models that link community function with composition is an emerging challenge in microbial ecology [11]. We are often limited in our ability to directly observe interactions between microbial species and thus are often unable to accurately infer functional relationships from community composition in complex, highly dynamic, communities. In these settings leveraging correlation-based methods to infer community interactions and structure-function relationships is tempting and can be informative, but is fraught with challenges [13]. The most definitive work on microbial interactions has been done experimentally. For example, microscopy and staining techniques, along with stable isotope labeling, have been employed to observe co-localization and cross-feeding between methanotrophic archaea and sulfate-reducing bacteria [14]. In addition to mutualistic interactions, direct bacterial antagonism through type VI secretion systems has been demonstrated using a combination of genomics, microscopy, and co-culturing assays [15]. Entire interaction networks have been determined in simplified microbial consortia consisting of a few species, where community membership can be manipulated to assess pairwise and higher-order interactions [16, 17]. These experimental approaches represent gold standards for inferring interactions between microorganisms, but they are often difficult and time consuming. Furthermore, laboratory-based studies can fail to capture the environmental context in which natural interactions occur. Recent work has demonstrated just how important this context can be in mediating interactions [18]. Thus, it is not practical to apply these experimental methods to all potential interactions between thousands of taxa, many of which cannot be cultured. As such, there is a strong incentive for carefully choosing simplified consortia for in-depth laboratory characterization and leveraging methods, both experimental and computational, that complement observational data collected from complex natural communities. Systems biology approaches, like the integration of metagenomics, transcriptomics, and metabolomics with gene regulatory and genome-scale metabolic models provide a means to address the challenge of linking community function with ecological composition [19–21].

1.3 SYSTEMS BIOLOGY APPROACHES TO CHARACTERIZE COMMUNITY FUNCTION

Systems biology seeks to understand the complexity of biological systems by considering the entirety of the system as an emergent property of its parts. Whether at the levels of cells, tissues, organs, or whole ecosystems, this approach attempts to integrate quantitative information about the system to gain a more holistic view of its function [22]. For microbial ecology, this means integrating molecular features (e.g., biomass, genes, transcripts, proteins, metabolites) to gain insights into the physiology of individual microbes and the functions of whole communities [19]. While systems biology is by no means a new concept, its practice has become more accessible in recent years as a result of advances in high-throughput experimental methodologies like metagenomics, transcriptomics, and metabolomics. These methodologies have enabled the collection of a vast amount of molecular data and provided fresh perspectives with which to examine complex biological phenomena [23, 24]. Alongside recent experimental advances, advances in computation and modeling have enabled increasingly accurate representations of cellular physiology for individual microbes and whole communities. By leveraging the available molecular data, contemporary modeling frameworks can capture environmental context-dependent changes in gene regulation, simulate metabolic fluxes, and even integrate both to better link cellular genotype to phenotype [21, 25, 26]. With these tools in hand, it is now more feasible than ever before to develop a predictive understanding of community function and dynamics and gain mechanistic insights into microbial interactions and their impacts on ecosystem function.

1.4 OUTLOOK

The work presented throughout this document demonstrates how systems biology approaches can be leveraged to gain mechanistic insights into microbial physiology and the roles played by microbes in the ecosystems they inhabit.

Chapter 2 highlights the influence reduced sulfur metabolites can have on the physiology of soil dwelling nitrate-reducing bacteria (NRB). Through comparative transcriptomic and metabolomic analyses, mechanisms of hydrogen sulfide- and cysteine-mediated inhibition of nitrate respiratory growth were identified for the NRB *Intrasporangium calvum C5*. The results of this analysis provide insights into the competitive interactions that can occur in anoxic environments between NRB and sulfate-reducing bacteria (SRB).

Chapter 3 highlights the role of pathway partitioning in nitrate-reducing communities and how environmental context can impact community phenotype. Through detailed physiological characterization and kinetic modeling of a cooperative synthetic community (SynCom) assembled by pairing NRB isolates from a field site heavily contaminated with nitrate, this analysis provides insight into the controls of nitrous oxide emissions. The results of this analysis highlight the balances that must be struck by denitrifying communities and the role nitrite toxicity can play in nitrous oxide production.

Chapter 4 highlights how microbial community-scale metabolic models (MCMMs) can be leveraged to accurately predict known instances of *Clostridiodies difficile* (*C. difficile*) colonization susceptibility or resistance *in vitro* and *in vivo*. Detailed mechanistic insights into the ecological interactions that govern *C. difficile* engraftment, like cross-feeding and competition are highlighted, as well as their variation from person to person. The analysis presented also shows how MCMMs can be leveraged to predict personalized engraftment and *C. difficile* growth suppression for a probiotic cocktail (VE303) designed to replace FMTs for the treatment recurrent *C. difficile* infection (rCDI).

Chapter 2. SULFUR METABOLITES PLAY KEY SYSTEMS-LEVEL ROLES IN MODULATING DENITRIFICATION

2.1 BACKGROUND

Over the past century, the nitrogen cycle has become increasingly perturbed by anthropogenic inputs of fixed nitrogen from synthetic fertilizers, industrial activity, and wastewater. This perturbation has global effects, contributing to increased groundwater contamination, eutrophication, and emissions of the greenhouse gas nitrous oxide (N_2O) [27]. In the environment, microbially driven redox reactions are mainly responsible for transforming nitrogen-containing compounds [4]. Microbial activities are impacted by factors that include resource concentration, pH, metal availability, and interactions between microorganisms in the environment [28–30]. Knowing how these factors differentially impact subsurface microbial nitrogen cycling is essential for developing a predictive understanding of the fate of different nitrogen species in natural and engineered systems. This has been a specific focus at the Oak Ridge Field Research Center (FRC), a Department of Energy legacy nuclear waste site with high levels of heavy metals and radionuclides [10]. Nitrate (NO_3^-) is a major co-contaminant at the site and has been shown to negatively impact the remediation potential of uranium (U) by hindering U(VI) reduction by sulfate-reducing bacteria (SRB) and iron-reducing bacteria [31] Recent work has also shown that groundwater NO_3^- levels are associated with N_2O production at the FRC and enrichment of denitrification pathway genes [10, 32]. Taken together, these observations suggest that dissimilatory nitrate reduction by nitrate-reducing bacteria (NRB) is a major metabolic process in nitrate-contaminated regions of the subsurface.

Both SRB and NRB play a variety of important roles in subsurface environments such as the FRC. In addition to reducing U(VI), SRB can stabilize heavy metals such as lead, zinc, and cadmium through the formation of insoluble metal sulfides, thus providing a means to sustainably sequester and immobilize these metals [33]. SRB also play important roles in marine systems, where they mineralize an estimated 29% of the organic matter deposited to the seafloor. Meanwhile, in sediments, NRB can facilitate the retention of nitrogen via dissimilatory nitrate reduction to ammonia (DNRA), converting mobile NO_3^- to ammonium (NH_4^+) retained by the soil matrix [34]. The activity of denitrifiers can also lead to significant losses of soil nitrogen through the reduction of NO_3^- to gaseous forms, including N_2O or dinitrogen gas (N_2) [35].

Previous studies have shown that intermediates derived from nitrogen and sulfur cycling (e.g., hydrogen sulfide, nitrite (NO_2^-), etc.) can impact the activities of SRB and NRB in oxygen-limited environments, such as sediments and the terrestrial subsurface. For example, as a consequence of the higher energy yield of NO_3^- respiration relative to that of sulfate respiration and NRB production of toxic NO_2^-, SRB are generally less competitive for carbon when NO_3^- is available [36]. For this reason, NO_3^- is often used to prevent the souring of oil reservoirs by stimulating NRB activity and depressing SRB activity [37]. However, the production of hydrogen sulfide by SRB is also known to impact the transformation of different nitrogen species. Heterotrophic denitrification is inhibited by hydrogen sulfide in Pseudomonas, and inhibition of two key enzymes in the pathway (nitric oxide and nitrous oxide reductase) was thought to be involved [38]. However, there has not been much progress on further characterizing the mechanism of sulfide-induced inhibition of NRB. Prior studies have also demonstrated that hydrogen sulfide can cause NRB to shift from denitrification to DNRA [39]. Additionally, inhibiting the activity of ammonia oxidizers indirectly influences both the NRB and SRB by depressing the conversion of NH_4 to

NO_2^- [40]. Despite much empirical evidence regarding the influences of SRB and NRB on one another, the specific mechanism(s) by which reduced sulfur species act to impact the activity of NRB remains unresolved. Here, we have used a systems approach to characterize the mechanisms by which reduced sulfur compounds modulate the growth and activity of NRB. We have characterized the effects of two naturally occurring reduced sulfur compounds, hydrogen sulfide (here referred to as sulfide) and cysteine, on *Intrasporangium calvum C5*. Whereas sulfide is encountered as an end product of sulfate respiration by SRB, cysteine is typically used as a reducing agent to achieve anoxic conditions in laboratory growth medium [28]. Interestingly, we observed that cysteine significantly lowered the growth rate, growth yield, and denitrification activity of *I. calvum C5*. Therefore, we sought to understand and contrast the mechanistic underpinnings of cysteine- and sulfide-induced inhibition of denitrification and growth of *I. calvum C5*, which has the genome-encoded metabolic capabilities to perform both denitrification and DNRA (Figure 2.1) [28]. Also, as an isolate recovered from the FRC, we anticipate its activities are representative of processes occurring at this contaminated field site. We profiled changes in the transcriptome and metabolome at a global level to characterize the physiological response of *I. calvum* to cysteine and sulfide [41]. Cysteine and sulfide had significant and distinct impacts on the physiology and growth characteristics of *I. calvum* that were correlated with dysregulation (i.e., abnormal changes in the activity) of multiple pathways, including branched-chain amino acid (BCAA) biosynthesis, carbon utilization, and cofactor metabolism. This work expands the scope of a previously established mechanism for growth inhibition of NRB by cysteine and provides new evidence for the mechanism of sulfide inhibition [42].

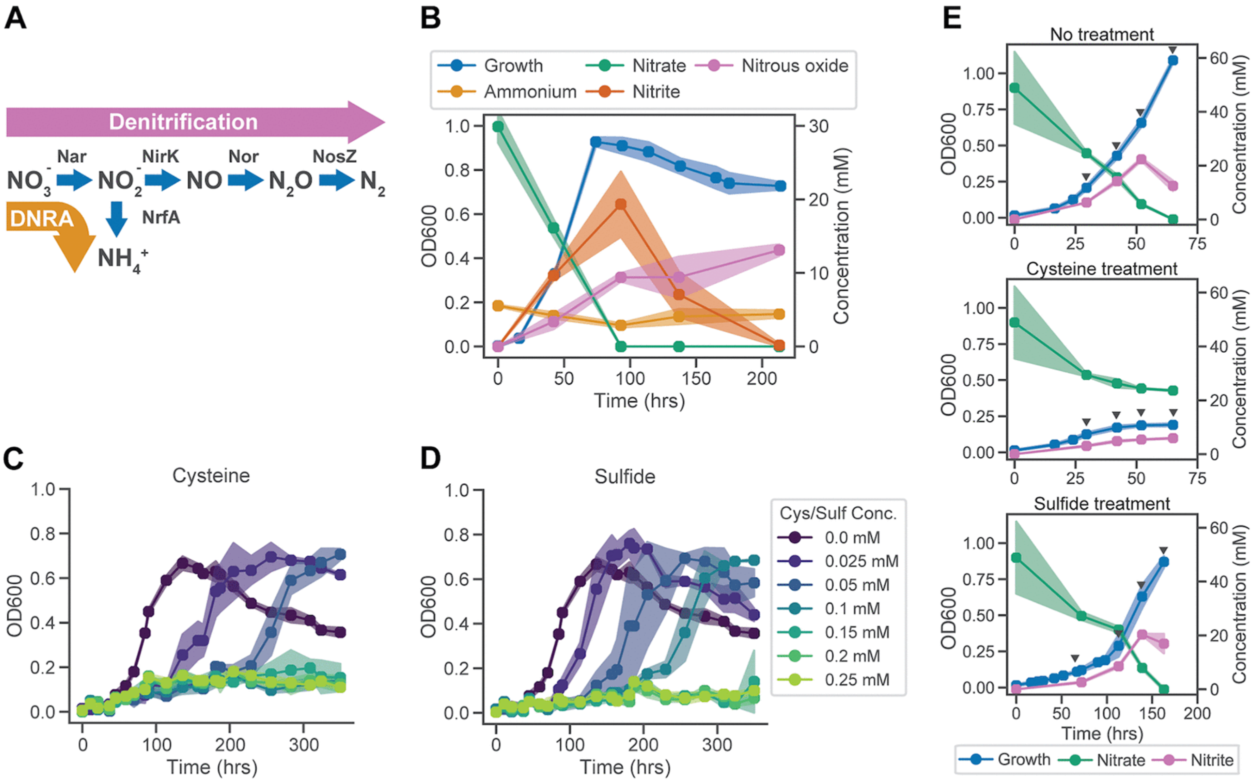

Figure 2.1. Nitrate-reducing phenotype of *I. calvum* and the growth inhibitory effects of cysteine and sulfide. (**A**) The genome sequence of *I. calvum* encodes two NO_3^- reduction pathways: partial denitrification to N_2O and dissimilatory reduction to ammonia (DNRA). (**B**) Growth characteristics and dynamics of NO_3^- reduction by *I. calvum* with 30 mM NO_3^- and 20 mM lactate. Growth characteristics of *I. calvum* in the presence of increasing concentrations of cysteine (**C**) and sulfide (**D**). (**E**) Growth characteristics of cultures sampled for transcriptomic and metabolomic profiling. Profiling timelines varied across conditions based on growth characteristics. Cysteine treatment cultures contained 0.25 mM cysteine. Sulfide treatment cultures contained 0.25 mM sulfide. Triangles indicate points at which samples were collected. Samples for transcriptomics were collected as biological triplicates at all four time points, whereas samples for metabolomics were collected as five biological replicates at the first two time points for each condition. Shaded regions in all plots represent standard deviation across biological replicates ($n \geq 3$).

2.2 RESULTS

2.2.1 *Phenotypic characterization of I. calvum during nitrate reduction and the impact of cysteine and sulfide on growth.*

To characterize the growth effects of reduced sulfur compounds, we first investigated the ability of *I. calvum* to grow anaerobically in the absence of cysteine, which is typically used as a reducing agent in anoxic culture medium [28, 43]. Consistent with its genome-encoded capabilities (Figure 2.1A), in the absence of cysteine, *I. calvum* utilized incomplete denitrification to grow on NO_3^-. Within the first 50 h, ~50% of the total NO_3^- in the growth medium was reduced to NO_2^-, which accumulated transiently, before complete reduction to N_2O after ~100 h (Figure 2.1B). We also found that varying the carbon-to-nitrogen ratio in the medium led to increased DNRA activity when NO_3^- was limited. Subsequently, we subjected *I. cavum* to growth assays in minimal medium with increasing concentrations of cysteine or sulfide. Both compounds inhibited the growth of *I. calvum* in a dose-dependent manner, with complete growth inhibition at >0.05 mM cysteine and >0.1 mM sulfide (Figure 2.1C and D). Notably, the inhibitory effect of cysteine and sulfide was observed only under anoxic conditions, when the primary mode of growth was by NO_3^- reduction. Furthermore, we confirmed that recovery from inhibition at lower concentrations of cysteine and sulfide was not due to oxygen leakage, suggesting potential physiological adaptation, including plausible activation of a tolerance mechanism(s).

2.2.2 *Transcriptome-wide changes induced by cysteine and sulfide treatment.*

To elucidate mechanisms of growth inhibition and recovery, we profiled temporal changes in both the transcriptome and metabolome of *I. calvum* cultured with and without cysteine or sulfide (Figure 2.1E). Cysteine revealed a substantial inhibitory effect, with ~20% of the growth and ~50% of the NO_3^- reduced compared to that of the control, while cultures treated with sulfide showed a similar growth and NO_3^- reduction/NO_2^- production extent but at a lower rate, requiring ~160 h compared to ~70 h in the control culture. Differentially expressed genes (DEGs) across treatments and regular growth conditions were identified using DESeq2-normalized transcriptome sequencing (RNA-seq) data and Boruta [44, 45]. Altogether, a total of 362 of 3,687 genes in the *I. calvum* genome were differentially regulated in response to cysteine or sulfide treatment, of which, 162 DEGs were associated with the response to cysteine, with 82 genes downregulated and 80 genes upregulated. A similar number of DEGs (139 in total) were associated with response to sulfide treatment; 90 of these genes were significantly downregulated and 49 were significantly upregulated. Finally, a total of 146 DEGs were different between cysteine and sulfide treatments; of these, 90 were downregulated by cysteine relative to sulfide treatment, and 56 were upregulated. We performed k-means clustering and pathway ontology enrichment analysis to elucidate global patterns in pathway regulation associated with growth inhibition and recovery in response to cysteine and sulfide treatment (Figure 2.2). We present the results of this analysis in the context of treatment type, highlighting similarities and differences across responses to cysteine and sulfide.

2.2.2.1 Transcriptional response to cysteine treatment.

We discovered that cysteine treatment resulted in the differential regulation of 162 genes within 94 of 294 pathways annotated by MetaCyc, 55 of 108 pathways annotated by SEED, and 117 of 548 GO biological process terms, including those associated with amino acid, fatty acid, and cofactor biosynthesis, transport processes, sulfur, and central carbon metabolism, and terminal cytochrome oxidase activity (Figure 2.2). Clustering and pathway enrichment analysis revealed that 28 MetaCyc pathways were significantly enriched in DEG clusters. While the differential regulation of 23 of 34 genes of sulfur metabolism was expected, there were distinct patterns of change that differed across related pathways and treatments. For instance, 19 of 25 total genes in the cysteine and methionine biosynthesis pathways and 1 of 2 predicted cysteine transporters were downregulated in response to cysteine but not sulfide treatment. Similarly, 3 of 5 genes involved in the biosynthesis of the glutathione analog mycothiol were also downregulated. In contrast, while the upregulation of sulfur oxidation (3 of 8 genes) in response to cysteine was consistent with concomitant downregulation of sulfate uptake (4 of 6 predicted transporters or uptake facilitators), there was a variable response of genes associated with disulfide bond reduction, with 1 upregulated and 1 downregulated among the 5 predicted to be involved in this process.

The dysregulation of cobalamin (vitamin B12) biosynthesis (17 of 20 genes were downregulated and 1 was upregulated) was consistent with the role of this cofactor in DNA metabolism and fatty acid biosynthesis (Figure 2.2C), which were also downregulated (13 of 27 genes of fatty acid biosynthesis; 15 of 55 DNA metabolism genes). The upregulation of 7 of 18 genes associated with iron and copper uptake potentially reflects the importance of these metals as enzyme cofactors across metabolic pathways that were differentially regulated and the need to replenish copper(II) and iron(III) reduced or sequestered by cysteine[46, 47]. Finally, cysteine treatment resulted in a shift in central carbon metabolism toward utilization of the glyoxylate cycle, based on the downregulation of 5 of 7 genes associated with the conversion of isocitrate to malate in the tricarboxylic acid (TCA) cycle and upregulation of 4 of 6 genes associated with the production of glyoxylate and its subsequent conversions (Figure 2.2C).

2.2.2.2 Transcriptional response to sulfide treatment.

Sulfide treatment resulted in the differential regulation of 139 genes across metal ion transport (19 of 38 genes), RNA metabolism (19 of 63 genes), glycerol uptake and utilization (4 of 6 genes), vitamin B12 biosynthesis (11 of 20 genes), central carbon metabolism (28 of 64 genes with DEGs distributed among the TCA cycle, glycolysis, and pyruvate metabolism), sulfite oxidation (YedY-

YedZ), and terminal cytochrome oxidase activity (6 of 11 genes) (Figure 2.2C). Some of these responses triggered by sulfide treatment, such as downregulation of the TCA cycle (15 of 24 genes), were similar to the response to cysteine treatment. However, several responses differed between the two treatments, including the absence of differential regulation of cysteine, methionine, and mycothiol biosynthesis genes in response to sulfide treatment. The upregulation of thiol-oxidizing cytochrome oxidases by cysteine and sulfide implicates these genes in relieving oxidative stress resulting from the two treatments. However, the specific cytochrome oxidases upregulated differed across the two treatments, with Ical_3589, Ical_3588, and Ical_3156 (annotated as thioredoxin/protein-disulfide isomerases and a cytochrome c biogenesis protein CcdA) upregulated by cysteine and Ical_0754, Ical_0755, and Ical_0756 (all cytochrome c biogenesis proteins CcdA and CcdB) upregulated by sulfide. Similarly, the expression of many transporters that were downregulated by cysteine were less affected by sulfide (e.g., lactate and sulfate). The up- and downregulation patterns of specific metal ion transporters also varied across the two treatments (e.g., cobalt, nickel, zinc, and iron). Ferrous iron transport genes, for instance, were downregulated in the presence of sulfide but not by cysteine treatment, while cobalt and nickel transporters were unaffected by sulfide but were upregulated by cysteine treatment. This was unexpected given the ability of sulfide to sequester these metals and potentially induce stress via their limitation. Finally, the upregulation of glycerol-3-phosphate transporters by sulfide treatment, even though glycerol and glycerol-3-phosphate were not added to the growth medium, suggested that high levels of sulfide might be coupled to the availability of these compounds in some niches occupied by *I. calvum*.

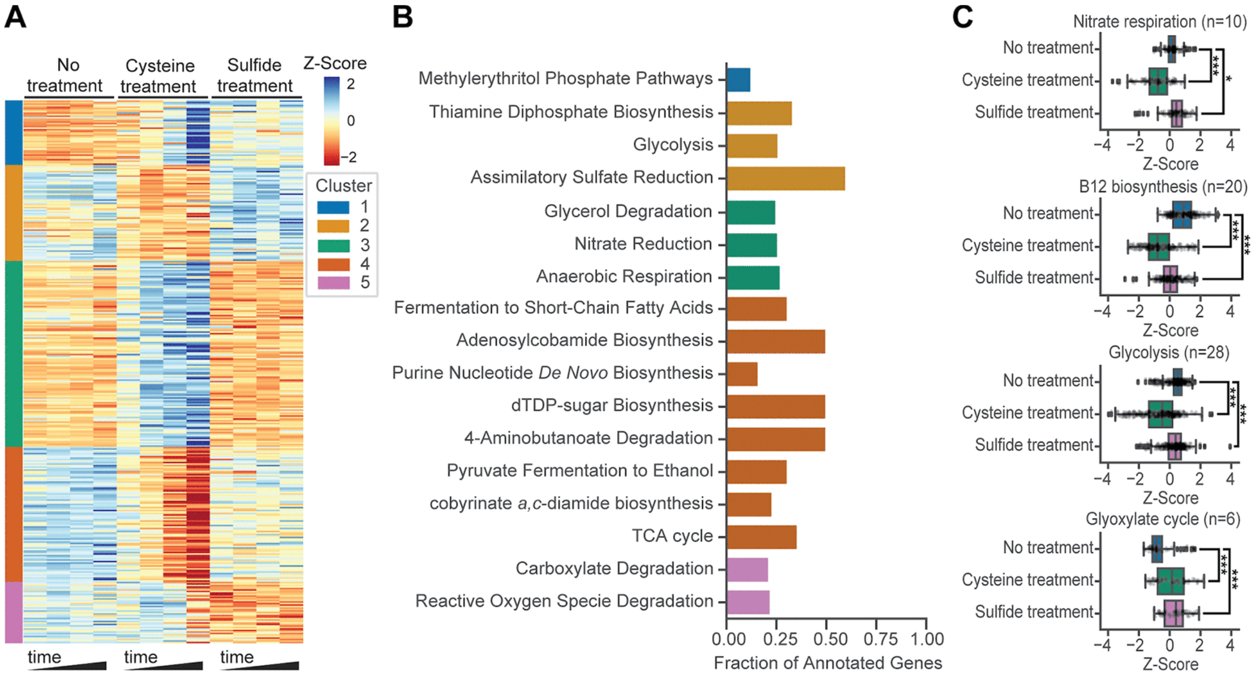

Figure 2.2. *I. calvum*'s transcriptional responses to cysteine and sulfide treatment. (**A**) Differentially expressed genes (DEGs) clustered into five groups using k-means. Expression levels displayed were normalized using the Z-score. MetaCyc pathway terms enriched in each cluster (**B**) and average Z-score normalized expression of select pathways and processes (**C**). Bars indicate comparisons for which differences were significant. *, $p < 0.05$; ***, $p < 0.001$. Boxplots display estimates of data minimum (left whisker), median (line within box), maximum (right whisker), and interquartile range (box dimensions) as well as possible outliers (points beyond whiskers). Data are overlaid as scatter points.

2.2.3 *Metabolome changes during cysteine and sulfide treatment.*

We investigated the physiological consequences of sulfide and cysteine treatment by performing untargeted metabolomics using liquid chromatography-electrospray ionization-quadrupole time of flight mass spectrometry (LC-ESI-QTOF-MS) for samples collected at early phases of growth (Figure 2.1E). Clustering of metabolomes into distinct groups by principal-component analysis (PCA) demonstrated high reproducibility across replicate measurements and suggested that treatment conditions had distinct effects on feature abundances. The mass spectra were analyzed using XCMS Online to identify mass spectral features that were significantly and reproducibly different in relative abundance of total ion intensities across replicates in each of the two treatments [48, 49]. Altogether, the untargeted metabolomics identified a total of 51 mass spectral features which mapped to 60 putative metabolites. Of these, 32 were differentially abundant in cysteine-treated cultures, and 39 were differentially abundant in sulfide-treated cultures. We mapped the putative metabolites to MetaCyc metabolic pathways encoded in the *I. calvum* genome using XCMS Online and identified 12 pathways, including isoleucine biosynthesis, glycerol degradation, and methionine biosynthesis, that were dysregulated in both treatments. Five pathways, including the glyoxylate cycle and TCA cycle, were dysregulated only in cysteine treatment, whereas 16 pathways, including valine biosynthesis, leucine biosynthesis, and alanine biosynthesis, were dysregulated in just sulfide treatment (Figure 2.3A and B).

To confirm the finding that amino acid metabolism was dysregulated across both treatments, we quantified the absolute abundance of 15 amino acids using targeted metabolomics and amino acid standards (Figure 2.3C). This analysis revealed that except for arginine and threonine, which changed in abundance only in sulfide treatment, the abundance of 9 amino acids, including 2 of 3 BCAAs (valine and isoleucine), was significantly reduced by both treatments but generally more so by sulfide treatment (except isoleucine). Dysregulation of the TCA cycle, glyoxylate cycle, mixed acid fermentation, and anaerobic respiration suggested that cysteine and sulfide treatment also resulted in a potential shift in the carbon metabolism of I. *calvum*.

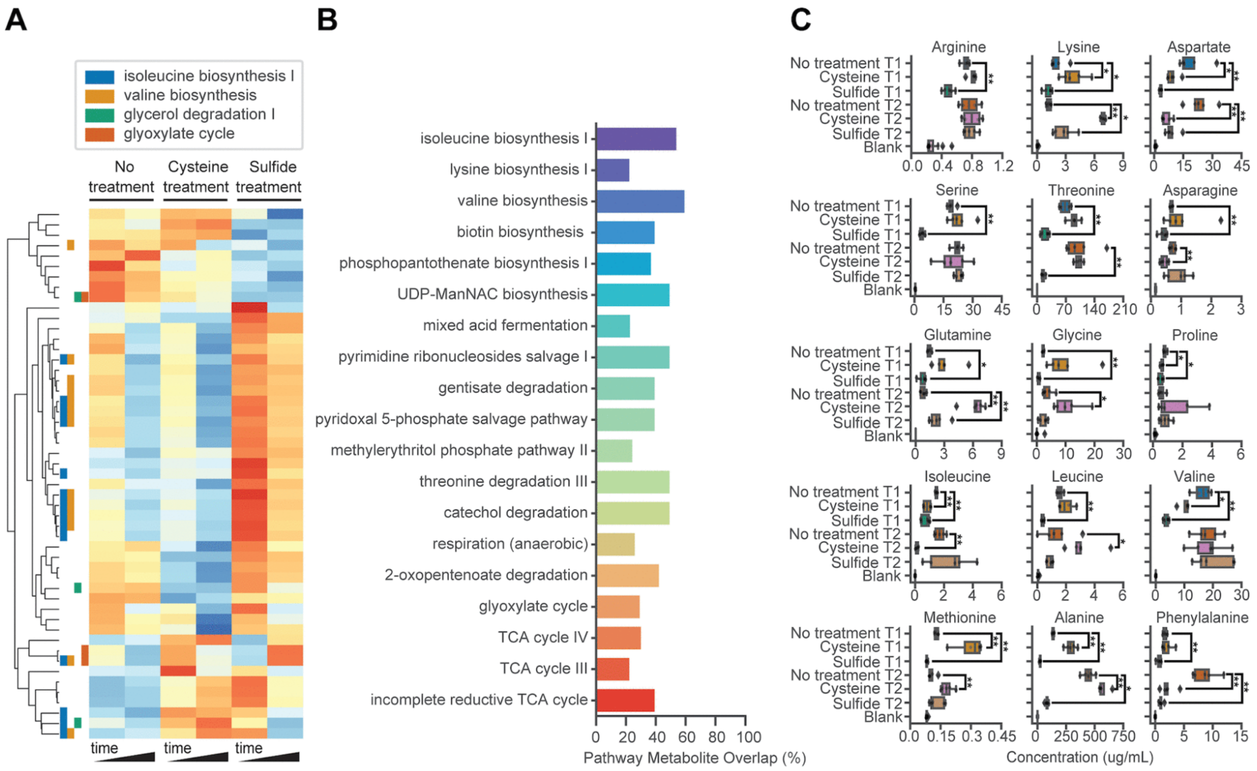

Figure 2.3. Metabolome changes during physiological adaptation to cysteine and sulfide treatment. (**A**) Hierarchically clustered metabolomic feature abundances for those with putative identities scaled using natural logarithm and normalized with Z-score. (**B**) Pathways identified as significantly dysregulated in cysteine or sulfide treatment relative to no treatment based on enrichment of differentially abundant putative metabolites. (**C**) Concentrations of amino acids quantified using targeted metabolomics are displayed. Bars indicate the significance level of select comparisons. *, $p < 0.05$; **, $p < 0.01$. Boxplots display estimates of data minimum (left whisker), median (line within box), maximum (right whisker), and interquartile range (box dimensions) as well as possible outliers (points beyond whiskers). Data are overlaid as scatter points.

2.2.4 *Putative mechanisms for growth inhibition and recovery of I. calvum during cysteine and sulfide treatment.*

The dysregulated pathways identified by metabolomics were consistent with the differential regulation of enzymes observed at the transcriptional level (e.g., mycothione biosynthesis, amino acid biosynthesis, glyoxylate cycle, and TCA cycle). For instance, consistent with transcriptional downregulation of mycothione synthase, the abundance of mycothione was decreased to an almost undetectable level in early phases of growth for cysteine treatment. This low-molecular-weight thiol is known to play a role in mitigating oxidative stress [50]. However, it is unclear what role this compound may have played in the response to cysteine and sulfide treatment given the reduced levels of its oxidized form and downregulation of biosynthesis genes.

All enzymes of both denitrification and DNRA pathways were progressively downregulated by cysteine through all stages of growth (Figure 2.4A). This pattern of downregulation was consistent with reduced NO_3^- and NO_2^- reductase activity as well as complete growth inhibition through the entire course of the experiment (Figure 2.1E). In contrast, sulfide treatment resulted in transient downregulation of the nitric oxide-producing nitrite reductase (Nirk; Ical_2449) and nitric oxide reductase (Nor; Ical_0054), both of which were impacted during the early phase of growth but later recovered. Meanwhile, both nitrate reductase (Nar; Ical_1176, Ical_1188, and Ical_1210) and the ammonium-producing nitrite reductase (NrfA; Ical_0747 and Ical_0748) were upregulated by sulfide treatment, especially in mid-to-late growth phases, suggesting a shift from denitrification to DNRA (Figure 2.4A). These expression patterns were consistent with growth inhibition and recovery patterns as well as changes in NO_3^- and NO_2^- levels during sulfide treatment (Figure 2.1E).

The uptake and metabolism of acetate and lactate were repressed in both treatments, but the specific regulation patterns of genes associated with the transport and metabolism of these compounds varied somewhat between treatments. Lactate transport was downregulated more severely in the presence of cysteine than with sulfide treatment, while acetate transport was less affected by cysteine treatment and more so by sulfide treatment (Figure 2.4B and C). These observations were consistent with perturbed abundance of intermediates of both the carbon and amino acid metabolism pathways (Figure 2.4B and C). Cysteine and sulfide treatments also resulted in the dysregulation of 8 and 3 of 15 genes, respectively, for BCAA biosynthesis (e.g., threonine deaminase; Ical_2218 and Ical_2634) and 2-isopropylmalate synthase (Ical_1041 and Ical_1304) (19). The differential regulation of these pathways was consistent with the altered abundance of intermediates and end products of this pathway (Figure 2.4B). For instance, the downregulation of threonine deaminase coincided with the reduced abundance of mass spectral features that matched end products of this enzyme (2-ketobutyrate and 2-iminobutanoate) as well as threonine (Figure 2.4B). Additionally, 11 of 20 genes associated with L-valine degradation were upregulated initially (first time point) or later in growth during cysteine treatment (e.g., Ical_1011 and Ical_1143), which may have served as a source of carbon (via production of propanoyl coenzyme A [propanoyl-CoA]) in lieu of repressed carbon uptake [51].

The global transcriptomics and metabolomics analyses (Figures 2.2 and 2.3) also suggested that glycerol degradation was dysregulated during cysteine and sulfide treatment. Exploring this possibility, we identified a pathway coupling glycerol and glycerophosphodiester oxidation to NO_3^- reduction (Figure 2.4D). Supporting the relevance of this pathway, transport of glycerol-3-phosphate and oxidation of glycerophosphodiesters to glycerol-3-phosphate were upregulated in the presence of sulfide (e.g., Ical_0677, Ical_0177, and Ical_0523). Subsequent steps linking oxidation of glycerol-3-phosphate to NO_3^- reduction were unaffected by sulfide treatment (i.e., neither upregulated or downregulated). Meanwhile, all pathway steps except the oxidation of glycerol-3-phosphate to glycerone-phosphate were downregulated by cysteine treatment (Figure 2.4D). Thus, glycerol or glycerophosphodiesters may provide an alternative or complementary means of producing reducing equivalents for NO_3^- reduction during sulfide-induced stress. In summary, the integrated analysis of changes at the levels of the transcriptome and the metabolome in the context of growth and phenotype characteristics identified a shift in carbon metabolism and the repression of branched-chain amino acid biosynthesis as potential mechanisms of growth inhibition by cysteine and sulfide.

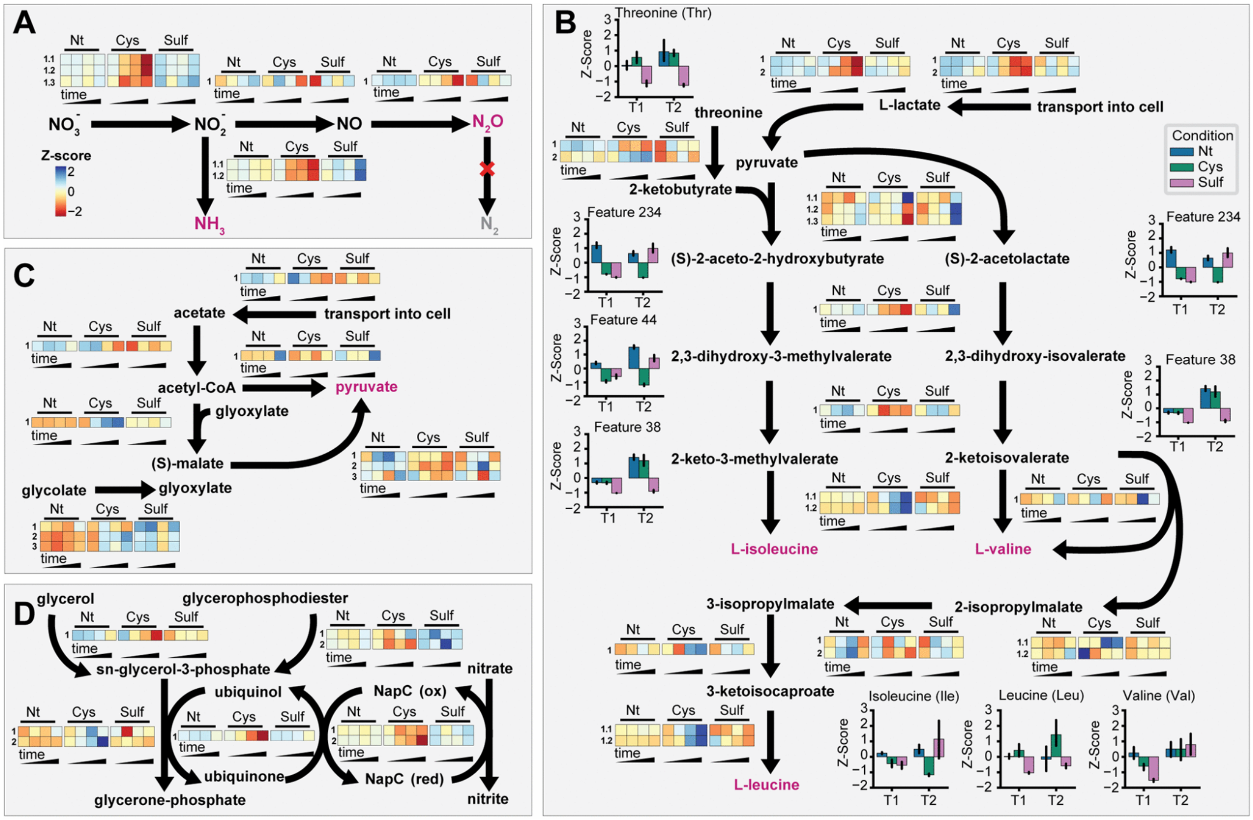

Figure 2.4. Integrated analysis of transcriptional and metabolic changes. Expression profiles for each condition are displayed as a heat map for genes predicted to be involved in NO_3^- reduction (**A**), branched-chain amino acid biosynthesis (**B**), acetate utilization (**C**), and glycerol-3-phosphate coupled NO_3^- reduction (**D**). In each panel, the expression level normalized by Z-score is displayed. In each heat map, rows are genes associated with the pathway step (integer numbers note individual enzymes, while decimals indicate subunits), and columns are condition time points. Nt, no treatment control; Cys, cysteine treatment; Sulf, sulfide treatment. Additionally, for specific metabolites for which the absolute concentration was quantified (e.g., amino acids) or for which a metabolomic feature was associated, the Z-score of the metabolite abundance (or relative abundance) is displayed. The metabolomic feature number is indicated for metabolites with

putative identification. Error bars represent the standard deviations from biological replicates ($n \geq 3$).

2.2.5 *Alternate carbon sources and BCAA supplementation rescue growth inhibition by cysteine and sulfide.*

Based on the findings of the systems analysis, we designed experiments to test hypotheses regarding specific mechanisms underlying the growth inhibition effects of cysteine and sulfide. We hypothesized that providing an alternative carbon source or supplementing with BCAAs would rescue the growth-inhibitory effects based on dysregulation of both pathways at the transcriptional and metabolic levels. To test these hypotheses, we assayed growth characteristics and NO_3^- metabolism of *I. calvum* in media supplemented with various carbon sources (lactate, acetate, and glycerol) or a mixture of BCAAs (Figure 2.5). Supplementation with 1 mM BCAAs (in the presence of lactate) relieved growth inhibition by both reduced sulfur compounds, with complete growth recovery of cultures treated with 0.2 mM cysteine. In the presence of BCAAs, cysteine-treated cultures also readily reduced NO_3^-, NO_2^-, and accumulated N_2O. In the absence of supplementation, cysteine treatment impaired these activities, in agreement with previous observations. BCAAs also relieved growth inhibition by 0.2 mM sulfide, contributing to a significant reduction in recovery time (t-test $p = 4 \times 10^{-4}$) compared to that of cultures with sulfide alone, which took an additional 60 h to recover on average. Sulfide-treated cultures supplemented with BCAAs also reduced NO_3^- at a greater rate within the first 100 h of growth than cultures treated with sulfide alone and accumulated greater quantities of N_2O following recovery (Figure 2.5).

Changing the carbon source from lactate to acetate resulted in slower growth and reduced carrying capacity (i.e., maximum optical density at 600 nm [OD600]). Acetate also mitigated the inhibitory effect of cysteine in the early and mid-growth phases but not in the stationary phase (Figure 2.5B). Acetate worsened the growth-inhibitory effect of sulfide. Interestingly, the inhibitory effects of cysteine and sulfide mirrored the patterns of differential regulation of acetate uptake genes during growth with lactate (Figure 2.4C). In contrast, glycerol by itself did not support growth of *I. calvum* (unpublished observation), but in combination with lactate, it reduced the inhibitory effects of sulfide but not cysteine. Sulfide-treated cultures supplemented with glycerol and lactate had significantly decreased recovery times relative to those of sulfide-treated cultures grown with lactate alone (t-test $p = 0.027$), although the effect was less significant than for supplementation with BCAAs. Meanwhile, glycerol supplementation had no detectable effect on the recovery from cysteine inhibition in cultures grown with lactate (no recovery observed over 200 h), which was consistent with expression patterns associated with glycerol utilization (Figure 2.4D). Altogether, these experiments corroborated our hypotheses for how dysregulation of carbon metabolism and BCAA biosynthesis mechanistically contribute to the inhibitory effects of cysteine and sulfide on the growth of *I. calvum*.

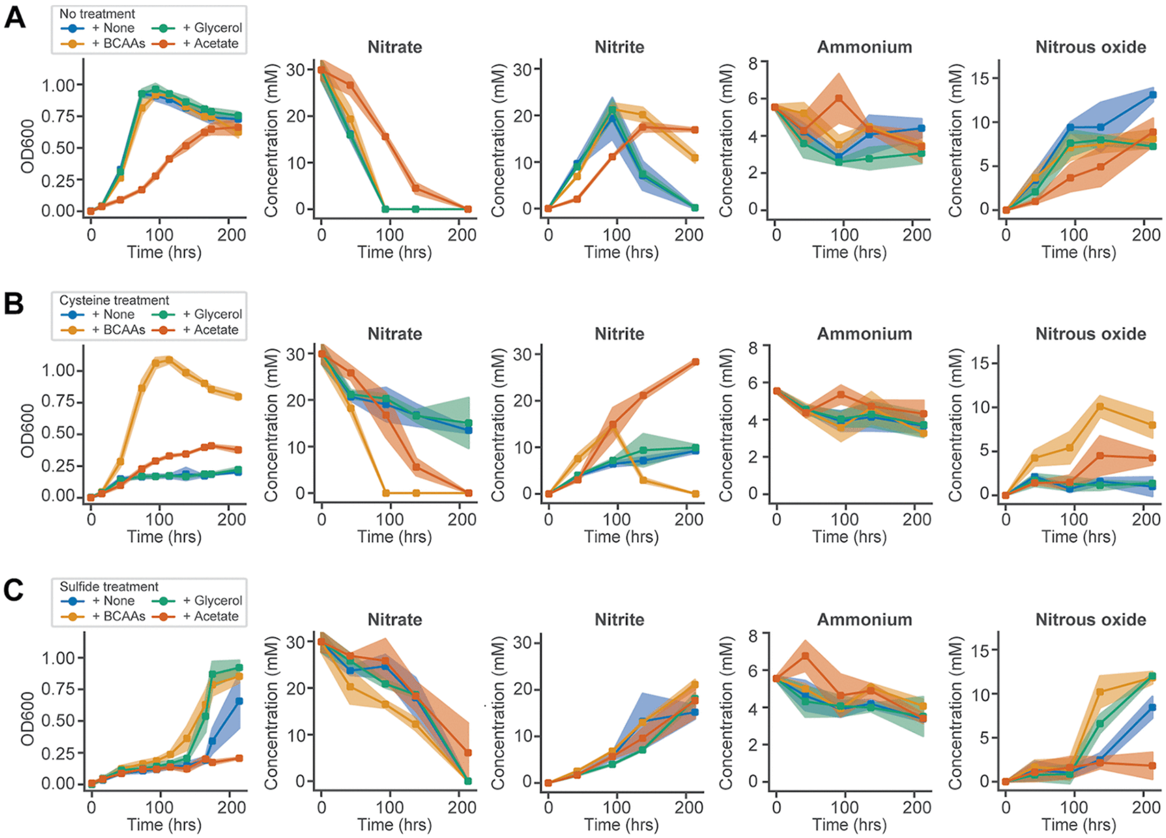

Figure 2.5. Supplementation experiments support the role of pathway dysregulation predicted by transcriptomics and metabolomics. Supplementation conditions consisted of a branched-chain amino acid mixture (Iso, Leu, and Val, each at 1 mM [+BCAAs]) plus 20 mM lactate, 30 mM glycerol (+Glycerol) plus 20 mM lactate, 20 mM lactate alone (+None), or 20 mM acetate alone (+Acetate). The effects of supplementation on growth and denitrification/DNRA for untreated (**A**), 0.2 mM cysteine treated (**B**), and 0.2 mM sulfide treated (**C**). Shaded regions depict the standard deviations from biological replicates ($n \geq 3$).

2.3 DISCUSSION

In this study, we characterized dose-dependent growth inhibition by cysteine and sulfide on a groundwater-associated NO_3^- reducer, *I. calvum*. Our systems analysis revealed that a shift in carbon metabolism and the repression of BCAA biosynthesis were two potential mechanisms of growth inhibition by reduced sulfur compounds, which we subsequently tested in a series of supplementation experiments. Growth inhibition by cysteine and sulfide in the presence of lactate was consistent with the downregulation patterns of lactate uptake and metabolism, which were severe in the presence of cysteine but were only observed in early stages of sulfide treatment (Figure 2.4B and 2.5). Acetate, on the other hand, relieved growth inhibition by cysteine, which upregulated acetate uptake and metabolism during early stages of growth. However, acetate failed to relieve growth inhibition during sulfide treatment, likely due to the downregulation of both its uptake and metabolism (Figures 2.4C and 2.5). In contrast, glycerol reduced the severity of growth inhibition by sulfide but not cysteine treatment, consistent with the expression patterns of the associated pathways and a modest increase in the rate of NO_3^- reduction for sulfide-treated cultures supplemented with glycerol (Figures 2.4D and 2.5). Collectively, these data demonstrate that the availability of an appropriate carbon source can play an important role in mediating the growth-inhibitory effects of reduced sulfur compounds and thus potentially also mediate the interplay between NRB and SRB.

Cysteine is known to have cytotoxic effects in various contexts, and studies in *Escherichia coli* established a mechanistic link between cysteine-induced growth inhibition and repression of BCAA biosynthesis [42, 52–54]. While our findings on the mechanism by which cysteine inhibits growth of *I. calvum* are consistent with previous literature, we have discovered that growth inhibition by sulfide also acts through dysregulation of BCAA biosynthesis. However, our results demonstrated that cysteine and sulfide may act on different enzymatic steps in BCAA biosynthesis (Figure 2.4B). Furthermore, our data showed that in addition to BCAA biosynthesis, carbon uptake played an important role in growth inhibition and that cysteine and sulfide treatment had broad and distinct impacts on the regulation of other pathway activities (e.g., TCA cycle and vitamin B12 and mycothiol biosynthesis). In addition, whereas growth inhibition of *I. calvum* by cysteine and sulfide treatments was observed only under anoxia, in *E. coli*, this phenomenon was observed under oxic conditions [53, 54]. Thus, for *I. calvum*, repression of genes involved in NO_3^- respiration likely also underlies the observed phenomenon. Indeed, all genes in the denitrification pathway were severely repressed during cysteine treatment, and pathway activity was significantly reduced (Figure 2.1E and 2.4A). Sulfide treatment also resulted in downregulation of the denitrification pathway prior to recovery, but interestingly, the DNRA pathway was upregulated (Figure 2.4A). While significant accumulation of NH_4^+ was not observed during sulfide treatment, the sulfide-induced shift from denitrification to DNRA is consistent with previous work on the effect of sulfide on the activities of nitrate-reducing communities in freshwater, wastewater, and sediment [39, 55–57].

I. calvum recovered from the inhibitory effects of cysteine and sulfide and achieved a normal growth rate within 300 h of treatment, with low-to-moderate doses of cysteine (< 0.1 mM) and sulfide (< 0.15 mM). The recovery from cysteine inhibition could have been mediated by the upregulation of a putative cystathionine gamma-lyase (EC 4.4.1.1, Ical_3515), which metabolizes cysteine into pyruvate, hydrogen sulfide, and NH_4^+. Overexpression of this enzyme, also known as cysteine desulfhydrase, confers resistance to cysteine in *E. coli* [58]. We also discovered that genes associated with pyridoxidal-5′-phosphate biosynthesis (a cofactor of cystathionine gamma-lyase), iron transport, cytochrome c assembly, and sulfide:quinone oxidoreductase (EC 1.8.5.4, Ical_1214) were upregulated. Taken together, these observations suggest a recovery mechanism in which cysteine is first reduced to sulfide by cystathionine gamma-lyase and then oxidized to polysulfide by sulfide:quinone oxidoreductase [59]. This putative cystathionine-mediated recovery mechanism might be utilized due to lack of availability of the mycothione-based primary redox stress response as a result of the downregulation of mycothiol synthase and other mycothiol-dependent enzymes during treatment with cysteine [60]. While cysteine is a reactant in mycothiol biosynthesis, the decrease in mycothione levels was likely due to the unavailability of precursors generated by BCAA biosynthesis, which was downregulated by cysteine treatment. It is also plausible that reducing conditions and increased thiol levels resulting from excess cysteine could have triggered sulfur relay pathways to decrease the synthesis of mycothione. Dysregulation of the sulfur pool can also alter protein translation via tRNA thiolation pathways and affect growth [61]. Like cystathione and mycothione, the thiols of thioredoxins, which were upregulated by both cysteine and sulfide, could have also played a role in recovery from reduced sulfur stress by scavenging excess cysteine through the formation of intermolecular disulfide bonds [62].

Interestingly, sulfide:quinone oxidoreductase and the other genes implicated in recovery from cysteine were not upregulated during sulfide treatment, suggesting the mechanism of recovery from sulfide inhibition is different. Terminal cytochrome c oxidases that were upregulated by sulfide treatment (e.g., Ical_0754, Ical_0755, Ical_0756, Ical_0747, and Ical_0748) could have facilitated its oxidation via the reduction of heme [63]. Among these cytochrome c oxidases are thioredoxin/protein-disulfide isomerases, cytochrome c biogenesis protein CcdA, and those in the DNRA pathway, which suggests that sulfide oxidation may have been coupled to dissimilatory NO_3^- reduction to ammonia, as observed previously in freshwater nitrate-reducing communities exposed to sulfide [39]. In stark contrast to the response to cysteine, mycothione biosynthesis and thioredoxin system genes were unaffected by sulfide treatment, suggesting their possible involvement in sulfide scavenging, clearing, or removal of free sulfide-induced disulfide bonds or nitroso groups [64]. Thus, the systems analysis demonstrates that the recovery of growth of *I. calvum* from cysteine and sulfide inhibition likely occurs through distinct mechanisms.

Our results broaden the scope of knowledge regarding the inhibitory effects of reduced sulfur compounds and provide additional insights into the competition between SRB and NRB in the environment. It is a well-accepted principle that NRB can outcompete SRB for carbon based on the thermodynamics of denitrification and sulfate reduction, which is the basis for NO_3^- injections into oil wells to prevent souring from sulfide production [37, 65, 66]. However, several studies have found exceptions to simple thermodynamic principles, which suggests that to better understand microbial community assembly and interactions in the environment, more research is needed [67–70]. In line with this notion, the interactions between NRB, SRB, and their metabolites are complex and not fully characterized. For instance, certain SRB have been shown to perform DNRA (a higher energy-yielding metabolism than sulfate reduction), but this activity is inhibited by the presence of sulfide [71]. Furthermore, previous work has shown that SRB that lack nitrite reductase are inhibited by NO_3^- and NO_2^- [36]. Other studies have shown that certain NRB are inhibited by sulfide [38]. Our findings, that growth and activity of the NRB *I. calvum* is inhibited by elevated levels of sulfide and cysteine, build on this prior work and provide a mechanistic explanation for the mutual exclusion of SRB and NRB in certain environments. Specifically, thermodynamic considerations and the inhibitory effects of NO_2^- likely exclude SRB from niches occupied by NRB and, conversely, elevated levels of sulfide may prevent growth of NRB in niches where SRB are active. These mechanistic hypotheses are supported by observations of the effects of sulfide on denitrification activity in freshwater and wastewater communities [39, 55] and are also consistent with an observed stratification of SRB and NRB activities down the vertical transect of sediment cores sampled and analyzed from the FRC. Specifically, transects of the cores that had supposed sulfate reduction activity (based on sulfide detection) and denitrification activity (based on acetylene block activity assays) were spatially separated [72]. However, there are notable exceptions, such as certain bacteria that couple NO_3^- reduction to sulfide oxidation and thrive under conditions where both NO_3^- and sulfide are present [73]. Additionally, the observation that *I. calvum* was capable of recovering from reduced sulfur stress suggests that NRB have developed mechanisms to overcome the inhibitory effects of moderate cysteine and sulfide levels and coexist with SRB in some environments. Functional genomic screens and laboratory evolution studies using mixed communities of NRB and SRB could help to better elucidate the biotic and abiotic constraints on competition between NRB and SRB [74].

2.4 METHODS

2.4.1 *Strains and medium preparation.*

I. calvum C5 was isolated from groundwater obtained from a well (GW 247) at the Oak Ridge Field Research Center (FRC) that is highly contaminated with NO_3^- (>200 mM). Growth studies were performed at 30°C in defined minimal medium containing 20 mM sodium lactate and 30 mM sodium nitrate at pH 7.2 with an 80:20 N2-CO2 headspace. For supplementation experiments, cysteine (L-cysteine; Sigma-Aldrich), hydrogen sulfide (sodium sulfide hydrated technical grade chips [Fisher Scientific] or sodium sulfide nonahydrate [Santa Cruz Biotechnology]), isoleucine (L-isoleucine; Sigma-Aldrich), leucine (L-leucine, Sigma-Aldrich), valine (L-valine; Sigma-Aldrich), sodium acetate (Sigma-Aldrich), and glycerol (molecular biology grade; Sigma-Aldrich) were added from sterile anoxic stocks. All growth assays were performed using cultures revived from freezer stocks, which were allowed at least one full growth cycle.

2.4.2 *Growth measurements.*

For all experiments, cell concentration was monitored in Balch tubes (10-ml culture volume) with periodic measurements of the optical density at 600 nm (OD600) using a Spectronic 200 spectrophotometer (Thermo Fisher). Blank subtraction was performed using Balch tubes containing medium with no cells. Initial cell densities were normalized to ~0.01 OD600 units for all experiments.

2.4.3 *Measurement of nitrogen species.*

NO_2^-, total oxidized nitrogen (TON; NO_2^- plus NO_3^-), and total ammonia (NH_4 and NH_4^+) content in media were quantified using the Gallery automated photometric analyzer (Thermo Fisher). Measurements were calibrated using sodium nitrite, sodium nitrate, and ammonium chloride. N_2O was measured by gas chromatography (model 8610; SRI Instruments) with nitrogen as the carrier gas, a 182.9-cm HayeSep D column (SRI Instruments), and an electron capture detector (ECD).

2.4.4 *Measurement of oxygen concentration.*

The concentration of dissolved oxygen was monitored in Balch tubes using a FireStingO2-Mini fiber optic oxygen meter with a temperature-compensated 3-mm-diameter optode and oxygen sensor spots (Pyroscience). Sensor spots were attached to the interior of Balch tubes below the 10-ml liquid level using silicone glue and calibrated using sodium sulfite solution and oxygen-saturated growth medium. Both temperature and the ionic strength of the medium were taken into account for calibration and subsequent measurements.

2.4.5 *Transcriptomics profiling and analysis.*

2.4.5.1 Sample collection and sequencing.

I. calvum C5 was cultured in biological triplicates, in a medium containing either 0.25 mM cysteine or 0.25 mM sulfide (Fisher). Aliquots (10 ml) were collected at four time points for transcriptome profiling (Figure 2.2E), and cell pellets were harvested under anoxic conditions by centrifugation at 4,000 × g. Cell pellets were flash frozen in liquid nitrogen, and total RNA was extracted using hot phenol-chloroform [75]. The Ribo-Zero bacterial kit (Illumina) was used for rRNA depletion, and the TruSeq Stranded mRNA library preparation kit (Illumina) was used for library preparation.

Sequencing was performed using the NextSeq platform (2 by 75 bp, Illumina) with 10 to 15 million reads per sample.

2.4.5.2 Read processing.

RNA sequencing reads were analyzed with FastQC according to Illumina's default quality filtering process and then trimmed using base quality scores by Trimmomatic[76, 77]. A quality score of 20 was used for read trimming and quality filtering. Reads were aligned to the genome of *Intrasporangium calvum C5* (NCBI BioProject PRJNA475609) (3, 52) using Spliced Transcripts Alignment to a Reference (STAR) [78] followed by tabulation of transcript abundances by HTSeq-count [79].

2.4.5.3 Differential expression analysis, clustering, and functional enrichment.

Normalized expression data obtained from DESeq2 [44] were filtered to remove genes with consistently low expression levels (among the lower 1% of normalized expression values across all conditions). Boruta, a random forest-based feature selection algorithm, was used to perform pairwise comparisons of transcriptional profiles across treatments to identify differentially expressed genes (DEGs) using Z-score-transformed normalized expression [45]. Gene ontology (GO) and MetaCyc annotations were obtained by comparing features in the genomes of *I. calvum C5* and the type strain *I. calvum 7 KIP* (for which these annotations were available) using reciprocal BLAST [80, 81]. Reciprocal BLAST default parameters were used. Sequence matches were required to be at least 70% identical within the aligned region, and high-scoring segment pair alignment had to cover at least 50% of the query sequence. SEED annotations were obtained using DIAMOND to map transcript reads against the SEED database of annotated microbial protein sequences [82, 83]. Default DIAMOND search parameters and the BLASTX option were used to query transcript sequences against the SEED protein database. DEGs identified by Boruta were clustered using the scikit-learn implementation of the k-means algorithm [84]. Additional statistical analyses, including principal component analysis (PCA), significance testing, and functional enrichment were performed using tools from the python scikit-learn, SciPy, and NumPy packages [84–86]. The significance of GO, MetaCyc, and SEED term enrichment among k-means clusters was assessed by comparing term frequencies within each cluster to their frequencies in the genome using a hypergeometric test. Significantly enriched terms ($p \leq 0.05$ and Benjamini-Hochberg false-discovery rate $q < 0.01$) were reported [87]. Gene level expression comparisons were performed for the pathways identified using SEED, GO, and MetaCyc process and pathway annotations [83, 88, 89]. Expression level differences were assessed by Welch's t test, and genes were reported as dysregulated if differences were significant ($p \leq 0.05$ and Benjamini-Hochberg false discovery rate $q < 0.01$).

2.4.6 *Metabolomics profiling and analysis.*

2.4.6.1 Extraction.

Metabolomic analysis was performed for five biological replicates over two time points during growth of *I. calvum C5* in medium containing 0.25 mM cysteine or 0.25 mM sulfide (Fisher) (see Text S1 for treatment of batch effects). Cell culture aliquots (10 ml) were harvested under anoxic conditions by centrifugation at 4,000 × g, flash frozen in liquid nitrogen and stored at −80°C. Frozen cell pellets were thawed on ice and resuspended in a 2:2:1 solution of acetonitrile, methanol, and water. Metabolites were extracted according to a standard approach, dried, and reconstituted in volumes of acetonitrile-water solution (1:1 [vol/vol]) normalized to protein content in the sample. Following extraction, samples were transferred to liquid chromatography-mass spectrometry autosampler vials and were stored at −80°C until analysis.

2.4.6.2 High-resolution mass spectrometry untargeted analysis.

Metabolite mixtures were analyzed with liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry (LC-ESI-QTOF-MS) (Bruker impact II) in both positive and negative electrospray ionization modes. Metabolites were separated by gradient elution with the mobile phase consisting of various ratios of water containing 0.1% formic acid and acetonitrile containing 0.1% formic acid. For the amide hydrophilic interaction chromatography analysis, metabolites were separated by gradient elution with the mobile phase composed of various ratios of water-acetonitrile (95:5 [vol/vol]) containing 20 mM ammonium acetate, 40 mM ammonium hydroxide, and water-acetonitrile (5:95 [vol/vol]).

2.4.6.3 Targeted metabolomics analysis of amino acid concentration.

Samples were analyzed on an Agilent 6495 triple quadrupole mass spectrometer coupled to an Agilent 1290 ultraperformance liquid chromatography stack. Separation was carried out using an Imtakt amino acid column. Targeted masses and retention times were selected based on previously established reference standards for each amino acid.

2.4.6.4 Data processing, metabolite annotation, and statistical analysis.

Raw data were processed and analyzed using XCMS Online [48, 49]. Metabolite features shared across all samples were identified using a multigroup analysis and compared for differential abundance. Differentially abundant metabolite features were identified with a total ion intensity of >10,000 and an associated abundance fold change of >3 in sulfide or cysteine-treated samples relative to the control sample ($p \leq 0.01$). The filtered feature data table was annotated via an accurate mass search against METLIN using a 25-ppm window and manual inspection of putative identification hits for each feature [90–92]. Dysregulated pathways were annotated by considering the enrichment of differentially abundant metabolites predicted by METLIN using the Systems Biology results feature of XCMS Online. Pathways with a significant enrichment of differentially abundant metabolite features ($p \leq 0.01$) and coverage of at least 20% were reported. Additional statistical analyses, including PCA, clustering, and significance testing, were performed using tools from the python scikit-learn, SciPy, and NumPy packages with Z-score-standardized log peak areas [84–86]. Pathways and putative metabolites were compared to the results of the corresponding samples in the transcriptomic analysis.

Chapter 3. NITRITE TOXICITY DRIVES NITROUS OXIDE ACCUMULATION IN A COOPERATIVE DENITRIFYING SYNTHETIC COMMUNITY

3.1 BACKGROUND

Since the development of the Haber-Bosch process, there has been an exponential increase in the global deposition of fixed nitrogen. This has largely resulted from the use of synthetic fertilizers [93]. Meeting the demands of a growing global population has contributed to a nearly tenfold increase in synthetic fertilizer use from 1960 to 2013 [94]. Recent estimates suggest anthropogenic contributions account for about half of the reactive nitrogen flux on Earth [94]. As a result, the global nitrogen cycle has become increasingly perturbed. The consequences of this perturbation are myriad and include eutrophication of terrestrial and aquatic systems, global acidification, and stratospheric ozone loss among other issues [93]. Emission of nitrous oxide (N_2O) in particular is a growing concern in agricultural systems and wastewater treatment [95–97]. The microbial processes of nitrification and denitrification are responsible for the bulk of agricultural N_2O emissions [96, 98]. Thus, understanding how microbial processes, interactions, and environmental factors such as resource concentration, pH, and metal availability influence the fate of nitrogen compounds, and the production of N_2O in particular, is essential for developing a predictive understanding of the fate of different nitrogen species in natural and engineered systems [28–30, 99].

Denitrification is the major biological process that returns fixed nitrogen to the atmosphere through reduction of nitrogen-oxides nitrate (NO_3^-) and nitrite (NO_2^-) to gaseous compounds nitric oxide (NO), N_2O, and dinitrogen (N_2). Therefore, it comprises a critical step in the global nitrogen cycle. Denitrification is also an anaerobic process that provides an alternative means to generate energy when oxygen (O_2) is not available. In complete denitrification, microorganisms utilize NO_3^- as an electron acceptor instead of O_2 and reduce NO_3^- to N_2 in a series of four reductive, energy-deriving steps. First, NO_3^- is reduced to NO_2^- by one of two dissimilatory nitrate reductases—the membrane-bound Nar complex or the periplasmic nitrate reductase, Nap. NO_2^- is then reduced to NO by one of two structurally unrelated nitrite reductases that contain different prosthetic groups—the copper-containing NirK or the cytochrome cd1 NirS [100]. NO is then reduced to N_2O by the nitric oxide reductase (NorB), and finally, N_2O is reduced to N_2 by the nitrous oxide reductase (NosZ), of which two distinct clades have been identified [100, 101].

Most studies have only examined complete denitrifiers [102, 103] in monoculture to study this respiratory process (enzymology, kinetics, regulation, etc.). However, ongoing genomic and metagenomic characterization of isolates and natural communities has revealed, remarkably, that the genomes of most organisms that encode denitrification enzymes lack the complete pathway [104]. More commonly, the microbial communities in soil that drive denitrification are composed of a complex mixture of species that individually encode partial (rather than complete) denitrification pathways [105–108]. The environmental significance of organismal patchiness of genes in the pathway for denitrification is mostly unexplored but may be associated with increased efficiency of denitrifying communities. Recent work has shown that pathway partitioning can decrease inter-enzyme competition in individual microbes and reduce the accumulation of toxic intermediates like NO_2^- when NO_2^- is exchanged, thus providing a potential selection mechanism for pathway partitioning in denitrifying communities [109]. However, understanding how this generalizes when multiple soluble nitrogen intermediates (i.e., NO_2^-, NO, N_2O) are exchanged as well as how the environment selects for natural communities with specific partial pathway combinations remains unexplored. Additionally, because many soil and groundwater-dwelling organisms are facultative anaerobes, the ability of these organisms to leverage multiple pathways and substrates for respiration may also contribute to denitrification pathway patchiness.

Here, we have investigated the role of pathway partitioning in a synthetic community (SynCom) composed of two isolates recovered from the same field site location at the Oak Ridge Integrated Field Research Center (OR-IFRC) (https://enigma.lbl.gov) [10], and representative of naturally occurring populations capable of exchange of pathway intermediates. This two-organism SynCom served as a model for examining the roles of biotic and abiotic factors controlling N_2O emissions in the low-oxygen and anaerobic zones of the subsurface environment of the OR-IFRC, which have high concentrations of NO_3^-, due to past nuclear waste disposal, and evidence of high levels of denitrification activity [10, 110].

Using this SynCom as a model system we have discovered how process partitioning of denitrification and exchange of multiple nitrogen intermediates improved community growth characteristics via changes in pathway kinetics (e.g., increased rate of NO_3^- reduction, removal of NO_2^-, and exchange of N_2O). We also show how community context can facilitate global changes in isolate physiology and growth dynamics, and how changes in environmental context can facilitate NO_2^- accumulation, break collaboration between the two members of the SynCom, and lead to N_2O emissions.

3.2 RESULTS

3.2.1 *Denitrification pathway is partitioned across members of microbial communities at the OR-IFRC.*

We sought to understand the structure of communities responsible for denitrification at the OR-IFRC and their relation to N_2O production. We first investigated the distribution of denitrification pathway enzymes within the genomes of 206 isolates obtained from sites (groundwater wells of various depths) spread across 3 major areas of the OR-IFRC. The sites spanned areas of high, medium, and low average groundwater NO_3^- concentration (between 0-250 mM NO_3^-). For the pathway analysis, we considered the genes Nar (or Nap), Nir, Nor, and Nos. Collectively, the reactions encoded by these genes convert NO_3^- to N_2 (Figure 3.1A). Genomic analysis revealed that the majority (~71%) of isolates were incomplete denitrifiers (i.e., missing one or more genes of the pathway). Of these, 40% likely contributed to N_2O production due to the presence of one or more genes encoding Nor, but none encoding Nos. The genomic analysis demonstrated that the denitrification pathway is largely partitioned across organisms at the OR-IFRC and suggests that complete denitrification likely occurs through interactions among organisms with complementary capabilities within microbial communities.

To better understand the role of pathway partitioning in denitrifying communities we assembled the field isolates *Rhodanobacter sp.* FW510-R12 (R12) and *Acidovorax sp.* GW101-3H11 (3H11) into a SynCom capable of complete denitrification through pathway intermediate exchange. The selection of organisms in this pairing was based on relationships between genus abundance, groundwater chemistry (e.g., NO_3^- and N_2O concentrations), and isolate co-occurrence determined through genomic and 16S rRNA profiling across the field site (Figure 3.1B and C). We found that OR-IFRC groundwater NO_3^- and N_2O concentrations were strongly correlated (Figure 3.1C, $r = 0.75$, $p < 10^{-6}$), in line with previous observations that high concentrations of NO_3^- can lead to N_2O emissions [10, 95, 111]. We also found that the abundance of genus *Rhodanobacter* was correlated with both NO_3^- and N_2O levels at the field site (Figure 3.1C, r =

0.41, $p = 3\text{x}10^{-4}$ for NO_3^- and $r = 0.27$, $p = 0.02$ for N_2O), a result that was consistent with findings from previous studies that *Rhodanobacter spp.* are most abundant in regions of the OR-IFRC highly contaminated with NO_3^- and heavy metals [29, 112]. While the abundance of genus *Acidovorax* was only weakly associated with NO_3^- and N_2O concentrations at the field site (Figure 3.1C, $r = 0.01$, $p = 0.9$ for NO_3^- and $r = -0.08$, $p = 0.5$ for N_2O), its abundance was positively correlated with that of *Rhodanobacter spp.* (Figure 3.1C, $r = 0.58$, $p = 4.9\text{x}10^{-4}$) in non-acidic groundwater ($pH \geq 7$). Flux balance analysis using constraints-based metabolic models indicated that the R12 and 3H11 SynCom could perform complete denitrification through the exchange of NO_2^- and N_2O (Figure 3.1D). Consistent with model predictions, we found that 3H11 and R12 were both capable of growth on an anaerobic minimal medium containing NO_3^- and acetate as the primary electron acceptor and donor pair, respectively. Additionally, consistent with the absence of a genome-encoded nitrite reductase (Nir), 3H11 consumed NO_3^- and accumulated NO_2^- (Figure 3.1E). By contrast, R12 took several days to initiate growth and reach saturation. R12 also consumed NO_3^-, and accumulated N_2O with no transient accumulation of NO_2^-, implying that the rate of NO_2^- reduction was equal to or greater than the rate of NO_3^- reduction (Figure 3.1E).

The SynCom growth rate was significantly greater than that of monocultures of either of the two isolates (Figure 3.1F, Welch's t-test $t = -10.7$, $p = 4\text{x}10^{-4}$; $t = -15.36$, $p = 1\text{x}10^{-4}$ for comparison of SynCom growth rate with R12 and 3H11 monocultures respectively). SynCom cultures achieved maximal biomass later than 3H11 but earlier than R12 with a maximum cell density comparable to that of R12 monocultures (Figure 3.1D). Additionally, the rate of NO_3^- reduction was increased relative to the individual monocultures, followed by a transient accumulation of NO_2^-, which was fully reduced to N_2 gas based on the lack of N_2O and ammonia accumulation (Figure 3.1E). In-depth growth characterization demonstrated that 3H11 and R12 complemented each other to perform complete denitrification, which manifested in synergistic improvement in overall growth dynamics relative to monocultures of the two organisms.

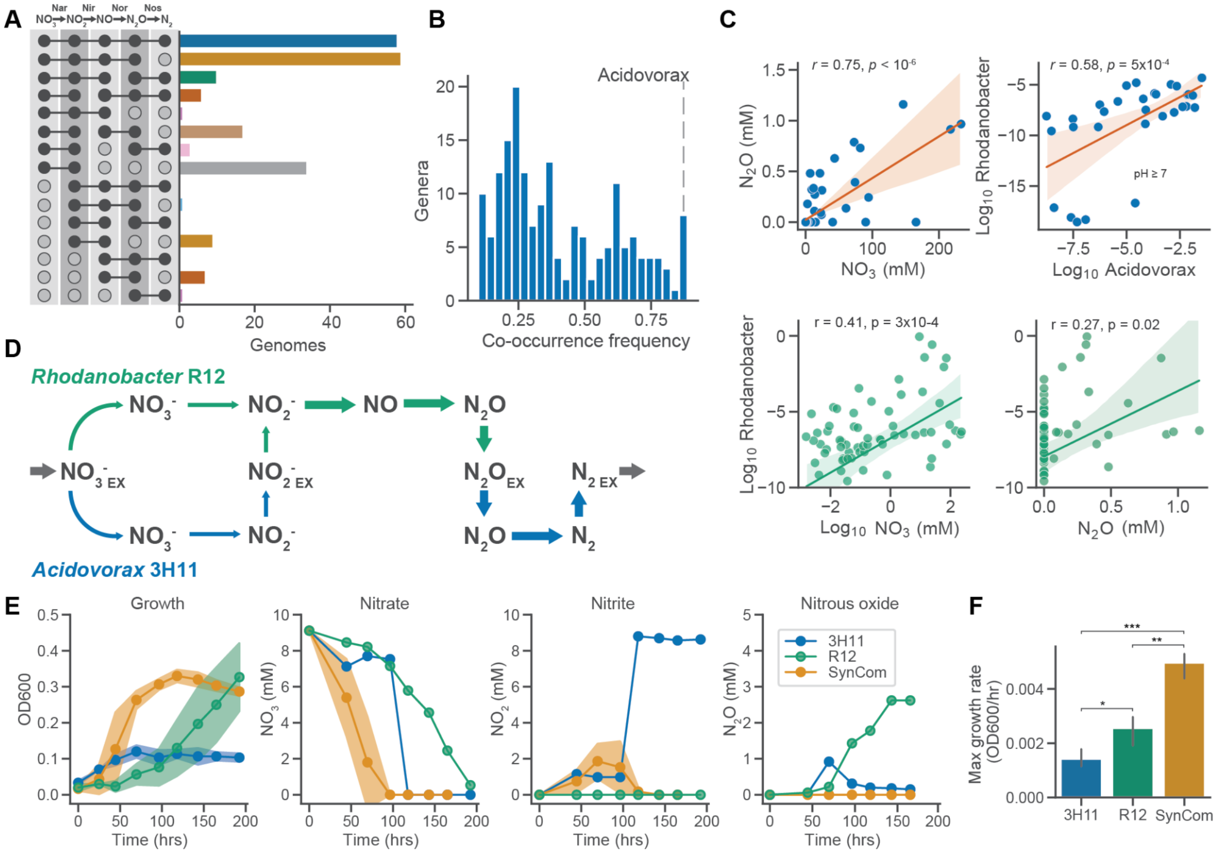

Figure 3.1. Field observations and metabolic modeling predict the exchange of intermediates in a denitrifying synthetic community. (**A**) Distribution of denitrification pathway compositions among OR-IFRC field isolates. (**B**) Genus level co-occurrence frequency of Rhodanobacter and other genera across OR-IFRC groundwater samples. (**C**) Association between groundwater chemistries and genus abundance. Pearson correlation r and p values are listed in each subplot. Linear fits to data using OLS are displayed as well as 95% confidence intervals (**D**) Denitrification flux patterns predicted by genome-scale metabolic model of R12 and 3H11 SynCom assuming equal abundances. (**E**) Growth characteristics of R12, 3H11, and R12-3H11 SynCom. Shading around trend lines represents standard deviation of samples and points represent averages. (**F**) Comparison of monoculture and SynCom maximum growth rates. Bars indicate comparisons for which differences were significant using Welch's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

3.2.2 *Exchange of intermediates in the denitrification pathway largely explain enhanced growth phenotype of the SynCom.*

We hypothesized that synergistic improvement in rate of NO_3^- reduction and overall growth characteristics of the SynCom were likely due to R12-mediated rescue of NO_2^--mediated growth inhibition of 3H11, which in turn reduced N_2O to N_2. To test this hypothesis, we characterized the growth kinetics of each monoculture in media supplemented with varied concentrations of NO_3^-, NO_2^-, and N_2O and developed kinetic models to predict the growth dynamics of each isolate and the SynCom. Growth kinetic measurements revealed that 3H11 achieved significantly greater growth rates across a wide range of NO_3^- concentrations relative to R12 (Figure 3.2A). 3H11 also achieved its maximum growth rate at much lower concentrations of NO_3^- relative to R12 ($Ks_{3H11} << Ks_{R12}$, Welch's *t*-test $t = 26.1$, $p < 10^{-6}$). While a low concentration of NO_2^- supported a significantly higher growth rate of R12 relative to its growth rate on NO_3^- (10-fold greater for 1-5 mM substrate, Figure 3.2A), the growth rate of R12 was reduced with higher concentrations ($Ki_{R12} = 13.23$ mM). No growth was detached when NO_2^- concentration was >10 mM. Additionally, assessment of biomass yield on NO_3^- and NO_2^- suggested that R12 may not conserve energy from the reduction of NO_3^- and that biomass accumulation may be entirely dependent on the reduction of NO_2^- and NO as indicated by nearly equivalent biomass yields on NO_3^- and NO_2^- ($\text{slope}_{NO3} = 0.033 \pm 0.003$, $\text{slope}_{NO2} = 0.032 \pm 0.006$ for ordinary least squares fits to the linear portion of substrate versus maximum OD600; substrate concentrations 1-10 mM; Welch's t-test $t = 1.1$, $p = 0.27$). NO_2^- also had a strong inhibitory effect on the growth rate and biomass yield of 3H11 in media containing 10 mM NO_3^-, totally abating growth at NO_2^- concentrations above 5 mM ($Ki_{3H11} = 9.11$ mM; Figure 3.2A). Finally, while 3H11 was able to accumulate biomass by reducing N_2O, its growth rate was constant and not dose-dependent (2-30 mM; $Ks_{3H11} \approx 0$, Figure 3.2A).

To assess whether improved growth characteristics of the SynCom could be explained by the growth kinetics of individual isolates on varied nitrogen oxide substrates, we developed kinetic models using a modified Monod framework by integrating logistic representation of carrying capacity into equations describing growth kinetics as a function of metabolite concentrations (Figure 3.2B; see Methods for more details) [113]. Equations in the model were parameterized using maximum growth rates as a function of substrate, half velocity constants, and carrying capacities extracted from the growth data using Logistic and Monod fits (Figure 3.2A). To represent carrying capacity as a function of substrate concentration, linear models were developed to capture relationships between NO_3^-, NO_2^-, and N_2O concentration and maximum OD600. With these parameters, the kinetic models were used in simulations to predict monoculture and SynCom growth kinetics in new experiments. The simulation was optimized to achieve better accuracy by using initial biomass as a free model parameter to effectively adjust lag phase. An initial model formulation for the reduction of NO_3^- to N_2, combining the reduction of NO_2^- and NO into a single reaction (i.e., reduction of NO_2^- to N_2O), yielded a reasonably accurate prediction of metabolite turnover and growth kinetics for R12. However, the period of maximum growth was overestimated for 3H11 and the carrying capacity achieved by R12 was not well represented. Furthermore, the prediction of SynCom dynamics was quite poor. Subsequently, the addition of the effect of NO_2^- inhibition into the models improved the accuracy of predicted growth dynamics of 3H11 and modest improvement in predicted growth dynamics of R12, which had higher tolerance to NO_2^-. While the introduction of NO_2^- inhibition improved the prediction of SynCom growth dynamics, overall model accuracy was still poor (Figure 3.2C). The period of exponential growth was overestimated and NO_2^- turnover in particular was still poorly predicted. These results challenged our assumptions underlying the model and forced us to reconsider the overall scheme of metabolite exchange across 3H11 and R12. The only plausible alternative was that NO released from R12 was reduced by 3H11. This scenario would be possible if the affinity and rate of NO reduction was considerably higher for 3H11 than R12 as was the case for NO_3^-. To represent this scenario in the model we doubled the contribution of N_2O for 3H11 (assuming a similar contribution to its growth) and halved the contribution of NO_2^- reduction for R12, which was estimated from the energetics of these processes due to the absence of measured rates of NO reduction. We found that this assumption significantly improved the accuracy of predicted growth dynamics of the SynCom and suggested that 3H11 might be the dominant member of the community, which was in stark contrast to its relatively poor growth in monoculture (Figure 3.2C). Thus, the kinetic modeling demonstrated that the scheme of metabolite exchange between 3H11 and R12 was critical for explaining the growth dynamics of the SynCom.

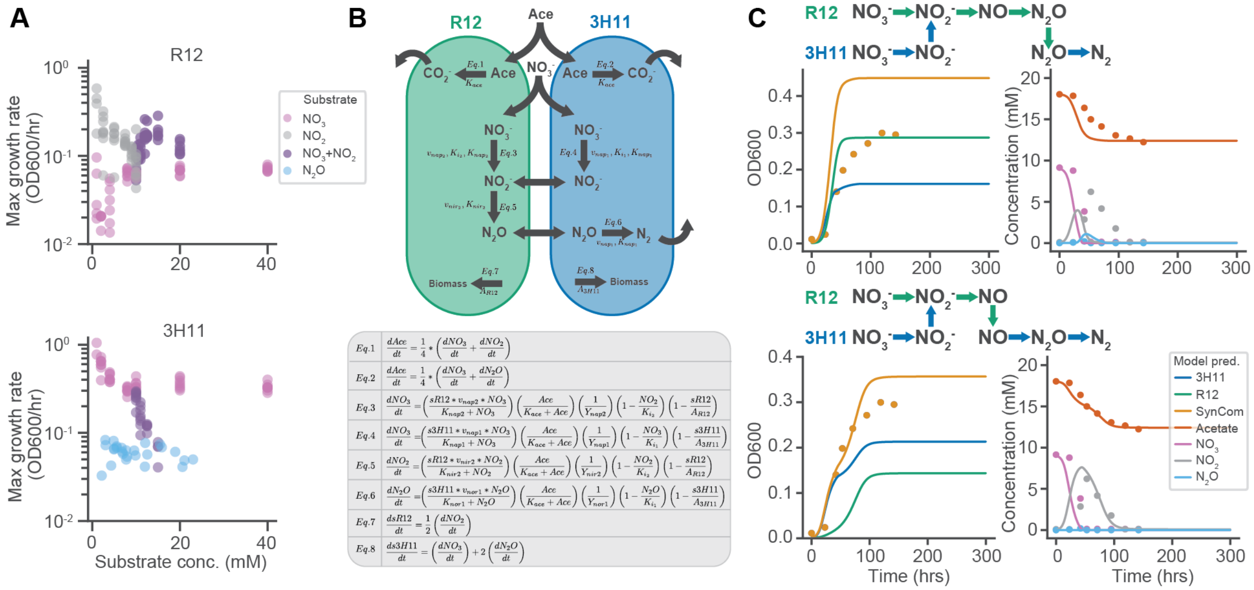

Figure 3.2. Exchange of pathway intermediates drives growth enhancement in R12-3H11 SynCom. (**A**) Maximum growth rates of R12 and 3H11 across variations in substrate concentrations. For conditions where NO_3^- and NO_2^- were present, 10 mM NO_3^- was used and NO_2^- was varied. (**B**) Schematic of R12-3H11 kinetic model and associated differential equations. The model represents growth resulting from the oxidation of acetate and reduction of NO_3^-, NO_2^-,

and N_2O via denitrification. Given challenges of measuring NO kinetics, NO reduction is not explicitly represented in the model, instead NO_2^- is reduced to N_2O (**C**) Two SynCom intermediate exchange scenarios and associated dynamics predicted by kinetic models displayed with empirical data. Model predictions are represented using solid lines, data are averages across samples and are represented using circular points.

3.2.3 *Community context drives global changes in transcriptional states of 3H11 and R12.*

To characterize the impact of paired growth on organismal physiology, we compared temporal changes in transcriptomes of each isolate grown as a monoculture and as a SynCom member. Transcriptomes were longitudinally profiled at four time points that corresponded with early-, mid-, and late-phases of growth for all cultures except for the 3H11 monoculture, which was profiled at two time points because of its fast growth rate and limited biomass (Figure 3.3A). Principal component analysis (PCA) indicated that growth in a SynCom context was associated with global changes in transcriptional states of both 3H11 and R12 (Figure 3.3B). Transcriptome states of R12 were distinct across the two contexts, with the monoculture transcriptomes differentiated along PC1, and SynCom transcriptomes aligned with PC2. The proximity of the monoculture T2 samples with those of SynCom T2 and T3 samples suggested similar transcriptional states were achieved at mid-log phase of growth across the two contexts. In contrast, transcriptional states of 3H11 in monoculture and SynCom contexts were similar in early log phase (T1), but they diverged in later stages of growth.

To better understand mechanisms of improved growth dynamics of the SynCom, we used the software package DESeq2 [44] to identify differentially expressed genes (DEGs) in each isolate across growth phases in the two contexts. Altogether, across all contexts, 321 of 4,670 genes were differentially regulated in 3H11 and 480 of 3,493 genes were differentially regulated in R12 (Figure 3.3C). We further explored global patterns in pathway regulation associated with growth context using k-means clustering and pathway ontology enrichment analysis (Figure 3.3D). Broadly, the DEGs identified fell into a few metabolic pathways, including denitrification, amino acid metabolism, central carbon metabolism, cofactor biosynthesis and transport, signaling and cellular processes, and genetic information processing (Figure 3.3D, E).

Expression patterns of genes in the denitrification pathway supported the hypothesis that 3H11 was the primary reducer of NO and N_2O when paired with R12 (Figure 3.3E). In particular, expression of *norB* and *nosZ* genes in 3H11 were significantly upregulated in the SynCom context (Welch's t-test $t = -11.8$, $p < 10^{-6}$ for *norB*; $t = -4.5$; $p = 0.003$ for *nosZ*). In addition to its overall higher average expression in the SynCom context, *norB* transcript levels were further upregulated during the period when NO_2^- was reduced, reaching maximum levels at T3. Transcript levels of 3H11 *nosZ* increased from early- (T1) to mid-log phase (T2) in the SynCom context and decreased thereafter, mirroring corresponding changes in NO_2^- levels. In R12, across all contexts, transcript levels were relatively stable for all denitrification genes (Figure 3.3E), except *nirK*, which was differentially expressed in the late-log phase of growth. Although the overall transcript level of *nirK* was lower on average in the SynCom context, its expression increased in the SynCom relative to monoculture growth during the period of active NO_2^- reduction ($\log_2$ fold change = 2.62, $p < 10^{-6}$ for comparison of T4 *nirK* expression). Additionally, while there was no significant change in the average expression of R12 *narG*, the dynamics changed with context. Maximum expression levels of R12 *narG* were achieved at 95 hours of growth in the SynCom, while in monoculture similar expression levels 24 hours later (Figure 3.3E) at 119 hours, which coincided with the timing of NO_2^- accumulation in each of these contexts. This observation was consistent with the observation that R12 generally achieved a higher maximum growth rate with a shorter lag phase during growth on NO_2^- relative to NO_3^- (Figure 3.2A) and the increased rate of NO_3^- reduction by the SynCom relative to monocultures.

Transcriptional changes across multiple pathways suggested additional mechanisms of competition and cooperation between R12 and 3H11 as a SynCom. For example, expression patterns for 6 of 34 tricarboxylic acid (TCA) cycle genes in 3H11 and 4 of 33 genes in R12 coincided with the timing of NO_3^- and NO/N_2O reduction across growth contexts (Figure 3.3E). Temporal expression patterns of TCA cycle genes in 3H11 coincided with the reduction of NO_3^- (T1 in both monoculture and SynCom contexts) and with NO_2^- (T4 in SynCom context). Unlike 3H11, expression of TCA cycle genes in R12 was lower in the initial time points, when NO_3^- was reduced, and increased in later phases of growth, when NO_2^- and N_2O were being reduced (e.g., R12_3138, *fumA*, downregulated in T1, $\log_2$ fold change = -2.14, $p < 10^{-6}$ and upregulated in T4, $\log_2$ fold change = 3.06, $p < 10^{-6}$). Additionally, 38 motility-associated genes (e.g., *cheA-Z*, *flgA-M*, *fliD-Q*, and *pilA*) were upregulated in R12 in the SynCom context, which may be a mechanism for interaction with 3H11 or competition for NO_3^- (Figure 3.3E). Similarly, a significant number of amino acid metabolism genes, 74 in total, were differentially regulated across 3H11 and R12. While the pathway for biosynthesis of branched-chain amino acids (BCAAs) was upregulated in both 3H11 and R12 in the SynCom context, the upregulation of genes encoding leucine efflux in 3H11 in conjunction with upregulation of BCAA degradation genes in R12 indicated a possible exchange of these amino acids. Supporting this, we found that allowing the exchange of individual amino acids in our SynCom metabolic model led to increased SynCom biomass yield by 2-3% for individual BCAAs and favored the direction of exchange suggested by the expression analyses. Collectively, the gene expression analysis provided support for the hypothesis that 3H11 is the primary reducer of NO and N_2O and identified potentially additional mechanisms of metabolic interplay that likely contribute to the improved growth characteristics of the SynCom.

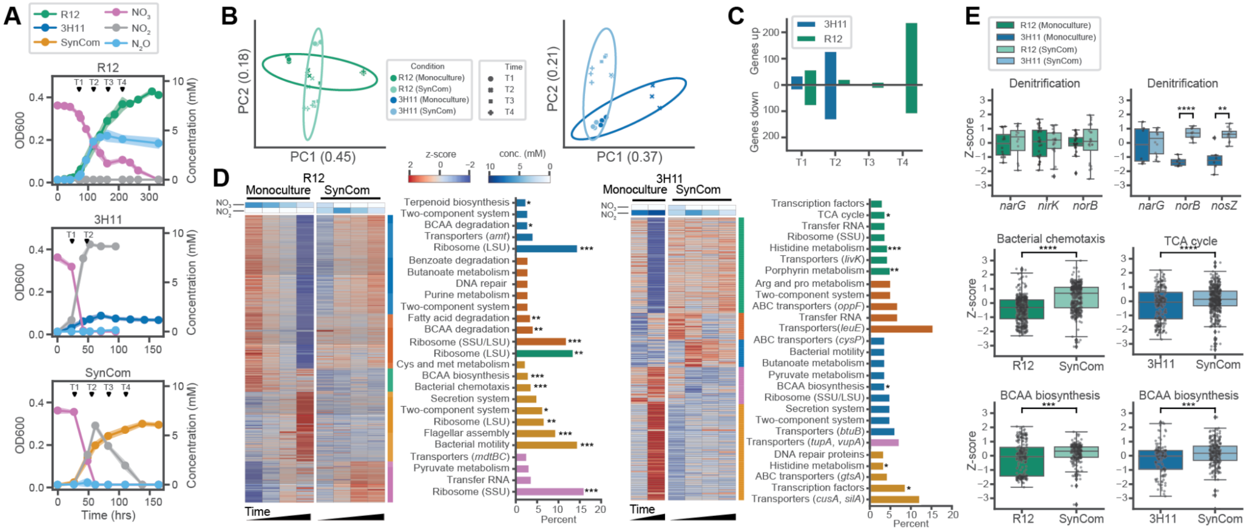

Figure 3.3. Community context drives changes in the global transcriptional states of 3H11 and R12. (**A**) Growth characteristics of R12, 3H11, the SynCom, and associated transcriptomics sampling points. Black triangles mark condition specific time points (T1-T4). Shading around trend lines represents standard deviations and points represent averages. (**B**) PCA of R12 and 3H11 gene expression across monoculture and co-culture growth contexts. DESeq2 log normalized gene expression Z-scores used. Z-scores computed across growth contexts for R12 and 3H11. (**C**) R12 and 3H11 specific DEG counts for comparisons across time and growth contexts. (**D**) K-means clustered R12 and 3H11 specific DEGs using k = 5. Expression Z-scores are displayed across time and growth contexts. Associated NO_3^- and NO_2^- concentrations are displayed above each heatmap. To the right of each heat map, the most abundant KEGG subpathway terms are displayed for each cluster. Bars are colored by the cluster they represent. For ribosome terms, the most abundant subunit is noted. Large subunit (LSU) or small subunit (SSU). For ABC transporter and transporter terms, a representative gene in each cluster is noted. Subpathway terms are noted which were significantly enriched in each cluster using the hypergeometric test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ (**E**). Boxplots of gene and pathway expression Z-scores. Bars indicate comparisons for which differences were significant using Welch's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$

3.2.4 *Proteomics analysis supports the dominant role of 3H11 in a SynCom across varied nitrate concentrations and reveals global physiological adaptations.*

We investigated how environmental context impacts both the composition of the SynCom and dynamics of denitrification by the two-member community. Specifically, we used mass spectrometry-based proteomics to characterize steady state community composition and functional characteristics of SynComs grown in media with varying initial concentrations of NO_3^- (1-40 mM NO_3^-, 50% 3H11) and varying initial community compositions (10 mM NO_3^-, 5-98% 3H11). We then leveraged the proteomics data to accurately estimate 3H11 and R12 composition within the mixed population and steady state protein abundances. Consistent with the kinetic model prediction (Figure 3.2C), 3H11 was the dominant member, representing ~65% of the overall SynCom composition (Figure 3.4A). Remarkably, the steady state composition of the SynCom converged to the same relative proportion of ~65% 3H11 and ~35% R12, regardless of the wide variation in initial ratio of the two organisms in the inoculum (Figure 3.4A).

PCA revealed that 36% of the variance in abundance profiles of 2,401 SynCom proteins was captured by variation in NO_3^- concentration of the growth medium, which was highly correlated with PC1 ($r = -0.94$, $p < 10^{-6}$; Figure 3.4B). Abundance changes in nearly half of the proteome (1,105 proteins) were significantly correlated with NO_3^- concentration ($r \leq 0.5$ or $r \geq 0.5$ and $q < 0.05$; Figure 3.4C). Globally, protein abundance changes were also significantly influenced by initial community composition, which showed moderate correlations with PC1 and PC2 ($r = 0.58$, $p < 10^{-6}$; $r = 0.32$, $p = 0.011$ for PC1 and PC2 respectively; Figure 3.4B). Changes in the abundance of 206 proteins were significantly correlated with community composition ($r \leq 0.5$ or $r \geq 0.5$ and $q < 0.05$). Broadly, the pathways and processes with the most proteins correlated with NO_3^- concentration and community compositions fell into a few categories. These included energy metabolism, central carbon metabolism, amino acid metabolism, cofactor biosynthesis and transport, and genetic information processing (Figure 3.4C).

Patterns in steady state abundance of enzymes across the denitrification pathway suggested that both R12 and 3H11 have distinct regulatory schemes that could benefit overall growth characteristics of the SynCom across a range of NO_3^- concentrations (Figure 3.4D). For instance, the inverse correlation of 3H11 NarG protein abundance and NO_3^- levels ($r < -0.6$, $p < 10^{-6}$ for all three copies of 3H11 NarG) suggests that regulation of this enzyme is highly sensitive to repression by NO_2^-. By contrast, expression of R12 NarG is stimulated by low concentrations of NO_3^- <10mM, but is repressed at concentrations >10mM, also likely due to NO_2^- accumulation. The abundance of R12 NirK was significantly correlated with NO_3^- concentration ($r = 0.82$, $p < 10^{-6}$), suggesting that its expression was likely activated by and proportional to NO_2^- levels. These patterns suggest that the SynCom may have regulatory mechanisms that balance the levels of NarG and NirK to minimize the accumulation of NO_2^- in the presence of higher concentrations of NO_3^-. Interestingly, trends in NorB were opposite for R12 and 3H11. Abundance of NorB increased for 3H11 as a function of NO_3^- ($r = 0.6$, $p = 1.5 \times 10^{-6}$), while for R12 it decreased ($r = -0.96$, $p < 10^{-6}$). Additionally, the abundance of 3H11 NosZ decreased with NO_3^- ($r = -0.63$, $p < 10^{-6}$; Figure 3.4D). Similar trends were observed for NarG ($r < -0.4$, $p < 10^{-3}$ for 2 of 3 3H11 NarG; $r = 0.03$, $p = 0.85$ for R12 NarG) and NirK ($r = 0.41$, $p = 0.005$) across variation in community composition but were less pronounced (Figure 3.4E). The decrease in NorB and NosZ in 3H11, with increase in its abundance in the inoculum ($r = -0.52$, $p = 1.9 \times 10^{-4}$; $r = -0.33$, $p = 0.026$ for NorB and NosZ respectively; Figure 3.4E) is likely a consequence of NO_2^- inhibition because the rate of 3H11-mediated NO_2^- production exceeded the rate of NO_2^- reduction by R12 NirK. Thus, correspondence in trends between NarG and NirK across conditions suggested that NO_2^- accumulation resulted from an increase in NO_3^- concentration as well as an increase in the proportion of 3H11 in the initial composition of the SynCom.

Protein abundance patterns across multiple pathways showed opposite trends between 3H11 and R12. For instance, whereas 40 of the 67 central carbon metabolism enzymes in R12 that were assayed were negatively correlated with NO_3^-, 68 of 87 assayed enzymes in the corresponding pathways in 3H11 were positively correlated with NO_3^- (Figure 3.4F). Similar opposing trends in protein abundance in 3H11 and R12 were also observed for initial community composition (e.g., proteins of TCA and glyoxylate cycles). Collectively these trends in protein abundance suggest that 3H11 outcompetes R12 for substrates (acetate, NO_3^-, NO etc.) in the SynCom. Furthermore, our results suggest that the competition for substrates becomes more pronounced with increases in the relative proportion of 3H11 in the inoculum, as well as with higher concentrations of NO_3^- (Figure 3.4G). Process partitioning is an alternative explanation for these trends with the two organisms in the SynCom complementing each other's needs and capabilities. For instance, the opposing trends in protein abundance were also observed for proteins in multiple amino acid biosynthesis pathways. In particular, while four enzymes in the methionine biosynthesis pathway in R12 had decreased abundance with increasing NO_3^- concentration, the abundance of a methionine importer, MetQ, increased (r = 0.7, $p < 10^{-6}$). This regulation in R12 was likely mediated at the transcriptional level by MetJ, a putative repressor of methionine biosynthesis, which was upregulated with increased NO_3^- (r = 0.53, $p = 5.6\text{x}10^{-5}$). By contrast, enzymes of methionine biosynthesis in 3H11 were upregulated with increased NO_3^- concentration ($r > 0.5$, $p < 10^{-6}$ for MetE, MetH and MetY). These trends suggested that methionine biosynthesis was likely happening in 3H11 and complementing the need of R12 for this amino acid. The NO_3^- -induced shifts in the abundance of enzymes of amino acid metabolism in both R12 and 3H11 were consistent with patterns in the transcriptomics analysis indicating that much of this regulation was mediated at the transcriptional level (Figure 3.4F). The proteomics analysis supported many of the conclusions from the transcriptomics analysis regarding the likely involvement of specific metabolic processes in enhancing growth characteristics of the SynCom, competition between 3H11 and R12, and that relative levels of NO_3^-, NO_2^-, NO, and N_2O likely drive observed changes in dynamics of denitrification and growth.

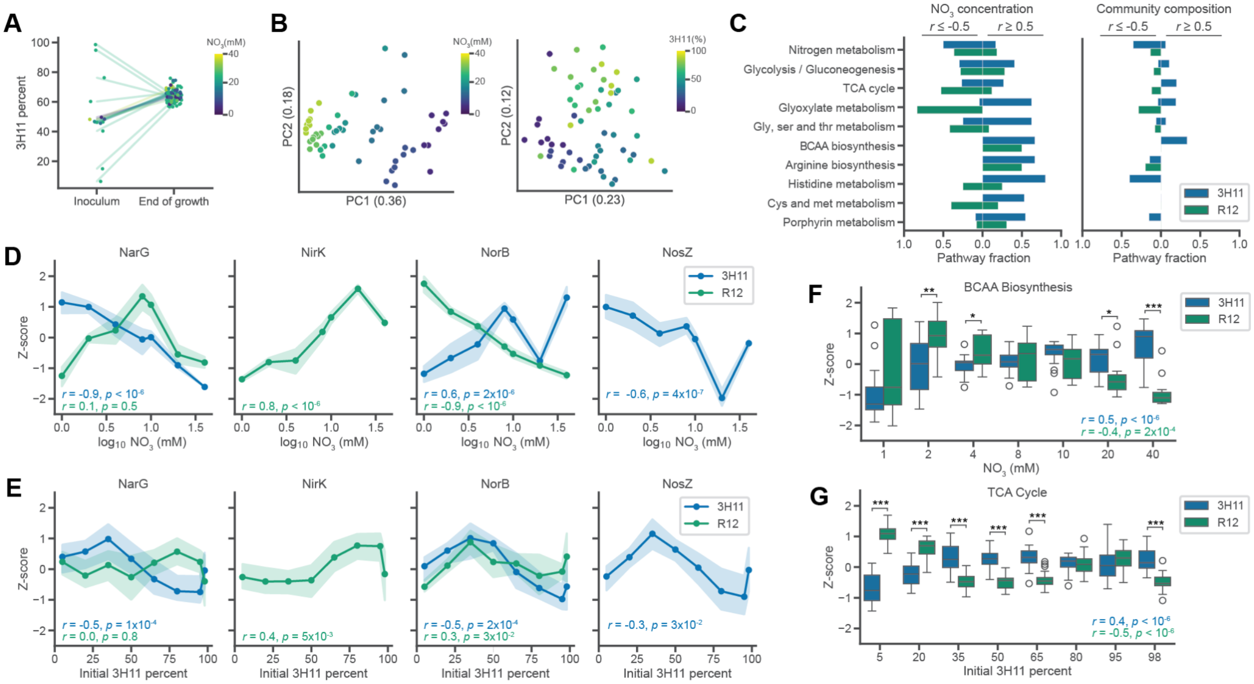

Figure 3.4. Variations in nitrate and community composition support the dominant role of 3H11 in the SynCom. (**A**) 3H11 SynCom percentage across variation in initial community composition and NO_3^- concentration. Community composition measured before inoculation and at the end of growth. The color indicates NO_3^- concentration used and lines connect the initial and end points of individual cultures. (**B**) PCA of SynCom end-point proteome Z-scores across variation in initial community composition and NO_3^- concentration. Color is used to represent growth conditions. (**C**) Fraction of detected 3H11 and R12 specific pathway proteins significantly correlated with NO_3^- and community composition. (**D**) End-point 3H11 and R12 denitrification protein Z-scores as a function of initial NO3 concentration. (**E**) End-point 3H11 and R12 denitrification protein Z-scores as a function of initial community composition. Points represent

an average of eight replicates, error bands display standard deviation. 3H11 and R12 specific Pearson *r* and *p* values are displayed on each subplot. (**F**) Boxplots of BCAA biosynthesis protein Z-scores across NO_3^- concentration. (**G**) Boxplots of TCA cycle protein Z-scores across community composition. Bars indicate comparisons for which differences were significant using Welch's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

3.2.5 *Nitrite accumulation drives nitrous oxide emissions across variation in initial nitrate concentration and community composition.*

Furthermore, we investigated how variation in extrinsic factors (specifically, NO_3^- concentration) and intrinsic properties (specifically, community composition) impact N_2O emissions. We measured the steady state levels of N_2O in SynCom cultures across a range of initial NO_3^- concentrations (1 mM vs 40 mM) and community composition (3H11:R12 ratio of 5%:95% vs 98%:2%). We expected that extremes in community composition and NO_3^- concentration might yield differences in community phenotype and steady state levels of N_2O. Consistent with our expectations, steady state levels of N_2O increased proportionally with the initial concentration of NO_3^- . However, we were surprised to find that it also increased with the relative proportions of 3H11 in the inoculum (Figure 3.5A). The proportional increase in N_2O off gassing with the amount of input NO_3^- was consistent with the stoichiometry of denitrification, which was expected to be limited by the availability of acetate when NO_3^- concentration exceeded 30 mM. That is, at high concentrations of NO_3^- , acetate would be exhausted before NO_3^- could be completely reduced to N_2 given that earlier steps in the pathway (e.g., NO_3^- and NO_2^- reduction) were prioritized by the SynCom. However, we observed N_2O accumulation at 20 mM NO_3^-, below the concentration at which acetate becomes limiting. Similarly, while N_2O accumulation was expected with high NO_3^- , we were surprised that despite its role as the primary N_2O reducer in the SynCom higher initial proportions of 3H11 also resulted in N_2O accumulation. Together these findings suggested a more complex interplay between 3H11 and R12 underlies the dynamics of denitrification by the SynCom.

To characterize possible mechanisms responsible for the production of N_2O by the SynCom, we looked at associations between N_2O concentration and protein abundance across all conditions. We identified proteins that were significantly correlated with steady state N_2O levels across contexts and had consistent trends across both datasets (e.g., r values that were consistently positive or negative across both datasets). We identified 13 proteins that were positively correlated with N_2O levels across NO_3^- concentrations and community composition ($r \geq 0.5$ and $q < 0.05$) and 12 proteins with abundance changes that were negatively correlated ($r \leq -0.5$ and $q < 0.05$).

Proteins that were positively correlated with N_2O were largely from R12 (10 of 13) and of unknown function (6 of 10). The four R12 proteins with predicted functions were NirK ($r = 0.69$, $p < 10^{-6}$), an iron transporter ($r = 0.64$, $p < 10^{-6}$), Bfr (bacterioferritin; $r = 0.83$, $p < 10^{-6}$), and Ahr (alcohol/geraniol dehydrogenase; $r = 0.76$, $p < 10^{-6}$). The three 3H11 proteins that were positively correlated with N_2O had predicted functions related to central carbon and amino acid metabolism and included AceB, GlcB (malate synthase; $r = 0.53$, $p = 1.1\text{x}10^{-4}$), Ddl (D-alanine-D-alanine ligase; $r = 0.53$, $p = 1.1\text{x}10^{-4}$), and ilvH, ilvN (acetolactate synthase; $r = 0.59$, $p = 1.0\text{x}10^{-5}$). In contrast, proteins with abundance changes negatively correlated with N_2O levels were primarily from 3H11 (10 of 12) and related to energy metabolism (7 of 10). Two of the seven energy metabolism proteins were cytochromes, while the rest were from the denitrification pathway, including two copies of NarG, NarH, NosZ, NosR (a transcriptional regulator of NosZ) and PhasZ (PHB depolymerase) (Figure 3.5B). Using a multi-linear regression framework, we investigated the ability of protein abundances within specific pathways to predict observed steady state N_2O

levels (Figure 3.5C, see Methods for additional details). As expected, protein abundances within the nitrogen metabolism pathway had the best performance in predicting steady state N_2O levels (out of sample $R^2 = 0.53$). The patterns of correlations between the abundance of denitrification enzymes in the two organisms and N_2O (e.g., positive correlation of R12 NirK and negative correlation of 3H11 NosZ), suggested that NO_2^- toxicity might be the driver of N_2O emissions in the SynCom. This hypothesis was supported by the strong correlation between both maximum concentration and cumulative (i.e., area under the curve (AUC)) levels of NO_2^- predicted by the kinetic model, and the observed steady state N_2O levels for each experimental context (i.e., initial NO_3^- levels and community composition). Relationships between steady state N_2O levels and both predicted NO_2^- metrics identified an inflection point that likely corresponds to the concentration at which NO_2^- became toxic, inhibiting key aspects of process partitioning between the two organisms, which ultimately manifested in increased N_2O emissions (Figure 3.5D). This inflection point for N_2O production occurred at around 7.5 mM NO_2^-, which was in line with the observation that growth of 3H11 monocultures were completely inhibited at NO_2^- concentrations > 5 mM (Figure 3.2A).

Finally, we investigated factors governing NO_2^- accumulation by using the SynCom kinetic model. We predicted the total amount of NO_2^- produced by the SynCom across a range of initial co-culture conditions, community compositions (0.1-99.9 % 3H11) and NO_3^- concentrations (1-30 mM) indicating that NO_3^- concentration and community composition had non-linear effects on the accumulation of NO_2^- to a level that would result in N_2O off gassing through growth inhibition of 3H11. Together our findings demonstrate that during process partitioning of denitrification, N_2O production is primarily driven by NO_2^- toxicity, which manifests from a complex interplay of environmental context and community composition.

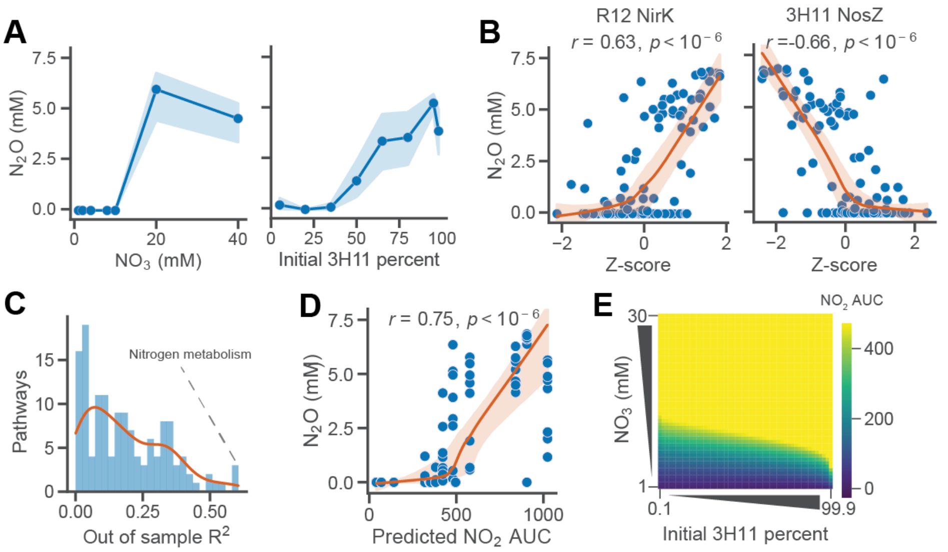

Figure 3.5. Nitrite inhibition drives nitrous oxide accumulation in the SynCom. (**A**) End-point SynCom N_2O concentration as a function of initial NO_3^- concentration and community composition. Points represent an average of eight replicates, error bands display standard deviation. (**B**) End-point SynCom N_2O concentration as a function of denitrification protein Z-scores across conditions. Lowess trendline and 95% confidence interval displayed. Pearson *r* and *p* values are displayed. (**C**) Distribution of predicted N_2O out of sample R^2 using pathway level multi-linear regression. R^2 of nitrogen metabolism proteins noted. (**D**) End-point SynCom N_2O concentration as a function of predicted SynCom NO_2^- AUC across conditions. Lowess trendline and OLS linear fit displayed with respective 95% confidence intervals. Pearson *r* and *p* values are displayed. (**E**) Heatmap of predicted SynCom NO_2^- AUC across variation in initial NO_3^- concentration and community composition.

3.3 DISCUSSION

Our findings demonstrate how two ecologically relevant partial denitrifiers cooperate as a SynCom to perform complete denitrification (Figure 3.1), with detailed physiological characterization of their interactions (Figures 3.2-3.4), which has detailed NO_2^- toxicity as a key driver of N_2O emissions (Figure 3.5). Our findings underscore how the exchange of specific intermediates completes denitrification through process partitioning, while also determining community composition and dynamics of denitrification (Figures 3.1 and 3.2) [104]. Consistent with our finding, Lija *et al.* showed that pathway partitioning eliminates inter-enzyme competition (e.g., between *narG* and *nirK*) to promote and maintain biodiversity in microbial communities by increasing the rate of consumption and reducing the accumulation of toxic intermediates like NO_2^- , [109]. Notably, we observed that R12 monocultures do not accumulate NO_2^- and can grow with significantly reduced lag time and higher growth rates on NO_2^- relative to NO_3^- (Figures 3.1E and 3.2A). This observation suggests that R12 has evolved a balance between NarG and Nirk, or between these enzymes and other rate limiting steps such as NO_3^- transport, that favors NO_2^- reduction over NO_3^-. This is in contrast to the strain of *Pseudomonas stutzeri* used by Lilja *et al.* which favors NO_3^- reduction and thus accumulates NO_2^- [109]. The environment R12 was isolated from provides an excess of NO_3^- and *Rhodanobacter* have been shown to dominate groundwater communities in regions of the OR-IFRC with low pH and high NO_3^- [29, 112, 114]. Thus, the groundwater conditions at OR-IFRC might provide a selection pressure that favors the balance in NO_3^- and NO_2^- reduction achieved by R12 and possibly other *Rhodanobacters*. While we note the natural variation that can be achieved by environmental isolates and communities, our findings are consistent with the principle of inter-enzyme competition described by Lilja *et al.*[109]. The observed increase in SynCom specific rate relative to R12 and 3H11 monocultures, as well as the increased rate of NO_3^- reduction (Figure 3.1E, and F), suggests that in SynCom inter-enzyme competition is reduced. However, we cannot rule out the possibility that the observed improvements in specific and NO_3^- reduction rates are due to the alleviation of NO_2^- toxicity. Indeed, the accumulation of NO_2^- observed in SynCom was largely attributed to the increased rate of NO_3^{--} reduction facilitated by 3H11, which effectively shifted the balance achieved by R12 toward NO_3^- reduction. (S3C). Perhaps, if inter-enzyme competition between NarG and NirK were eliminated in R12 by knocking out NarG, the rate of NO_2^- removal in the SynCom could be further increased and NO_2^- accumulation minimized. Chemical removal NO_2^- from the system might also allow better interrogation of the role of NO_2^- toxicity. Our results suggest that in addition to inter-enzyme competition, natural variation in community member denitrification kinetics and intermediate toxicity are important additional factors to consider and likely contribute to selection mechanisms for community assembly in the environment. Future work should explore how inter-enzyme competition and natural variation in enzyme kinetics interact to shape dynamics in environmentally relevant communities.

In addition to supporting the relevance of pathway partitioning in denitrifying communities, our analysis highlights how emergent phenotypes of communities can arise from differences in isolate substrate kinetics (Figure 3.2). Specifically, we discovered four key mechanisms that contributed to the overall improvement in the growth dynamics of the SynCom. First, the fast rate of 3H11-mediated reduction of NO_3^- into NO_2^- promoted the growth of R12, a finding supported by evidence that monocultures of R12 grew better on NO_2^-, relative to NO_3^-, with a shorter lag phase and increased growth rate (Figure 3.2A). Second, by consuming NO_2^-, R12 facilitated increased biomass of 3H11, based on evidence that 3H11 was severely growth inhibited due to accumulation of NO_2^- when grown as a monoculture, but not in a SynCom context wherein NO_2^- levels increased transiently and returned to undetectable levels (Figure 3.2C). Third, the exchange of NO (instead of N_2O) from R12 to 3H11 turned out to be a critical model-enabled discovery, which was supported by the subsequent observation that 3H11 *norB* was significantly upregulated in the SynCom context (Figure 3.3E). Finally, the finding that N_2O can itself serve as a source of energy and biomass for 3H11, added evidence for a fourth mechanism. Collectively, the four mechanisms incorporated into a kinetic model explained how the interplay between the two isolates improved the overall growth characteristics of the SynCom (Figure 3.2C). The kinetic model was crucial in elucidating the importance of NO exchange from R12 to 3H11 (as opposed to N_2O exchange predicted by the reconstructed genome-scale metabolic model). This important change significantly improved model performance, and accurately predicted the dominant role of 3H11 in the SynCom, as well as the role of NO_2^- inhibition in the production of N_2O (Figure 3.4A and 3.5D). Thus, while stoichiometric models have proven themselves valuable tools for understanding microbial interactions at a systems scale, we found that kinetics were an important feature for understanding the emergent properties of our system [25, 115–117]. Indeed, kinetic models have been instrumental in characterizing multiple phenomena, including growth dynamics of bacteria in the subsurface of oil wells, dynamics of N_2O emissions in soils, cycling of iron in the ocean, and fundamental ecological principles [109, 118–121]. Future work should integrate denitrification kinetics into monoculture and SynCom metabolic models to better understand the dynamics of community metabolism and how it changes with growth context [122, 123]. Experimentally derived biomass composition and integration of gene regulation may also provide a means to improve metabolic model accuracy and further explore questions related to the regulation of metabolism [26, 124].

While the kinetic model provided a better mechanistic explanation for the improved growth characteristics of the SynCom, it also suggested functional interactions between the two organisms extended beyond the core process of denitrification. Indeed, gene expression and proteomics analyses uncovered widespread physiological changes across both organisms when they were grown as a SynCom (Figures 3.3 and 3.4). Upregulation of BCAA and histidine biosynthesis by 3H11 in SynCom and the positive correlations with proteins in these pathways with NO_3^- and N_2O levels suggest that increases in the activity of these amino acid biosynthesis pathways were not related to the presence of R12 necessarily but instead the availability and utilization of NO and N_2O by 3H11. In line with this hypothesis, recent studies have shown that BCAA biosynthesis provides an important means to balance intracellular redox in the purple non-sulfur bacteria *Rhodospirillum rubrum* when growing under photoheterotrophic conditions [125, 126]. BCAA biosynthesis was also shown to play an important role in mitigating the inhibitory effects of reduced sulfur for the nitrate reducing bacteria *Intrasporangium calvum* [99] (See Chapter 2). Thus, BCAA biosynthesis may also provide a means to balance redox for denitrifiers like 3H11 and R12. Additionally, upregulation of leucine export by 3H11 and BCAA degradation genes by R12 (e.g., leucine dehydrogenase, and propionyl-CoA carboxylase) in the SynCom suggested the exchange of leucine between 3H11 and R12. We also identified trends in protein abundance that suggested methionine might be exchanged by 3H11 and R12. A recent study by Ramoneda *et al.* characterizing the role of flagellar motility genes in soil bacterial communities and found that these genes were often associated with soil carbon availability [127]. Thus, the upregulation of flagellar motility genes by R12 in SynCom could have been a response to amino acid export by 3H11. The SynCom stoichiometric model also supported the possibility of amino acid exchange, but suggested the overall contribution to SynCom biomass was small (<10% total biomass). While these potential exchanges are in line with recent work suggesting that the exchange of amino acids, vitamins, and other metabolites is ubiquitous in natural microbial communities, additional characterization is needed to support the exchange of amino acids and other byproducts of metabolism between R12 and 3H11 [128].

Interestingly, our analyses point toward NO_2^- accumulation as an important factor that can contribute to N_2O emissions in some contexts. We found that the inhibitory effect of NO_2^- on 3H11 could be compounded by the increased rate of NO_3^- reduction and accumulation of NO_2^- (Figure

3.2A) when initial community compositions were skewed in favor of 3H11 (Figure 3.5). Supporting the generalizability of this phenomenon, studies have linked NO_2^- with N_2O emissions in other contexts. Maharjan and Ventera showed that NO_2^- accumulation was a major driver of N_2O emissions across different nitrogen fertilizer regimes in an agricultural setting and found that mitigation of NO_2^- accumulation reduced N_2O emissions [129]. NO_2^- has also been implicated in N_2O emissions in the context of wastewater treatment, where free nitrous acid (HNO_2) has been shown to drive emissions for denitrifying enhanced biological phosphorus removal sludge via inhibition [130]. In this context the authors point toward free HNO_2 rather than NO_2^- itself as the source of inhibition, thus implicating pH as an important factor for the link between nitrite and N_2O emissions in denitrifying communities. Supporting this, other studies have pointed to at least two pH dependent inhibitory mechanisms for NO_2^-. The first involves HNO_2, which uncouples proton translocation when formed at lower pH [109, 131, 132]. The other involves the spontaneous generation of NO radicals that disrupt enzyme activity via the generation of metal-nitrosyl complexes [133]. So, while the inhibitory effects of NO_2^- are well characterized and have been associated with N_2O emission in other contexts here we provide insights into additional factors that can contribute to NO_2^- accumulation (e.g., NO_3^- concentration, community composition, substrate kinetics) and N_2O emissions.

The SynCom presented here provides a simplified, environmentally relevant, model of pathway partitioning in denitrifying communities and a framework with which to explore how environmental context can impact cooperation and lead to the production of N_2O. Future studies will leverage this SynCom to further explore biotic and abiotic controls of denitrification, such as pH, substrate concentrations, carbon source, oxygen, and other factors associated with niche differentiation.

3.4 METHODS

3.4.1 *OR-IFRC isolate denitrification pathway composition analysis*

Annotated OR-IFRC isolate genomes were obtained from KBase (https://narrative.kbase.us/narrative/41372) and the presence of the denitrification genes Nar/Nap, Nir, Nor, and Nos were characterized. Isolates with one or more denitrification genes were considered. Denitrification pathway composition of the isolates was tabulated using gene presence/absence.

3.4.2 *OR-IFRC groundwater genus abundance and chemistry analysis*

Groundwater OTU abundance, taxonomy, and chemical data previously described by Smith et al. 2015 were obtained and re-analyzed [10]. OTU relative abundances were summarized at the genus level and $\log_{10}$ transformed. Chemical concentrations were converted to mM and $\log_{10}$ transformed where appropriate. Genus presence/absence was determined by examining the overall distribution of $\log_{10}$ transformed genus relative abundances and setting a threshold. Genera with a $\log_{10}$ abundance > -6 (e.g., relative abundance $> 10^{-6}$) were considered present. This threshold separated the peaks of the bi-model $\log_{10}$ genus relative abundance distribution. Statistical analyses, including linear regression, and significance testing, were performed using tools from the python Scikit-Learn, SciPy, and NumPy packages [84–86].

3.4.3 *Isolate and SynCom and metabolic models.*

Metabolic models were constructed using the KBase suit of genome analysis tools [134] and the genome sequences of *Rhodanobacter sp.* FW510-R12 and *Acidovorax sp.* GW101-3H11. Isolate genome sequences were annotated using RAST [135] and genome scale metabolic models were built using KBase's ModelSEED tool [136]. Gap filling was performed with a representation of the NO_3^- and acetate based minimal media used to grow the isolates and SynComs. SynCom growth simulations were performed using KBase's Flux balance analysis (FBA) tool, an equal proportion of each isolate, and the NO_3^- and acetate based minimal media and exchange of denitrification pathway intermediates. Amino acid exchange simulations were performed by allowing the exchange (e.g., export into media) of individual amino acids and carbon sources.

3.4.4 *Strains and medium preparation.*

R12 and 3H11 were both isolated from contaminated wells at the OR-IFRC as reported in Hemme et al. 2016 and Price et al. 2018 respectively [74, 137]. NO_3 reduction growth studies were performed in Balch tubes (10-mL culture volume) at 30°C in a defined minimal medium as described previously containing 20 mM sodium acetate and 10 mM sodium NO_3 at pH 7.2 with a 80:20 N_2-CO_2 headspace [99]. A small amount of yeast extract (0.1 g/L) was added to the medium before autoclaving to support the growth of 3H11 and R12. For experiments that characterized growth on variable amounts of NO_3, NO_2^- , and N_2O media was first prepared in serum bottles. Vitamin and phosphate solutions were added to serum bottle media and then aliquoted into N_2-CO_2 flushed and pressurized Balch tubes (10 mL each, ~1 psi overpressure). For experiments characterizing growth of 3H11 on N_2O, injections of pure N_2O into Balch tubes were performed to achieve desired concentrations.

3.4.5 *Preparation of R12 and 3H11 co-cultures and monocultures.*

From freezer stocks, cells were plated on R2A agar and incubated at 30°C. Following the growth of isolated colonies, colonies were picked and used for inoculation of R2A liquid medium cultures. Liquid R2A cultures were incubated at 30°C with shaking at 100 rpm. Following overnight growth, optical density at 600 nm (OD600) was measured and cell density was normalized to using R2A liquid medium if necessary. With starting cultures at ~0.5 OD600 units, 100 uL of R12 and 100 uL of 3H11 were inoculated into a NO_3 reduction medium to establish co-cultures. For pure cultures, 200 uL of either strain was utilized to normalize starting cell densities. Initial cell densities were normalized to ∼0.01 OD600 units across all experiments and conditions. Biological replicates varied from 3-8 across experiments. For all experiments cell concentration was monitored in Balch tubes with periodic measurements of the optical density. Initial experiments (Figure. 3.1E, 3.2C, and 3.3A) measured optical density 1-3 times daily using a Spectronic 200 spectrophotometer (Thermo Fisher). Later experiments (Figures 3.2A, 3.4, and 3.5) leverage an automated OD600 measurement with shaking at 100 rpm previously described [138]. In this context, measurements were taken every 5-30 minutes and instrument voltage was converted to OD600 by calibration. Sample OD600 was measured before and after each experiment to calibrate individual sample voltage ranges.

3.4.6 *Measurement of nitrogen species.*

NO_3, NO_2^-, and acetate concentrations were analyzed using the ion-chromatography Dionex ICS-5000 system with the IonPac® ICE-AS6 column (Thermo Fisher). Medium N_2O concentration was quantified by sampling 200 uL medium into GC vials containing 200 uL 8% PFA in 1x PBS (to achieve a final concentration of 4% PFA). Vials were allowed to equilibrate and headspace N_2O was measured using a Shimadzu GC-2014 with an AOC-6000 autosampler. Vials were equilibrated for 60 seconds in a 35 °C oven with shaking, and then a 1000 μL sample was aspirated from the headspace of the vial with a syringe preheated to 100 °C. Run conditions were as follows: inlet temperature 325 °C, oven temperature 80 °C, column flow rate 20 mL/min, TCD temperature: 150 °C, current: 80 mA, ECD temperature: 325 °C, current: 2 nA, run time: 20 minutes, carrier gas: argon, makeup gas: nitrogen. GC vial headspace N_2O ppm values were converted to culture vessel liquid N_2O concentration (mM) using Henry's law. Total culture vessel N_2O concentration reported was obtained using mass balance and assuming equilibrium.

3.4.7 *Monoculture and SynCom kinetic modesl.*

Kinetic models were developed using a modified Monod framework by integrating Logistic representation of carrying capacity into equations describing growth kinetics as a function of metabolite concentrations [113]. Equations in the model were of form $\frac{1}{X}\frac{dX}{dt} = \mu_m\left(\frac{S}{S+K_m}\right)\left(1 - \frac{X}{X_m}\right)$, where X is the cell population size, S is the substrate concentration, K_m and μ_m are substrate specific Monod kinetic parameters, and X_m is cell specific capacity. Substrate specific rates were assumed to be additive (e.g., $\mu_{R12} = \mu_{NO3} + \mu_{NO2} + \mu_{NO}$). Equations were parameterized using maximum growth rates as a function of substrate, half velocity constants, and carrying capacities extracted from species specific growth data using Logistic and Monod fits. To represent carrying capacity as a function of substrate concentration, linear models were developed to capture relationships between NO_3^-, NO_2^-, N_2O concentration, and maximum OD600. Carrying capacity models were of the form $OD_{R12} = \beta_{NO3}X_{NO3} + \beta_{NO2}X_{NO2}$ and $OD_{3H11} = \beta_{NO3}X_{NO3} + \beta_{NO2}X_{NO2} + \beta_{N2O}X_{N2O}$ for R12 and 3H11 respectively. Linear model parameters were estimated by OLS fit of monoculture maximum OD600 and initial substrate concentration data. Kinetic model simulations were performed using the Python package Tellurium [139]. Statistical analyses, including linear regression, and significance testing, were performed using tools from the Python Scikit-Learn, SciPy, and NumPy packages [84–86].

3.4.8 *Transcriptomics profiling and analysis.*

3.4.8.1 Sample collection and sequencing.

R12-3H11 co-cultures and pure cultures were harvested for transcriptome profiling in biological triplicates across time points (Figure 3.3). Cell pellets were harvested by centrifugation at 4,000 × g and flash frozen in liquid N_2. Total RNA was extracted using the MasterPure™ Complete DNA and RNA Purification Kit from Epicentre. All samples were treated with Invitrogens TURBO DNA-Free kit to remove DNA contamination. The Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus (Illumina) was used for rRNA depletion and library preparation. Sequencing was performed using the NextSeq 500 platform (2 by 75 bp, Illumina) with 10 to 15 million reads per sample.

3.4.8.2 Read processing.

RNA sequencing reads were analyzed with FastQC according to Illumina's default quality filtering process and then trimmed using base quality scores by Trimmomatic [76, 77]. A quality score of 20 was used for read trimming and quality filtering. Reads were aligned to the concatenated genomes of *Rhodanobacter sp. FW510-R12* (NCBI BioProject PRJNA255897) and *Acidovorax sp. GW101-3H11* (NCBI BioProject PRJNA314893) using Spliced Transcripts Alignment to a Reference (STAR) [78] followed by a tabulation of transcript abundances by HTSeq-count [79].

3.4.8.3 Differential expression analysis, clustering, and functional enrichment.

Normalized expression data and differentially expressed genes (DEGs) were obtained from DESeq2 [44] using pairwise comparisons of growth phase matched monoculture and SynCom transcriptomes. Genes were reported as dysregulated if the $\log_2$ fold change magnitude ≥ 2 and significant ($p < 0.05$ and Benjamini-Hochberg false discovery rate $q < 0.01$). DEGs were clustered using the scikit-learn implementation of the k-means algorithm [84]. Statistical analyses, including principal component analysis (PCA), significance testing, and functional enrichment were performed using tools from the Python Scikit-Learn, SciPy, and NumPy packages [84–86]. The significance of KEGG subpathway term enrichment among k-means clusters was assessed by comparing term frequencies within each cluster to their frequencies in the genome using the hypergeometric test [140]. Significantly enriched terms ($p < 0.05$ and Benjamini-Hochberg false-discovery rate $q < 0.01$) were reported [87]. KEGG pathway level expression comparisons were assessed using DESeq2 normalized gene expression Z-scores and Welch's t-test.

3.4.9 *Proteomics profiling and analysis.*

3.4.9.1 *Sample collection and peptide quantification.*

R12-3H11 co-cultures and pure cultures were harvested before inoculation and at the end of growth (stationary phase) for proteomics profiling using 8 biological replicates (Figure 3.4). Cell pellets were harvested by centrifugation at 4,000 × g and flash frozen in liquid N_2. Samples were shipped to collaborators who quantified peptide abundances in each sample.

3.4.9.2 *Species relative abundance estimates.*

Isolate peptide profiles were used to identify peptides unique to each isolate. Isolate specific peptide peak areas were used to estimate isolate relative abundances across samples. Inoculum samples with defined mixtures of each isolate (based on OD600) were used to validate proteomics based relative abundance estimates.

3.4.9.3 *Statistical analyses.*

Statistical analyses, including PCA, linear regression, and significance testing, were performed using tools from the Python Scikit-Learn, SciPy, and NumPy packages [84–86]. The significance of individual protein abundance changes was assessed by correlating protein abundance with initial NO_3^- levels, inoculum 3H11 percentage, and steady state N_2O levels. Pearson r values were used to identify significant hits, proteins with $r \leq 0.5$ or $r \geq 0.5$ and $q < 0.05$ (Benjamini-Hochberg false-discovery rate correction) were considered [87]. KEGG subpathway level protein abundance correlations with steady state N_2O levels were performed using data randomly split into training (70% of samples) and testing (30% samples) sets. KEGG subpathway terms were used to identify groups of proteins and multi-linear regression was performed with the training set. Model accuracy was then assessed using the test set and out of sample R^2 was reported.

Chapter 4. PERSONALIZED *CLOSTRIDIOIDES DIFFICILE* ENGRAFTMENT RISK PREDICTION AND PROBIOTIC THERAPY ASSESSMENT IN THE HUMAN GUT

4.1 INTRODUCTION

The human gut microbiome plays important roles in shaping host metabolism, in the development of chronic diseases, and in preventing opportunistic pathogen colonization and infection [9, 141, 142]. The metabolic versatility of gut bacteria allows for the stable coexistence of hundreds of commensal species within the gastrointestinal tract [7]. Some species extract energy and nutrition directly from indigestible dietary substrates, like plant fibers or recalcitrant proteins, while others subsist largely on host-derived mucosal glycans or the vast array of metabolic byproducts produced by primary fiber, protein, and mucus degraders [143, 144]. Saturation of these metabolic niches by commensal microbes can prevent colonization and engraftment by external microbes that may share a similar niche, including pathobionts [145–147].

Perturbations to the gut microbiome (e.g., antibiotic use or diarrheal events) provide a window of opportunity for pathobiont colonization [148], which could in turn lead to the development of disease following subsequent perturbations [149, 150]. One such pathobiont, *Clostridioides difficile*, is the most common hospital acquired gastrointestinal infection in the U.S. [151, 152]. *C. difficile* colonizes as much as 30-40% of community-dwelling adults without causing disease, lying in wait until the opportunity for infection arises [153, 154]. During active *C. difficile* infection (CDI), antibiotic treatment can be effective in suppressing *C. difficile* growth, but antibiotics also disrupt the ecology of the commensal microbiota and potentiate reinfection if

C. difficile is not completely cleared by the treatment [149, 150]. Thus, an intact gut microbiota that prevents *C. difficile* colonization and engraftment is critical to the host's defense against CDIs [142]. This understanding has led to the widespread use of fecal microbiota transplants (FMTs) as a means of combating cases of recurrent CDI (rCDI), where antibiotic treatment proves insufficient [155]. While the biology of *C. difficile* has been fairly well-characterized in the context of disease, the pre-disease mechanisms of *C. difficile* colonization and engraftment are still poorly understood, as are the factors that govern *C. difficile* decolonization and FMT efficacy [148].

There are currently no mechanistically grounded, generalizable approaches to accurately predicting the engraftment of an exogenous bacterial taxon in the context of a given microbiota. Previous work has leveraged machine learning (ML) to predict the engraftment of FMT donor strains in FMT recipients [156]. While effective and relatively accurate, this kind of quasi-black-box ML approach does not provide a means of understanding the molecular mechanisms that facilitate or prevent engraftment. Here, we present an alternative approach to engraftment prediction that leverages microbial community-scale metabolic models (MCMMs), which provide detailed, mechanistic information on the ecological interactions within individual microbiota that prevent or facilitate engraftment, in addition to generating accurate engraftment predictions.

Genome-scale metabolic models and classical flux balance analysis (FBA) have been invaluable tools for exploring how environmental conditions impact the metabolic capacities of individual bacterial taxa grown *in vitro* [157]. However, extending these methods to complex, multi-species communities has proved to be a challenge. Recently, we developed a community-scale metabolic modeling platform called MICOM, which leverages microbiome compositional and dietary constraints, along with a cooperative-tradeoff flux balance analysis (ctFBA) approach that balances computational speed and biological accuracy, to rapidly estimate steady-state community-scale metabolic fluxes [25, 158]. Here, we use publicly available 16S amplicon and shotgun metagenomic data from *in vitro* and *in vivo* studies with both known and unknown *C. difficile* colonization dynamics, along with our MICOM platform [25], to build and test MCMMs to estimate *C. difficile* engraftment potential within a given microbiome and dietary/environmental context. We present novel insights into how *C. difficile* can occupy three discrete metabolic niches across commensal communities, what metabolic interactions within gut bacterial communities promote or prevent colonization, and we show how we can predict potential responders and non-responders to a defined probiotic cocktail that has recently shown efficacy in the treatment of rCDI [159]. Overall, MCMMs provide a novel path toward predicting *C. difficile* engraftment risk. Furthermore, these models can be leveraged to design precision dietary or probiotic interventions aimed at decolonizing individuals who are already carrying *C. difficile* and preventing engraftment in those who are not. Finally, we suggest that MCMMs could enable precision engineering of the gut microbiome through personalized engraftment predictions for other pathobionts beyond *C. difficile*, probiotic bacterial strains, or for entire microbial consortia (e.g., FMTs from different donors), in the context of a specific diet.

4.2 RESULTS

4.2.1 *Development of an in silico invasion assay to simulate C. difficile colonization.*

To simulate the colonization of *C. difficile* we developed an *in silico* invasion assay that leverages microbiome relative abundance data, manually curated genome-scale metabolic models of gut bacteria from the AGORA database, constraints on the diet/environment, and the MICOM modeling framework (Figure 4.1A) [25, 160]. Here, we used several existing 16S amplicon and shotgun metagenomic sequencing data sets to validate our approach [148, 156, 161–164]. Amplicon sequencing data is often limited to genus-level resolution in the taxonomic classifications of amplicon sequence variants (ASVs). Even with metagenomic sequencing data, species level resolution can be poor when mapping to available species-level metabolic models. Generally, only the most abundant or prevalent commensal bacterial species are well represented. Therefore, we focused on genus-level MCMMs, to approximate the community metabolic context, for our invasion assays (see Methods) [165]. However, we also evaluated the use of species-level MCMMs using *in vitro* 16S amplicon and stool shotgun metagenomic data. Briefly, strain-level metabolic models from AGORA were combined at the genus or species level, to account for the potential coexistence of multiple strains from a given genus or species within an individual and to reduce potential bias from arbitrarily selecting individual strain models. Using this approach ~75% of reads, on average, could be mapped to a genus-level metabolic model within the AGORA database from 16S amplicon data, while ~90% of reads, on average, could be mapped using shotgun metagenomic data. At the species level, only ~50% of reads could be mapped, on average, using the same *in vivo* metagenomics data, while 100% of reads could be mapped for *in vitro* community profiles using 16S amplicon sequencing (i.e., because type strains were used in these experiments), illustrating the inherent limitation of using species level models for *in vivo* applications and further justifying our focus on genus-level analyses. To simulate the invasion of *C. difficile* into these model communities, a pan-species model of *Clostridioides*, representing all four common *C. difficile* strains present in the AGORA database (including hypervirulent and non-epidemic strains), was introduced at a relative abundance of 10% (see below for justification for this percentage), while other community relative abundances were decreased proportionally to approximate a minor perturbation in community-wide biomass (Figure 4.1A). Growth simulations were then performed using a medium representing an average European diet (i.e., a standard developed-world diet appropriate to the cohorts studied here), with fluxes of metabolites known to be absorbed in the small intestine decreased by 90%, as previously described for *in vivo* samples [25]. To represent the anaerobic basal broth (ABB) used to culture the *in vitro* communities, a curated medium was constructed using a previously described representation of Luria-Bertani (LB) broth as a template and the known composition of ABB (see Methods for further details) [164, 166]. Growth rates were estimated using ctFBA, as implemented in MICOM, which uses a regularization step and allows for a suboptimal community growth rate to achieve a more realistic growth rate distribution across the community [25, 158]. Import and export fluxes were estimated using parsimonious enzyme usage FBA (pFBA) [25].

Personalized MCMMs were constructed for each sample and the potential for *C. difficile* engraftment was quantified as the model-inferred growth rate. ctFBA has a single free parameter that needs to be chosen, the tradeoff between community-wide growth rates and individual, taxon-specific growth rates. Assuming that most taxa detected at appreciable abundances in a gut microbiome are actively growing *in vivo* and *in vitro*, a trade-off value was selected by choosing the minimal deviation from optimal community growth for which >70% of genera obtained non-zero growth rates on average. We found that with a trade-off value of 0.8 (i.e., 80% of maximal community biomass production) resulted in a mean fraction of genera with non-zero growth of >70% using genus-level models constructed from 16S amplicon data. This same non-zero growth fraction was achieved with a trade-off value of 0.9 for the genus-level models constructed from metagenomics data. The 16S and metagenomics specific tradeoff values were also used for the species level models derived from these data types, respectively. At the species level, these values were associated with much higher growth fractions (~90% and 95% for 16S amplicon and shotgun metagenomics, respectively).

We first validated our approach using an *in vitro* 16S amplicon data set where *C. difficile* was co-cultured with communities of human commensal gut bacteria at differing levels of species richness [164]. We found that MCMM-inferred *C. difficile* biomass (calculated using the MCMM-inferred *C. difficile* growth rates, assuming exponential growth over a fixed interval, see Methods) accurately reflected a negative trend in measured *C. difficile* abundance along a gradient of increasing species richness and was significantly correlated with empirically measured *C. difficile* biomass after 48 hours of growth (Figure 4.1B; ordinary least squares (OLS) $R^2 = 0.45$, $p < 10^{-6}$). Similarly, we found that the normalized *C. difficile* growth rate (normalized to the overall MCMM-inferred community growth rate, see Methods) decreased as a function of community richness and was an even stronger predictor of the empirically-observed *C. difficile* biomass after 48 hours of growth (Figure 4.1C; ordinary least squares (OLS) $R^2 = 0.5$, $p < 10^{-6}$). Biclustering of the inferred species growth rates suggested that growth of *C. difficile* was largely suppressed by *B. thetaiotaomicron, P. vulgatus*, and *E. lenta in vitro*, as indicated by the strong negative associations between the growth rates of these species and the growth rate of *C. difficile* (Pearson r <-0.45, $p <$ 0.01 for species). Positive associations between the growth rates of *C. difficile, C. scindens*, and *P. hiranonis* (Pearson $r > 0.8$, $p < 10^{-6}$ for all species) were in line with the observation that these species tend to share overlapping metabolic niches. The previously-observed suppressive effects of *C. scindens* and *P. hiranonis* on *C. difficile* were not apparent within our modeling framework [164], likely because pH and its effects on growth are not easily captured by FBA models. However, while *in vitro* acidification is a strong environmental force in anaerobic batch culture, we do not expect variation in pH to be as prominent in the buffered environment of the human colon.

To further validate our approach *in vivo* we leveraged a time series with a known *C. difficile* colonization event (Figure 4.1E) [148, 167]. We found that estimated *C. difficile* growth rates were at or below the limit of solver accuracy ($< 10^{-6}$, which effectively indicates an absence of growth) in samples collected before colonization and comparable to growth rates of other dominant genera in samples taken after the initial colonization event (Welch's t-test t = -3.29, p = 0.003, for comparison of $\log_{10}$ growth rates before and after the known colonization event; Figure 4.1E). Furthermore, we saw patchy engraftment predictions in a second individual that was known to be colonized by *C. difficile* at a low level (i.e., near the limit of detection) throughout a time series (Figure 4.1F). We also assessed the importance of propagule pressure [168] (i.e., the relative abundance at which the invasive taxon is introduced into the models) and found that below 10% relative abundance, agreement between growth rate estimates and measured abundances was poor. Thus, propagule pressure plays an important role in predicted engraftment success [164].

Based on these initial results, we decided to use a fixed tradeoff value of 0.8 and a *C. difficile* invasion fraction of 10% for all subsequent analyses of 16S amplicon data (i.e., the highest value at which >70% of the community showed a positive growth rate). Using the same heuristic, a tradeoff value of 0.9 was chosen for the metagenomics analyses, along with a 10% *C. difficile* invasion fraction

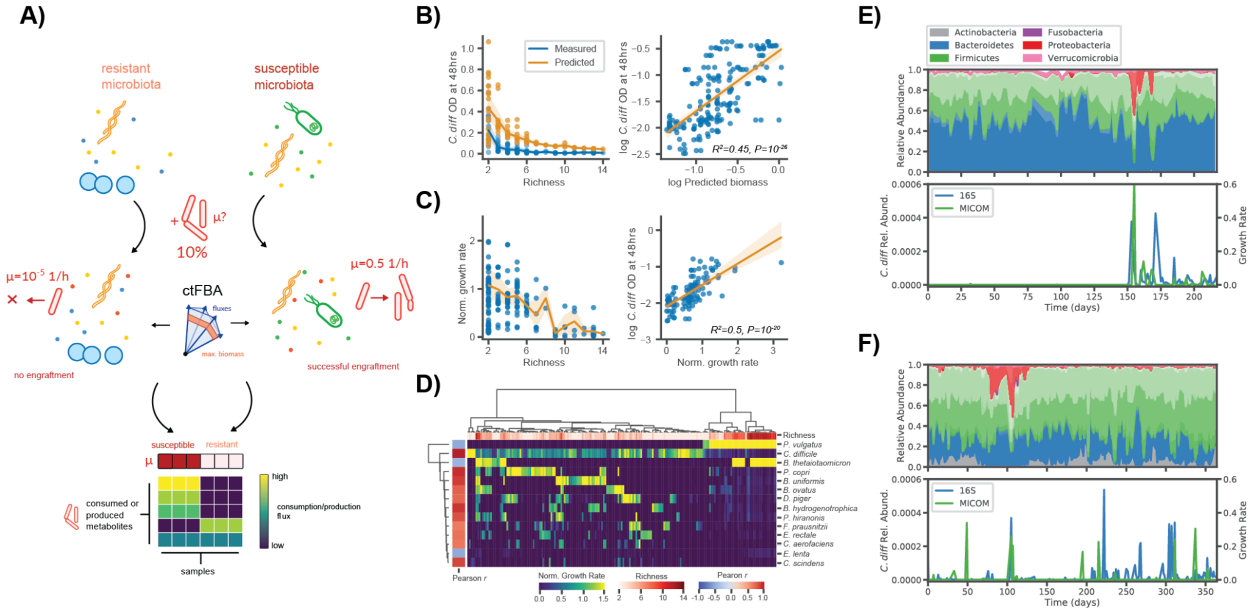

Figure 4.1. *In silico* invasion assay accurately predicts *C. difficile* engraftment *in vitro* and *in vivo*. (**A**) Schematic illustrating the *in silico* invasion assay workflow leveraged in this study. Personalized microbial community-scale metabolic models (MCMMs) are supplemented with 10% of a pan-genus *Clostridioides* model to simulate an invasion event and ctFBA was used to predict *C. difficile* engraftment and metabolic fluxes. (**B**) Measured *C. difficile* biomass at 48 hours from Hromada et al. and predicted biomass, computed from MCMM-predicted *C. difficile* growth rate and an exponential growth model, across a gradient of in vitro community richness. Mean trend lines and standard deviation are displayed for both measured (blue) and predicted (yellow) biomass. Relationship between $\log_{10}$ predicted biomass and $\log_{10}$ biomass. Ordinary least squares fit and 95% confidence interval are displayed, as well as regression R^2 and *p*-value (**C**) Normalized *C. difficile* growth rate along a community richness gradient and the relationship between normalized *C. difficile* growth rate and measured $\log_{10}$ biomass. Mean trend line and standard deviation are displayed for the relationship between normalized *C. difficile* growth rate and community richness. Ordinary least squares fit and 95% confidence interval are displayed, as well as regression R^2 and *p*-value, for the relationship between normalized *C. difficile* growth rate and measured $\log_{10}$ biomass. (**D**) Biclustered normalized growth rates for species across conditions. Species normalized growth rate is displayed using blue-to-yellow heatmap coloring for each sample. Sample community richness is displayed using white-to-red heatmap coloring on the top row of the plot. Pearson correlation coefficients between *C. difficile* normalized growth rate and the normalized growth rate of each other species are displayed using blue-to-red heatmap coloring

in the leftmost column of the plot. Colorbars at the bottom of the blot indicate the scales for the various measures. (**E**) Donor A time series taken from David et al. displaying daily fluctuations in microbiome composition over several months. Composition is displayed, colored by phylum-level annotations (different shading indicates taxonomic families). At day 150, Donor A experienced a diarrheal event and was subsequently colonized by *C. difficile*. Estimates of *C. difficile* relative abundance from 16S sequencing and MICOM-predicted *C. difficile* growth rates are displayed. (**F**) Time series from Donor B from the same study, who was colonized by *C. difficile* (at very low relative abundances, near the limit of detection) throughout the sampling period.

4.2.2 *In silico invasion assay accurately predicts C. difficile colonization potential in rCDI patients pre- and post- FMT.*

We applied our *in silico* invasion model to two separate datasets (16S amplicon and metagenomic shotgun sequencing, respectively) of rCDI patients who received FMTs and were subsequently followed over time [156, 163]. These data provided an additional test of MCMM performance and a means to explore the metabolic features associated with community-scale colonization susceptibility or resistance across a larger population. Given that all individuals in the rCDI cohorts had experienced multiple recurrent infections, we expected pre-FMT microbiomes from these patients to be more susceptible to invasion. Additionally, both of the original studies found that nearly all the patient microbiomes transitioned to a compositional state that was much closer to the healthy controls post-FMT than to their pre-FMT compositional states [156, 163]. Thus, we expected post-FMT samples would be less susceptible to invasion. For the sake of comparison of results across studies, genus-level MCMMs were used for both analyses. Across both data sets, rCDI patients had significantly higher MCMM-predicted *C. difficile* growth rates before FMT treatment than they did post-FMT treatment (Figure 4.2A; Welch’s t-test $t = -3.19$, $p = 0.001$ for comparison of pre-FMT vs. post-FMT). Predicted *C. difficile* growth rates were negatively associated with Shannon diversity, albeit weakly (Figure 4.2B; OLS $R^2 = 0.05$, $p = 0.01$), which is in line with prior empirical observations indicating that lower diversity communities are more susceptible to *C. difficile* colonization and the development of rCDI [147, 164, 169, 170].

The community-scale import flux profile before the *in silico* invasion was predictive of *C. difficile* growth rate following the invasion in Weingarden *et al.* (Figure 4.2C). High-dimensional community-scale import flux profiles were projected into a two-dimensional space using the Uniform Manifold Approximation and Projection (UMAP) technique (Figure 4.2C) [171]. The UMAP projection provides a visual means of identifying patterns in the high dimensional import flux space. The closer points are to one another in this ordination the more similar their import flux profiles are. Thus, clusters of points in the UMAP can represent distinct metabolic environments across samples. The ordination plot indicated that *C. difficile* appears to grow well in more than one metabolic environment when colonizing different individuals. Indeed, we saw that the predicted metabolic environments occupied by *C. difficile* could vary within an individual over time (Figure 4.2C). For most patients in the Weingarden *et al.* cohort, there was a transition from colonization-susceptibility pre-FMT to colonization-resistance post-FMT (Figure 4.2A, C). We next examined the different metabolic niches that *C. difficile* was able to exploit when colonizing individuals across both rCDI-FMT cohorts, to better understand this phenotypic plasticity.

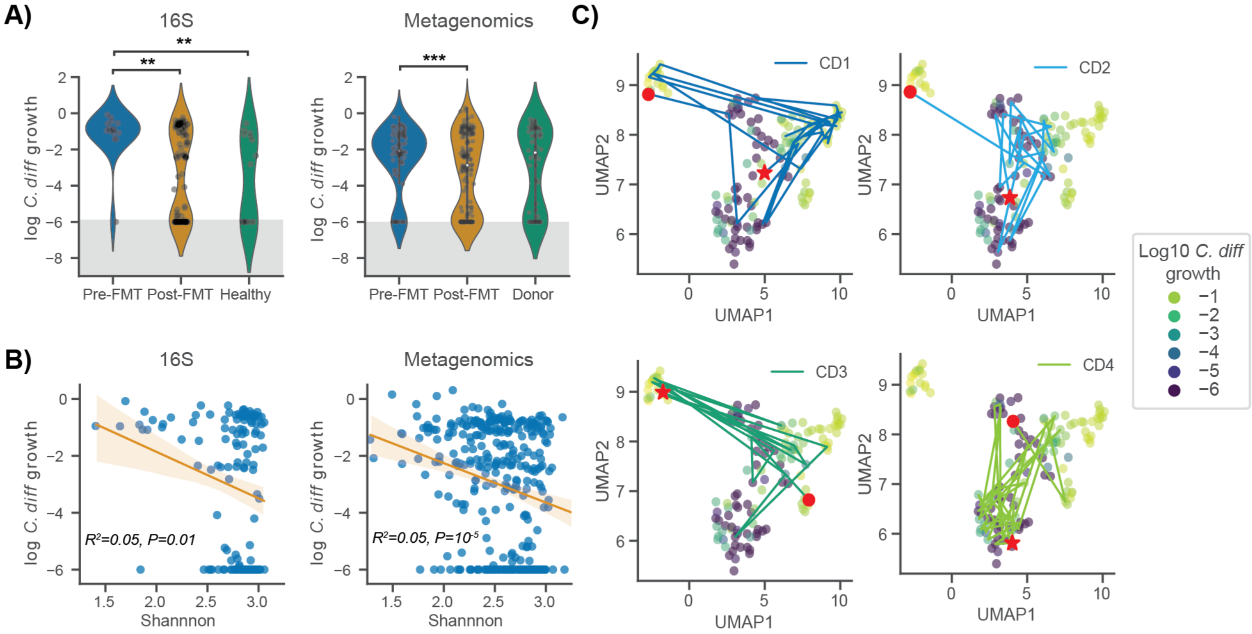

Figure 4.2. ***C. difficile*** **growth rate predictions capture importance of community context for patient recovery from CDI.** (**A**) Violin plots displaying predicted *C. difficile* $\log_{10}$ growth rate distributions across patient disease status categories. Data from Weingarden et al. and Ianiro et al. were leveraged to validate the approach using both 16S amplicon and metagenomic sequencing. Gray shading indicates the numerical accuracy of the solver (values below 10^{-6} cannot be distinguished from zero and are considered negligible). Bars indicate comparisons for which differences were significant using Welch's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (**B**) Relationship between predicted *C. difficile* $\log_{10}$ growth rate and Shannon diversity across datasets. Ordinary least squares fit and 95% confidence interval are displayed, as well as regression R^2 and p-value. (**C**) Two-dimensional representation of community import fluxes before in silico invasion using UMAP colored by $\log_{10}$ growth rates of *C. difficile* following in silico invasion. Patient trajectories are displayed, each with a red circle representing the patient's starting point (before FMT), and a red star representing the patient's endpoint (post-recovery).

4.2.3 *C. difficile is predicted to occupy three distinct metabolic niches within the human gut microbiome.*

To characterize *C. difficile* colonization-associated niches and identify the potential for multiple metabolic strategies associated with its growth, we examined *C. difficile* import fluxes with high variance ($\log_{10}$ flux variance ≥ 4.5) across the rCDI-FMT cohorts. Using this criterion the high variance metabolites identified across the 16S and metagenomic data sets were highly consistent (Figure 4.3). We identified 31 and 44 high-variance metabolite fluxes from the 16S and metagenomics analyses, respectively. Of these, 27 were shared, 4 were unique to the Weingarden *et al.* (16S) analysis, and 17 were unique to Iraniro *et al.* (metagenomics) analysis. Biclustering of the high variance import flux data and an examination of how the apparent clusters were associated with growth rates revealed that *C. difficile* makes use of multiple metabolic strategies (Figure 4.3). Three major clusters were observed across patient samples and cohorts. We designate these three clusters as “high growth”, “moderate growth” and “no growth” (Figure 4.3). The high growth cluster in both cohorts included many of the pre-FMT samples and was characterized by consistently high import fluxes for all the metabolites identified as most strongly coupled to *C. difficile* growth across all models. The moderate growth cluster showed a sparser metabolite consumption profile. For example, ornithine and fructose were rapidly consumed in the high-growth cluster but showed almost no consumption in the moderate-growth cluster (Figure 4.3). As expected, very few metabolites were consumed by *C. difficile* above the zero threshold of 10^{-6} in the no-growth cluster (Figure 4.3).

The metabolic strategies employed by *C. difficile* within the MCMMs showed convergence with several observations from the literature. For example, we found that metabolites known to promote the growth of *C. difficile in vivo* (e.g., succinate, ornithine, and trehalose) were preferentially utilized when available and were associated with higher pathobiont growth rates [172–174]. In addition, the consumption of amino acids valine, leucine, glycine, glutamate, glutamine, and proline were associated with higher *C. difficile* growth rates in the MCMMs, implying that *C. difficile* employs Strickland fermentation in one of its growth modes, which has been observed empirically [175].

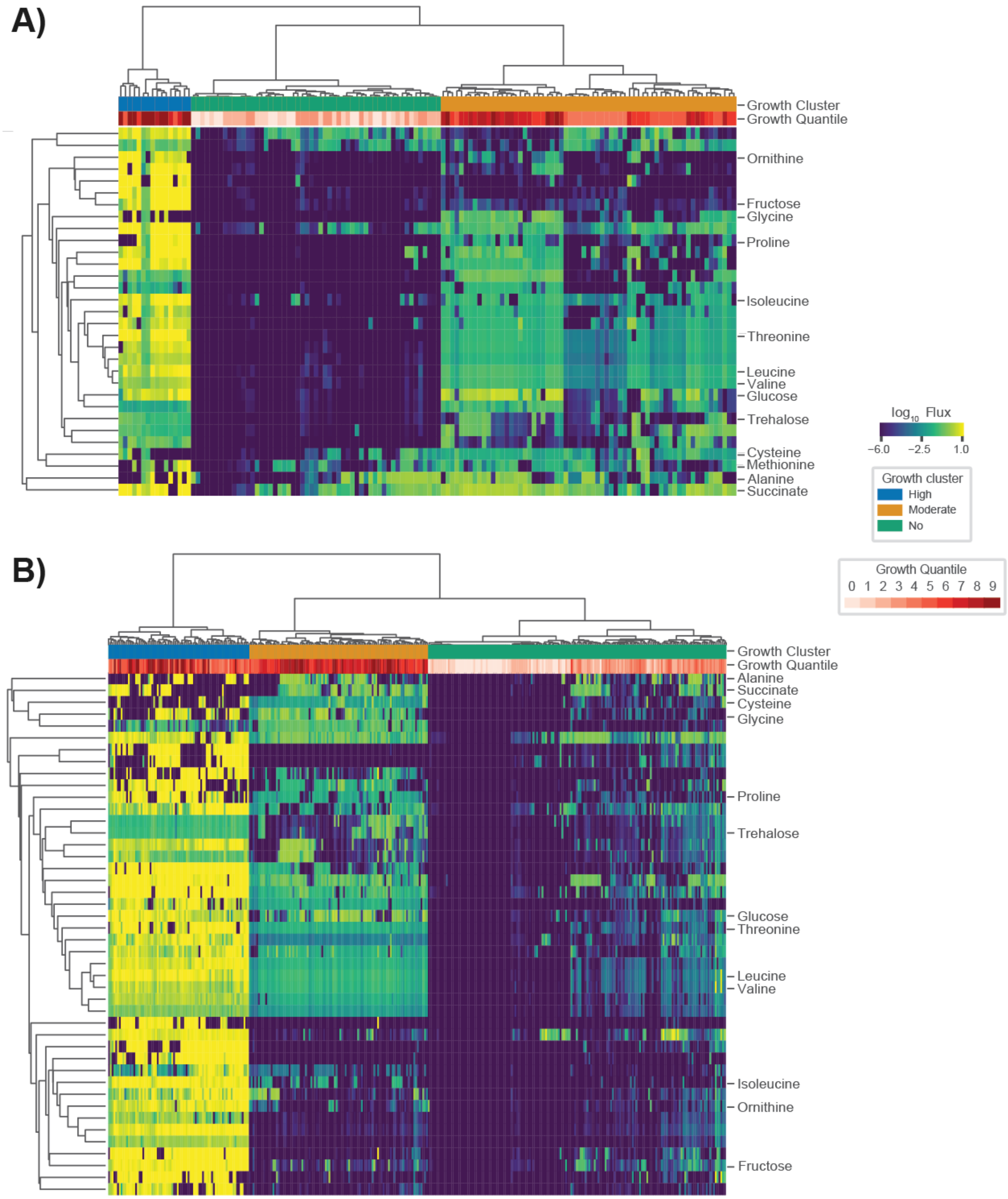

Figure 4.3. *C. difficile* occupies multiple metabolic niches across communities. (**A**) Violin plots displaying predicted *C. difficile* $\log_{10}$ growth rate distributions across patient disease. Biclustered *C. difficile* $\log_{10}$ import fluxes from Weingarden et al., where each row is the import flux of a particular metabolite and each column is a patient sample. (**B**) Biclustered *C. difficile* $\log_{10}$ import fluxes from Ianiro et al., where each row is the import flux of a particular metabolite and each

column is a patient sample. In each heatmap, imports for which the log variance across samples was ≥ 4.5 are displayed. Fluxes across samples are displayed using blue-to-yellow heatmap coloring. *C. difficile* $\log_{10}$ growth rate quantiles are displayed in white-to-red heatmap coloring for each patient sample in the top row of each plot. Additionally, three coarse grain growth clusters are noted. These growth clusters represent “high growth”, “moderate growth”, and “no growth” phenotypes.

Following up on these findings we examined how cooperative and competitive interactions within MCMMs contributed to *C. difficile* colonization. To accomplish this, we examined the import and export fluxes of metabolites associated with *C. difficile* colonization (e.g., amino acids, ornithine, succinate, etc.). Genera that produced metabolites consumed by *C. difficile* likely promote its growth, while those consuming *C. difficile* growth-associated metabolites may be in direct competition. For ornithine and succinate, we found that cooperative and competitive interactions are context-dependent, varying across samples. The genus *Phocaeicola*, for instance, produces ornithine in some samples, which is in turn consumed by *C. difficile*, while in other contexts it consumes ornithine, competing with *C. difficile* (Figure 4.4A). Meanwhile, *Roseburia*, and *Faecalibacterium* compete with *C. difficile* for ornithine, but these genera also produce succinate and leucine in some contexts, which *C. difficile* consumes (Figure 4.4A-B). Thus, the overall community context, rather than the presence or absence of any single taxon, appears to be the most important factor in determining the metabolic strategies used by *C. difficile* and can lead to emergent competitive or cooperative interactions, which can either hinder or promote colonization.

Finally, we assessed whether or not compositional variation in the microbiome alone could explain observed differences in MCMM-predicted *C. difficile* growth rates. We found that compositional variation was a modest predictor of estimated *C. difficile* growth rate (out-of-sample least absolute shrinkage and selection operator (LASSO) regression $R^2 = 0.37$ for the Weingarden *et al.* rCDI-FMT cohort, see Methods). Meanwhile, the import flux derived growth clusters (e.g., “high growth”, “medium growth”, and “no growth” groups) could explain the vast majority of the variance in predicted *C. difficile* growth rates (analysis of variance (ANOVA) $R^2 = 0.94$ for the CDI-FMT cohort). However, this comparison between model-estimated growth rates and growth clusters is a bit circular (i.e., both are derived from the model), so as mentioned above, we also observed that MCMM-estimated *C. difficile* growth rates were able to explain 50% of the variance in empirical *C. difficile* biomass measurements *in vitro* (Figure 4.1C). In summary, these results suggest that compositional information alone is not sufficient for consistently accurate engraftment predictions.

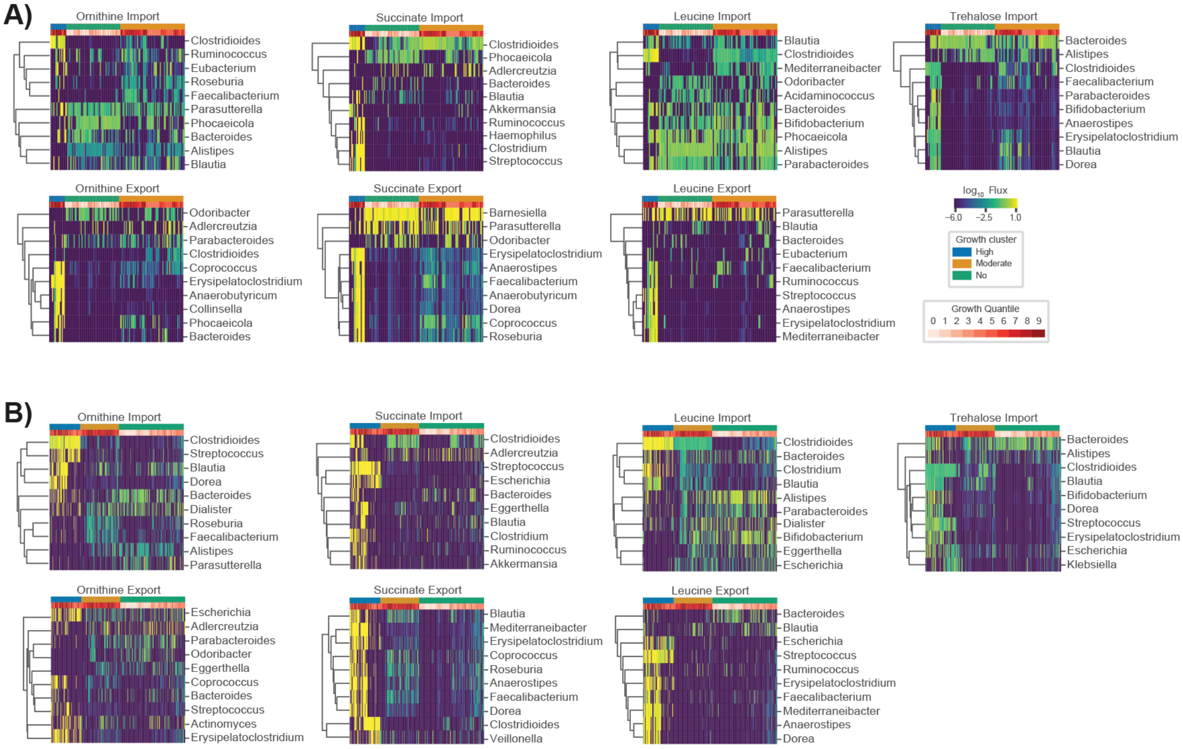

Figure 4.4. Import and export fluxes of key metabolites across communities highlight the context-dependency of key *C. difficile* competitors and cooperators. (A) Biclustered community import and export fluxes of specific metabolites associated with *C. difficile*

colonization from Weingarden et al., where each row is a genus and each column is a patient sample. (**B**) Biclustered community import and export fluxes of specific metabolites associated with *C. difficile* colonization from Ianiro et al., where each row is a genus and each column is a patient sample. For each heatmap, the top 10 genera with the highest mean import or export flux across samples are displayed. Fluxes across samples are displayed using blue-to-yellow heatmap coloring. *C. difficile* $\log_{10}$ growth rate quantiles are displayed in white-to-red heatmap coloring for each patient sample in the top row of each plot. Additionally, three coarse grain growth clusters are noted. These growth clusters represent “high growth”, “moderate growth”, and “no growth” phenotypes.

4.2.4 *Associations with C. difficile growth provide insights into the role of community context.*

To assess the consistency of the *C. difficile* growth clusters, we leveraged four independent *in vivo* 16S amplicon data sets, including the time series and the Weingarden *et al.* rCDI-FMT studies presented above (Figures. 4.1-4.4), along with two large cross-sectional cohorts (i.e., the American Gut and Arivale cohorts), covering a total of 14,862 individuals [161, 167, 176, 177]. We evaluated growth and flux predictions generated across all four data sets and found that *C. difficile* fell into the same three clusters that were identified in the rCDI-FMT data sets, representing no growth, moderate growth, and high growth (Figure 4.5A).

To further contextualize the metabolic niche of *C. difficile*, we integrated model outputs for all four 16S data sets. Specifically, we looked at import fluxes across all genera. Most genera formed unique clusters in the UMAP projection, suggesting that each genus had a single metabolic niche that was consistent across datasets (Figure 4.5B). Within this community context, we found that *C. difficile* still fell into three very distinct clusters (Figure 4.5B). Several genera that showed some of the strongest competitive and cooperative interactions with *C. difficile* across data sets, *Blautia*, *Faecalibacterium*, *Ruminococcus* and *Dorea*, clustered near one another in import flux space, with some overlap, indicating that these commensal taxa shared a similar metabolic niche (Figure 4.5B). Meanwhile, *Bacteroides* and *Phocaeicola*, two apparent *C. difficile* competitors, formed clusters largely separate from those formed by the other genera. Several of these taxa also occupied niche space near the *C. difficile* moderate growth cluster, supporting the potential for competitive interactions between *C. difficile* and these genera. The *C. difficile* high-growth cluster was more separated from other commensal genera, suggesting that this growth mode is more specialized in the context of the surrounding community (Figure 4.5B).

We next explored gut community diversity and predicted *C. difficile* growth rates across the four data sets. Specifically, we looked at Shannon diversity, which integrates species richness and evenness and is commonly used to quantify gut microbiome alpha diversity. Lower Shannon diversity is commonly associated with disease states, like diarrhea, while higher diversity has generally been associated with diverse plant-based diets and overall better health [169]. However, constipated individuals generally have higher gut microbiome alpha diversity as well, suggesting that there may be an optimal range of alpha diversity across healthy individuals [176, 178]. Our initial analysis using the rCDI-FMT cohort suggested a monotonically negative relationship between the predicted *C. difficile* growth rate and Shannon (Figure 4.2B). However, the integrated data sets, which spanned a wider range of diversity, showed a U-shaped relationship between Shannon diversity and predicted *C. difficile* growth rate (Figure 4.5C). Intermediate levels of Shannon diversity were associated with the lowest predicted growth rates, on average, with higher average growth at the upper and lower tails of the distribution (Figure 4.5C). The relationship between Shannon diversity and predicted growth rate suggests extremes in either direction on the diversity scale are, on average, more permissive to *C. difficile* engraftment.

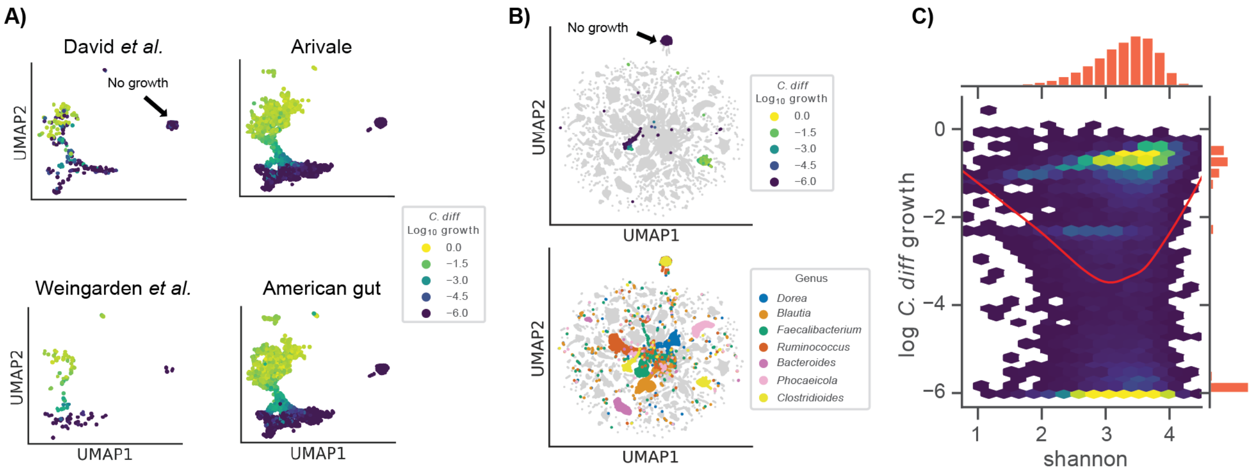

Figure 4.5. Growth niches across large four independent cohorts challenged in silico with *C. difficile.* (**A**) Two-dimensional representation of $\log_{10}$ *C. difficile* import fluxes using UMAP across four independent data sets. Colors denote *C. difficile* growth rate ranging from low (blue) to high (yellow). The position of the no-growth cluster is indicated. (**B**) Two-dimensional representation of $\log_{10}$ genus import fluxes using UMAP across all datasets. The top panel displays $\log_{10}$ *C. difficile* growth rate within the context of all other genera. Bottom panel colors *C. difficile* along with six genera of interest: *Dorea*, *Blautia*, *Faecalibacterium*, *Ruminococcus*, Bacteroides, and *Phocaeicola*. The position of the no-growth cluster is indicated. (**C**) Two-dimensional hexagonal binning of $\log_{10}$ *C. difficile* growth rate and community alpha diversity (Shannon index). The red trend line indicates a LOWESS fit to the $\log_{10}$ *C. difficile* growth rate and community Shannon diversity data.

4.2.5 *Blood metabolites and clinical labs associated with MCMM-predicted C. difficile colonization susceptibility.*

We next sought to identify potential blood-based markers that were significantly associated with MCMM-predicted *C. difficile* growth rates. Previous work has shown that circulating blood metabolites can be leveraged to predict gut microbiome alpha-diversity [176]. We identified several blood metabolites and clinical chemistries significantly associated with MCMM-predicted *C. difficile* growth rates in the Arivale cohort, after adjusting for common covariates (i.e., sex, age, and BMI) and correcting for multiple tests (FDR $q < 0.05$). These included two secondary bile acids, an unannotated metabolite previously associated with the abundance of the family *Eggerthellacea*, and several red blood cell-associated clinical chemistries [179]. Unfortunately, while significant, these blood-based markers, along with sex, age, and BMI, collectively accounted for only ~5% of the variance in MCMM-predicted *C. difficile* growth rates. Thus, it appears MCMM-based estimates of *C. difficile* engraftment, constrained by fecal microbiome data, cannot be readily replaced with commonly measured clinical chemistries or blood metabolites.

4.2.6 *MCMMs predicts engraftment heterogeneity of probiotic cocktail designed to treat rCDI.*

As a final proof-of-concept for our modeling framework, we simulated a probiotic intervention using a previously validated probiotic cocktail designed to treat rCDI [159, 180]. The probiotic, referred to as VE303, was composed of 8 commensal Clostridia strains and was shown to be effective at treating CDI in mice [180]. This probiotic was also shown to be safe, well-tolerated, and effective in reducing rCDI incidence in humans [159]. An earlier study in both mice and humans found that engraftment of VE303 strains was optimal following antibiotic pretreatment [180]. With these facts in hand, we designed a simulated intervention that mimicked the treatment found to be most effective by Dsouza *et al* [180]. We were only able to identify metabolic models for 6 of the 8 strains in VE303 in the AGORA database [160]. We leveraged the Weingarden *et al.* rCDI-FMT dataset to test this six-member probiotic cocktail, paired with *in silico* invasion by *C. difficile*. The probiotic cocktail was introduced to patient samples, alongside 10% *C. difficile*, at a total relative abundance of 50%, which was evenly distributed among the six strains. We also simulated vancomycin treatment by reducing the abundance of *C. difficile* and all commensal genera known to be impacted by vancomycin treatment [181] by 90%. We found that a combined probiotic and antibiotic intervention most effectively suppressed the growth of *C. difficile* in both the moderate and high *C. difficile* growth rate clusters (Figure 4.6A).

To better understand the mechanism of action of the probiotic cocktail, we assessed the growth characteristics and the niche proximity of the probiotic strains in relation to *C. difficile.* We found that suppression of *C. difficile* growth occurred more readily when the average growth of the probiotic strains was high ($> 10^{-4}$, Figure 4.6B) and when the average niche distance between the probiotic strains and *C. difficile* was low (< 25 import flux Euclidean distance, Figure 4.6C). We also found that relative to other genera, several of the probiotics strains occupied niches closer to *C. difficile*, although these niche distances could vary widely for each organism depending on their community context. Finally, we compared the import fluxes of the probiotic strains and *C. difficile* for the metabolites identified as important for *C. difficile* growth (Figure 4.3). This analysis showed that, in addition to occupying niches similar to *C. difficile*, several of the probiotic strains directly competed for metabolites important for *C. difficile* growth, such as succinate, ornithine, and trehalose (Figure 4.6D). Cumulatively these results suggest that metabolic competition is the mechanism by which the probiotic cocktail suppressed *C. difficile* growth, which is consistent with the emerging consensus in the field [147, 180, 182]. Finally, we found that certain probiotic strains were more or less likely to engraft in an individual (Figure 4.6D), and that this engraftment/growth was generally associated with *C. difficile* suppression (Figure 4.6B), which indicates that MCMMs can be leveraged to identify responders and non-responders before these kinds of probiotic interventions.

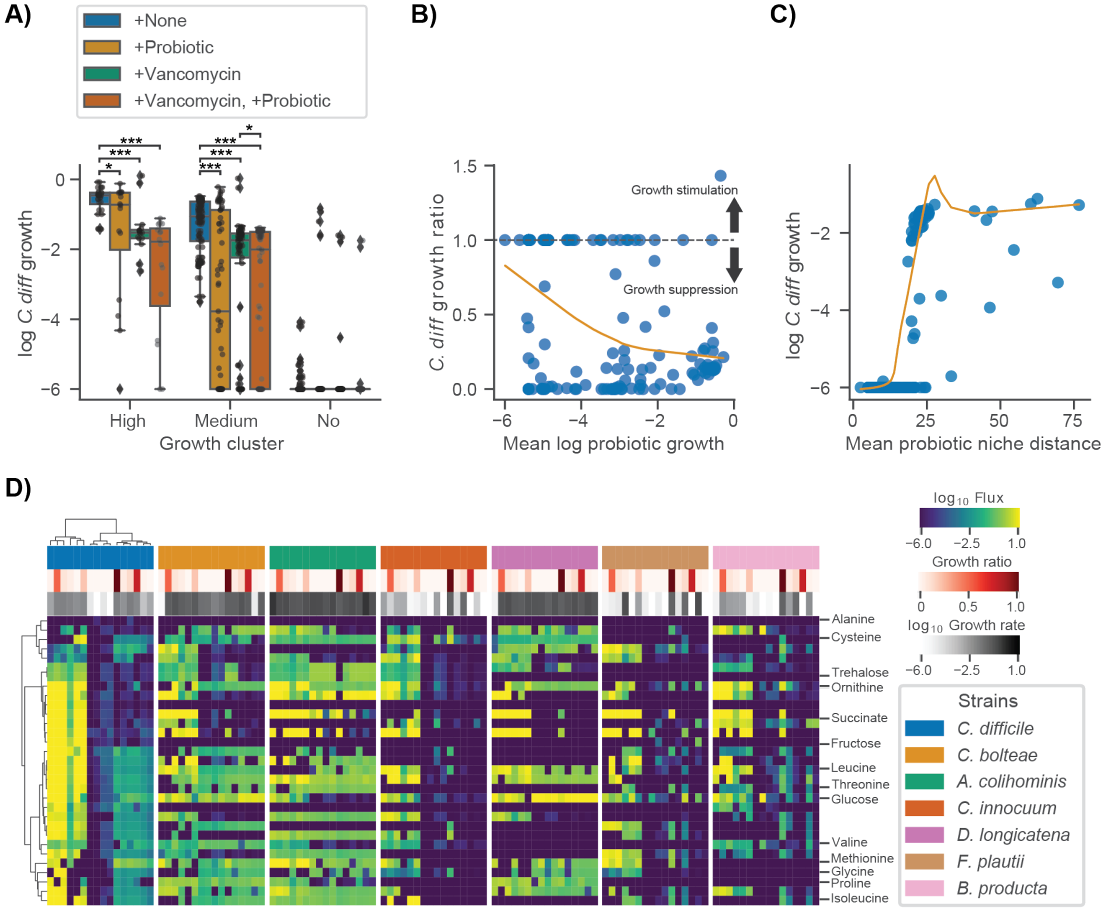

Figure 4.6. Simulated probiotic intervention effectively suppresses *C. difficile* growth *in silico*. (**A**) Box plots displaying $\log_{10}$ *C. difficile* growth rate across growth clusters and simulated interventions. Growth clusters are those identified by bi-clustering of *C. difficile* import fluxes using the Weingarden data (Figure 4.3). Conditions include +None (no intervention control), +Probiotic (introduction of 6 strain probiotic previously identified as an effective treatment for CDI at a total relative abundance of 50% equally distributed across the strains), +Vancomycin (90% reduction of *C. difficile* relative abundance as well as all genera known to be impacted by Vancomycin), and +Vancomycin, +Probiotic (introduction of 6 strain probiotic in combination with simulated vancomycin treatment). Bars indicate comparisons for which differences were significant using the Wilcoxon signed-rank test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (**B**) Relationship between *C. difficile* growth ratio and mean $\log_{10}$ probiotic growth rate. *C. difficile* growth ratio is the growth rate of samples in the +Vancomycin, +Probiotic intervention relative to +None. Values below 1 indicate growth suppression by the probiotic and values above 1 indicate growth stimulation. The dashed line marks the value at which no effect is observed (1). The orange

trend line indicates a LOWESS fit to the *C. difficile* growth ratio and mean $\log_{10}$ probiotic growth rate. (**C**) Relationship between $\log_{10}$ *C. difficile* growth rate and mean probiotic niche distance. Niche distance was calculated using the euclidean distance of $\log_{10}$ import flux vectors of each probiotic strain relative to C. difficile on a per sample basis. Orange trend line indicates a LOWESS fit to the $\log_{10}$ *C. difficile* growth rate and mean probiotic niche distance. (**D**) Biclustered $\log_{10}$ import fluxes for C. difficile and probiotic strains for samples previously identified as “high growth”, where each row is the import flux of a particular metabolite and each column is a patient sample. Imports displayed are those previously identified as important for *C. difficile*. Color bars indicate sample *C. difficile* growth ratio and strain specific $\log_{10}$ growth rate. Ordering of samples and metabolites is the same across heatmaps and based on biclustering of *C. difficile* data.

4.3 DISCUSSION

In this study, we provide a framework for predicting *C. difficile* engraftment risk in the human gut microbiome using MCMMs. While we focus on *C. difficile*, due to its clinical importance, this approach could be extended to other opportunistic bacterial pathogens, probiotic organisms, or even entire communities, in the case of FMTs. We were able to show how our approach predicts expected longitudinal and cross-sectional variation in *C. difficile* colonization potential *in vitro* and *in vivo* (Figures 4.1-4.2), using both shotgun metagenomic and 16S amplicon sequencing data sets, and we provide insights into the metabolic strategies leveraged by *C. difficile* in different ecological contexts (Figure 4.3-4.4). Our analysis not only recapitulates known metabolic associations with *C. difficile* growth (e.g., consumption of trehalose, ornithine, and succinate; Figure 4.3), it suggests additional associations (e.g., importance of reduced sulfur compounds like cysteine, Stickland fermentation reactants, and utilization of dietary sugars, like fructose; Figure 4.3). Additionally, we show that competition and cooperation with community members can prevent or promote colonization of *C. difficile*, and that many of these associations are highly context-dependent (Figure 4.4).

Consistent with the idea that simple metrics of community structure and composition alone are not effective predictors of colonization susceptibility, we found that community compositional variation was a modest predictor of estimated *C. difficile* growth rate and the relationship between alpha diversity and estimated *C. difficile* growth rate was nonlinear (Figure 4.5C). Not only did low diversity communities tend to be more invasible, as might be expected due to putative non-saturation of the metabolic niche space [147], high-diversity communities were also more prone to *C. difficile* engraftment (Figure 4.5C). In high-diversity communities, successful colonization may be due to changes in the interaction landscape, like a higher rate of niche construction in higher-diversity systems, which would be in line with the diversity-begets-diversity hypothesis [183]. An intermediate range of alpha diversity seems to be optimal for mitigating *C. difficile* colonization potential (Figure 4.5C). Overall, these complex mappings between community composition and pathobiont engraftment risk underscore the necessity of systems-scale tools, like MCMMs, that are capable of synthesizing this complexity.

Several genera were found to engage in cooperative and competitive interactions with *C. difficile* across MCMMs. *Blautia*, *Faecalibacterium*, and *Dorea* were all shown to benefit *C. difficile* through production of key metabolites that it consumes, like succinate and leucine, but were also capable of competition for other metabolic resources (Figure 4.4). Meanwhile, *Rumminococcus*, *Bacteroides* and *Phocaeicola* growth rates were often negatively associated with *C. difficile* engraftment, competing for some of the same metabolites that *C. difficile* consumed at high rates (Figure 4.1D, 4.4). Contextualizing these results through analysis of individual taxon import fluxes across studies we found that *Blautia*, *Faecalibacterium*, *Phocaeicola, Rumminococcus,* and *Dorea* share similar niches, with *Phocaeicola* being the most divergent from the others. In most cases, these niches did not overlap with *C. difficile*, but in a subset of individuals all five occupied niches states overlapping or similar to *C. difficile* (Figure 4.5B). Thus, while we observed competition for some key metabolites, on a global scale, the majority of the metabolic niche space used by *C. difficile* tends not to overlap with its apparent competitors (Figure 4.5B). These results highlight how flexible commensal gut bacteria are in adjusting their import fluxes to the communities in which they reside, which in turn suggests why so many taxa can coexist. These results also suggest that colonization resistance is an emergent feature of multi-species metabolic interactions, and not strongly driven by any particular pairwise interaction.

In addition to developing a simulation framework to predict engraftment, we sought to identify blood-based clinical chemistries and blood metabolites that were associated with MCMM-inferred *C. difficile* growth rate. We found three blood metabolites that were independently associated with predicted *C. difficile* growth rates. These included two secondary bile acids and an unannotated metabolite. One of the secondary bile acids, isoursodeoxycholate, has previously been positively associated with the abundance of *Bacteroides* [176], and was negatively associated with predicted *C. difficile* growth rates. This result is in line with the apparent competition between *Bacteroides* and *C. difficile* in our MCMMs. We also identified several clinical labs negatively associated with predicted growth rates. However, together with age, sex, and BMI, these features only accounted for ~5% of the variance in predicted growth rates. Thus, while these features may be signatures for colonization susceptibility in the blood, their clinical relevance is limited at this time.

Finally, we demonstrated that a probiotic intervention (i.e., 6/8 strains from VE303), which recently showed positive efficacy results in a double-blinded, placebo-controlled clinical trial for the treatment of rCDI [159], suppresses the growth of *C. difficile in silico* in most people in the Weingarden *et al.* rCDI-FMT study (Figure 4.6). We also showed that the mechanism of action for this particular probiotic is likely competition for metabolites essential for the growth of *C. difficile*, as many of the probiotic strains occupy niches close to *C. difficile,* and directly compete for these metabolites, such as succinate and ornithine, in samples where growth suppression was observed (Figure 4.6). Furthermore, analysis of niche distances between *C. difficile* and other genera across donors suggests selecting strains from *Blautia and Dorea* (e.g., including *B. producta and D. longicatana,* from VE303), in addition to *Anaerostipes*, *Roseburia*, and *Faecalibacterium*, could be leveraged to design better individual-specific probiotic cocktails capable of suppressing *C. difficile* and rescuing VE303 non-responders (Figure 4.6B). These results illustrate how MCMMs may be powerful tools for assessing the individual-specific efficacy of clinically relevant probiotics, in addition to understanding personalized pathobiont colonization susceptibility.

Future work should test this MCMM framework in the design of precision interventions to prevent engraftment of *C. difficile* and other pathobionts, to design precision probiotic interventions, and to improve the production of beneficial microbial metabolites, like short-chain-fatty-acids, or reduce the production of undesired metabolites, such as hydrogen sulfide or trimethylamine [184–186]. In summary, MCMMs present a promising new path forward in engineering the ecological composition and metabolic outputs of microbiota to prevent or treat disease.

4.4 METHODS

4.4.1 *External data collection and processing.*

Data used in this study came from six sources. This included both cross-sectional and time series 16S amplicon and shotgun metagenomics sequencing data from Hromada *et al.*, David *et al.*, Weingarden, A. *et al.*, Ianiro *et al.,* the American Gut (McDonald, D. *et al.*), and a former scientific wellness program run by Arivale, Inc.[156, 161, 163, 164, 167, 177]. Publicly available 16S amplicon sequence data and associated metadata were downloaded from the sequence read archive (SRA). Additionally, de-identified 16S amplicon sequence data, associated metadata, and paired blood-based clinical chemistries and metabolomics were obtained for 2,687 research-consenting individuals who were formerly participants in the Arivale wellness program. Raw 16S amplicon sequence data were processed using QIIME2 (v2020.11.1). In brief, the QIIME2 workflow consisted of read demultiplexing using the command *qiime tools import*, and an associated manifest table for each study describing read metadata followed by read quality assessment using *qiime demux summarize*. Read quality assessment was used to determine trimming parameters for subsequent denoising using the QIIME2 implementation of DADA2 via the command, *qiime dada2 denoise-single* or *qiime dada2 denoise-paired*, for single and paired reads respectively. The first 10 bases were trimmed from all reads and reads were truncated to a length where median quality score was >20 (100-150 base pair for the data leveraged). Following denoising, data were reformatted into a table format using the command *qiime metadata tabulate*, and representative sequence taxonomy was inferred using a custom NCBI classifier with the command *qiime feature-classifier classify-sklearn*. The NCBI classifier was trained using 16S 515f-806r V4 regions extracted from all available bacterial NCBI genomes. To train the classifier 515f-806r regions were extracted from NCBI sequences using the command *qiime feature-classifier extract-reads*, followed by the command *qiime feature-classifier fit-classifier-naive-bayes* using the extracted V4 sequences and a table of known taxonomies. Pre-processed data from Iraniro *et al.* were provided by the authors upon request and included sample taxonomic relative abundance inferred using Metaphlan4 and associated sample metadata. See Iraniro *et al.* for further details on their data processing pipelines [156]. Sample taxonomic relative abundance and associated metadata from Hromada *et al.* were obtained from their supplemental data. See Hromada *et al.* for further details [164]. For source code and tables of processed data refer to the Github repository listed below in Data and Source Code Availability.

4.4.2 *Model construction and growth simulations.*

To construct community-level metabolic models sample specific taxonomic abundance profiles inferred from 16S amplicon or shotgun metagenomic sequencing were summarized at the genus or species level and mapped to genus or species level metabolic models from the AGORA database (v1.03) using MICOM (v0.25.1). Taxa with a relative abundance of less than 0.1% were omitted from community models. An *in silico* media previously designed to represent an average western diet was applied which defined the bounds for metabolic imports by the model communities for all *in vivo* analyses [25, 160]. For the *in vitro* communities, a custom medium was developed to reflect the ABB medium used experimentally. The custom medium was developed using a previously defined template designed for the representation of Luria-Bertani (LB) medium, this contained metabolite concentration and fluxes for both yeast extract and tryptone [166]. These details were used to convert the quantities of yeast extract and peptone present in the experimental medium anaerobic basal broth (ABB) to fluxes. The additional defined components were included with fluxes equal to the millimolar concentration per hour. Finally, additional components were added to the constructed medium so that all 13 species used in the *in vitro* communities could achieve a growth rate of 10^{-3} or greater using a routine implemented in the MICOM function *complete_db_medium*. Across all conditions, growth rates were inferred using cooperative tradeoff flux balance analysis (ctFBA). In brief, this is a two-step optimization scheme, where the first step finds the largest possible biomass production rate for the full microbial community and the second step infers taxon-specific growth rates and fluxes, while maintaining community growth within a fraction of the theoretical maximum (i.e., the tradeoff parameter), thus balancing individual growth rates and the community-wide growth rate[25]. For all models in the manuscript that leveraged 16S amplicon data, we used a tradeoff parameter of 0.8. A tradeoff parameter of 0.9 was used for all the results derived from shotgun metagenomics data. These parameter values were chosen by identifying the largest tradeoff value (corresponding to the smallest deviation from maximal community biomass) that allowed > 70% taxa to grow (growth rate > 10^{-6}). Import and export fluxes were estimated using parsimonious enzyme usage FBA (pFBA) and a defined medium constructed to represent an average European diet or ABB [25]. pFBA further constrained simulation results by requiring genera to utilize the lowest overall flux through their networks to achieve maximal growth [157]. For source code and tables of processed data refer to the Github repository listed below in Data and Source Code Availability.

4.4.3 *Biomass estimation and growth rate normalization for in vitro communities.*

To allow better comparison between model predictions and experimental results two growth rate transformations were employed. These methods were implemented to account for the fact that batch growth (in which cultures grow on a limited quantity of nutrients) is not compatible with the steady state assumption employed by the modeling framework (which assumes a continuous supply of nutrients). Estimation of biomass was leveraged to compare growth rates with experimentally measured endpoint optical densities. This consisted of using a simple exponential growth model of the form $OD_f = OD_i e^{kt}$, where OD_f is the final predicted biomass, OD_i is the initial biomass, k is the MCMM inferred growth rate, and t is time in hours. The second transformation employed was a growth rate normalization procedure, which accounted for differences in community growth rates across samples. Growth rates were divided by the MCMM inferred community growth rate, which was computed from the sum of relative growth rates of each taxon in the system (e.g., the product of growth rate and relative abundance), $k_{norm} = \frac{k}{\sum_i^n k_i r_i}$.

4.4.4 *Probiotic intervention.*

To model probiotic intervention a combination of strains previously shown to be effective at suppressing the growth of *C. difficile* in mice were used [180]. Metabolic models for six of the eight stains in the VE303 cocktail described by Dsouza *et al.* were identified in the AGORA database and intervention was simulated by introducing them along with *C. difficile* to individual samples. A total probiotic fraction of 50% was used, which was evenly distributed among the six strains. This fraction was determined to be the most effective at suppressing the growth of *C. difficile* growth *in silico* for the samples tested (data not shown). Vancomycin treatment was simulated by reducing the abundance of *C. difficile* and all genera known to be impacted by vancomycin by 90% [181]. Growth simulations were performed as described above. For source code and tables of processed data refer to the Github repository listed below in Data and Source Code Availability.

4.4.5 *Statistical analysis.*

Statistical analyses were performed using functions from the Python SciPy (v1.7.1), Seaborn (v0.11.2), Sklearn-Learn (v0.24.2), Umap-Learn (v0.5.1), and Statsmodels (v0.13.1) packages. Linear associations were performed using the *statsmodels* ordinary least squares function *OLS*, and visualized using the *seaborn* function *regplot*. Least absolute shrinkage and selection operator (LASSO) was performed using the *sklearn Lasso* function and a training-test framework. Data were split into training and test sets (70% of samples were randomly assigned to the training set) and model performance was assessed across a range of regularization values. LASSO training and test set R^2 values were used to select the model with the best test set R^2 that did not overfit training data (training R^2 >> test R^2). Analysis of variance (ANOVA) was performed using the *statsmodels OLS* and *anova_lm* functions. UMAP dimensionality reduction was performed using the *umap* function from the *umap* package and associated methods with default parameters (i.e., *n_components=2*, *n_neighbors=15*, *metric='euclidean'*, etc.). Biclustering was performed using the *seaborn* function *clustermap* and the Ward clustering algorithm. Hexagonal binning and associated histograms were generated using the *seaborn* function *jointplot.* Locally weighted scatterplot smoothing (LOWESS) curves were generated using the *lowess* function from *statsmodels* with default parameters. Additional statistical tests included the t-test and Wilcox rank sum test implemented in scipy as *ttest_ind* and *wilcoxon* respectively. For source code and tables of processed data refer to the Github repository listed below in Data and Source Code Availability.

4.5 DATA AND SOURCE CODE AVAILABILITY

Processed data tables and source code to reproduce the findings presented in this manuscript can be found at https://github.com/Gibbons-Lab/cdiff_invasion. Raw 16S amplicon sequence data from David *et al.*, Weingarden, A. *et al.*, the American Gut (McDonald, D. *et al.*) can be downloaded using the sequence read archive (SRA) accession numbers PRJEB6518, PRJEB19996, and PRJEB11419 respectively. Data and metadata from Iraniro *et al.* were provided from the authors upon request. Data from Hromada *et al.* were downloaded from the manuscript supplementary information. Metadata were obtained from manuscript supplementary information, where available. Qualified researchers can access the full Arivale deidentified dataset, including all raw data, supporting the findings in this study for research purposes through signing a Data Use Agreement (DUA). Inquiries to access the data can be made at data-access@isbscience.org and will be responded to within 7 business days.

Chapter 5. CONCLUSION

In summary, the body of work presented here highlights how systems biology approaches can be leveraged to infer microbial community interaction mechanisms and understand ecological phenomena. Chapters 2 and 3 highlight how environmental context and community dynamics can influence the production of the greenhouse gas nitrous oxide (N_2O) by nitrate-reducing bacteria (NRB) and the communities they form. Growth inhibition by the products of nitrate (NO_3^-) and sulfate respiration are a key feature of the work presented in these chapters. We found that both hydrogen sulfide (H_2S) and nitrite (NO_2^-) can act as inhibitors of nitrate respiratory growth and either suppress or enhance the production of N_2O as a consequence. These findings indicate that ecological interactions can have a significant impact on the production of N_2O in natural environments. Competition between sulfate-reducing bacteria (SRB) and NRB can suppress the production of N_2O as a result of H_2S production [38, 55], while the inhibitory effects of NO_2^- can lead to N_2O production when the kinetics of denitrification are unbalanced in nitrate-reducing communities [129, 130]. However, both of these effects are highly dependent on context. For instance, previous work has demonstrated that NO_3^- can be leveraged to stimulate NRB, suppress SRB, and prevent the souring of oil reservoirs [37, 65]. Meanwhile, NO_2^- inhibition can either stimulate or suppress N_2O production depending on the specific context [120, 129, 130]. These chapters also highlight strategies for characterizing interaction mechanisms using isolates and Synthetic Communities (SynComs). Laboratory characterization of isolates and SynComs provide a framework with which detailed information about growth dynamics and cellular physiology can be extracted. This framework provides an ideal means with which to infer and validate interactions. However, laboratory studies are also limited in many ways. For instance, the studies described in chapters 2 and 3 present vastly oversimplified models of the natural ecosystems the study designs were intended to represent. While these bottom-up approaches do not capture the full complexity of the natural ecosystems they seek to approximate, they do provide a means to dissect specific details of these systems and develop insights that can be used to better understand the dynamics of natural ecosystems.

In contrast to the bottom-up approaches presented in Chapters 2 and 3, Chapter 4 highlights a top-down strategy to infer microbial community interactions. In many ways, this strategy complements the study designs presented in Chapters 2 and 3. Instead of breaking down the complexity of the human gut microbiome into simpler model systems, we leveraged microbial community-scale metabolic models (MCMMs) to predict community phenotypes *in vivo*. This approach was used to better understand mechanisms of *Clostridiodes difficile* (*C. difficile*) colonization susceptibility and resistance. With the mechanistic information provided by MCMMs, we developed insights into the metabolic niche of *C. difficile* as well as the cooperative and competitive interactions that govern *C. difficile* colonization. Cross-feeding and competition involving metabolites like succinate, trehalose, ornithine, and leucine were found to be important determinants of colonization. Furthermore, the extent of cross-feeding and competition for these metabolites was highly dependent on the community context. For example, we found that *Roseburia*, and *Faecalibacterium* often compete with *C. difficile* for ornithine, but in some contexts provide *C. difficile* with succinate and leucine. This result alongside the findings presented in Chapters 2 and 3 illustrates an important feature of microbial interactions; that context matters and it can be difficult at times to make generalizations about the strategies used by individual microbes. A consequence of this is that it is often difficult to bin commensal microbes into simple binary categories like “good bugs” and “bad bugs” because whether an individual commensal plays a role that is beneficial or detrimental to the host can be highly context-dependent. *Akkermansia muciniphila* (*A. muciniphila*) provides another example of the context-dependent nature of commensal-host bacterial relationships. A growing body of literature has identified *A. muciniphila* as a promising therapeutic probiotic against metabolic disorders such as obesity, type 2 diabetes, and cardiovascular diseases [187]. However, in specific intestinal microenvironments, its excessive enrichment may be not beneficial. Conditions like inflammatory bowel disease (IBD), *Salmonella typhimurium* infection, and post-antibiotic reconstitution may not benefit from *A. muciniphila* supplementation [188]. This is because of *A. muciniphila*'s roles as both a fiber degrading short chain fatty acid (SCFA) producer and a mucin degrader [189]. When fiber is available *A. muciniphila* will produce SCFAs that contribute to intestinal epithelial barrier maintenance, lower inflammation, and suppress the growth of some pathogens [190–192]. However, when sufficient levels of fiber are not present *A. muciniphila* will instead degrade host mucin, which can compromise intestinal barrier integrity and increase levels of inflammation [193, 194]. MCMMs have the potential to provide insights into the dietary and community contexts in which probiotics like *A. muciniphila* may be efficacious. Moving forward, MCMMs may also be able to aid the design of personalized probiotic and prebiotic interventions to promote the production of SCFAs and other beneficial metabolites.

As it becomes more feasible to map community composition to function and we develop a deeper collective understanding of the phenotypic landscapes of microbial consortia, engineering desired community functions may be realized [195]. Achieving this goal will require overcoming significant hurdles in our ability to characterize and model microbial interactions. The work presented here highlights advances in these areas but microbiome engineering remains a nascent field. Alongside the approaches described, advances in high throughput characterization of microbial interactions and enzyme kinetics [196, 197], integration of gene regulation and kinetics into MCMMs [26, 198], and better techniques for delivery, tracking, and maintenance of engineered consortia are getting us closer to engineering microbiomes [199, 200]. As the field develops, it may provide solutions that will allow us to address challenges in human health and the health of natural ecosystem

BIBLIOGRAPHY

1. Mony C, Vandenkoornhuyse P, Bohannan BJM, Peay K, Leibold MA. A Landscape of Opportunities for Microbial Ecology Research. *Front Microbiol* 2020; **11**: 561427.
2. Morris A, Meyer K, Bohannan B. Linking microbial communities to ecosystem functions: what we can learn from genotype-phenotype mapping in organisms. *Philos Trans R Soc Lond B Biol Sci* 2020; **375**: 20190244.
3. Gougoulias C, Clark JM, Shaw LJ. The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *J Sci Food Agric* 2014; **94**: 2362–2371.
4. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. *Nat Rev Microbiol* 2018; **16**: 263–276.
5. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017; **551**: 457–463.
6. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007; **449**: 804–810.
7. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 2016; **14**: 20–32.
8. Vuong HE, Yano JM, Fung TC, Hsiao EY. The Microbiome and Host Behavior. *Annu Rev Neurosci* 2017; **40**: 21–49.
9. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 2021; **19**: 55–71.
10. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, et al. Natural bacterial

communities serve as quantitative geochemical biosensors. *MBio* 2015; **6**: e00326-15.

11. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* 2016; **10**: 2557–2568.
12. Antwis RE, Griffiths SM, Harrison XA, Aranega-Bou P, Arce A, Bettridge AS, et al. Fifty important research questions in microbial ecology. *FEMS Microbiol Ecol* 2017; **93**.
13. Carr A, Diener C, Baliga NS, Gibbons SM. Use and abuse of correlation analyses in microbial ecology. *ISME J* 2019; **13**: 2647–2655.
14. Orphan VJ, Turk KA, Green AM, House CH. Patterns of 15N assimilation and growth of methanotrophic ANME-2 archaea and sulfate-reducing bacteria within structured syntrophic consortia revealed by FISH-SIMS. *Environ Microbiol* 2009; **11**: 1777–1791.
15. Russell AB, Wexler AG, Harding BN, Whitney JC, Bohn AJ, Goo YA, et al. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host Microbe* 2014; **16**: 227–236.
16. Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol Syst Biol* 2018; **14**: e8157.
17. Friedman J, Higgins LM, Gore J. Community structure follows simple assembly rules in microbial microcosms. *Nat Ecol Evol* 2017; **1**: 109.
18. Liu A, Archer AM, Biggs MB, Papin JA. Growth-altering microbial interactions are responsive to chemical context. *PLoS One* 2017; **12**: e0164919.
19. Qian Y, Lan F, Venturelli OS. Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models. *Curr Opin Microbiol* 2021; **62**: 84–92.

20. Durot M, Bourguignon P-Y, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev* 2009; **33**: 164–190.

21. Brooks AN, Reiss DJ, Allard A, Wu W-J, Salvanha DM, Plaisier CL, et al. A system-level model for the microbial regulatory genome. *Mol Syst Biol* 2014; **10**: 740.

22. Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. *Essays Biochem* 2018; **62**: 487–500.

23. Meyer P, Saez-Rodriguez J. Advances in systems biology modeling: 10 years of crowdsourcing DREAM challenges. *Cell Syst* 2021; **12**: 636–653.

24. O'Donnell ST, Ross RP, Stanton C. The Progress of Multi-Omics Technologies: Determining Function in Lactic Acid Bacteria Using a Systems Level Approach. *Front Microbiol* 2019; **10**: 3084.

25. Diener C, Gibbons SM, Resendis-Antonio O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* 2020; **5**.

26. Immanuel SRC, Arrieta-Ortiz ML, Ruiz RA, Pan M, Lopez Garcia de Lomana A, Peterson EJR, et al. Quantitative prediction of conditional vulnerabilities in regulatory and metabolic networks using PRIME. *NPJ Syst Biol Appl* 2021; **7**: 43.

27. Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, et al. The global nitrogen cycle in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci* 2013; **368**: 20130164.

28. Vuono DC, Read RW, Hemp J, Sullivan BW, Arnone JA 3rd, Neveux I, et al. Resource Concentration Modulates the Fate of Dissimilated Nitrogen in a Dual-Pathway Actinobacterium. *Front Microbiol* 2019; **10**: 3.

29. Carlson HK, Price MN, Callaghan M, Aaring A, Chakraborty R, Liu H, et al. The selective pressures on the microbial community in a metal-contaminated aquifer. *ISME J* 2019; **13**:

937–949.

30. Carlson HK, Lui LM, Price MN, Kazakov AE, Carr AV, Kuehl JV, et al. Selective carbon sources influence the end products of microbial nitrate respiration. *ISME J* 2020; **14**: 2034–2045.

31. Finneran KT, Housewright ME, Lovley DR. Multiple influences of nitrate on uranium solubility during bioremediation of uranium-contaminated subsurface sediments. *Environ Microbiol* 2002; **4**: 510–516.

32. Hemme CL, Tu Q, Shi Z, Qin Y, Gao W, Deng Y, et al. Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. *Front Microbiol* 2015; **6**: 1205.

33. Ayangbenro AS, Olanrewaju OS, Babalola OO. Sulfate-Reducing Bacteria as an Effective Tool for Sustainable Acid Mine Bioremediation. *Front Microbiol* 2018; **9**: 1986.

34. Zhang J, Lan T, Müller C, Cai Z. Dissimilatory nitrate reduction to ammonium (DNRA) plays an important role in soil nitrogen conservation in neutral and alkaline but not acidic rice soil. *J Soils Sediments* 2015; **15**: 523–531.

35. Smirnov PM, Kidin VV, Pedishyus RK. Loss of nitrogen by denitrification. *Biol Bull Acad Sci USSR* 1979; **6**: 450–459.

36. He Q, He Z, Joyner DC, Joachimiak M, Price MN, Yang ZK, et al. Impact of elevated nitrate on sulfate-reducing bacteria: a comparative study of Desulfovibrio vulgaris. *ISME J* 2010; **4**: 1386–1397.

37. Xue Y, Voordouw G. Control of Microbial Sulfide Production with Biocides and Nitrate in Oil Reservoir Simulating Bioreactors. *Front Microbiol* 2015; **6**: 1387.

38. Sørensen J, Tiedje JM, Firestone RB. Inhibition by sulfide of nitric and nitrous oxide reduction by denitrifying Pseudomonas fluorescens. *Appl Environ Microbiol* 1980; **39**: 105–

108.

39. Brunet RC, Garcia-Gil LJ. Sulfide-induced dissimilatory nitrate reduction to ammonia in anaerobic freshwater sediments. *FEMS Microbiol Ecol* 1996; **21**: 131–138.
40. Joye SB, Hollibaugh JT. Influence of Sulfide Inhibition of Nitrification on Nitrogen Regeneration in Sediments. *Science* 1995; **270**: 623–625.
41. Zhang W, Li F, Nie L. Integrating multiple "omics" analysis for microbial biology: application and methodologies. *Microbiology* 2010; **156**: 287–301.
42. Harris CL. Cysteine and growth inhibition of Escherichia coli: threonine deaminase as the target enzyme. *J Bacteriol* 1981; **145**: 1031–1035.
43. Fukushima RS, Weimer PJ, Kunz DA. Use of photocatalytic reduction to hasten preparation of culture media for saccharolytic Clostridium species. *Braz J Microbiol* 2003; **34**: 22–26.
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.
45. Kursa M, Jankowski A, Rudnicki W. Boruta - A System for Feature Selection. *Fund Inform* 2010; **101**: 271–285.
46. Folsom JP, Parker AE, Carlson RP. Physiological and proteomic analysis of Escherichia coli iron-limited chemostat growth. *J Bacteriol* 2014; **196**: 2748–2761.
47. Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C. Metal and redox modulation of cysteine protein function. *Chem Biol* 2003; **10**: 677–693.
48. Huan T, Forsberg EM, Rinehart D, Johnson CH, Ivanisevic J, Benton HP, et al. Systems biology guided by XCMS Online metabolomics. *Nat Methods* . 2017. , **14**: 461–462
49. Forsberg EM, Huan T, Rinehart D, Benton HP, Warth B, Hilmers B, et al. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nat*

Protoc 2018; **13**: 633–651.

50. Ung KSE, Av-Gay Y. Mycothiol-dependent mycobacterial response to oxidative stress. *FEBS Lett* 2006; **580**: 2712–2716.
51. Massey LK, Sokatch JR, Conrad RS. Branched-chain amino acid catabolism in bacteria. *Bacteriol Rev* 1976; **40**: 42–54.
52. Takumi K, Nonaka G. Bacterial Cysteine-Inducible Cysteine Resistance Systems. *J Bacteriol* 2016; **198**: 1384–1392.
53. Korshunov S, Imlay KRC, Imlay JA. Cystine import is a valuable but risky process whose hazards Escherichia coli minimizes by inducing a cysteine exporter. *Mol Microbiol* 2020; **113**: 22–39.
54. Chonoles Imlay KR, Korshunov S, Imlay JA. Physiological Roles and Adverse Effects of the Two Cystine Importers of Escherichia coli. *J Bacteriol* 2015; **197**: 3629–3644.
55. Pan Y, Ye L, Yuan Z. Effect of H2S on N2O reduction and accumulation during denitrification by methanol utilizing denitrifiers. *Environ Sci Technol* 2013; **47**: 8408–8415.
56. Bowles MW, Nigro LM, Teske AP, Joye SB. Denitrification and environmental factors influencing nitrate removal in Guaymas Basin hydrothermally altered sediments. *Front Microbiol* 2012; **3**: 377.
57. Jones ZL, Jasper JT, Sedlak DL, Sharp JO. Sulfide-Induced Dissimilatory Nitrate Reduction to Ammonium Supports Anaerobic Ammonium Oxidation (Anammox) in an Open-Water Unit Process Wetland. *Appl Environ Microbiol* 2017; **83**.
58. Soutourina J, Blanquet S, Plateau P. Role of D-cysteine desulfhydrase in the adaptation of Escherichia coli to D-cysteine. *J Biol Chem* 2001; **276**: 40864–40872.
59. Chan L-K, Morgan-Kiss RM, Hanson TE. Functional analysis of three sulfide:quinone

oxidoreductase homologs in Chlorobaculum tepidum. *J Bacteriol* 2009; **191**: 1026–1034.

60. Fahey RC. Glutathione analogs in prokaryotes. *Biochim Biophys Acta* 2013; **1830**: 3182–3198.
61. Paritala H, Carroll KS. New targets and inhibitors of mycobacterial sulfur metabolism. *Infect Disord Drug Targets* 2013; **13**: 85–115.
62. Van Laer K, Hamilton CJ, Messens J. Low-molecular-weight thiols in thiol-disulfide exchange. *Antioxid Redox Signal* 2013; **18**: 1642–1653.
63. Vitvitsky V, Miljkovic JL, Bostelaar T, Adhikari B, Yadav PK, Steiger AK, et al. Cytochrome c Reduction by H2S Potentiates Sulfide Signaling. *ACS Chem Biol* 2018; **13**: 2300–2307.
64. Peng H, Shen J, Edmonds KA, Luebke JL, Hickey AK, Palmer LD, et al. Sulfide Homeostasis and Nitroxyl Intersect via Formation of Reactive Sulfur Species in Staphylococcus aureus. *mSphere* 2017; **2**.
65. Dolfing J, Hubert CRJ. Using Thermodynamics to Predict the Outcomes of Nitrate-Based Oil Reservoir Souring Control Interventions. *Front Microbiol* 2017; **8**: 2575.
66. Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* 2011; **75**: 361–422.
67. Chen J, Hanke A, Tegetmeyer HE, Kattelmann I, Sharma R, Hamann E, et al. Impacts of chemical gradients on microbial community structure. *ISME J* 2017; **11**: 920–931.
68. Takaya N, Catalan-Sakairi MAB, Sakaguchi Y, Kato I, Zhou Z, Shoun H. Aerobic denitrifying bacteria that produce low levels of nitrous oxide. *Appl Environ Microbiol* 2003; **69**: 3152–3157.
69. Gao H, Schreiber F, Collins G, Jensen MM, Svitlica O, Kostka JE, et al. Aerobic

denitrification in permeable Wadden Sea sediments. *ISME J* 2010; **4**: 417–426.

70. Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, et al. A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* 2010; **330**: 1375–1378.
71. Dalsgaard T, Bak F. Nitrate Reduction in a Sulfate-Reducing Bacterium, Desulfovibrio desulfuricans, Isolated from Rice Paddy Soil: Sulfide Inhibition, Kinetics, and Regulation. *Appl Environ Microbiol* 1994; **60**: 291–297.
72. Moon J-W, Paradis CJ, Joyner DC, von Netzer F, Majumder EL, Dixon ER, et al. Characterization of subsurface media from locations up- and down-gradient of a uranium-contaminated aquifer. *Chemosphere* 2020; **255**: 126951.
73. Villahermosa D, Corzo A, Garcia-Robledo E, González JM, Papaspyrou S. Kinetics of Indigenous Nitrate Reducing Sulfide Oxidizing Activity in Microaerophilic Wastewater Biofilms. *PLoS One* 2016; **11**: e0149096.
74. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 2018; **557**: 503–509.
75. Khodursky AB, Bernstein JA, Peter BJ, Rhodius V, Wendisch VF, Zimmer DP. Escherichia coli spotted double-strand DNA microarrays: RNA extraction, labeling, hybridization, quality control, and data management. *Methods Mol Biol* 2003; **224**: 61–78.
76. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.
77. Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 16 Apr 2024.
78. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast

universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15–21.

79. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2014; **31**: 166–169.

80. Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST+ integrated into Galaxy. *Gigascience* 2015; **4**: 39.

81. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 421.

82. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015; **12**: 59–60.

83. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005; **33**: 5691–5702.

84. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011; **abs/1201.0490**.

85. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020; **17**: 261–272.

86. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature* 2020; **585**: 357–362.

87. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc* 1995; **57**: 289–300.

88. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:

25–29.

89. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2016; **44**: D471-80.

90. Kite GC. Comment on METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal Chem* . 2018. , **90**: 13126–13127

91. Xue J, Guijas C, Benton HP, Warth B, Siuzdak G. METLIN MS2 molecular standards database: a broad chemical and biological resource. *Nat Methods* 2020; **17**: 953–954.

92. Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti GJ, Siuzdak G. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat Biotechnol* 2012; **30**: 826–828.

93. Erisman JW, Sutton MA, Galloway J, Klimont Z, Winiwarter W. How a century of ammonia synthesis changed the world. *Nat Geosci* 2008; **1**: 636–639.

94. Battye W, Aneja VP, Schlesinger WH. Is nitrogen the next carbon? *Earths Future* 2017; **5**: 894–904.

95. Tian H, Xu R, Canadell JG, Thompson RL, Winiwarter W, Suntharalingam P, et al. A comprehensive quantification of global nitrous oxide sources and sinks. *Nature* 2020; **586**: 248–256.

96. Reay DS, Davidson EA, Smith KA, Smith P, Melillo JM, Dentener F, et al. Global agriculture and nitrous oxide emissions. *Nat Clim Chang* 2012; **2**: 410–416.

97. Duan H, Zhao Y, Koch K, Wells GF, Zheng M, Yuan Z, et al. Insights into Nitrous Oxide Mitigation Strategies in Wastewater Treatment and Challenges for Wider Implementation. *Environ Sci Technol* 2021; **55**: 7208–7224.

98. Bremner JM. Sources of nitrous oxide in soils. *Nutr Cycling Agroecosyst* 1997; **49**: 7–16.

99. Otwell AE, Carr AV, Majumder ELW, Ruiz MK, Wilpiszeski RL, Hoang LT, et al. Sulfur Metabolites Play Key System-Level Roles in Modulating Denitrification. *mSystems* 2021; **6**.

100. Shapleigh JP. The denitrifying prokaryotes. *The prokaryotes* 2006; **2**: 769–792.

101. Hallin S, Philippot L, Löffler FE, Sanford RA, Jones CM. Genomics and Ecology of Novel N2O-Reducing Microorganisms. *Trends Microbiol* 2018; **26**: 43–55.

102. Zumft WG. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* 1997; **61**: 533–616.

103. Gaimster H, Alston M, Richardson DJ, Gates AJ, Rowley G. Transcriptional and environmental control of bacterial denitrification and N2O emissions. *FEMS Microbiol Lett* 2018; **365**.

104. Zhang IH, Sun X, Jayakumar A, Fortin SG, Ward BB, Babbin AR. Partitioning of the denitrification pathway and other nitrite metabolisms within global oxygen deficient zones. *ISME Commun* 2023; **3**: 76.

105. 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.

106. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 2016; **7**: 13219.

107. Roco CA, Bergaust LL, Bakken LR, Yavitt JB, Shapleigh JP. Modularity of nitrogen-oxide reducing soil bacteria: linking phenotype to genotype. *Environ Microbiol* 2017; **19**: 2507–2519.

108. Nelson WC, Graham EB, Crump AR, Fansler SJ, Arntzen EV, Kennedy DW, et al. Distinct temporal diversity profiles for nitrogen cycling genes in a hyporheic microbiome. *PLoS One* 2020; **15**: e0228165.

109. Lilja EE, Johnson DR. Segregating metabolic processes into different microbial cells accelerates the consumption of inhibitory substrates. *ISME J* 2016; **10**: 1568–1578.

110. Wu X, Gushgari-Doyle S, Lui LM, Hendrickson AJ, Liu Y, Jagadamma S, et al. Distinct Depth-Discrete Profiles of Microbial Communities and Geochemical Insights in the Subsurface Critical Zone. *Appl Environ Microbiol* 2023; **89**: e0050023.

111. Ramzan S, Rasool T, Bhat RA, Ahmad P, Ashraf I, Rashid N, et al. Agricultural soils a trigger to nitrous oxide: a persuasive greenhouse gas and its management. *Environ Monit Assess* 2020; **192**: 436.

112. Green SJ, Prakash O, Jasrotia P, Overholt WA, Cardenas E, Hubbard D, et al. Denitrifying bacteria from the genus Rhodanobacter dominate bacterial communities in the highly contaminated subsurface of a nuclear legacy waste site. *Appl Environ Microbiol* 2012; **78**: 1039–1047.

113. Xu P. Analytical solution for a hybrid Logistic-Monod cell growth model in batch and continuous stirred tank reactor culture. *Biotechnol Bioeng* 2020; **117**: 873–878.

114. Peng M, Wang D, Lui LM, Nielsen T, Tian R, Kempher ML, et al. Genomic Features and Pervasive Negative Selection in Rhodanobacter Strains Isolated from Nitrate and Heavy Metal Contaminated Aquifer. *Microbiol Spectr* 2022; **10**: e0259121.

115. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol* 1994; **60**: 3724–3731.

116. Feist AM, Scholten JCM, Palsson BØ, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Mol Syst Biol* 2006; **2**: 2006.0004.

117. Brochado AR, Matos C, Møller BL, Hansen J, Mortensen UH, Patil KR. Improved vanillin production in baker's yeast through in silico design. *Microb Cell Fact* 2010; **9**: 84.

118. Yao C, Meng X, Qu X, Cheng T, Da Q, Zhang K, et al. Kinetic Model and Numerical Simulation of Microbial Growth, Migration, and Oil Displacement in Reservoir Porous Media. *ACS Omega* 2022; **7**: 32549–32561.

119. Croot PL, Heller MI. The importance of kinetics and redox in the biogeochemical cycling of iron in the surface ocean. *Front Microbiol* 2012; **3**: 219.

120. Frostegård Å, Vick SHW, Lim NYN, Bakken LR, Shapleigh JP. Linking meta-omics to the kinetics of denitrification intermediates reveals pH-dependent causes of N2O emissions and nitrite accumulation in soil. *ISME J* 2022; **16**: 26–37.

121. Dolinšek J, Ramoneda J, Johnson DR. Initial community composition determines the long-term dynamics of a microbial cross-feeding interaction by modulating niche availability. *ISME Commun* 2022; **2**: 77.

122. Mahadevan R, Edwards JS, Doyle FJ 3rd. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophys J* 2002; **83**: 1331–1340.

123. Antoniewicz MR. Dynamic metabolic flux analysis--tools for probing transient states of metabolic networks. *Curr Opin Biotechnol* 2013; **24**: 973–978.

124. Lachance J-C, Lloyd CJ, Monk JM, Yang L, Sastry AV, Seif Y, et al. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput Biol* 2019; **15**: e1006971.

125. Bayon-Vicente G, Marchand E, Ducrotois J, Dufrasne FE, Hallez R, Wattiez R, et al. Analysis of the Involvement of the Isoleucine Biosynthesis Pathway in Photoheterotrophic Metabolism of Rhodospirillum rubrum. *Front Microbiol* 2021; **12**: 731976.

126. Cabecas Segura P, De Meur Q, Tanghe A, Onderwater R, Dewasme L, Wattiez R, et al. Effects of Mixing Volatile Fatty Acids as Carbon Sources on Rhodospirillum rubrum Carbon Metabolism and Redox Balance Mechanisms. *Microorganisms* 2021; **9**.

127. Ramoneda J, Fan K, Lucas JM, Chu H, Bissett A, Strickland MS, et al. Ecological relevance of flagellar motility in soil bacterial communities. *ISME J* 2024.

128. Kost C, Patil KR, Friedman J, Garcia SL, Ralser M. Metabolic exchanges are ubiquitous in natural microbial communities. *Nat Microbiol* 2023; **8**: 2244–2252.

129. Maharjan B, Venterea RT. Nitrite intensity explains N management effects on N2O emissions in maize. *Soil Biol Biochem* 2013; **66**: 229–238.

130. Zhou Y, Pijuan M, Zeng RJ, Yuan Z. Free nitrous acid inhibition on nitrous oxide reduction by a denitrifying-enhanced biological phosphorus removal sludge. *Environ Sci Technol* 2008; **42**: 8260–8265.

131. Sijbesma WF, Almeida JS, Reis MA, Santos H. Uncoupling effect of nitrite during denitrification by Pseudomonas fluorescens: An in vivo (31)P-NMR study. *Biotechnol Bioeng* 1996; **52**: 176–182.

132. Zhou Y, Oehmen A, Lim M, Vadivelu V, Ng WJ. The role of nitrite and free nitrous acid (FNA) in wastewater treatment plants. *Water Res* 2011; **45**: 4672–4682.

133. Zumft WG. The biological role of nitric oxide in bacteria. *Arch Microbiol* 1993; **160**: 253–264.

134. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase:

The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 2018; **36**: 566–569.

135. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008; **9**: 386.

136. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res* 2021; **49**: D575–D588.

137. Hemme CL, Green SJ, Rishishwar L, Prakash O, Pettenato A, Chakraborty R, et al. Lateral Gene Transfer in a Heavy Metal-Contaminated-Groundwater Microbial Community. *MBio* 2016; **7**: e02234-15.

138. Hunt KA, Forbes J, Taub F, Elliott N, Hardwicke J, Petersen R, et al. An automated multiplexed turbidometric and data collection system for measuring growth kinetics of anaerobes dependent on gaseous substrates. *J Microbiol Methods* 2021; **188**: 106294.

139. Choi K, Medley JK, König M, Stocking K, Smith L, Gu S, et al. Tellurium: An extensible python-based modeling environment for systems and synthetic biology. *Biosystems* 2018; **171**: 74–79.

140. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; **45**: D353–D361.

141. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized Nutrition by Prediction of Glycemic Responses. *Cell* 2015; **163**: 1079–1094.

142. Schubert AM, Sinani H, Schloss PD. Antibiotic-Induced Alterations of the Murine Gut

Microbiota and Subsequent Effects on Colonization Resistance against Clostridium difficile. *MBio* 2015; **6**: e00974.

143. Bell A, Juge N. Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology* 2021; **31**: 691–696.

144. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. *MBio* 2019; **10**.

145. Gibbons SM. Defining Microbiome Health through a Host Lens. *mSystems* 2019; **4**.

146. Cheng AG, Ho P-Y, Aranda-Díaz A, Jain S, Yu FB, Meng X, et al. Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* 2022; **185**: 3617-3636.e19.

147. Spragge F, Bakkeren E, Jahn MT, B N Araujo E, Pearson CF, Wang X, et al. Microbiome diversity protects against pathogens by nutrient blocking. *Science* 2023; **382**: eadj3502.

148. VanInsberghe D, Elsherbini JA, Varian B, Poutahidis T, Erdman S, Polz MF. Diarrhoeal events can trigger long-term Clostridium difficile colonization with recurrent blooms. *Nat Microbiol* 2020; **5**: 642–650.

149. Shields K, Araujo-Castillo RV, Theethira TG, Alonso CD, Kelly CP. Recurrent Clostridium difficile infection: From colonization to cure. *Anaerobe* 2015; **34**: 59–73.

150. Song JH, Kim YS. Recurrent Clostridium difficile Infection: Risk Factors, Treatment, and Prevention. *Gut Liver* 2019; **13**: 16–24.

151. Dubberke ER, Olsen MA. Burden of Clostridium difficile on the healthcare system. *Clin Infect Dis* 2012; **55 Suppl 2**: S88-92.

152. Banaei N, Anikst V, Schroeder LF. Burden of Clostridium difficile infection in the

United States. *N Engl J Med* . 2015. , **372**: 2368–2369

153. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Péchiné S, Wilcox MH, et al. Understanding Clostridium difficile Colonization. *Clin Microbiol Rev* 2018; **31**.

154. Baron SW, Ostrowsky BE, Nori P, Drory DY, Levi MH, Szymczak WA, et al. Screening of carriers in an urban academic medical center: Understanding implications of disease. *Infect Control Hosp Epidemiol* 2020; **41**: 149–153.

155. Khoruts A, Staley C, Sadowsky MJ. Faecal microbiota transplantation for Clostridioides difficile: mechanisms and pharmacology. *Nat Rev Gastroenterol Hepatol* 2021; **18**: 67–80.

156. Ianiro G, Punčochář M, Karcher N, Porcari S, Armanini F, Asnicar F, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. *Nat Med* 2022; **28**: 1913–1923.

157. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 2010; **6**: 390.

158. Diener C, Gibbons SM. More is Different: Metabolic Modeling of Diverse Microbial Communities. *mSystems* 2023; e0127022.

159. Louie T, Golan Y, Khanna S, Bobilev D, Erpelding N, Fratazzi C, et al. VE303, a Defined Bacterial Consortium, for Prevention of Recurrent Clostridioides difficile Infection: A Randomized Clinical Trial. *JAMA* 2023.

160. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* 2017; **35**: 81–89.

161. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al.

American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* 2018; **3**.

162. Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat Metab* 2021; **3**: 274–286.

163. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, et al. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. *Microbiome* . 2015. , **3**

164. Hromada S, Qian Y, Jacobson TB, Clark RL, Watson L, Safdar N, et al. Negative interactions determine Clostridioides difficile growth in synthetic human gut communities. *Mol Syst Biol* 2021; **17**: e10355.

165. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; **37**: 852–857.

166. Marinos G, Kaleta C, Waschina S. Defining the nutritional input for genome-scale metabolic models: A roadmap. *PLoS One* 2020; **15**: e0236890.

167. David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol* 2014; **15**: R89.

168. Lockwood JL, Cassey P, Blackburn T. The role of propagule pressure in explaining species invasions. *Trends Ecol Evol* 2005; **20**: 223–228.

169. Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, et al. Decreased Diversity of the Fecal Microbiome in Recurrent Clostridium difficile—

Associated Diarrhea. *J Infect Dis* 2008; **197**: 435–438.

170. Pakpour S, Bhanvadia A, Zhu R, Amarnani A, Gibbons SM, Gurry T, et al. Identifying predictive features of Clostridium difficile infection recurrence before, during, and after primary antibiotic treatment. *Microbiome* 2017; **5**: 148.

171. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [statML]* . 2018.

172. Ferreyra JA, Wu KJ, Hryckowian AJ, Bouley DM, Weimer BC, Sonnenburg JL. Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. *Cell Host Microbe* 2014; **16**: 770–777.

173. Pruss KM, Enam F, Battaglioli E, DeFeo M, Diaz OR, Higginbottom SK, et al. Oxidative ornithine metabolism supports non-inflammatory C. difficile colonization. *Nat Metab* 2022; **4**: 19–28.

174. Buckley AM, Moura IB, Arai N, Spittal W, Clark E, Nishida Y, et al. Trehalose-Induced Remodelling of the Human Microbiota Affects Clostridioides difficile Infection Outcome in an In Vitro Colonic Model: A Pilot Study. *Front Cell Infect Microbiol* 2021; **11**: 670935.

175. Bouillaut L, Self WT, Sonenshein AL. Proline-dependent regulation of Clostridium difficile Stickland metabolism. *J Bacteriol* 2013; **195**: 844–854.

176. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, et al. Blood metabolome predicts gut microbiome α-diversity in humans. *Nat Biotechnol* 2019; **37**: 1217–1228.

177. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun* 2020; **11**: 5206.

178. Tian H, Chen Q, Yang B, Qin H, Li N. Analysis of Gut Microbiome and Metabolite Characteristics in Patients with Slow Transit Constipation. *Dig Dis Sci* 2021; **66**: 3026–3035.

179. Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, et al. A reference map of potential determinants for the human serum metabolome. *Nature* 2020; **588**: 135–140.

180. Dsouza M, Menon R, Crossette E, Bhattarai SK, Schneider J, Kim Y-G, et al. Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host Microbe* 2022; **30**: 583-598.e8.

181. Nazzal L, Soiefer L, Chang M, Tamizuddin F, Schatoff D, Cofer L, et al. Effect of Vancomycin on the Gut Microbiome and Plasma Concentrations of Gut-Derived Uremic Solutes. *Kidney Int Rep* 2021; **6**: 2122–2133.

182. Horrocks V, King OG, Yip AYG, Marques IM, McDonald JAK. Role of the gut microbiota in nutrient competition and protection against intestinal pathogen colonization. *Microbiology* 2023; **169**.

183. Madi N, Vos M, Murall CL, Legendre P, Shapiro BJ. Does diversity beget diversity in microbiomes? *Elife* 2020; **9**.

184. Liu Y, Dai M. Trimethylamine N-Oxide Generated by the Gut Microbiota Is Associated with Vascular Inflammation: New Insights into Atherosclerosis. *Mediators Inflamm* 2020; **2020**: 4634172.

185. Dalile B, Van Oudenhove L, Vervliet B, Verbeke K. The role of short-chain fatty acids in microbiota-gut-brain communication. *Nat Rev Gastroenterol Hepatol* 2019; **16**: 461–478.

186. Murros KE. Hydrogen Sulfide Produced by Gut Bacteria May Induce Parkinson's

Disease. *Cells* 2022; **11**.

187. Si J, Kang H, You HJ, Ko G. Revisiting the role of Akkermansia muciniphila as a therapeutic bacterium. *Gut Microbes* 2022; **14**: 2078619.

188. Chiantera V, Laganà AS, Basciani S, Nordio M, Bizzarri M. A Critical Perspective on the Supplementation of Akkermansia muciniphila: Benefits and Harms. *Life* 2023; **13**.

189. Zhang Y, Hu J, Tan H, Zhong Y, Nie S. Akkermansia muciniphila, an important link between dietary fiber and host health. *Current Opinion in Food Science* 2022; **47**: 100905.

190. Xiong R-G, Zhou D-D, Wu S-X, Huang S-Y, Saimaiti A, Yang Z-J, et al. Health Benefits and Side Effects of Short-Chain Fatty Acids. *Foods* 2022; **11**.

191. Li Z, Hu G, Zhu L, Sun Z, Jiang Y, Gao M-J, et al. Study of growth, metabolism, and morphology of Akkermansia muciniphila with an in vitro advanced bionic intestinal reactor. *BMC Microbiol* 2021; **21**: 61.

192. Zhang S, Dogan B, Guo C, Herlekar D, Stewart K, Scherl EJ, et al. Short Chain Fatty Acids Modulate the Growth and Virulence of Pathosymbiont Escherichia coli and Host Response. *Antibiotics (Basel)* 2020; **9**.

193. Nishiwaki H, Ito M, Hamaguchi T, Maeda T, Kashihara K, Tsuboi Y, et al. Short chain fatty acids-producing and mucin-degrading intestinal bacteria predict the progression of early Parkinson’s disease. *NPJ Parkinsons Dis* 2022; **8**: 65.

194. Parrish A, Boudaud M, Grant ET, Willieme S, Neumann M, Wolter M, et al. Akkermansia muciniphila exacerbates food allergy in fibre-deprived mice. *Nat Microbiol* 2023; **8**: 1863–1879.

195. Sanchez A, Bajic D, Diaz-Colunga J, Skwara A, Vila JCC, Kuehn S. The community-function landscape of microbial consortia. *Cell Syst* 2023; **14**: 122–134.

196. Hsu RH, Clark RL, Tan JW, Ahn JC, Gupta S, Romero PA, et al. Microbial Interaction Network Inference in Microfluidic Droplets. *Cell Syst* 2019; **9**: 229-242.e4.

197. Neun S, van Vliet L, Hollfelder F, Gielen F. High-Throughput Steady-State Enzyme Kinetics Measured in a Parallel Droplet Generation and Absorbance Detection Platform. *Anal Chem* 2022; **94**: 16701–16710.

198. van Rosmalen RP, Smith RW, Martins Dos Santos VAP, Fleck C, Suarez-Diez M. Model reduction of genome-scale metabolic models as a basis for targeted kinetic models. *Metab Eng* 2021; **64**: 74–84.

199. Praveschotinunt P, Dorval Courchesne N-M, den Hartog I, Lu C, Kim JJ, Nguyen PQ, et al. Tracking of Engineered Bacteria In Vivo Using Nonstandard Amino Acid Incorporation. *ACS Synth Biol* 2018; **7**: 1640–1650.

200. Albright MBN, Louca S, Winkler DE, Feeser KL, Haig S-J, Whiteson KL, et al. Solutions in microbiome engineering: prioritizing barriers to organism establishment. *ISME J* 2022; **16**: 331–338.

VITA

2013	Transfer to University of California San Diego Chemistry program, Fullerton College, Fullerton
2015	B.S. in Chemistry, University of California San Diego, San Diego
2015-2017	Research Associate, Division of Environmental Genomics and Systems Biology at Lawrence Berkeley National Laboratory, Berkeley
2020	M.S. in Molecular Engineering, University of Washington, Seattle
2024	Ph.D. in Molecular Engineering: Data Science, University of Washington, Seattle