DISSERTATION | DOCTORAL THESIS

Titel | Title

Insights into community composition, function and structure of the human gut microbiome: Keystone species and mucus degraders as case studies

verfasst von | submitted by

Franziska Bauchinger BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Biologie
Betreut von | Supervisor:	Univ.-Prof. David Berry Privatdoz. PhD

"…and all the science, I don't understand
it's just my job 5 days a week…"

"Rocket Man" by Elton John

ACKNOWLEDGEMENTS

Working on and writing this thesis and the publications within would not have been possible without the support of an absolute boatload of people. I am incredibly fortunate and very grateful that I had all of them at my back. And I apologize in advance if I forgot someone. I tried to think of everyone, but it was a long time and I am very tired.

First and foremost, I want to thank my supervisor David. This endeavor most certainly took much longer than either of us anticipated and there were a lot of ups and downs along the way. His continuous trust in my abilities as a researcher was a much-needed counterbalance to my increasing imposter syndrome and the sometimes hare-brained ideas we came up with in our nerdy data analysis discussions, usually some sort of incomprehensible ratio of a ratio, reminded me that research can be fun and inspiring.

I would like to thank the entire gut group, new members and old, for being a lovely, funny, supportive research community inside and outside of the lab. Thank you to Dimitri, Jovana and Fatima for great discussions and lots of encouragement and to Annelieke and Huimin for always having a sympathetic ear. A special thank you goes to David Seki for his scientific expertise and his steadfast friendship, both of which were invaluable.

A massive thank you and many, many hugs go to my lovely pack. We started out as a bunch of people who shared an office and now, although we all have moved on to a different office, building, city or even country, we still celebrate Thanksgiving together. They are a special kind of crazy and luckily, I fit right in.

Over the last couple of years, I have successfully transitioned more and more into administrative and organizational work and I have my incredible colleagues in the admin team to thank for that. I started out with little to no experience and their support was and still is very much needed and appreciated. They are just genuinely fantastic people and I am grateful and happy that they welcomed me into their midst. Working with them kept me sane when my thesis was about to drive me crazy.

The most special “Thank you” goes to the best colleague of them all, my one and only work wife Nadja. Not only is she a great friend and colleague, she is also the most supportive doctoral school coordinator imaginable. I am beyond lucky to work so closely with someone that I admire that much and whose friendship means so much to me.

I want to extend my thanks to all of DOME and CeMESS. I have now spent a lot of years in this community of people and watched it grow and learn as I grew and learned with it. I met so many great people during this time and had so many incredible conversations, from the smallest small talk to the deepest discussion. Thank you in particular to Geli, Cristi, Andrew,

Steph, Stephan, Petra, Laura, Sandra, Kerrin, Julia, Sean, Eva, Mark, Patrick, Dima and Ben for good talks, fun times and much appreciated support.

Thank you also to Claudia, Lukas and Bernadette. I am convinced that we were peak WG awesomeness and their friendship is incredibly special to me.

Many thanks go to my friends and extended family back in upper Austria. They are all a loud and crazy bunch and I am forever grateful to have them.

Lastly, I want to thank my father, mother, sister and brother. I genuinely do not know how I got so lucky to call them my family. They have supported me in everything I've done and helped me back up whenever something knocked me down. Their unconditional love is at the basis of everything I have achieved.

Papa, Mama, Desi und Joesi, danke für eure ununterbrochene Unterstützung und eure bedingungslose Liebe. Ich bin jeden Tag dankbar, euch in meinem Leben zu haben.

TABLE OF CONTENTS

SUMMARY

The human gastrointestinal tract is inhabited by a diverse community of microorganisms that together form the human gut microbiome. These microbes play a pivotal role in human health, from training our immune system in early development to metabolizing dietary compounds or providing colonization resistance to pathogens. The composition of this microbial community can vary strongly between individuals and is influenced by host factors such as diet, lifestyle or genetic background. Despite these compositional differences and the environmental fluctuations, the gut microbiome is able to maintain its important functionality. This ecological interplay between individual microbes, community dynamics and host influence leading to the observed complexity still leaves many open questions.

In this thesis I apply data analysis approaches to elucidate some of these aspects. I focus on microbial community composition, transcriptional profiling, and network analysis and use correlation network keystone taxa as well as the mucosal glycan degrading community as case studies. Keystone taxa are thought to have a disproportionately strong effect on community structure and consequently attracted a lot of interest in gut microbiome research, in particular in a medical context. Frequently, correlation network analysis is employed to identify keystone taxa in microbial communities, but the ecological significance of such keystones is still unclear. In my work on correlation network keystone taxa I find that such networks are highly sensitive to taxonomic resolution and that putative keystone taxa are only observable at high resolution. These keystone taxa seem to act as stabilizing agents in individual clusters of co-occurring taxa and exhibit distinct transcriptional states. The mucosal glycan degrading community degrades the loose outer mucosal layer produced by gut epithelial cells. While it has been shown that a large number of gut microbial taxa encode genes to degrade mucosal glycans, the *in vivo* dynamics of this community are less well known. In my research I find that in fact only a small proportion of species that encode the necessary genes actively transcribe them and that the transcription of individual genes is frequently dominated by a single species in a given sample. I further find apparent differences in the proportion of abundance-driven transcription which suggests the presence of competitive and opportunistic degraders.

My analyses revealed interesting similarities between the keystone taxa and mucosal glycan degraders, such as a prevalence of negative associations between species and compositional gradients across communities. However, different ecological principles may underlie the observed dynamics and further research is needed to untangle this complexity.

ZUSAMMENFASSUNG

Der menschliche Verdauungstrakt dient einer diversen Gemeinschaft von Mikroorganismen als Habitat, die zusammen das menschliche Darmmikrobiom bilden. Diese Mikroben spielen eine zentrale Rolle für die menschliche Gesundheit, von der Ausbildung unseres Immunsystems in der frühen Entwicklung bis hin zur Verstoffwechselung von Nahrungsbestandteilen oder der Kolonisationsresistenz gegenüber Krankheitserregern. Die Zusammensetzung dieser mikrobiellen Gemeinschaft kann von Mensch zu Mensch stark variieren und wird durch Wirtsfaktoren wie Ernährung, Lebensstil oder genetischen Hintergrund beeinflusst. Trotz dieser Unterschiede in der Zusammensetzung und der Fluktuationen in der Umwelt ist das Darmmikrobiom in der Lage, seine wichtigen Funktionen aufrechtzuerhalten. Dieses ökologische Zusammenspiel zwischen individuellen Mikroben, Dynamiken der mikrobiellen Gemeinschaft und Wirtseinfluss, das zu der beobachteten Komplexität führt, lässt aber noch viele Fragen offen.

In dieser Arbeit wende ich Ansätze der Datenanalyse an, um einige dieser Aspekte zu beleuchten. Ich konzentriere mich auf die Zusammensetzung mikrobieller Gemeinschaften, die Erstellung von Transkriptionsprofilen und Netzwerkanalysen und verwende Schlüsselarten in Korrelationsnetzwerken sowie die Glykan abbauende Gemeinschaft in der Darmschleimhaut als Fallstudien. Es wird angenommen, dass Schlüsselarten einen unverhältnismäßig starken Einfluss auf die Struktur der Gemeinschaft haben und sie stoßen daher in der Darmmikrobiomforschung, insbesondere im medizinischen Kontext, auf großes Interesse. Häufig werden Korrelationsnetzwerkanalysen eingesetzt, um Schlüsselarten in mikrobiellen Gemeinschaften zu identifizieren, aber deren ökologische Bedeutung ist noch unklar. In meiner Arbeit an Schlüsselarten in Korrelationsnetzwerken stelle ich fest, dass die Struktur dieser Netzwerke stark von der verwendeten taxonomischen Ebene abhängig ist und potentielle Schlüsselarten nur auf Gattungs- und Artenebene identifizierbar sind. Diese Schlüsselarten scheinen eine stabilisierende Wirkung auf Gruppen gemeinsam vorkommender Arten zu haben und weisen unterschiedliche Transkriptionszustände auf. Glykan abbauende Mikroorganismen zersetzen die lose äußere Schicht der Darmschleimhaut, die von den Darmepithelzellen gebildet wird. Während bereits gezeigt wurde, dass eine große Anzahl mikrobieller Taxa im Darm Gene zum Abbau von mukosalen Glykanen kodieren, ist die *in vivo* Dynamik dieser Gemeinschaft weniger gut bekannt. In meiner Forschung stelle ich fest, dass tatsächlich nur wenige Arten, die die erforderlichen Gene kodieren, diese auch aktiv transkribieren und dass die Transkription einzelner Gene in individuellen Proben häufig von einer einzigen Art dominiert wird. Darüber hinaus stelle ich fest, dass sich Arten im Anteil der Abundanz abhängigen Transkription unterscheiden, was auf die Existenz von kompetitiven und opportunistischen Zersetzern schließen lässt.

Meine Analysen zeigen interessante Gemeinsamkeiten zwischen Schlüsselarten und den Darmschleimhaut abbauenden Mikroben auf, wie z. B. eine Prävalenz negativer Assoziationen

zwischen den Arten und Gradienten in der Zusammensetzung von Gemeinschaften. Den beobachteten Dynamiken könnten jedoch unterschiedliche ökologische Prinzipien zugrunde liegen und es bedarf weiterer Forschung, um diese Komplexität ausführlicher zu untersuchen.

(Übersetzung unterstützt durch DeepL - deepl.com)

INTRODUCTION

My scientific journey has been a meandering one. I started studying thinking I would go into behavioral biology but soon discovered that my idea of this field had little to do with reality. Realizing that relationships between organisms and their environment and the resulting interdependencies interested me most, I first moved into ecology and then into environmental sciences. Here I came into contact with mathematical modelling in the form of community modelling and modelling of groundwater flows. The concept of abstracting and simplifying the natural world yet still recovering similar dynamics fascinated me. I got the chance to dive into microbial community modelling during my Master's thesis and worked on an individual based model of the human gut microbiome. My PhD then started out in a very similar fashion and I also worked on community models of the human gut microbiome. However, as is the nature of research, I hit a bit of a dead end and it became apparent that I needed to change direction or at least focus. My research started to move more and more towards data analysis and data science. The more I started to work with big datasets the more intrigued I became. There is a wealth of information contained in all this data researchers constantly produce, one just needs to uncover it. In particular in microbial ecology the amount of data that is nowadays accessible to us is tremendous and the possibilities it offers seem endless. And yes, if you look across a lot of samples and time points and maybe even environments you miss a lot of details, but it can also reveal general patterns and overarching dynamics. Ultimately, I think it takes both a detailed and a coarse-grained point of view to thoroughly understand any system, but personally I have always been a big picture kind of person. This brings me right back to my fascination with abstraction and simplification. In order to extract information out of vast amounts of data, you need to think in an abstract manner and you need to simplify. You need to twist and turn the data, compress it or change the angle to find your view of interest.

While my thesis has a methodological emphasis, it is also very much about the human gut microbiome. Partially it was pure coincidence that my enthusiasm for modelling and data analysis brought me to this particular research field. On the other hand, it did not take much to pique my interest either. After all, we are all pretty much personally concerned with this topic on a daily basis. Gut microbiome research is a massive field, especially in humans, which is highlighted by the sheer number of publications. The peak was thus far reached in 2024 with 8816 hits on PubMed (https://pubmed.ncbi.nlm.nih.gov) linked to the search term "human gut microbiome". Despite this immense research effort there are still countless open questions, already starting at the basic definition of a healthy gut microbiome (Shanahan et al., 2021). Five years ago, Li and colleagues even made the argument that the research field is still in its infancy (Li et al., 2020). The gut microbiome exists in a complex relationship with its host and both sides in this relationship are reacting and adapting to each other (Nichols and Davenport, 2021; Peluzio et al., 2021; Corbin et al., 2023; Xiong et al., 2023). This microbial community is spatially heterogeneous (Earle et al., 2015; Tropini et al., 2017;

McCallum and Tropini, 2024) and varies considerably between people (Huttenhower et al., 2012; Chen et al., 2021; Kurilshikov et al., 2021). Gut microbes can also show different behavior and functionality in different communities and contexts (Dey and Ray Chaudhuri, 2023; Weiss et al., 2023). These characteristics make an ecological mapping of the human gut microbiome very challenging. The research community meets this challenge with a combined effort that utilizes a multitude of approaches and techniques that reach, to name a few, from *in silico* metabolic models (Quinn-Bohmann et al., 2024) to *in vitro* co-cultivation experiments (Weiss et al., 2023), *in situ* visualization techniques (Brugiroux et al., 2022), mouse models (Brugiroux et al., 2016) or even ingestible biosensing systems (De la Paz et al., 2022). In my own little corner of this large research body I focused on a few data science approaches that I applied to two research projects. The now ubiquitous term data science for a unique scientific field differing from statistics emerged in the 1980s and was heavily based on what John Tukey simply termed "data analysis" in the 1960s (Donoho, 2017). At its core data science and data analysis is the extraction of knowledge from data (Dhar, 2013), though the methods of extraction and the nature of the data are incredibly varied and can be very specific to a field. In microbiome research we frequently deal with "omics" data, such as metagenomics or metatranscriptomics. We can use this data to further our understanding of how microbiomes are assembled and structured and how they interact with and react to their environments. Here I highlight three data analysis approaches that can utilize the information contained in microbiome data to answer ecological questions.

Microbial community composition and stratification

The human gut microbiome is a diverse microbial community that shows high interindividual variability (Huttenhower et al., 2012). Its heterogeneity and complexity make it challenging to infer general ecological principles that govern its observed dynamics. One approach to tackle this problem is to cluster communities based on properties such as similarity and thereby reduce the complexity. The concept of microbial community stratification has gained a lot of traction in the last 15 years. The introduction of enterotypes by Arumugam and colleagues (Arumugam et al., 2011) sparked a lot of interest, but also controversy. Enterotypes are thought to represent preferred states of the human gut microbiome that stratify the large space of microbial community composition. However, they are likely not discrete community clusters but rather local optima in a continuous compositional space (Knights et al., 2014; Costea et al., 2018). These recurring compositional patterns are of interest both from an ecological as well as a medical perspective. Ecologically the preferred community compositions may represent optimized states of community effectiveness and deviations from these optima may be imposed by the environment, i.e. the host (Costea et al., 2018). This structure in the compositional space has implications on properties such as resilience and the effect of perturbation. For example, small perturbations such as short-term dietary interventions generally seem to not alter the enterotype (Wu et al., 2011) while antibiotic treatment, a considerably stronger interference into community composition, frequently leads to altered states (Dethlefsen and Relman, 2011). From a medical perspective,

enterotype signatures have been associated with various diseases (Jernberg et al., 2010; Larsen et al., 2010; Zeller et al., 2014) but stratification of patients can also lead to the discovery of medically relevant associations (Claesson et al., 2012; Schubert et al., 2014). More recently Frioux and colleagues identified so-called enterosignatures, co-existing clusters of gut microbial species that together describe a given community (Frioux et al., 2023). They observe diverse metabolic potentials between these enterosignatures that are often complementary and suggest that the established framework can be used to identify dysbiotic states. In a study on pigs Larzul and colleagues investigated the effect of host genetics on enterotype assembly and found that, through specific selection over three generations, the selected enterotype increased in relative abundance (Larzul et al., 2024). The authors suggest that pressure on the host genetics translates into selection for an enterotype as a whole and that enterotypes are functional ecosystems. Combined, this research suggests the existence of an ecologically informative and relevant gradient across gut microbiomes that can be characterized through compositional analysis.

Transcriptional patterns and profiling

Apart from composition the functional potential and versatility of microbiomes adds another level of complexity to understanding these systems. The functional potential of the gut microbiome is generally considered to be more stable across individuals than the community composition and substantial functional redundancy is encoded in the metagenome (Huttenhower et al., 2012; Zhernakova et al., 2016). Interestingly, although a metatranscriptomic "core" that is consistently transcribed can be identified, this core seems to be distributed among different microorganisms between hosts and transcriptional patterns beyond this core are much more dynamic and context-dependent (Abu-Ali et al., 2018). Franzosa and colleagues liken the human microbiome to our own genomes in that the genomic content is significantly similar across people and transcriptional control adapts it to distinct ecological niches (Franzosa et al., 2014). Consequently, investigating transcriptional patterns is a crucial step in understanding *in vivo* dynamics. Metatranscriptomic analysis for example revealed that microbial gene expression recovers faster from dysbiosis in a colitis mouse model than community composition (Schwab et al., 2014). Morgan and colleagues contrastingly found microbial function to be more persistently perturbed than microbial composition in inflammatory bowel disease (Morgan et al., 2012). Similarly, Schirmer et al. showed that certain microbial characteristics associated with inflammatory bowel disease were not detectable in metagenomic data, but only on transcript level (Schirmer et al., 2018). In a study on pediatric asthma that analyzed transcription in both host and microbiota of the respiratory airways metabolic functions showed greater differences between health and disease than microbial composition (Pérez-Losada et al., 2015). When investigating the human oral microbiome Nowicki and colleagues found that significantly differentially expressed virulence-related transcripts were associated with early stages of a transition to gingivitis (Nowicki et al., 2018). While it is by no means a new insight that functional analysis plays a crucial part in understanding microbial communities, I would like to emphasize that in

particular large-scale transcriptional profiling can reveal interesting dynamics that purely compositional analyses fail to detect.

Network construction and analysis

Lastly, I would like to point to network analysis as a powerful tool in understanding microbial communities and the human gut microbiome in particular. Luo and colleagues differentiate between four general types of networks in gut microecology: co-occurrence networks that use co-abundance to estimate associations and interactions, causal networks that often follow a concept of observation, intervention and counterfactual analysis to detect cause-effect relationships, dynamic networks that aim to portray changes in network structure and connectivity through the usage of time-resolved data, and multi-omics networks that aggregate and integrate data from multiple sources (Luo et al., 2024). From a more practical and applied point of view the type of data needed to construct a network is often a determining factor, as abundance data is still the most frequently available. This abundance data may need to be time-resolved for certain algorithms. Stein and colleagues, for example, use time-resolved metagenomics data to infer microbial interactions and assess community stability (Stein et al., 2013). They extend generalized Lotka-Volterra equations to include perturbations and use it to model antibiotic-mediated *Clostridium difficile* infections. The LIMITS algorithm by Fisher and Mehta employs sparse linear regression on metagenomic time series and infers a discrete-time Lotka-Volterra model (Fisher and Mehta, 2014). They then use this approach to detect keystone species in two gut microbial communities. When time-resolved data is not available researchers often turn to correlation or co-occurrence network analysis. Broadly speaking a correlation network infers association between microbial taxa if they co-occur across communities more or less frequently than expected stochastically, but the specific algorithms and applications can of course vary. For example, SparCC (Sparse Correlations for Compositional Data), a popular tool for inferring correlations (Friedman and Alm, 2012), and its parallelizable C++ implementation FastSpar (Watts et al., 2019) utilize log-ratio transformed abundance data and a bootstrapping approach and thereby also tackle the inherent issues of compositional data that can result in the detection of spurious correlations. Another established network inference algorithm, SPIEC-EASI (sparse inverse covariance estimation for ecological association inference), uses inverse covariances to address these issues (Kurtz et al., 2015). These and other algorithms have been widely applied in gut microbiome research to detect similar community structures in different human populations (Jackson et al., 2018), map correlation networks across human body sites (Faust et al., 2012), assess spatial distribution patterns of the mucosa-associated gut microbial community (Zhang et al., 2014) or to delineate healthy and diseased association networks (Lam and Ye, 2022). This broad array of applications highlights the versatility of correlation network analysis and demonstrates the insights we can gain from it.

Identifying putative keystone taxa and understanding the mucosal glycan degrading community

In my PhD thesis I analyze correlation networks at different taxonomic resolutions to identify putative keystone taxa and I utilize transcriptional profiles to discern associations between mucosal glycan degrading gut microbes. A common denominator between these two projects is not only that they are located in the human gut microbiome but the specific data I use. I combine two publicly available datasets of paired metagenomes and metatranscriptomes for my analyses in both research projects: a cohort of adult men (Mehta et al., 2018) and a study on inflammatory bowel disease where I focus on the healthy control samples (Schirmer et al., 2018). I process the reads with HUMAnN 3 (Beghini et al., 2021) which maps them to reference genomes and provides both compositional as well as functional profiles.

Methodologically, the emphasis in chapter 1 of my thesis is on correlation network analysis, specifically the detection of keystone taxa based on these networks. Keystone taxa are thought to have a disproportionately large effect on the structure and stability of their community (Banerjee et al., 2018). In the context of gut microbiome research this makes them interesting targets for medical intervention and substantial research effort was put into the detection of keystone taxa, frequently through the interrogation of correlation networks (Ze et al., 2012; Liu et al., 2019; Rodrigues et al., 2021; Amit and Bashan, 2023). However, the ecological relevance of keystone taxa identified through correlation network analysis is still an open question. I use previously described network features (Berry and Widder, 2014) to propose putative keystone taxa and evaluate the effect of taxonomic resolution on these networks. Additionally, I investigate community stratification and analyze transcriptional profiles of keystone taxa to further our understanding of their potential ecological role.

In chapter 2 of my thesis I focus on the mucosal glycan degrading community. Here the emphasis is on the analysis of transcriptional profiles from degrading microbes. The mucosal layer coating the gut epithelium is constantly being degraded by a subset of the gut microbiome (Bell and Juge, 2021). While research has shown that many members of the gut microbial community have the genomic potential to degrade mucosal glycans (Ravcheev and Thiele, 2017) the actively degrading community is not yet well described. I interrogate metatranscriptomes to identify species actively transcribing mucosal glycan degradation genes and analyze transcriptional patterns to understand their different ecological strategies. Furthermore, I utilize transcription-based networks to understand associations between degraders, and investigate stratification patterns in the community based on co-occurrence.

REFERENCES

Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., et al. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* 3, 356–366. doi: 10.1038/s41564-017-0084-4

Amit, G., and Bashan, A. (2023). Top-down identification of keystone taxa in the microbiome. *Nat. Commun.* 14, 3951. doi: 10.1038/s41467-023-39459-5

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180. doi: 10.1038/nature09944

Banerjee, S., Schlaeppi, K., and van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* 16, 567–576. doi: 10.1038/s41579-018-0024-1

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10, e65088. doi: 10.7554/eLife.65088

Bell, A., and Juge, N. (2021). Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology* 31, 691–696. doi: 10.1093/glycob/cwaa097

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2014.00219 (Accessed April 4, 2023).

Brugiroux, S., Berry, D., Ring, D., Barnich, N., Daims, H., and Stecher, B. (2022). Specific Localization and Quantification of the Oligo-Mouse-Microbiota (OMM12) by Fluorescence In Situ Hybridization (FISH). *Curr. Protoc.* 2, e548. doi: 10.1002/cpz1.548

Brugiroux, S., Beutler, M., Pfann, C., Garzetti, D., Ruscheweyh, H.-J., Ring, D., et al. (2016). Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat. Microbiol.* 2, 1–12. doi: 10.1038/nmicrobiol.2016.215

Chen, L., Wang, D., Garmaeva, S., Kurilshikov, A., Vila, A. V., Gacesa, R., et al. (2021). The long-term genetic stability and individual specificity of the human gut microbiome. *Cell* 184, 2302-2315.e12. doi: 10.1016/j.cell.2021.03.024

Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178–184. doi: 10.1038/nature11319

Corbin, K. D., Carnero, E. A., Dirks, B., Igudesman, D., Yi, F., Marcus, A., et al. (2023). Host-diet-gut microbiome interactions influence human energy balance: a randomized clinical trial. *Nat. Commun.* 14, 3161. doi: 10.1038/s41467-023-38778-x

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., et al. (2018). Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* 3, 8–16. doi: 10.1038/s41564-017-0072-8

De la Paz, E., Maganti, N. H., Trifonov, A., Jeerapan, I., Mahato, K., Yin, L., et al. (2022). A self-powered ingestible wireless biosensing system for real-time in situ monitoring of gastrointestinal

tract metabolites. *Nat. Commun.* 13, 7405. doi: 10.1038/s41467-022-35074-y

Dethlefsen, L., and Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl. Acad. Sci.* 108, 4554–4561. doi: 10.1073/pnas.1000087107

Dey, P., and Ray Chaudhuri, S. (2023). The opportunistic nature of gut commensal microbiota. *Crit. Rev. Microbiol.* 49, 739–763. doi: 10.1080/1040841X.2022.2133987

Dhar, V. (2013). Data science and prediction. *Commun. ACM* 56, 64–73. doi: 10.1145/2500499

Donoho, D. (2017). 50 Years of Data Science. *J. Comput. Graph. Stat.* 26, 745–766. doi: 10.1080/10618600.2017.1384734

Earle, K. A., Billings, G., Sigal, M., Lichtman, J. S., Hansson, G. C., Elias, J. E., et al. (2015). Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* 18, 478–488. doi: 10.1016/j.chom.2015.09.002

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial Co-occurrence Relationships in the Human Microbiome. *PLOS Comput. Biol.* 8, e1002606. doi: 10.1371/journal.pcbi.1002606

Fisher, C. K., and Mehta, P. (2014). Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. *PLOS ONE* 9, e102451. doi: 10.1371/journal.pone.0102451

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111

Friedman, J., and Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLOS Comput. Biol.* 8, e1002687. doi: 10.1371/journal.pcbi.1002687

Frioux, C., Ansorge, R., Özkurt, E., Ghassemi Nedjad, C., Fritscher, J., Quince, C., et al. (2023). Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* 31, 1111-1125.e6. doi: 10.1016/j.chom.2023.05.024

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234

Jackson, M. A., Bonder, M. J., Kuncheva, Z., Zierer, J., Fu, J., Kurilshikov, A., et al. (2018). Detection of stable community structures within gut microbiota co-occurrence networks from different human populations. *PeerJ* 6, e4303. doi: 10.7717/peerj.4303

Jernberg, C., Löfmark, S., Edlund, C., and Jansson, J. K. (2010). Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* 156, 3216–3223. doi: 10.1099/mic.0.040618-0

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D., et al. (2014). Rethinking “Enterotypes.” *Cell Host Microbe* 16, 433–437. doi: 10.1016/j.chom.2014.09.013

Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., et al.

(2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* 53, 156–165. doi: 10.1038/s41588-020-00763-1

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Comput. Biol.* 11, e1004226. doi: 10.1371/journal.pcbi.1004226

Lam, T. J., and Ye, Y. (2022). Meta-analysis of microbiome association networks reveal patterns of dysbiosis in diseased microbiomes. *Sci. Rep.* 12, 17482. doi: 10.1038/s41598-022-22541-1

Larsen, N., Vogensen, F. K., Berg, F. W. J. van den, Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., et al. (2010). Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. *PLOS ONE* 5, e9085. doi: 10.1371/journal.pone.0009085

Larzul, C., Estellé, J., Borey, M., Blanc, F., Lemonnier, G., Billon, Y., et al. (2024). Driving gut microbiota enterotypes through host genetics. *Microbiome* 12, 116. doi: 10.1186/s40168-024-01827-8

Li, D., Gao, C., Zhang, F., Yang, R., Lan, C., Ma, Y., et al. (2020). Seven facts and five initiatives for gut microbiome research. *Protein Cell* 11, 391–400. doi: 10.1007/s13238-020-00697-8

Liu, F., Li, Z., Wang, X., Xue, C., Tang, Q., and Li, R. W. (2019). Microbial Co-Occurrence Patterns and Keystone Species in the Gut Microbial Community of Mice in Response to Stress and Chondroitin Sulfate Disaccharide. *Int. J. Mol. Sci.* 20, 2130. doi: 10.3390/ijms20092130

Luo, M., Zhu, J., Jia, J., Zhang, H., and Zhao, J. (2024). Progress on network modeling and analysis of gut microecology: a review. *Appl. Environ. Microbiol.* 90, e00092-24. doi: 10.1128/aem.00092-24

McCallum, G., and Tropini, C. (2024). The gut microbiota and its biogeography. *Nat. Rev. Microbiol.* 22, 105–118. doi: 10.1038/s41579-023-00969-0

Mehta, R. S., Abu-Ali, G. S., Drew, D. A., Lloyd-Price, J., Subramanian, A., Lochhead, P., et al. (2018). Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* 3, 347–355. doi: 10.1038/s41564-017-0096-0

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79. doi: 10.1186/gb-2012-13-9-r79

Nichols, R. G., and Davenport, E. R. (2021). The relationship between the gut microbiome and host gene expression: a review. *Hum. Genet.* 140, 747–760. doi: 10.1007/s00439-020-02237-0

Nowicki, E. M., Shroff, R., Singleton, J. A., Renaud, D. E., Wallace, D., Drury, J., et al. (2018). Microbiota and Metatranscriptome Changes Accompanying the Onset of Gingivitis. *mBio* 9, e00575-18. doi: 10.1128/mBio.00575-18

Peluzio, M. do C. G., Martinez, J. A., and Milagro, F. I. (2021). Postbiotics: Metabolites and mechanisms involved in microbiota-host interactions. *Trends Food Sci. Technol.* 108, 11–26. doi: 10.1016/j.tifs.2020.12.004

Pérez-Losada, M., Castro-Nallar, E., Bendall, M. L., Freishtat, R. J., and Crandall, K. A. (2015). Dual Transcriptomic Profiling of Host and Microbiota during Health and Disease in Pediatric Asthma. *PLOS ONE* 10, e0131819. doi: 10.1371/journal.pone.0131819

Quinn-Bohmann, N., Wilmanski, T., Sarmiento, K. R., Levy, L., Lampe, J. W., Gurry, T., et al. (2024). Microbial community-scale metabolic modelling predicts personalized short-chain fatty acid production profiles in the human gut. *Nat. Microbiol.* 9, 1700–1712. doi: 10.1038/s41564-024-01728-4

Ravcheev, D. A., and Thiele, I. (2017). Comparative Genomic Analysis of the Human Gut Microbiome Reveals a Broad Distribution of Metabolic Pathways for the Degradation of Host-Synthetized Mucin Glycans and Utilization of Mucin-Derived Monosaccharides. *Front. Genet.* 8. doi: 10.3389/fgene.2017.00111

Rodrigues, R. R., Gurung, M., Li, Z., García-Jaramillo, M., Greer, R., Gaulke, C., et al. (2021). Transkingdom interactions between Lactobacilli and hepatic mitochondria attenuate western diet-induced diabetes. *Nat. Commun.* 12, 101. doi: 10.1038/s41467-020-20313-x

Schirmer, M., Franzosa, E. A., Lloyd-Price, J., McIver, L. J., Schwager, R., Poon, T. W., et al. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* 3, 337–346. doi: 10.1038/s41564-017-0089-z

Schubert, A. M., Rogers, M. A. M., Ring, C., Mogle, J., Petrosino, J. P., Young, V. B., et al. (2014). Microbiome Data Distinguish Patients with Clostridium difficile Infection and Non-C. difficile-Associated Diarrhea from Healthy Controls. *mBio* 5, 10.1128/mbio.01021-14. doi: 10.1128/mbio.01021-14

Schwab, C., Berry, D., Rauch, I., Rennisch, I., Ramesmayer, J., Hainzl, E., et al. (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J.* 8, 1101–1114. doi: 10.1038/ismej.2013.223

Shanahan, F., Ghosh, T. S., and O'Toole, P. W. (2021). The Healthy Microbiome—What Is the Definition of a Healthy Gut Microbiome? *Gastroenterology* 160, 483–494. doi: 10.1053/j.gastro.2020.09.057

Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., et al. (2013). Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLOS Comput. Biol.* 9, e1003388. doi: 10.1371/journal.pcbi.1003388

Tropini, C., Earle, K. A., Huang, K. C., and Sonnenburg, J. L. (2017). The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe* 21, 433–442. doi: 10.1016/j.chom.2017.03.010

Watts, S. C., Ritchie, S. C., Inouye, M., and Holt, K. E. (2019). FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* 35, 1064–1066. doi: 10.1093/bioinformatics/bty734

Weiss, A. S., Niedermeier, L. S., von Strempel, A., Burrichter, A. G., Ring, D., Meng, C., et al. (2023). Nutritional and host environments determine community ecology and keystone species in a synthetic gut bacterial community. *Nat. Commun.* 14, 4780. doi: 10.1038/s41467-023-40372-0

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., et al. (2011). Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344

Xiong, R., Gunter, C., Fleming, E., Vernon, S. D., Bateman, L., Unutmaz, D., et al. (2023). Multi-'omics of gut microbiome-host interactions in short- and long-term myalgic encephalomyelitis/chronic fatigue syndrome patients. *Cell Host Microbe* 31, 273-287.e5. doi: 10.1016/j.chom.2023.01.001

Ze, X., Duncan, S. H., Louis, P., and Flint, H. J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* 6, 1535–1543. doi: 10.1038/ismej.2012.4

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

Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., et al. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* 8, 881–893. doi: 10.1038/ismej.2013.185

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352, 565–569. doi: 10.1126/science.aad3369

CHAPTER 1

Characteristics of putative keystones in the healthy adult human gut microbiota as determined by correlation network analysis

Franziska Bauchinger[1, 2], David Seki [1], David Berry[1,3*]

[1]University of Vienna, Centre for Microbiology and Environmental Systems Science CeMESS, Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, Vienna, Austria

[2]University of Vienna, Doctoral School in Microbiology and Environmental Science, Vienna, Austria

[3]Joint Microbiome Facility of the Medical University of Vienna and the University of Vienna, Vienna, Austria

Frontiers in Microbiology, 20 November 2024

Section Microorganisms in Vertebrate Digestive Systems

Volume 15 – 2024 | https://doi.org/10.3389/fmicb.2024.1454634

ABSTRACT

Keystone species are thought to play a critical role in determining the structure and function of microbial communities. As they are important candidates for microbiome-targeted interventions, the identification and characterization of keystones is a pressing research goal. Both empirical as well as computational approaches to identify keystones have been proposed, and in particular correlation network analysis is frequently utilized to interrogate sequencing-based microbiome data. Here, we apply an established method for identifying putative keystone taxa in correlation networks. We develop a robust workflow for network construction and systematically evaluate the effects of taxonomic resolution on network properties and the identification of keystone taxa. We are able to identify correlation network keystone species and genera, but could not detect taxa with high keystone potential at lower taxonomic resolution. Based on the correlation patterns observed, we hypothesize that the identified putative keystone taxa have a stabilizing effect that is exerted on correlated taxa. Correlation network analysis further revealed subcommunities present in the dataset that are remarkably similar to previously described patterns. The interrogation of available metatranscriptomes also revealed distinct transcriptional states present in all putative keystone taxa. These results suggest that keystone taxa may have stabilizing properties in a subset of community members rather than global effects. The work presented here contributes to the understanding of correlation network keystone taxa and sheds light on their potential ecological significance.

INTRODUCTION

Microbiomes are characterized by diverse interspecific and microbe-environment interactions. The net outcome of microbial activities and interactions produces the observed community structure and function, and interactions are thought to be crucial for maintaining community stability and conferring resistance and resilience in the face of disturbance. Borrowing a concept from macro-ecology (Paine, 1969), the idea of keystone species has intrigued many microbiologists as a potential ecological mechanism that facilitates community stability and resilience. Microbial keystones, whether defined at a species level or a higher taxonomic grouping, are characterized as highly interconnected taxa that exert a strong influence on the entire community, shaping the microbiome irrespective of their abundance (Hausmann et al., 2019). Loss of a keystone taxon from the community would be expected to disrupt the microbiome and cause major shifts in community composition and function. Their crucial role in maintaining community structure is what makes keystone taxa a particularly interesting target for human gut microbiome research.

The gut microbiome is essential for human physiological function and health by performing many important services including metabolizing complex food molecules (Hamaker and Tuncil, 2014), training the immune system (Zheng et al., 2020), and providing colonization resistance against pathogens (Brugiroux et al., 2016). The composition of the gut microbial community is highly individualized (Lloyd-Price et al., 2017) and in particular effected by diet and geographical location (Wilson et al., 2020). A so-called westernized diet, high in fat but low in fiber, has been linked to diseases such as type 2 diabetes (Sharma and Tripathi, 2019) and colorectal cancer (Arita and Inagaki-Ohara, 2019). Identifying keystone taxa in the healthy gut microbiome and understanding their role in the community could provide us with novel microbiome-targeted intervention strategies for altered community compositions related to diseases such as inflammatory bowel disease, cardiovascular disease, and obesity (Carding et al., 2015). Some species have been proposed as keystones in the human gut microbiome based on empirical evidence, such as *Akkermansia muciniphila* (Belzer et al., 2017) and *Christensenella minuta* (Mazier et al., 2021).

A. muciniphila was shown to facilitate the growth of butyrate producers by degrading host-derived mucosal sugars in a mucus-dependent cross-feeding community (Belzer et al., 2017). *C. minuta* showed protective properties against diet-induced obesity in mouse models and modulated the intestinal microbiota in an in vitro model inoculated with fecal samples from obese individuals (Mazier et al., 2021). However, the experimental identification of keystone taxa in the human gut microbiome is challenging as intervention studies are difficult to conduct and animal and in vitro models can have limited translatability for the human gut microbiome. Consequently, researchers have moved towards bioinformatics and data analysis to identify putative keystone taxa, utilizing methods such as presence-absence analysis (Amit and Bashan, 2023), linear regression (Fisher and Mehta, 2014) and machine learning algorithms (Wang et al., 2024).

A popular approach to identify keystone taxa in microbiomes is to use sequencing-based microbial abundance profiles to construct and analyze correlation or co-occurrence networks (Faust and Raes, 2012). The assumption is that positive or negative interspecific interactions, be they direct or indirect, lead to positive or negative correlations between the abundances of the respective taxa. It has been shown by Berry and Widder (Berry and Widder, 2014) that in simulated communities certain network features, namely node degree, closeness centrality, betweenness centrality and transitivity, are indicative of a taxon's keystone potential and can be used to identify keystone taxa with 85% accuracy.

Several studies have identified putative keystone taxa based on correlation network analysis (Fisher and Mehta, 2014; Goodrich et al., 2014; Liu et al., 2019) and proposed them as targets for subsequent experimental studies. However, we still lack an understanding of what ecological features these keystones share or which functional niches they occupy that lead to their observed characteristics in correlation networks (Banerjee et al., 2018). We can speculate that a keystone taxon provides a conserved, specific function in the gut microbiome that is not provided by other taxa and is therefore crucial in maintaining community structure and function (Ze et al., 2012; Cartmell et al., 2018). Alternatively, keystone taxa might be functionally versatile and able to fill available ecological niches, thereby providing the needed functional redundancy and resilience to microbial communities (Weiss et al., 2023). One can also imagine a mixture of both functions, either provided by a single taxon or a keystone guild. Keystone taxa may be part of the common core microbiome, the component found across a large proportion of communities, but are closer in concept to an ecological core microbiome, the component disproportionally important for shaping the community (Risely, 2020).

With this study, we aim to further our understanding of putative keystone taxa in the human gut microbiome. We establish a workflow to construct robust and statistically significant correlation networks with FastSpar (Watts et al., 2019) and identify correlation network keystones based on their network features, as proposed by Berry and Widder (Berry and Widder, 2014). We then utilize this workflow to interrogate the prokaryotic gut microbiota using publicly available metagenomes and metatranscriptomes from large human gut microbiome studies based in the USA. We find that detection of network keystones is highly sensitive to taxonomic resolution and are able to identify putative keystone taxa only at species and genus level. Correlation network analysis further suggests a community stabilizing effect of putative keystone taxa and reveals co-occurring subcommunities present across gut microbiomes.

MATERIAL AND METHODS

Processing of raw data

We analyzed metagenome and metatranscriptome reads from two publicly available datasets, a cohort of patients with inflammatory bowel disease as well as healthy control patients (Schirmer et al., 2018) and a cohort of adult men (Mehta et al., 2018). The sequencing data is available on the Sequence Read Archive under BioProject PRJNA389280 and BioProject PRJNA354235, respectively. The raw reads were preprocessed with trimmomatic (Bolger et al., 2014) (with settings leading: 3, trailing: 3, slidingwindow: 4:15, minlen: 50). We removed samples with fewer than one million reads from further analyses and, in the Schirmer dataset, used only samples from participants not diagnosed with IBD. We then randomly selected one sample per participant to ensure statistical independence of the samples. The trimmed reads were processed using HUMAnN 3.0.0 (Beghini et al., 2021). In further analyses we only considered reads mapped to reference genomes and discarded unmapped reads (leading to a total of 123,109,687 metagenome reads and 115,784,680 metatranscriptome reads). Briefly, HUMAnN 3.0.0 first estimates community composition with MetaphlAn 3.0.7 (Beghini et al., 2021) and then maps reads to a community pangenome with bowtie2 (Langmead and Salzberg, 2012). This results in both taxonomic as well as functional profiles of the metagenome and metatranscriptome reads. We used the setting rel_ab_w_read_stats in MetaphlAn 3.0.7 to estimate both relative abundances and total reads mapped to a reference genome. In the functional profiles we further grouped reads mapped to gene families into enzyme commission numbers (ECs) using the HUMAnN 3.0.0 function humann_regroup_table with - -groups uniref90_level4ec.

Principal coordinate analysis of community profiles

In order to visualize community similarity and overlap between the two analyzed datasets, we computed the robust Aitchison distance between samples with the function vegandist in the R package vegan (Oksanen et al., 2022), using the total estimated reads output from MetaphlAn 3.0.7 on species resolution. We then performed a principal coordinate analysis on these distances using the function cmdscale in the R package stats (Dalgaard, 2010).

Computation of correlation networks

We computed binary Jaccard similarities of community composition at species resolution between samples to estimate whether the communities were similar enough to compute correlation networks. Berry and Widder (Berry and Widder, 2014) suggest an overall site similarity (Jaccard similarity) of at least 20% and we observed a mean Jaccard similarity of 36%. We computed correlations between taxa using FastSpar (Watts et al., 2019) with the settings - -iterations 100 (corresponds to rounds of SparCC correlation estimations), - -exclude_iterations 20 (number of times highly correlated taxa pairs are discovered and excluded), - -threshold 0.1 (minimum threshold to exclude correlated taxa pairs) and - -number 1000 (number of bootstraps). FastSpar is a C++ implementation of the correlation

network analysis algorithm SparCC (Friedman and Alm, 2012). In short, these algorithms use log-transformed components to estimate linear Pearson correlations to account for the compositional nature of metagenomic data. In order to estimate the significance of observed correlations, FastSpar uses a permutation-based approach to generate a null distribution that observed correlations are then compared against. We used the total estimated reads output from MetaphlAn 3.0.7 (Beghini et al., 2021) as abundance estimates for each taxon. We computed correlations between taxa at the taxonomic resolution of order, family, genus and species. To ensure robust results, we subsampled our combined datasets. Specifically, we used FastSpar to compute correlation networks on a random subsample of 50 samples and repeated this process 1000 times. We then built a consensus network using only correlations with a p-value <= 0.05 in at least 200 iterations. The correlation strength in the consensus network was estimated as the mean correlation strength of significant correlations across all iterations. This workflow is visualized in Figure 1. We calculated network features, namely modularity, cohesion, relative node degree, closeness centrality, betweenness centrality and transitivity, using the R package igraph (Csárdi et al., 2023). Following the suggestion of Berry and Widder (Berry and Widder, 2014) we estimated each taxon's potential to act as a keystone using the network features relative node degree, betweenness centrality and transitivity. We excluded closeness centrality from our calculations in order to reduce collinearity of features. As shown by Berry and Widder (Berry and Widder, 2014), correlation network keystone taxa show a high relative node degree and transitivity as well as low betweenness centrality. We therefore computed the keystone potential (= KP) of each taxon as (relative node degree * transitivity) / betweenness centrality. We consider taxa a keystone if they show a particularly high KP as compared to all other prevalent taxa (taxa present in at least 20% of samples). Namely, we chose a cut-off of median + 5x median absolute deviation of KP after visually examining the entire distribution. We furthermore grouped the prevalent taxa into clusters using the function cluster_fast_greedy with default settings in the R package igraph (Csárdi et al., 2023) on positive correlation networks, in which only correlations with a correlation strength > 0 are considered. We visualized correlation networks using Cytoscape version 3.8.0 (Shannon et al., 2003).

Functional potential of keystone taxa

We investigated the functional potential of keystone taxa using the pathabundances output from HUMAnN 3.0.0 (Beghini et al., 2021). In short, pathway abundances are computed on community level as well as for individual taxa based on the abundances of the individual component reactions constituting an entire pathway. The pathway definitions used are based on MetaCyc (Caspi et al., 2016).

Transcriptional stability and transcriptional states of keystone taxa

In order to estimate functional stability of taxa we computed pairwise Sorenson similarities on the presence and absence of gene families using the function betadiver in the R package vegan (Oksanen et al., 2022). We performed k-means clustering to further analyze the

functional profiles of the keystone taxa and identify different transcriptional states. For this analysis we opted to use the functional resolution of ECs provided by HUMAnN 3.0.0 (Beghini et al., 2021) (function humann_regroup_table with - - groups uniref90_level4ec) and normalized copies per million to the total transcription attributed to each taxon per sample. We use the function fviz_nbclust in the R package factoextra (Kassambara and Mundt, 2020) to choose the optimal number of clusters (with a k.max - maximum number of clusters - of 10) by computing the average silhouette width for each number of clusters and choosing the highest score. For this number of clusters, we then performed k-means clustering using the function kmeans in the R packages stats (Dalgaard, 2010). However, if we observed clusters containing only 1 or 2 samples, we excluded these samples from the analysis and repeated the process. In order to gain some insights into which ECs are particularly important for distinguishing the observed transcriptional states, we performed random forest analyses. A random forest analysis is a classifying algorithm based on decision trees that calculates importance metrics for the features used to cluster a dataset. In order to estimate the significance of these importance metrics we used the function rfPermute in the R package rfPermute (Archer, 2023). This function computes a null distribution of importance metrics by permuting the response variable and calculates p-values for the observed importances. We used the mean decrease in accuracy as the importance metric for our analysis.

RESULTS

Establishing a workflow to build significant and robust correlation networks

In this study, we established a workflow that results in correlation networks based on both highly significant as well as robust correlations (Fig 1). A major concern when constructing correlation networks is the inherent compositionality of community abundance profiles derived from sequencing data, which can lead to spurious correlations. We utilized FastSpar (Watts et al., 2019), a C++ implementation of the well-established tool SparCC (Friedman and Alm, 2012), which identifies significant correlations and excludes spurious correlations based on a bootstrapping algorithm. We constructed correlation networks with FastSpar on subsamples from two publicly available datasets (Mehta et al., 2018; Schirmer et al., 2018) that we combined for our analysis. Subsequently we constructed a consensus network based only on correlations that were identified as significant in at least 20% of the individually constructed networks. This workflow ensures that correlations kept in the consensus network are representative of both the entire dataset as well as subsamples. We used this procedure to identify taxa that robustly exhibit correlation network keystone features in the gut microbiomes studied.

Community structure and species properties

In order to evaluate keystone taxa in the healthy human gut microbiome, we leveraged two large studies of healthy human adults, which included 580 paired metagenome and metatranscriptome samples from 131 individuals (Mehta et al., 2018; Schirmer et al., 2018). We first computed community composition profiles using HUMAnN 3.0.0 (Beghini et al., 2021). We then estimated whether the community compositions of the two datasets were similar enough to be combined by performing an ordination analysis and computing the similarity between communities on a species level. This revealed that while samples from the two datasets are somewhat separated, they still exhibit a strong overlap (Fig S1A). To ensure statistical independence between samples we randomly selected one sample per participant, resulting in an even larger overlap between the two datasets (Fig S1B). To further confirm the validity of combining datasets, we investigated the site similarity of the included samples by computing the overall Jaccard similarity between the communities (presence and absence of species), as suggested by Berry and Widder (Berry and Widder, 2014). Overall, the studied communities exhibit a similarity of 36%, exceeding the suggested lower limit of 20%. These results confirmed the suitability of combining the datasets for computing correlation networks and estimating keystone potential. To gain a better understanding of the data used for this study, we analyzed relative abundance, prevalence and transcriptional contribution of prevalent species (Fig S1C). We observed a large variance in relative abundance across samples, both within and between species. As expected, several species of *Bacteroides*, such as *Bacteroides vulgatus* (85% sample prevalence, 9% mean relative abundance [sd 10%]), *Bacteroides uniformis* (93% sample prevalence, 8% mean relative abundance [sd 7%])

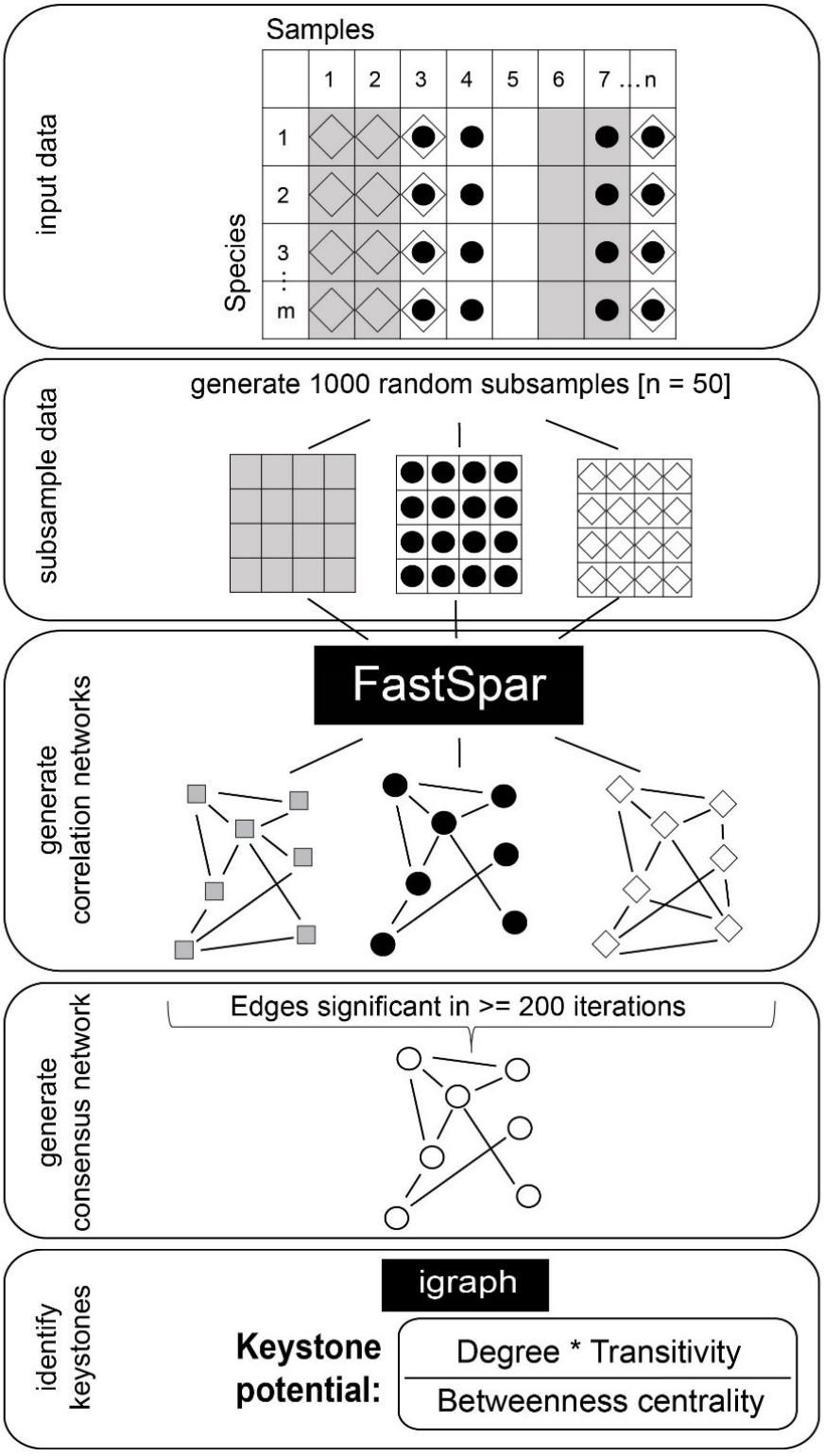

Figure 1 Workflow for the construction of robust and significant correlation networks. The dataset is subsampled 1,000 times and correlation networks are constructed from subsamples using FastSpar. Correlations identified as significant in >= 200 iterations are used to construct a consensus network. The mean correlation strength of all significant correlations is used as the correlation strength for the consensus network. Igraph is used to calculate relative node degree, transitivity and betweenness centrality.

and *Bacteroides stercoris* (51% sample prevalence, 4% mean relative abundance [sd 9%]), as well as *Faecalibacterium prausnitzii* (99% sample prevalence, 7% mean relative abundance [sd 5%]) and *Prevotella copri* (20% sample prevalence, 5% mean relative abundance [sd 14%]) are highly abundant in the dataset. More abundant species also tend to contribute more strongly to metatranscriptomes. This could reflect actual contribution to transcription, but may also be a result of better annotation quality in reference genomes from highly abundant species. While many highly abundant species are also highly prevalent across the dataset, there are multiple exceptions. Most notable and well known is the low prevalence of *P. copri*, a species that has previously been observed to be abundant in certain gut microbiomes while mostly absent in many others, particularly in cohorts from Europe and North America (Tett et al., 2019). In order to avoid spurious correlations and ensure high sensitivity (Berry and Widder, 2014) we focused our analysis on taxa exhibiting a prevalence of at least 20% across all studied samples and constructed correlation networks as described above.

Correlation networks reveal a loss of structure and keystone potential at lower taxonomic resolution

Previous work has shown that closely related taxa can form networks to utilize complex substrates in the human gut (Rakoff-Nahoum et al., 2014), but stable interactions between distantly related taxa have been also observed (Rao et al., 2021). It is, however, poorly understood how taxonomic resolution affects correlation analysis and specifically the identification of putative keystones. We therefore computed correlation networks at different taxonomic resolutions (order, family, genus and species) and investigated their structure. Overall, networks showed reduced modularity with increased taxonomic resolution (genus and species), while cohesion decreased from species- to order-level networks (Fig 2A). Modularity measures the extent to which a given network is divided into modules, with a higher modularity indicating greater divisions within the network. In contrast, the cohesion of a network estimates the minimum number of nodes that need to be removed to result in a weakly connected network, and a higher cohesion therefore indicates a more tightly connected network. Constructing correlation networks at low taxonomic resolution produced weakly connected networks while networks constructed at genus- and species-level are more cohesive and less modular. This observation is further confirmed by the observed distributions of network features, namely relative node degree (ND), transitivity (T) and betweenness centrality (BC). The relative degree of a node (= a taxon) is the number of edges a node has (correlations to another node) relative to the size of the network (total number of nodes). Transitivity indicates whether all nodes correlated with a given node are in turn correlated. The betweenness centrality of a node estimates how many shortest paths between two nodes (smallest number of edges connecting two nodes within a network) go through this given node. Correlation network keystones are characterized by high node degree, high transitivity and low betweenness centrality (Berry and Widder, 2014). The probability density functions of these network features, particularly transitivity and betweenness centrality, are flatter at lower taxonomic resolution (Fig 2B). In the node degree

distribution, we observe a slight flattening as well as a general shift towards a higher degree at lower taxonomic resolution. Flatter probability density functions, and therefore a more equal distribution of network features across all taxa, suggest that any one taxon is less likely to exhibit a high keystone potential. We confirm this by computing the keystone potential (= KP), using the formula KP = (ND*T)/BC, and comparing it across taxonomic resolutions (Fig 2C). As expected, we observe very low KP at family and order level, indicating that such broad taxonomic groups are unlikely to exhibit structuring effects on the overall community (Table S1). In contrast, we see both genera and species with a high potential to act as correlation network keystones. Interestingly, the pattern of KP is not consistent from genus to species, with some taxa exhibiting a higher potential at genus level and seemingly losing it at species level and vice versa. We also observe that degree and betweenness centrality tend to be positively correlated in both genera and species and consequently taxa with a high keystone potential show a surprisingly low degree, but high transitivity (Fig 2D and Fig S2). In summary, these results suggest that correlation network keystones are solely found within genera or species, and we thus focused our downstream analyses on these two taxonomic levels.

Few taxa have a relatively high keystone potential

We next analyzed the distribution of keystone potential in genera and species to better understand the network properties of correlation network keystones. Both on genus as well as on species level the distribution of keystone potential is strongly skewed to the right and only few taxa show a particularly high KP when compared to all other taxa (Fig 3A and B, respectively). This aligns well with our understanding of keystones and is to be expected based on the probability density functions of the network features used to compute KP (Fig 2B). Specifically, 8.8% of prevalent genera and 3.6% of prevalent species show a KP above 22 (see Materials and Methods section) and are therefore considered putative keystones in our analysis. On genus level these are *Methanobrevibacter* (KP = 92), *Agathobaculum* (KP = 39), *Bilophila* (KP = 30) and *Holdemania* (KP = 30) (Table 1) have a notably high potential to act as correlation network keystones. We also observe a group of unclassified *Bacillota* (KP = 41) with high KP, but as this group of species is likely not phylogenetically coherent, we did not include it in further analyses. On species level, *Firmicutes bacterium CAG* 83 (unclassified *Bacillota*, KP = 169), *Eisenbergiella tayi* (KP = 126), *Veillonella atypica* (KP = 52) and *Ruminococcus lactaris* (KP = 27) are putative keystones (Table 1). With the exception of the genera *Methanobrevibacter* and *Bilophila*, which belong to the phylum *Euryarchaeota* and *Thermodesulfobacteriota*, respectively, all of these taxa are *Bacillota*. When we compare the KP of each species with its respective KP on genus level, there is no clear pattern (Fig 3C). This suggests that certain genera may act as a keystone towards other genera but lose this potential on species level, and vice versa. We observe this for *Agathobaculum*, *Bilophila*, *Holdemania* and *Methanobrevibacter*. These four genera show a high KP, but the respective species exhibit low KP (Fig 2C). This is a particularly surprising observation considering that only one species belonging to each of these genera is present in the datasets, namely

Agathobaculum butyriciproducens, *Bilophila wadsworthia*, *Holdemania filiformis* and *Methanobrevibacter smithii*. These keystone genera are effectively keystone species that seem to have stronger network keystone properties when considering intergenic correlations. The correlations might be diluted on a species level and intergenic correlations may be distributed amongst multiple species with differing abundance patterns. Conversely, particular species have a high KP that is diluted when observed on genus level, as can be seen for *V*. *atypica* and *E*. *tayi*. The species within one genus may rarely co-occur, resulting in opposing correlation patterns and leading to a low KP on genus level. Ultimately, only few taxa in species and genus correlation networks exhibit a high potential to act as keystones, while the vast majority of taxa do not show the characteristics of putative keystones.

Correlation patterns indicate that keystone taxa may facilitate community stability

It has been suggested that in a highly diverse and functionally redundant system, such as the human gut microbiome, stability is facilitated by a high number of negative interactions, preventing the formation of positive feedback loops (Coyte et al., 2015). However, we observe that more than 50% of correlations in the analyzed networks are positive (Fig 4A). This is even more pronounced in species that belong to the same genus, where almost 75% of observed within-genus correlations are positive (Fig 4B). Putative keystone genera show a contrasting pattern, with around 60% negative correlations (Fig 4C) and genera that are strongly correlated with a keystone exhibit weaker within-genus correlations (Fig 4D). These results suggest that single-species keystone genera might provide community stability by counteracting positive feedback loops within other genera and dampening correlations. We also observe that larger genera exhibit weaker positive correlations within a respective genus (Fig S3A) which may be a result of functional redundancy in these more closely related taxa. On both genus as well as species level we observe that taxa correlated with a keystone taxon (first neighbors of keystones) are in general correlated to more taxa than those not correlated to a keystone. First neighbors of keystones also tend to have weaker correlations (Fig S3B and C, respectively). These observations further point towards a potential stabilizing effect of putative keystone taxa.

Correlation patterns reveal co-occurring clusters of taxa

To gain a better understanding of the community assembly characteristics of putative keystone taxa, we investigated the co-occurrence patterns of all prevalent species and genera. We constructed positive correlation networks, considering only correlations with a correlation strength > 0, and performed a clustering analysis on these networks. This analysis revealed four distinct clusters of species that are highly positively correlated with each other and that show fewer positive correlations to species of other clusters (Fig 5A). Each keystone species is a member of a different species cluster, indicating that keystone species are not highly correlated with each other. In contrast, keystone genera do not fall into separate clusters in the genus network, despite the presence of again four clusters (Fig 5B). Specifically, *Bilophila* and *Holdemania* are part of one cluster while *Agathobaculum* and

Methanobrevibacter are located in another. The remaining two clusters are a cluster including *Bacteroides*, a highly diverse and abundant genus, and a cluster including *Prevotella*, a genus with, as previously mentioned, distinct abundance patterns. It should further be noted that high abundance is not a necessity for a correlation network keystone (Fig S1C). In fact, amongst all prevalent taxa (at least 20% prevalence), keystones exhibit medium to low mean relative abundance. Examining the distribution of all network clusters across samples, we observe a gradient of community compositions both on species and genus level (Fig S4). At both taxonomic resolutions there are two dominant clusters that generally comprise the majority of the community. However, all four respective clusters co-occur in most samples and the composition ranges from heavily dominated by one cluster to almost equal contribution of each cluster. This observation is in accordance with recently described enterosignatures of gut microbial communities (Frioux et al., 2023) and may point towards a modular gut microbiome composed of distinct groups of co-occurring taxa.

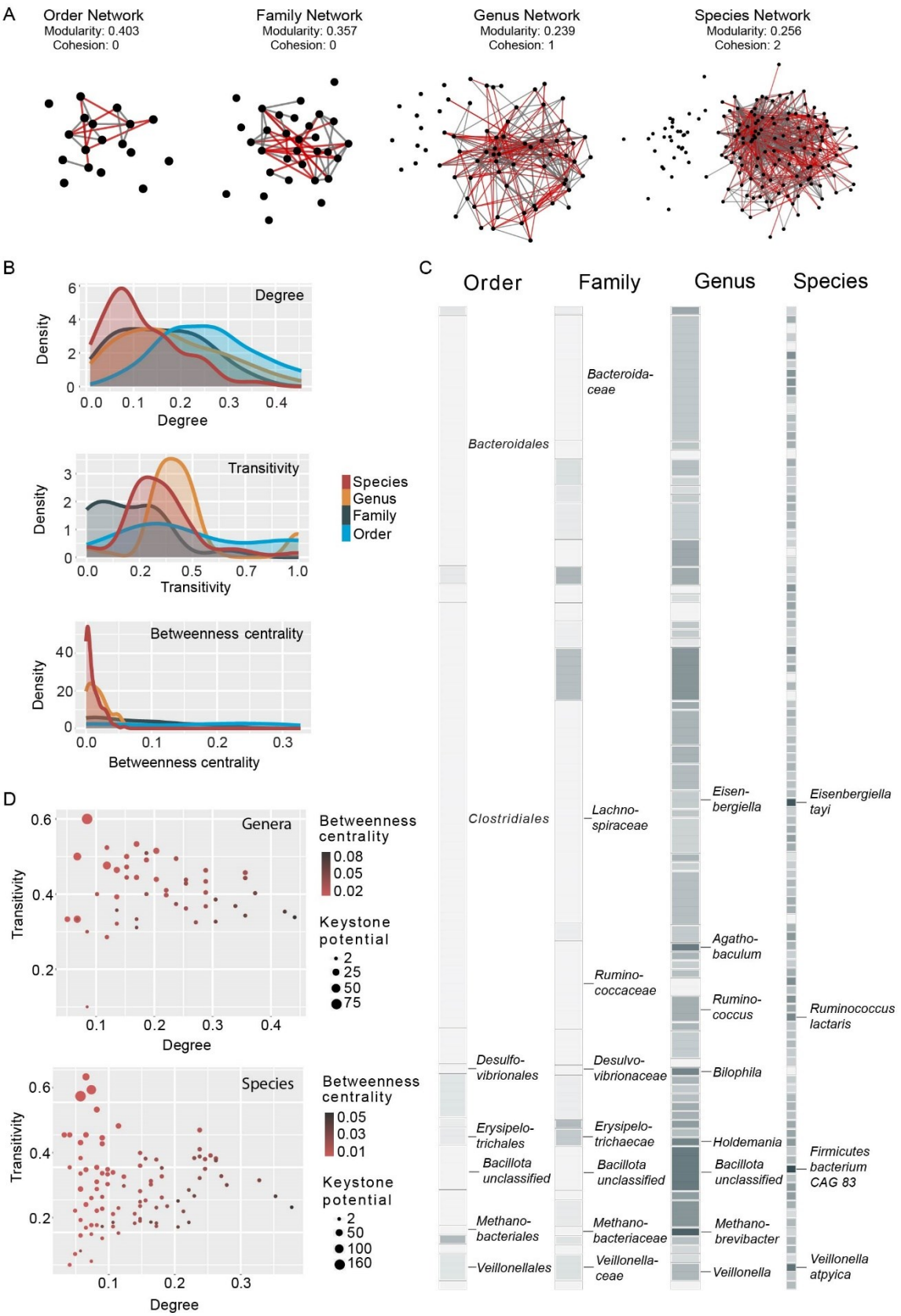
A
Order Network
Modularity: 0.403
Cohesion: 0
Family Network
Modularity: 0.357
Cohesion: 0
Genus Network
Modularity: 0.239
Cohesion: 1
Species Network
Modularity: 0.256
Cohesion: 2
B
Degree
Density
Degree
Transitivity
Density
Transitivity
Species
Genus
Family
Order
Betweenness centrality
Density
Betweenness centrality
C
Order
Family
Genus
Species
Bacteroidales
Clostridiales
Desulfovibrionales
Erysipelotrichales
Bacillota unclassified
Methanobacteriales
Veillonellales
Bacteroidaceae
Lachnospiraceae
Ruminococcaceae
Desulfovibrionaceae
Erysipelotrichaecae
Bacillota unclassified
Methanobacteriaceae
Veillonellaceae
Eisenbergiella
Agathobaculum
Ruminococcus
Bilophila
Holdemania
Bacillota unclassified
Methanobrevibacter
Veillonella
Eisenbergiella tayi
Ruminococcus lactaris
Firmicutes bacterium CAG 83
Veillonella atypica
D
Genera
Transitivity
Degree
Betweenness centrality
0.08
0.05
0.02
Keystone potential
2
25
50
75
Species
Transitivity
Degree
Betweenness centrality
0.05
0.03
0.01
Keystone potential
2
50
100
160

Figure 2 Distribution of correlation network features and keystone potential. (**A**) Correlation networks constructed on order, family, genus and species level. Network nodes (black circles) indicate taxa, network edges connecting the circles indicate positive (gray) and negative (red) correlations between taxa. Nodes without edges did not show significant correlations in >= 200 subsamples (see Materials and Methods section). The network layout was computed in Cytoscape (Shannon et al., 2003) using the layout option "Prefuse Force Directed Layout". In this layout nodes (taxa) with a higher number of significant correlations between them are placed closer together. (**B**) Density functions of relative node degree (relative to total network size), transitivity and betweenness centrality of taxa within a correlation network, colored by taxonomic resolution. (**C**) Keystone potential of taxa present in >= 20% of analyzed samples. The keystone potential of each taxon is indicated as a colored square, with white indicating very low keystone potential and dark gray indicating high keystone potential. (**D**) Relationship between mean degree and transitivity of network nodes. Circle size represents keystone potential and red (grey) circle color indicates low (high) betweenness centrality.

Table 1 Characteristics of identified correlation network keystone taxa

Agathobaculum butyriciproducens is a gram-positive, strictly anaerobic *Bacillota*. It is a known butyrate producer in the human gut microbiome (Ahn et al., 2016) and has previously been suggested to facilitate protection against Alzheimer's disease (Go et al., 2021) as well as Parkinson's disease (Lee et al., 2022) in mouse models. (83% sample prevalence; 0.31% mean relative abundance [sd 0.4%]; 25% negative correlations)
Bilophila wadsworthia is a gram-negative, obligate anaerobe bacterium of the phylum *Thermodesulfobacteriota* (Baron et al., 1989). It has previously been described as an opportunistic pathogen that is able to metabolize taurine to H_2S (Peck et al., 2019). In mouse models it has furthermore been suggested to aggravate symptoms caused by high fat diet (Natividad et al., 2018) as well as cause inflammation (Feng et al., 2017). Furthermore, higher relative abundances of *B. wadsworthia* have been observed in young colorectal cancer patients (Kharofa et al., 2023). (54% sample prevalence; 0.04% mean relative abundance [sd 0.07%]; 75% negative correlations)
Eisenbergiella tayi is a strictly anaerobic *Bacillota* that has been described as gram-stain-variable (Amir et al., 2014; Bernard et al., 2017). It has furthermore been suggested as a potential pathogen, being found at generally sterile human body sites (Bernard et al., 2017). The genus *Eisenbergiella* has been found to correlate with the uptake of macronutrients in a study investigating the microbiome of stunted children (Surono et al., 2021) and increased relative abundances of *E. tayi* have been observed in the gut microbiome of multiple sclerosis patients (Zhou et al., 2022). (26% sample prevalence; 0.05% mean relative abundance [sd 0.3%]; 11% negative correlations)
Firmicutes bacterium CAG 83 (unclassified *Bacillota*)has previously been described as a likely slow-growing member of the genus *Oscillospira* that is able to utilize host glycans and to produce butyrate (Gophna et al., 2017). It was found to be enriched in young colorectal cancer patients (Kharofa et al., 2023) as well as in patients suffering from asthma, after probiotic intervention (Liu et al., 2021). (69% sample prevalence; 0.34% mean relative abundance [sd 0.78%]; 57% negative correlations)
Holdemania filiformis is a strictly anaerobic, gram-positive *Bacillota*. Its main fermentation end products from glucose are acetic acids and lactic acids (Willems et al., 1997). Higher relative abundance of *H. filiformis* has been observed in young colorectal cancer patients (Kharofa et al., 2023) and patients with alopecia areata (Moreno-Arrones et al., 2020), but also in metastatic melanoma patients that responded to treatment with immune checkpoint inhibitors (Frankel et al., 2017; Moreno-Arrones et al., 2020). (37% sample prevalence; 0.01% mean relative abundance [sd 0,03%]; 50% negative correlations)
Methanobrevibacter smithii is a mesophilic, anaerobic archaeon in the phylum *Euryarchaeota* (Miller et al., 1982). It is the most prevalent methanogen in the human gastrointestinal tract (Dridi et al., 2009) and plays a pivotal role by consuming end products of bacterial fermentation to produce methane. In this context it has previously been described as a keystone taxon of the gut microbiome (Dridi et al., 2009; Horz and Conrads, 2010). (31% sample prevalence; 0.2% mean relative abundance [sd 0.74%]; 75% negative correlations)

Ruminococcus lactaris is an obligately anaerobic, gram-positive *Bacillota* named for its rapid fermentation of lactose (Moore et al., 1976). It is a prominent member of the human gut microbiome and able to produce short-chain fatty acids as well as lactate (Peck et al., 2019). Metagenome analysis has furthermore suggested that it can produce various B vitamins (Magnúsdóttir et al., 2015). Elevated abundances of *R. lactaris* have been reported in the gut microbiome of children with tic disorders (Xi et al., 2021). (53% sample prevalence; 0.49% mean relative abundance [sd 1.3%]; 70% negative correlations)
Veillonella atypica is a gram-negative, anaerobic *Bacillota* (Mays et al., 1982). It has been indicated as performance-enhancing in rodents through its lactate metabolism (Scheiman et al., 2019) as well as a potential agent for the bioremediation of sodium selenite (Pearce et al., 2009). *V. atpyica* is also frequently observed in human oral dental plaque biofilms where it may be metabolically strongly linked to *Streptococcus gordonii* (Egland et al., 2004). (23% sample prevalence; 0.04% mean relative abundance [sd 0.16%]; 23% negative correlations)

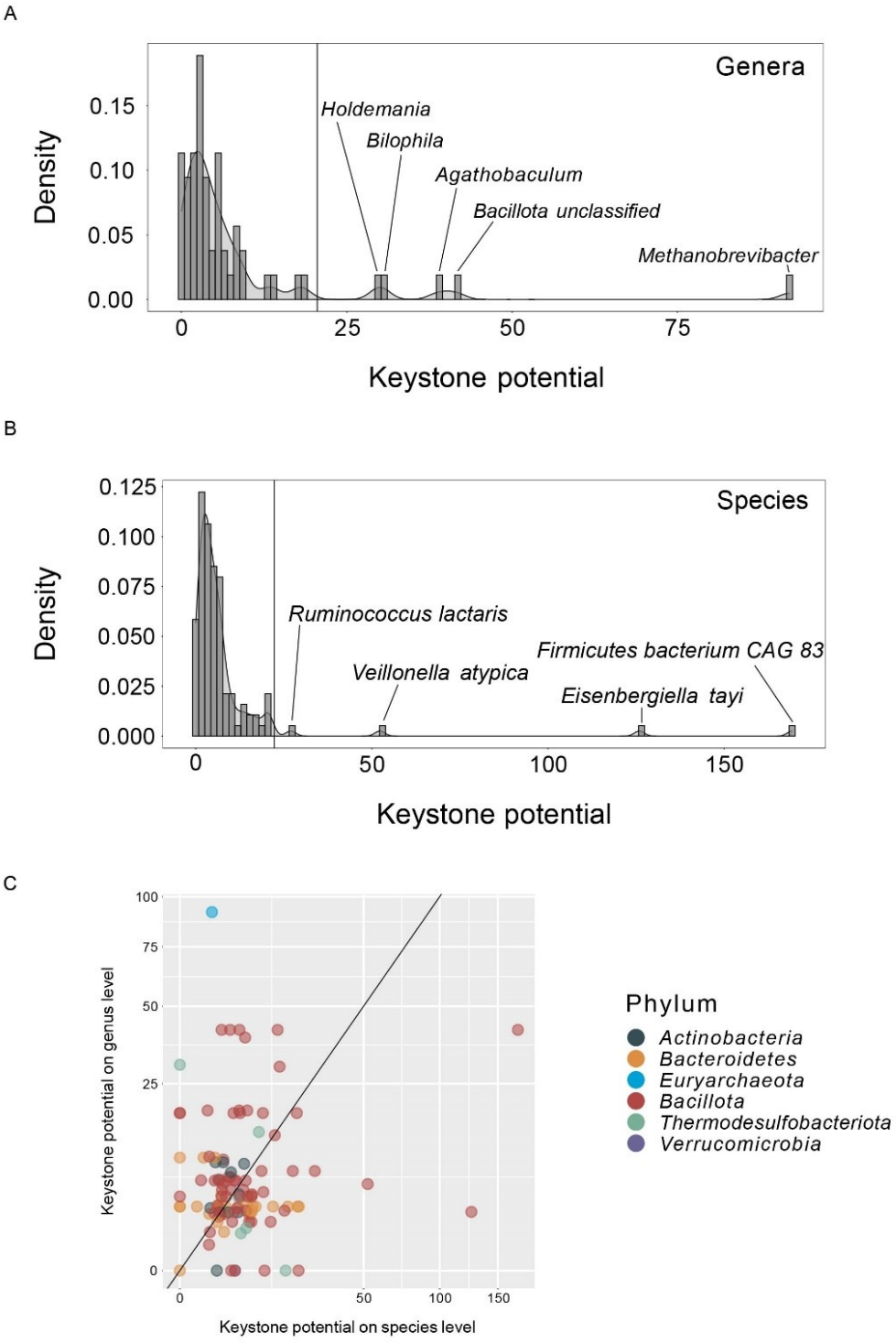

Figure 3 Density functions of keystone potential on (**A**) genus and (**B**) species level. Inserts show the taxa with the highest keystone potential. Horizontal lines indicate the keystone potential cut-off for taxa considered a keystone. The cut-off is set at median + 5x median absolute deviation. (**C**) Keystone potential on species level versus keystone potential on genus level, shown on a logarithmic scale. The diagonal line indicates a 1:1 linear relationship. Colors indicate the phylum of each taxon.

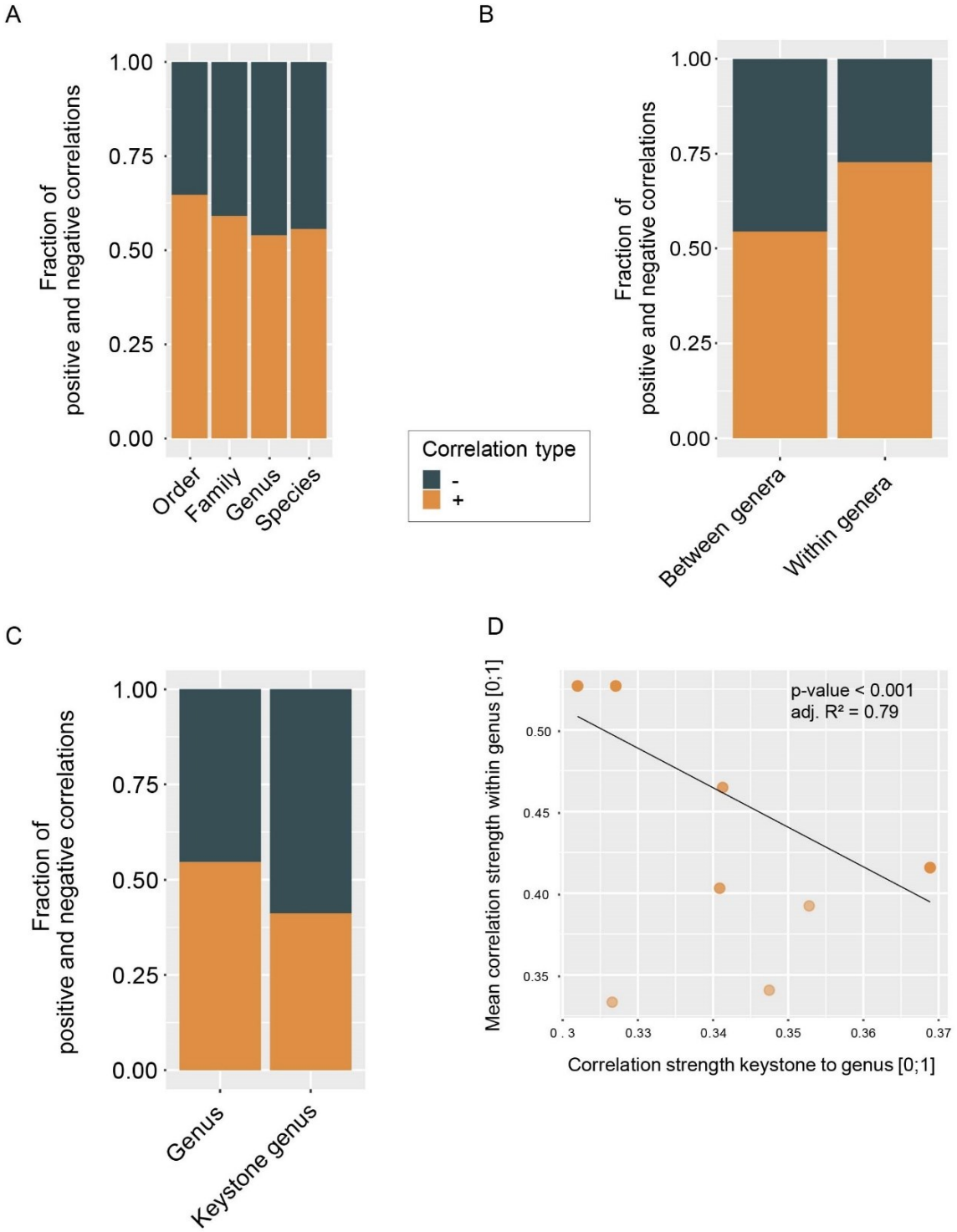

Figure 4 Fractions of negative and positive correlations (**A**) between taxa at different taxonomic resolution, (**B**) between genera and within genera and (**C**) of keystone genera and non-keystone genera. (**D**) Absolute correlation strength from keystone genera to other genera versus the mean absolute correlation strength within the respective genera (Linear model: p-value < 0.001, adj R^2 = 0.79).

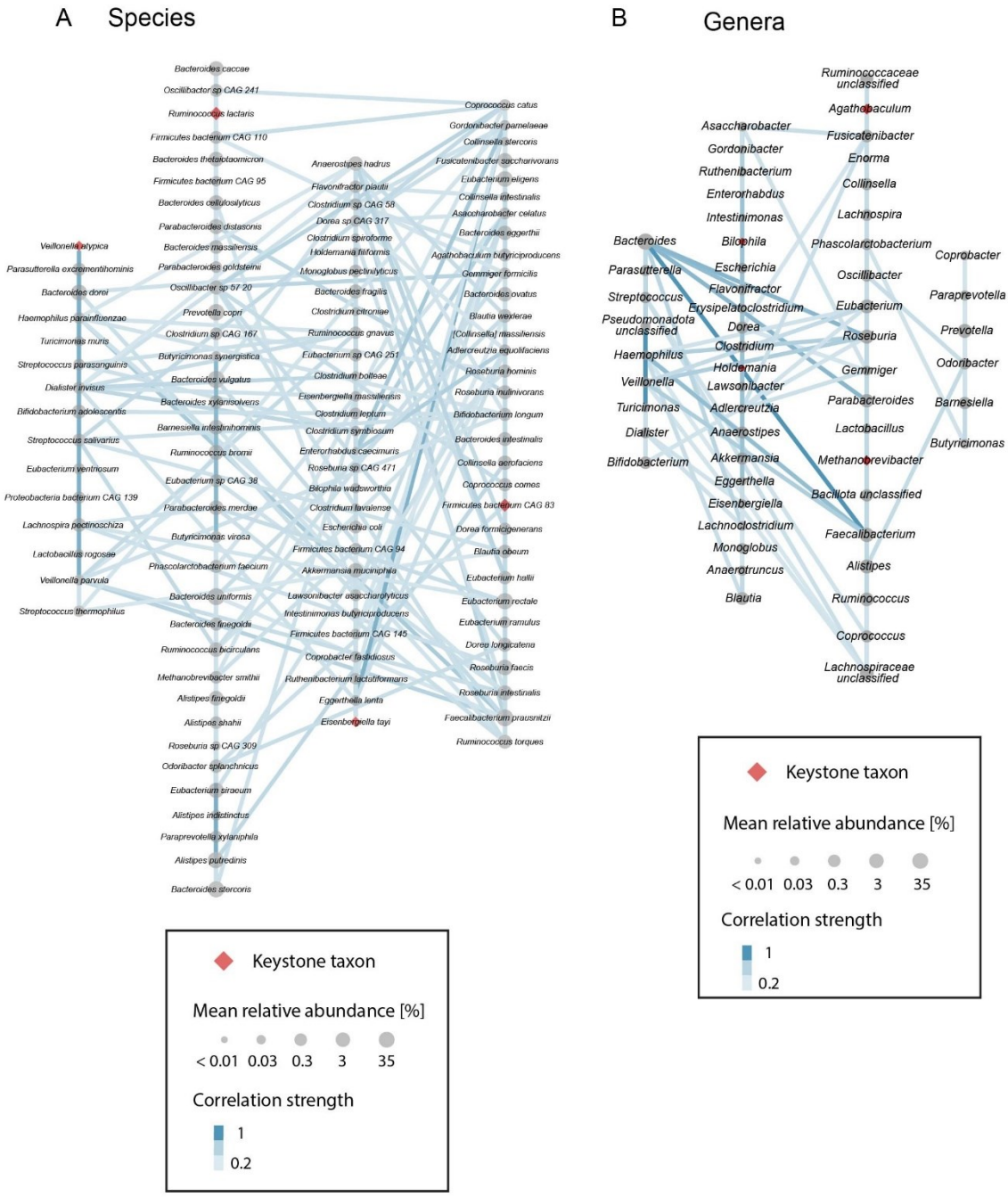

Figure 5 Positive correlation networks of prevalent (present in >= 20% of analyzed samples) (**A**) species and (**B**) genera. Gray circles indicate taxa, red diamonds indicate identified correlation network keystone taxa. Circle and diamond size represent the mean relative abundance of taxa. Edges between circles indicate positive correlations. Light blue indicates a weak correlation, darker blue indicates a stronger correlation. Taxa are organized into 4 clusters of strongly positively correlated species and genera, respectively, as identified with a fast greedy clustering algorithm.

Keystone taxa are transcriptionally versatile and show distinct transcriptional states

In an effort to elucidate the functional role of putative keystone taxa, we investigated their genomic potential and transcriptional versatility. No genomic features were identified to be distinctive of keystone taxa (Fig S5), so we focused on analyzing their transcriptional versatility and potential transcriptional states. We estimated the transcriptional versatility of prevalent taxa by computing pairwise Sorensen similarities on the presence and absence of gene families in the metatranscriptomes. The average transcriptional similarity of keystone taxa ranges from fairly high (*Holdemania*: 0.7, *R. lactaris*: 0.64) to the lower end of the observed range (*Bilophila*: 0.22, *E. tayi*: 0.18) (Fig S6). However, as is the case for most of the prevalent taxa, transcriptional similarity within each keystone taxon varies strongly across all analyzed samples. This suggests that keystone taxa do not occupy a conserved ecological niche - as inferred by transcriptional profiles - but are rather functionally versatile and potentially occupy different ecological niches within different gut microbial communities. Based on these observations we performed clustering analyses on the transcriptional profiles of the keystone taxa to identify potentially differing transcriptional states. We were able to identify two to three discrete transcriptional states for most keystone taxa (Figs 6, 7), with the exceptions of *Holdemania* and *V. atypica*. In these two cases very few reads mapped to known gene families and fewer still could be grouped into known enzyme commission numbers (ECs), indicating low annotation rates in their reference genomes. Consequently, a clustering analysis based on transcriptional profiles could not be performed for these two taxa.

We next performed Random Forest machine-learning analyses to identify individual ECs that distinguish the transcriptional states (Figs 6, 7, Table S1). For *Bilophila*, this revealed two ECs with contrasting transcription profiles (Fig 6C and D). Taurine-pyruvate aminotransferase (EC 2.6.1.77), is involved in taurine and hypotaurine metabolism and shows higher relative transcription in the samples in transcriptional state 1. Taurine metabolism is well described in *Bilophila* and may be linked to disease phenotypes (Natividad et al., 2018; Peck et al., 2019). Glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.157), which shows higher relative transcription in transcriptional state 2, is involved in the metabolism of various amino sugars and nucleotide sugars, such as glucose or extracellular N-Acetyl-D-glucosamine. For *Methanobrevibacter*, we could identify three transcriptional states that are likely related to methane metabolism (Fig 6E). Coenzyme-B sulfoethylthiotransferase (EC 2.8.4.1), the enzyme that catalyzes the final step of methanogenesis, was the most important in distinguishing these states (Fig 6F). Our analysis revealed three additional ECs involved in methane metabolism (EC 2.3.1.101, Formylmethanofuran-tetrahydromethanopterin formyltransferase, EC 1.17.1.9, Formate dehydrogenase and EC 4.1.2.43, 3-hexulose-6-phosphate synthase) as important in distinguishing transcriptional states (Table S2). Two transcriptional states were identified in *E. tayi* (Fig 7C). Notably, all three ECs identified as important in distinguishing these transcriptional states are involved in the biosynthesis of lysine (Fig 7D). The proportion of *E. tayi* transcription dedicated to these enzymes seems to be higher in transcriptional state 2, pointing towards an upregulation of lysine biosynthesis.

It is important to note that a large proportion of metatranscriptome reads could not be grouped into any known ECs. We still chose the higher resolution of ECs over gene families to gain more insight into the function of the transcripts at the expense of losing a larger proportion of reads. We would also like to point out that the presence of transcriptional states is not unique to keystone taxa and that we could observe similar patterns in many other taxa. Despite these limitations, the data point towards the presence of functionally versatile keystone taxa in the gut microbial community.

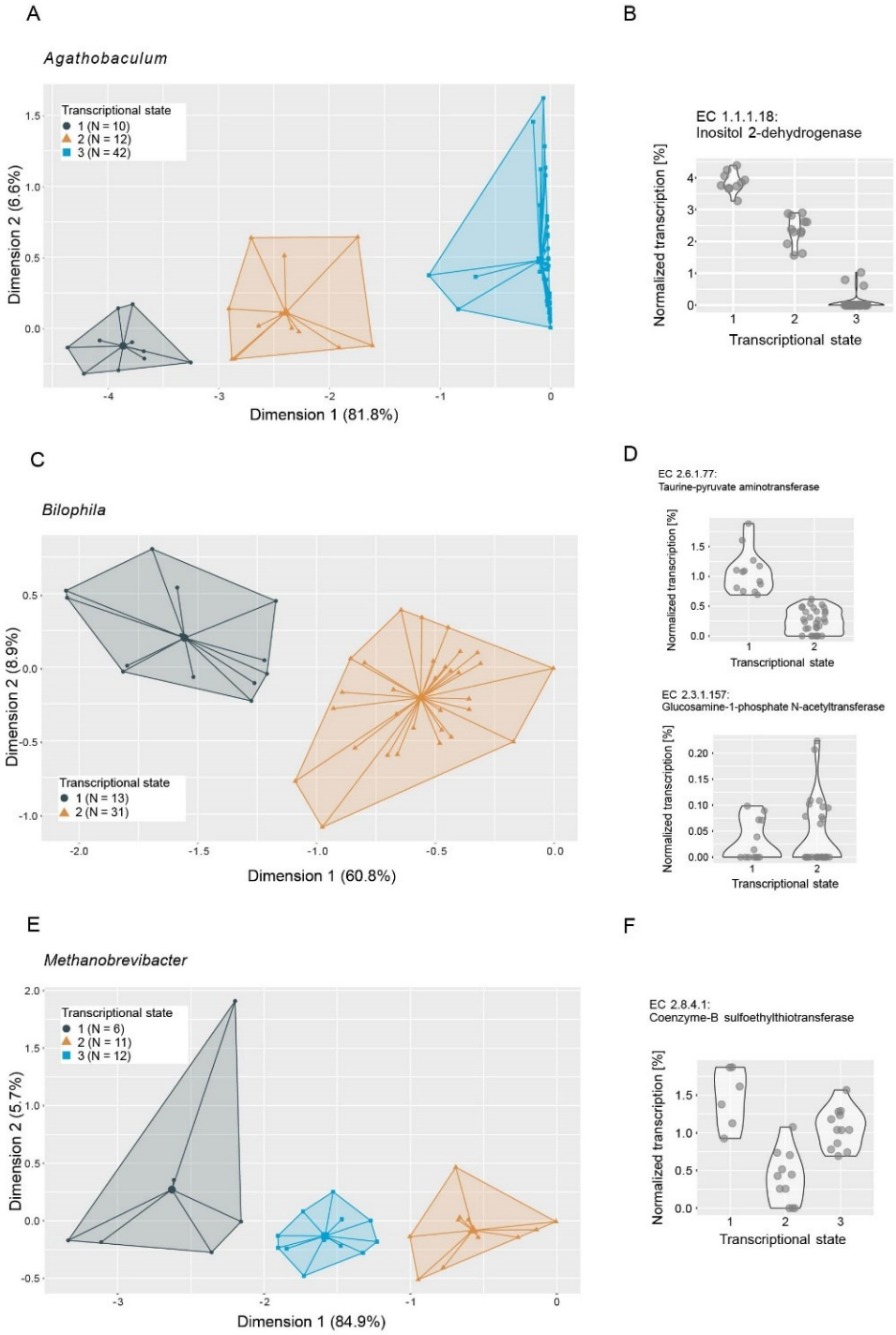

Figure 6 Ordination plots depicting principal component analyses of metatranscriptomes of (**A**) *Agathobaculum*, (**C**) *Bilophila* and (**E**) *Methanobrevibacter*. Metatranscriptome reads were mapped to reference genomes, grouped to Enzyme commission numbers (ECs) and normalized to the total number of reads mapped to each taxon per sample. Colors and icons indicate transcriptional states identified through k-means clustering (average silhouette width: *Agathobaculum*: 0.87, *Bilophila*: 0.92, *Methanobrevibacter*: 0.86). Normalized transcription of example ECs identified as important in distinguishing the transcriptional states through a random forest model are depicted in panels **B**, **D** and **F**, respectively.

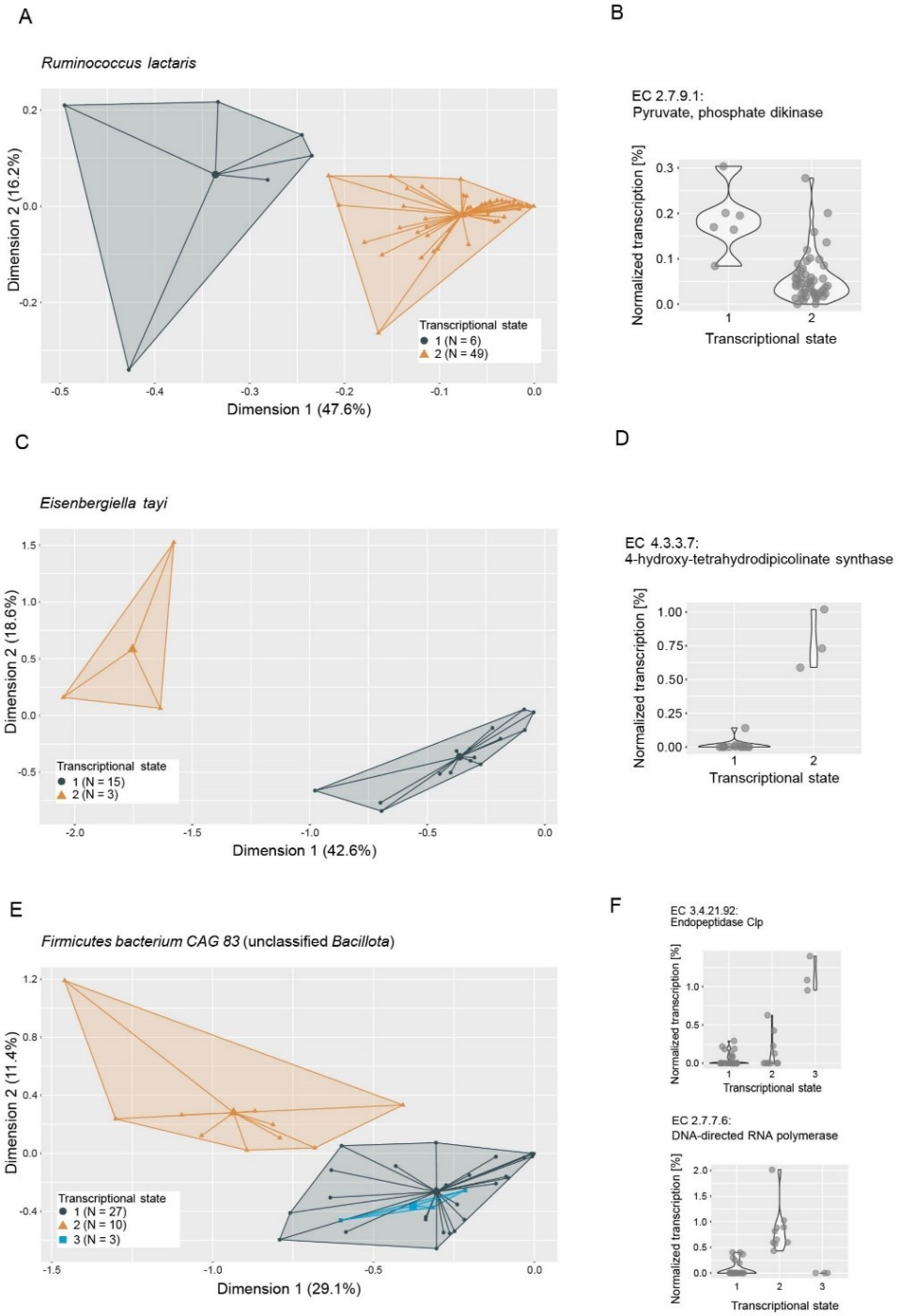

Figure 7 Ordination plots depicting principal component analyses of metatranscriptomes of (**A**) *Ruminococcus lactaris*, (**C**) *Eisenbergiella tayi* and (**E**) *Firmicutes bacterium CAG* 83 *(*unclassified *Bacillota)*. Metatranscriptome reads were mapped to reference genomes, grouped to Enzyme commission numbers (ECs) and normalized to the total number of reads mapped to each taxon per sample. Colors and icons indicate transcriptional states identified through k-means clustering (average silhouette width*: R. lactaris*: 0.92, *E. tayi*: 0.78*, Firmicutes bacterium CAG* 83: 0.81). Normalized transcription of example ECs identified as important in distinguishing the transcriptional states through a random forest model are depicted in panels **B**, **D** and **F**, respectively.

DISCUSSION

The aim of the present study is to elucidate the potential ecological role of correlation network keystone taxa of the human gut microbiome and gain a better understanding of how these analyses tie into other avenues of human gut microbiome research. In order to identify correlation network keystones, we developed a workflow based on a bootstrapping approach and a subsampling regime that enables the construction of robust and statistically significant correlation networks. We only observe high keystone potential of any taxa in genus- and species-level networks (Fig 2C), in contrast to previous suggestions of higher-order keystones (Trosvik and de Muinck, 2015). Due to the combination of different network features in our approach and the observed positive relationship between mean degree and betweenness centrality, the identified putative keystone taxa show a surprisingly low mean degree (Fig 2D). This result contrasts a more traditional understanding of keystone taxa as network hubs (nodes with a high number of edges) (Paine, 1969; Faust and Raes, 2012; Liu et al., 2019; Varsadiya et al., 2021). While we were able to identify taxa with a high keystone potential in the genus-level correlation network, they are interestingly all single-species genera, meaning that no other species belonging to these genera were detected in the analyzed samples. However, the respective species of these single-species genera are not identified as keystone species in the species-level network. This indicates that they exert their influence on other genera rather than other species. We hypothesize that the correlations to other species may be weak because they are spread amongst many species of individual genera, but on genus-level these correlations aggregate to stronger correlations which lead to the detection of single-species keystone genera. Conversely, we may miss correlations due to intraspecific genetic variability that is only detectable at strain level. These dilutions of the correlation signal may occur to a different extent across taxa due to different genetic boundaries across taxonomic clades. A lower taxonomic resolution inevitably increases the functional potential of a taxon and broadens its ecological niche, which thereby likely weakens direct and indirect interactions between taxa. This would also explain the more evenly distributed correlations between orders and families when compared to genera and species and the lower keystone potentials. Keystone taxa are of great interest for systematic microbiome engineering, for example for the restoration of a perturbed gut microbiome through the introduction or targeted manipulation of keystones. In applied fields a microbial species or genus may well present a more practical target, as opposed to broader taxonomic groups.

Overall, the correlation networks have a high percentage of positive correlations (Fig 4A). This is somewhat surprising, as work by Coyte and colleagues (Coyte et al., 2015) has suggested that a diverse microbial community can only be stable when the majority of interactions are negative. Interestingly, the single-species keystone genera identified through our analysis have a considerably higher percentage of negative correlations to other genera (Fig 4C and Table 1). These high numbers of negative correlations with keystone taxa may provide stability to the microbial community, especially considering the particularly high percentage of positive correlations between species of the same genus (Fig 4B). The negative correlations

with keystone genera potentially counteract these positive correlations that could otherwise result in positive feedback loops and lead to instability. Coyte and colleagues (Coyte et al., 2015) identify the dampening of these positive feedback loops as one mechanism for stabilization of a community. They furthermore suggest weaker ecological interactions as an additional stabilization mechanism. We indeed observe that the stronger the correlation of a keystone to a genus, the weaker the correlations between the species of the respective genus (Fig 4D), suggesting that the keystone genera may be both dampening positive feedback loops and weakening intergenic correlations. We additionally find that species' positive correlations tend to be weaker in larger genera (Fig S3A). This result supports another observation by Coyte et al. (Coyte et al., 2015) that functional redundancy in closely related taxa provides stability by replacing few strong interactions with many weaker interactions. In general, taxa correlated to a keystone taxon tend to have more, but weaker correlations (Fig S3B and C), again indicating a stabilizing effect from keystone taxa.

In both the species- as well as genus-level network we identified four clusters of co-occurring taxa (Fig 5). The analyzed gut microbiomes all consisted of more than one cluster, both on species as well as genus level, resulting in a compositional gradient ranging from strongly dominated by a particular cluster to a fairly even distribution of all four clusters (Fig S3). Frioux and colleagues (Frioux et al., 2023) recently described similar patterns in the genus composition of gut microbial communities. They identified five so-called enterosignatures (commonly co-occurring genera) that, in combination, can accurately describe most healthy gut microbiomes. These five enterosignatures show strong similarities to the genus clusters we were able to identify in our correlation network. Specifically, ES-Firm (enterosignature with a high contribution of various *Bacillota*) and ES-Prev (dominated by *Prevotella*) correspond to cluster 3 and cluster 4 in Figure 4, respectively. An enterosignature mainly found in the gut microbiome of infants, ES-Bifi, does not have a corresponding genus cluster in the cohort we analyzed, which consists solely of adults. However, the main genera of ES-Bifi, namely *Bifidobacterium*, *Streptococcus*, *Veillonella*, *Enterococcus* and *Haemophilus*, are members of genus cluster 1, together with *Bacteroides*, one of the main genera in ES-Bact. The remaining cluster 2 includes *Escherichia*, the main contributor to ES-Esch, and *Blautia*, a strong contributor to ES-Firm, but most genera in this cluster were not reported as strongly contributing to an enterosignature. These differences may be due to differences in the applied methods (correlation analysis vs. non-negative matrix factorization) as well as size and diversity of the analyzed cohorts. While not a perfect match to the described enterosignatures, the genus clusters we identified are still remarkably similar in composition. In contrast to the study by Frioux and colleagues (Frioux et al., 2023) we focused on the analysis of samples from healthy adults and can therefore not draw any conclusions on the relationship between the identified keystones and disease states. The presence of the keystones does not imply a healthy microbiome nor does their absence imply a perturbed microbiome.

We observe distinct transcriptional states in all putative keystone taxa that can be attributed to differential transcription of particular ECs. In a small synthetic gut microbial community, Shetty and colleagues (Shetty et al., 2022) observed shifting functional roles of species in an otherwise compositionally and functionally fairly stable community. These observations suggest functional versatility that enables these taxa to adjust their metabolism to available resources and shifting ecological niches. Particularly striking is the high number of ECs from *Methanobrevibacter* involved in methane metabolism that we could identify as important in distinguishing the transcriptional states of *Methanobrevibacter*. In *Bilophila*, we identified an EC involved in taurine and hypotaurine metabolism as important in distinguishing transcriptional states and an EC potentially involved in host glycan degradation that shows an opposing transcription pattern, suggesting a metabolic shift. These findings are consistent with the idea that the gut microbiome needs to be able to accommodate a diverse and variable set of available resources as well as dynamic interplay with the host. Together with the stabilizing effects previously discussed, functional versatility may be another mechanism enabling keystone taxa to exert their influence on the microbial community.

In an intriguing study on gnotobiotic mice, Weiss and colleagues (Weiss et al., 2023) observe that functional roles of individual taxa are strongly dependent on the environment and question the existence of universally valid keystone species in the gut microbiome. While we also observe that the transcriptional activity of keystones is variable, our results suggest that the complex intestinal microbial community is modular and keystone taxa exert a stabilizing effect on subsets of the microbiome. These subsets are relatively independent and co-occur in individual gut microbiomes at various frequencies. However, experimental validation is needed to strengthen this hypothesis and in particular to further our understanding of the metabolic interplay underlying the observed dynamics (Tudela et al., 2021).

LIMITATIONS OF THE STUDY

The present study has several limitations. We focused exclusively on the analysis of publicly available data and no experimental validation was carried out. Metagenome and metatranscriptome reads were mapped to reference genomes and analyzed on different taxonomic resolutions up to species. Consequently, strain-level variation has not been taken into account and the reliance on reference genomes further reduces the genetic variation represented in this study. We also recognize that correlation networks can reflect other ecological processes beyond interactions such as habitat filtering. Additionally, our analyses focused only on prevalent taxa and therefore would miss any less common keystone taxa. Our observations are furthermore limited to the distal colon, while community composition as well as functionality may differ in more proximal parts of the intestines. Lastly, the data we used for this study consisted of two cohorts with limited demographic diversity and different putative keystones may be present in other populations. Our study focused on healthy adults from the USA and does not represent and cannot readily be translated to other world populations.

CONCLUSION

In this study we developed a pipeline (available on github) for the robust construction and analysis of correlation networks and the identification of putative keystone taxa, namely *Agathobaculum butyriciproducens, Bilophila wadsworthia, Eisenbergiella tayi, Firmicutes bacterium CAG* 83 (unclassified *Bacillota, Holdemania filiformis, Methanobrevibacter smithii, Ruminococcus lactaris* and *Veillonella atypica*. Through a comprehensive analysis of the constructed correlation networks we are able to show that correlation networks and their properties are highly sensitive to taxonomic resolution. Furthermore, we identified signatures of potentially community stabilizing interaction patterns reflected in the correlation networks as well as co-occurring sub-communities in the human gut microbiome that show a high similarity to previously described enterosignatures. The ecological significance of correlation network keystones is still an open question. Nevertheless, we see indications that keystone taxa may have a local stabilizing effect on sub-communities of the gut microbiome. This suggests that it is unlikely that there are individual taxa that globally affect the entire community. We rather hypothesize that keystone taxa act as stabilizing agents in relatively independently co-occurring subsets of the microbiome.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at:
https://www.ncbi.nlm.nih.gov/bioproject/PRJNA354235/ and
https://www.ncbi.nlm.nih.gov/bioproject/PRJNA389280/.
Code and a wrapper function for the established workflow to construct correlation networks and compute keystone potential of individual taxa can be found at:
https://github.com/fbauchinger/correlation.network.keystones_workflow.

ETHICS STATEMENT

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

FB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. DS: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. DB: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

FUNDING

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by ERC starting grant 714623 "FunKeyGut" and

the Austrian Science Fund (FWF) (10.55776/COE7). Work in the laboratory of DB is supported by the FWF project MAINTAIN DOC 69 doc.fund (10.55776/DOC69).

ACKNOWLEDGEMENTS

A previous version of this work was deposited on a preprint server (Bauchinger et al., 2023).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

REFERENCES

Ahn, S., Jin, T.-E., Chang, D.-H., Rhee, M.-S., Kim, H. J., Lee, S. J., et al. (2016). Agathobaculum butyriciproducens gen. nov. sp. nov., a strict anaerobic, butyrate-producing gut bacterium isolated from human faeces and reclassification of Eubacterium desmolans as Agathobaculum desmolans comb. nov. *Int. J. Syst. Evol. Microbiol.* 66, 3656–3661. doi: 10.1099/ijsem.0.001195

Amir, I., Bouvet, P., Legeay, C., Gophna, U., and Weinberger, A. (2014). Eisenbergiella tayi gen. nov., sp. nov., isolated from human blood. *Int. J. Syst. Evol. Microbiol.* 64, 907–914. doi: 10.1099/ijs.0.057331-0

Amit, G., and Bashan, A. (2023). Top-down identification of keystone taxa in the microbiome. *Nat. Commun.* 14, 3951. doi: 10.1038/s41467-023-39459-5

Archer, E. (2023). rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. Available at: https://cran.r-project.org/web/packages/rfPermute/index.html (Accessed September 22, 2023).

Arita, S., and Inagaki-Ohara, K. (2019). High-fat-diet–induced modulations of leptin signaling and gastric microbiota drive precancerous lesions in the stomach. *Nutrition* 67–68, 110556. doi: 10.1016/j.nut.2019.110556

Banerjee, S., Schlaeppi, K., and van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* 16, 567–576. doi: 10.1038/s41579-018-0024-1

Baron, E. J., Summanen, P., Downes, J., Roberts, M. C., Wexler, H., and Finegold, S. M. (1989). Bilophila wadsworthia, gen. nov. and sp. nov., a Unique Gram-negative Anaerobic Rod Recovered from Appendicitis Specimens and Human Faeces. *Microbiology* 135, 3405–3411. doi: 10.1099/00221287-135-12-3405

Bauchinger, F., Seki, D., and Berry, D. (2023). Characteristics of putative keystones in the healthy adult human gut microbiome as determined by correlation network analysis. 2023.11.20.567895. doi: 10.1101/2023.11.20.567895

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10, e65088. doi: 10.7554/eLife.65088

Belzer, C., Chia, L. W., Aalvink, S., Chamlagain, B., Piironen, V., Knol, J., et al. (2017). Microbial Metabolic Networks at the Mucus Layer Lead to Diet-Independent Butyrate and Vitamin B12 Production by Intestinal Symbionts. *mBio* 8, 10.1128/mbio.00770-17. doi: 10.1128/mbio.00770-17

Bernard, K., Burdz, T., Wiebe, D., Balcewich, B. M., Zimmerman, T., Lagacé-Wiens, P., et al. (2017). Characterization of isolates of Eisenbergiella tayi, a strictly anaerobic Gram-stain variable bacillus recovered from human clinical materials in Canada. *Anaerobe* 44, 128–132. doi: 10.1016/j.anaerobe.2017.03.005

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2014.00219 (Accessed April 4, 2023).

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brugiroux, S., Beutler, M., Pfann, C., Garzetti, D., Ruscheweyh, H.-J., Ring, D., et al. (2016). Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat. Microbiol.* 2, 1–12. doi: 10.1038/nmicrobiol.2016.215

Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., and Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Health Dis.* 26, 10.3402/mehd.v26.26191. doi: 10.3402/mehd.v26.26191

Cartmell, A., Muñoz-Muñoz, J., Briggs, J. A., Ndeh, D. A., Lowe, E. C., Baslé, A., et al. (2018). A surface endogalactanase in Bacteroides thetaiotaomicron confers keystone status for arabinogalactan degradation. *Nat. Microbiol.* 3, 1314–1326. doi: 10.1038/s41564-018-0258-8

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164

Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science* 350, 663–666. doi: 10.1126/science.aad2602

Csárdi, G., Nepusz, T., Müller, K., Horvát, S., Traag, V., Zanini, F., et al. (2023). igraph for R: R interface of the igraph library for graph theory and network analysis. doi: 10.5281/zenodo.8240644

Dalgaard, P. (2010). R Development Core Team (2010): R: A language and environment for statistical computing. Available at: https://research.cbs.dk/en/publications/r-development-core-team-2010-r-a-language-and-environment-for-sta (Accessed September 22, 2023).

Dridi, B., Henry, M., Khéchine, A. E., Raoult, D., and Drancourt, M. (2009). High Prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae Detected in the Human Gut Using an Improved DNA Detection Protocol. *PLOS ONE* 4, e7063. doi: 10.1371/journal.pone.0007063

Egland, P. G., Palmer, R. J., and Kolenbrander, P. E. (2004). Interspecies communication in Streptococcus gordonii–Veillonella atypica biofilms: Signaling in flow conditions requires juxtaposition. *Proc. Natl. Acad. Sci.* 101, 16917–16922. doi: 10.1073/pnas.0407457101

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832

Feng, Z., Long, W., Hao, B., Ding, D., Ma, X., Zhao, L., et al. (2017). A human stool-derived Bilophila wadsworthia strain caused systemic inflammation in specific-pathogen-free mice. *Gut Pathog.* 9, 59. doi: 10.1186/s13099-017-0208-7

Fisher, C. K., and Mehta, P. (2014). Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. *PLOS ONE* 9, e102451. doi: 10.1371/journal.pone.0102451

Frankel, A. E., Coughlin, L. A., Kim, J., Froehlich, T. W., Xie, Y., Frenkel, E. P., et al. (2017). Metagenomic Shotgun Sequencing and Unbiased Metabolomic Profiling Identify Specific Human Gut Microbiota and Metabolites Associated with Immune Checkpoint Therapy Efficacy in Melanoma Patients. *Neoplasia N. Y. N* 19, 848–855. doi: 10.1016/j.neo.2017.08.004

Friedman, J., and Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLOS Comput. Biol.* 8, e1002687. doi: 10.1371/journal.pcbi.1002687

Frioux, C., Ansorge, R., Özkurt, E., Ghassemi Nedjad, C., Fritscher, J., Quince, C., et al. (2023). Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* 31, 1111-1125.e6. doi: 10.1016/j.chom.2023.05.024

Go, J., Chang, D.-H., Ryu, Y.-K., Park, H.-Y., Lee, I.-B., Noh, J.-R., et al. (2021). Human gut microbiota Agathobaculum butyriciproducens improves cognitive impairment in LPS-induced and APP/PS1 mouse models of Alzheimer's disease. *Nutr. Res.* 86, 96–108. doi: 10.1016/j.nutres.2020.12.010

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., et al. (2014). Human genetics shape the gut microbiome. *Cell* 159, 789–799. doi: 10.1016/j.cell.2014.09.053

Gophna, U., Konikoff, T., and Nielsen, H. B. (2017). Oscillospira and related bacteria – From metagenomic species to metabolic features. *Environ. Microbiol.* 19, 835–841. doi: 10.1111/1462-2920.13658

Hamaker, B. R., and Tuncil, Y. E. (2014). A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota. *J. Mol. Biol.* 426, 3838–3850. doi: 10.1016/j.jmb.2014.07.028

Hausmann, B., Pelikan, C., Rattei, T., Loy, A., and Pester, M. (2019). Long-Term Transcriptional Activity at Zero Growth of a Cosmopolitan Rare Biosphere Member. *mBio* 10, 10.1128/mbio.02189-18. doi: 10.1128/mbio.02189-18

Horz, H.-P., and Conrads, G. (2010). The Discussion Goes on: What Is the Role of Euryarchaeota in Humans? *Archaea* 2010, 967271. doi: 10.1155/2010/967271

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30.

Kassambara, A., and Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. Available at: https://cran.r-project.org/web/packages/factoextra/index.html (Accessed September 22, 2023).

Kharofa, J., Apewokin, S., Alenghat, T., and Ollberding, N. J. (2023). Metagenomic analysis of the fecal microbiome in colorectal cancer patients compared to healthy controls as a function of age. *Cancer Med.* 12, 2945–2957. doi: 10.1002/cam4.5197

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

Lee, D. W., Ryu, Y.-K., Chang, D.-H., Park, H.-Y., Go, J., Maeng, S.-Y., et al. (2022). Agathobaculum butyriciproducens Shows Neuroprotective Effects in a 6-OHDA-Induced Mouse Model of Parkinson’s Disease. *J. Microbiol. Biotechnol.* 32, 1168–1177. doi: 10.4014/jmb.2205.05032

Liu, A., Ma, T., Xu, N., Jin, H., Zhao, F., Kwok, L.-Y., et al. (2021). Adjunctive Probiotics Alleviates Asthmatic Symptoms via Modulating the Gut Microbiome and Serum Metabolome. *Microbiol. Spectr.* 9, e00859-21. doi: 10.1128/Spectrum.00859-21

Liu, F., Li, Z., Wang, X., Xue, C., Tang, Q., and Li, R. W. (2019). Microbial Co-Occurrence Patterns and Keystone Species in the Gut Microbial Community of Mice in Response to Stress and Chondroitin Sulfate Disaccharide. *Int. J. Mol. Sci.* 20, 2130. doi: 10.3390/ijms20092130

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550, 61–66. doi: 10.1038/nature23889

Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V., and Thiele, I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front. Genet.* 6, 148. doi: 10.3389/fgene.2015.00148

Mays, T. D., Holdeman, L. V., Moore, W. E. C., Rogosa, M., and Johnson, J. L. (1982). Taxonomy of the Genus Veillonella Prévot. *Int. J. Syst. Evol. Microbiol.* 32, 28–36. doi: 10.1099/00207713-32-1-28

Mazier, W., Le Corf, K., Martinez, C., Tudela, H., Kissi, D., Kropp, C., et al. (2021). A New Strain of Christensenella minuta as a Potential Biotherapy for Obesity and Associated Metabolic Diseases. *Cells* 10, 823. doi: 10.3390/cells10040823

Mehta, R. S., Abu-Ali, G. S., Drew, D. A., Lloyd-Price, J., Subramanian, A., Lochhead, P., et al. (2018). Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* 3, 347–355. doi: 10.1038/s41564-017-0096-0

Miller, T. L., Wolin, M. J., de Macario, E. C., and Macario, A. J. (1982). Isolation of Methanobrevibacter smithii from human feces. *Appl. Environ. Microbiol.* 43, 227–232.

Moore, W. E. C., Johnson, J. L., and Holdeman, L. V. (1976). Emendation of Bacteroidaceae and Butyrivibrio and Descriptions of Desulfomonas gen. nov. and Ten New Species in the Genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus. *Int. J. Syst. Evol. Microbiol.* 26, 238–252. doi: 10.1099/00207713-26-2-238

Moreno-Arrones, O. m., Serrano-Villar, S., Perez-Brocal, V., Saceda-Corralo, D., Morales-Raya, C., Rodrigues-Barata, R., et al. (2020). Analysis of the gut microbiota in alopecia areata: identification of bacterial biomarkers. *J. Eur. Acad. Dermatol. Venereol.* 34, 400–405. doi: 10.1111/jdv.15885

Natividad, J. M., Lamas, B., Pham, H. P., Michel, M.-L., Rainteau, D., Bridonneau, C., et al. (2018). Bilophila wadsworthia aggravates high fat diet induced metabolic dysfunctions in mice. *Nat. Commun.* 9, 2802. doi: 10.1038/s41467-018-05249-7

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., et al. (2022). vegan: Community Ecology Package. Available at: https://cran.r-project.org/web/packages/vegan/index.html (Accessed September 22, 2023).

Paine, R. T. (1969). A Note on Trophic Complexity and Community Stability. *Am. Nat.* 103, 91–93.

Pearce, C. I., Pattrick, R. A. D., Law, N., Charnock, J. M., Coker, V. S., Fellowes, J. W., et al. (2009). Investigating different mechanisms for biogenic selenite transformations: Geobacter sulfurreducens, Shewanella oneidensis and Veillonella atypica. *Environ. Technol.* 30, 1313–1326. doi: 10.1080/09593330902984751

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

Rakoff-Nahoum, S., Coyne, M. J., and Comstock, L. E. (2014). An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts. *Curr. Biol.* 24, 40–49. doi: 10.1016/j.cub.2013.10.077

Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R., and Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature* 591, 633–638. doi: 10.1038/s41586-021-03241-8

Risely, A. (2020). Applying the core microbiome to understand host–microbe systems. *J. Anim. Ecol.* 89, 1549–1558. doi: 10.1111/1365-2656.13229

Scheiman, J., Luber, J. M., Chavkin, T. A., MacDonald, T., Tung, A., Pham, L.-D., et al. (2019). Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nat. Med.* 25, 1104–1109. doi: 10.1038/s41591-019-0485-4

Schirmer, M., Franzosa, E. A., Lloyd-Price, J., McIver, L. J., Schwager, R., Poon, T. W., et al. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* 3, 337–346. doi: 10.1038/s41564-017-0089-z

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharma, S., and Tripathi, P. (2019). Gut microbiome and type 2 diabetes: where we are and where to go? *J. Nutr. Biochem.* 63, 101–108. doi: 10.1016/j.jnutbio.2018.10.003

Shetty, S. A., Kostopoulos, I., Geerlings, S. Y., Smidt, H., de Vos, W. M., and Belzer, C. (2022). Dynamic metabolic interactions and trophic roles of human gut microbes identified using a minimal microbiome exhibiting ecological properties. *ISME J.* 16, 2144–2159. doi: 10.1038/s41396-022-01255-2

Surono, I. S., Jalal, F., Bahri, S., Romulo, A., Kusumo, P. D., Manalu, E., et al. (2021). Differences in immune status and fecal SCFA between Indonesian stunted children and children with normal nutritional status. *PLOS ONE* 16, e0254300. doi: 10.1371/journal.pone.0254300

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., et al. (2019). The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host Microbe* 26, 666-679.e7. doi: 10.1016/j.chom.2019.08.018

Trosvik, P., and de Muinck, E. J. (2015). Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome* 3, 44. doi: 10.1186/s40168-015-0107-4

Tudela, H., Claus, S. P., and Saleh, M. (2021). Next Generation Microbiome Research: Identification of Keystone Species in the Metabolic Regulation of Host-Gut Microbiota Interplay. *Front. Cell Dev. Biol.* 9. Available at: https://www.frontiersin.org/articles/10.3389/fcell.2021.719072 (Accessed April 4, 2023).

Varsadiya, M., Urich, T., Hugelius, G., and Bárta, J. (2021). Fungi in Permafrost-Affected Soils of the Canadian Arctic: Horizon- and Site-Specific Keystone Taxa Revealed by Co-Occurrence Network. *Microorganisms* 9, 1943. doi: 10.3390/microorganisms9091943

Wang, X.-W., Sun, Z., Jia, H., Michel-Mata, S., Angulo, M. T., Dai, L., et al. (2024). Identifying keystone species in microbial communities using deep learning. *Nat. Ecol. Evol.* 8, 22–31. doi: 10.1038/s41559-023-02250-2

Watts, S. C., Ritchie, S. C., Inouye, M., and Holt, K. E. (2019). FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* 35, 1064–1066. doi: 10.1093/bioinformatics/bty734

Weiss, A. S., Niedermeier, L. S., von Strempel, A., Burrichter, A. G., Ring, D., Meng, C., et al. (2023). Nutritional and host environments determine community ecology and keystone species in a synthetic gut bacterial community. *Nat. Commun.* 14, 4780. doi: 10.1038/s41467-023-40372-0

Willems, A., Moore, W. E. C., Weiss, N., and Collins, M. D. (1997). Phenotypic and Phylogenetic Characterization of Some Eubacterium-Like Isolates Containing a Novel Type B Wall Murein from Human Feces: Description of Holdemania filiformis gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* 47, 1201–1204. doi: 10.1099/00207713-47-4-1201

Wilson, A. S., Koller, K. R., Ramaboli, M. C., Nesengani, L. T., Ocvirk, S., Chen, C., et al. (2020). Diet and the Human Gut Microbiome: An International Review. *Dig. Dis. Sci.* 65, 723–740. doi: 10.1007/s10620-020-06112-w

Xi, W., Gao, X., Zhao, H., Luo, X., Li, J., Tan, X., et al. (2021). Depicting the composition of gut microbiota in children with tic disorders: an exploratory study. *J. Child Psychol. Psychiatry* 62, 1246–1254. doi: 10.1111/jcpp.13409

Ze, X., Duncan, S. H., Louis, P., and Flint, H. J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* 6, 1535–1543. doi: 10.1038/ismej.2012.4

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

Zhou, X., Baumann, R., Gao, X., Mendoza, M., Singh, S., Sand, I. K., et al. (2022). Gut microbiome of multiple sclerosis patients and paired household healthy controls reveal associations with disease risk and course. *Cell* 185, 3467-3486.e16. doi: 10.1016/j.cell.2022.08.021

CHAPTER 2

Metatranscriptomic-driven insights into mucosal glycan degradation by the human gut microbiota

Franziska Bauchinger[1, 2], David Berry[1,3*]

[1]University of Vienna, Centre for Microbiology and Environmental Systems Science CeMESS, Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, Vienna, Austria

[2]University of Vienna, Doctoral School in Microbiology and Environmental Science, Vienna, Austria

[3]Joint Microbiome Facility of the Medical University of Vienna and the University of Vienna, Vienna, Austria

Manuscript submitted for review on January 30th, 2025

ABSTRACT

The secreted mucus layer in the human gastrointestinal tract constitutes both a protective boundary between gut lumen and epithelium as well as an important nutrient source for members of the gut microbiota. While many gut microbes possess the genetic potential to degrade mucin it is still unclear which degrade mucosal glycans *in vivo*. Here, we systematically analyze publicly available metagenome and metatranscriptome datasets to characterize the gut microbial community involved in mucosal glycan degradation. We utilize co-occurrence network analysis and linear regression to elucidate the ecological strategies of, and relationship between, mucus degraders. We find that although approximately 60% of species carrying genes encoding for mucosal-glycan-degrading enzymes have detectable transcription of these genes, only 21 species prevalently transcribe more than 1 gene. Furthermore, the transcription of individual genes is frequently dominated by single species in individual samples. Transcription patterns suggest the presence of competitive mucosal glycan degraders characterized by abundance-driven transcription that are negative predictors for the transcription of other degraders as well as opportunistic species with decoupled abundance and transcription profiles. These findings provide insights into the ecology of the mucosal glycan degradation niche in the human gut microbiota.

INTRODUCTION

The human gastrointestinal (GI) tract is inhabited by a diverse community of microorganisms that form the human gut microbiota. These microbes provide essential functions to their human host such as metabolizing dietary compounds (Hamaker and Tuncil 2014), conferring resistance to pathogens (Brugiroux et al. 2016) and immunomodulation (Zheng, Liwinski and Elinav 2020). The microbiota reside in the gut lumen and are prevented from invading the gut epithelium in part via the action of the secreted mucus layer. Mucin, the main component of secreted mucus, is a heavily-glycosylated glycoprotein that is secreted by goblet cells in the epithelium of the GI tract. An inner, denser mucus layer is virtually free of microbes while a looser outer layer functions as a habitat for members of the gut microbiota (Johansson et al. 2008). Some gut microbes can also utilize the glycans that form the mucosal structure as an energy source (Paone and Cani 2020). The mucus layer of the GI tract is formed from branched, cross-linked glycan chains consisting of galactose (Gal), N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), fucose (Fuc) and sialic acid (Neu5Ac) (Podolsky 1985; Robbe et al. 2004). These glycan chains, which represent around 80% of mucin mass (Gendler and Spicer 1995), are largely linked through GalNAc moieties to serine or threonine residues of the mucus protein backbone (Brockhausen, Schachter and Stanley 2009). While the exact structure and length of the glycan chains can vary considerably, the terminal sugars are either GalNAc, Fuc or Neu5Ac (Podolsky 1985; Brockhausen, Schachter and Stanley 2009; Tailford et al. 2015). Gal or GlcNAc are additionally frequently sulfated (Luis et al. 2021). Degradation of these complex structures requires a multitude of different carbohydrate-active enzymes (CAZymes). Bioinformatic analysis of the genomic potential of gut microbiota to degrade mucosal glycans has predicted that up to 86% of this microbial community harbor genes linked to mucus degradation (Ravcheev and Thiele 2017). Few species, however, encode a larger array of the necessary enzymes and are able to extensively degrade mucin, such as *Akkermansia muciniphila* (Shuoker et al. 2023) or *Bacteroides thetaiotaomicron* (Kostopoulos et al. 2021). As has been shown experimentally, mucus degradation is usually a communal process and individual species are capable of breaking only specific linkages between glycans (Pudlo et al. 2015; Raimondi et al. 2021). Individual glycan-derived sugars can also be available to non-degraders as many mucosal glycan degradation enzymes act extracellularly (Yamaguchi and Yamamoto 2023). *A. muciniphila*, for example, produces a number of sialidases and fucosidases, enabling it to remove terminal sugars and thereby not only facilitating further degradation of glycans, but also sharing nutrients with co-occurring microbes (Shuoker et al. 2023). *Ruminococcus gnavus* has been shown to liberate mucosal glycans and make them accessible to *B. thetaiotaomicron* (Schaus et al. 2024), a species viewed as a glycan generalist that encodes multiple mucus degradation CAZymes but shows a preference for other substrates (Pudlo et al. 2015). An imbalance between mucus production by the epithelial cells and degradation by the microbiome has been linked to various health issues such as colitis (Fang et al. 2021) and colorectal cancer (Coleman and Haller 2021), and thinning of the mucus layer may be linked to dietary fiber deprivation

(Jegatheesan, Moorthy and Eberl 2024). Understanding both the composition of the microbial community that degrades mucosal glycans as well as the ecological principles governing the mucus degradation guild is consequently of high interest. The complex mucosal structure makes it a nutrient source that may support an equally complex microbial community and allow for the coexistence of both highly specialized and competitive mucus degraders as well as opportunistic microbes. Mucus degradation is an interplay between the production of glycans by the gut epithelium, competition of gut microbes for these glycans (Kostopoulos et al. 2021) and the necessity for collaborative degradation due to individual microbes' inability to degrade the complete mucosal structure (Arike and Hansson 2016; Shuoker et al. 2023). As such it can provide interesting insights into the ecological properties of gut microbial communities. Previous work made it apparent that a surprisingly large number of gut microbes are theoretically able to degrade mucosal structures, while work on individual mucin degraders and co-cultures furthered our understanding of the underlying ecological strategies. However, we lack knowledge on the properties of *in vivo* mucus degrading communities. It is as of yet unclear which community members actively transcribe mucosal glycan degradation enzymes in the GI tract and how these species navigate the apparent overlap in ecological niche. Here, we aim to address some of these open questions through the analysis of publicly available metatranscriptomes from stool samples of healthy adults. We investigate which mucin degradation enzymes can be detected in the transcriptomes and identify microbial species that prevalently transcribe these enzymes. We then analyze whether abundance and transcriptional patterns of mucosal glycan degraders are linked to elucidate differences in ecological strategy.

MATERIAL AND METHODS

Processing of raw data

We analyzed the paired metagenome and metatranscriptome reads from 574 stool samples of two publicly available datasets: a cohort of male health professionals (Mehta et al., 2018), and a study on inflammatory bowel disease (Schirmer et al., 2018), focusing on the healthy control participants in the second dataset. We processed the raw reads with trimmomatic (Bolger et al., 2014) (leading: 3, trailing: 3, slidingwindow: 4:15, minlen: 50), removed samples with under one million reads and analyzed the trimmed reads with HUMAnN 3.9 (Beghini et al., 2021). Within HUMAnN 3.9, we here used MetaphlAn 3.1 (Beghini et al., 2021) to estimate community composition, map reads to a community pangenome with bowtie2 (version 2.5) (Langmead and Salzberg, 2012) and align unmapped reads to a protein database using DIAMOND 2.1.9 (Buchfink et al., 2015). This combination allows for an analysis of both the taxonomic and the functional profiles of metagenome and metatranscriptome reads. We used the MetaphlAn setting rel_ab_w_read_stats to obtain both relative abundances as well as estimated read counts in the taxonomic profiles. In the functional profiles, reads were mapped to gene families and we further grouped them into enzyme commission numbers (ECs) with the HUMAnN function humann_regroup_table and the flag - -groups uniref90_level4ec. Gene family and EC values are given as copies per million.

Identification of ECs involved in mucin glycan degradation

We identified ECs of interest by cross-referencing glycosyl hydrolase families presumed to be involved in mucin glycan degradation, as reviewed by Ravcheev and Thiele (Ravcheev and Thiele 2017), Berkhout and colleagues (Berkhout, Plugge and Belzer 2022) as well as Glover et al. (Glover, Ticer and Engevik 2022), with the corresponding entries in the Carbohydrate Active Enzymes (CAZy) database (https://www.cazy.org/) (Drula et al. 2022) and in InterPro (Paysan-Lafosse et al. 2023). In addition to glycosyl hydrolases we also looked at sulfatases involved in mucin degradation (Berkhout, Plugge and Belzer 2022). The corresponding ECs of the sulfatases were obtained from the Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000). The distribution of reads grouping to the respective ECs was visualized utilizing k-means clustering with the function kmeans in the R package stats (Dalgaard 2010). In order to compare a species' mucosal glycan degradation gene repertoire with their overall capacity to hydrolyze glycans we extracted all ECs corresponding to GH families (EC 3.2.1.-) (Drula et al. 2022) from the metagenomes and metatranscriptomes. The investigated ECs and corresponding GH families can be found in Table S1.

Identification of mucosal glycan degraders

We compiled all species encoding any of the identified ECs and then went on to filter for species transcribing these ECs. We removed species only transcribing alpha- and/or beta-galactosidase from further analysis due to the broad dissemination of galactose. We also did not further consider species only transcribing one of the identified ECs and removed

species transcribing ECs in less than 10% of samples with detectable transcription. These steps resulted in a list of versatile (transcribing more than 1 EC) as well as prevalent (transcribing at least 1 EC in at least 10% of samples with detectable transcription) mucosal glycan degraders. The distribution of both the encoding and transcribing species amongst phylogenetic groups was visualized in a radial cladogram. The cladogram was built with the function ggtree from the R package of the same name (Xu et al. 2022) and species were grouped based on order, family and genus.

Co-occurrence network of mucus degraders

We computed species co-occurrence networks to better understand abundance patterns of mucus degraders. These networks are based on Spearman correlations between the centered log-ratio (clr) transformed estimated read counts of species. We used centered log-ratio transformation to account for the compositional nature of metagenome sequencing data (Kucera and Malmgren 1998). For each sample, we divided the estimated read counts by the geometric mean and then log-transformed the ratio, adding a pseudocount of 1 to avoid non-finite numbers. Then we computed pairwise Spearman correlations between species with these log-transformed ratios using the R function cor() with method = “spearman”. We estimated the significance of the observed correlations by comparing them to a null distribution. This null distribution was obtained by randomly shuffling read counts and calculating Spearman correlations from the shuffled data 1000 times. We considered observed correlations to be significant if they were outside the 99% confidence interval of the null distribution. We then removed all negative correlations as well as positive correlations < 0.2 and filtered for species correlated to any of the investigated mucosal glycan degraders (as defined above). This resulted in a network of species exhibiting a significant and strong co-occurrence pattern with the mucosal glycan degraders. We used the function cluster_fast_greedy from the R package igraph (Csárdi et al. 2023) on the adjacency matrix of the network (based on presence/absence of correlations) to further cluster these species according to their co-occurrence patterns. The co-occurrence network was visualized with Cytoscape 3.10.1 (Shannon et al. 2003), with network nodes representing species and edges (connections between nodes) representing significant positive correlations between mucosal glycan degraders and their first neighbors (species with a positive correlation to a mucosal glycan degrader). A node and edge table of the computed co-occurrence network is available in Table S2.

Linear models between species abundance and transcription

In order to understand the relationship between the abundance of mucosal glycan degraders and their contribution to transcription of degradation enzymes we utilized linear models. For each investigated EC we computed linear models between the abundance of each prevalently contributing species (clr transformed estimated read numbers) and their respective contribution to EC transcription (% of total EC transcription per sample). Linear model statistics are reported in Table S3. We considered a model with a p-value < 0.05 as significant

and used the adjusted R^2 values to estimate how much variation in transcription is explained by the abundance patterns of the respective species.

Response-predictor networks of EC transcription

We further utilized linear models to understand the relationship between mucosal glycan degraders. Specifically, we computed linear models for each investigated EC to estimate how the individual abundance patterns of prevalently transcribing species (detectable transcription in at least 10% of samples where a respective EC is transcribed) in turn influence the transcription patterns of each of those transcribing species. For each EC the linear model looks as follows: *Transcription species 1 [% of total EC transcription] ~ abundance species 2 [clr transformed read counts] + … + abundance species n [clr transformed read counts]*. Significant coefficients in the linear model indicate that transcription of the response species (species 1) is predicted by the abundance of a predictor species (species 2 - n). The detailed linear model statistics can be found in Table S4. We again used Cytoscape 3.10.1 (Shannon et al. 2003) to visualize these linear model coefficients as edges in response-predictor networks. Lastly, we built a consensus network combining all edges across response-predictor networks of investigated mucosal-glycan-degrading enzymes.

RESULTS

Ten ECs involved in the degradation of mucosal glycans are prevalent in the analyzed metagenomes and metatranscriptomes

In previous work a large suite of enzymes has been identified as being potentially involved in the degradation of mucosal glycans (Ravcheev and Thiele 2017; Berkhout, Plugge and Belzer 2022; Glover, Ticer and Engevik 2022). In this study we analyzed 574 paired metagenomes and metatranscriptomes (Mehta et al. 2018; Schirmer et al. 2018) at the level of enzyme commission numbers (ECs) to evaluate whether the genes and transcripts encoding these enzymes could be detected in the stool microbiomes of healthy adults. We screened for 18 ECs, comprising 13 glycosyl hydrolases, 1 lyase and 4 sulfatases (Table S1) and analyzed their abundance and prevalence. In only 10 of the analyzed metatranscriptome samples none of the ECs were detected. From the total of 18 ECs, 11 were detectable in the analyzed dataset (Table 1) and 10 ECs were prevalently (prevalence > 5%) detected (Fig 1). Reads aligning to EC 3.1.6.14 were only found in ~1% of samples and could not be mapped to any reference genomes. We therefore did not include this enzyme in further analyses.

A small subset of bacteria encoding mucosal-glycan-degrading enzymes transcribe these genes

We identified 395 species, belonging to 49 different families, that encoded one or more of the detected ECs. However, only 238 of these species, belonging to 33 different families, actively transcribed at least 1 gene in the analyzed dataset (Fig 2A and Fig S1). We excluded species transcribing only one of the compiled ECs to focus on a subset of species that we consider versatile mucin degraders. Additionally, we excluded species only transcribing alpha-galactosidase and/or beta-galactosidase. Both of these enzymes cleave off galactose, and considering that galactose is a sugar widely occurring in dietary compounds and the focus of this study is the degradation of mucosal glycans, we chose not to further consider these species. Lastly, we only focused on species transcribing at least 1 EC in at least 10% of the analyzed samples. This resulted in 21 versatile as well as prevalent mucosal glycan degraders (Fig 2B and Fig S2B), belonging to the families *Akkermansiaceae* (1 species)*, Bacteroidaceae* (10 species), *Eubacteriaceae* (1 species)*, Lachnospiraceae* (4 species)*, Rikenellaceae* (2 species)*, Ruminococcaceae* (1 species), and *Tannerellaceae* (2 species) (Fig 2B).

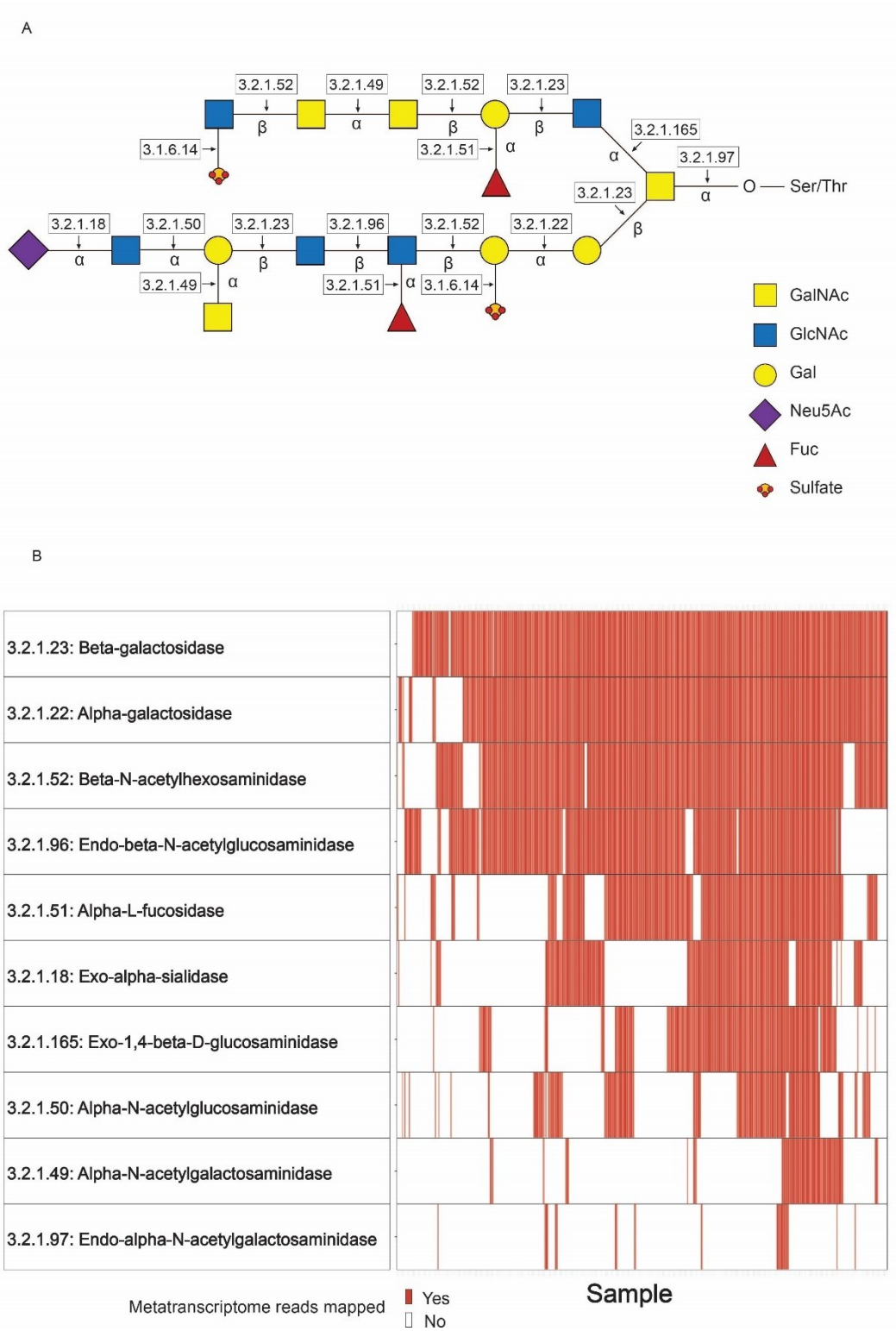

Figure 1 Enzymes involved in the degradation of mucosal glycans. (A) Hypothetical mucosal glycan structure and respective degradative enzymes (indicated with the corresponding enzyme commission numbers = ECs). (B) Distribution of ECs across analyzed metatranscriptomes. Samples with reads grouping to a respective EC are shown in red. Samples are ordered based on a k-means clustering algorithm.

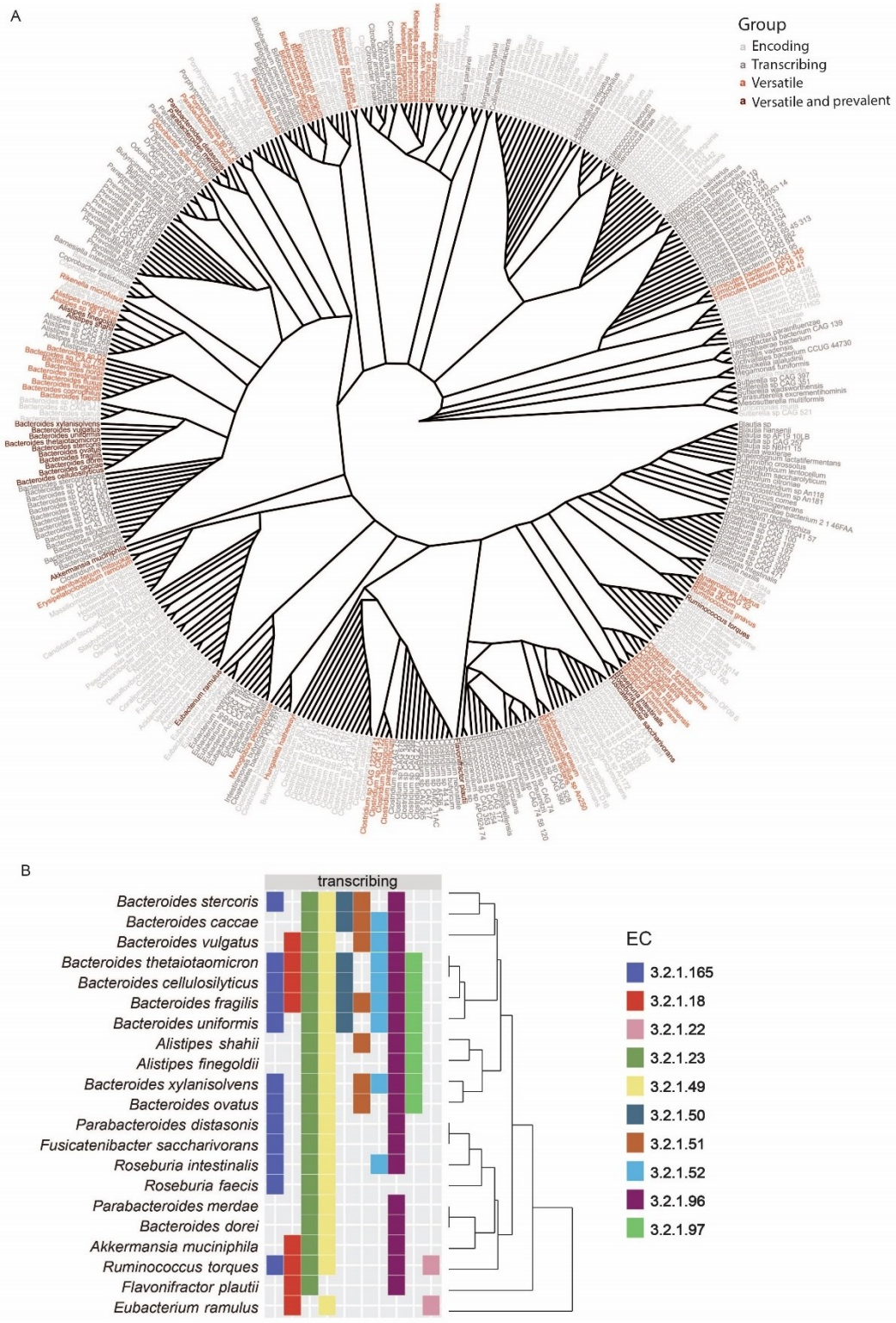

Figure 2 Species encoding and transcribing mucosal glycan degradation enzymes. (A) Radial cladogram of species encoding the analyzed ECs (light gray), transcribing any of the ECs (dark gray), transcribing more than 1 EC (light red), and prevalently transcribing (transcription prevalence >= 10%) more than one EC (dark red). (B) Species prevalently transcribing more than one EC. Species are ordered based on their EC transcription profile. The dendrogram indicates which species have the most similar EC profiles.

Mucosal glycan degraders form distinct clusters with co-occurring species

We next investigated with which species mucosal glycan degraders co-occur. We computed a co-occurrence network based on Spearman correlations and performed a clustering analysis to identify the species with which glycan degraders consistently co-occur. This revealed 4 distinct clusters (Fig 3A). Cluster 1 is the largest cluster and contains 4 mucosal glycan degraders, namely *Roseburia faecis, Fusicatenibacter saccharivorans, Ruminococcus torques* and *Eubacterium ramulus*, all of which belong to the order *Clostridiales*. In total, 44 out of 71 species in this cluster are *Clostridiales*. Cluster 2 contains 3 degraders, *Flavonifractor plautii, Bacteroides fragilis* and *Roseburia intestinalis*. Again, most species, 25 out of 44, belong to the order *Clostridiales*. Cluster 3 is the smallest cluster, but contains the largest number of degraders (10 species), all of which belong to the order *Bacteroidales*. The mucosal glycan degraders in this cluster are *Bacteroides uniformis, Parabacteroides distasonis, Bacteroides xylanisolvens, Bacteroides ovatus, Bacteroides caccae, Bacteroides vulgatus, Parabacteroides merdae, Bacteroides stercoris, Bacteroides thetaiotaomicron* and *Bacteroides dorei.* Out of 33 species in this cluster, 15 belong to the order *Bacteroidales* and 12 to the order *Clostridiales*. The last cluster, cluster 4, contains 4 mucin degraders, *Bacteroides cellulosilyticus, Alistipes finegoldii, Alistipes shahii* and *Akkermansia muciniphila. Clostridia* and *Bacteroidales* are again the dominant orders in this cluster of 61 species, with 20 and 16 species, respectively. We observed the highest number of between-cluster positive correlations between cluster 2 and cluster 4 while cluster 1 has the lowest number of positive correlations to any other cluster (Fig 3A). None of the clusters are mutually exclusive, and in fact they frequently all co-occur in the analyzed samples (Fig 3B). However, we observed a gradient from samples heavily dominated by a single cluster to samples with a very evenly distributed composition.

Transcription of individual genes is often dominated by a single species

In order to investigate what percentage of EC transcription is attributed to individual species, we normalized the transcription levels per species to total transcription attributed to the respective EC per sample. This analysis revealed an interesting common pattern in all investigated ECs: transcription is frequently entirely dominated by a single species, but the dominant species can vary between samples (Fig 4A and Figs S4A-S11A). We then evaluated if this pattern of mono-domination is simply reflecting the abundance of the respective species. If this were the case, species abundance should be a strong predictor for transcription levels. Linear models between species abundance and species transcription of individual ECs (in percentage of total transcription attributed to the respective EC) resulted in explained proportions of variance (adjusted R^2) ranging from -0.02 to 0.53, with an overall median of 0.16 (Table S3). In general, species abundance explains the highest proportions of transcriptional variance (median > 0.2) in *B. dorei*, *A. muciniphila*, *B. stercoris*, *B. uniformis*, *A. finegoldii, R. intestinalis, B. cellulosilyticus* and *B. fragilis* (Table S3). This may indicate that these species are particularly competitive mucosal glycan degraders that tend to transcribe these enzymes whenever they are abundant. However, while abundance is a good predictor

for the contribution to transcription in some species, it can only partially explain the observed transcription patterns.

The abundance patterns of other degraders partially explain transcription patterns

We next evaluated if the abundance of other species transcribing the same enzyme can further explain or predict the transcription of a given species. To evaluate this, we computed linear models of the form: *transcription species 1 [% of total EC transcription] ~ abundance species 2 [clr transformed abundance] + abundance species 3 [clr transformed abundance] + … + abundance species n [clr transformed abundance].* If the abundance of another transcribing species (species 2 - n = predictor species) can explain some of the variance in the transcription level of the species in question (species 1 = response species) this species will be a significant predictor in the linear model with the coefficient indicating the effect size. A positive coefficient indicates that the transcription of the response species will be high if the abundance of the predictor species is high. A negative coefficient indicates that the transcription of the response species is low if the abundance of the predictor species is high. In order to visualize all significant predictors in these linear models, we built response-predictor networks for each EC (Fig 4B and Figs S4B-S11B). Statistical details on these linear models can be found in Table S4. In general, we predominantly observed negative associations between species, where the abundance of one species is a negative predictor for the transcription of another species. Positive relationships between species occur particularly in alpha- and beta-galactosidase as well as alpha-N-acetylglucosaminidase, perhaps indicating a higher potential for facilitating interactions in more widely disseminated glycans. In order to summarize all response-predictor networks of individual ECs we built a consensus network (Fig 5A). This consensus network visualizes the number of times a significant relationship between two species was observed in the individual networks and whether they are consistently positive, negative or mixed. Interestingly, there was only one case of a mixed relationship across all 10 individual networks: the abundance of *R. intestinalis* is a negative predictor for *B. caccae's* transcription of alpha-L-fucosidase and a positive predictor for its transcription of alpha-galactosidase. This may indicate generally stable relationships between mucosal glycan degraders. In four individual networks there was a negative relationship from *B. uniformis* to *B. thetaiotaomicron,* from *B. thetaiotaomicron* to *B. vulgatus* and from *B. fragilis* to *B. uniformis*, suggesting strong competition between these degraders. Positive relationships between species are observed much less frequently, occurring in a maximum of two individual networks (Fig 5B). We observe reoccurring positive relationships from *B. xylanisolvens* to *B. ovatus,* from *B. thetaiotaomicron* to *B. xylanisolvens*, from *B. ovatus* to *B. stercoris* and from *B. caccae* to *A. shahii*. This may indicate an either direct or indirect enhancement of transcription by the presence of another species. In summary we see very stable relationships between mucosal glycan degraders when investigating the transcription of multiple enzymes.

Competitive mucosal glycan degraders are associated with reduced transcriptional activity of other degraders

To further elucidate how abundance and transcription patterns of mucosal glycan degraders are connected, we compared the predictive power of a species' own abundance on its transcription patterns with this species effect on other species' transcription (Fig 6A). This revealed that more competitive degraders (high percentage of abundance-driven transcription) tend to be associated with reduced transcriptional activity of a larger number of species. Conversely, the abundance of less competitive degraders is rarely a negative predictor for the transcription of other species. Additionally, we found that negative and positive associations between species are inversely correlated (Fig 6B), meaning that the less competitive a degrader is the more likely it is to be associated with increased transcription of other degraders. Finally, we evaluated whether competitive mucosal glycan degraders have distinctive genomic features associated with polysaccharide or glycan degradation. We found a consistent correlation between the number of mucosal glycan degradation enzymes and the number of other glycosyl hydrolase families that degraders encode and transcribe (Fig S12A and B, respectively). This indicates that all investigated species seem to have a similar proportion of mucosal glycan degradation enzymes in comparison to their repertoire of glycosyl hydrolases.

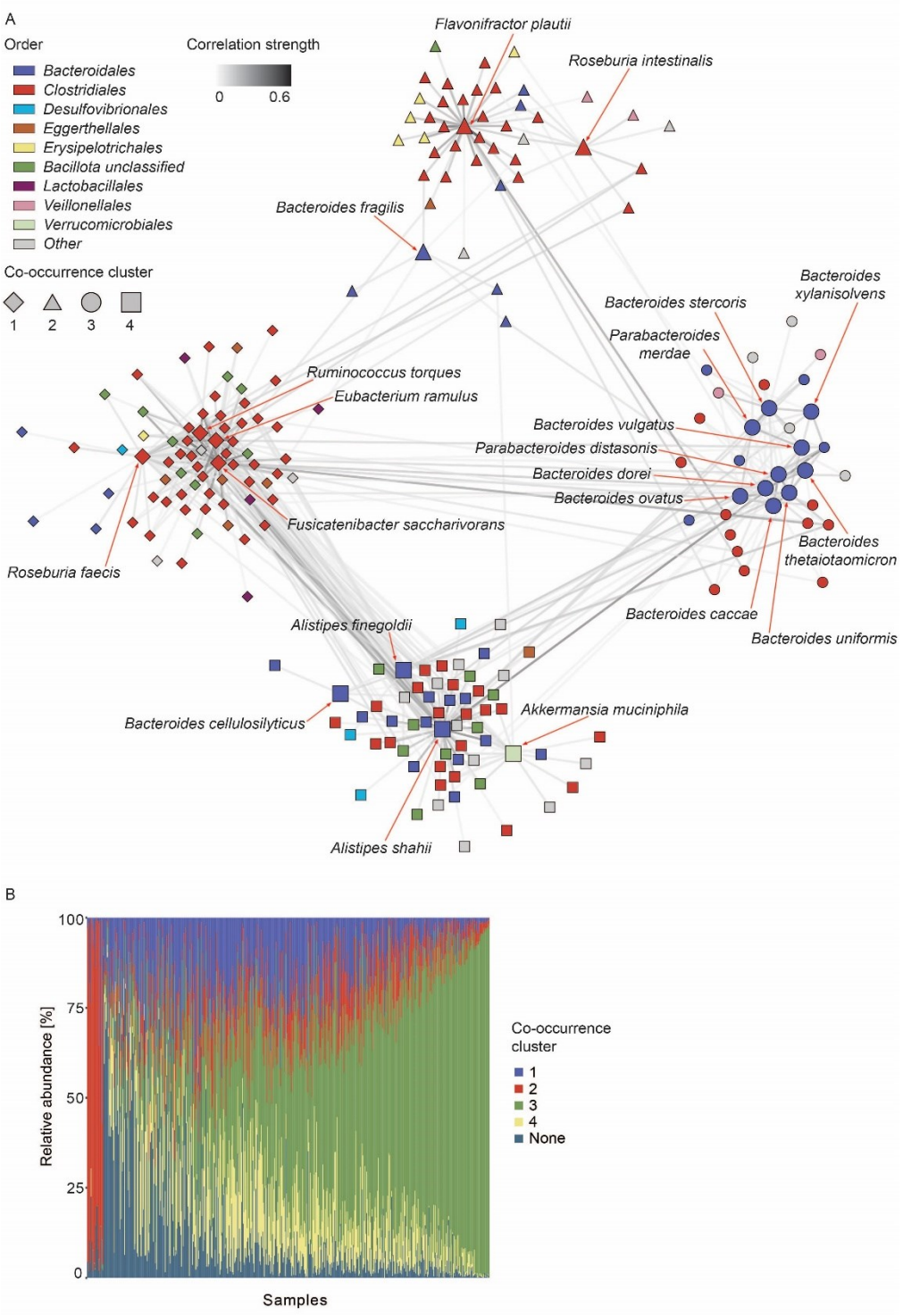

Figure 3 Co-occurrence clusters of mucosal glycan degraders and their first neighbors. (A) Co-occurrence network of mucosal glycan degraders and their first neighbors (species positively correlated with a mucosal glycan degrader). Network nodes depict species and network edges are significant positive correlations. Node shape indicates the co-occurrence cluster, node color indicates phylogenetic order of a species, and edge color indicates correlation strength. Nodes of mucosal glycan degraders are larger and labeled. (B) Relative abundance [%] of co-occurrence network clusters across samples. Colors indicate the respective cluster that species were grouped into. Samples were ordered based on a k-means clustering algorithm.

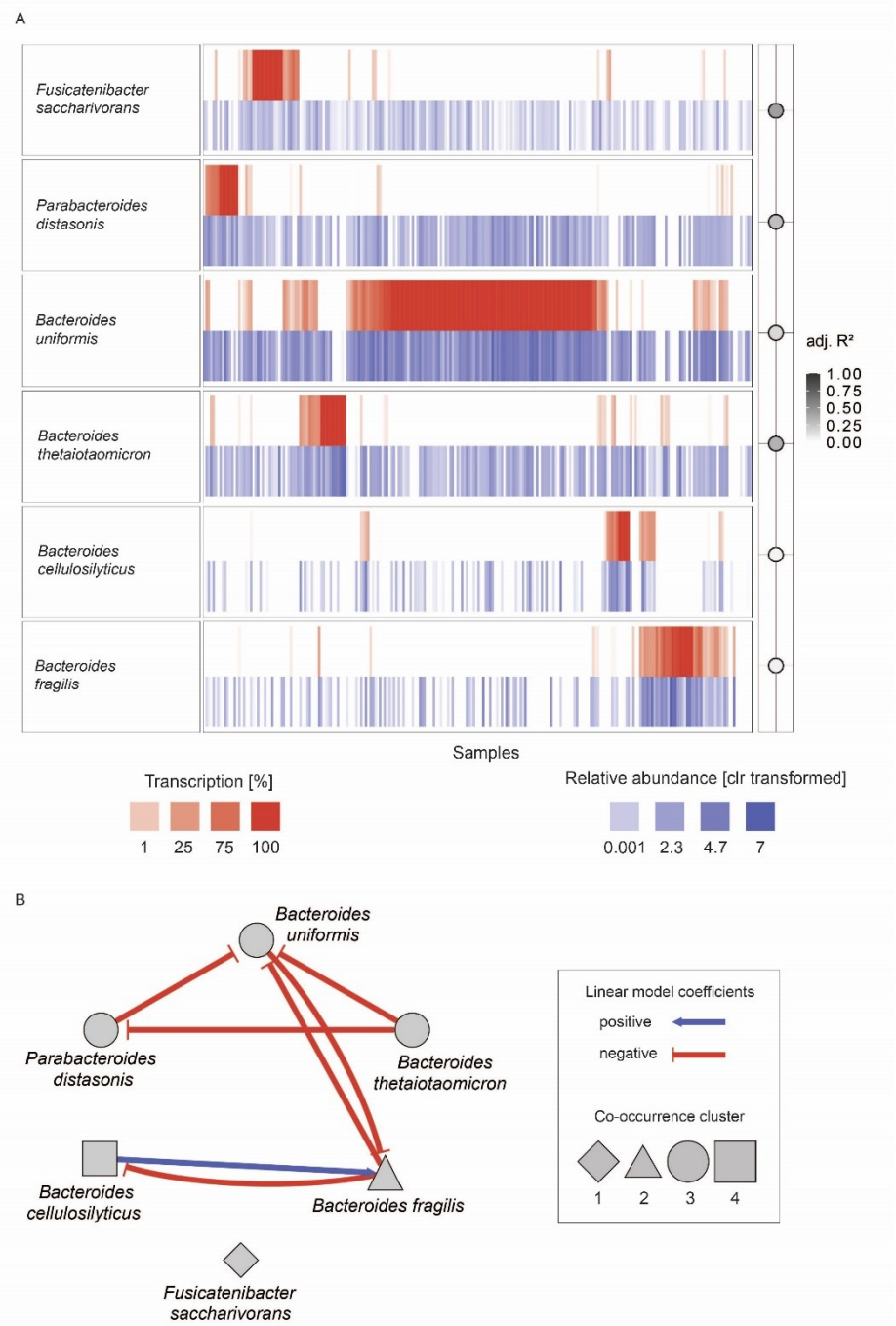

Figure 4 Contribution of species to transcription of EC 3.2.1.165: Exo-1,4-beta-D-glucosaminidase. (A) Transcription of EC 3.2.1.165 mapped to individual species (in % of total transcription grouped to EC 3.2.1.165) is shown in red. Relative abundance (clr transformed) of prevalently transcribing (>= 10%) species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (B) Response-predictor network of EC 3.2.1.165 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

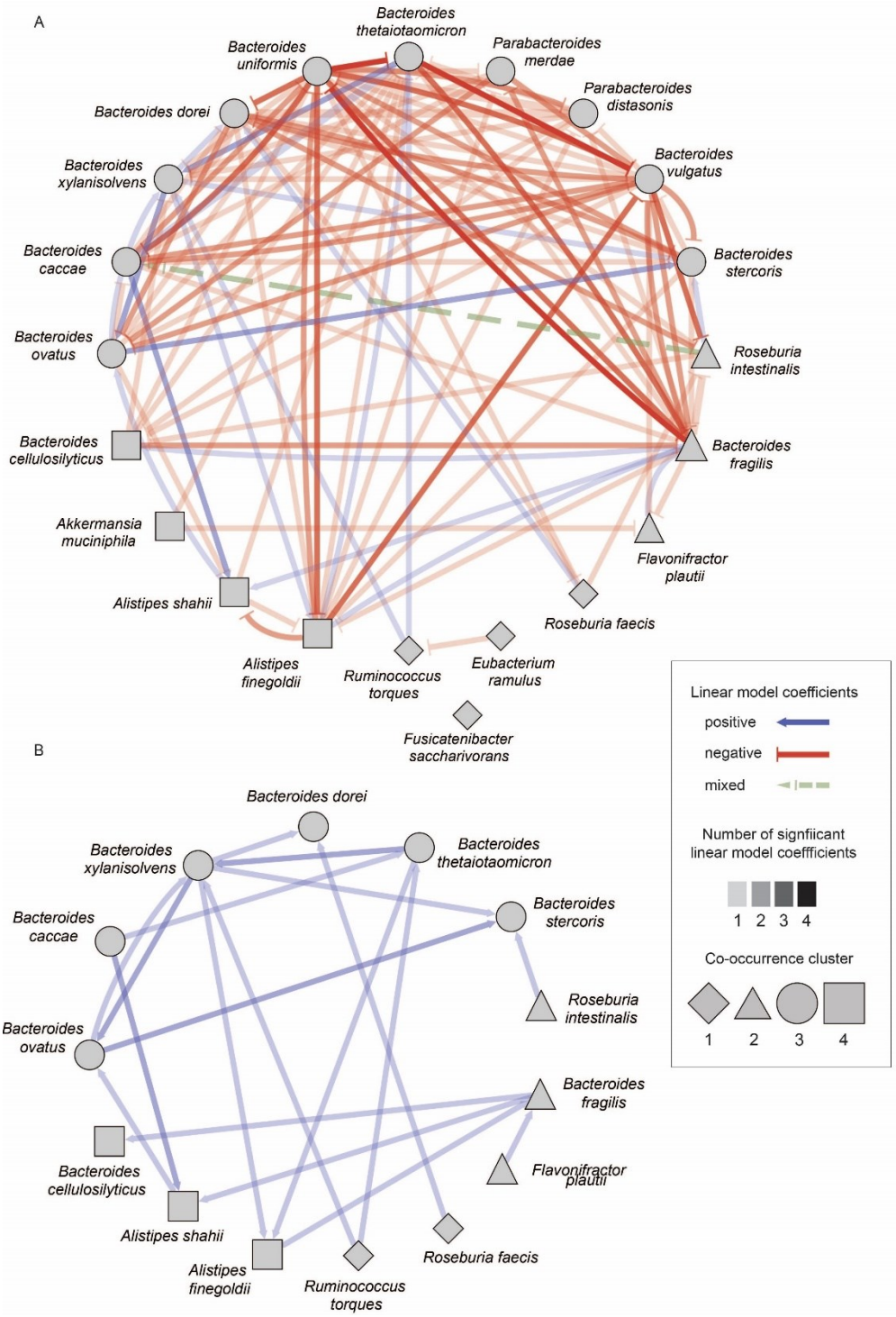

Figure 5 Consensus response-predictor networks of all analyzed ECs. Nodes are mucosal glycan degraders and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are directed from predictor to response and edge color and shape indicates whether the association between species across networks is consistent (blue - always positive, red - always negative) or mixed (green). Edge transparency depicts the number of significant linear model coefficients across response-predictor networks (light gray - low, dark gray - high). (A) Consensus network summarizing positive and negative associations between species. (B) Consensus network depicting only positive associations between species.

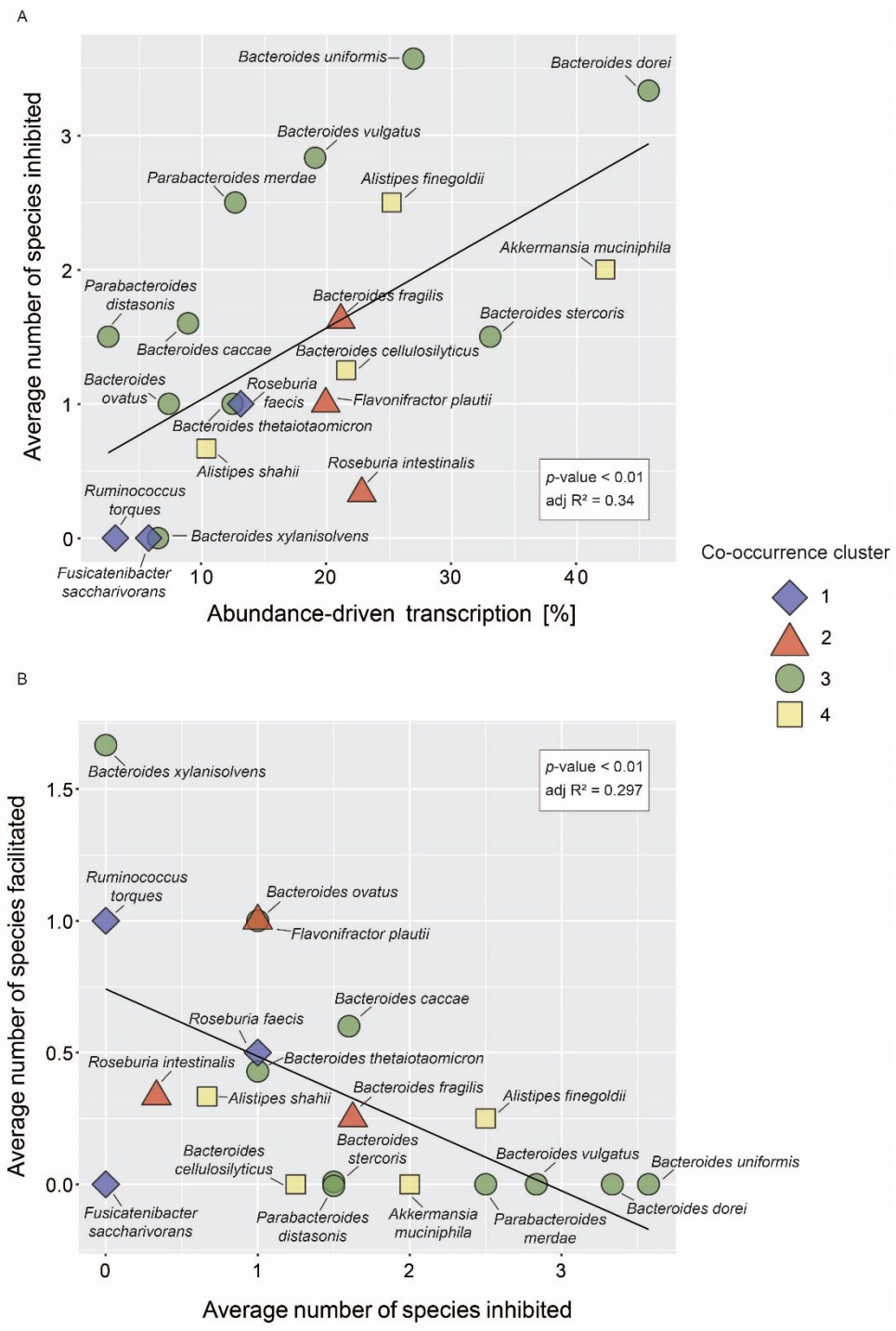

Figure 6 Abundance-driven transcription and transcription inhibition and facilitation in prevalent mucosal glycan degrading species. Shape and color of data points refers to the co-occurrence cluster the species belong to (Fig 3A) (A) Percentage of abundance-driven transcription (= median R^2 of linear models between species abundance and transcription) versus average number of species inhibited (= average number of significant negative linear model coefficients in response-predictor networks) (Linear model: p-value < 0.01, adjusted R^2 = 0.34). (B) Average number of species inhibited (= average number of significant negative linear model coefficients in response-predictor networks) versus average number of species facilitated (= average number of significant positive linear model coefficients in response-predictor networks) (Linear model: p-value < 0.01, adjusted R^2 = 0.297).

DISCUSSION

Many members of the human gut microbiome encode enzymes that are predicted to have mucosal glycan degradation activities (Ravcheev and Thiele 2017; Glover, Ticer and Engevik 2022). However, we observe that a much smaller portion is actively transcribing these genes in a given community and even fewer species transcribe them prevalently across the analyzed samples (Fig 2B). Frequently the transcription of a particular gene is attributed to a single species in individual samples, leading to patterns of mono-domination (Fig 4A, Figs S3A-S11A). In a proportion of prevalent mucosal glycan degraders this mono-domination seems linked to their abundance patterns. For other species their own abundance is not a good predictor of transcription, but their transcription patterns are frequently linked to the absence or low abundance of other species (Fig 5A). We speculate that this reflects a continuum between competitive and opportunistic degraders. Competitive degraders tend to show a high percentage of abundance-driven transcription (Fig 6A), meaning transcription of mucosal glycan degradation genes is high when a species is abundant. Opportunistic degraders, on the other hand, do not transcribe these genes whenever they are abundant, but seem to be only sporadically involved in mucosal glycan degradation. This decouples their abundance and transcription patterns, resulting in a low percentage of abundance-driven transcription (Fig 6A).

Additionally, the presence of competitive degraders tends to be associated with inhibited transcription of other degraders (Fig 6A) while opportunistic degraders tend to facilitate it (Fig 6B). While negative associations seem to be predominant in the mucosal glycan degrading community (Fig 5A), we do find indications for facilitation. Notably, this is the case for *R. torques*. The abundance of *R. torques* is a positive predictor for the transcription of beta-galactosidase in both *B. thetaiotaomicron* and *B. xylanisolvens* (Fig S6B). It has previously been shown that *R. torques* can liberate products from larger mucin structure, making them accessible for *B. thetaiotaomicron* (Schaus et al. 2024). Our results suggest that not only may other species, such as *B. xylanisolvens*, also benefit from this behavior, relationships like these may be more widespread. In particular, we find that the abundance patterns of *B. xylanisolvens, B. caccae, B. ovatus* and *B. fragilis* are positive predictors for transcription patterns of multiple other species, indicating that they may facilitate mucus degradation in other members of the community. Another well-known mucin degrader that shows a strong prevalence for mucosal glycans is *A. muciniphila*. It can express beta-galactosidase and beta-hexosaminidase (Kostopoulos et al. 2021) as well as sialidases and fucosidases (Shuoker et al. 2023). *A. muciniphila* has also been shown to upregulate defense related genes in the presence of other mucin degraders (Kostopoulos et al. 2021), suggesting that it is a competitive mucus degrader. Although we only found *A. muciniphila* to prevalently transcribe exo-alpha-sialidase in the analyzed metatranscriptomes we still observe a strong link between the species' abundance and its transcription of exo-alpha-sialidase (Table S3). Additionally, the abundance of *A. muciniphila* is a negative predictor for the transcription of this enzyme by both *B. thetaiotaomicron* and *F. plautii* (Fig S3B), perhaps reflecting *A. muciniphila's*

competitive advantage in mucus degradation. We observe similar patterns in *B. fragilis*, a species that exhibits a preference for O-glycans over starch and inulin (Pudlo et al. 2015), although *B. fragilis* seems to be less competitive than *A. muciniphila* as it shows a lower percentage of abundance-driven transcription and on average inhibits fewer species (Fig 6A). *B. thetaiotaomicron*, on the other hand, was previously described as a mucin degradation generalist (Kostopoulos et al. 2021) and is able to rapidly respond to multiple glycans (Rogers et al. 2013). Correspondingly, we find that *B. thetaiotaomicron* prevalently transcribes 7 out of the 10 investigated mucin degradation enzymes, underlining its generalist strategy. It has also been shown that *B. thetaiotaomicron* downregulates polysaccharide utilization loci associated with mucin degradation in the presence of other polysaccharides (Kostopoulos et al. 2021), suggesting an opportunistic rather than competitive role in mucin degradation. The abundance of *B. thetaiotaomicron* is, in contrast to *B. fragilis* and *A. muciniphila*, a poor predictor for its transcription patterns (Fig 6A, Table S3), again congruent with its more opportunist role. *B. thetaiotaomicron* may have to resort to active mucin degradation when its preferred nutrient sources are scarce and competition for mucosal glycans consequently higher.

The observation that the abundance of strong competitors for mucosal glycans, as determined by their high percentage of abundance-driven transcription, tends to frequently be a negative predictor for the transcription patterns of other species (Fig 6A) might indicate that these competitive degraders are more specialized for mucin glycan degradation and may additionally have an inhibiting effect on species degrading the same glycans. Conversely, negative associations might also indicate that opportunistic degraders take advantage of mucus products and monosaccharides liberated by the activity of competitive degraders and are therefore not transcribing any of their own mucus degradation enzymes. They may however be forced to actively degrade mucin in the absence, or low abundance, of more competitive degraders. We observe a particularly high number of such negative associations between mucin degraders of the genus *Bacteroides* (Fig 5A), suggesting either strong competition for mucin glycans, common cross-feeding or even cooperative degradation. The majority of these closely related species are in the same co-occurrence cluster, cluster 3 (Fig 3), meaning that they are frequently observed in the same samples. Tuncil and colleagues could show that *B. ovatus* and *B. thetaiotaomicron* exhibit different priorities for various dietary glycans, promoting stable coexistence despite their similar ecological niches (Tuncil et al. 2017). We find that the species in co-occurrence cluster 3 are spread out along the observed continuum between competitive and opportunistic mucosal glycan degraders (Fig 6A), pointing towards differing ecological strategies that allow for co-occurrence. The previous work on nutrient prioritization (Rogers et al. 2013; Schwalm III, Townsend II and Groisman 2017; Tuncil et al. 2017) as well as work on diauxic growth of microbes (Wang et al. 2021) indicates that at least some members of the human gut microbiome strongly favor particular glycans and that this prioritization is preserved in communities as well. Interestingly, the species investigated in these studies show characteristics of mucin degradation opportunists in our analysis. We observe recurring associations of the same

nature between species across transcription patterns of several enzymes. These recurring associations are mostly negative and particularly driven by species showing characteristics of competitive mucosal glycan degraders. In a complex community the effect of nutrient prioritization in opportunistic degraders may be less apparent due to the presence of more competitive community members. However, it has to be noted that in a complex gut microbial community the glycosyl hydrolases investigated are not exclusively used to degrade mucosal glycans. In particular GlcNAc and Gal are widely disseminated sugars that can also be found in dietary glycans (Ravcheev and Thiele 2017). Consequently, there is an overlap between the mucus and dietary polysaccharide degradation niches and communities and associations between mucus degraders could therefore also be a product of a shift between these niches. We identified a higher number of species to prevalently transcribe the GlcNAc- and Gal-cleaving enzymes as opposed to other enzymes, aligning with the likely higher prevalence of these sugars in the gut. We also primarily observe the positive associations in the response-predictor networks of these enzymes. Positive associations could therefore be the result of a shift to dietary polysaccharides when other mucin degraders are more abundant. Interestingly, we find a strong linear relationship between mucosal glycan degradation genes and other glycosyl hydrolase degradation genes across all analyzed species, both in terms of encoded as well as transcribed genes (Fig S12). Some competitive degraders, such as *B. vulgatus* or *B. dorei* encode and transcribe a large number of glycosyl hydrolase degradation genes, suggesting a generalist rather than specialist strategy. Other competitive degraders, such as *A. muciniphila* and *A. finegoldii*, seem to have a much smaller number of glycosyl hydrolase degradation genes and a comparatively large number of mucosal glycan degradation genes, suggesting a specialization towards mucosal glycan degradation.

In this work we identify, in line with previous results, various members of the genus *Bacteroides* as well as other species such as *A. muciniphila* and *R. torques* as prevalent mucosal glycan degraders (Ravcheev and Thiele 2017; Berkhout, Plugge and Belzer 2022). In particular *B. fragilis, B. thetaiotaomicron* and *B. cellulosilyticus* contribute to transcription of many mucin degradation enzymes (Fig 2C). Other known mucin degraders, such as *Ruminococcus gnavus* and *Bifidobacterium spp.* (Crost et al. 2016; Bell et al. 2019; Glover, Ticer and Engevik 2022), were found to transcribe various mucin degradation genes, but at very low prevalence. Furthermore, we did not observe transcription of fucosidase in *A. muciniphila*, despite experimental evidence to the contrary (Shuoker et al. 2023). These inconsistencies may be attributed to strain variability within species or low annotation quality of certain genes and are a clear limitation of mapping metatranscriptome reads to reference genomes. We would also like to emphasize that the analyzed publicly available datasets are from two cohorts with limited geographical diversity and therefore do not accurately represent the global diversity found in human gut microbiomes. The analyzed data does also not include samples from infants, where *Bifidobacterium spp.* would likely play a much larger role in the mucus degrading community (Yamaguchi and Yamamoto 2023). Samples from the distal colon also do not reflect the entirety of the GI tract and neither can they provide insight into spatial variability, which certainly has an impact on mucus degradation (Duncan, Carey-

Ewend and Vaishnava 2021; Mondragón-Palomino et al. 2022). Lastly, the analyses presented here rely on bioinformatic analysis and experimental validation is ultimately needed to elucidate predicted inter-species relationships and ecological mechanisms.

CONCLUSION

In this study we systematically interrogated metagenome and metatranscriptome datasets for the presence of reads associated with mucosal glycan degradation enzymes. A large group of species encodes such genes, but only few species with limited diversity, mainly *Bacteroidales* and *Clostridiales*, prevalently transcribe them. We could show that transcription of the individual enzymes is frequently dominated by a single species in a given sample. In a subgroup of degraders transcription is mostly driven by abundance suggesting that they are competitive mucosal glycan degraders while the transcription of opportunistic degraders is decoupled from their abundance. Competitive degraders are additionally frequent negative predictors for transcription of other degraders while opportunistic degraders show indications for facilitating transcription of other species. These observations further our understanding of the ecological strategies of mucosal glycan degraders and suggest a continuum between competitive and opportunistic degraders that allows co-occurrence despite a seemingly strongly overlapping ecological niche.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

DB and FB designed the study. FB performed analyses. DB and FB interpreted the results and wrote the manuscript.

ACKNOWLEDGEMENTS

Computational results of this work have been achieved using the Life Science Compute Cluster (LiSC) of the University of Vienna.

FUNDING

This work was supported by ERC starting grant 714623 “FunKeyGut” and the Austrian Science Fund (FWF) (10.55776/COE7). Work in the laboratory of DB was supported by the FWF project MAINTAIN DOC 69 (10.55776/DOC69).

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data is deposited in the Sequence Read Archive under BioProject ID PRJNA354235 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA354235/) and BioProject ID PRJNA389280 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA389280/).

REFERENCES

Arike L, Hansson GC. The Densely O-Glycosylated MUC2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *J Mol Biol* 2016;**428**:3221–9.

Bell A, Brunt J, Crost E *et al.* Elucidation of a sialic acid metabolism pathway in mucus-foraging Ruminococcus gnavus unravels mechanisms of bacterial adaptation to the gut. *Nat Microbiol* 2019;**4**:2393–404.

Berkhout MD, Plugge CM, Belzer C. How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. *Glycobiology* 2022;**32**:182–200.

Brockhausen I, Schachter H, Stanley P. O-GalNAc Glycans. *Essentials of Glycobiology. 2nd Edition*. Cold Spring Harbor Laboratory Press, 2009.

Brugiroux S, Beutler M, Pfann C *et al.* Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat Microbiol* 2016;**2**:1–12.

Coleman OI, Haller D. Microbe–Mucus Interface in the Pathogenesis of Colorectal Cancer. *Cancers* 2021;**13**:616.

Crost EH, Tailford LE, Monestier M *et al.* The mucin-degradation strategy of Ruminococcus gnavus: The importance of intramolecular trans-sialidases. *Gut Microbes* 2016;**7**:302–12.

Csárdi G, Nepusz T, Müller K *et al.* igraph for R: R interface of the igraph library for graph theory and network analysis. 2023, DOI: 10.5281/zenodo.8240644.

Dalgaard P. R Development Core Team (2010): R: A language and environment for statistical computing. 2010.

Drula E, Garron M-L, Dogan S *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 2022;**50**:D571–7.

Duncan K, Carey-Ewend K, Vaishnava S. Spatial analysis of gut microbiome reveals a distinct ecological niche associated with the mucus layer. *Gut Microbes* 2021;**13**:1874815.

Fang J, Wang H, Zhou Y *et al.* Slimy partners: the mucus barrier and gut microbiome in ulcerative colitis. *Exp Mol Med* 2021;**53**:772–87.

Gendler SJ, Spicer AP. Epithelial Mucin Genes. *Annu Rev Physiol* 1995;**57**:607–34.

Glover JS, Ticer TD, Engevik MA. Characterizing the mucin-degrading capacity of the human gut microbiota. *Sci Rep* 2022;**12**:8456.

Hamaker BR, Tuncil YE. A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota. *J Mol Biol* 2014;**426**:3838–50.

Jegatheesan T, Moorthy AS, Eberl HJ. A mathematical model of competition between fiber and mucin degraders in the gut provides a possible explanation for mucus thinning. *J Theor Biol* 2024;**587**:111824.

Johansson MEV, Phillipson M, Petersson J *et al.* The inner of the two Muc2 mucin-dependent mucus

layers in colon is devoid of bacteria. *Proc Natl Acad Sci* 2008;**105**:15064–9.

Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;**28**:27–30.

Kostopoulos I, Aalvink S, Kovatcheva-Datchary P *et al.* A Continuous Battle for Host-Derived Glycans Between a Mucus Specialist and a Glycan Generalist in vitro and in vivo. *Front Microbiol* 2021;**12**.

Kucera M, Malmgren BA. Logratio transformation of compositional data: a resolution of the constant sum constraint. *Mar Micropaleontol* 1998;**34**:117–20.

Luis AS, Jin C, Pereira GV *et al.* A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* 2021;**598**:332–7.

Mehta RS, Abu-Ali GS, Drew DA *et al.* Stability of the human faecal microbiome in a cohort of adult men. *Nat Microbiol* 2018;**3**:347–55.

Mondragón-Palomino O, Poceviciute R, Lignell A *et al.* Three-dimensional imaging for the quantification of spatial patterns in microbiota of the intestinal mucosa. *Proc Natl Acad Sci* 2022;**119**:e2118483119.

Paone P, Cani PD. Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut* 2020;**69**:2232–43.

Paysan-Lafosse T, Blum M, Chuguransky S *et al.* InterPro in 2022. *Nucleic Acids Res* 2023;**51**:D418–27.

Podolsky DK. Oligosaccharide structures of human colonic mucin. *J Biol Chem* 1985;**260**:8262–71.

Pudlo NA, Urs K, Kumar SS *et al.* Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *mBio* 2015;**6**:10.1128/mbio.01282-15.

Raimondi S, Musmeci E, Candeliere F *et al.* Identification of mucin degraders of the human gut microbiota. *Sci Rep* 2021;**11**:11094.

Ravcheev DA, Thiele I. Comparative Genomic Analysis of the Human Gut Microbiome Reveals a Broad Distribution of Metabolic Pathways for the Degradation of Host-Synthetized Mucin Glycans and Utilization of Mucin-Derived Monosaccharides. *Front Genet* 2017;**8**, DOI: 10.3389/fgene.2017.00111.

Robbe C, Capon C, Coddeville B *et al.* Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 2004;**384**:307–16.

Rogers TE, Pudlo NA, Koropatkin NM *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* 2013;**88**:876–90.

Schaus SR, Vasconcelos Pereira G, Luis AS *et al.* Ruminococcus torques is a keystone degrader of intestinal mucin glycoprotein, releasing oligosaccharides used by Bacteroides thetaiotaomicron. *mBio* 2024;**15**:e00039-24.

Schirmer M, Franzosa EA, Lloyd-Price J *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* 2018;**3**:337–46.

Schwalm III ND, Townsend II GE, Groisman EA. Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. *Mol Microbiol* 2017;**104**:32–45.

Shannon P, Markiel A, Ozier O *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 2003;**13**:2498–504.

Shuoker B, Pichler MJ, Jin C *et al.* Sialidases and fucosidases of Akkermansia muciniphila are crucial for growth on mucin and nutrient sharing with mucus-associated gut bacteria. *Nat Commun* 2023;**14**:1833.

Tailford LE, Owen CD, Walshaw J *et al.* Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* 2015;**6**:7624.

Tuncil YE, Xiao Y, Porter NT *et al.* Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. *mBio* 2017;**8**:10.1128/mbio.01068-17.

Wang Z, Goyal A, Dubinkina V *et al.* Complementary resource preferences spontaneously emerge in diauxic microbial communities. *Nat Commun* 2021;**12**:6661.

Xu S, Li L, Luo X *et al.* Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* 2022;**1**:e56.

Yamaguchi M, Yamamoto K. Mucin glycans and their degradation by gut microbiota. *Glycoconj J* 2023;**40**:493–512.

Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res* 2020;**30**:492–50

DISCUSSION

In this thesis I investigate the human gut microbiome at two different scales. In the first chapter I explore the concept of correlation network keystone taxa, which are thought to have a disproportionately strong impact on the whole community. Here I find that correlation networks are highly sensitive to taxonomic resolution and that putative keystone taxa are only observable at high resolution. These keystone taxa seem to act as stabilizing agents in individual clusters of co-occurring taxa and exhibit distinct transcriptional states. In the second chapter I analyze the structure of the comparably small sub-community of mucosal glycan degraders. Here I show that only a small number of gut microbial species that encode genes for mucosal glycan degradation actively transcribe them and that the transcription of individual genes is frequently dominated by a single species in a given sample. I further find apparent differences in the proportion of abundance-driven transcription which suggests the presence of competitive and opportunistic degraders. Competitive degraders are also more likely to be negative predictors for the transcription of other degraders. I come to these conclusions by examining the respective gut microbial communities from three different angles: I utilize network analysis to understand the role of individual taxa in the community, I interrogate transcriptional patterns to explore the interplay of functional redundancy of the community and functional versatility of individual taxa, and I analyze co-occurrence patterns to elucidate community composition and stratification.

The human gut microbiome is a highly diverse community and early theoretical work suggested that such systems should be unstable in nature (May, 1972). However, the gut microbiome of individuals is quite stable over time (David et al., 2014) and robust ecological theory has been presented to ground these observations (McCann, 2000). In particular it seems that negative interactions between species, e.g. competition, stabilize a highly diverse community and are needed to prevent positive feedback loops that would lead to overgrowth (Coyte et al., 2015). The putative keystone taxa I identified may provide these stabilizing properties as they tend to correlate negatively with other taxa and on a genus level seem to dampen positive correlations within other genera. The association patterns in the mucosal glycan degrading community are also dominated by negative associations where the abundance of one degrader is a negative predictor for the transcription of other degraders. These negative associations seem to be mainly driven by competitive degraders which may therefore serve as stabilizing agents akin to the keystone taxa. Host derived compounds such as mucosal glycans are ecologically particularly interesting as they provide a constant source of nutrition to the community members able to degrade these glycans. Foster and colleagues argue that the host utilizes these resources to exert control over the microbial community and postulate the gut microbiome as an “ecosystem on a leash” balancing between microbial competition and host control (Foster et al., 2017). Secreted mucins may also be necessary to assemble the gut microbiome in the first place. Coyte et al. investigated potential assembly paths that could lead to observed microbiome communities. They argue that, while the gut

microbiome is a stable community resilient to fluctuation, large perturbations that remove multiple species, such as antibiotic interventions, can lead to a collapse and formation of an alternative stable state. Initial assembly of such a community is therefore unfavorable but can be aided by the host through the provision of resources (Coyte et al., 2021). Mucosal glycans produced by the gut epithelium may serve both the role of initially “recruiting” gut microbes as well as allowing the host to exert control over the microbiome. The degradation of these complex mucosal structures is often a community effort as most degraders do not possess the necessary enzymes to cleave off all the different glycans (Raimondi et al., 2021). This seems to be supported by my observation that degraders with similar sets of mucosal glycan degradation genes frequently co-occur. The transcriptional profiles suggest that they split up the degradation process as different species dominate transcription of specific genes in subsets of samples despite the other degraders being present. In fact, microbes with similar nutritional profiles frequently co-occur and are often closely related (Levy and Borenstein, 2013), as is the case with *Bacteroides* species in my mucosal glycan degradation study. Counterintuitively, I mainly observe negative associations between degraders despite mucus degradation seemingly being a perfect example for collaborative degradation. As already mentioned, the ecological theory presented above as well as my observations in the keystone study could indicate that the few highly competitive degraders, which show the highest number of negative associations, are an ecological necessity and stabilize the community. In support of the theoretical work, Herren and McMahon find that the presence of highly connected taxa as well as the strength and placement of negative interactions is linked to community stability (Herren and McMahon, 2018). Goldford and colleagues further argue that stability in large microbial communities may in fact be a generic emergent property rather than a result of system-specific dynamics (Goldford et al., 2018). They also find cumulative rather than pairwise interactions in cross-feeding networks, supporting the importance of higher-order interactions in complex microbial communities (Levine et al., 2017). The correlation-based approaches I utilize in my research likely capture a mixture of pairwise interactions, higher-order interactions, environmental drivers as well as stochastic processes. It may therefore be that the dynamics I can recover are a combination of what is described in different theoretical work. Additionally, the gut microbiome is a heterogeneous community with complex interactions ranging from competitive exclusion (Wiles et al., 2016) to cross-feeding on complex polysaccharides (Rakoff-Nahoum et al., 2016) and such interactions are highly context dependent (Weiss et al., 2023). The influence of putative keystones on the whole community may be governed by different ecological principles than the role of competitive degraders in the mucosal glycan degradation community, despite recurring negative associations being a common observable feature.

When I investigate the transcriptional patterns of mucosal glycan degraders an apparent difference is the proportion of transcriptional variation that can be explained by a given degrader’s abundance patterns. This may reflect species' different capability or necessity to switch their realized nutrient niche. In a complex nutrient environment such as the gut microbiome, with high temporal variability as well as variety in carbon sources constantly

shaping nutrient availability, successful species are likely quite versatile and able to switch their nutritional niche frequently (Pereira and Berry, 2017). This versatility also allows for the high functional redundancy observable in the gut microbiome that seems counterintuitive at first glance. In fact, Louca and colleagues argue that functional redundancy is inherent to open microbial systems and that the decoupling between taxonomy and function cannot solely be attributed to stochastic processes (Louca et al., 2018). I see indications that mucosal glycan degraders navigate this functional redundancy of the community in different ways. There seem to be highly successful degraders, as indicated by their correlated abundance and transcription patterns, as well as more opportunistic species with decoupled abundance and transcription. It would seem intuitive that competitive mucosal glycan degraders have a smaller suite of glycoside hydrolases at their disposal and therefore needed to evolve strategies to stay competitive in this nutritional niche while opportunists could easily switch niches due to their large assortment of glycoside hydrolases. However, I see no distinction between competitive and opportunistic degraders in the number of glycoside hydrolases encoded in their genome. This may again indicate that different ecological strategies result in similar observable patterns in correlation-based analyses. Competitive degraders with a large number of other glycoside hydrolases may for example be able to utilize multiple substrates simultaneously (Okano et al., 2021) while competitive degraders with fewer glycoside hydrolases may rather invest in defense mechanisms against competitors (Kostopoulos et al., 2021). Opportunists with a large array of glycoside hydrolases may switch their nutritional niche (Tuncil et al., 2017) while those with few glycoside hydrolases at their disposal may benefit from mucosal glycans freed by other degraders (Schaus et al., 2024). Similarly, the transcriptional and co-occurrence patterns I observe in the putative keystone taxa could be the result of different ecological mechanisms. In general, I identified distinct transcriptional states in these keystone taxa which again points towards functional versatility and potential niche switching. Combined with the observation that keystone taxa have a higher number of negative correlations and seem to dampen correlations between species of the same genus this seems to point towards keystone taxa actively keeping other species “in check” and thereby stabilizing the community, akin to some competitive mucosal glycan degraders. Considering that the observed negative correlations are not directional they could also indicate that keystone taxa switch to ecological niches that open up due to the absence of other species. This could mean that the keystones take a more passive role and constitute a buffering system that provides resilience and stability to a fluctuating community. Here they may share similarities with the opportunistic mucosal glycan degraders moving in and out of that niche as needed and/or possible. However, due to the keystone definition I used in my analysis, where the compositional change of the community upon removal of a taxon determines their keystone potential (Berry and Widder, 2014), a more active role of the putative keystone taxa identified with this method seems probable. Correspondingly, some of the putative keystone taxa I identify likely have rather distinct roles in the gut microbiome. *Methanobrevibacter smithii* is the most prevalent methanogen in the human gut microbiome (Dridi et al., 2009) and *Bilophila wadsworthia* is able to metabolize taurine to hydrogen sulfide

(Peck et al., 2019). Currently, neither of these metabolic pathways seem to be widespread in the gut microbiome and a transcriptional change in these taxa may primarily reflect a change in the host environment as opposed to a navigation of community dynamics. Even though these two aspects are inextricably interlinked, the ability to occupy unique ecological niches may be the main driver behind these taxa's high keystone potential. I hypothesize that using different analysis and/or modeling approaches, applying them to different datasets and also the specific definition of keystones used is likely going to lead to the identification of not only different taxa but in fact to the identification of ecologically distinct kinds of keystones. If everyone postulates different taxa as keystones dependent on methodology and dataset used, we can of course ask the question if genuine keystone taxa that have a disproportionately strong effect on the whole microbial community actually exist.

In chapter 1 of my thesis I hypothesize that there are in fact no keystone taxa that globally affect the entire microbial community. Instead it seems that the identified keystones have a stabilizing effect on sub-communities of frequently co-occurring taxa. Together these sub-communities constitute the gut microbial community although the relative fraction each of them occupies in a given sample varies strongly. Across the analyzed samples a compositional gradient emerges that ranges from samples strongly dominated by individual clusters to highly mixed samples. The observed co-occurring clusters on genus level are remarkably similar to the enterosignatures described by Frioux and colleagues who also report an apparent gradient of enterosignature composition across communities (Frioux et al., 2023). They furthermore describe these enterosignatures as metabolically complimentary. The different possible enterosignature compositions however suggest that the overall functionality of a given gut microbial community can be achieved with various combinations and could reflect differences in the host environment or functional versatility of enterosignatures. Many dependencies between taxa in the gut microbiome are likely not very specific and could therefore be fulfilled by different members (Pereira and Berry, 2017). Similarly, cooperation may not be obligatory but rather only realized under the right conditions with the cooperating taxa being able to switch to other ecological niches when advantageous (Pereira and Berry, 2017). Combining the observed sub-communities or enterosignatures with this individual flexibility and versatility creates the image of a gut microbial community that is built up from combinable "modules". It seems that a large number of combinations can still provide the necessary functionality. I also observe clusters of co-occurring species when looking at mucosal glycan degraders and their first neighbors in a correlation network. However, these clusters are fairly different from the clusters based on the whole community network. In part this can certainly be attributed to methodological differences between my keystone and mucosal glycan degradation study, as the focus in the former is on inferring interactions while the focus in the latter is simply observing co-occurrences. Additionally, the distinct nutritional and ecological niche mucosal glycan degraders occupy may further explain the different clusters. Hildebrand and colleagues identify three major dispersal strategies of gut microbes that underlie their persistence in the gut: tenacious (persistent with strong family and geographic associations), spatiopersistant

(persistent with strong geographic associations) and heredipersistent (persistent with strong family associations) (Hildebrand et al., 2021). Many mucosal glycan degraders, in particular *Bacteroides* species, seem to fall into the tenacious category which suggest that they are highly adapted to the host, e.g. through their ability to use host-derived compounds such as mucosal glycans. Contrastingly, the keystone taxa I identified were not found to be tenacious in this study, but some are characterized as heredipersistent. Heredipersistent microbes may not be as well adapted to the host, but they are also less likely to be lost from the community as they seem to be able to re-colonize. The co-occurring clusters of the mucosal glycan degrading species and their directly correlated neighbors may therefore differ from whole community clusters as they are derived from a particular group of species with distinct ecological properties, at least in terms of dispersal strategy. Co-occurrence clusters could likely be identified in other functional sub-communities of the gut microbiome as well, for example short-chain fatty acid producers or degraders of plant-derived polysaccharides, and these clusters would probably again differ as they sample a different sub-community with different ecological properties. The enterosignatures found on whole community level may constitute a sort of emergent property derived from these smaller sub-community clusters. However, beyond the ecological strategies of microbes the environment also plays a major role in shaping the gut microbiome. The pool of heredipersistent microbes that can re-colonize a particular person for example is strongly influenced by their immediate environment. Valles-Colomer and colleagues could show that cohabitation leads to substantial strain sharing between individuals and has a stronger effect than age or genetics (Valles-Colomer et al., 2023). We also have substantial evidence that diet probably has the strongest influence on microbial composition on a day-to-day basis (David et al., 2014; Ross et al., 2024). Despite these strong external determinants host genetics certainly still plays a role, for example the ability to metabolize lactose (Lopera-Maya et al., 2022), although most associations are still unresolved (Sanna et al., 2022). Larzul and colleagues studied community stratification in pigs and found that host genetics can select for a certain enterotype as a whole, at least under the controlled conditions of a breeding program (Larzul et al., 2024), and view enterotypes as a functional ecosystem. The enterotype model suggests that certain community compositions constitute particularly favorable local optima in the compositional space. These may be comparable to the communities dominated by a single enterosignature, while the mixed communities may represent the transitional states between enterotypes. However, if enterotypes are selected by the host as a whole the gut microbiome is less modular than I argue above. This is also supported in a study by Tap and colleagues. They performed a large-scale analysis to stratify the human gut microbiome and identified local partitions within large global gradients they refer to as microbiome branches (Tap et al., 2023). Additionally, they analyzed time series data and observed partition switches mostly between neighboring partitions suggesting favorable ecological paths in this compositional space and thereby a constraint on how freely sub-communities such as enterosignatures may be combinable.

CONCLUDING REMARKS

In this thesis I analyzed publicly available metagenomes and metatranscriptomes from the healthy adult human gut microbiome and the work I present is deeply rooted in a desire to exploit the vast information contained in this type of data. The aim of this work was to further elucidate community structure and dynamics as well as ecological principles governing this complex microbial community. Comparing the observations on keystone taxa as well as mucosal glycan degraders revealed common patterns and allowed for a broader discussion on possible ecological strategies. However, the methods I deployed are largely based on correlation and can ultimately not resolve the ecological questions I pose in the discussion as various underlying dynamics can lead to the patterns I observe. This type of coarse-grained data analysis can nevertheless highlight commonalities such as the prevalent negative correlations of keystone taxa and competitive mucosal glycan degraders or the observed compositional gradients. Uncovering and investigating these patterns can help us to explore the ecological principles behind them and aid directional experimental work.

REFERENCES

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2014.00219 (Accessed April 4, 2023).

Coyte, K. Z., Rao, C., Rakoff-Nahoum, S., and Foster, K. R. (2021). Ecological rules for the assembly of microbiome communities. *PLOS Biol.* 19, e3001116. doi: 10.1371/journal.pbio.3001116

Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science* 350, 663–666. doi: 10.1126/science.aad2602

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820

Dridi, B., Henry, M., Khéchine, A. E., Raoult, D., and Drancourt, M. (2009). High Prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae Detected in the Human Gut Using an Improved DNA Detection Protocol. *PLOS ONE* 4, e7063. doi: 10.1371/journal.pone.0007063

Foster, K. R., Schluter, J., Coyte, K. Z., and Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. *Nature* 548, 43–51. doi: 10.1038/nature23292

Frioux, C., Ansorge, R., Özkurt, E., Ghassemi Nedjad, C., Fritscher, J., Quince, C., et al. (2023). Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* 31, 1111-1125.e6. doi: 10.1016/j.chom.2023.05.024

Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., et al. (2018). Emergent simplicity in microbial community assembly. *Science* 361, 469–474. doi: 10.1126/science.aat1168

Herren, C. M., and McMahon, K. D. (2018). Keystone taxa predict compositional change in microbial communities. *Environ. Microbiol.* 20, 2207–2217. doi: 10.1111/1462-2920.14257

Hildebrand, F., Gossmann, T. I., Frioux, C., Özkurt, E., Myers, P. N., Ferretti, P., et al. (2021). Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host Microbe* 29, 1167-1176.e9. doi: 10.1016/j.chom.2021.05.008

Kostopoulos, I., Aalvink, S., Kovatcheva-Datchary, P., Nijsse, B., Bäckhed, F., Knol, J., et al. (2021). A Continuous Battle for Host-Derived Glycans Between a Mucus Specialist and a Glycan Generalist in vitro and in vivo. *Front. Microbiol.* 12. Available at: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.632454 (Accessed February 24, 2024).

Larzul, C., Estellé, J., Borey, M., Blanc, F., Lemonnier, G., Billon, Y., et al. (2024). Driving gut microbiota enterotypes through host genetics. *Microbiome* 12, 116. doi: 10.1186/s40168-024-01827-8

Levine, J. M., Bascompte, J., Adler, P. B., and Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* 546, 56–64. doi: 10.1038/nature22898

Levy, R., and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci.* 110, 12804–12809.

doi: 10.1073/pnas.1300926110

Lopera-Maya, E. A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., et al. (2022). Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *Nat. Genet.* 54, 143–151. doi: 10.1038/s41588-021-00992-y

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., et al. (2018). Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2, 936–943. doi: 10.1038/s41559-018-0519-1

May, R. M. (1972). Will a Large Complex System be Stable? *Nature* 238, 413–414. doi: 10.1038/238413a0

McCann, K. S. (2000). The diversity–stability debate. *Nature* 405, 228–233. doi: 10.1038/35012234

Okano, H., Hermsen, R., and Hwa, T. (2021). Hierarchical and simultaneous utilization of carbon substrates: mechanistic insights, physiological roles, and ecological consequences. *Curr. Opin. Microbiol.* 63, 172–178. doi: 10.1016/j.mib.2021.07.008

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

Pereira, F. C., and Berry, D. (2017). Microbial nutrient niches in the gut. *Environ. Microbiol.* 19, 1366–1378. doi: 10.1111/1462-2920.13659

Raimondi, S., Musmeci, E., Candeliere, F., Amaretti, A., and Rossi, M. (2021). Identification of mucin degraders of the human gut microbiota. *Sci. Rep.* 11, 11094. doi: 10.1038/s41598-021-90553-4

Rakoff-Nahoum, S., Foster, K. R., and Comstock, L. E. (2016). The evolution of cooperation within the gut microbiota. *Nature* 533, 255–259. doi: 10.1038/nature17626

Ross, F. C., Patangia, D., Grimaud, G., Lavelle, A., Dempsey, E. M., Ross, R. P., et al. (2024). The interplay between diet and the gut microbiome: implications for health and disease. *Nat. Rev. Microbiol.* 22, 671–686. doi: 10.1038/s41579-024-01068-4

Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J., and Zhernakova, A. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nat. Genet.* 54, 100–106. doi: 10.1038/s41588-021-00983-z

Schaus, S. R., Vasconcelos Pereira, G., Luis, A. S., Madlambayan, E., Terrapon, N., Ostrowski, M. P., et al. (2024). Ruminococcus torques is a keystone degrader of intestinal mucin glycoprotein, releasing oligosaccharides used by Bacteroides thetaiotaomicron. *mBio* 15, e00039-24. doi: 10.1128/mbio.00039-24

Tap, J., Lejzerowicz, F., Cotillard, A., Pichaud, M., McDonald, D., Song, S. J., et al. (2023). Global branches and local states of the human gut microbiome define associations with environmental and intrinsic factors. *Nat. Commun.* 14, 3310. doi: 10.1038/s41467-023-38558-7

Tuncil, Y. E., Xiao, Y., Porter, N. T., Reuhs, B. L., Martens, E. C., and Hamaker, B. R. (2017). Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. *mBio* 8, 10.1128/mbio.01068-17. doi: 10.1128/mbio.01068-17

Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., et al. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* 614, 125–135. doi: 10.1038/s41586-022-05620-1

Weiss, A. S., Niedermeier, L. S., von Strempel, A., Burrichter, A. G., Ring, D., Meng, C., et al. (2023). Nutritional and host environments determine community ecology and keystone species in a synthetic gut bacterial community. *Nat. Commun.* 14, 4780. doi: 10.1038/s41467-023-40372-0

Wiles, T. J., Jemielita, M., Baker, R. P., Schlomann, B. H., Logan, S. L., Ganz, J., et al. (2016). Host Gut Motility Promotes Competitive Exclusion within a Model Intestinal Microbiota. *PLOS Biol.* 14, e1002517. doi: 10.1371/journal.pbio.1002517

SUPPLEMENTARY MATERIAL

CHAPTER 1

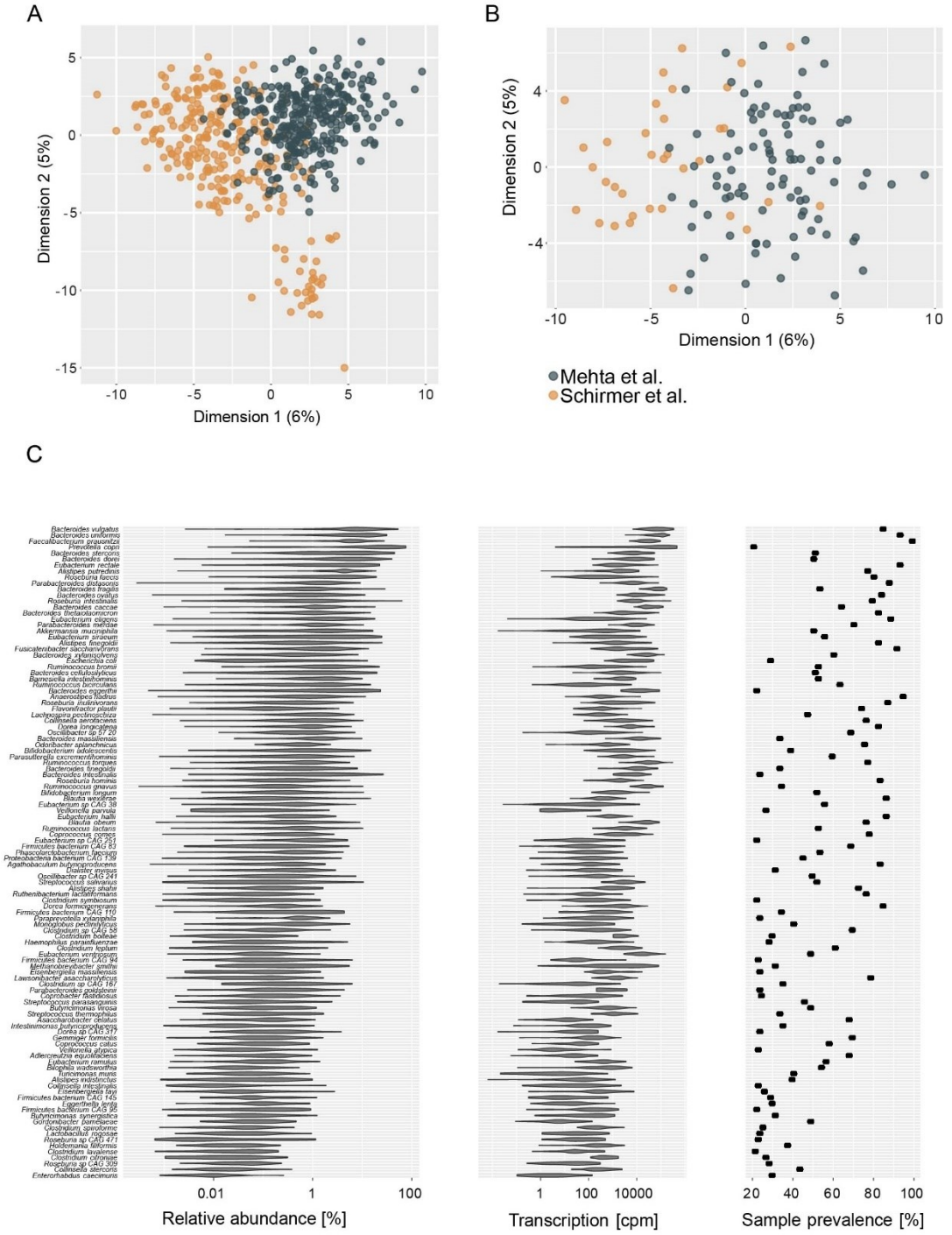

Figure S1 Principal coordinate analysis of (A) all processed samples and (B) samples used for further analysis (one randomly selected sample per participant). Depicted is the robust Aitchison distance of estimated total reads mapped to species reference genomes. Colors indicate the study that published the raw data. Panel C shows relative abundance [%], transcription [cpm - copies per million] and sample prevalence [%] of species present in at least 20% of analyzed samples. Species are ordered by mean relative abundance.

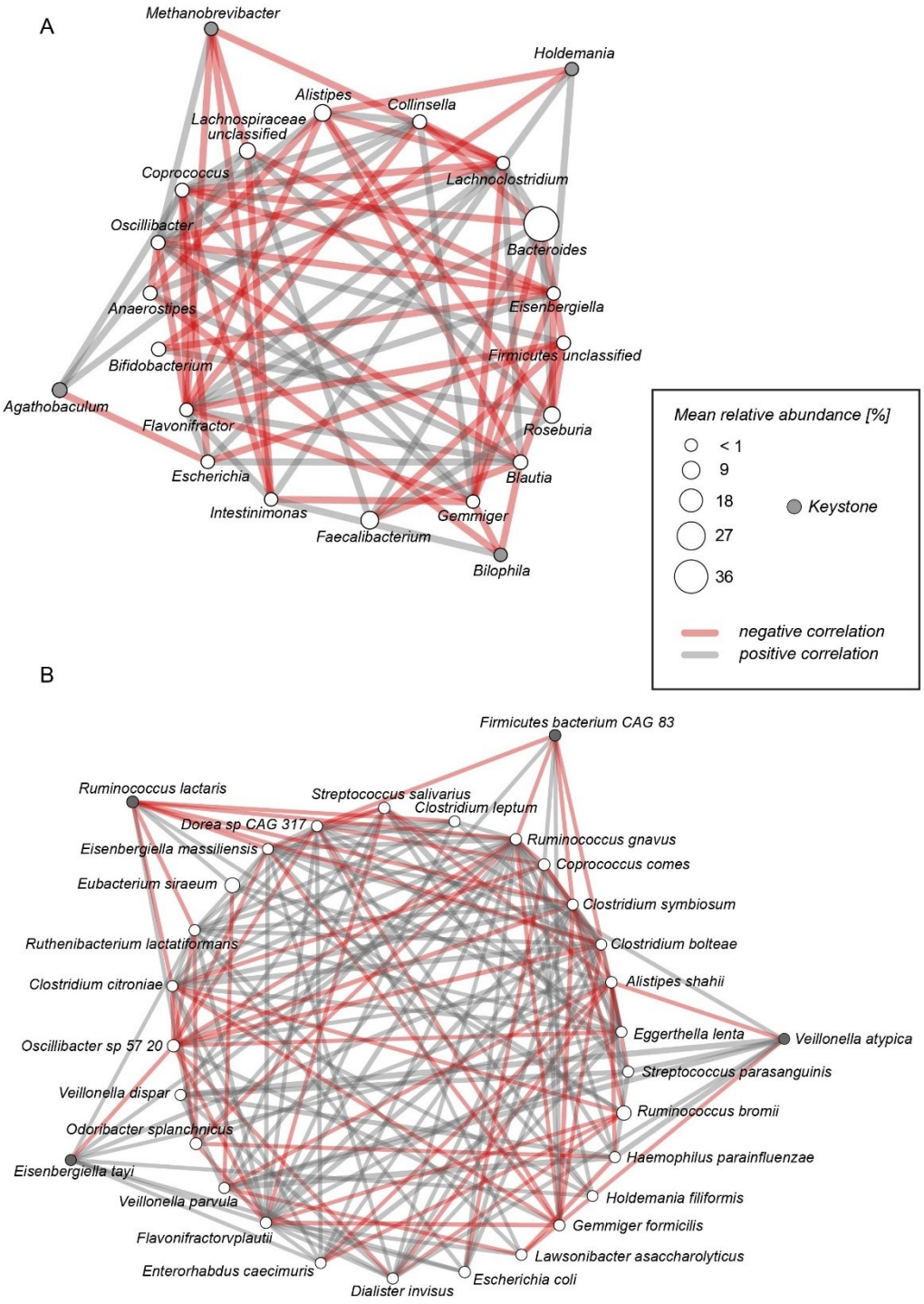

Figure S2 Correlation networks showing keystone taxa and their first neighbors (taxa directly correlated to a keystone taxon) on (A) genus and (B) species level. Network nodes in grey (white) indicate keystone taxa (non-keystone taxa), network edges connecting the circles indicate positive (grey) and negative (red) correlations between taxa. Node size indicates mean relative abundance of a given taxon across analyzed samples. Exact correlation coefficients can be found in Table S2.

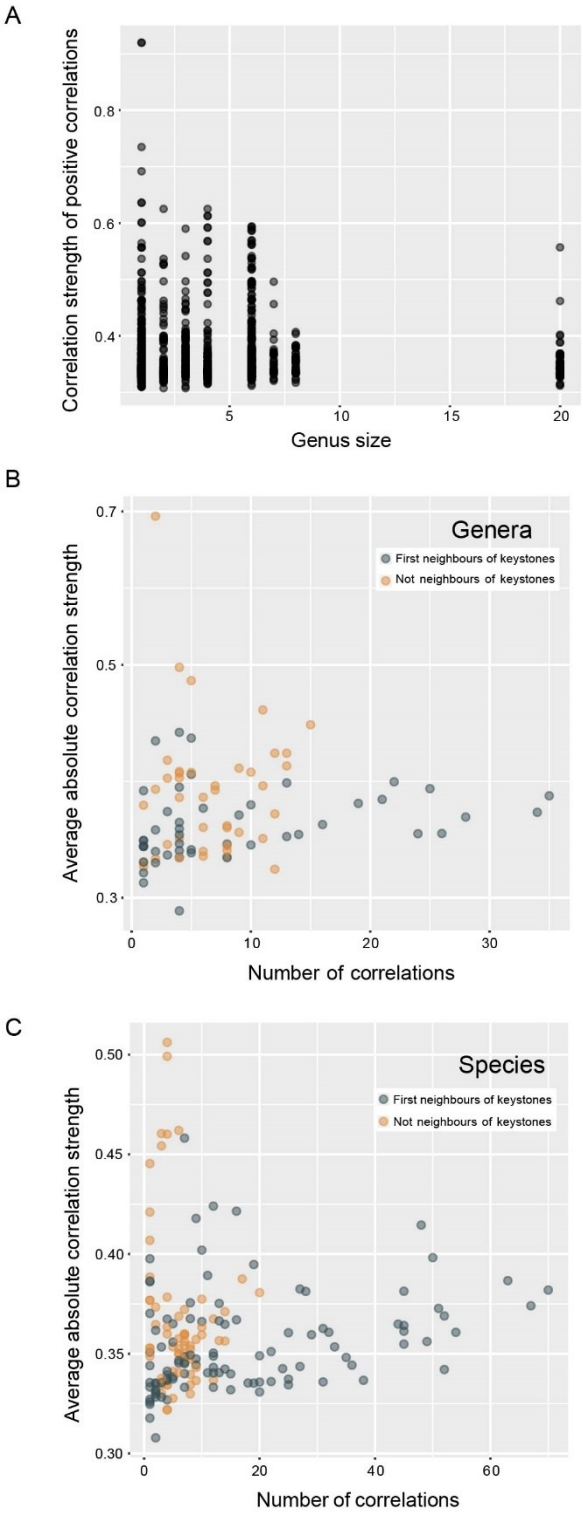

Figure S3 Distributions of network correlations in (A, C) the species-level network and (B) the genus-level network. Panel A shows the within-genus correlation strength of positive correlations as a factor of genus size. Panels B and C depict the average absolute correlation strength compared to the total number of correlations in (B) the genus-level network and (C) the species-level network. Gray circles indicate taxa that are first neighbors of (= directly correlated with) a keystone taxon, orange circles indicate taxa that are not first neighbors of any keystone taxa.

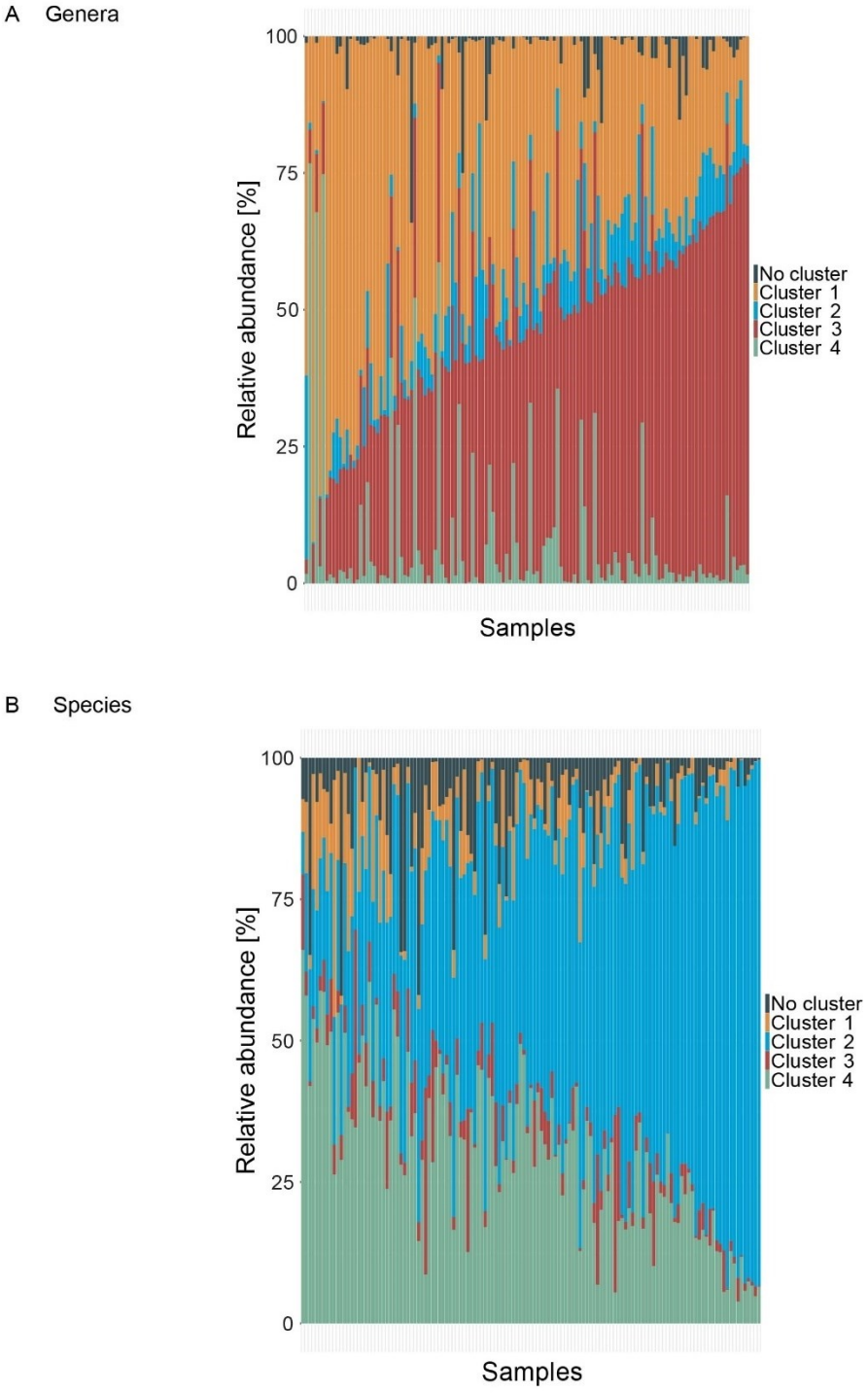

Figure S4 Relative abundance [%] of co-occurrence network clusters of (A) genera and (B) species. Colors indicate the respective cluster that taxa were grouped into (Fig 5), gray indicates taxa not grouped into any cluster. Samples were ordered based on a k-means clustering algorithm. (A) Genus network cluster 2 includes the keystone genera *Bilophila* and *Holdemania* and cluster 3 includes the keystone genera *Agathobaculum* and *Methanobrevibacter*. (B) Species network cluster 1 includes the keystone species *Veillonella atypica*, cluster 2 includes *Ruminococcus lactaris*, cluster 3 includes *Eisenbergiella tayi* and cluster 4 includes *Firmicutes bacterium CAG 83* (unclassified *Bacillota*).

Figure S5 Presence of metabolic pathways in identified correlation network keystone taxa. Gray color indicates that a pathway was found to be present in the metagenome reads mapped to the respective taxon.

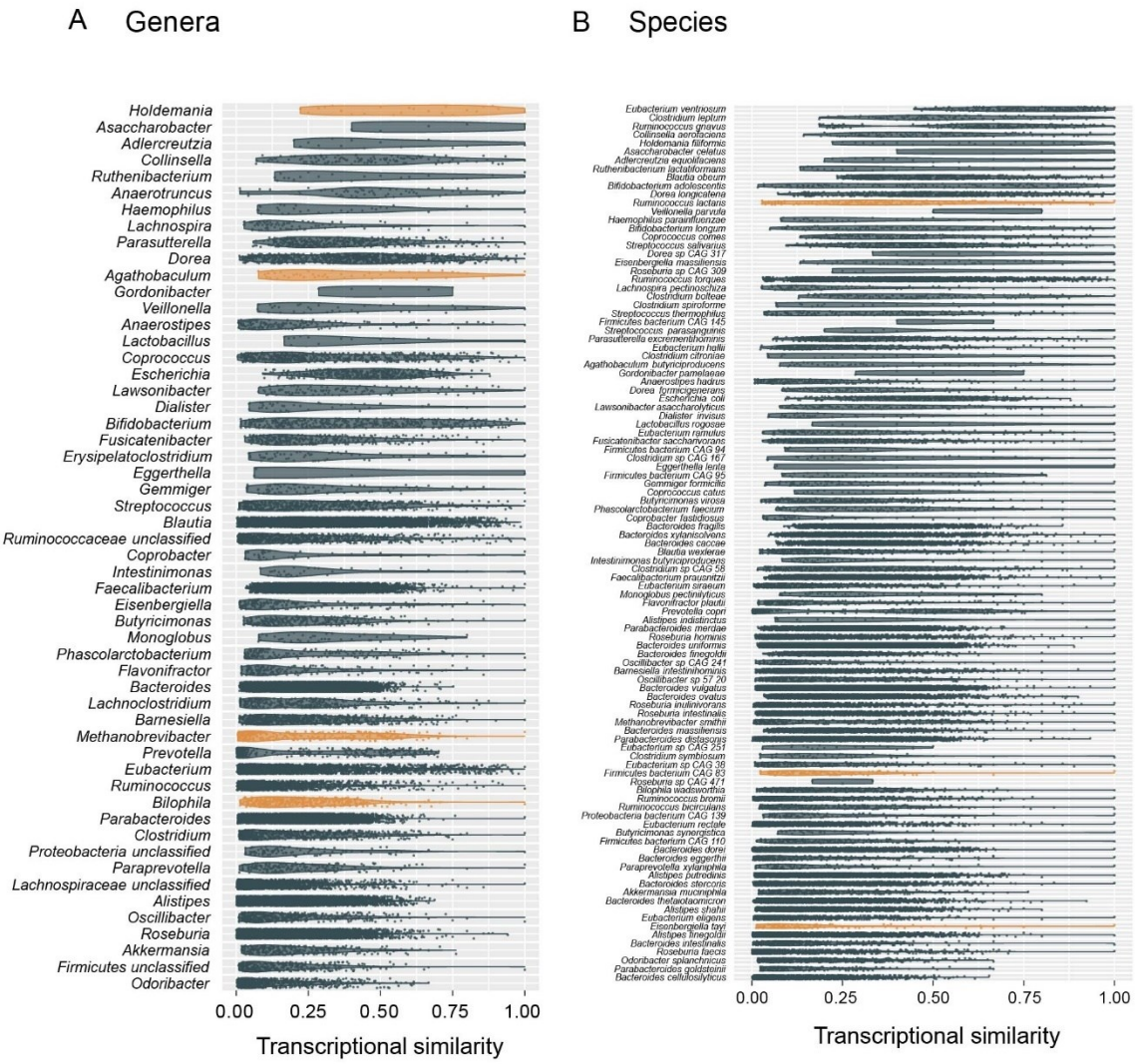

Figure S6 Transcriptional similarity of (A) genera and (B) species, with 1 indicating identical transcriptional profiles and 0 indicating no overlap. Transcriptional similarity was calculated as pairwise Sorenson distance from the presence and absence of gene families in the metatranscriptomes. Orange color marks the identified keystone taxa. Taxa are ordered by mean Sorenson distance.

Table S1 Network features of all analyzed taxa

Species	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
[Collinsella] massiliensis	0.003	23	0.057	0.190	0.005	2
Adlercreutzia equolifaciens	0.107	68	0.213	0.265	0.024	2
Agathobaculum butyriciproducens	0.311	83	0.074	0.222	0.003	6
Akkermansia muciniphila	1.065	50	0.049	0.133	0.001	5
Alistipes finegoldii	1.322	82	0.057	0.190	0.002	7
Alistipes indistinctus	0.057	40	0.041	0.100	0.001	3
Alistipes putredinis	3.764	77	0.148	0.405	0.008	7
Alistipes shahii	0.243	73	0.205	0.217	0.035	1
Anaerostipes hadrus	1.016	95	0.189	0.281	0.019	3
Asaccharobacter celatus	0.124	68	0.221	0.231	0.037	1
Bacteroides caccae	1.429	64	0.090	0.273	0.003	8
Bacteroides cellulosilyticus	1.399	51	0.049	0.000	0.005	0
Bacteroides dorei	3.386	50	0.066	0.214	0.006	2
Bacteroides eggerthii	1.001	22	0.025	0.000	0.001	0
Bacteroides finegoldii	0.376	34	0.066	0.500	0.002	21
Bacteroides fragilis	0.804	53	0.090	0.182	0.007	2
Bacteroides intestinalis	0.683	24	0.033	0.500	0.001	17
Bacteroides massiliensis	0.552	34	0.090	0.491	0.002	21
Bacteroides ovatus	1.179	84	0.066	0.286	0.001	13
Bacteroides stercoris	4.114	51	0.148	0.235	0.016	2
Bacteroides thetaiotaomicron	1.358	82	0.057	0.143	0.019	0
Bacteroides uniformis	8.100	93	0.082	0.267	0.004	5
Bacteroides vulgatus	8.890	85	0.139	0.257	0.012	3
Bacteroides xylanisolvens	0.546	60	0.098	0.379	0.004	9
Barnesiella intestinihominis	1.169	53	0.098	0.348	0.005	7
Bifidobacterium adolescentis	0.731	39	0.115	0.275	0.011	3
Bifidobacterium longum	0.348	52	0.115	0.198	0.012	2
Bilophila wadsworthia	0.035	54	0.016	0.000	0.000	0
Blautia obeum	0.546	76	0.148	0.444	0.010	6
Blautia wexlerae	0.552	86	0.090	0.309	0.006	5
Butyricimonas synergistica	0.033	31	0.082	0.333	0.004	8
Butyricimonas virosa	0.110	49	0.090	0.309	0.011	3
Clostridium bolteae	0.022	30	0.328	0.360	0.029	4
Clostridium citroniae	0.011	27	0.262	0.427	0.016	7
Clostridium lavalense	0.010	21	0.238	0.515	0.010	12

Species	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Clostridium leptum	0.111	61	0.115	0.527	0.003	21
Clostridium sp CAG 167	0.245	35	0.082	0.356	0.003	11
Clostridium sp CAG 58	0.145	69	0.156	0.298	0.010	4
Clostridium spiroforme	0.018	25	0.090	0.382	0.003	10
Clostridium symbiosum	0.042	22	0.254	0.437	0.015	8
Collinsella aerofaciens	0.948	76	0.156	0.427	0.019	3
Collinsella intestinalis	0.027	23	0.049	0.267	0.003	5
Collinsella stercoris	0.015	44	0.074	0.194	0.006	2
Coprobacter fastidiosus	0.087	24	0.008	0.000	0.000	0
Coprococcus catus	0.101	58	0.262	0.421	0.029	4
Coprococcus comes	0.394	78	0.238	0.367	0.025	4
Dialister invisus	0.153	31	0.205	0.297	0.045	1
Dorea formicigenerans	0.240	85	0.164	0.358	0.010	6
Dorea longicatena	0.757	82	0.221	0.330	0.028	3
Dorea sp CAG 317	0.092	24	0.270	0.373	0.026	4
Eggerthella lenta	0.050	30	0.246	0.395	0.019	5
Eisenbergiella massiliensis	0.047	24	0.238	0.433	0.023	4
Eisenbergiella tayi	0.051	26	0.074	0.639	0.000	126
Enterorhabdus caecimuris	0.005	30	0.180	0.372	0.011	6
Escherichia coli	0.287	29	0.123	0.371	0.005	9
Eubacterium eligens	2.714	89	0.057	0.476	0.001	20
Eubacterium hallii	0.469	86	0.180	0.307	0.013	4
Eubacterium ramulus	0.100	56	0.107	0.385	0.004	10
Eubacterium rectale	4.340	93	0.172	0.276	0.021	2
Eubacterium siraeum	1.960	56	0.082	0.222	0.005	4
Eubacterium sp CAG 251	0.277	22	0.000	0.000	0.000	0
Eubacterium sp CAG 38	0.320	56	0.057	0.238	0.003	5
Eubacterium ventriosum	0.080	49	0.008	0.000	0.000	0
Faecalibacterium prausnitzii	7.150	99	0.172	0.219	0.017	2
*Firmicutes bacterium CAG 110 **	0.297	34	0.180	0.398	0.028	3
*Firmicutes bacterium CAG 145 **	0.046	29	0.164	0.347	0.011	5
*Firmicutes bacterium CAG 83 **	0.335	69	0.057	0.619	0.000	169
*Firmicutes bacterium CAG 94 **	0.106	23	0.066	0.214	0.004	4
*Firmicutes bacterium CAG 95 **	0.041	22	0.066	0.393	0.002	14
Flavonifractor plautii	0.226	74	0.352	0.311	0.044	2
Fusicatenibacter saccharivorans	1.657	92	0.230	0.312	0.034	2
Gemmiger formicilis	0.120	69	0.238	0.456	0.014	8

Species	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Gordonibacter pamelaeae	0.049	49	0.148	0.281	0.011	4
Haemophilus parainfluenzae	0.109	28	0.148	0.327	0.009	5
Holdemania filiformis	0.015	37	0.082	0.400	0.002	15
Intestinimonas butyriciproducens	0.040	35	0.172	0.424	0.015	5
Lachnospira pectinoschiza	0.634	47	0.090	0.218	0.030	1
Lactobacillus rogosae	0.020	24	0.033	0.500	0.001	13
Lawsonibacter asaccharolyticus	0.129	79	0.148	0.242	0.020	2
Methanobrevibacter smithii	0.226	31	0.172	0.229	0.026	2
Monoglobus pectinilyticus	0.259	40	0.148	0.294	0.034	1
Odoribacter splanchnicus	0.450	76	0.148	0.235	0.017	2
Oscillibacter sp 57 20	0.688	69	0.246	0.425	0.016	6
Oscillibacter sp CAG 241	0.315	50	0.090	0.473	0.003	16
Parabacteroides distasonis	1.695	88	0.139	0.228	0.018	2
Parabacteroides goldsteinii	0.113	24	0.016	0.000	0.000	0
Parabacteroides merdae	1.211	70	0.107	0.231	0.029	1
Paraprevotella xylaniphila	0.144	24	0.090	0.273	0.009	3
Parasutterella excrementihominis	0.404	60	0.041	0.500	0.001	17
Phascolarctobacterium faecium	0.346	53	0.074	0.111	0.006	1
Prevotella copri	5.028	21	0.057	0.429	0.003	8
Proteobacteria bacterium CAG 139	0.170	45	0.057	0.381	0.003	7
Roseburia faecis	2.684	80	0.123	0.343	0.016	3
Roseburia hominis	0.398	83	0.107	0.256	0.005	5
Roseburia intestinalis	1.484	79	0.107	0.218	0.007	3
Roseburia inulinivorans	0.722	87	0.082	0.267	0.003	8
Roseburia sp CAG 309	0.015	28	0.074	0.306	0.003	7
Roseburia sp CAG 471	0.032	23	0.016	1.000	0.000	0
Ruminococcus bicirculans	1.099	63	0.057	0.333	0.001	19
Ruminococcus bromii	1.734	53	0.098	0.394	0.004	10
Ruminococcus gnavus	0.209	34	0.377	0.277	0.058	2
Ruminococcus lactaris	0.492	53	0.082	0.578	0.002	27
Ruminococcus torques	0.624	77	0.230	0.429	0.023	4
Ruthenibacterium lactatiformans	0.101	76	0.279	0.346	0.032	3

Species	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Streptococcus parasanguinis	0.108	46	0.098	0.348	0.005	7
Streptococcus salivarius	0.303	52	0.164	0.232	0.033	1
Streptococcus thermophilus	0.093	34	0.074	0.222	0.003	5
Turicimonas muris	0.041	40	0.041	0.400	0.003	6
Veillonella atypica	0.037	23	0.066	0.679	0.001	52
Veillonella parvula	0.046	27	0.148	0.294	0.016	3

* unclassified *Bacillota*

Genus	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Adlercreutzia	0.107	68	0.220	0.397	0.015	6
Agathobaculum	0.311	83	0.068	0.500	0.001	39
Akkermansia	1.065	50	0.034	1.000	0.000	0
Alistipes	5.513	93	0.356	0.343	0.053	2
Anaerostipes	1.020	95	0.169	0.444	0.009	9
Anaerotruncus	0.168	27	0.186	0.400	0.033	2
Asaccharobacter	0.124	68	0.237	0.374	0.032	3
Bacillota unclassified	0.959	91	0.119	0.476	0.001	41
Bacteroides	35.904	99	0.271	0.325	0.030	3
Barnesiella	1.196	53	0.102	0.400	0.013	3
Bifidobacterium	1.283	69	0.136	0.393	0.006	8
Bilophila	0.035	54	0.068	0.333	0.001	30
Blautia	1.987	98	0.288	0.434	0.022	6
Butyricimonas	0.143	50	0.119	0.286	0.008	4
Clostridium	0.632	86	0.051	0.000	0.002	0
Collinsella	0.990	80	0.186	0.509	0.039	2
Coprobacter	0.096	33	0.034	1.000	0.000	0
Coprococcus	1.082	82	0.288	0.463	0.021	6
Dialister	0.280	40	0.136	0.357	0.045	1
Dorea	1.089	96	0.153	0.444	0.015	5
Eggerthella	0.050	30	0.254	0.438	0.027	4
Eisenbergiella	0.098	40	0.305	0.386	0.048	2
Enorma	0.012	25	0.034	0.000	0.005	0
Enterorhabdus	0.005	30	0.186	0.491	0.011	8
Erysipelatoclostridium	0.097	38	0.356	0.443	0.036	4
Escherichia	0.287	29	0.169	0.533	0.007	14
Eubacterium	4.500	98	0.203	0.515	0.006	18
Faecalibacterium	7.150	99	0.288	0.404	0.020	6
Flavonifractor	0.227	75	0.424	0.353	0.068	2
Fusicatenibacter	1.657	92	0.373	0.403	0.049	3
Gemmiger	0.120	69	0.254	0.429	0.026	4
Gordonibacter	0.049	49	0.153	0.472	0.010	7
Haemophilus	0.134	31	0.254	0.362	0.024	4
Holdemania	0.015	37	0.068	0.500	0.001	30
Intestinimonas	0.040	35	0.186	0.400	0.027	3
Lachnoclostridium	0.112	53	0.441	0.338	0.088	2

Genus	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Lachnospira	0.634	47	0.119	0.524	0.011	6
Lachnospiraceae unclassified	4.352	94	0.305	0.327	0.048	2
Lactobacillus	0.155	31	0.051	0.333	0.001	13
Lawsonibacter	0.129	79	0.119	0.286	0.014	2
Methanobrevibacter	0.226	31	0.085	0.600	0.001	92
Monoglobus	0.259	40	0.085	0.100	0.018	0
Odoribacter	0.555	76	0.169	0.333	0.033	2
Oscillibacter	1.003	79	0.339	0.368	0.049	3
Parabacteroides	3.235	95	0.203	0.439	0.010	9
Paraprevotella	0.169	24	0.169	0.311	0.049	1
Parasutterella	0.404	60	0.034	1.000	0.000	0
Phascolarctobacterium	0.556	66	0.153	0.500	0.008	9
Prevotella	5.351	27	0.136	0.321	0.017	3
Pseudomonadota unclassified	0.170	45	0.085	0.300	0.020	1
Roseburia	5.444	98	0.288	0.368	0.027	4
Ruminococcaceae unclassified	2.099	82	0.034	1.000	0.000	0
Ruminococcus	3.558	89	0.356	0.457	0.023	7
Ruthenibacterium	0.101	76	0.237	0.495	0.021	6
Streptococcus	0.543	69	0.136	0.464	0.003	18
Turicimonas	0.041	40	0.068	0.333	0.023	1
Veillonella	0.209	38	0.220	0.410	0.017	5

Family	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Acidaminococcaceae	0.701	69	0.217	0.100	0.150	0
Akkermansiaceae	1.065	50	0.000	0.000	0.000	0
Bacillota unclassified	0.959	91	0.043	0.000	0.000	0
Bacteroidaceae	35.904	99	0.043	0.000	0.000	0
Barnesiellaceae	1.292	63	0.174	0.000	0.087	0
Bifidobacteriaceae	1.283	69	0.174	0.667	0.023	5
Clostridiaceae	0.656	89	0.087	0.000	0.009	0
Clostridiales unclassified	0.546	88	0.217	0.100	0.139	0
Coriobacteriaceae	1.003	81	0.043	0.000	0.000	0
Desulfovibrionaceae	0.084	60	0.000	0.000	0.000	0
Eggerthellaceae	0.374	77	0.217	0.100	0.103	0
Enterobacteriaceae	0.531	35	0.130	0.333	0.009	5
Erysipelotrichaceae	0.218	73	0.130	0.333	0.016	3
Eubacteriaceae	4.501	99	0.130	0.333	0.011	4
Lachnospiraceae	18.688	100	0.000	0.000	0.000	0
Lactobacillaceae	0.155	31	0.043	0.000	0.000	0
Methanobacteriaceae	0.229	32	0.043	0.000	0.000	0
Odoribacteraceae	0.699	82	0.217	0.300	0.072	1
Oscillospiraceae	1.003	79	0.304	0.143	0.203	0
Pasteurellaceae	0.134	31	0.261	0.333	0.122	1
Peptostreptococcaceae	0.018	15	0.000	0.000	0.000	0

Family	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Prevotellaceae	5.520	38	0.174	0.167	0.062	0
Pseudomonadota unclassified	0.170	45	0.087	0.000	0.010	0
Rikenellaceae	5.513	93	0.348	0.179	0.195	0
Ruminococcaceae	13.757	100	0.000	0.000	0.000	0
Streptococcaceae	0.554	69	0.261	0.267	0.200	0
Sutterellaceae	0.460	72	0.130	0.000	0.058	0
Synergistaceae	0.019	11	0.000	0.000	0.000	0
Tannerellaceae	3.235	95	0.087	0.000	0.087	0
Veillonellaceae	0.535	60	0.261	0.333	0.108	1
Victivallaceae	0.014	11	0.000	0.000	0.000	0

Order	Mean relative abundance [%]	Prevalence [%]	Degree	Transitivity	Betweenness centrality	Keystone potential
Acidaminococcales	0.701	69	0.273	0.333	0.182	1
Bacillota unclassified	0.959	91	0.182	1.000	0.000	0
Bacteroidales	52.174	100	0.000	0.000	0.000	0
Bifidobacteriales	1.283	69	0.364	0.333	0.273	0
Burkholderiales	0.488	76	0.182	0.000	0.182	0
Coriobacteriales	1.004	82	0.182	1.000	0.000	0
Desulfovibrionales	0.084	60	0.000	0.000	0.000	0
Eggerthellales	0.374	77	0.364	0.333	0.200	1
Enterobacterales	0.532	35	0.273	0.333	0.327	0
Erysipelotrichales	0.218	73	0.273	0.333	0.273	0
Lactobacillales	0.746	79	0.182	1.000	0.000	0
Methanobacteriales	0.229	32	0.000	0.000	0.000	0
Pasteurellales	0.134	31	0.273	0.667	0.036	5
Pseudomonadota unclassified	0.170	45	0.091	0.000	0.000	0
Synergistales	0.019	11	0.000	0.000	0.000	0
Veillonellales	0.535	60	0.455	0.300	0.255	1
Verrucomicrobiales	1.065	50	0.000	0.000	0.000	0
Victivallales	0.014	11	0.000	0.000	0.000	0

Table S2 Keystone taxa and first neighbors (significantly correlated taxa)

Keystone genus	First neighbor	Correlation coefficient
Bacillota unclassified	*Oscillibacter*	0.36
Agathobaculum	*Coprococcus*	0.35
Holdemania	*Eisenbergiella*	0.35
Agathobaculum	*Collinsella*	0.34
Agathobaculum	*Lachnospiraceae unclassified*	0.34
Holdemania	*Lachnoclostridium*	0.33
Methanobrevibacter	*Oscillibacter*	0.33
Bilophila	*Intestinimonas*	0.32
Methanobrevibacter	*Intestinimonas*	-0.32
Bilophila	*Gemmiger*	-0.32
Methanobrevibacter	*Lachnoclostridium*	-0.32
Bilophila	*Anaerostipes*	-0.32
Bacillota unclassified	*Flavonifractor*	-0.33
Methanobrevibacter	*Flavonifractor*	-0.33
Bacillota unclassified	*Escherichia*	-0.33
Holdemania	*Alistipes*	-0.34
Methanobrevibacter	*Lachnospiraceae unclassified*	-0.35
Holdemania	*Bifidobacterium*	-0.35
Agathobaculum	*Escherichia*	-0.35
Bilophila	*Blautia*	-0.37
Bacillota unclassified	*Roseburia*	-0.37
Bacillota unclassified	*Blautia*	-0.39
Bacillota unclassified	*Faecalibacterium*	-0.39
Bacillota unclassified	*Bacteroides*	-0.41

Keystone species	First neighbor	Correlation coefficient
Veillonella atypica	*Veillonella dispar*	0.59
Veillonella atypica	*Veillonella parvula*	0.57
Veillonella atypica	*Haemophilus parainfluenzae*	0.54
Veillonella atypica	*Streptococcus parasanguinis*	0.49
Eisenbergiella tayi	*Clostridium bolteae*	0.39
Veillonella atypica	*Streptococcus salivarius*	0.36
Veillonella atypica	*Dialister invisus*	0.36
*Firmicutes bacterium CAG 83 **	*Coprococcus comes*	0.35
Eisenbergiella tayi	*Holdemania filiformis*	0.35
*Firmicutes bacterium CAG 83 **	*Odoribacter splanchnicus*	0.34
Eisenbergiella tayi	*Ruminococcus gnavus*	0.34
Ruminococcus lactaris	*Eubacterium siraeum*	0.34
Eisenbergiella tayi	*Escherichia coli*	0.34
Ruminococcus lactaris	*Odoribacter splanchnicus*	0.34
*Firmicutes bacterium CAG 83 **	*Gemmiger formicilis*	0.34
Ruminococcus lactaris	*Ruminococcus bromii*	0.33
Eisenbergiella tayi	*Eggerthella lenta*	0.33
Eisenbergiella tayi	*Enterorhabdus caecimuris*	0.32
Eisenbergiella tayi	*Flavonifractor plautii*	0.32
Eisenbergiella tayi	*Ruthenibacterium lactatiformans*	0.32
Ruminococcus lactaris	*Clostridium bolteae*	-0.32
*Firmicutes bacterium CAG 83 **	*Ruminococcus gnavus*	-0.33
Ruminococcus lactaris	*Clostridium symbiosum*	-0.33
Veillonella atypica	*Lawsonibacter asaccharolyticus*	-0.33
Ruminococcus lactaris	*Clostridium citroniae*	-0.33

Keystone species	First neighbor	Correlation coefficient
Ruminococcus lactaris	*Clostridium leptum*	-0.33
Ruminococcus lactaris	*Eisenbergiella massiliensis*	-0.33
*Firmicutes bacterium CAG 83 **	*Eggerthella lenta*	-0.33
Ruminococcus lactaris	*Ruminococcus gnavus*	-0.34
*Firmicutes bacterium CAG 83 **	*Clostridium symbiosum*	-0.34
Veillonella atypica	*Alistipes shahii*	-0.35
Ruminococcus lactaris	*Ruthenibacterium lactatiformans*	-0.35
Eisenbergiella tayi	*Oscillibacter sp 57 20*	-0.35
*Firmicutes bacterium CAG 83 **	*Dorea sp CAG 317*	-0.36

* unclassified Bacillota

Table S3 Enzyme commission numbers (ECs) identified as important in distinguishing transcriptional states of keystone taxa. Shown are significant predictors (p-value <= 0.05). Mean decrease in accuracy is scaled through division by standard error.

Taxon	**EC**	**Mean decrease accuracy**	**Indicated in the following pathways (*)**
Agathobaculum	1.1.1.18: Inositol 2-dehydrogenase	6.87	Streptomycin biosynthesis Inositol phosphate metabolism
Agathobaculum	2.7.13.3: Histidine kinase	3.204	-
Agathobaculum	1.3.8.1: Short-chain acyl-CoA dehydrogenase	2.036	-
Agathobaculum	5.3.1.9: Glucose-6-phosphate isomerase	2.028	Glycolysis/Gluconeogenesis Pentose phosphate pathway Starch and sucrose metabolism Amino sugar and nucleotide sugar metabolism
Agathobaculum	2.4.2.3: Uridine phosphorylase	1.005	Pyrimidine metabolism
Bilophila	2.6.1.77: Taurine-pyruvate aminotransferase	4.049	Taurine and hypotaurine metabolism
Bilophila	2.3.1.157: Glucosamine-1-phosphate N-acetyltransferase	1.754	Amino sugar and nucleotide sugar metabolism
Bilophila	7.1.3.1: H+-exporting disphosphatase	1.604	-
Bilophila	4.1.99.12: 3,4-dihydroxy-2-butanone-4-phosphate synthase	1.005	Riboflavin metabolism
Bilophila	2.7.4.9: dTMP kinase	1.005	Pyrimidine metabolism
Eisenbergiella tayi	1.17.1.8: 4-hydroxy-tetrahydrodipicolinate reductase	1.742	Monobactam biosynthesis Lysine biosynthesis
Eisenbergiella tayi	4.3.3.7: 4-hydroxy-tetrahydrodipicolinate synthase	1.477	Monobactam biosynthesis Lysine biosynthesis
Eisenbergiella tayi	2.7.2.4: Aspartate kinase	1.005	Glycine, serine and threonine metabolism Monobactam biosynthesis Cysteine and methionine metabolism Lysine biosynthesis
Firmicutes bacterium CAG 83 °	EC 2.7.7.6: DNA-directed RNA polymerase	6.068	-
Firmicutes bacterium CAG 83 °	3.4.21.92: Endopeptidase Clp	4.772	-
Methanobrevibacter	2.8.4.1: coenzyme-B sulfoethylthiotransferase	2.937	Methane metabolism
Methanobrevibacter	2.3.1.101: formylmethanofuran-tetrahydromethanopterin N-formyltransferase	2.291	Methane metabolism
Methanobrevibacter	1.17.1.9: formate dehydrogenase	1.903	Glyoxylate and dicarboxylate metabolism Methane metabolism

Taxon	EC	Mean decrease accuracy	Indicated in the following pathways (*)
Methanobrevibacter	4.1.2.43: 3-hexulose-6-phosphate synthase	1.722	Pentose phosphate pathway Methane metabolism
Methanobrevibacter	1.1.1.1: alcohol dehydrogenase	1.722	Glycolysis/Gluconeogenesis Fatty acid degradation Glycine, serine and threonine metabolism Tyrosine metabolism alpha-Linolenic acid metabolism Pyruvate metabolism Chloroalkane and chloroalkene degradation Naphthalene degradation Retinol metabolism Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450
Methanobrevibacter	2.3.1.174: 3-oxoadipyl-CoA thiolase	1.719	Phenylalanine metabolism Benzoate degradation
Methanobrevibacter	1.1.1.261: sn-glycerol-1-phosphate dehydrogenase	1.005	Glycerophospholipid metabolism
Ruminococcus lactaris	2.7.9.1: (pyruvate, phosphate dikinase)	2.276	Glycolysis/Gluconeogenesis Pyruvate metabolism Carbon fixation pathways in prokaryotes

* Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30. doi: 10.1093/nar/28.1.27

° unclassified *Bacillota*

SUPPLEMENTARY MATERIAL

CHAPTER 2

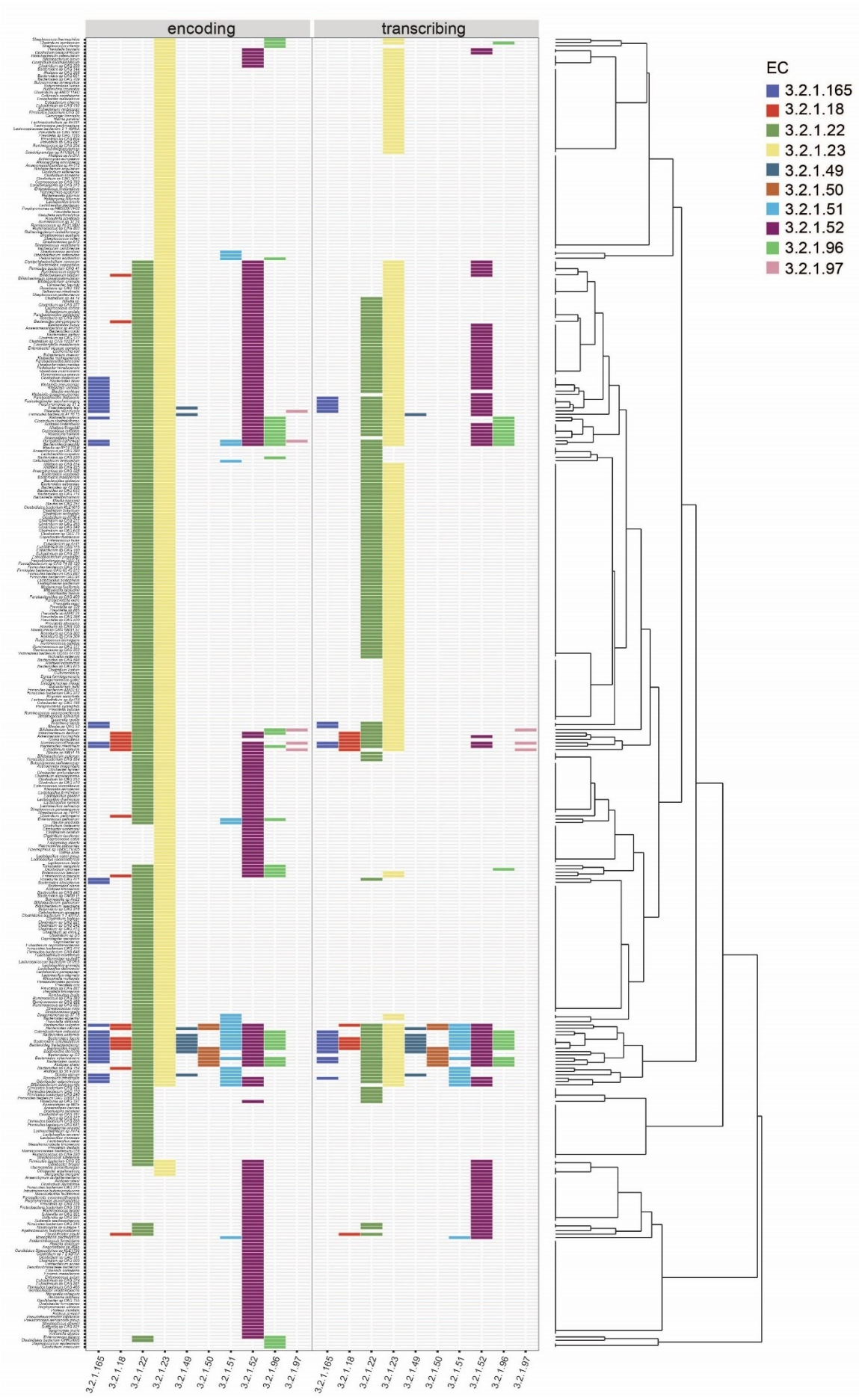
encoding
transcribing
EC
3.2.1.165
3.2.1.18
3.2.1.22
3.2.1.23
3.2.1.49
3.2.1.50
3.2.1.51
3.2.1.52
3.2.1.96
3.2.1.97

Figure S1 Species encoding (left column) and transcribing (right column) mucosal glycan degradation enzymes. Species are ordered based on their EC profiles in the metagenomes and metatranscriptomes. The dendrogram indicates which species have the most similar EC profiles.

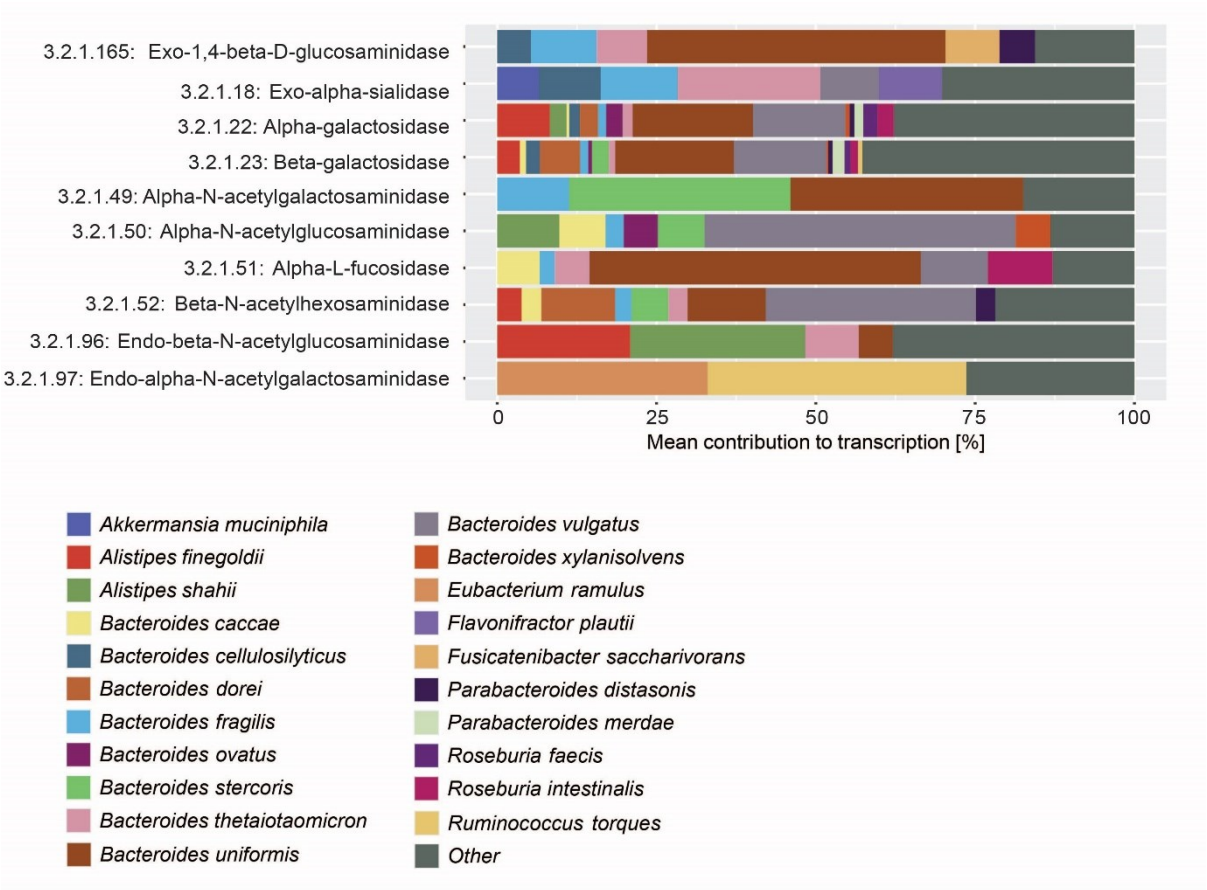

Figure S2 Average contribution of species to transcription of mucosal glycan degradation enzymes across analyzed samples [%]. Contribution of species transcribing an enzyme in < 10% of samples is summarized in the category “Other”.

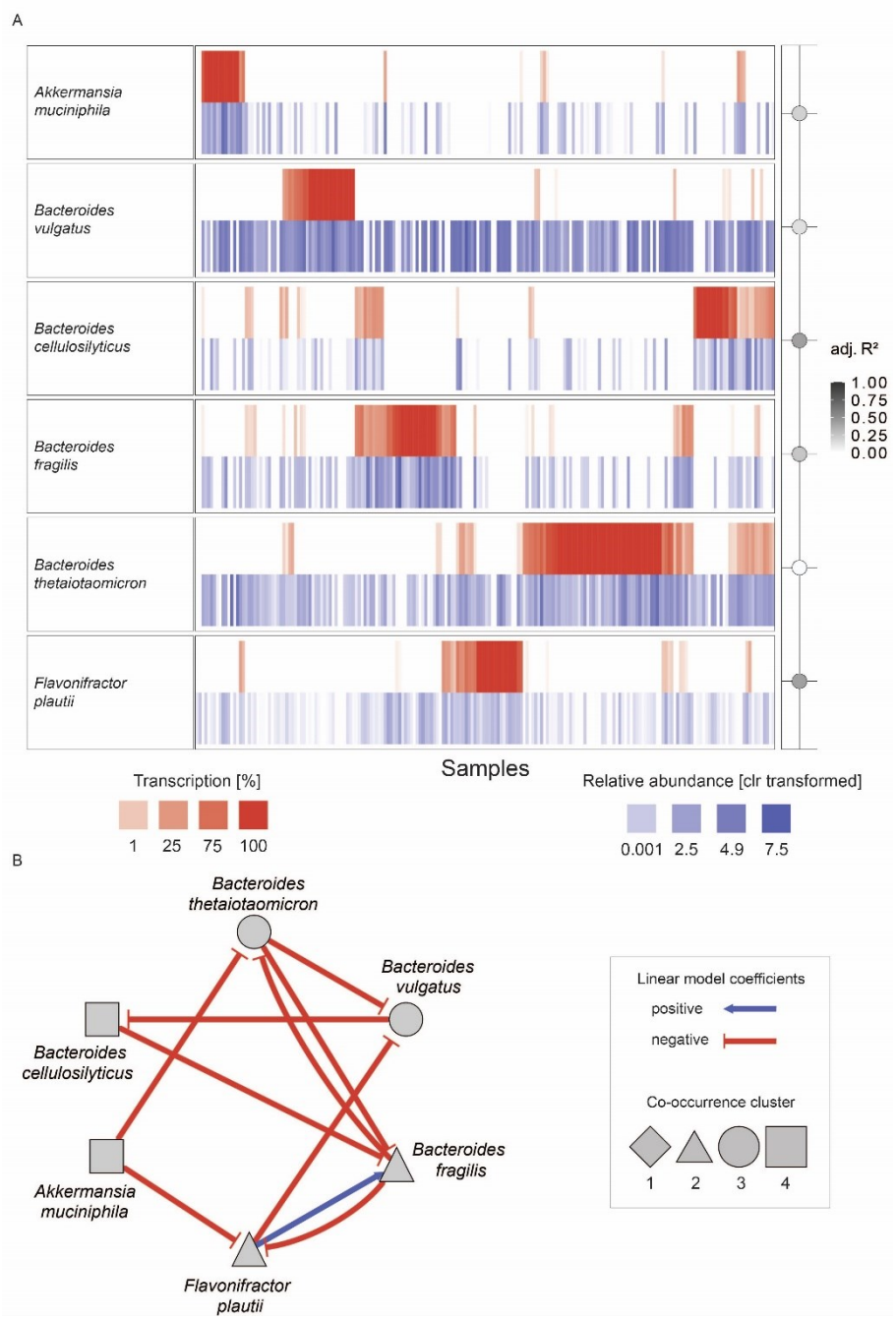

Figure S3 Contribution of species to transcription of EC 3.2.1.18: Exo-alpha-sialidase. (**A**) Transcription of EC 3.2.1.18 mapped to individual species (in % of total transcription grouped to EC 3.2.1.18) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.18 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

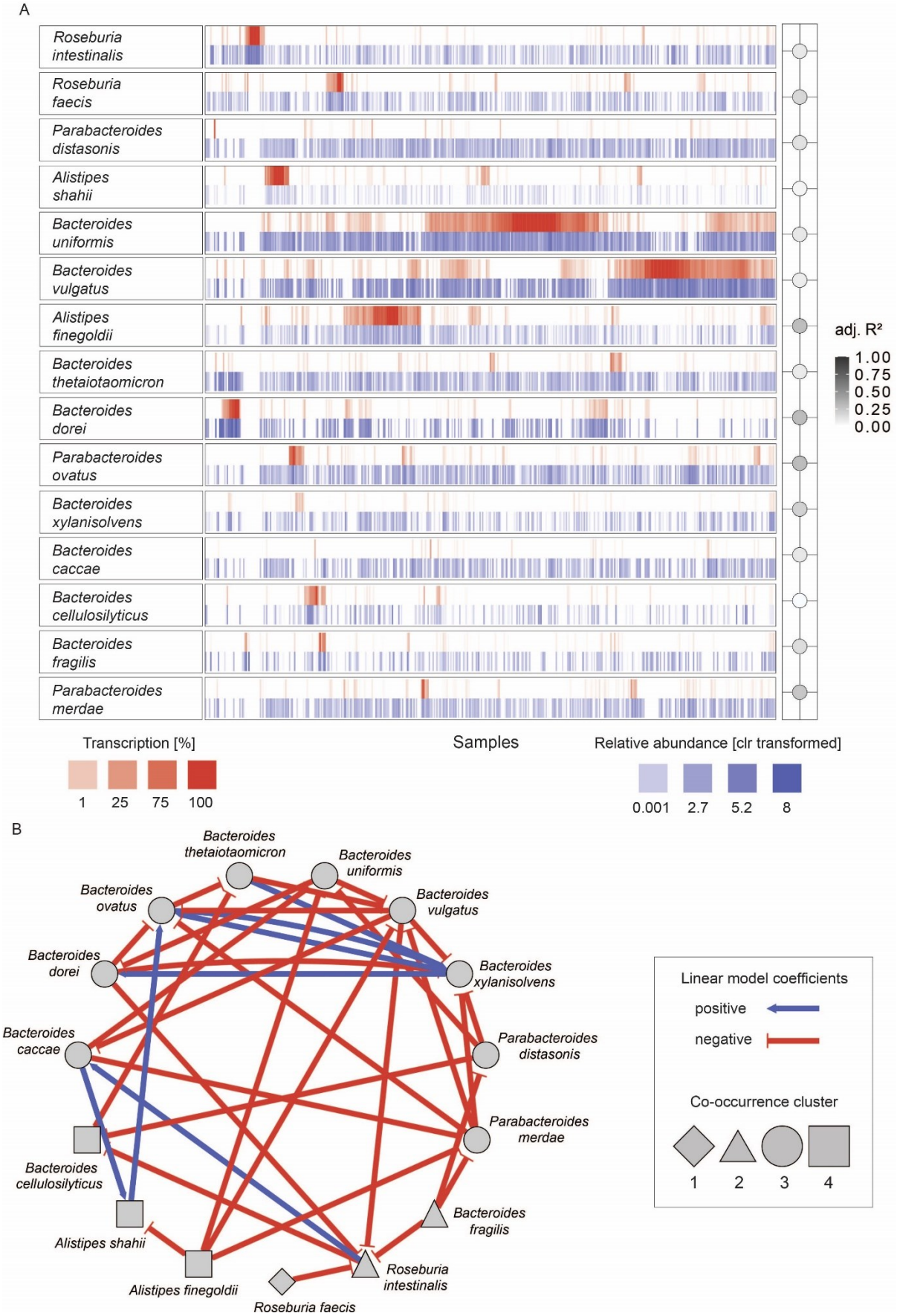
A
Roseburia intestinalis
Roseburia faecis
Parabacteroides distasonis
Alistipes shahii
Bacteroides uniformis
Bacteroides vulgatus
Alistipes finegoldii
Bacteroides thetaiotaomicron
Bacteroides dorei
Parabacteroides ovatus
Bacteroides xylanisolvens
Bacteroides caccae
Bacteroides cellulosilyticus
Bacteroides fragilis
Parabacteroides merdae
adj. R²
1.00
0.75
0.50
0.25
0.00
Transcription [%]
1
25
75
100
Samples
Relative abundance [clr transformed]
0.001
2.7
5.2
8
B
Bacteroides thetaiotaomicron
Bacteroides uniformis
Bacteroides ovatus
Bacteroides vulgatus
Bacteroides dorei
Bacteroides xylanisolvens
Bacteroides caccae
Parabacteroides distasonis
Parabacteroides merdae
Bacteroides cellulosilyticus
Bacteroides fragilis
Alistipes shahii
Roseburia intestinalis
Alistipes finegoldii
Roseburia faecis
Linear model coefficients
positive
negative
Co-occurrence cluster
1
2
3
4

Figure S4 Contribution of species to transcription of EC 3.2.1.22: Alpha-galactosidase. (**A**) Transcription of EC 3.2.1.22 mapped to individual species (in % of total transcription grouped to EC 3.2.1.22) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.22 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

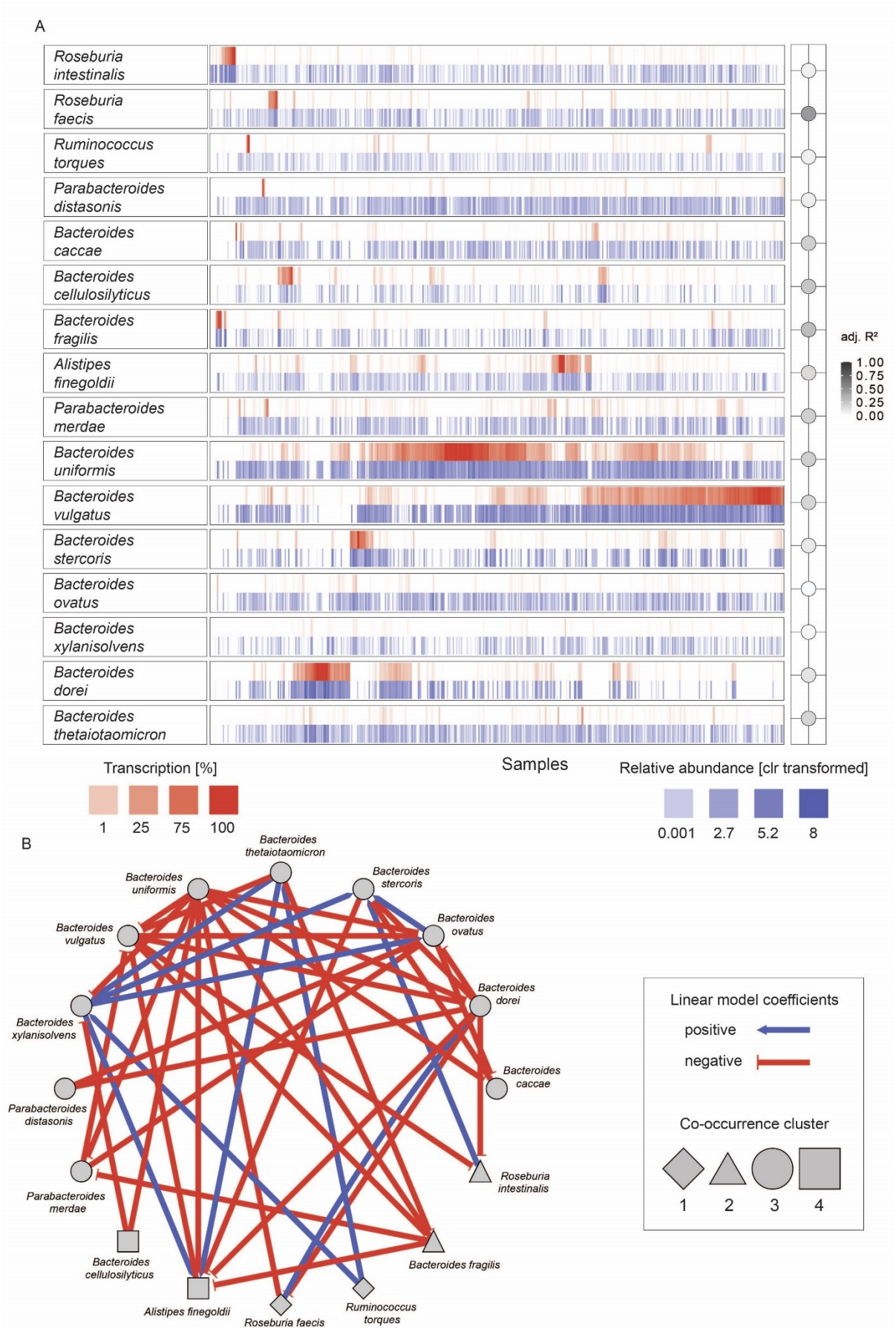
A
Roseburia intestinalis
Roseburia faecis
Ruminococcus torques
Parabacteroides distasonis
Bacteroides caccae
Bacteroides cellulosilyticus
Bacteroides fragilis
Alistipes finegoldii
Parabacteroides merdae
Bacteroides uniformis
Bacteroides vulgatus
Bacteroides stercoris
Bacteroides ovatus
Bacteroides xylanisolvens
Bacteroides dorei
Bacteroides thetaiotaomicron
adj. R²
1.00
0.75
0.50
0.25
0.00
Transcription [%]
1 25 75 100
Samples
Relative abundance [clr transformed]
0.001 2.7 5.2 8
B
Linear model coefficients
positive
negative
Co-occurrence cluster
1 2 3 4

Figure S5 Contribution of species to transcription of EC 3.2.1.23: Beta-galactosidase. (**A**) Transcription of EC 3.2.1.23 mapped to individual species (in % of total transcription grouped to EC 3.2.1.23) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.23 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

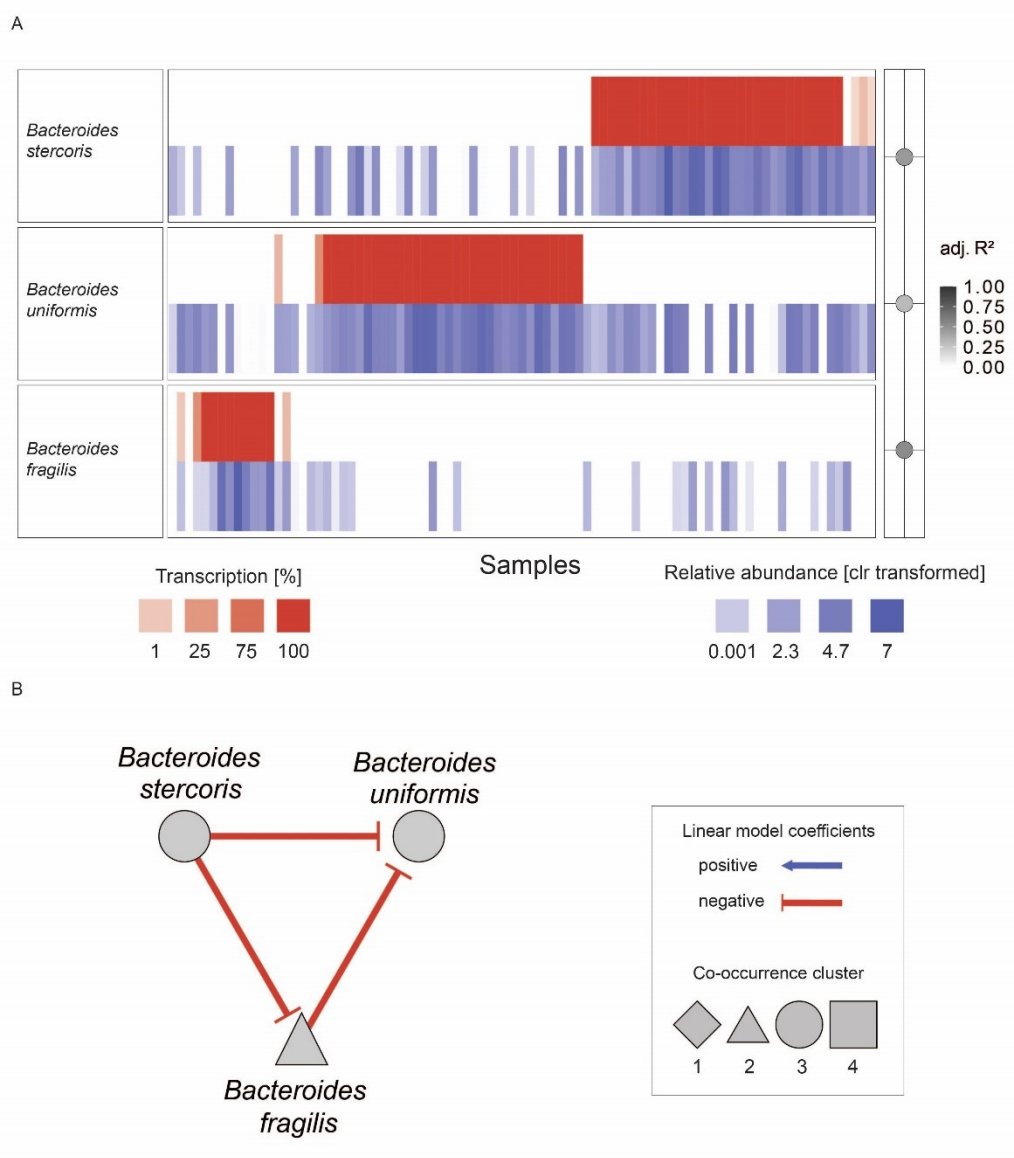

Figure S6 Contribution of species to transcription of EC 3.2.1.49: Alpha-N-acetylgalactosaminidase. (**A**) Transcription of EC 3.2.1.49 mapped to individual species (in % of total transcription grouped to EC 3.2.1.49) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.49 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

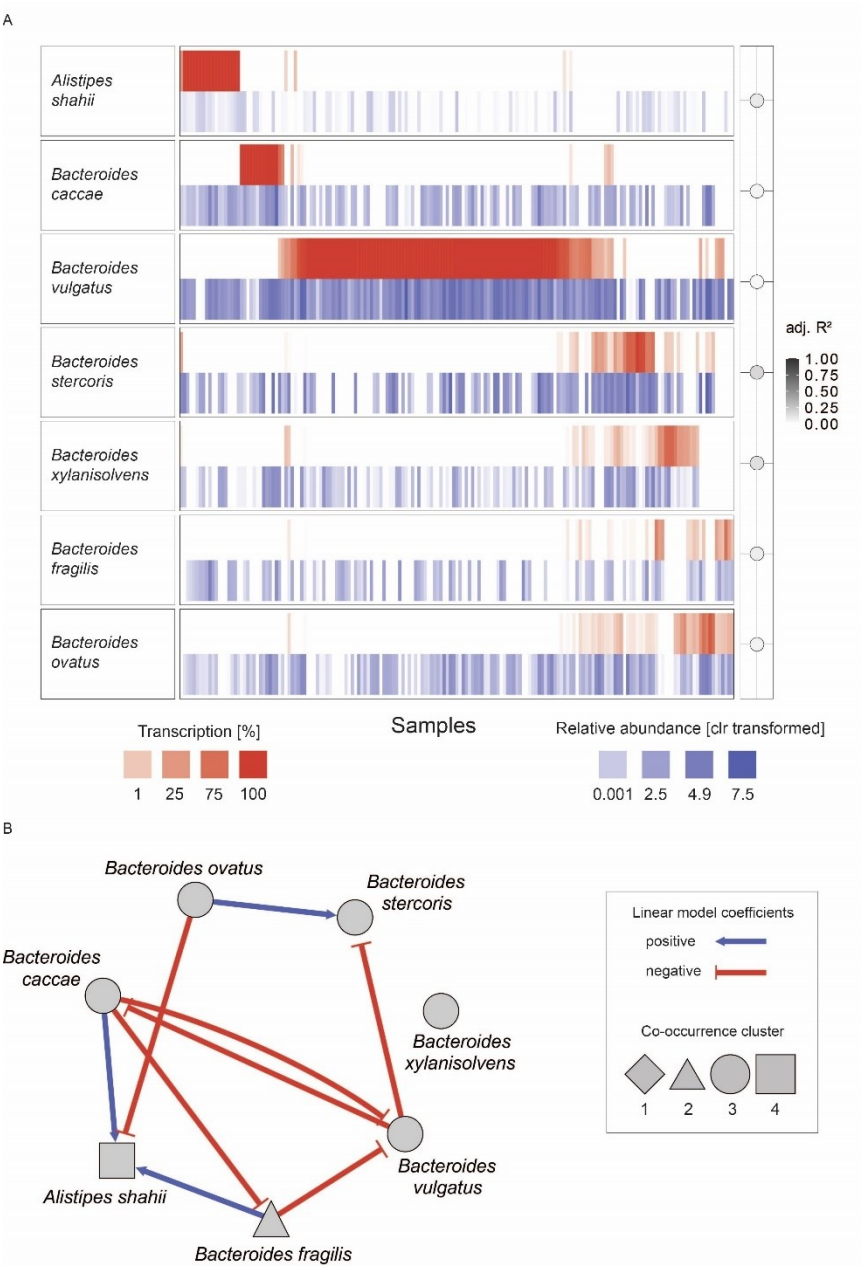

Figure S7 Contribution of species to transcription of EC 3.2.1.50: Alpha-N-acetylglucosaminidase. (**A**) Transcription of EC 3.2.1.50 mapped to individual species (in % of total transcription grouped to EC 3.2.1.50) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.50 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

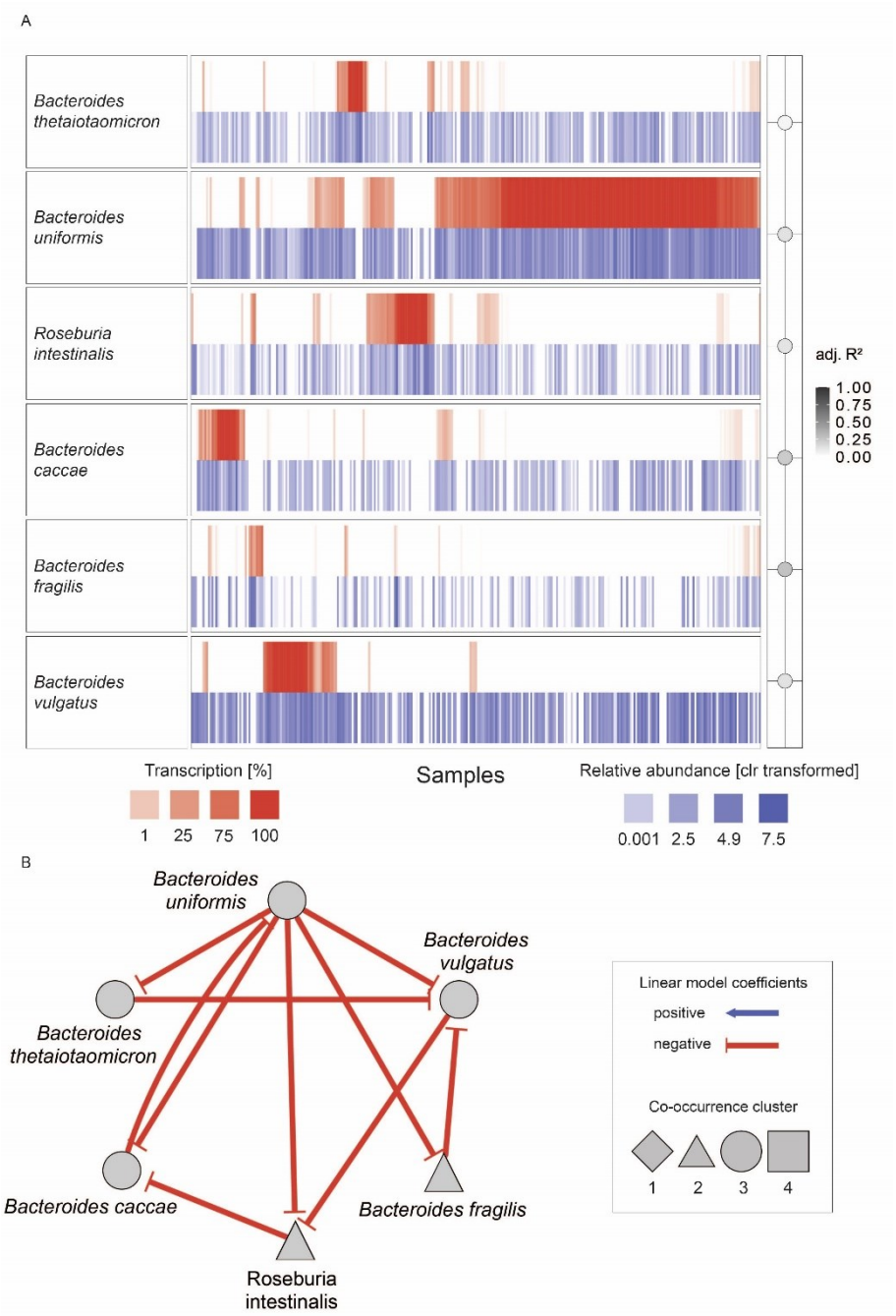

Figure S8 Contribution of species to transcription of EC 3.2.1.51: Alpha-L-fucosidase. (**A**) Transcription of EC 3.2.1.51 mapped to individual species (in % of total transcription grouped to EC 3.2.1.51) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.51 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

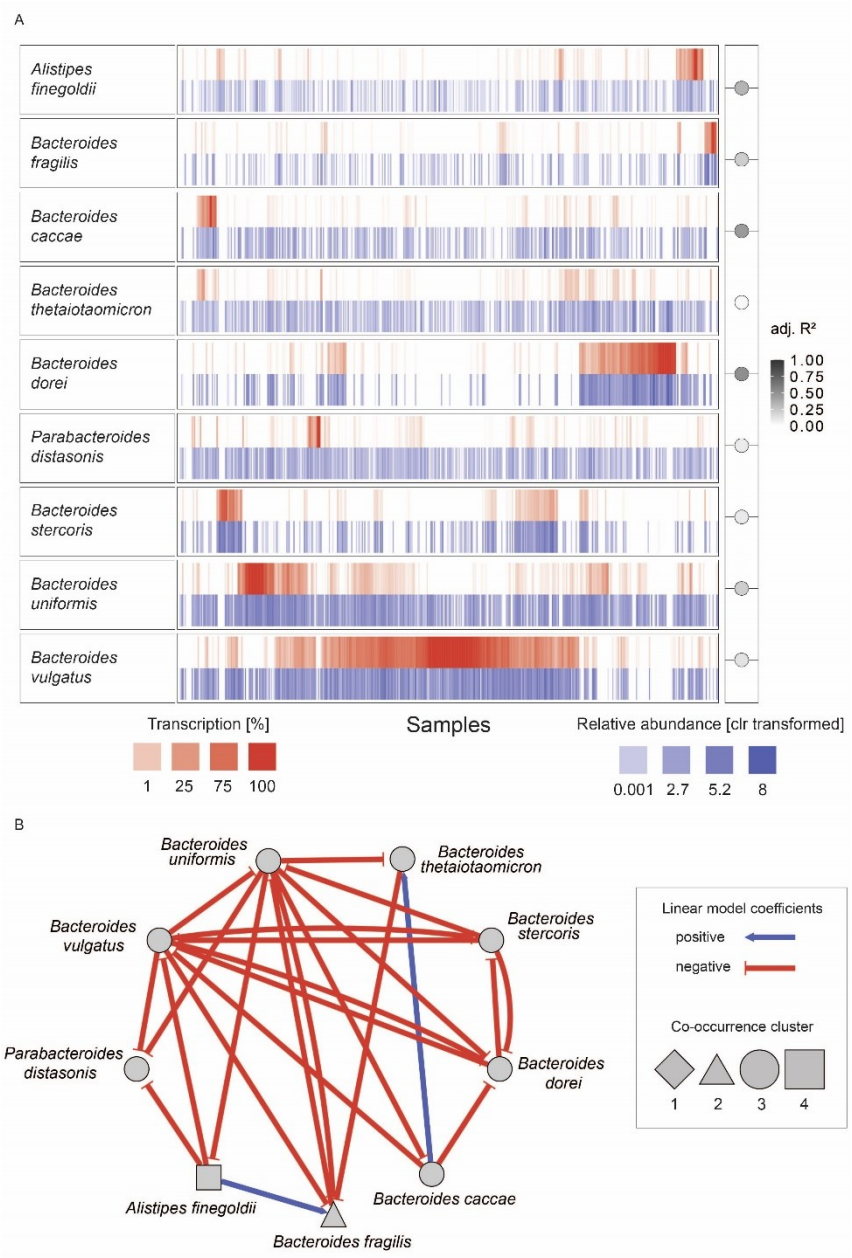

Figure S9 Contribution of species to transcription of EC 3.2.1.52: Beta-N-acetylhexosaminidase. (**A**) Transcription of EC 3.2.1.52 mapped to individual species (in % of total transcription grouped to EC 3.2.1.52) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.52 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

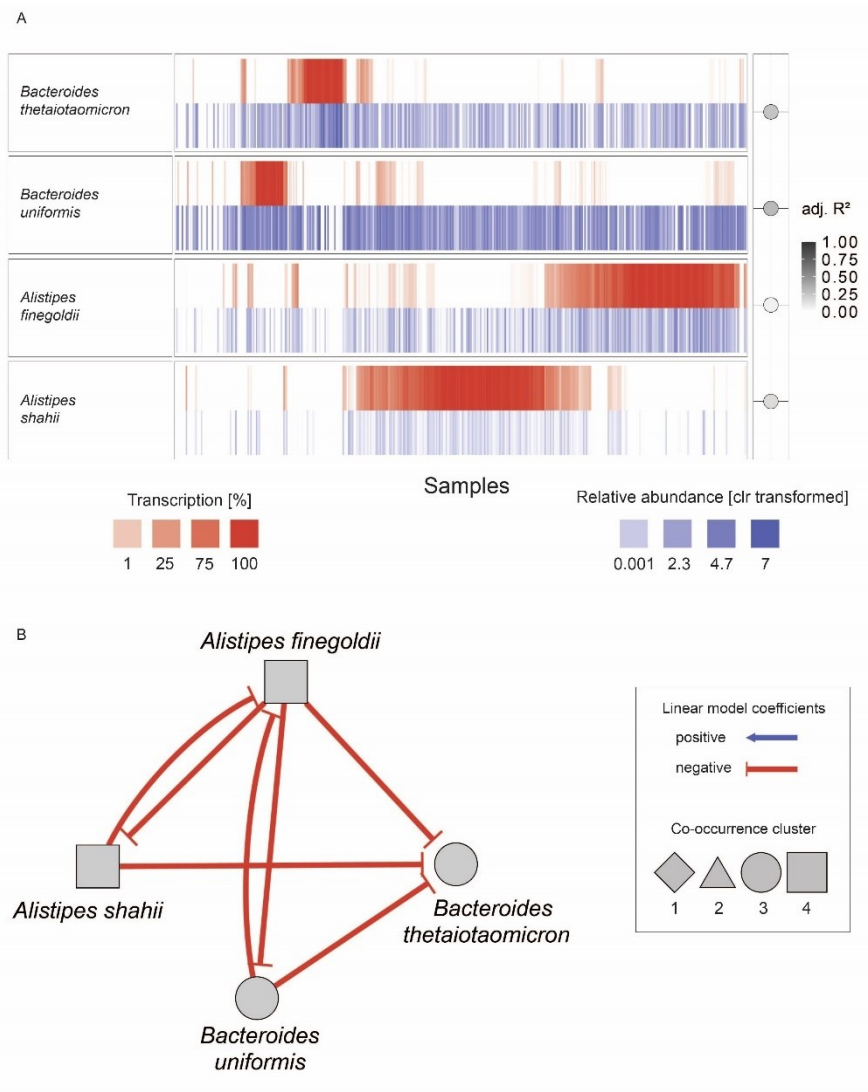

Figure S10 Contribution of species to transcription of EC 3.2.1.96: Endo-beta-N-Acetylglucosaminidase. (**A**) Transcription of EC 3.2.1.96 mapped to individual species (in % of total transcription grouped to EC 3.2.1.96) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.96 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

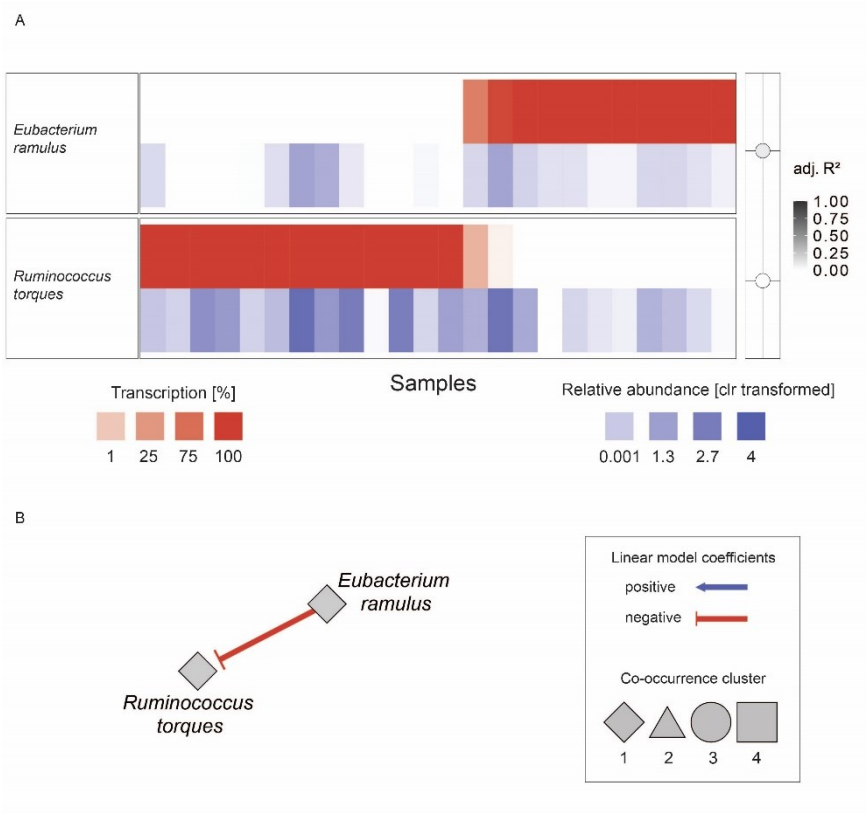

Figure S11 Contribution of species to transcription of EC 3.2.1.97: Endo-alpha-N-acetylgalactosaminidase. (**A**) Transcription of EC 3.2.1.97 mapped to individual species (in % of total transcription grouped to EC 3.2.1.97) is shown in red. Relative abundance (clr transformed) of prevalently (>= 10%) transcribing species is shown in blue. Percentage of abundance-driven transcription (= adjusted R^2 of linear models between species abundance and transcription) is shown in gray circles on the right. (**B**) Response-predictor network of EC 3.2.1.97 transcription. Nodes are prevalently transcribing species and node shape indicates the co-occurrence cluster a species belongs to (Fig 3A). Edges are significant coefficients of linear models (*Transcription response species [% of total EC transcription] ~ abundance predictor species 1 [clr transformed read counts] + … + abundance predictor species n [clr transformed read counts]*). Edges are directed from predictor to response species. Inhibition (negative coefficient) is indicated in red and facilitation (positive coefficient) is indicated in blue.

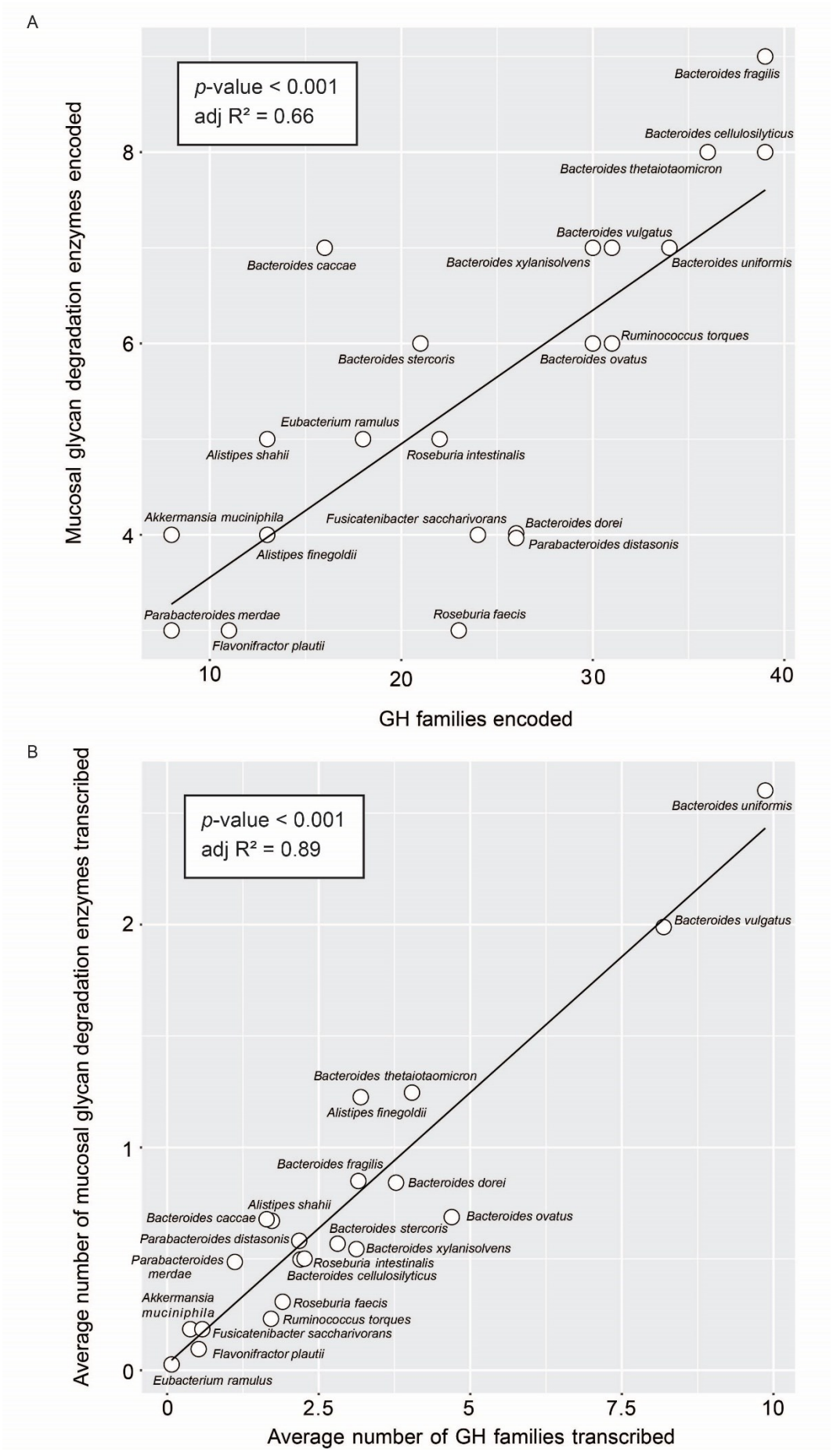
A
p-value < 0.001
adj R² = 0.66
Mucosal glycan degradation enzymes encoded
GH families encoded
Bacteroides fragilis
Bacteroides cellulosilyticus
Bacteroides thetaiotaomicron
Bacteroides vulgatus
Bacteroides caccae
Bacteroides xylanisolvens
Bacteroides uniformis
Ruminococcus torques
Bacteroides stercoris
Bacteroides ovatus
Eubacterium ramulus
Alistipes shahii
Roseburia intestinalis
Akkermansia muciniphila
Fusicatenibacter saccharivorans
Bacteroides dorei
Parabacteroides distasonis
Alistipes finegoldii
Parabacteroides merdae
Roseburia faecis
Flavonifractor plautii
B
p-value < 0.001
adj R² = 0.89
Average number of mucosal glycan degradation enzymes transcribed
Average number of GH families transcribed
Bacteroides uniformis
Bacteroides vulgatus
Bacteroides thetaiotaomicron
Alistipes finegoldii
Bacteroides fragilis
Bacteroides dorei
Alistipes shahii
Bacteroides caccae
Bacteroides ovatus
Bacteroides stercoris
Parabacteroides distasonis
Bacteroides xylanisolvens
Parabacteroides merdae
Roseburia intestinalis
Bacteroides cellulosilyticus
Akkermansia muciniphila
Roseburia faecis
Ruminococcus torques
Fusicatenibacter saccharivorans
Flavonifractor plautii
Eubacterium ramulus

Figure S12 Glycosyl hydrolase suite and mucosal glycan degradation enzyme suite of species. (**A**) Number of GH families encoded versus number of mucosal glycan degradation enzymes encoded (Linear model: p-value <0.001, adjusted R^2 = 0.66) (**B**) Average number of GH families transcribed versus average number of mucosal glycan degradation enzymes transcribed (Linear model: p-value < 0.001, adjusted R^2 = 0.89).

Table S1 Enzyme commission numbers, names, corresponding GH families and description of catalyzed reactions of mucosal glycan degradation enzymes

EC *	Name	Corresponding GH families **	Catalyzed reaction
3.1.6.4 °	Galactose-6-sulfate sulfatase	Sulfatase	hydrolyzes sufate groups of GalNAc and Gal (Kanehisa and Goto 2000)
3.1.6.8 °	Cerebroside-sulfatase	Sulfatase	hydrolyzes Gal-3-sulfate residues (Kanehisa and Goto 2000)
3.1.6.11 °	Disulfoglucosamine-6-sulfatase	Sulfatase	hydrolyzes sulfate groups of GlcNAc, may be identical to 3.2.1.14 (Kanehisa and Goto 2000)
3.1.6.14	N-acetylglucosamine-6-sulfatase	Sulfatase	hydrolyzes sulfate groups of GlcNAc, may be identical to 3.2.1.11 (Kanehisa and Goto 2000)
3.2.1.102 °	Blood-group-substance Endo-1,4-beta-galactosidase	GH98	hydrolyzes beta-linkages of Gal with GlcNAc in blood group A and B glycoconjugates (Berkhout, Plugge and Belzer 2022)
3.2.1.111 °	1,3-alpha-L-fucosidase	GH29	hydrolyzes alpha-linkages between Fuc and Gal or GlcNAc (Ravcheev and Thiele 2017)
3.2.1.165	Exo-1,4-beta-D-glucosaminidase	GH2, GH9, GH35	hydrolyzes alpha-linkages between GlcNAc and GalNAc (Berkhout, Plugge and Belzer 2022)
3.2.1.18	Exo-alpha-sialidase	GH1, GH33, GH34, GH83, GH156, GH177, GH181	cleaves off Neu5Ac from mucosal glycan structures (Ravcheev and Thiele 2017)
3.2.1.22	Alpha-galactosidase	GH4, GH27, GH31, GH36, GH97	hydrolysis alpha-linkages between Gal and Gal (Ravcheev and Thiele 2017)
3.2.1.23	Beta-galactosidase	GH1, GH2, GH3, GH5, GH35, GH42	hydrolyzes beta-linkages of Gal with GalNAc or GlcNAc (Ravcheev and Thiele 2017)
3.2.1.49	Alpha-N-acetylgalactosaminidase	GH27, GH36, GH109, GH129	hydrolyzes alpha-linkages between GalNAc and Gal as well as GalNAc and GalNAc (Ravcheev and Thiele 2017)
3.2.1.50	Alpha-N-acetylglucosaminidase	GH89	hydrolyzes alpha-linkages between GlcNAc and Gal (Ravcheev and Thiele 2017)
3.2.1.51	Alpha-L-fucosidase	GH18, GH29, GH95, GH141, GH151	hydrolyzes alpha-linkages between Fuc and Gal or GlcNAc (Ravcheev and Thiele 2017)
3.2.1.52	Beta-N-acetylhexosaminidase	GH2, GH3, GH5, GH18, GH20, GH84	hydrolyzes beta-linkages between GalNAc and Gal as well as between GlcNAc and Gal or GalNAc (Ravcheev and Thiele 2017)
3.2.1.63 °	1,2-alpha-L-fucosidase	GH29, GH95	hydrolyzes alpha-linkages between Fuc and Gal or GlcNAc (Ravcheev and Thiele 2017)
3.2.1.96	Endo-beta-N-Acetylglucosaminidase	GH18, GH20, GH73, GH85	hydrolyzes beta-linkages between GlcNAc and GlcNAc (Berkhout, Plugge and Belzer 2022, Kanehisa and Goto 2000)
3.2.1.97	Endo-alpha-N-acetylgalactosaminidase	GH101	hydrolyzes alpha-linkages between GalNAc and Ser/Thr residues of mucin peptide chains (Ravcheev and Thiele 2017)
4.2.2.15 °	Anhydrosialidase	Lyase	releases anhydro-Neu5Ac from mucosal glycan structures (Berkhout, Plugge and Belzer 2022)

* Glover, Ticer and Engevik 2022; Berkhout, Plugge and Belzer 2022; Ravcheev and Thiele 2017

** Drula et al. 2022; Paysan-Lafosse et al. 2023

° not detected in the analyzed metagenome and metatranscriptome samples

Table S2 Node and edge table of co-occurrence network between mucosal glycan degraders and their first neighbors

Node table

Species	Mucosal glycan degrader	Co-occurrence cluster
Acidaminococcus intestini	NO	3
Agathobaculum butyriciproducens	NO	1
Akkermansia muciniphila	YES	4
Alistipes finegoldii	YES	4
Alistipes indistinctus	NO	4
Alistipes inops	NO	4
Alistipes obesi	NO	4
Alistipes onderdonkii	NO	1
Alistipes putredinis	NO	4
Alistipes shahii	YES	4
Anaeromassilibacillus sp An250	NO	2
Anaerostipes caccae	NO	2
Anaerostipes hadrus	NO	1
Anaerotignum lactatifermentans	NO	3
Anaerotruncus colihominis	NO	2
Asaccharobacter celatus	NO	1
Azospirillum sp CAG 260	NO	1
Bacteroides caccae	YES	3
Bacteroides cellulosilyticus	YES	4
Bacteroides clarus	NO	1
Bacteroides dorei	YES	3
Bacteroides eggerthii	NO	1
Bacteroides finegoldii	NO	3
Bacteroides fragilis	YES	2
Bacteroides galacturonicus	NO	1
Bacteroides intestinalis	NO	2
Bacteroides massiliensis	NO	3
Bacteroides nordii	NO	2
Bacteroides ovatus	YES	3
Bacteroides salyersiae	NO	3
Bacteroides sp D2	NO	3
Bacteroides stercoris	YES	3
Bacteroides thetaiotaomicron	YES	3
Bacteroides uniformis	YES	3
Bacteroides vulgatus	YES	3
Bacteroides xylanisolvens	YES	3
Barnesiella intestinihominis	NO	4
Bifidobacterium adolescentis	NO	3
Bifidobacterium bifidum	NO	3
Bilophila wadsworthia	NO	4
Blautia obeum	NO	1
Blautia producta	NO	2
Blautia sp	NO	1
Blautia sp AF19 10LB	NO	3
Blautia sp CAG 257	NO	2
Blautia sp CAG 52	NO	1
Blautia wexlerae	NO	1
Butyricimonas synergistica	NO	4

Species	Mucosal glycan degrader	Co-occurrence cluster
Butyricimonas virosa	NO	4
Butyrivibrio crossotus	NO	4
Candidatus Gastranaerophilales bacterium HUM 10	NO	4
Candidatus Gastranaerophilales bacterium HUM 11	NO	4
Candidatus Methanomassiliicoccus intestinalis	NO	4
Candidatus Stoquefichus sp KLE1796	NO	2
Clostridiales bacterium KLE1615	NO	1
Clostridiales Family XIII bacterium	NO	4
Clostridium aldenense	NO	2
Clostridium asparagiforme	NO	2
Clostridium bolteae	NO	2
Clostridium citroniae	NO	2
Clostridium clostridioforme	NO	2
Clostridium disporicum	NO	1
Clostridium innocuum	NO	2
Clostridium lavalense	NO	2
Clostridium leptum	NO	3
Clostridium scindens	NO	2
Clostridium sp AF36 4	NO	3
Clostridium sp AM22 11AC	NO	1
Clostridium sp CAG 1013	NO	3
Clostridium sp CAG 122	NO	3
Clostridium sp CAG 12237 41	NO	3
Clostridium sp CAG 138	NO	4
Clostridium sp CAG 167	NO	4
Clostridium sp CAG 169	NO	1
Clostridium sp CAG 217	NO	1
Clostridium sp CAG 226	NO	4
Clostridium sp CAG 245	NO	1
Clostridium sp CAG 269	NO	1
Clostridium sp CAG 273	NO	4
Clostridium sp CAG 302	NO	4
Clostridium sp CAG 451	NO	4
Clostridium sp CAG 505	NO	3
Clostridium sp CAG 58	NO	1
Clostridium sp CAG 75	NO	1
Clostridium spiroforme	NO	2
Clostridium symbiosum	NO	2
Collinsella aerofaciens	NO	1
Collinsella intestinalis	NO	2
Coprobacter fastidiosus	NO	4
Coprococcus catus	NO	1
Coprococcus comes	NO	1
Coprococcus eutactus	NO	4

Species	Mucosal glycan degrader	Co-occurrence cluster
Cryptobacterium sp CAG 338	NO	4
Cutibacterium acnes	NO	3
Desulfovibrio fairfieldensis	NO	4
Desulfovibrio piger	NO	4
Desulfovibrionaceae bacterium	NO	1
Dialister invisus	NO	2
Dielma fastidiosa	NO	2
Dorea formicigenerans	NO	1
Dorea longicatena	NO	1
Dorea sp CAG 317	NO	2
Dysgonomonas gadei	NO	2
Dysgonomonas mossii	NO	2
Dysgonomonas sp 37 18	NO	2
Eggerthella lenta	NO	2
Eggerthella sp CAG 209	NO	1
Eggerthella sp CAG 298	NO	1
Eisenbergiella massiliensis	NO	2
Erysipelatoclostridium ramosum	NO	2
Eubacterium eligens	NO	1
Eubacterium hallii	NO	1
Eubacterium ramulus	YES	1
Eubacterium rectale	NO	1
Eubacterium siraeum	NO	4
Eubacterium sp 36 13	NO	1
Eubacterium sp An11	NO	3
Eubacterium sp CAG 115	NO	4
Eubacterium sp CAG 161	NO	1
Eubacterium sp CAG 251	NO	4
Eubacterium sp OM08 24	NO	1
Eubacterium ventriosum	NO	1
Faecalibacterium prausnitzii	NO	2
Faecalibacterium sp CAG 74	NO	4
Faecalibacterium sp CAG 74 58 120	NO	4
Firmicutes bacterium AF16 15	NO	1
Firmicutes bacterium CAG 103	NO	1
Firmicutes bacterium CAG 110	NO	1
Firmicutes bacterium CAG 114	NO	1
Firmicutes bacterium CAG 124	NO	4
Firmicutes bacterium CAG 129	NO	4
Firmicutes bacterium CAG 137	NO	4
Firmicutes bacterium CAG 145	NO	1
Firmicutes bacterium CAG 176	NO	4
Firmicutes bacterium CAG 240	NO	4
Firmicutes bacterium CAG 24053 14	NO	4
Firmicutes bacterium CAG 272	NO	4
Firmicutes bacterium CAG 321	NO	4

Species	Mucosal glycan degrader	Co-occurrence cluster
Firmicutes bacterium CAG 41	NO	1
Firmicutes bacterium CAG 466	NO	2
Firmicutes bacterium CAG 56	NO	1
Firmicutes bacterium CAG 65 45 313	NO	1
Firmicutes bacterium CAG 83	NO	1
Firmicutes bacterium CAG 95	NO	4
Flavonifractor plautii	YES	2
Fusicatenibacter saccharivorans	YES	1
Gemmiger formicilis	NO	1
Gordonibacter pamelaeae	NO	1
Haemophilus parainfluenzae	NO	2
Holdemanella biformis	NO	1
Hungatella hathewayi	NO	2
Intestinimonas butyriciproducens	NO	2
Lachnospira pectinoschiza	NO	1
Lachnospiraceae bacterium OF09 6	NO	1
Lactococcus lactis	NO	1
Lawsonibacter asaccharolyticus	NO	3
Lentisphaerae bacterium	NO	4
Megasphaera sp DISK 18	NO	3
Megasphaera sp MJR8396C	NO	3
Methanobrevibacter smithii	NO	4
Monoglobus pectinilyticus	NO	3
Odoribacter splanchnicus	NO	3
Oscillibacter sp 57 20	NO	4
Oscillibacter sp CAG 241	NO	1
Parabacteroides distasonis	YES	3
Parabacteroides goldsteinii	NO	4
Parabacteroides johnsonii	NO	4
Parabacteroides merdae	YES	3
Paraprevotella xylaniphila	NO	4
Phascolarctobacterium faecium	NO	4
Phascolarctobacterium succinatutens	NO	4
Porphyromonas sp 31 2	NO	4
Prevotella corporis	NO	2
Proteobacteria bacterium CAG 495	NO	1
Proteus mirabilis	NO	2
Roseburia faecis	YES	1
Roseburia hominis	NO	1
Roseburia intestinalis	YES	2
Roseburia inulinivorans	NO	2
Roseburia sp CAG 197	NO	1
Roseburia sp CAG 309	NO	4
Roseburia sp CAG 471	NO	1
Ruminococcaceae bacterium D16	NO	3
Ruminococcaceae bacterium D5	NO	4

Species	Mucosal glycan degrader	Co-occurrence cluster
Ruminococcaceae bacterium KLE1738	NO	1
Ruminococcus bicirculans	NO	1
Ruminococcus bromii	NO	4
Ruminococcus callidus	NO	1
Ruminococcus gnavus	NO	2
Ruminococcus lactaris	NO	1
Ruminococcus sp 37 24	NO	4
Ruminococcus sp CAG 177	NO	4
Ruminococcus torques	YES	1
Ruthenibacterium lactatiformans	NO	2
Sanguibacteroides justesenii	NO	4
Sellimonas intestinalis	NO	2
Senegalimassilia anaerobia	NO	4
Slackia isoflavoniconvertens	NO	1
Streptococcus australis	NO	1
Streptococcus salivarius	NO	1
Streptococcus sp A12	NO	1
Subdoligranulum sp	NO	1
Subdoligranulum sp APC924 74	NO	1
Sutterella wadsworthensis	NO	4
Tyzzerella nexilis	NO	2
Tyzzerella sp	NO	1
Veillonella dispar	NO	2
Victivallales bacterium CCUG 44730	NO	4
Victivallis vadensis	NO	4

Edge table

Mucosal glycan degrader	First neighbor	Correlation strength
Bacteroides stercoris	*Acidaminococcus intestini*	0.22
Bacteroides dorei	*Agathobaculum butyriciproducens*	0.22
Eubacterium ramulus	*Agathobaculum butyriciproducens*	0.40
Fusicatenibacter saccharivorans	*Agathobaculum butyriciproducens*	0.46
Roseburia faecis	*Agathobaculum butyriciproducens*	0.30
Ruminococcus torques	*Agathobaculum butyriciproducens*	0.27
Ruminococcus torques	*Akkermansia muciniphila*	0.24
Bacteroides ovatus	*Alistipes finegoldii*	0.22
Bacteroides uniformis	*Alistipes finegoldii*	0.29
Alistipes shahii	*Alistipes indistinctus*	0.30
Eubacterium ramulus	*Alistipes indistinctus*	0.21
Parabacteroides distasonis	*Alistipes indistinctus*	0.23
Alistipes shahii	*Alistipes inops*	0.38
Alistipes shahii	*Alistipes obesi*	0.45
Fusicatenibacter saccharivorans	*Alistipes obesi*	0.31
Ruminococcus torques	*Alistipes obesi*	0.26
Fusicatenibacter saccharivorans	*Alistipes onderdonkii*	0.21
Alistipes shahii	*Alistipes putredinis*	0.46
Bacteroides cellulosilyticus	*Alistipes putredinis*	0.23
Bacteroides dorei	*Alistipes putredinis*	0.21
Bacteroides uniformis	*Alistipes putredinis*	0.23
Bacteroides xylanisolvens	*Alistipes putredinis*	0.27
Bacteroides stercoris	*Alistipes shahii*	0.26
Flavonifractor plautii	*Anaeromassilibacillus sp An250*	0.26
Flavonifractor plautii	*Anaerostipes caccae*	0.36
Roseburia intestinalis	*Anaerostipes caccae*	0.24
Fusicatenibacter saccharivorans	*Anaerostipes hadrus*	0.31
Ruminococcus torques	*Anaerostipes hadrus*	0.21
Bacteroides uniformis	*Anaerotignum lactatifermentans*	0.22
Flavonifractor plautii	*Anaerotignum lactatifermentans*	0.36
Parabacteroides distasonis	*Anaerotignum lactatifermentans*	0.25
Flavonifractor plautii	*Anaerotruncus colihominis*	0.36
Eubacterium ramulus	*Asaccharobacter celatus*	0.31
Fusicatenibacter saccharivorans	*Asaccharobacter celatus*	0.53
Roseburia faecis	*Asaccharobacter celatus*	0.22
Ruminococcus torques	*Asaccharobacter celatus*	0.28
Fusicatenibacter saccharivorans	*Azospirillum sp CAG 260*	0.26
Bacteroides ovatus	*Bacteroides caccae*	0.27
Bacteroides stercoris	*Bacteroides caccae*	0.24
Bacteroides uniformis	*Bacteroides caccae*	0.43
Bacteroides vulgatus	*Bacteroides caccae*	0.24
Parabacteroides distasonis	*Bacteroides caccae*	0.40
Parabacteroides merdae	*Bacteroides caccae*	0.21

Mucosal glycan degrader	First neighbor	Correlation strength
Bacteroides thetaiotaomicron	*Bacteroides cellulosilyticus*	0.21
Bacteroides uniformis	*Bacteroides cellulosilyticus*	0.21
Roseburia faecis	*Bacteroides clarus*	0.21
Bacteroides thetaiotaomicron	*Bacteroides dorei*	0.32
Fusicatenibacter saccharivorans	*Bacteroides dorei*	0.30
Roseburia faecis	*Bacteroides eggerthii*	0.24
Bacteroides stercoris	*Bacteroides finegoldii*	0.25
Parabacteroides merdae	*Bacteroides finegoldii*	0.25
Roseburia faecis	*Bacteroides galacturonicus*	0.20
Fusicatenibacter saccharivorans	*Bacteroides intestinalis*	0.23
Ruminococcus torques	*Bacteroides intestinalis*	0.20
Bacteroides stercoris	*Bacteroides massiliensis*	0.26
Bacteroides stercoris	*Bacteroides ovatus*	0.20
Bacteroides uniformis	*Bacteroides ovatus*	0.34
Bacteroides vulgatus	*Bacteroides ovatus*	0.21
Parabacteroides distasonis	*Bacteroides ovatus*	0.25
Parabacteroides merdae	*Bacteroides ovatus*	0.29
Bacteroides stercoris	*Bacteroides salyersiae*	0.23
Bacteroides xylanisolvens	*Bacteroides salyersiae*	0.20
Bacteroides vulgatus	*Bacteroides stercoris*	0.26
Bacteroides xylanisolvens	*Bacteroides stercoris*	0.30
Parabacteroides merdae	*Bacteroides stercoris*	0.36
Bacteroides uniformis	*Bacteroides thetaiotaomicron*	0.27
Parabacteroides distasonis	*Bacteroides thetaiotaomicron*	0.29
Bacteroides vulgatus	*Bacteroides uniformis*	0.28
Bacteroides xylanisolvens	*Bacteroides uniformis*	0.22
Parabacteroides distasonis	*Bacteroides uniformis*	0.42
Bacteroides xylanisolvens	*Bacteroides vulgatus*	0.26
Parabacteroides distasonis	*Bacteroides vulgatus*	0.29
Parabacteroides merdae	*Bacteroides vulgatus*	0.33
Parabacteroides merdae	*Bacteroides xylanisolvens*	0.23
Fusicatenibacter saccharivorans	*Barnesiella intestinihominis*	0.20
Ruminococcus torques	*Barnesiella intestinihominis*	0.25
Parabacteroides merdae	*Bifidobacterium adolescentis*	0.21
Roseburia faecis	*Bifidobacterium adolescentis*	0.26
Parabacteroides merdae	*Bifidobacterium bifidum*	0.23
Eubacterium ramulus	*Blautia obeum*	0.30
Fusicatenibacter saccharivorans	*Blautia obeum*	0.42
Ruminococcus torques	*Blautia obeum*	0.45
Flavonifractor plautii	*Blautia producta*	0.34
Eubacterium ramulus	*Blautia sp*	0.28
Fusicatenibacter saccharivorans	*Blautia sp*	0.33
Fusicatenibacter saccharivorans	*Blautia sp AF19 10LB*	0.24
Flavonifractor plautii	*Blautia sp CAG 257*	0.39
Eubacterium ramulus	*Blautia sp CAG 52*	0.36

Mucosal glycan degrader	First neighbor	Correlation strength
Fusicatenibacter saccharivorans	*Blautia sp CAG 52*	0.39
Ruminococcus torques	*Blautia sp CAG 52*	0.33
Fusicatenibacter saccharivorans	*Blautia wexlerae*	0.26
Ruminococcus torques	*Butyricimonas virosa*	0.21
Flavonifractor plautii	*Candidatus Stoquefichus sp KLE1796*	0.27
Eubacterium ramulus	*Clostridiales Family XIII bacterium*	0.20
Eubacterium ramulus	*Clostridiales bacterium KLE1615*	0.40
Fusicatenibacter saccharivorans	*Clostridiales bacterium KLE1615*	0.46
Roseburia faecis	*Clostridiales bacterium KLE1615*	0.23
Ruminococcus torques	*Clostridiales bacterium KLE1615*	0.34
Flavonifractor plautii	*Clostridium aldenense*	0.38
Flavonifractor plautii	*Clostridium asparagiforme*	0.27
Flavonifractor plautii	*Clostridium bolteae*	0.60
Flavonifractor plautii	*Clostridium citroniae*	0.30
Flavonifractor plautii	*Clostridium clostridioforme*	0.38
Ruminococcus torques	*Clostridium disporicum*	0.23
Flavonifractor plautii	*Clostridium innocuum*	0.32
Flavonifractor plautii	*Clostridium lavalense*	0.39
Parabacteroides distasonis	*Clostridium leptum*	0.22
Flavonifractor plautii	*Clostridium scindens*	0.28
Fusicatenibacter saccharivorans	*Clostridium sp AF36 4*	0.32
Eubacterium ramulus	*Clostridium sp AM22 11AC*	0.34
Fusicatenibacter saccharivorans	*Clostridium sp AM22 11AC*	0.32
Parabacteroides distasonis	*Clostridium sp CAG 1013*	0.21
Roseburia faecis	*Clostridium sp CAG 122*	0.21
Roseburia faecis	*Clostridium sp CAG 12237 41*	0.21
Ruminococcus torques	*Clostridium sp CAG 138*	0.20
Fusicatenibacter saccharivorans	*Clostridium sp CAG 167*	0.24
Ruminococcus torques	*Clostridium sp CAG 167*	0.21
Eubacterium ramulus	*Clostridium sp CAG 169*	0.32
Fusicatenibacter saccharivorans	*Clostridium sp CAG 217*	0.25
Ruminococcus torques	*Clostridium sp CAG 217*	0.21
Eubacterium ramulus	*Clostridium sp CAG 245*	0.21
Fusicatenibacter saccharivorans	*Clostridium sp CAG 245*	0.26
Ruminococcus torques	*Clostridium sp CAG 245*	0.27
Eubacterium ramulus	*Clostridium sp CAG 269*	0.28
Fusicatenibacter saccharivorans	*Clostridium sp CAG 269*	0.27
Ruminococcus torques	*Clostridium sp CAG 269*	0.25
Ruminococcus torques	*Clostridium sp CAG 302*	0.21
Flavonifractor plautii	*Clostridium sp CAG 505*	0.28
Parabacteroides distasonis	*Clostridium sp CAG 505*	0.20
Eubacterium ramulus	*Clostridium sp CAG 58*	0.26
Fusicatenibacter saccharivorans	*Clostridium sp CAG 58*	0.33
Ruminococcus torques	*Clostridium sp CAG 58*	0.20
Eubacterium ramulus	*Clostridium sp CAG 75*	0.23

Mucosal glycan degrader	First neighbor	Correlation strength
Fusicatenibacter saccharivorans	*Clostridium sp CAG 75*	0.27
Roseburia faecis	*Clostridium sp CAG 75*	0.20
Ruminococcus torques	*Clostridium sp CAG 75*	0.24
Flavonifractor plautii	*Clostridium spiroforme*	0.23
Roseburia intestinalis	*Clostridium spiroforme*	0.21
Flavonifractor plautii	*Clostridium symbiosum*	0.56
Eubacterium ramulus	*Collinsella aerofaciens*	0.49
Fusicatenibacter saccharivorans	*Collinsella aerofaciens*	0.44
Roseburia faecis	*Collinsella aerofaciens*	0.26
Ruminococcus torques	*Collinsella aerofaciens*	0.34
Flavonifractor plautii	*Collinsella intestinalis*	0.20
Eubacterium ramulus	*Coprococcus catus*	0.40
Fusicatenibacter saccharivorans	*Coprococcus catus*	0.45
Roseburia faecis	*Coprococcus catus*	0.23
Ruminococcus torques	*Coprococcus catus*	0.46
Eubacterium ramulus	*Coprococcus comes*	0.40
Fusicatenibacter saccharivorans	*Coprococcus comes*	0.48
Roseburia faecis	*Coprococcus comes*	0.30
Ruminococcus torques	*Coprococcus comes*	0.40
Eubacterium ramulus	*Desulfovibrionaceae bacterium*	0.23
Flavonifractor plautii	*Dialister invisus*	0.20
Flavonifractor plautii	*Dielma fastidiosa*	0.32
Eubacterium ramulus	*Dorea formicigenerans*	0.48
Fusicatenibacter saccharivorans	*Dorea formicigenerans*	0.50
Roseburia faecis	*Dorea formicigenerans*	0.21
Ruminococcus torques	*Dorea formicigenerans*	0.39
Eubacterium ramulus	*Dorea longicatena*	0.36
Fusicatenibacter saccharivorans	*Dorea longicatena*	0.51
Roseburia faecis	*Dorea longicatena*	0.29
Ruminococcus torques	*Dorea longicatena*	0.46
Flavonifractor plautii	*Dorea sp CAG 317*	0.28
Flavonifractor plautii	*Dysgonomonas gadei*	0.27
Roseburia intestinalis	*Dysgonomonas gadei*	0.24
Flavonifractor plautii	*Dysgonomonas mossii*	0.26
Roseburia intestinalis	*Dysgonomonas mossii*	0.23
Flavonifractor plautii	*Dysgonomonas sp 37 18*	0.27
Roseburia intestinalis	*Dysgonomonas sp 37 18*	0.24
Flavonifractor plautii	*Eggerthella lenta*	0.24
Fusicatenibacter saccharivorans	*Eggerthella lenta*	0.23
Fusicatenibacter saccharivorans	*Eggerthella sp CAG 209*	0.24
Roseburia faecis	*Eggerthella sp CAG 209*	0.20
Ruminococcus torques	*Eggerthella sp CAG 209*	0.25
Eubacterium ramulus	*Eggerthella sp CAG 298*	0.21
Fusicatenibacter saccharivorans	*Eggerthella sp CAG 298*	0.34
Flavonifractor plautii	*Eisenbergiella massiliensis*	0.43

Mucosal glycan degrader	First neighbor	Correlation strength
Flavonifractor plautii	*Erysipelatoclostridium ramosum*	0.37
Fusicatenibacter saccharivorans	*Eubacterium eligens*	0.41
Roseburia faecis	*Eubacterium eligens*	0.23
Eubacterium ramulus	*Eubacterium hallii*	0.39
Fusicatenibacter saccharivorans	*Eubacterium hallii*	0.54
Roseburia faecis	*Eubacterium hallii*	0.24
Ruminococcus torques	*Eubacterium hallii*	0.35
Fusicatenibacter saccharivorans	*Eubacterium ramulus*	0.41
Roseburia faecis	*Eubacterium ramulus*	0.22
Ruminococcus torques	*Eubacterium ramulus*	0.32
Fusicatenibacter saccharivorans	*Eubacterium rectale*	0.33
Roseburia faecis	*Eubacterium rectale*	0.23
Fusicatenibacter saccharivorans	*Eubacterium siraeum*	0.20
Fusicatenibacter saccharivorans	*Eubacterium sp 36 13*	0.27
Roseburia faecis	*Eubacterium sp 36 13*	0.21
Parabacteroides distasonis	*Eubacterium sp An11*	0.21
Fusicatenibacter saccharivorans	*Eubacterium sp CAG 161*	0.24
Ruminococcus torques	*Eubacterium sp CAG 161*	0.21
Fusicatenibacter saccharivorans	*Eubacterium ventriosum*	0.25
Ruminococcus torques	*Eubacterium ventriosum*	0.21
Roseburia faecis	*Faecalibacterium prausnitzii*	0.27
Roseburia intestinalis	*Faecalibacterium prausnitzii*	0.24
Fusicatenibacter saccharivorans	*Faecalibacterium sp CAG 74*	0.21
Ruminococcus torques	*Faecalibacterium sp CAG 74*	0.27
Fusicatenibacter saccharivorans	*Faecalibacterium sp CAG 74 58 120*	0.22
Ruminococcus torques	*Faecalibacterium sp CAG 74 58 120*	0.23
Fusicatenibacter saccharivorans	*Firmicutes bacterium AF16 15*	0.39
Roseburia faecis	*Firmicutes bacterium AF16 15*	0.29
Ruminococcus torques	*Firmicutes bacterium AF16 15*	0.26
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 103*	0.32
Ruminococcus torques	*Firmicutes bacterium CAG 103*	0.22
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 110*	0.21
Ruminococcus torques	*Firmicutes bacterium CAG 110*	0.34
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 114*	0.44
Ruminococcus torques	*Firmicutes bacterium CAG 114*	0.29
Ruminococcus torques	*Firmicutes bacterium CAG 124*	0.23
Ruminococcus torques	*Firmicutes bacterium CAG 137*	0.20
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 145*	0.25
Ruminococcus torques	*Firmicutes bacterium CAG 24053 14*	0.27
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 41*	0.44
Ruminococcus torques	*Firmicutes bacterium CAG 41*	0.26
Flavonifractor plautii	*Firmicutes bacterium CAG 466*	0.24
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 56*	0.37
Roseburia faecis	*Firmicutes bacterium CAG 56*	0.32
Ruminococcus torques	*Firmicutes bacterium CAG 56*	0.33

Mucosal glycan degrader	First neighbor	Correlation strength
Roseburia faecis	*Firmicutes bacterium CAG 65 45 313*	0.24
Fusicatenibacter saccharivorans	*Firmicutes bacterium CAG 83*	0.23
Roseburia faecis	*Firmicutes bacterium CAG 83*	0.22
Roseburia faecis	*Fusicatenibacter saccharivorans*	0.21
Ruminococcus torques	*Fusicatenibacter saccharivorans*	0.45
Roseburia faecis	*Gemmiger formicilis*	0.26
Ruminococcus torques	*Gemmiger formicilis*	0.46
Roseburia intestinalis	*Haemophilus parainfluenzae*	0.23
Ruminococcus torques	*Holdemanella biformis*	0.24
Roseburia faecis	*Lachnospira pectinoschiza*	0.23
Ruminococcus torques	*Lachnospira pectinoschiza*	0.20
Ruminococcus torques	*Lactococcus lactis*	0.21
Parabacteroides merdae	*Megasphaera sp DISK 18*	0.21
Parabacteroides merdae	*Megasphaera sp MJR8396C*	0.27
Ruminococcus torques	*Methanobrevibacter smithii*	0.24
Parabacteroides distasonis	*Odoribacter splanchnicus*	0.26
Parabacteroides merdae	*Odoribacter splanchnicus*	0.21
Roseburia faecis	*Oscillibacter sp 57 20*	0.21
Ruminococcus torques	*Oscillibacter sp CAG 241*	0.24
Roseburia intestinalis	*Proteus mirabilis*	0.27
Ruminococcus torques	*Roseburia sp CAG 471*	0.24
Ruminococcus torques	*Ruminococcaceae bacterium KLE1738*	0.29
Ruminococcus torques	*Ruminococcus bromii*	0.24
Ruminococcus torques	*Ruminococcus lactaris*	0.23
Ruminococcus torques	*Ruminococcus sp CAG 177*	0.22
Akkermansia muciniphila	*Bacteroides intestinalis*	0.21
Akkermansia muciniphila	*Butyricimonas virosa*	0.20
Akkermansia muciniphila	*Candidatus Gastranaerophilales bacterium HUM 10*	0.24
Akkermansia muciniphila	*Candidatus Gastranaerophilales bacterium HUM 11*	0.23
Akkermansia muciniphila	*Clostridium sp CAG 273*	0.24
Akkermansia muciniphila	*Clostridium sp CAG 451*	0.22
Akkermansia muciniphila	*Eubacterium siraeum*	0.22
Akkermansia muciniphila	*Firmicutes bacterium CAG 129*	0.25
Akkermansia muciniphila	*Firmicutes bacterium CAG 176*	0.25
Akkermansia muciniphila	*Firmicutes bacterium CAG 240*	0.21
Akkermansia muciniphila	*Firmicutes bacterium CAG 24053 14*	0.25
Akkermansia muciniphila	*Firmicutes bacterium CAG 95*	0.24
Akkermansia muciniphila	*Methanobrevibacter smithii*	0.23
Akkermansia muciniphila	*Ruminococcaceae bacterium D5*	0.23
Akkermansia muciniphila	*Ruminococcus bromii*	0.25
Akkermansia muciniphila	*Ruminococcus sp 37 24*	0.24
Akkermansia muciniphila	*Subdoligranulum sp*	0.21
Akkermansia muciniphila	*Subdoligranulum sp APC924 74*	0.26
Alistipes finegoldii	*Alistipes indistinctus*	0.24

Mucosal glycan degrader	**First neighbor**	**Correlation strength**
Alistipes finegoldii	*Alistipes obesi*	0.26
Alistipes finegoldii	*Alistipes putredinis*	0.26
Alistipes finegoldii	*Bilophila wadsworthia*	0.22
Alistipes finegoldii	*Firmicutes bacterium AF16 15*	0.23
Alistipes finegoldii	*Firmicutes bacterium CAG 103*	0.25
Alistipes finegoldii	*Firmicutes bacterium CAG 124*	0.23
Alistipes finegoldii	*Firmicutes bacterium CAG 83*	0.23
Alistipes finegoldii	*Lawsonibacter asaccharolyticus*	0.28
Alistipes finegoldii	*Odoribacter splanchnicus*	0.30
Alistipes finegoldii	*Ruminococcaceae bacterium D5*	0.21
Alistipes finegoldii	*Ruminococcus bicirculans*	0.21
Alistipes finegoldii	*Ruminococcus bromii*	0.29
Alistipes finegoldii	*Ruminococcus lactaris*	0.21
Alistipes finegoldii	*Subdoligranulum sp APC924 74*	0.22
Alistipes shahii	*Asaccharobacter celatus*	0.21
Alistipes shahii	*Bacteroides massiliensis*	0.26
Alistipes shahii	*Barnesiella intestinihominis*	0.42
Alistipes shahii	*Bilophila wadsworthia*	0.22
Alistipes shahii	*Blautia obeum*	0.25
Alistipes shahii	*Butyricimonas synergistica*	0.35
Alistipes shahii	*Butyricimonas virosa*	0.40
Alistipes shahii	*Butyrivibrio crossotus*	0.27
Alistipes shahii	*Candidatus Methanomassiliicoccus intestinalis*	0.31
Alistipes shahii	*Clostridiales Family XIII bacterium*	0.30
Alistipes shahii	*Clostridiales bacterium KLE1615*	0.25
Alistipes shahii	*Clostridium sp CAG 138*	0.37
Alistipes shahii	*Clostridium sp CAG 167*	0.31
Alistipes shahii	*Clostridium sp CAG 226*	0.27
Alistipes shahii	*Clostridium sp CAG 245*	0.21
Alistipes shahii	*Clostridium sp CAG 269*	0.24
Alistipes shahii	*Clostridium sp CAG 302*	0.32
Alistipes shahii	*Clostridium sp CAG 58*	0.20
Alistipes shahii	*Collinsella aerofaciens*	0.27
Alistipes shahii	*Coprobacter fastidiosus*	0.23
Alistipes shahii	*Coprococcus catus*	0.20
Alistipes shahii	*Coprococcus eutactus*	0.25
Alistipes shahii	*Cryptobacterium sp CAG 338*	0.21
Alistipes shahii	*Desulfovibrio fairfieldensis*	0.21
Alistipes shahii	*Desulfovibrio piger*	0.21
Alistipes shahii	*Dorea longicatena*	0.31
Alistipes shahii	*Eubacterium hallii*	0.22
Alistipes shahii	*Eubacterium siraeum*	0.34
Alistipes shahii	*Eubacterium sp 36 13*	0.22
Alistipes shahii	*Eubacterium sp CAG 115*	0.21

Mucosal glycan degrader	First neighbor	Correlation strength
Alistipes shahii	*Eubacterium sp CAG 251*	0.21
Alistipes shahii	*Faecalibacterium sp CAG 74*	0.37
Alistipes shahii	*Faecalibacterium sp CAG 74 58 120*	0.35
Alistipes shahii	*Firmicutes bacterium CAG 110*	0.38
Alistipes shahii	*Firmicutes bacterium CAG 124*	0.31
Alistipes shahii	*Firmicutes bacterium CAG 129*	0.40
Alistipes shahii	*Firmicutes bacterium CAG 137*	0.29
Alistipes shahii	*Firmicutes bacterium CAG 176*	0.25
Alistipes shahii	*Firmicutes bacterium CAG 240*	0.34
Alistipes shahii	*Firmicutes bacterium CAG 24053 14*	0.40
Alistipes shahii	*Firmicutes bacterium CAG 272*	0.23
Alistipes shahii	*Firmicutes bacterium CAG 321*	0.23
Alistipes shahii	*Firmicutes bacterium CAG 95*	0.27
Alistipes shahii	*Gemmiger formicilis*	0.35
Alistipes shahii	*Lentisphaerae bacterium*	0.26
Alistipes shahii	*Methanobrevibacter smithii*	0.38
Alistipes shahii	*Odoribacter splanchnicus*	0.44
Alistipes shahii	*Oscillibacter sp 57 20*	0.32
Alistipes shahii	*Oscillibacter sp CAG 241*	0.35
Alistipes shahii	*Parabacteroides goldsteinii*	0.34
Alistipes shahii	*Paraprevotella xylaniphila*	0.21
Alistipes shahii	*Phascolarctobacterium faecium*	0.24
Alistipes shahii	*Phascolarctobacterium succinatutens*	0.21
Alistipes shahii	*Porphyromonas sp 31 2*	0.23
Alistipes shahii	*Roseburia sp CAG 309*	0.26
Alistipes shahii	*Ruminococcaceae bacterium D5*	0.28
Alistipes shahii	*Ruminococcus bromii*	0.29
Alistipes shahii	*Ruminococcus sp CAG 177*	0.27
Alistipes shahii	*Sanguibacteroides justesenii*	0.25
Alistipes shahii	*Senegalimassilia anaerobia*	0.23
Alistipes shahii	*Subdoligranulum sp*	0.33
Alistipes shahii	*Subdoligranulum sp APC924 74*	0.33
Alistipes shahii	*Sutterella wadsworthensis*	0.21
Alistipes shahii	*Victivallales bacterium CCUG 44730*	0.30
Alistipes shahii	*Victivallis vadensis*	0.29
Bacteroides caccae	*Bacteroides finegoldii*	0.25
Bacteroides caccae	*Prevotella corporis*	0.24
Bacteroides caccae	*Ruminococcaceae bacterium D16*	0.21
Bacteroides cellulosilyticus	*Butyricimonas virosa*	0.25
Bacteroides cellulosilyticus	*Firmicutes bacterium CAG 137*	0.20
Bacteroides cellulosilyticus	*Lawsonibacter asaccharolyticus*	0.26
Bacteroides cellulosilyticus	*Parabacteroides johnsonii*	0.23
Bacteroides dorei	*Blautia sp AF19 10LB*	0.23
Bacteroides dorei	*Clostridium sp AF36 4*	0.25
Bacteroides dorei	*Clostridium sp CAG 1013*	0.21

Mucosal glycan degrader	First neighbor	Correlation strength
Bacteroides dorei	*Clostridium sp CAG 122*	0.21
Bacteroides dorei	*Clostridium sp CAG 12237 41*	0.21
Bacteroides dorei	*Eubacterium eligens*	0.20
Bacteroides dorei	*Eubacterium siraeum*	0.20
Bacteroides dorei	*Eubacterium sp 36 13*	0.21
Bacteroides dorei	*Eubacterium sp An11*	0.21
Bacteroides dorei	*Intestinimonas butyriciproducens*	0.22
Bacteroides dorei	*Lawsonibacter asaccharolyticus*	0.21
Bacteroides dorei	*Monoglobus pectinilyticus*	0.26
Bacteroides fragilis	*Bacteroides intestinalis*	0.23
Bacteroides fragilis	*Bacteroides nordii*	0.23
Bacteroides fragilis	*Clostridium sp CAG 58*	0.21
Bacteroides fragilis	*Dorea sp CAG 317*	0.24
Bacteroides fragilis	*Prevotella corporis*	0.21
Bacteroides fragilis	*Tyzzerella nexilis*	0.20
Bacteroides ovatus	*Bacteroides sp D2*	0.25
Bacteroides ovatus	*Cutibacterium acnes*	0.22
Bacteroides ovatus	*Megasphaera sp MJR8396C*	0.21
Bacteroides ovatus	*Ruminococcaceae bacterium D16*	0.21
Bacteroides stercoris	*Cutibacterium acnes*	0.29
Bacteroides stercoris	*Megasphaera sp DISK 18*	0.22
Bacteroides stercoris	*Megasphaera sp MJR8396C*	0.29
Bacteroides thetaiotaomicron	*Clostridium sp CAG 122*	0.21
Bacteroides uniformis	*Clostridium sp CAG 1013*	0.21
Bacteroides uniformis	*Clostridium sp CAG 505*	0.23
Bacteroides uniformis	*Eubacterium sp An11*	0.21
Bacteroides uniformis	*Lawsonibacter asaccharolyticus*	0.30
Bacteroides uniformis	*Odoribacter splanchnicus*	0.29
Bacteroides uniformis	*Ruminococcaceae bacterium KLE1738*	0.24
Bacteroides vulgatus	*Megasphaera sp MJR8396C*	0.20
Bacteroides xylanisolvens	*Odoribacter splanchnicus*	0.31
Eubacterium ramulus	*Eubacterium sp OM08 24*	0.21
Eubacterium ramulus	*Eubacterium ventriosum*	0.26
Eubacterium ramulus	*Firmicutes bacterium AF16 15*	0.28
Eubacterium ramulus	*Firmicutes bacterium CAG 110*	0.23
Eubacterium ramulus	*Firmicutes bacterium CAG 114*	0.27
Eubacterium ramulus	*Firmicutes bacterium CAG 124*	0.26
Eubacterium ramulus	*Firmicutes bacterium CAG 240*	0.23
Eubacterium ramulus	*Firmicutes bacterium CAG 24053 14*	0.27
Eubacterium ramulus	*Firmicutes bacterium CAG 41*	0.30
Eubacterium ramulus	*Firmicutes bacterium CAG 56*	0.29
Eubacterium ramulus	*Firmicutes bacterium CAG 65 45 313*	0.21
Eubacterium ramulus	*Firmicutes bacterium CAG 83*	0.24
Eubacterium ramulus	*Gemmiger formicilis*	0.39
Eubacterium ramulus	*Holdemanella biformis*	0.26

Mucosal glycan degrader	First neighbor	Correlation strength
Eubacterium ramulus	*Lachnospira pectinoschiza*	0.22
Eubacterium ramulus	*Lachnospiraceae bacterium OF09 6*	0.24
Eubacterium ramulus	*Lawsonibacter asaccharolyticus*	0.22
Eubacterium ramulus	*Odoribacter splanchnicus*	0.27
Eubacterium ramulus	*Oscillibacter sp CAG 241*	0.21
Eubacterium ramulus	*Proteobacteria bacterium CAG 495*	0.20
Eubacterium ramulus	*Roseburia hominis*	0.20
Eubacterium ramulus	*Roseburia inulinivorans*	0.27
Eubacterium ramulus	*Ruminococcaceae bacterium KLE1738*	0.26
Eubacterium ramulus	*Slackia isoflavoniconvertens*	0.23
Eubacterium ramulus	*Streptococcus australis*	0.23
Eubacterium ramulus	*Subdoligranulum sp*	0.30
Eubacterium ramulus	*Subdoligranulum sp APC924 74*	0.31
Eubacterium ramulus	*Tyzzerella sp*	0.25
Flavonifractor plautii	*Hungatella hathewayi*	0.50
Flavonifractor plautii	*Intestinimonas butyriciproducens*	0.26
Flavonifractor plautii	*Proteus mirabilis*	0.29
Flavonifractor plautii	*Ruminococcaceae bacterium D16*	0.25
Flavonifractor plautii	*Ruminococcus gnavus*	0.49
Flavonifractor plautii	*Ruthenibacterium lactatiformans*	0.32
Flavonifractor plautii	*Sellimonas intestinalis*	0.37
Flavonifractor plautii	*Tyzzerella nexilis*	0.27
Fusicatenibacter saccharivorans	*Gemmiger formicilis*	0.44
Fusicatenibacter saccharivorans	*Gordonibacter pamelaeae*	0.30
Fusicatenibacter saccharivorans	*Lachnospira pectinoschiza*	0.21
Fusicatenibacter saccharivorans	*Lawsonibacter asaccharolyticus*	0.35
Fusicatenibacter saccharivorans	*Methanobrevibacter smithii*	0.22
Fusicatenibacter saccharivorans	*Monoglobus pectinilyticus*	0.27
Fusicatenibacter saccharivorans	*Odoribacter splanchnicus*	0.20
Fusicatenibacter saccharivorans	*Oscillibacter sp CAG 241*	0.20
Fusicatenibacter saccharivorans	*Roseburia hominis*	0.22
Fusicatenibacter saccharivorans	*Roseburia sp CAG 471*	0.25
Fusicatenibacter saccharivorans	*Ruminococcaceae bacterium KLE1738*	0.43
Fusicatenibacter saccharivorans	*Ruminococcus bicirculans*	0.29
Fusicatenibacter saccharivorans	*Ruminococcus bromii*	0.29
Fusicatenibacter saccharivorans	*Ruminococcus callidus*	0.22
Fusicatenibacter saccharivorans	*Ruminococcus lactaris*	0.37
Fusicatenibacter saccharivorans	*Ruminococcus sp 37 24*	0.21
Fusicatenibacter saccharivorans	*Slackia isoflavoniconvertens*	0.20
Fusicatenibacter saccharivorans	*Streptococcus australis*	0.21
Fusicatenibacter saccharivorans	*Streptococcus salivarius*	0.27
Fusicatenibacter saccharivorans	*Streptococcus sp A12*	0.20
Fusicatenibacter saccharivorans	*Subdoligranulum sp*	0.44
Fusicatenibacter saccharivorans	*Subdoligranulum sp APC924 74*	0.34
Fusicatenibacter saccharivorans	*Tyzzerella sp*	0.24

Mucosal glycan degrader	**First neighbor**	**Correlation strength**
Parabacteroides distasonis	*Phascolarctobacterium faecium*	0.20
Parabacteroides distasonis	*Ruthenibacterium lactatiformans*	0.20
Roseburia faecis	*Roseburia hominis*	0.25
Roseburia faecis	*Roseburia inulinivorans*	0.24
Roseburia faecis	*Roseburia sp CAG 197*	0.25
Roseburia faecis	*Ruminococcus bicirculans*	0.25
Roseburia faecis	*Subdoligranulum sp APC924 74*	0.26
Roseburia intestinalis	*Roseburia inulinivorans*	0.27
Roseburia intestinalis	*Veillonella dispar*	0.27
Ruminococcus torques	*Streptococcus salivarius*	0.24
Ruminococcus torques	*Subdoligranulum sp*	0.46
Ruminococcus torques	*Subdoligranulum sp APC924 74*	0.32

Table S3 P-values, adjusted R^2 values and median adjusted R^2 values of significant (p-value < 0.05) abundance-transcription linear models

Species	adjusted R^2	p-value	EC	Species median R^2 +/- sd
Akkermansia muciniphila	0.42	1.3E-25	3.2.1.18	0.42 +/- NA
Alistipes finegoldii	0.31	7.7E-36	3.2.1.96	0.25 +/- 0.09
Alistipes finegoldii	0.30	1.3E-36	3.2.1.22	0.25 +/- 0.09
Alistipes finegoldii	0.20	2.5E-26	3.2.1.23	0.25 +/- 0.09
Alistipes finegoldii	0.12	1.5E-14	3.2.1.52	0.25 +/- 0.09
Alistipes shahii	0.26	2.1E-29	3.2.1.96	0.1 +/- 0.11
Alistipes shahii	0.10	1.8E-12	3.2.1.22	0.1 +/- 0.11
Alistipes shahii	0.05	1.9E-03	3.2.1.50	0.1 +/- 0.11
Bacteroides caccae	0.11	1.1E-09	3.2.1.51	0.09 +/- 0.03
Bacteroides caccae	0.09	5.6E-12	3.2.1.23	0.09 +/- 0.03
Bacteroides caccae	0.09	4.6E-11	3.2.1.52	0.09 +/- 0.03
Bacteroides caccae	0.07	2.8E-04	3.2.1.50	0.09 +/- 0.03
Bacteroides caccae	0.04	1.6E-05	3.2.1.22	0.09 +/- 0.03
Bacteroides cellulosilyticus	0.28	8.4E-19	3.2.1.165	0.22 +/- 0.07
Bacteroides cellulosilyticus	0.25	2.3E-14	3.2.1.18	0.22 +/- 0.07
Bacteroides cellulosilyticus	0.18	5.5E-23	3.2.1.23	0.22 +/- 0.07
Bacteroides cellulosilyticus	0.13	3.0E-15	3.2.1.22	0.22 +/- 0.07
Bacteroides dorei	0.51	8.2E-73	3.2.1.52	0.46 +/- 0.12
Bacteroides dorei	0.46	3.0E-67	3.2.1.23	0.46 +/- 0.12
Bacteroides dorei	0.29	9.0E-35	3.2.1.22	0.46 +/- 0.12
Bacteroides fragilis	0.46	4.2E-13	3.2.1.49	0.21 +/- 0.16
Bacteroides fragilis	0.43	4.4E-26	3.2.1.18	0.21 +/- 0.16
Bacteroides fragilis	0.42	7.4E-30	3.2.1.165	0.21 +/- 0.16
Bacteroides fragilis	0.21	1.2E-25	3.2.1.52	0.21 +/- 0.16
Bacteroides fragilis	0.21	4.3E-27	3.2.1.23	0.21 +/- 0.16
Bacteroides fragilis	0.17	1.9E-20	3.2.1.22	0.21 +/- 0.16
Bacteroides fragilis	0.12	1.5E-10	3.2.1.51	0.21 +/- 0.16
Bacteroides fragilis	0.04	4.9E-03	3.2.1.50	0.21 +/- 0.16
Bacteroides ovatus	0.09	3.3E-05	3.2.1.50	0.07 +/- 0.01
Bacteroides ovatus	0.07	3.3E-09	3.2.1.22	0.07 +/- 0.01
Bacteroides ovatus	0.07	1.8E-09	3.2.1.23	0.07 +/- 0.01
Bacteroides stercoris	0.53	1.0E-15	3.2.1.49	0.33 +/- 0.18
Bacteroides stercoris	0.45	1.3E-61	3.2.1.52	0.33 +/- 0.18
Bacteroides stercoris	0.21	4.1E-27	3.2.1.23	0.33 +/- 0.18
Bacteroides stercoris	0.17	1.0E-08	3.2.1.50	0.33 +/- 0.18
Bacteroides thetaiotaomicron	0.17	5.2E-11	3.2.1.165	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.16	7.8E-18	3.2.1.96	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.14	5.0E-08	3.2.1.18	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.12	1.1E-10	3.2.1.51	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.09	1.6E-10	3.2.1.22	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.08	1.2E-09	3.2.1.52	0.12 +/- 0.04
Bacteroides thetaiotaomicron	0.07	2.9E-09	3.2.1.23	0.12 +/- 0.04
Bacteroides uniformis	0.34	7.6E-23	3.2.1.165	0.27 +/- 0.1

Species	adjusted R^2	p-value	EC	Species median R^2 +/- sd
Bacteroides uniformis	0.31	1.3E-08	3.2.1.49	0.27 +/- 0.1
Bacteroides uniformis	0.29	1.7E-38	3.2.1.23	0.27 +/- 0.1
Bacteroides uniformis	0.27	6.6E-23	3.2.1.51	0.27 +/- 0.1
Bacteroides uniformis	0.22	1.2E-26	3.2.1.52	0.27 +/- 0.1
Bacteroides uniformis	0.20	4.5E-24	3.2.1.22	0.27 +/- 0.1
Bacteroides uniformis	0.06	2.4E-07	3.2.1.96	0.27 +/- 0.1
Bacteroides vulgatus	0.33	3.3E-42	3.2.1.52	0.19 +/- 0.13
Bacteroides vulgatus	0.30	1.9E-36	3.2.1.22	0.19 +/- 0.13
Bacteroides vulgatus	0.25	3.0E-32	3.2.1.23	0.19 +/- 0.13
Bacteroides vulgatus	0.13	3.7E-07	3.2.1.50	0.19 +/- 0.13
Bacteroides vulgatus	0.04	3.3E-04	3.2.1.51	0.19 +/- 0.13
Bacteroides vulgatus	0.02	2.9E-02	3.2.1.18	0.19 +/- 0.13
Bacteroides xylanisolvens	0.10	2.5E-12	3.2.1.22	0.06 +/- 0.04
Bacteroides xylanisolvens	0.06	6.2E-09	3.2.1.23	0.06 +/- 0.04
Bacteroides xylanisolvens	0.03	1.7E-02	3.2.1.50	0.06 +/- 0.04
Flavonifractor plautii	0.20	2.2E-11	3.2.1.18	0.2 +/- NA
Fusicatenibacter saccharivorans	0.06	1.3E-04	3.2.1.165	0.06 +/- NA
Parabacteroides distasonis	0.04	7.5E-04	3.2.1.165	0.02 +/- 0.01
Parabacteroides distasonis	0.03	1.9E-04	3.2.1.52	0.02 +/- 0.01
Parabacteroides distasonis	0.02	1.0E-03	3.2.1.22	0.02 +/- 0.01
Parabacteroides distasonis	0.02	2.8E-03	3.2.1.23	0.02 +/- 0.01
Parabacteroides merdae	0.16	2.4E-20	3.2.1.23	0.13 +/- 0.05
Parabacteroides merdae	0.09	1.6E-11	3.2.1.22	0.13 +/- 0.05
Roseburia faecis	0.15	6.9E-18	3.2.1.22	0.13 +/- 0.03
Roseburia faecis	0.11	2.3E-14	3.2.1.23	0.13 +/- 0.03
Roseburia intestinalis	0.25	6.2E-30	3.2.1.22	0.23 +/- 0.03
Roseburia intestinalis	0.23	3.3E-19	3.2.1.51	0.23 +/- 0.03
Roseburia intestinalis	0.18	2.0E-23	3.2.1.23	0.23 +/- 0.03
Ruminococcus torques	0.03	5.6E-05	3.2.1.23	0.03 +/- NA

Table S4 Significant (p-value < 0.05) linear model coefficients of response-predictor networks

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides fragilis	*Bacteroides cellulosilyticus*	3.86	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides cellulosilyticus	*Bacteroides fragilis*	-4.01	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides uniformis	*Bacteroides fragilis*	-5.36	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides uniformis	*Bacteroides thetaiotaomicron*	-4.38	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Parabacteroides distasonis	*Bacteroides thetaiotaomicron*	-3.64	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides fragilis	*Bacteroides uniformis*	-5.12	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides uniformis	*Parabacteroides distasonis*	-6.28	3.2.1.165: Exo-1,4-beta-D-glucosaminidase
Bacteroides vulgatus	*Bacteroides cellulosilyticus*	-4.77	3.2.1.18: Exo-alpha-sialidase
Bacteroides cellulosilyticus	*Bacteroides fragilis*	-3.76	3.2.1.18: Exo-alpha-sialidase
Bacteroides thetaiotaomicron	*Bacteroides fragilis*	-7.17	3.2.1.18: Exo-alpha-sialidase
Flavonifractor plautii	*Bacteroides fragilis*	6.23	3.2.1.18: Exo-alpha-sialidase
Akkermansia muciniphila	*Bacteroides thetaiotaomicron*	-5.40	3.2.1.18: Exo-alpha-sialidase
Bacteroides fragilis	*Bacteroides thetaiotaomicron*	-5.19	3.2.1.18: Exo-alpha-sialidase
Bacteroides thetaiotaomicron	*Bacteroides vulgatus*	-4.55	3.2.1.18: Exo-alpha-sialidase
Flavonifractor plautii	*Bacteroides vulgatus*	-11.58	3.2.1.18: Exo-alpha-sialidase
Akkermansia muciniphila	*Flavonifractor plautii*	-3.37	3.2.1.18: Exo-alpha-sialidase
Bacteroides fragilis	*Flavonifractor plautii*	-5.97	3.2.1.18: Exo-alpha-sialidase
Alistipes finegoldii	*Alistipes shahii*	-2.77	3.2.1.22: Alpha-galactosidase
Bacteroides caccae	*Alistipes shahii*	2.65	3.2.1.22: Alpha-galactosidase
Bacteroides vulgatus	*Bacteroides caccae*	-0.43	3.2.1.22: Alpha-galactosidase
Roseburia intestinalis	*Bacteroides caccae*	0.51	3.2.1.22: Alpha-galactosidase
Parabacteroides distasonis	*Bacteroides cellulosilyticus*	-2.44	3.2.1.22: Alpha-galactosidase
Roseburia intestinalis	*Bacteroides cellulosilyticus*	-2.01	3.2.1.22: Alpha-galactosidase
Bacteroides uniformis	*Bacteroides dorei*	-5.08	3.2.1.22: Alpha-galactosidase

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides xylanisolvens	Bacteroides dorei	2.65	3.2.1.22: Alpha-galactosidase
Alistipes shahii	Bacteroides ovatus	2.02	3.2.1.22: Alpha-galactosidase
Bacteroides dorei	Bacteroides ovatus	-1.04	3.2.1.22: Alpha-galactosidase
Bacteroides vulgatus	Bacteroides ovatus	-0.97	3.2.1.22: Alpha-galactosidase
Bacteroides xylanisolvens	Bacteroides ovatus	1.42	3.2.1.22: Alpha-galactosidase
Parabacteroides merdae	Bacteroides ovatus	-1.72	3.2.1.22: Alpha-galactosidase
Bacteroides cellulosilyticus	Bacteroides thetaiotaomicron	-0.78	3.2.1.22: Alpha-galactosidase
Bacteroides ovatus	Bacteroides thetaiotaomicron	-0.73	3.2.1.22: Alpha-galactosidase
Alistipes finegoldii	Bacteroides uniformis	-3.72	3.2.1.22: Alpha-galactosidase
Bacteroides caccae	Bacteroides uniformis	-2.24	3.2.1.22: Alpha-galactosidase
Parabacteroides distasonis	Bacteroides uniformis	-3.36	3.2.1.22: Alpha-galactosidase
Alistipes finegoldii	Bacteroides vulgatus	-4.46	3.2.1.22: Alpha-galactosidase
Bacteroides thetaiotaomicron	Bacteroides vulgatus	-2.42	3.2.1.22: Alpha-galactosidase
Bacteroides uniformis	Bacteroides vulgatus	-6.65	3.2.1.22: Alpha-galactosidase
Parabacteroides merdae	Bacteroides vulgatus	-3.18	3.2.1.22: Alpha-galactosidase
Bacteroides dorei	Bacteroides xylanisolvens	-0.39	3.2.1.22: Alpha-galactosidase
Bacteroides ovatus	Bacteroides xylanisolvens	0.55	3.2.1.22: Alpha-galactosidase
Bacteroides thetaiotaomicron	Bacteroides xylanisolvens	0.44	3.2.1.22: Alpha-galactosidase
Bacteroides vulgatus	Bacteroides xylanisolvens	-0.55	3.2.1.22: Alpha-galactosidase
Parabacteroides distasonis	Bacteroides xylanisolvens	-1.08	3.2.1.22: Alpha-galactosidase
Parabacteroides merdae	Bacteroides xylanisolvens	-0.57	3.2.1.22: Alpha-galactosidase
Bacteroides fragilis	Parabacteroides distasonis	-0.54	3.2.1.22: Alpha-galactosidase
Alistipes finegoldii	Parabacteroides merdae	-1.13	3.2.1.22: Alpha-galactosidase
Bacteroides caccae	Parabacteroides merdae	-1.08	3.2.1.22: Alpha-galactosidase
Bacteroides fragilis	Parabacteroides merdae	-0.89	3.2.1.22: Alpha-galactosidase
Bacteroides dorei	Roseburia intestinalis	-2.09	3.2.1.22: Alpha-galactosidase

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides fragilis	Roseburia intestinalis	-1.40	3.2.1.22: Alpha-galactosidase
Bacteroides vulgatus	Roseburia intestinalis	-2.95	3.2.1.22: Alpha-galactosidase
Roseburia faecis	Roseburia intestinalis	-1.83	3.2.1.22: Alpha-galactosidase
Bacteroides dorei	Alistipes finegoldii	-1.06	3.2.1.23: Beta-galactosidase
Bacteroides fragilis	Alistipes finegoldii	-1.62	3.2.1.23: Beta-galactosidase
Bacteroides stercoris	Alistipes finegoldii	-1.25	3.2.1.23: Beta-galactosidase
Bacteroides thetaiotaomicron	Alistipes finegoldii	1.58	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Alistipes finegoldii	-1.55	3.2.1.23: Beta-galactosidase
Bacteroides xylanisolvens	Alistipes finegoldii	1.70	3.2.1.23: Beta-galactosidase
Bacteroides ovatus	Bacteroides caccae	-1.55	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Bacteroides caccae	-1.12	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Bacteroides dorei	-5.58	3.2.1.23: Beta-galactosidase
Bacteroides vulgatus	Bacteroides dorei	-2.57	3.2.1.23: Beta-galactosidase
Parabacteroides distasonis	Bacteroides dorei	-3.51	3.2.1.23: Beta-galactosidase
Roseburia faecis	Bacteroides dorei	2.39	3.2.1.23: Beta-galactosidase
Bacteroides thetaiotaomicron	Bacteroides fragilis	-1.45	3.2.1.23: Beta-galactosidase
Bacteroides vulgatus	Bacteroides fragilis	-2.08	3.2.1.23: Beta-galactosidase
Bacteroides dorei	Bacteroides ovatus	-0.23	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Bacteroides ovatus	-0.40	3.2.1.23: Beta-galactosidase
Bacteroides vulgatus	Bacteroides ovatus	-0.54	3.2.1.23: Beta-galactosidase
Bacteroides xylanisolvens	Bacteroides ovatus	0.29	3.2.1.23: Beta-galactosidase
Parabacteroides distasonis	Bacteroides ovatus	-0.31	3.2.1.23: Beta-galactosidase
Parabacteroides merdae	Bacteroides ovatus	-0.32	3.2.1.23: Beta-galactosidase
Bacteroides caccae	Bacteroides stercoris	-2.60	3.2.1.23: Beta-galactosidase
Bacteroides dorei	Bacteroides stercoris	-1.35	3.2.1.23: Beta-galactosidase
Bacteroides ovatus	Bacteroides stercoris	2.55	3.2.1.23: Beta-galactosidase

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides xylanisolvens	Bacteroides stercoris	1.98	3.2.1.23: Beta-galactosidase
Roseburia intestinalis	Bacteroides stercoris	1.71	3.2.1.23: Beta-galactosidase
Ruminococcus torques	Bacteroides thetaiotaomicron	0.65	3.2.1.23: Beta-galactosidase
Bacteroides cellulosilyticus	Bacteroides uniformis	-3.56	3.2.1.23: Beta-galactosidase
Bacteroides fragilis	Bacteroides uniformis	-2.84	3.2.1.23: Beta-galactosidase
Roseburia faecis	Bacteroides uniformis	-2.29	3.2.1.23: Beta-galactosidase
Alistipes finegoldii	Bacteroides vulgatus	-3.09	3.2.1.23: Beta-galactosidase
Bacteroides thetaiotaomicron	Bacteroides vulgatus	-2.93	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Bacteroides vulgatus	-4.94	3.2.1.23: Beta-galactosidase
Parabacteroides merdae	Bacteroides vulgatus	-2.70	3.2.1.23: Beta-galactosidase
Bacteroides cellulosilyticus	Bacteroides xylanisolvens	-0.27	3.2.1.23: Beta-galactosidase
Bacteroides thetaiotaomicron	Bacteroides xylanisolvens	0.28	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Bacteroides xylanisolvens	-0.27	3.2.1.23: Beta-galactosidase
Ruminococcus torques	Bacteroides xylanisolvens	0.58	3.2.1.23: Beta-galactosidase
Bacteroides fragilis	Parabacteroides merdae	-1.01	3.2.1.23: Beta-galactosidase
Bacteroides uniformis	Parabacteroides merdae	-1.33	3.2.1.23: Beta-galactosidase
Bacteroides dorei	Roseburia faecis	-0.79	3.2.1.23: Beta-galactosidase
Bacteroides dorei	Roseburia intestinalis	-1.61	3.2.1.23: Beta-galactosidase
Bacteroides vulgatus	Roseburia intestinalis	-2.05	3.2.1.23: Beta-galactosidase
Bacteroides stercoris	Bacteroides fragilis	-7.42	3.2.1.49: Alpha-N-acetylgalactosaminidase
Bacteroides fragilis	Bacteroides uniformis	-6.82	3.2.1.49: Alpha-N-acetylgalactosaminidase
Bacteroides stercoris	Bacteroides uniformis	-9.85	3.2.1.49: Alpha-N-acetylgalactosaminidase
Bacteroides caccae	Alistipes shahii	6.88	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides fragilis	Alistipes shahii	6.83	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides ovatus	Alistipes shahii	-7.25	3.2.1.50: Alpha-N-acetylglucosaminidase

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides vulgatus	Bacteroides caccae	-3.76	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides caccae	Bacteroides fragilis	-3.72	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides ovatus	Bacteroides stercoris	4.67	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides vulgatus	Bacteroides stercoris	-7.00	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides caccae	Bacteroides vulgatus	-5.94	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides fragilis	Bacteroides vulgatus	-5.54	3.2.1.50: Alpha-N-acetylglucosaminidase
Bacteroides uniformis	Bacteroides caccae	-4.49	3.2.1.51: Alpha-L-fucosidase
Roseburia intestinalis	Bacteroides caccae	-2.73	3.2.1.51: Alpha-L-fucosidase
Bacteroides uniformis	Bacteroides fragilis	-1.95	3.2.1.51: Alpha-L-fucosidase
Bacteroides uniformis	Bacteroides thetaiotaomicron	-6.18	3.2.1.51: Alpha-L-fucosidase
Bacteroides caccae	Bacteroides uniformis	-4.75	3.2.1.51: Alpha-L-fucosidase
Bacteroides fragilis	Bacteroides vulgatus	-4.78	3.2.1.51: Alpha-L-fucosidase
Bacteroides thetaiotaomicron	Bacteroides vulgatus	-3.61	3.2.1.51: Alpha-L-fucosidase
Bacteroides uniformis	Bacteroides vulgatus	-4.61	3.2.1.51: Alpha-L-fucosidase
Bacteroides uniformis	Roseburia intestinalis	-5.24	3.2.1.51: Alpha-L-fucosidase
Bacteroides vulgatus	Roseburia intestinalis	-1.67	3.2.1.51: Alpha-L-fucosidase
Bacteroides uniformis	Alistipes finegoldii	-1.43	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides uniformis	Bacteroides caccae	-1.51	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides caccae	Bacteroides dorei	-2.81	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides stercoris	Bacteroides dorei	-2.20	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides uniformis	Bacteroides dorei	-3.98	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides vulgatus	Bacteroides dorei	-2.87	3.2.1.52: Beta-N-acetylhexosaminidase
Alistipes finegoldii	Bacteroides fragilis	1.41	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides thetaiotaomicron	Bacteroides fragilis	-2.04	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides uniformis	Bacteroides fragilis	-2.28	3.2.1.52: Beta-N-acetylhexosaminidase

Predictor species	Response species	Linear model coefficient	Enzyme
Bacteroides vulgatus	Bacteroides fragilis	-2.34	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides dorei	Bacteroides stercoris	-1.78	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides vulgatus	Bacteroides stercoris	-2.66	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides caccae	Bacteroides thetaiotaomicron	1.30	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides uniformis	Bacteroides thetaiotaomicron	-0.93	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides fragilis	Bacteroides uniformis	-1.59	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides stercoris	Bacteroides uniformis	-1.23	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides vulgatus	Bacteroides uniformis	-1.94	3.2.1.52: Beta-N-acetylhexosaminidase
Alistipes finegoldii	Bacteroides vulgatus	-2.83	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides caccae	Bacteroides vulgatus	-2.52	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides dorei	Bacteroides vulgatus	-4.14	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides stercoris	Bacteroides vulgatus	-1.60	3.2.1.52: Beta-N-acetylhexosaminidase
Alistipes finegoldii	Parabacteroides distasonis	-1.03	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides uniformis	Parabacteroides distasonis	-1.17	3.2.1.52: Beta-N-acetylhexosaminidase
Bacteroides vulgatus	Parabacteroides distasonis	-1.20	3.2.1.52: Beta-N-acetylhexosaminidase
Alistipes shahii	Alistipes finegoldii	-22.01	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Bacteroides uniformis	Alistipes finegoldii	-3.76	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Alistipes finegoldii	Alistipes shahii	-8.61	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Alistipes finegoldii	Bacteroides thetaiotaomicron	-6.55	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Alistipes shahii	Bacteroides thetaiotaomicron	-6.56	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Bacteroides uniformis	Bacteroides thetaiotaomicron	-3.35	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Alistipes finegoldii	Bacteroides uniformis	-4.65	3.2.1.96: Endo-beta-N-Acetylglucosaminidase
Eubacterium ramulus	Ruminococcus torques	-41.43	3.2.1.97: Endo-alpha-N-acetylgalactosaminidase