Diversity in Chemical Subunits and Linkages: A Key Molecular Determinant of Microbial Richness, Community Structure, Interactions, and Substrate Utilization

Hugh C. McCullough[a,b], Hyun-Seob Song[a,b,c#] and Jennifer M. Auchtung[a,b#]

[a]Nebraska Foods for Health Center, University of Nebraska-Lincoln, Lincoln, NE USA
[b]Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE USA
[c]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE USA

Running title: Chemical complexity impacts microbial community structure

#Address correspondence to Hyun-Seob Song, hsong5@unl.edu and Jennifer M. Auchtung, jauchtung2@unl.edu

Word count, abstract: 231 words

Word count, text: 4247 words

ABSTRACT

Diets play a significant role in shaping the composition and function of microbial communities in the human colon. Although previous studies have shown that a diverse nutrient intake can lead to a richer microbiota that promotes human health, our understanding of the specific chemical traits of nutrients that influence microbial diversity, interactions, and function remains limited. Towards filling this knowledge gap, we conducted *in vitro* experiments by culturing human fecal bacteria under controlled, continuous flow conditions using media that systematically varied in carbohydrate compositions. Our analysis revealed that existing chemical diversity metrics did not sufficiently correlate with microbial richness, leading us to develop a novel measure, termed Chemical Subunits and Linkages (CheSL). The proposed CheSL Shannon Index demonstrated a higher correlation (~0.9) with microbial richness, compared to traditional diversity metrics that show the correlations ranging from 0.41 to 0.65. Additionally, we observed that microbial communities in media with higher CheSL scores exhibited greater peptide utilization, and more connected structure in computationally inferred microbial interaction networks. Together, these findings suggest that the proposed CheSL Shannon Index can serve as a strong indicator of dietary complexity that supports microbial diversity, metabolic potential, and interactions. Furthermore, our work highlights how robust and stable community data can be generated by engineering media composition and structure. These studies provide a valuable framework for future research on microbial community interactions and their potential impacts on host health.

IMPORTANCE

For the human adult gut microbiota, higher microbial diversity strongly correlates with positive health outcomes. This correlation is likely due to increased community resilience that results from functional redundancy that can occur within diverse communities. Dietary recommendations to improve gut microbiota diversity include the addition of microbiota-accessible carbohydrates. However, we lack a mechanistic understanding of how differences in the composition of microbiota-accessible carbohydrates are likely to functionally impact microbiota diversity. To address this need, we developed Chemical Subunits and Linkages (CheSL), a novel measure that describes carbohydrate diversity. Using this measure, we were able to demonstrate how carbohydrate complexity led to changes in microbial diversity and interspecies interactions. Overall, these analyses provide new perspectives on dietary optimization strategies to improve human health.

INTRODUCTION

Members of the gut microbiota are known to contribute to different aspects of human health, including metabolizing otherwise inaccessible dietary substrates, educating the immune system, and providing protection from pathogen infection (1–4). With various evidence of gut microbiota influencing human health, great interest has developed in the past few decades to develop

treatments and interventions to engineer the gut microbiota to improve host health (5–8). The microbiome in the gut is modified through addition of new dietary substrates (e.g., prebiotics and dietary fibers (9–11)), new microbes (e.g., probiotics and live microbial therapeutics (1, 12, 13)), or combinations of substrates and microbes (e.g., synbiotics (2, 14)).

Among factors that contribute to differences in microbial composition, diets are a major contributor (15). In mice, for example, consumption of a westernized diet (high fat, high sugar) is known to cause large changes in microbiome composition compared to consumption of a low fat, low sugar diet enriched in complex polysaccharides (16, 17). Similarly, correlations have been observed between long-term dietary patterns and microbiota composition in humans (18). This study showed that consumption of diets enriched in fat and protein selected for microbial communities with higher levels of *Bacteroides* species, whereas diets enriched in carbohydrates and sugars selected for microbial communities with higher levels of *Prevotella* species. Dietary fibers have long been studied for their roles in modulating gut microbiota composition and function (19, 20). Because these fibers cannot be metabolized by human enzymes, the gut microbiota plays an important role in metabolism of these substrates to products (e.g., butyrate) that can be metabolized by the host (21, 22). The subset of dietary fibers that can be metabolized by the gut microbiota are often referred to as microbiota-accessible carbohydrates (MACs) (17). As these MACs can only be metabolized by a subset of members of the gut microbiota, they play an important role in shaping microbiome composition (17, 21).

The "nutrient-niche theory" proposed by Freter et al. (1983) can serve as a basic framework to understand the impacts of diets on microbiota composition. It states that the metabolic potential of microbial community members and the availability of nutrients dictate how niches are occupied (23, 24). Indeed, prebiotics and synbiotics can exploit nutrient niches directly to either enrich members within a niche or open an entirely new nutrient niche (8, 25). Other work has also shown that nutrient context strongly influences the stability of populations and the ecological interactions among members of microbial communities (17, 26–30), highlighting a key role of the nutrient-niche dynamics in shaping microbiomes.

Despite extensive research in the literature, we still have a limited understanding of the mechanistic interplay between nutrients and gut microbiomes, leaving several questions remaining unanswered, including what chemical traits of nutrients determine microbiota composition and microbial interactions, how the outcomes of diets are affected by initial the nutrient niche and microbiota compositions already existing before treatments, and whether one could find general principles translatable across systems (29, 30). Limitations include difficulties in controlling the entire substrate landscape *in vivo*, which is important, as we will demonstrate in this work that the same substrate can have drastically different effects on microbial composition depending on the relative availability of other substrates. Additionally, *in vivo* it becomes difficult to fully characterize microbial diversity and other sources of variability that may interact with the diet, such as host-derived factors, host behavior, and exposure to new microbiota.

In this study, we aimed to address questions regarding microbial community responses to dietary nutrients by analyzing the response of human fecal microbial communities to changes in the diversity and concentration of MACs through a combination of *in vitro* experiments and computational modeling. Inspired by the nutrient-niche theory, we proposed a novel molecular-level diversity metric of carbohydrates termed Chemical Subunit and Linkage (CheSL), which quantifies nutrient diversity based on their internal linkages and degradation products. The proposed CheSL index demonstrated strong relationships with microbial richness, peptide utilization (measured and predicted through Stickland fermentation), and microbial interactions in computationally inferred networks. Through these combined data analyses and model predictions, we also found that central taxa were associated with more reproducible cultures between replicates and were able to draw parallels between enrichment of specific microbes in the presence of carbohydrate mixtures that were consistent with predicted abundances of metabolic pathways and enrichment previously observed *in vivo.*

RESULTS

Coordinated design of *in vitro* experiments and computational network inference

We designed our study by combining *in vitro* experiments and computational modeling to identify the impacts of MACs on gut microbiomes, particularly on their composition and microbial interactions. Our initial hypotheses were that distinct carbohydrates provide unique niches that would influence microbiota richness and microbial interactions. Our workflow designed to achieve this goal was composed of three parts (**Fig. 1**): (1) data generation and analysis to identify key traits of carbohydrates determining microbial richness, (2) post-analysis of metabolic potential and activity, and (3) network inference to predict microbial interactions and central microbiota

Data was initially generated by culturing microbial communities from the fecal sample of a healthy human (Fecal Sample B (FSB) in continuous-flow minibioreactor arrays (MBRAs) with media that differed in composition and abundance of MACs. MBRAs are a continuous-flow bioreactor model that simulates nutritional conditions of the lumen of the distal colon (anoxic, low nutrient availability, neutral pH) in a parallelized format that allows controlled study of complex microbial communities in the absence of host factors that potentially increase variation (26). Replicate cultures were inoculated into four media (NS, MM, 5MM, and IL) that varied in composition and abundance of MAC but were composed of the same base of proteins, lipids, trace minerals (Table S1). All media contain low levels of the MACs inulin and arabinogalactan; NS contains low levels of mono and disaccharides, MM contains low levels of mucosal monosaccharides and soluble starch, 5MM contains higher levels of mucosal monosaccharides and soluble starch compared to MM, and IL is supplemented with ileostomy effluent, which likely contains a complex mixture of carbohydrates and other factors (Table S1).

Analysis of data demonstrated that community composition varied by media type (Fig. S1A) and that 5MM supported communities with the lowest richness, whereas media IL supported communities with highest richness (Fig. S1B). These results conflicted with our original hypothesis that media with higher carbohydrate richness (measured as the total number of unique carbohydrate compounds present in the growth medium) would support microbial communities with higher richness, as 5MM had higher carbohydrate richness than NS and the same carbohydrate richness as MM. We evaluated other potential measures of carbohydrate complexity (Fig. S1C-2E) and found the strongest linear correlation between carbohydrate Shannon index and microbial richness ($R = 0.91$, $p = 4.5x10^{-5}$; Fig. S1E) followed closely by Shannon evenness ($R=0.9$, $p = 5.5x10^{-5}$, Fig. S1D). An additional fecal sample (Fecal Sample A (FSA) was cultured in three of the four media conditions (NS, MM, and IL), for which similar results can be found in Supplementary Figure 2. Combined data from both fecal samples can be found in Supplementary Figure 3. These two results together suggested that the evenness of carbohydrates within a medium (rather than richness) was a key driver of microbial richness.

Iterative experimental design to reevaluate the impacts of traditional carbohydrate diversity indices on microbial richness

The hypothesis that evenness of carbohydrates within a medium was a key driver of microbial richness was obtained based upon one type of medium with low evenness, 5MM, which had high levels of soluble starch (0.2%; Table S1). Starch is a relatively simple polysaccharide, composed of amylose and amylopectin, and contains only two unique carbohydrate linkages (Table 1) (31). *Bifidobacterium longum* (32) and *Bacteroides thetaiotamicron*, two common members of the gut microbiota, possess enzymes necessary for complete degradation of starch (33) In contrast, the more complex polysaccharide arabinogalactan is composed of five different types of glycosidic bonds that connect primarily arabinose and galactose monomers (Table 1). Neither *B. longum* nor *B. caccae* can fully metabolize arabinogalactan and differences are observed in ability to cleave glycosidic linkages between strains (34).These previous studies indicated that polysaccharides may provide one or more unique niches depending on their complexity and the likelihood that enzymes required for degradation are encoded within a single microbe may vary based upon linkage complexity.

This led us to hypothesize that *both* carbohydrate evenness and carbohydrate complexity (i.e., number of unique linkages in polysaccharides) influence microbial richness. To test this new hypothesis, we designed additional media that varied in carbohydrate evenness and complexity (Table 2). Medium B contained a mixture of carbohydrates at low concentrations as indicated (0.084% w/v, Table 2) and had an evenness measure of 0.947 (Figure 2B) and was similar to NS and MM (Table S1). ILC included a mixture of carbohydrates inferred to be present in IL (35, 36); this medium has an evenness measure of 0.988 (Figure 2B). The remaining four media all had lower levels of evenness but varied in the complexity of the carbohydrate that was uneven. These media were GP (high glucose polysaccharide (starch)), GM (high glucose monosaccharide), AP (high arabinogalactan polysaccharide) and AM (high

arabinogalactan monomers (i.e., galactose and arabinose). We cultured microbial communities from a second fecal sample from a healthy human, Fecal Sample A (FS-A), in these different media in continuous flow MBRA and analyzed changes in microbial community composition.

For this expanded set of media, we observed that there was a weak linear correlation with the richness of the carbohydrates (R = 0.41, p = 4.8×10^{-2}, Fig. 2D) and modest linear correlation with the evenness of carbohydrates (R = 0.65, p = 5.6×10^{-4}, Fig. 2E). These differences were unlikely due to the use of FS-A, as we saw similar levels of diversity for FS-A cultured in NS, MM, and ILC media (Figure S2). The correlation between microbiota richness and carbohydrate Shannon Diversity was weaker than evenness alone (R = 0.62, p = 1.3×10^{-3}, Fig. 4F). Therefore, our results were inconsistent with our earlier hypothesis that carbohydrate evenness (irrespective of carbohydrate complexity) was the primary driver of microbial richness and pointed to our alternative hypothesis that carbohydrate complexity and evenness may both be important for microbial richness.

Diversity in chemical subunits and linkages connects carbohydrate complexity to microbial richness

To quantify the potential contribution of carbohydrate complexity to unique niches within a medium, we developed a novel metric to quantify carbohydrate diversity based on the number of unique free monosaccharides and linkages present in each carbohydrate, termed **Che**mical **S**ubunits and **L**inkages (CheSL) diversity (Fig. 3A). Data on unique linkages and monomers within several polysaccharides was obtained from previously published. Because pathways to import and metabolize monomeric glucose are conserved across many species of bacteria (37), it was assigned a CheSL value of 0, as glucose metabolism was unlikely to provide a unique niche for microbes within our community. Other unique monosaccharides (e.g., arabinose, N-acetylglucosamine, D-glucuronic acid) were assigned CheSL richness values of one, as pathways for their uptake and utilization were not expected to be widely conserved and could represent unique niches. CheSL values for polysaccharides were primarily assigned based upon the number of unique linkages, as the capability to metabolize monosaccharides liberated from polysaccharides was expected to be part of the same unique niche.

The primary equation for calculating CheSL is the fractional composition of carbohydrates and linkages in the medium. To determine this measure for each media, we began by calculating a Media matrix **M** representing media carbohydrate composition.

$$\mathbf{M} = \begin{bmatrix} m_{1,1} & \cdots & m_{1,n_M} \\ \vdots & \ddots & \vdots \\ m_{n_C,1} & \cdots & m_{n_C,n_M} \end{bmatrix} \tag{1}$$

In M, n_c and n_M denote the total numbers of unique carbohydrates and media used in experiments, respectively; the $(j,k)^{th}$ element of **M**, i.e., $m_{j,k}$, denotes the concentration of a specific carbohydrate *j* in a medium *k*.

We then constructed a Linkage matrix, **L**, which represented the linkages present across the carbohydrates under consideration in our media matrix, **M**.

$$\mathbf{L} = \begin{bmatrix} l_{1,1} & \cdots & l_{1,n_C} \\ \vdots & \ddots & \vdots \\ l_{n_L,1} & \cdots & l_{n_L,n_C} \end{bmatrix} \tag{2}$$

In **L**, n_L denotes the total number of unique linkages contained in all carbohydrates; the $(i,j)^{th}$ element of **L**, i.e., $l_{i,j}$, denotes the proportion of linkage *i* within carbohydrate *j*. Therefore, each row *i* represents a linkage and each column *j* represents a carbohydrate's composition.

We used these matrices to calculate $\mathbf{s_k}$, a vector that was the transposed product of our linkage ratio matrix **L**, and $\mathbf{m_k}$, column *k* from the **M** media matrix as shown below.

$$\mathbf{s_k} \equiv (\mathbf{L}\mathbf{m_k})^{\mathrm{T}} \tag{3}$$

Within $\mathbf{s_k}$ were elements, such as $s_{i,k}$, which represents i^{th} linkage/subunit concentration for the kth medium.

We then calculated the fractional proportions of each linkage abundance within the medium *k*. where $f_{i,k}$ represents the fraction of the ith linkage/subunit concentration of the kth medium's total carbohydrate pool.

$$f_{i,k} \equiv \frac{s_{i,k}}{\sum_{i'=1}^{n_L} s_{i',k}} \tag{4}$$

Fractional values can then be used to calculate CheSL Shannon index and CheSL evenness using the equations shown below.

$$CheSL\ Shannon\ Index\ of\ the\ k^{th}\ media\ (CSI_k) = -\sum_{i \in I_{NZ,k}} f_{i,k} \ln\left(f_{i,k}\right) \tag{5}$$

$$CheSL\ Shannon\ Evenness\ Index\ of\ the\ k^{th}\ media\ (CSE_k) = \frac{CSI_k}{\ln(|I_{NZ,k}|)} \tag{6}$$

For these equations, $I_{NZ,k}$ denoted the indices of non-zero elements of $\mathbf{s_k}$, For, Shannon Evenness Index of CheSL we divide CSI_k by the cardinality of the set of non-zero elements of $\mathbf{s_k}$, $I_{NZ,k}$. An example tutorial is available in Supplementary Text.

When we assessed effects of CheSL richness, evenness and Shannon diversity index on microbiota richness, we found that there was a moderate correlation with CheSL richness (R = 0.51, p = 0.01, Fig. 3B), but strong correlations with Shannon diversity (R = 0.90, p = $2.6\text{x}10^{-9}$; Fig. 3C) and evenness (R = 0.92, p = $1.1\text{x}10^{-10}$; Fig. 3D).

CheSL Shannon diversity predicts peptide utilization by medium

To investigate potential functional differences between communities that may be drivers of community assembly or function, we used PICRUSt2 to infer genome content from 16S rRNA gene sequences and STAMP to identify differentially abundant functional pathways between communities cultured in different media. Of note, we observed that taxa potentially encoding genes for L-arginine degradation (Stickland Fermentation; Fig. 4A) and fermentation of pyruvate to butanoate (butyrate; Fig. 4B) were enriched in microbial communities with higher shared richness (cultured in B, GM, AP, and ILC) and lower in microbial communities with lower shared richness (cultured in AM and GP). We hypothesized that differential abundance of these pathways could indicate differences in peptide utilization between communities. Using a commercial assay to measure levels of peptide remaining in spent culture supernatants, we found that peptide utilization varied between communities cultured in different types of media (Fig. 4C). When we performed a correlation between peptide utilization and CheSL, we found a strong correlation ($R = -0.8$, $p = 2.9 \times 10^{-6}$, Fig. 4D), greater than that between peptide utilization and microbiota richness alone ($R = -0.78$, $p = 6.7 \times 10^{-6}$, Fig. S4A) and less than Shannon evenness ($R = -0.82$, $p = 8.7 \times 10^{-7}$, Fig. S4C). From this data, we conclude that either be seeing that with certain community member establishment, peptide can serve as an additional substrate niche or that with greater capability to utilize carbohydrates across multiple niches, more peptide utilization also arises.

Bacteroides and *Lachnospiraceae* are central members in conserved, rich communities

To determine the extent to which variations in microbial and carbohydrate composition influence microbial interactions, we applied LIMITS to our sequence data. LIMITS is a computational algorithm that infers microbial interactions by fitting a discrete Lotka-Volterra model against the temporal profiles of ASV relative abundance data (38, 39). The LIMITS equation was derived by implicitly assuming that total community biomass does not significantly vary in time. Thus, performance is improved when LIMITS is conducted with data collected from communities with minimal variations of the total biomass, such as those found when culturing microbial communities under continuous-flow conditions approaching steady state. However, LIMITS ignores temporal variations of interspecies interactions by assuming interaction coefficients to be constant in time. While relevant in certain cases where microbial interactions are expected not to significantly change, this assumption is invalid in general (40, 41). To address this limitation, we employed a moving window approach by: (1) dividing the data into multiple subsets, each representing a shorter time period; (2) using LIMITS to identify (constant) microbial interactions within these shorter periods; and (3) iteratively inferring the network by shifting the time window to subsequent timepoints. By analyzing the goodness of fit of these interaction networks (mean cross-correlation between inferred networks and observed networks) across time windows, we observed that mean cross-correlation across media types increased after the early phase of *in vitro* culture (Fig. S5).

We initially compared ensemble networks at the ASV level across replicates cultured in the same media. However, we observed very few interactions were conserved. We hypothesized that this may be due to functional redundancy within the gut microbiota from the complex fecal sample allowing different dominant ASVs from the same genus to emerge across replicate communities as they adapt to growth in continuous-flow culture. To test whether this may be the case, we assessed conserved interactions at the genus level, as conserved functions have been previously observed to occur at the genus level in human fecal communities (42). We found that interactions were conserved between replicates within a medium at the genus level.

Among the ensemble models, we saw a distinct trend in the number of nodes, representing interacting taxonomic groups, and the shared richness of the cultures. (Fig. 5A). These networks were not driven solely by correlations between genera, but rather population dynamics fit to generalized Lotka Volterra models by LIMITS (38), as was demonstrated by different network features between media (e.g., AP, ILC, and B). However, it is important to note that not all taxonomic groups were represented in the networks, as not all groups had conserved (or any) interactions detected in some conditions.

From the ensemble models, we found three consistently central taxonomic groups, *Bacteroides*, *Lachnospiraceae*, and *Ruminococcaceae* of unclassified genera (Fig. 5B-F). Particularly *Lachnospiraceae* and *Ruminococcacecae* of unclassified genera exhibit conserved centrality in media with high CheSL Shannon Diversity which produce complex ensemble networks. While the loss of *Lachnospiraceae spp.* in low CheSL Shannon Diversity networks could be explained by differences in abundance between media types (e.g., the negative $\log_2$ fold change in AM and GP compared to B), differences in abundance of *Ruminococcaeceae spp.* could not explain the absence of interactions between communities cultured in different media, as *Ruminococcaceae* spp were not differentially enriched in AM and GP when compared to B. Rather, the presence of absence of interactions with *Ruminococcaceae* spp was due to the consistency of the interactions identified (Fig. 6). When these combined models were compared, we found greater connectivity in conserved networks among media types which also supported the highest microbiota richness. To further support future uses of ILC, we found communities cultured in this medium exhibited more conserved, positive interactions between the three central taxonomic groups mentioned earlier (Fig. 5F).

For media with low conservation of interactions between replicate networks (i.e., GP and AP), we can gain some insights into the effects of media on community composition through differential enrichment analysis. We observed that most taxa decreased in abundance in GP compared to B, except for *Clostridium spp.* and *Bifidobacterium spp.* (Fig 6). For AM, we saw a similar selection against many genera, but strong enrichment of *Collinsella spp.* (Fig 6). Based upon the ensemble network, this enrichment may be driven by *Bacteroides spp.* (Fig. 5B, 6). Alternatively, enrichment may be to the comparatively low ratio of monosaccharides to polysaccharides, as was previously found in human fecal samples (43).

DISCUSSION

In this work, we found that the presence of a mixture of carbohydrates with higher levels of Chemical substrate and linkage (CheSL) diversity and evenness supported growth of human fecal microbial communities with higher richness and diversity. The relative importance of CheSL evenness was contrary to our initial hypothesis that CheSL richness would be the primary driver of microbial richness. We hypothesize that CheSL evenness rather than richness may contribute more strongly to increased microbiota richness due to competition for shared resources. Specifically, when species occupy who occupy different carbohydrate niches compete for shared resources, generation of higher levels of biomass by the species who occupies the carbohydrate niche with higher concentrations may deplete other limiting nutrients that will lead to the extinction of other species unable to compete for this limiting nutrient. Thus, these observations are consistent with nutrient-niche theory (23, 24) and competitive fitness in the context of community stability and carbohydrate diversity.

Degradation of complex polysaccharides is dependent upon the presence of specific carbohydrate active enzymes (CAZymes) within microbial community members. In the absence of enzymes able to break unique linkages within complex polysaccharides, monosaccharides remain inaccessible and cannot provide unique niches. The presence of specific CAZymes which allow degradation of linkages are the gateway to energy for the microbes, and in turn, linkages serve as a representation of the latent energy granting niche space. Thus, it was not entirely surprising when we observed that when more carbohydrate mixtures were tested, CheSL diversity was better able to predict the ability of media to enhance shared richness within a community.

From our observations, among samples cultured in media with the greatest CheSL Shannon index (B, AP, and ILC), there is higher similarity between microbial interaction network structure and shared microbiota richness. This may suggest the diversity of nutrient niches inferred by the abundance and evenness of linkages may control the reproducibility of community assembly from complex microbial samples in continuous flow bioreactors, thereby impacting the shared microbiota richness of the samples. From our findings, the initial assembly of a community around *Bacteroides*, *Lachnospriaceae,* and *Ruminococcaceae* may result in a network of interactions supporting reproducible cultures with greater microbiota diversity.

We also found enrichment and microbial associations previously observed *in vivo* systems exposed to similar treatments– another indicator of our *in vitro* system's capability to support representative communities. Particularly the relationship between *Bacteroides* and *Lachnospiraeceae* are supported through studies demonstrating cross feeding between *Bacteroides thetaiotaomicron* and *Anaerostipes caccae* (44). Additionally, if a culture fails to establish a network with central *Bacteroides* and *Lachnospiraceae*, *Clostridium sensu stricto* tends to become more relevant, and along with it in the case of GP *Bifidobacterium spp.* We the enrichment of *Clostridium sensu stricto* and *Bifidobacterium* has previously been observed in

humans consuming diets in starch aligning with our GP medium which over-represents soluble starch (32, 45, 46).

Further, those communities cultured in media with the lowest CheSL Shannon index (such as AM) lack conserved network structures and can even lead to the formation of microbial communities between replicate reactors that hold variable community composition. We find when these cultures fail to construct communities around *Bacteroides*, *Lachnospiraceae*, and *Ruminococcaceae*, the functional profiles lack Stickland fermentation capabilities. This difference in Stickland fermentation is further supported by a large difference in the peptide concentrations between media types with and without the Stickland fermentation pathways represented. In addition, we find that communities without these stable *Bacteroides*, *Lachnospiraceae*, *Ruminococcaeae* central members, may lack beneficial pathways with potential for improved host health, such as butanoate (butyrate) production. Therefore, through high CheSL diversity we may be able to improve the structure to support more diverse communities capable of beneficial metabolic activity.

While our work demonstrates strong correlations, we do expect there to be limitations in applicability. Particularly that the CheSL metric should be applied across media which are controlled with respect to other substrates. Additionally, it may not be applicable in cases where the inherent genetic potential of a microbiome indicates that counted substrates cannot be utilized. We find that even between our two fecal samples, within the same media there may be higher or lower microbiota richness under the same media due to differences in the functional potential in the initial communities. While our trends are strong and consistent between communities derived from two fecal samples across two experiments, we were unable to test both original fecal samples on the second generation of media (B, AP, AM, GP, GM, and ILC) to more fully validate these observations. We also believe CheSL may have less utility for defined or other communities with lower likelihood of functional redundancy, which may lead to more constraints in community assembly.

Like any ecological set of interactions, interactions between GI microbiota are directly determined by the biotic and abiotic properties of the environment to which the community is exposed. Using approaches that adhere to limitations in data analysis and modeling as described in this work, we gained additional resolution into the relationships between fecal bacteria and the potential interaction landscapes which can be realized from the availability of specific substrates. From progression in process-based approaches to interaction prediction, we further informed and connected population dynamics with metabolic potential and ecological theories of nutrient niche partitioning (23, 24). Using CheSL Shannon Index as a guide when designing *in vitro* culture media has the potential to improve reproducibility between replicate bioreactors and increase the similarity of *in vitro* communities to those found in the fecal samples without introducing unknown dietary components that would be found in media supplements, such as ileostomy effluents or foods that have undergone simulated digestion. Our initial test at an improved carbohydrate medium seemingly supported this potential, with ILC granting similar network

structure to IL and the greatest shared richness of media which had a known substrate concentration. By improving the reproducibility of community assembly in continuous flow conditions, we can better detect differences between samples and reactors driven by other interventions. Additionally, the experimental conditions we used, the improved quality of fit realized using a continuous flow bioreactor, strict filtering, and division of our time series into windows which represent transitory states in the community, all together may aid in inferring interactions for future mechanistic investigation of microbial metabolic interactions.

MATERIALS AND METHODS

Samples used in this study

Fecal samples were previously collected from two adult volunteers who were self-described as healthy and had not consumed antibiotics within the previous two months (26). Samples were homogenized under anaerobic conditions, aliquoted into anaerobic tubes and stored at -80°C until cultivation. Ileostomy effluents were collected from participants who had ostomies placed due to previous illness or injury as previously described (47). Samples were stored at -20°C until use. Informed consent was obtained prior to sample collection according to a protocol approved by the Institutional Review Board of Michigan State University (10-736SM).

Media used in this study

Microbes from fecal samples were cultured in bioreactor basal medium that varied in carbohydrate composition. 1L of basal medium contained 1g tryptone, 2g proteose peptone #3, 2g yeast extract, 0.4g sodium chloride, 0.5g bovine bile, 0.01g magnesium sulfate, 0.01g calcium chloride, and 2ml Tween 80, which was sterilized by autoclaving. 2g sodium bicarbonate, 0.2g inulin from chicory, 0.1g arabinogalactan from larch wood, 0.2 g soluble starch, 0.04g potassium phosphate dibasic, 0.04g potassium phosphate monobasic, and 0.001g vitamin K3 were filter sterilized and added to each liter of basal medium after autoclaving. Different variations of MACs added to basal medium were used to test the impact of carbohydrate composition and concentration on microbial communities (Table S1). NS (No starch) lacked soluble starch and contained 0.15g/L D-cellobiose, 0.15g/L maltose, and 0.04g/L D-glucose in addition to basal medium. MM (mucosal monosaccharides) contained an additional 0.2g/L of soluble starch, 0.05g/L N-acetylglucosamine, 0.05g/L N-acetylneuraminic acid, 0.05g/L D-glucuronic acid, and 0.05g/L L-fucose. 5MM (5X higher mucosal monosaccharides) contained an additional 1.8g/L soluble starch, 0.25g/L N-acetylglucosamine, 0.25g/L N-acetylneuraminic acid, 0.25g/L D-glucuronic acid, and 0.25g/L L-fucose. GP (glucose polysaccharide) contained an additional 1.8g/L starch (2g/L final concentration). GM (glucose monomer) contained an additional 1.96g/L D-glucose (2 g/L final concentration). AP (arabinogalactan polymer) contained an additional 1.9 g/L of arabinogalactan (2 g/L final concentration). AM (arabinogalactan monomers) lacked

arabinogalactan and instead contained the two primary monosaccharides present in arabinogalactan (48)- 1.79g/L galactose and 0.21g/L arabinose. IL (ileostomy effluent) contained an additional 0.2g/L soluble starch, 0.05g/L N-acetylglucosamine, 0.05g/L N-acetylneuraminic acid, 0.05g/L D-glucuronic acid, 0.05g/L L-fucose, and 100 ml of ileostomy effluent. Ileostomy effluent was thawed, pooled from multiple donors, and filtered through a coffee filter to remove particulates prior to sterilization along with base medium by autoclaving. Carbohydrate diversity of ileostomy effluent was estimated from references (35, 36). Finally, ILC (ileostomy carbohydrates) reduced the concentrations of arabinogalactan, inulin, and soluble starch to 0.05 g/L for each of these carbohydrates and also contained 0.05g cellobiose, 0.05g fructose, 0.05g galactose, 0.05g N-acetylneuraminic acid, 0.05g N-acetylglucosamine, 0.05g L-fucose, and 0.05g D-glucuronic acid.

Culturing human GI microbes in continuous flow bioreactors

We used minibioreactor arrays (MBRAs)(26), continuous-flow culture vessels that simulate environmental parameters of the human colon and maintain communities in steady-state based on the carrying capacity of the reactors, to culture communities of human GI microbes from fecal samples. MBRAs are parallelized, therefore, we could culture replicate communities of human GI microbes in medium that varied by carbohydrate composition and concentration and determine how these changes impacted microbial interactions. Prior to inoculation into MBRAs, fecal samples were thawed and homogenized in anaerobic phosphate buffered saline (PBS) as a 20% w/v slurry as previously described (26). In the first experiment, Fecal Sample A (FS-A) was cultured in BBM-NS (n=2 replicates), BBM-MM (n=3 replicates), or BBM-IL (n=3 replicates) and Fecal sample B (FS-B) was cultured in BBM-NS (n=3 replicates), BBM-MM (n=3 replicates), or BBM-5MM (n=3 replicates), or BBM-IL (n=3 replicates). FS-A was cultured in BBM-NS in duplicate (third replicate lost due to contamination), and in triplicate in BBM-MM and BBM-IL; FS-B was cultured in all media in triplicate. Media contamination prevented cultivation of FS-A in BBM-5MM. Both FS-A and FS-B formerly referred to as Fecal Sample B and Fecal Sample C respectively, were previously characterized when cultured in BRM2, a medium highly similar to BRM3, with an 8 hour retention time (26). For these experiments, 16 hours after inoculation, continuous flow was initiated at 0.94 ml/hr (16 hr retention time for 15 ml minibioreactor). 1 ml samples were collected from each culture daily for 10 days as previously described (26).

In the second experiment, FS-A was selected for studies because of higher shared richness across replicate samples compared to FS-B (26), see Figure 3A. FS-A was cultured in BBM-B (BBM (basal)), BBM-GP, BBM-GM, BBM-AP, BBM-AM, and BBM-ILS with n=4 replicates/medium. As before, continuous flow was initiated 16 hr after inoculation at 0.94 ml/hr. 1 mL samples were collected every 12 hours for 8 days. The increased frequency of sampling and number of replicates was to improve our ability to subsequently infer microbial interactions from data. In both experiments, MBRAs were maintained in an anaerobic chamber (Coy Laboratories) heated to 37°C with a gas atmosphere of 90% N_2, 5% CO_2, and 5% H_2.

16S rRNA gene amplicon sequencing

The V4 region of the 16S rRNA gene was amplified directly from mechanically disrupted cells with broad range 16S rRNA primers 515F/805R as previously described (26). Purified amplicons were pooled in equimolar concentration and sequenced using Illumina MiSeq v2 according to Manufacturer's protocol. Sequence data were processed using the mothur V1.35.0 as described. Sequences were aligned using the Silva r132 16S rRNA gene sequence database (49). Any chimeric sequences were removed following identification by UCHIME, and any contaminating sequences identified as chloroplasts, mitochondria, Archaea or Eukarya are removed as well (50). Sequences were denoised using mothur's pre.cluster function, giving ASVs, and taxonomy was assigned using the Silva r138 16S rRNA gene sequence database.

Data pre-processing for computational network inference

Since Lotka-Volterra equations model niche-based effects determining population dynamics, multiple filtering steps were required to mitigate spurious interactions. During the early phases of culture, two factors could influence community stability: the loss of ASV presence as non-viable cells were removed by dilution and partnership switching as communities adapted to the culture conditions. To improve LIMITS inferred model quality of fit, we tested whether a sliding window approach, sampling time regions across the time series, could improve consistency of interactions over time. We tested windows of varying sizes from 6 to 15 time points, across the entire time series. To inhibit performance loss from transient ASVs, we filtered ASVs by minimum presence across all time points within a window, requiring presence of a single read from an ASV across all time points. This pre-processing avoided issues associated with taxa falling and rising above the threshold of detection, which would influence the model similarly to migration and extinction events, and disturb inferences of a fixed ecosystem's interactions (51). From this filtering we obtained what we call a "shared richness" which represents the count of ASVs found consistently across time, excluding ASVs that may be alternating above and below threshold of detection.

Inference of interspecies interactions using LIMITS

LIMITS analysis was performed using the seqtime package in R by Faust et al. (39). Prior to running LIMITS, we preprocessed the data as described above. From there, we applied the LIMITS algorithm using a sliding window approach as described above. Inferred interactions were then used to simulate time series for the given window of time and the resulting simulated data were compared to the experimental data. Mean cross correlations were used to quantify the quality of fit of the simulated data to the experimental data. Specifically, mean cross correlations were calculated by determining the correlation between each taxon's simulated and experimental data over the time series; the mean correlation of all taxa was then calculated; this method of quality of fit analysis was also implemented through the seqtime package (39). This sliding window method granted greater quality of fit than applying LIMITS to the entire experiment's

time series. In addition to improved quality of fit, a sliding window approach allowed us to select the time windows which most likely represent interactions driven by deterministic effects, rather than including data from the first day(s) of cultivation which likely included influences of founder effects, partner selection, and adjustment to the bioreactor system.

From each replicate and window, a network was inferred using LIMITS. To identify conserved interactions, we calculated consensus across networks from replicates of the same fecal sample and media culture conditions. using the same fecal sample inoculum. After calculating average interaction values and filtering by a consensus threshold of 2, meaning at least 2 networks identified the interaction with the same directionality, we obtained ensemble networks for each media type.

Characterization of chemical subunit and linkage diversity

We hypothesized that different carbohydrates present within our media could provide unique niches to support microbial growth. To quantify how differences in carbohydrate composition of our media may provide unique niches to support growth, we developed a measure that we termed Chemical Subunit and Linkage (CheSL) diversity, that calculated the complexity of carbohydrate substrates based upon distinct linkages within di- and polysaccharides or the presence of unique monosaccharides. We used these parameters due to our assumption that the ability to metabolize a monosaccharide liberated from a polysaccharide linkage would be required for an organism to inhabit this niche, therefore no additional weighting would be needed for metabolism of two monosaccharides liberated from unique polysaccharide linkages. In addition, because glucose is a carbon source that is likely to be used by most microbes within our community (37), it was given a weight of 0, as its utilization was unlikely to provide a unique niche for microbial growth. For example, the polysaccharide arabinogalactan is primarily composed of arabinose (80%) and galactose (15%) monomers (48) that are linked in eight unique types linkages within the polysaccharide (Table 1). The relative abundances of these linkages are also known (48), which allows ecological measures of alpha diversity (Richness, Evenness, and Shannon Diversity) to be used to describe the overall complexity of arabinogalactan. We additionally informed the ratios of our linkages for amylopectin from literature, being predominantly α-1–4 linkages, with 5% of linkages being α-1–6 linkages (31). For disaccharides, which have a single linkage type, there is no effective change in diversity when applying CheSL if it is the sole carbohydrate, but it is important to decompose into linkage types in cases where there is overlap. For example, we find that maltose, with α-1–4 linkages, contributes to the same pool as amylose (52). While cellobiose and inulin, have their own unique linkage pools, β(1—4) and β(2—1) respectively (53, 54).These values are used to calculate CheSL evenness, since to calculate CheSL, one must weigh each linkage pool by both the abundance of the linkage in each carbohydrate and the abundance of the associated carbohydrate. Table 1 below shows values for other di- and polysaccharides used within our media. Figure 2 provides a representation of how CheSL diversity can be calculated across a medium containing multiple carbohydrates.

Statistical analysis – comparison of means and correlations

To make comparisons for media shared richness, we used the R package MASS's implementation of a negative binomial GLM, glm.nb, as the richness did not fit a normal distribution determined by the Shapiro-Wilk test, and the data was over-dispersed for a Poisson distribution. Pairwise comparisons were made with the R package emmeans and Tukey's HSD was used to adjust for family-wise error rate.

Inferred metabolic functional potential with PICRUSt

After characterization of conserved interactions and visualization using Cytoscape, we inferred Enzyme Commission Pathway abundance for our cultures from fully sequenced isolates with similar 16S rRNA gene sequences using PICRUSt2 (55). The potential significance of Pfam pathway abundance was assessed using STAMP (56).

Measurement of amylose/amylopectin ratio

To measure the ratio of Amylose/Amylopectin for this study we use a NEOGEN Megazyme Amylose/Amylopectin Assay kit, product code K-AMYL. The assay was performed according to the manufacturer's instructions. This kit was used to determine the proportions of amylose and amylopectin comprising the soluble starch used in various media formulation. This ratio was then used to calculate CheSL diversity measures.

Differential analysis of microbial abundances

We then performed differential abundance analysis with DESeq2 (57) between communities cultured in basal bioreactor medium (B) and BBM with differing carbohydrate composition. The DESeq2 model we used to explain composition as a function of media type, with time as a covariate. Additionally, we performed DESeq2 analysis between communities cultured in media composed of the polymer form (starch (GP), arabinogalactan (AP)) and respective monomer forms (glucose (GM), galactose and arabinose (AM)), as they represent potential differences between monosaccharide-based niches and linkage-based niches at the genus level. Changes in abundance that were identified by DESeq2 as significant and had an absolute fold-change in abundance greater than two were plotted.

Data Availability

16S rRNA gene data is available in the NCBI Sequence Read Archive under BioProject ID PRJNA1146550.

Acknowledgements

The authors thank Dr. Robert A. Britton (Baylor College of Medicine, Houston, Texas USA) for access to sequence abundance data collected in his laboratory by JMA and shown in Figures S1 and S2.

H.C.M and J.M.A. were responsible for conceptualization and design of the studies. Data collection was performed by H.C.M. and J.M.A. Analysis was performed by H.C.M. with guidance provided by H.-S.S. and J.M.A. J.M.A. acquired funding for the studies. The initial draft of the manuscript was written by H.C.M. and J.M.A. and all authors revised and edited the final manuscript. Bioinformatics was done using the Holland Computing Center, supported by the UNL Office of Research and Economic Development and the Nebraska Research Initiative.

Funding

H.C.M. was supported by a graduate fellowship from the Nebraska Food for Health Center. This work was partially supported by NSF grant OIA-2044049 to JMA, and by Nebraska Tobacco Settlement Biomedical Research Enhancement Funds to H.-S.S and J.M.A.

REFERENCES

1. Crovesy L, Ostrowski M, Ferreira DMTP, Rosado EL, Soares-Mota M. 2017. Effect of Lactobacillus on body weight and body fat in overweight subjects: A systematic review of randomized controlled clinical trials. Int J Obes 41:1607–1614.

2. Kok CR, Quintero DFG, Niyirora C, Rose D, Li A, Hutkins R. 2019. An in vitro enrichment strategy for formulating synergistic synbiotics. Appl Environ Microbiol 85:1–15.

3. Tofalo R, Cocchi S, Suzzi G. 2019. Polyamines and gut microbiota. Front Nutr 6:1–5.

4. O'Toole PW, Cooney JC. 2008. Probiotic Bacteria Influence the Composition and Function of the Intestinal Microbiota. Interdiscip Perspect Infect Dis 2008:1–9.

5. Harcombe WR, Betts A, Shapiro JW, Marx CJ. 2016. Adding biotic complexity alters the metabolic benefits of mutualism. Evol Int J Org Evol 70:1871–1881.

6. Litvak Y, Mon KKZ, Nguyen H, Chanthavixay G, Liou M, Velazquez EM, Kutter L, Alcantara MA, Byndloss MX, Tiffany CR, Walker GT, Faber F, Zhu Y, Bronner DN,

Byndloss AJ, Tsolis RM, Zhou H, Bäumler AJ. 2019. Commensal Enterobacteriaceae Protect against Salmonella Colonization through Oxygen Competition. Cell Host Microbe 25:128-139.e5.

7. Simon JC, Marchesi JR, Mougel C, Selosse MA. 2019. Host-microbiota interactions: From holobiont theory to analysis. Microbiome 7:1–5.

8. Van Herreweghen F, De Paepe K, Marzorati M, Van de Wiele T. 2020. Mucin as a Functional Niche Is a More Important Driver of In Vitro Gut Microbiota Composition and Functionality than Akkermansia muciniphila Supplementation. Appl Environ Microbiol AEM.02647-20.

9. Prosky L, Hoebregs H. 1999. Nutritional and Health Benefits of Inulin and Oligofructose Methods to Determine Food Inulin and Oligofructose 1. J Nutr 129:1418–1423.

10. Ahmadi S, Mainali R, Nagpal R, Sheikh-Zeinoddin M, Soleimanian-Zad S, Wang S, Deep G, Kumar Mishra S, Yadav H. 2017. Dietary Polysaccharides in the Amelioration of Gut Microbiome Dysbiosis and Metabolic Diseases. Obes Control Ther Open Access 4:10.15226/2374-8354/4/2/00140.

11. Chleilat F, Klancic T, Ma K, Schick A, Nettleton JE, Reimer RA. 2020. Human Milk Oligosaccharide Supplementation Affects Intestinal Barrier Function and Microbial Composition in the Gastrointestinal Tract of Young Sprague Dawley Rats. Nutrients 12:1532.

12. Lee NK, Kim WS, Paik HD. 2019. Bacillus strains as human probiotics: characterization, safety, microbiome, and probiotic carrier. Food Sci Biotechnol 28:1297–1305.

13. Seekatz AM, Theriot CM, Molloy CT, Wozniak KL, Bergin IL, Young VB. 2015. Fecal microbiota transplantation eliminates Clostridium difficile in a murine model of relapsing disease. Infect Immun 83:3838–3846.

14. Krumbeck JA, Rasmussen HE, Hutkins RW, Clarke J, Shawron K, Keshavarzian A, Walter J. 2018. Probiotic Bifidobacterium strains and galactooligosaccharides improve intestinal barrier function in obese adults but show no synergism when used together as synbiotics. Microbiome 6:121.

15. Friedman J, Alm EJ. 2012. Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol 8:1–11.

16. Yan S, Shi R, Li L, Ma S, Zhang H, Ye J, Wang J, Pan J, Wang Q, Jin X, Liu X, Liu Z. 2019. Mannan Oligosaccharide Suppresses Lipid Accumulation and Appetite in Western-Diet-Induced Obese Mice Via Reshaping Gut Microbiome and Enhancing Short-Chain Fatty Acids Production. Mol Nutr Food Res 63:1–13.

17. Sonnenburg ED, Sonnenburg JL. 2014. Starving our microbial self: The deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab 20:779–786.

18. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–108.

19. Makki K, Deehan EC, Walter J, Bäckhed F. 2018. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. Cell Host Microbe 23:705–715.

20. Van Hul M, Cani PD. 2019. Targeting Carbohydrates and Polyphenols for a Healthy Microbiome and Healthy Weight. Curr Nutr Rep 8:307–316.

21. Hryckowian AJ, Van Treuren W, Smits SA, Davis NM, Gardner JO, Bouley DM, Sonnenburg JL. 2018. Microbiota-Accessible carbohydrates suppress Clostridium difficile infection in a murine model. Nat Microbiol 3:662–669.

22. Chen J, Vitetta L. 2020. The Role of Butyrate in Attenuating Pathobiont-Induced Hyperinflammation 20:1–18.

23. Freter R, Brickner H, Botney M, Cleven D, Aranki A. 1983. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. Infect Immun 39:676–685.

24. Freter R, Brickner H, Fekete J, Vickerman MM, Carey KE. 1983. Survival and implantation of Escherichia coli in the intestinal tract. Infect Immun 39:686–703.

25. Pereira FC, Berry D. 2017. Microbial nutrient niches in the gut. Environ Microbiol 19:1366–1378.

26. Auchtung JM, Robinson CD, Britton RA. 2015. Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). Microbiome 3:1–15.

27. Brand A, Barnes JD, Mackenzie KS, Odds FC, Gow NAR. 2008. Cell wall glycans and soluble factors determine the interactions between the hyphae of *Candida albicans* and *Pseudomonas aeruginosa*. FEMS Microbiol Lett 287:48–55.

28. Yao T, Chen MH, Lindemann SR. 2020. Structurally complex carbohydrates maintain diversity in gut-derived microbial consortia under high dilution pressure. FEMS Microbiol Ecol 96:1–16.

29. Chung WSF, Walker AW, Vermeiren J, Sheridan PO, Bosscher D, Garcia-Campayo V, Parkhill J, Flint HJ, Duncan SH. 2018. Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota. FEMS Microbiol Ecol 95:1–13.

30. Yousi F, Kainan C, Junnan Z, Chuanxing X, Lina F, Bangzhou Z, Jianlin R, Baishan F. 2019. Evaluation of the effects of four media on human intestinal microbiota culture in vitro. AMB Express 9.

31. Ai Y, Jane J l. 2016. Starch: Structure, Property, and Determination, p. 165–174. *In* Caballero, B, Finglas, PM, Toldrá, F (eds.), Encyclopedia of Food and Health. Academic Press, Oxford.

32. Liu S, Ren F, Zhao L, Jiang L, Hao Y, Jin J, Zhang M, Guo H, Lei X, Sun E, Liu H. 2015. Starch and starch hydrolysates are favorable carbon sources for Bifidobacteria in the human gut. BMC Microbiol 15:1–9.

33. Anderson KL, Salyers AA. 1989. Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J Bacteriol 171:3192–3198.

34. Wang Y, Lapointe G. 2020. Arabinogalactan utilization by bifidobacterium longum subsp. Longum ncc 2705 and bacteroides caccae atcc 43185 in monoculture and coculture. Microorganisms 8:1–16.

35. Barrett JS, Gearry RB, Muir JG, Irving PM, Rose R, Rosella O, Haines ML, Shepherd SJ, Gibson PR. 2010. Dietary poorly absorbed, short-chain carbohydrates increase delivery of water and fermentable substrates to the proximal colon. Aliment Pharmacol Ther 31:874–882.

36. Berghouse L, Hori S, Hill M, Hudson M, Lennard-Jones JE, Rogers E. 1984. Comparison between the bacterial and oligosaccharide content of ileostomy effluent in subjects taking diets rich in refined or unrefined carbohydrate. Gut 25:1071–1077.

37. Gil R, Silva FJ, Peretó J, Moya A. 2004. Determination of the Core of a Minimal Bacterial Gene Set. Microbiol Mol Biol Rev 68:518–537.

38. Fisher CK, Mehta P. 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLoS ONE 9:1–10.

39. Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, Gonze D, Widder S. 2018. Signatures of ecological processes in microbial community time series. Microbiome 6:1–13.

40. Song H-S, Lee N-R, Kessell AK, McCullough HC, Park S-Y, Zhou K, Lee D-Y. 2024. Kinetics-based inference of environment-dependent microbial interactions and their dynamic variation. mSystems 9:e01305-23.

41. Song H-S, Lee J-Y, Haruta S, Nelson WC, Lee D-Y, Lindemann SR, Fredrickson JK, Bernstein HC. 2019. Minimal Interspecies Interaction Adjustment (MIIA): Inference of Neighbor-Dependent Interactions in Microbial Communities. Front Microbiol 10:1264.

42. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M, Wang L, Wang G, Wang F, Xu J, Cao H, Xu H, Lv Q, Zhong Z, Chen Y, Qimuge S, Menghe B, Zheng Y, Zhao L, Chen W, Zhang H. 2015. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. ISME J 9:1979–1990.

43. Gomez-Arango LF, Barrett HL, Wilkinson SA, Callaway LK, McIntyre HD, Morrison M, Dekker Nitert M. 2018. Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. Gut Microbes 9:189–201.

44. Chia LW, Mank M, Blijenberg B, Aalvink S, Bongers RS, Stahl B, Knol J, Belzer C. 2020. Bacteroides thetaiotaomicron fosters the growth of butyrate-producing anaerostipes caccae in the presence of lactose and total human milk carbohydrates. Microorganisms 8:1–13.

45. Wang Y, Leong LEX, Keating RL, Kanno T, Abell GCJ, Mobegi FM, Choo JM, Wesselingh SL, Mason AJ, Burr LD, Rogers GB. 2019. Opportunistic bacteria confer the ability to ferment prebiotic starch in the adult cystic fibrosis gut. Gut Microbes 10:367–381.

46. Jung DH, Seo DH, Kim YJ, Chung WH, Nam YD, Park CS. 2020. The presence of resistant starch-degrading amylases in Bifidobacterium adolescentis of the human gut. Int J Biol Macromol 161:389–397.

47. Collins J, Robinson C, Danhof H, Knetsch CW, Van Leeuwen HC, Lawley TD, Auchtung JM, Britton RA. 2018. Dietary trehalose enhances virulence of epidemic Clostridium difficile. Nature 553:291–294.

48. Goellner EM, Utermoehlen J, Kramer R, Classen B. 2011. Structure of arabinogalactan from Larix laricina and its reactivity with antibodies directed against type-II-arabinogalactans. Carbohydr Polym 86:1739–1744.

49. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541.

50. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection 27:2194–2200.

51. Poos MS, Jackson DA. 2012. Addressing the removal of rare species in multivariate bioassessments: The impact of methodological choices. Ecol Indic 18:82–90.

52. Nawaz MA, Pervez S, Jamal M, Jan T, Khan W, Rauf A, Aman A, Qader SAU. 2019. Maltose deterioration approach: Catalytic behavior optimization and stability profile of maltase from Bacillus licheniformis KIBGE-IB4. Biotechnol Rep 24:e00400.

53. French AD. 2017. Glucose, not cellobiose, is the repeating unit of cellulose and why that is important. Cellulose 24:4605–4609.

54. Mensink MA, Frijlink HW, van der Voort Maarschalk K, Hinrichs WLJ. 2015. Inulin, a flexible oligosaccharide I: Review of its physicochemical characteristics. Carbohydr Polym 130:405–419.

55. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. 2020. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38:685–688.

56. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics 30:3123–3124.

57. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

TABLES AND FIGURES

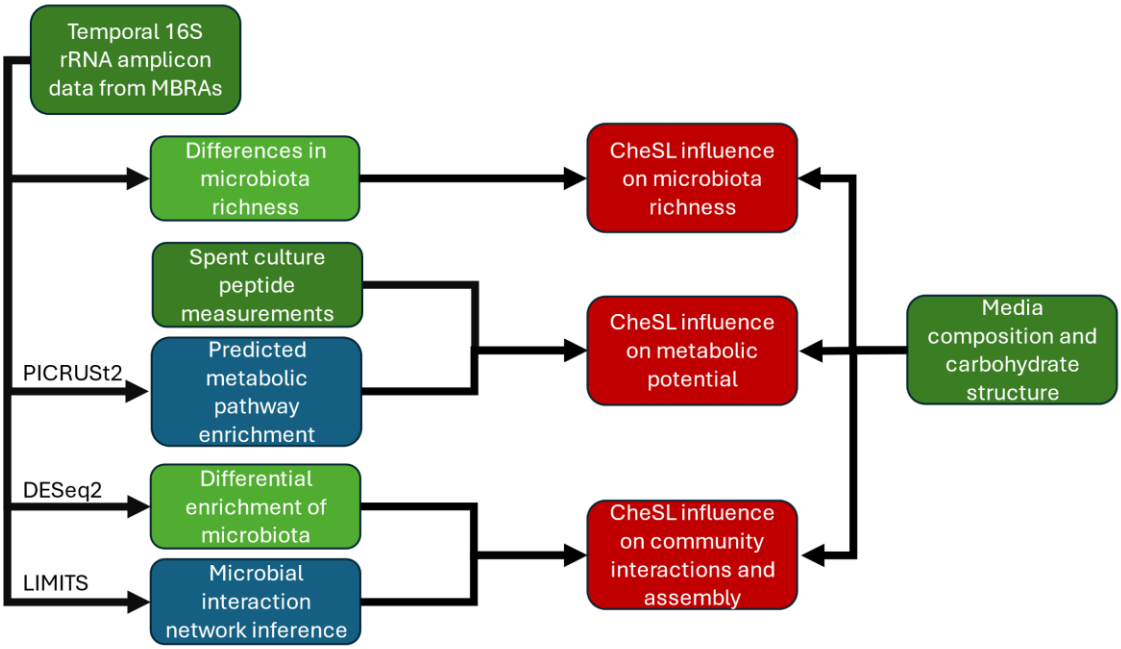


Figure 1: Graphical Abstract A conceptual overview of this study, centered on applicability of CheSL to different community culture traits. Empirical data is shown in dark green, post-analysis traits are shown in light green, inferred characteristics are show in blue, and the interpretations are shown in red.

Table 1: Linkages of Disaccharides and Polysaccharides tested. Linkages from the different polysaccharides used in this study, with estimated linkage abundances from literature, and in the case of starch, fractions of amylose and amylopectin were determined from analysis of our own soluble starch as determined a Megazyme Amylose Amylopectin Assay KitFor linkages in arabinogalactan, t indicates terminal, 3,4,6, refers to which carbon the bond is on, f indicates it is in the furanose form and p, the pyranose form.

Polymeric Carbohydrate	Linkage	Linkage Abundance	Monosaccharide	Reference
Arabinogalactan	Gal tp	.257	Galactose	Goellner et al. (48)
	Gal 3p	.049	Galactose	
	Gal 6p	.258	Galactose	
	Gal 3,4p	.01	Galactose	
	Gal 3,6p	.289	Galactose	
	Ara tf	.045	Arabinose	
	Ara tp	.048	Arabinose	
	Ara 3f	.044	Arabinose	
Starch - Amylose	α-(l-4)	1	Glucose	Ai and Jane (31)
Starch - Amylopectin	α-(l-4)	.95	Glucose	Ai and Jane (31)
	α-(l-6)	.05	Glucose	
Inulin	β(2—1)	1	Fructose	Mensink et al. (54)
Cellobiose	β(1—4)	1	Glucose	French (53)
Maltose	α-(l-4)	1	Glucose	Nawaz et al. (52)

Table 2: Variations in Bioreactor Medium

Carbohydrate (g/L)	**B**[1]	**ILC**[1]	**GP**[1]	**GM**[1]	**AP**[1]	**AM**[1]
Arabinogalactan	0.1	0.05	0.1	0.1	2.0	-
Inulin	0.2	0.05	0.2	0.2	0.2	0.2
Soluble starch	0.2	0.05	2.0	0.2	0.2	0.2
Cellobiose	0.15	0.05	0.15	0.15	0.15	0.15
Maltose	0.15	-	0.15	0.15	0.15	0.15
Glucose	0.04	-	0.04	2.0	0.04	0.04
Fructose	-	0.05	-	-	-	-
Galactose	-	0.05	-	-	-	1.71
Arabinose	-	-	-	-	-	0.29
N-acetylneuraminic acid	-	0.05	-	-	-	-
N-acetylglucosamine	-	0.05	-	-	-	-
L-fucose	-	0.05	-	-	-	-
D-glucuronic acid	-	0.05	-	-	-	-
Total carbohydrate (g/L)	0.84	0.5	2.64	2.8	2.74	2.74

[1]Abbreviations for basal media variations: B: basal medium; ILC: ileostomy carbohydrates; GP: glucose polymer (soluble starch); GM: glucose monomer; AP: arabinogalactan polymer; AM: arabinogalactan monomers.

[2]Concentration of carbohydrates estimated based upon published values for ileostomy effluent (35, 36).

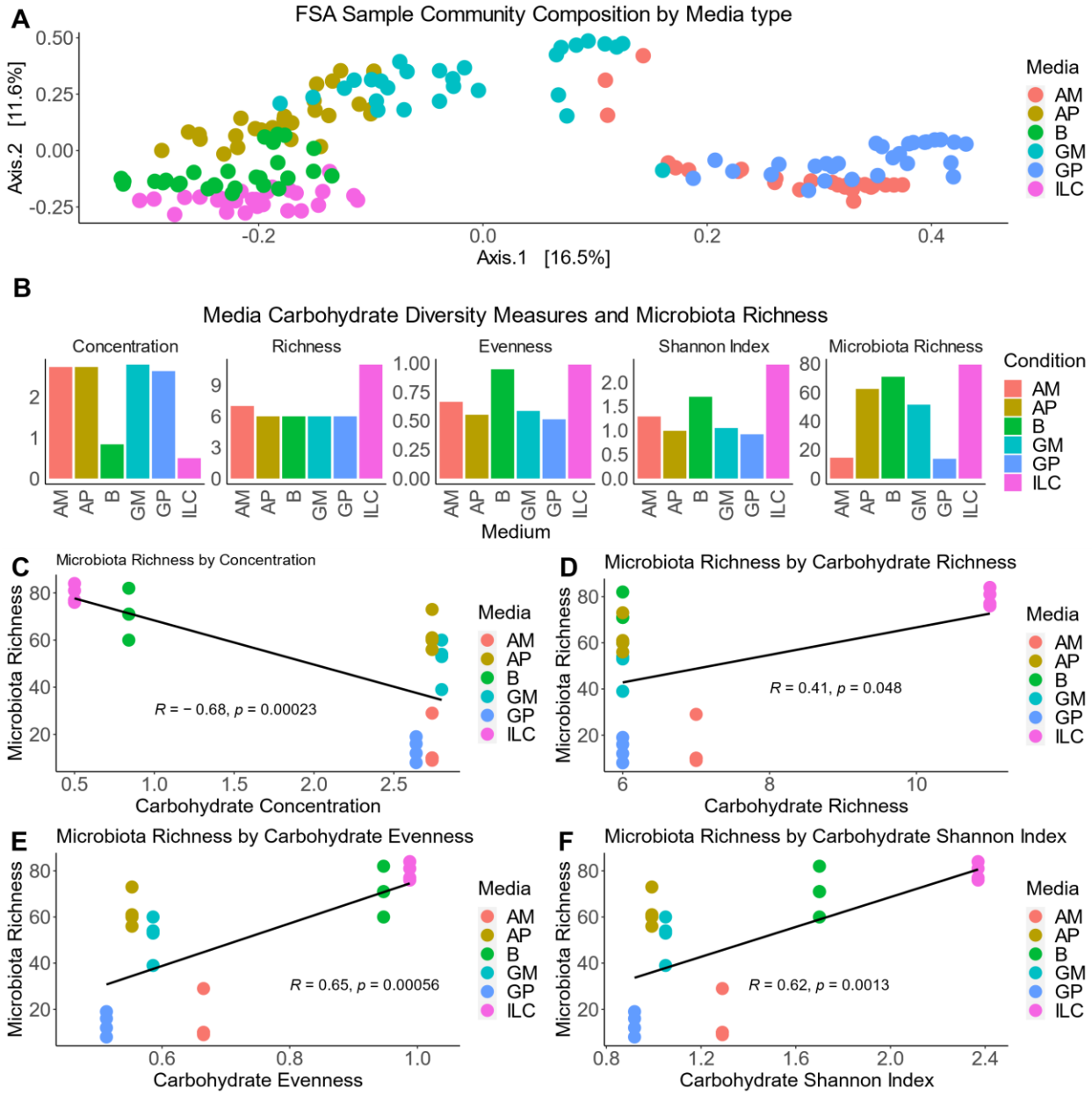


Figure 2: Effects of carbohydrate concentrations, richness, evenness, and diversity on microbiota richness (A) PCoA of Bray-Curtis Dissimilarity of replicate cultures on days 3-10 of Fecal Sample A colored by media type (B) Barplots of carbohydrate concentration (in grams), carbohydrate richness, carbohydrate evenness, carbohydrate Shannon index, and mean microbiota shared richness across replicate reactors for each media type (C) Correlation of microbiota richness with total carbohydrate concentration (R = -0.68, p = 0.00023) (D) Correlation of microbiota richness with carbohydrate richness (R = 0.41, p = 0.048) (E) Correlation of microbiota richness with carbohydrate evenness (R = 0.65, p = 0.00056) (F) Correlation of microbiota richness with carbohydrate Shannon index (R = 0.62, p = 0.0013).

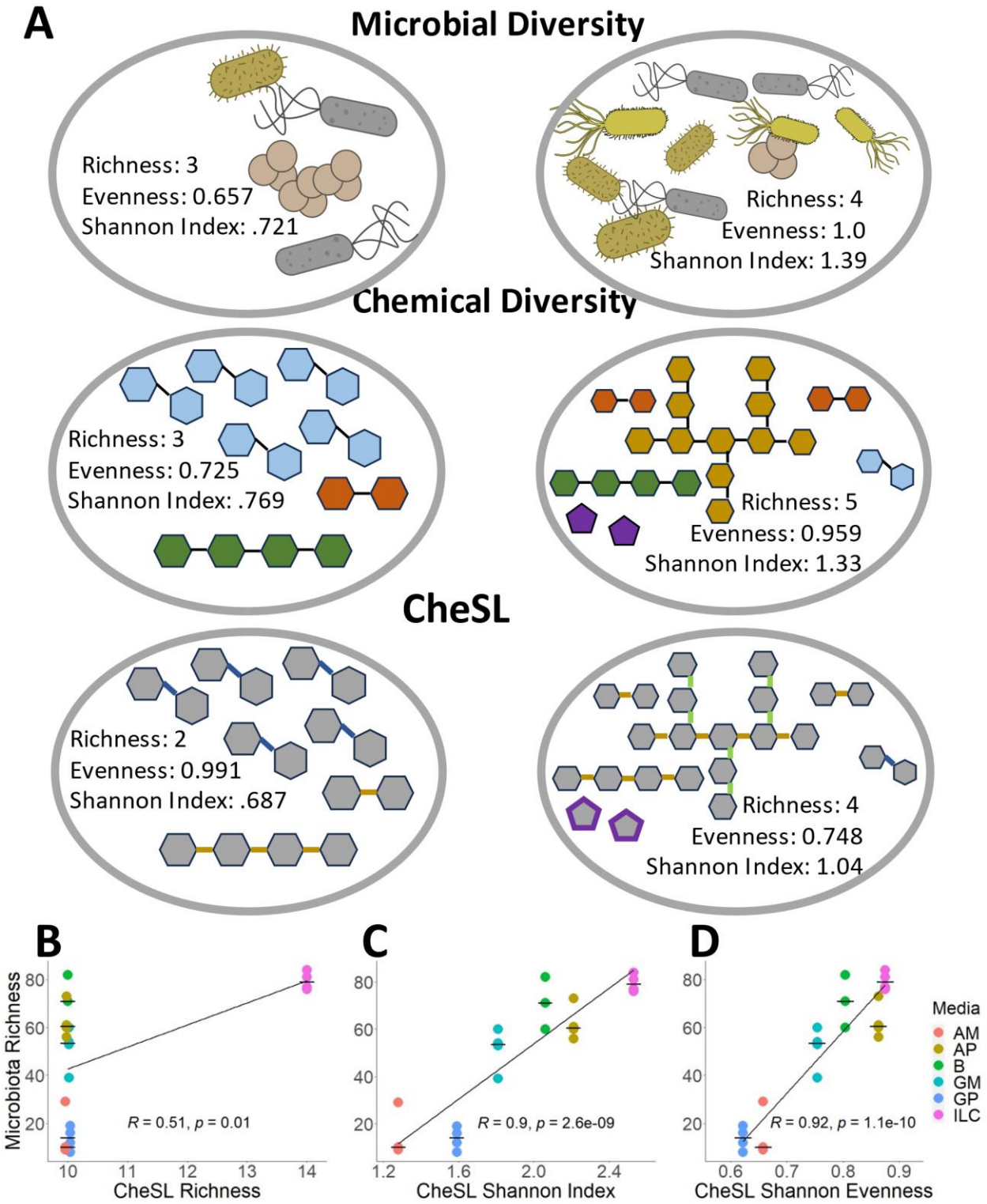


Figure 3: Effects of Carbohydrate Chemical Subunits and Linkages (CheSL) on microbiota richness. A. Graphical representations of approaches to calculate microbial, carbohydrate and CheSL richness, evenness and Shannon Diversity. In these examples, different microbes are represented by different shapes and colors for

microbial measures, different carbohydrates are represented by different colors for carbohydrate measures, and different linkages are color coded with unique substrates represented by different shapes for CheSL measures. Richness, Evenness, and Shannon Diversity are shown for two different cases for each measure. **B.** Correlations of microbiota richness and CheSL Richness (R=0.51, p = 0.01), CheSL Shannon index (R=0.9, p = 2.6×10^{-9}), and Shannon evenness (R=0.92, p = 1.1×10^{-10}).

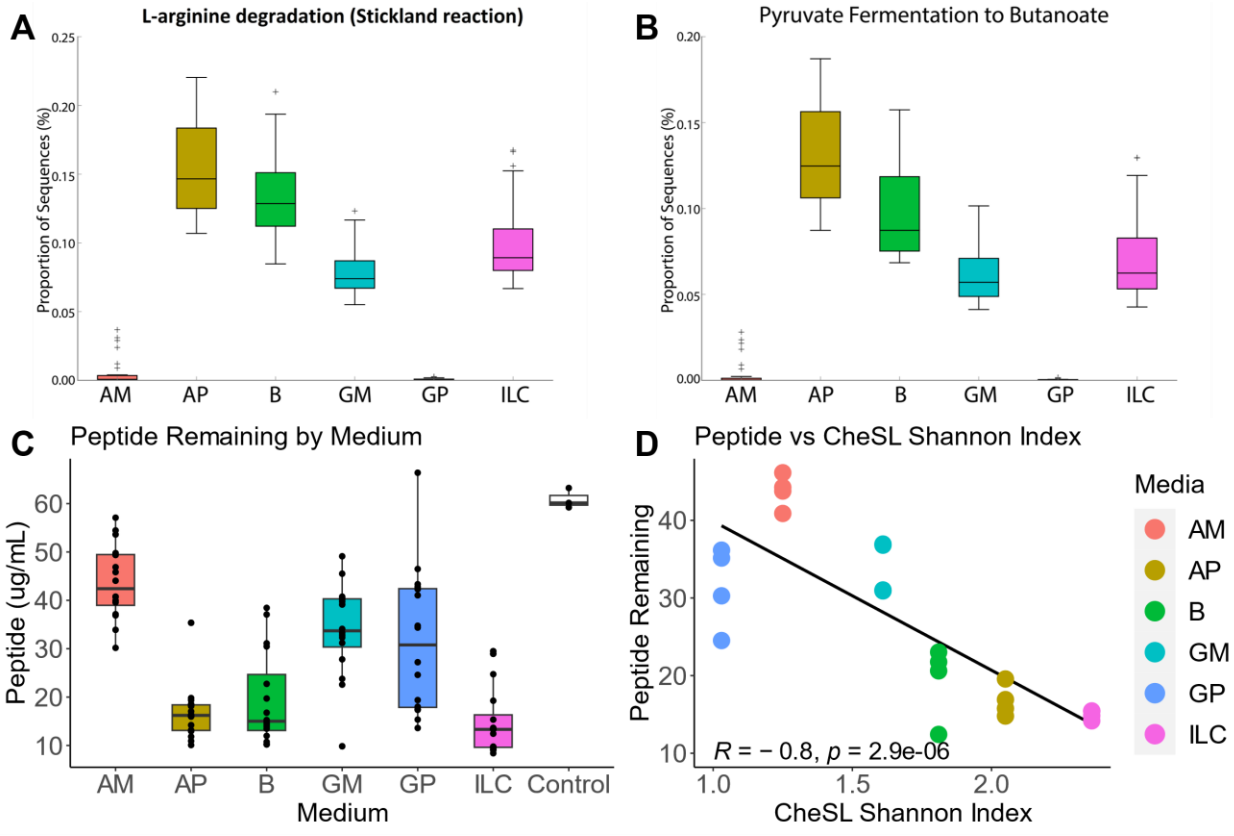
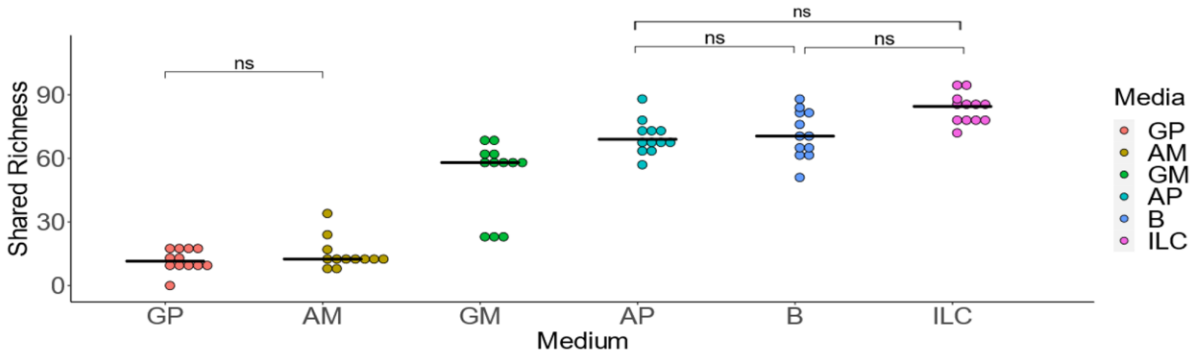
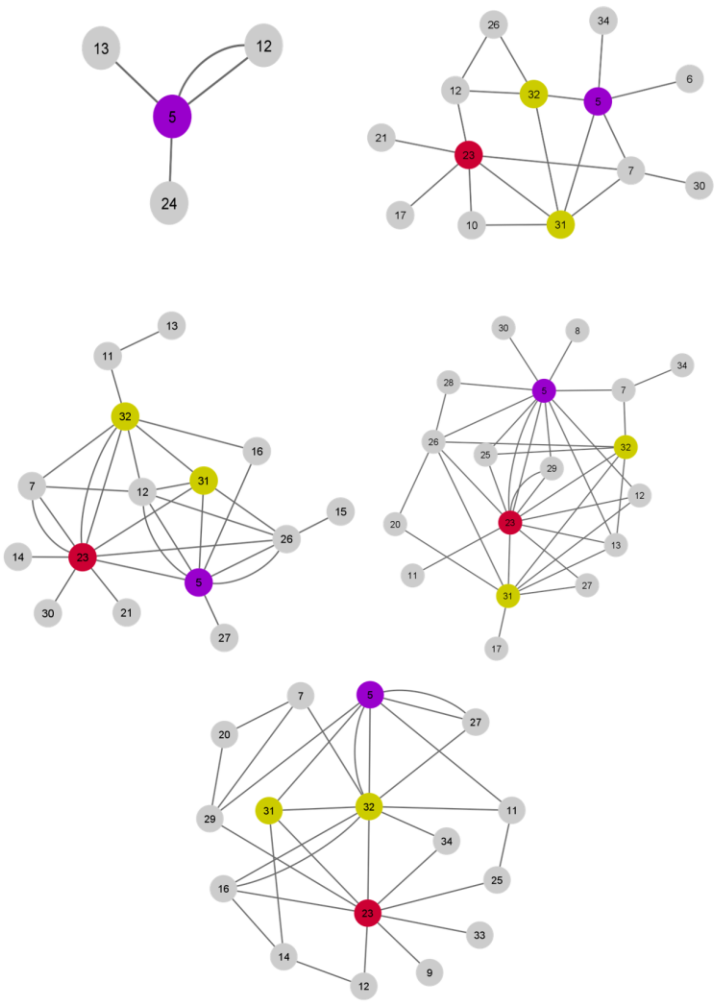


Figure 4: Differentially abundant functional pathways suggested divergent metabolism and production of beneficial compounds (A-B) STAMP generated plots of pathways differentially abundant between media types of high complexity and low complexity cultures (C) Total peptide concentrations measured in steady state culture of each media type exhibiting differences in peptide utilization by communities compared to fresh media control (D) Correlation between CheSL Shannon Index and the peptide concentration in spent medium (R=-.80, p = 2.9×10^{-6})





Node Number	Taxonomy
1	Akkermansia
2	Anaerofustis
3	Anaerostipes
4	Anaerotruncus
5	Bacteroides
6	Bifidobacterium
7	Blautia
8	Clostridiaceae
9	Clostridiaceae_unclassified
10	Clostridiales
11	Clostridiales_unclassified
12	Clostridium
13	Collinsella
14	Coprococcus
15	Dialister
16	Dorea
17	Eggerthella
18	Enterobacteriaceae
19	Enterobacteriaceae_unclassified
20	Erysipelotrichaceae
21	Faecalibacterium
22	Finegoldia
23	Lachnospiraceae
24	Lactobacillus
25	Mogibacteriaceae
26	Oscillospira
27	Peptoniphilus
28	Peptostreptococcaceae
29	Rikenellaceae
30	Roseburia
31	Rumiococcaceae
32	Ruminococcus
33	Sutterella
34	Veillonella

Figure 5: Ensemble Networks of Conserved Interactions demonstrate conserved structure of communities cultured in media that support high shared microbiota richness and within-group similarity (A) Non-significant differences between shared microbiota abundances supported in late culture time windows in each medium displayed as all other are significant with an adjusted p-value < 5e-5 from Tukey's HSD (B) Ensemble

Network from highest quality of fit windows for each replicate of AM, (C) GM, (D) AP, (E) B, and (F) ILC. An ensemble networks for GP could not be determined due to lack of conserved interactions.

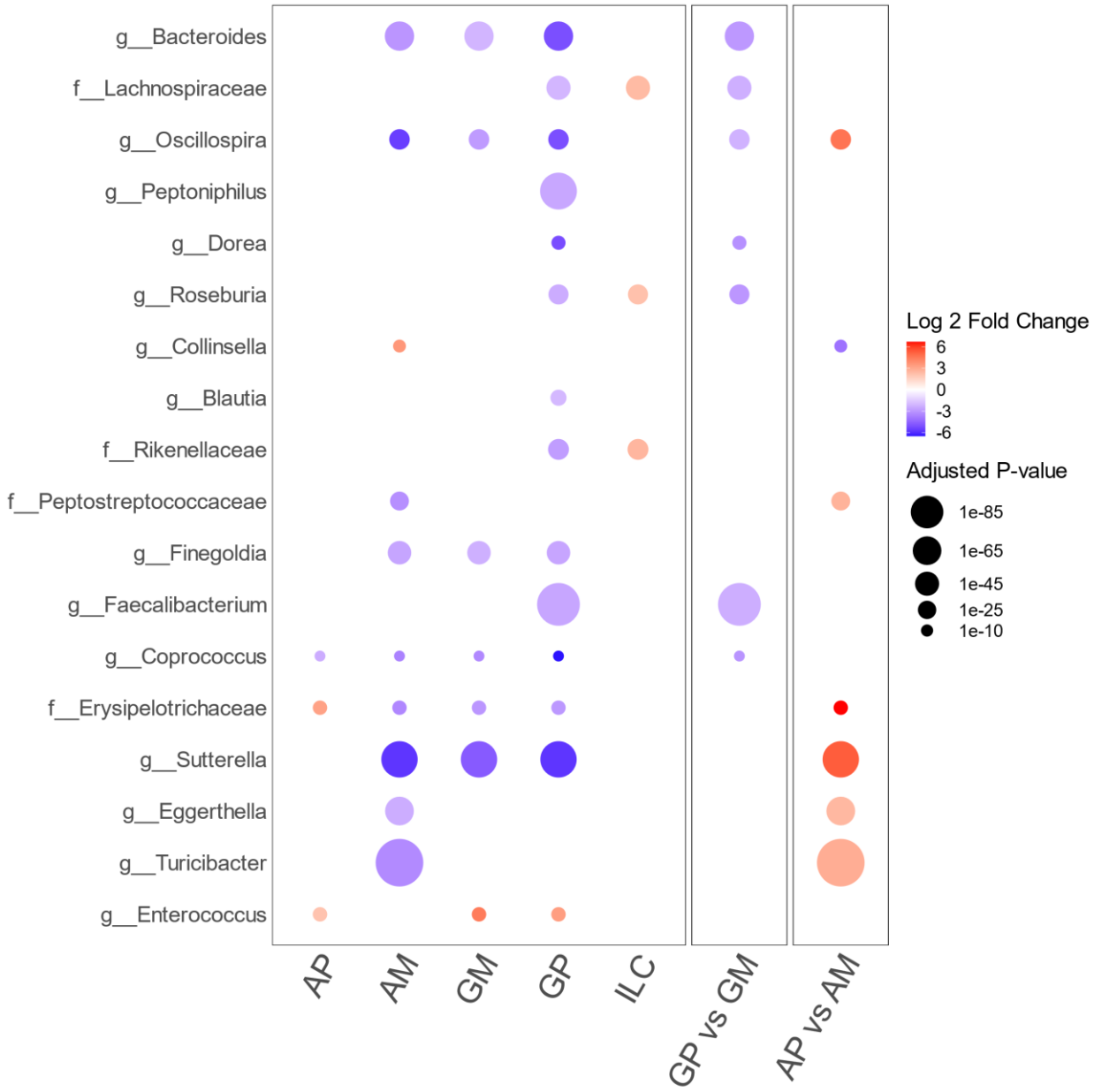


Figure 6: Differential Abundance of Taxa Between FS-A Communities Cultured in Different Types of Media ASVs were binned into shared taxa at the genus level (or lowest level of taxonomic resolution available for taxa whose classification could not be determined at the genus level). Differential abundance of these taxa between communities cultured in different media types was determined using DESeq2. Color indicates the log2 fold change in abundance between cultures and the size of the points indicate the statistical significance of the abundance difference. Unless otherwise indicated, media were compared to medium B as baseline. Comparisons were also

made between communities cultured in overrepresented polysaccharide media (GP and AP) and their respective overrepresented monosaccharide media (GM and AM).