Characterization of Dietary and Genetic Influences on the Gastrointestinal Microbiota

Dissertation to obtain the doctoral degree of Natural Sciences

(Dr. rer. nat.)

Faculty of Natural Sciences

University of Hohenheim

Institute of Nutritional Science

submitted by

Alena Marie Bubeck née Fricker

from Stuttgart

2023

Faculty Dean:	Prof. Dr. Uwe Beifuß
1st Reviewer, Examiner:	Prof. Dr. W. Florian Fricke Institute of Nutritional Science
2nd Reviewer, Examiner:	Prof. Dr. Thomas Kufer Institute of Nutritional Science
3rd Examiner:	Jun.-Prof. Dr. Amélia Camarinha Silva Institute of Animal Science

Submitted on:	January 18, 2023
Oral examination:	May 23, 2023

This work was accepted by the Faculty of Natural Sciences at the University of Hohenheim on 23rd of January 2023 for Obtaining the Doctoral Degree of Natural Science.

Preamble

Substantial parts of this thesis have been submitted for publication. Apart from the advice of my supervisor Prof. Dr. W. Florian Fricke, this thesis is the result of my own work.

Submitted for publication in relation to the present work

A.M. Bubeck, P. Urbain, C. Horn, A.S. Jung, L. Ferrari, H.K. Ruple, D. Podlesny, S. Zorn, J. Laupsa-Borge, C. Jensen, I. Lindseth, G. Arslan Lied, J. Dierkes, G. Mellgren, H. Bertz, S. Matysik, S. Krautbauer, G. Liebisch, H.F. Schoett, S.N. Dankel, W.F. Fricke. (2022). *Microbial cholesterol conversion and converter type-dependent LDL-C response to low-carbohydrate high-fat diet*. Under revision at *Cell Press Community*

In addition to the work presented in this thesis, the following research articles have been published with me as an author during my doctorate. These publications are independent of the work presented in this thesis. My contributions to these publications are listed in the Appendix.

Additional publications

A.M. Bubeck, L. Preiss, A. Jung, E. Dörner, D. Podlesny, M. Kulis, C. Maddox, C. Arze, C. Zörb, N. Merkt & W.F. Fricke (2020). Bacterial microbiota diversity and composition in red and white wines correlate with plant-derived DNA contributions and botrytis infection. *Scientific Reports* 10, 13828 (2020). https://doi.org/10.1038/s41598-020-70535-8

A.M. Fricker*, D. Podlesny* & W.F. Fricke (2019). What is new and relevant for sequencing-based microbiome research? A mini-review. *Journal of Advanced Research*, Vol 19, https://doi.org/10.1016/j.jare.2019.03.006

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Acetyl-CoA	Acetyl coenzyme A
ACVD	Atherosclerotic cardiovascular disease
AIC	Akaike information criterion
AMP	Antimicrobial host defense peptides
ANOSIM	Analysis of similarities
ASV	Amplicon sequence variant
AUC	Area under the curve
AOM/DSS	Azoxymethane/dextran sulfate sodium
BH	Benjamnini-Hochberg procedure
BHB	β-Hydroxybutyric acid
BMI	Body Mass Index
bp	Base pair
CARBFUNC	Norwegian CARBFUNC study
CCL	Chemokine C ligand
CD	Crohn's disease
CED	Chronisch entzündliche Darmerkrankungen
CI	Confidence interval
DAMP	Danger associated molecular patterns
DC	Dendritic cells
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSS	Dextran sulfate sodium
DW	Dry weight
E%	Energy Percent
EMM	Estimated marginal mean
FDR	False discovery rate
GABA	Gamma amino butyric acid
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GC-MS	Gas-chromatography tandem mass spectrometry
(GC)-MS/MS	Triple quadrupole Gas Chromatography
GCN	16S rRNA gene copy number
GLMM	Generalized linear mixed effect model
$GLMM^{FDR}$	Generalized linear mixed effect model with FDR corrected q-value according to Benjamnini-Hochberg procedure
Gro-beta	Growth-regulated protein beta
GWAS	Genome-wide association study
HDL-C	High-density lipoprotein cholesterol
HIV	Human Immunodeficiency Virus
HMG-CoA reductase	3-hydroxy-3-methyl-glutaryl-coenzyme A reductase
IBD	Inflammatory bowel disease
IL-10 / 1 / 2 / 12	Interleukin-10 / 1 / 2 / 12
IgA	Immunoglobulin A
JAK/STAT	Janus kinase/signal transducer and activator of transcription
KETO	Ketogenic diet study
KO	knockout
LCHF	Low-carbohydrate high-fat diet
LC-MS/HRMS	Liquid chromatography-MS/ high resolution MS
LDL-C	Low-density lipoprotein cholesterol
LPS	Lipopolysaccharide
MAMP	Microbe associated molecular patterns

MS	Mass spectrometry
MUC2	Mucin-2 glycoprotein
NC	Natural killer cells
NLR	Nucleotide-binding and oligomerization domain-like receptor
NOD	Nucleotide Binding Oligomerization Domain Containing
OR	Odds ratio
OTU	Operational taxonomic unit
PBS	Phosphate-buffered saline
PCoA	Principal Coordinates Analysis
PCR	Polymerase chain reaction
PI3K/Akt	Phosphatidylinositol 3-kinase/protein kinase B
PRR	Pattern-recognition receptors
qPCR	Quantitative polymerase chain reaction
QIIME	Quantitative insights into microbial ecology
QTI	Quantitative trait loci
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
ROC curve	Receiver operating characteristics curve
RT-PCR	Reverse transcriptase polymerase chain reaction
SCFA	Short-chain fatty acid
SCR	Spearman's rank correlation test
SNP	Single nucleotide polymorphisms
SPF	Specific pathogen free
TAG	Triacylglycerides
TGFβ	Transforming growth factor beta
Th1 / 2 / 17	Helper T cells 1 / 2 / 17
TLR	Transmembrane PRRs like Toll-like receptor
TNF	Tumor necrosis factor
TMAO	Trimethylamine-N-oxide
UC	Ulcerative colitis
$WRST^{FDR}$	Wilcoxon rank-sum test with FDR corrected q-value according to Benjamnini-Hochberg procedure
WT	Wild-type
3NPH	3-nitrophenylhydrazones
-/-	Gene knockout

Zusammenfassung

Das Darmmikrobiom leistet einen wesentlichen Beitrag zur menschlichen Gesundheit, indem es z. B. die Reifung des Immunsystems und die Darmhomöostase fördert. Dennoch sind die Faktoren, die seine Zusammensetzung bestimmen, nur unzureichend erforscht. Bedingungen außerhalb und innerhalb des Körpers können das streng kontrollierte Zusammenspiel von Mikrobiom und Wirt stören und eine Dysbiose hervorrufen, die mit verschiedenen Erkrankungen wie Adipositas, atherosklerotischen Herz-Kreislauf-Erkrankungen und chronisch entzündlichen Darmerkrankungen (CED) assoziiert ist. Daher spielen das Erforschen der Einflussfaktoren, die zu mikrobiellen Veränderungen im Darm führen, und die Vorhersage der damit verbundenen gesundheitlichen Folgen eine zentrale Rolle in der Verbesserung der Prävention und Behandlung dieser Erkrankungen.

In der vorliegenden Arbeit wurde der Einfluss von diätetischen und genetischen Faktoren auf die Zusammensetzung der gastrointestinalen Mikrobiota untersucht, wobei die Ernährung in einer Humankohorte als extrinsischer, veränderbarer mikrobiom-relevanter Faktor und ein genetisches Knock-out Mausmodell für gastrointestinale Entzündungen als intrinsischer, nicht veränderbarer mikrobiom-relevanter Faktor jeweils exemplarisch diente. In beiden Studien wurden mikrobielle kompositionelle Zusammensetzungen, durch 16S rRNA-Genamplikon-Sequenzierung taxonomisch charakterisiert und im Zusammenhang mit metabolischen und entzündlichen Auswirkungen auf den Wirt analysiert.

Die Behandlung von atherosklerotischen Herz-Kreislauf-Erkrankungen zielt in erster Linie auf die Senkung eines Cholesterinüberschusses im Blut, der ein Hauptrisikofaktor ist, ab. Diese wird entweder durch eine Verringerung der Cholesterinaufnahme aus exogenen, diätetischen Quellen oder der Hemmung der endogenen Cholesterinbiosynthese erreicht. Die Umwandlung von Cholesterin in das nicht-absorbierbare Coprostanol durch die intestinale Mikrobiota soll ebenfalls die Cholesterinverfügbarkeit im Darm und im Serum verringern. Jedoch sind die Abhängigkeiten der Cholesterinumwandlung von spezifischen Bakterien, ihr Zusammenhang mit den Serumlipidwerten und Ernährungsgewohnheiten, die die Umwandlung bedingen, noch weitestgehend unbekannt. Um den Einfluss der Mikrobiota auf den menschlichen Cholesterinstoffwechsel unter verschiedenen Bedingungen zu untersuchen, wurden fäkale Mikrobiom- und Lipidprofile sowie Lipid-Biomarker im Serum in zwei unabhängigen Humankohorten bestimmt, darunter Personen mit (CARBFUNC-Studie) und ohne Adipositas (KETO-Studie), die sich drei bis sechs Monate bzw. sechs Wochen lang einer sehr kohlenhydratarmen, fettreichen

Ernährungsintervention (LCHF) unterzogen. Die Analyse von Personen mit und ohne Fettleibigkeit aus zwei geografisch unabhängigen Kohorten, zeigte eine einheitliche Verteilung der Cholesterinumwandlung in high- und low-converter Typen. In beiden Kohorten war die Cholesterinumwandlung am stärksten mit der relativen Häufigkeit des cholesterinumwandelnden Bakteriums *Eubacterium coprostanoligenes* assoziiert, dessen Vorkommen durch die LCHF-Diät in den low-convertern erhöht wurde und sie somit in einen high-conversion-ähnlichen Zustand versetzte. Die high-converter ohne Adipositas, die bereits vor der LCHF-Diät durch ungünstige Serumlipidprofile gekennzeichnet waren, reagierten auf die Intervention mit einem Anstieg der LDL-C Konzentration im Serum unabhängig von deren Verzehr von Fett, Cholesterin und gesättigten Fettsäuren. Diese Ergebnisse zeigen, dass der Cholesterin high-converter Typ ein potenzieller prädiktiver Biomarker für eine erhöhte LDL-C-Antwort auf eine LCHF-Diät bei stoffwechselgesunden, schlanken Personen ist.

Obwohl die Ätiologie von CED noch nicht vollständig geklärt ist, wird davon ausgegangen, dass ein Zusammenspiel zwischen der Darmmikrobiota, Umweltfaktoren und der genetischen Anfälligkeit eines Individuums besteht, welche chronische Entzündungen durch eine Dysregulation der Immunantwort im Darm auslöst. Um Colitis-assoziierte Mikrobiota-Veränderungen während der Entwicklung von CED zu identifizieren, wurden Mäuse mit einem genetischen Defekt des entzündungshemmenden Zytokines Interleukin-10 (IL-10), die aus verschiedenen Würfen stammen, zusammen mit Wildtyp-Mäusen in Käfige gesetzt und 20 Wochen lang beobachtet. Die Bewertung der Mäuse anhand ihres Phänotyps und ihrer Stuhlkonsistenz spiegelte den Zustand der Schleimhautentzündung wider, welche anhand histopathologischer Untersuchungen und Zytokinexpressionsprofile bestätigt wurde. Globale mikrobielle Veränderungen, welche die Entwicklung der Colitis kennzeichneten, sowie die Anfälligkeit für Colitis, hingen zudem von der Mikrobiom-Zusammensetzung ab, welche die Mäuse schon früh im Leben erhalten hatten. Die erhöhte Colitis-Anfälligkeit in Abhängigkeit vom Wurf wurde außerdem mit der Präsenz der Gattung *Akkermansia* kurz vor dem Auftreten von Symptomen in Verbindung gebracht. Die Präsenz dieser Gattung war zudem ein guter Prädiktor für das frühe Colitis-bedingte Ausscheiden der Mäuse, was darauf hindeutet, dass *Akkermansia* als Marker für eine früh einsetzende, subklinische Kolitis dienen könnte.

Zusammenfassend unterstreichen die Charakterisierungen des Mikrobiomes durch diätetische Modulation einer LCHF-Diät im Menschen und während der spontanen Entwicklung einer Darmentzündung in einem Colitis-Mausmodell das Potenzial von

mikrobiombasierten Patientenklassifizierungen. Diese könnten verwendet werden, um den klinischen Verlauf eines einzelnen Patienten vorherzusagen und personalisierte Behandlungsansätze zu verbessern.

Abstract

Although the gut microbiota is known to contribute fundamentally to human health, e.g. by promoting the maturation of the immune system and intestinal homeostasis, the factors shaping its composition are only poorly understood. Extrinsic and intrinsic influences can disturb the tightly controlled equilibrium between the microbiome and the host and induce dysbiosis, which has been linked to diverse health conditions such as obesity, atherosclerotic cardiovascular disease (ACVD) and inflammatory bowel disease (IBD). Therefore, understanding events leading to microbial perturbations and the prediction of associated health outcomes could aid in the prevention and treatment of these conditions.

In this work, the impact of dietary and genetic factors on gastrointestinal microbiota compositions were determined, with the diet serving as an exemplary extrinsic, modifiable microbiota-relevant factor and with a genetic deficiency in a mouse model for intestinal inflammation serving as an exemplary intrinsic, non-modifiable microbiota-relevant factor. In both studies, microbial communities obtained from either a human or a murine cohort, respectively, were taxonomically characterized by 16S rRNA gene amplicon sequencing and analyzed in the context of metabolic and inflammatory implications for the host.

In ACVD, the reduction of excess blood cholesterol, which is a main risk factor, is tackled by clinical interventions aiming to reduce cholesterol uptake from exogenous, dietary sources or by inhibiting endogenous cholesterol biosynthesis. Cholesterol-to-coprostanol conversion by the intestinal microbiota has also been suggested to reduce intestinal and serum cholesterol availability, but the dependencies of cholesterol conversion on specific bacterial taxa and dietary habits, as well as its association with serum lipid levels remain largely unknown. To study microbiota contributions to human cholesterol metabolism under varying conditions, fecal microbiota and lipid profiles, as well as serum lipid biomarkers, were determined in two independent human cohorts, including individuals with (CARBFUNC study) and without obesity (KETO study) on very low-carbohydrate high-fat diets (LCHF) for three to six months and six weeks, respectively. Across these two geographically independent studies, conserved distributions of cholesterol high and low-converter types were measured. Also, cholesterol conversion was most dominantly linked to the relative abundance of the cholesterol-converting bacterial species *Eubacterium coprostanoligenes*, which was further increased in low-converters by LCHF diets, shifting them towards a high-conversion state. Lean cholesterol high-converters, which were characterized by adverse serum lipid profiles even before the LCHF diet, responded to the

intervention with increased LDL-C, independently of fat, cholesterol and saturated fatty acid intake. These findings identify the cholesterol high-converter type as a potential predictive biomarker for an increased LDL-C response to LCHF diet in metabolically healthy lean individuals.

Although the etiology of IBD has not been fully resolved, an interplay between the intestinal microbiota, environmental factors and an individual's genetic susceptibility is thought to trigger chronic inflammation by a dysregulation of the immune response in the gut. To identify colitis-associated microbiota alterations throughout the development of spontaneous colitis, mice with a genetic deficiency of the anti-inflammatory cytokine Interleukin-10 (IL-10) from different litters were co-housed with wild-type mice and monitored for 20 weeks. The scoring of mice based on their phenotype and stool consistency mirrored the state of mucosal inflammation as assessed based on histopathological examinations and cytokine expression profiles. Also, the state of colitis was characterized by global microbiota alterations and susceptibility to colitis was dependent on litter-specific microbiome compositions that mice adopted early on in their lives. Colitis development was further associated with the presence of the bacterial genus *Akkermansia* in mature mice shortly before symptoms manifested. This genus was also a good predictor of colitis-related mice withdrawal, suggesting the potential of *Akkermansia* to serve as an early onset, subclinical colitis marker.

In summary, fecal microbiota characterizations in response to LCHF diets in humans and throughout the development of intestinal inflammation in a colitis mouse model highlight the potential of personalized microbiome-based patient classifications to predict clinical outcomes and improve treatment approaches.

I. Introduction

A. Gut microbiome-host interactions contribute to human health

Mammalians are thought to have evolved in the presence of microorganisms such as bacteria, which results in a natural population of commensals on various parts of the human body e.g. the skin, nose, and gastrointestinal tract. Human and bacterial cells are even represented in equal numbers in and on the human body with the intestine being the most densely colonized compartment harboring ~10^{13} bacteria in the colon (Sender et al., 2016), which adds an estimated 2 million genes to the functional capacity of the host (Gilbert et al., 2018). Under healthy conditions, the host and the gut commensals live in a symbiotic state with the host providing nutrients and niches for bacteria to grow, such as endogenous products like bile acids, cholesterol or intestinal mucus polysaccharides (Bull et al., 2002; Johansson et al., 2013; Midtvedt, 1974). The commensal bacteria in return prevent colonization and growth of pathogens, e.g. by competing for nutrients or altering the intestinal milieu via pH modifications or the production of antimicrobial components (Schoultz & Keita, 2020).

The gastrointestinal tract and the residential gut microbiota are both essential for maintaining nutrient supply and immune homeostasis to the host. The digestive tract, being comprised of an upper (mouth, esophagus, stomach, duodenum, jejunum and ileum) and lower (colon, rectum and anus) compartment, plays a pivotal role in the digestion and absorption of ingested nutrients (Carabotti et al., 2015). Whereas the intestinal microbiota further extends nutrient availability to the host, e.g. by metabolizing indigestible food components into exploitable metabolites to the host or by synthesizing vitamins such as biotin (LeBlanc et al., 2013).

In addition, the gastrointestinal tract harbors the largest number of endocrine and immune cells in the human body which highlights its importance in intestinal immune regulation (Latorre et al., 2016; Mason et al., 2008). The intestinal barrier, separating the gastrointestinal lumen and the host's systemic circulation, is designed to discriminate between harmful and non-harmful agents, which enables the absorption of nutrients and self-antigens, while protecting against the transition of harmful or infectious agents, e.g. antigens, toxins and pathogens (Camilleri et al., 2012; Furness et al., 1999). The epithelial layer, composed of enterocytes and secretory cells such as Paneth and goblet cells, is held together by tight junction proteins and represents a physical barrier towards the intestinal luminal milieu (Schoultz et al., 2020). On its basolateral site, the lamina propria harbors a vast population of endocrine and immune cells, e.g. neutrophils, macrophages, T-regulatory

and mast cells, acting as a first line defense against foreign substances and regulating immune responses, e.g. to dietary or microbial antigens (Keita & Söderholm, 2010). The apical site of the colonic epithelium is covered by a thick, dynamic and complex mucus layer consisting largely (~80%) of mucin-2 glycoprotein (MUC2), which is produced and secreted by goblet cells. Additionally, the mucus is enriched in Immunoglobulin A (IgA) and antimicrobial host defense peptides (AMP), which are produced and secreted respectively by plasma and Paneth cells (Bevins & Salzman, 2011; Gutzeit et al., 2014; Johansson & Hansson, 2016). With these properties, the mucus serves as a physical and chemical barrier against harmful luminal content but also normal microbial residents of the intestine. The mucus layer can further be divided into an inner tighter section, which is free of bacteria under healthy conditions, and a looser outer layer, which is continuously hydrolyzed by bacterial and host enzymes and represents a nutrient-rich niche for gut commensals capable of utilizing *O*-linked glycans, such as *Akkermansia municiphila* (Johansson et al., 2013; van Passel et al., 2011).

At birth, humans are first colonized by microbial communities, which are influenced by the mode of delivery (Blaser & Dominguez-Bello, 2016; Dominguez-Bello et al., 2010). The importance of this initial colonization for the development of the immune system has been demonstrated in studies with germ-free animals, born and housed under sterile conditions, which exhibit impaired immune maturation in the absence of commensals, e.g. reduced levels of IgA antibodies and helper T cells 17 (Th17). Interestingly, these impairments have been shown to be restored after microbial colonization (Dominguez-Bello et al., 2019; Gensollen et al., 2016; Hapfelmeier et al., 2010). Subsequent to the initial colonization, within the first two to three years of life microbial communities gradually develop towards an adult-like microbiome (Cho & Blaser, 2012), which is characterized by microbial persistence within individuals under healthy conditions (Faith et al., 2013; Podlesny et al., 2022; Schloissnig et al., 2012; Turnbaugh et al., 2007). The use of antibiotics in this early developmental window is linked to an increased risk of inflammatory and metabolic diseases in later stages of life (Mueller et al., 2014; Torow & Hornef, 2017), highlighting the pivotal role of microbiome-host interactions for human health.

B. Dysregulations of microbiome-host interactions

Imbalances of the microbial community equilibrium which prevent the recovery to the resilient state, are referred to as dysbiosis. This term does not describe a defined state of the microbial composition, but characterizes community alterations with reduced microbial diversity (referred to as species richness or α-diversity), a loss of beneficial bacteria and a

rise of pathobionts, bacteria that exhibit harmful effects under certain circumstances (Belizário & Faintuch, 2018; G. A. Weiss & Thierry Hennet, 2017). These alterations of gut microbiota profiles have been linked to numerous health conditions ranging from metabolic problems, such as obesity (Turnbaugh et al., 2006b), and intestinal immune disorders, including inflammatory bowel disease (IBD) (Halfvarson et al., 2017), to systemic inflammatory problems, such as allergy (Fujimura & Lynch, 2015), and even neurodevelopmental disorders like autism (Kang et al., 2019), and cancer (Ohigashi et al., 2013).

Mechanistically, the microbiota mediates changes affecting the host either by altering biochemical profiles or directly when physical barriers separating host and microbe environments are disrupted. For example, the production of neurotransmitters such as gamma amino butyric acid (GABA) and serotonin, have been attributed to *Bacteroides*, *Parabacteroides* and *Escherichia*, linking relative abundance alterations to neurological conditions like depression and anxiety (Bear et al., 2021; Strandwitz et al., 2018). On the other hand, disruption of the protective gut barrier increases intestinal permeability, which facilitates the translocation of bacteria, toxins and other luminal agents into the underlying tissue or the host's systemic circulation (Bjarnason et al., 1983; Hollander et al., 1986). This breach, also known as leaky gut, and consequent transition of luminal components trigger immune responses by recognition of microbe-associated molecular patterns (MAMPs) or damage-associated molecular patterns (DAMPs) via pattern-recognition receptors (PRR), which are presented on immune and non-immune cells in the gut (Janeway & Medzhitov, 2003; Matzinger, 2002). Transmembrane PRRs like Toll-like receptors (TLRs) and cytosolic PPRs, such as Nucleotide-binding and oligomerization domain (NOD)-like receptors (NLRs), can stimulate innate immune responses not only against bacterial AMPs but also against host-derived agents or dietary antigens leading to autoimmune diseases like celiac disease (Khaleghi et al., 2016), IBD (Fasano & Shea-Donohue, 2005), rheumatoid arthritis (C. J. Edwards, 2008) or type 1 diabetes (Tlaskalová-Hogenová et al., 2011). However, to what extent the gastrointestinal microbiota contributes to these diseases remains unclear, and the factors which provoke microbial community composition alterations have not been fully understood yet.

C. Factors shaping the intestinal microbiota composition

One of the major challenges in microbiome research is the high interindividual heterogeneity in and between human cohorts. These interindividual microbiota variations are determined by intrinsic factors such as genetics (Hall et al., 2017; Imhann et al., 2018; Lim et al., 2017;

Ussar et al., 2015) and even more so by environmental conditions, e.g. dietary habits, sharing a household or drug intake (Rothschild et al., 2018; Vujkovic-Cvijin et al., 2020; Zhernakova et al., 2016).

1. Exogenous determinants of intestinal microbiota composition

As the survival and growth of the intestinal microbiota relies on the exploitation of food residues, gut bacteria have acquired diverse capabilities to metabolize ingested food components. Therefore, dietary habits have been shown to shape the intestinal community favoring colonization of well attuned commensals, which are able to utilize foods consumed in high frequency (Ang et al., 2020; Bolyen et al., 2019; Smith et al., 2019; A. W. Walker, Ince, et al., 2011). The microbial community even rapidly responds to changes in dietary habits, as shown within five days of entirely animal or plant-based product dietary interventions to represent shifts in microbial activity either towards protein or carbohydrate fermentation depending on the diet (David et al., 2014). These dietary modifications of the microbiota have been linked to beneficial as well as adverse effects on health, e.g. as described for high/low-fiber diets. Short-chain fatty acids (SCFA), metabolites obtained by the microbial breakdown of undigestible fiber in the human intestine, strengthen the gut barrier by serving as energy source for colonocytes, increasing mucus production and secretion, and altering pH levels, which limits pathogen growth (Koh et al., 2016). Whereas a high consumption of undigestible fibers is revealed to promote host health, their deprivation reduces SCFA concentrations in the gut and facilitates successive degradation of the mucus barrier, which increases the risk of pathogen invasion and bacterial translocation (Desai et al., 2016; Schroeder et al., 2018; Sonnenburg & Sonnenburg, 2014).

Lifestyle determinants, such as smoking (Fluhr et al., 2021; Lee et al., 2018; Lim et al., 2016), drug intake (Falony et al., 2016; Imhann et al., 2018; M. A. Jackson et al., 2018), physical activity (Estaki et al., 2016; Scheiman et al., 2019; Zhao et al., 2018) and housing situations have also been identified to characterize the microbiota (Vujkovic-Cvijin et al., 2020). Interestingly, the gut microbiomes of genetically unrelated individuals are more similar to each other when they share the same household compared to individuals, which they do not cohabitate with (Rothschild et al., 2018; Yatsunenko et al., 2012). Humans even share microbial community similarities with their pets (Caugant et al., 1984; Song et al., 2013) stressing the importance of exposure to similar environments, e.g. by diet, being in contact with the same microbial sources and surfaces, or by direct contact between individuals. The bacterial transmission between unrelated individuals is even more important to consider in animal models, e.g. murine studies, as mice are coprophagic, which

facilitates the exchange of intestinal microbiomes when being cohoused (Ebino et al., 1987; Keubler et al., 2015).

2. Endogenous traits modulating the gut microbiota

Compared to environmental factors, intrinsic conditions such as genetics are being thought to play a minor role in determining the microbial community composition (Turnbaugh et al., 2008; Yatsunenko et al., 2012). However, they should not be neglected, as food preferences, aspects of immunity or gut physiology, could be genetically determined (Blekhman et al., 2015; Bonder et al., 2016). Also, the heritability of certain gut microbiota members has been suggested by multiple independent studies (Goodrich et al., 2014, 2016; Podlesny & Fricke, 2021; Stewart et al., 2005; van de Merwe et al., 1983; Xie et al., 2016; Zoetendal et al., 2001). For example, profiling the taxonomic composition of gut microbiomes from 416 twin pairs of the UK Twin study, found microbial communities to be more similar between monozygotic twins, sharing an identical genetic profile and environment *in utero*, than between dizygotic twins, which carry different genetic profiles. Additionally, the authors were able to model estimates of heritability, *"[…] which is the proportion of variance in the phenotype that can be attributed to genetic differences between hosts [...]"* (Goodrich et al., 2016) identifying the family *Christensenellaceae* to be the most heritable among dominant members of the gut microbiota (Goodrich et al., 2014). The fecal transplantation of microbiomes from twins of this study into germ-free mice revealed that the addition of the species *Christensenella minuta* to a microbial consortium, which is restricted in *Christensenellaceae* abundance, even altered the mice's phenotype as shown by a limited increase of adiposity (Goodrich et al., 2014). When the UK twin study dataset was expanded to 1,019 twin pairs, highly heritable taxa were additionally associated with a higher compositional stability within individuals over time (Goodrich et al., 2016). Deep metagenomic sequencing of a subgroup from the UK twin study further validated this heritable microbial stability, as measured by shared microbial single nucleotide polymorphisms (SNPs) between twins, to slowly but gradually deviate with increased time spent living apart (Xie et al., 2016), which highlights the interplay between host-related factors and environmental conditions to determine microbial compositions.

The mapping of quantitative trait loci (QTL), chromosomal regions that determine variations of phenotypic traits, in mice further revealed associations between taxon abundances and genetic loci to be involved in the immune response (Benson et al., 2010; Leamy et al., 2014). For example, genetic regions encoding *TGFβ-3* (Transforming Growth Factor Beta 3), an immune-suppressive cytokine contributing to intestinal homeostasis, have been

found to overlap a QTL on chromosome 12, which correlates with relative abundances of *Prevotellaceae* (McKnite et al., 2012). Taken together, this emphasizes the microbiota's role to determine phenotypic traits relevant for host health.

3. Current challenges in determining microbiota-modulating factors

Whereas environmental conditions can be modified relatively easy, in multiple ways and investigated in various model organisms, e.g. by dietary or lifestyle interventions in human cohorts or animal studies, the investigation of intrinsic modulations of the microbiome are mostly limited to animal models. Both approaches imply unique advantages and restrictions, depending on the research hypothesis. Although findings of human studies are preferred over animal models in terms of direct translatability, they have a limited range of application due to the exclusion of vulnerable groups such as pregnant women and children. Additionally, not every environmental condition can be examined due to ethical restrictions, e.g. if it involves the exposure to harmful effects. In particular for microbiome studies, the consideration of biases due to uncontrolled environmental conditions is critical. To overcome these disadvantages, animal especially mouse studies are often used, which allow for controlled environments and reduced external influences. Their validity has been increased by the generation of humanized mice, mice that have been engrafted with human material, e.g. DNA, tissue, cells or the fecal microbiome (Fujiwara, 2018). Methods to genetically modify mice by gene knockout (KO) introducing a loss of function are also used to create models mirroring human diseases, e.g. immune-deficient conditions such as IBD or neurodegenerative diseases (Doyle et al., 2012).

To improve the success of treatment and prevention strategies for microbiota-associated diseases, it is pivotal to understand which factors and to what extent they shape microbial communities, and despite their limitations, human and mouse models play a central role in their uncovering. Therefore, in the following two main factors were examined for their influence on the microbiota: (1) dietary modulations in human cohorts as an environmental microbiota-relevant factor, and (2) genetic susceptibility to intestinal inflammation in a mouse model as an intrinsic, microbiota-modulation factor.

4. Dietary modulation of the intestinal microbiota

Low-carbohydrate high-fat (LCHF) diets in which more than 75 energy percent (E %) are derived from fat and less than 10 E % from carbohydrates, introduce specific metabolic alterations due to a lack of glucose, the primary energy source in humans. Among LCHF

diets, ketogenic diets represent an extreme form with very-low carbohydrate intake of less than 50 g per day for an adult, which enhances shifts towards a catabolic state. Deprivation of carbohydrates forces the body to use alternative energy production pathways such as gluconeogenesis and ketogenesis (Garber et al., 1974). Gluconeogenesis, the exploitation of glucose from endogenous glycogen stores in the liver and muscles or from lactate, glycerol, glucogenic amino acids, and odd chain fatty acids, is an energy-demanding process that facilitates energy recovery only temporarily. During long-term fasting the energy supply is maintained primarily by ketogenesis, the breakdown of fatty acids into the three ketone bodies acetone, acetoacetate, and beta-hydroxybutyrate (BHB) in the liver. Ketogenic pathways are constantly active resulting in low ketone levels even when glucose is available and are heavily increasing in times of carbohydrate deprivation (Scheinberg, 1965; X. Zhang et al., 2019). While acetoacetate and BHB are metabolized into acetyl coenzyme A (acetyl-CoA), a central molecule in energy production pathways, acetone is excreted via the urine or exhaled (Scharrer, 1999; Sengupta et al., 2010).

LCHF diets have been associated with variable metabolic health consequences, especially in the context of atherosclerotic cardiovascular disease (ACVD), the leading cause of mortality worldwide (Barquera et al., 2015). In ACVD, cholesterol, an amphipathic sterol lipid, plays a central role. Cholesterol is an essential structural component of human and animal cell membranes and serves as a precursor for the biosynthesis of steroid hormones, bile acids and vitamin D. To exhibit its beneficial properties, cholesterol needs to be distributed in the periphery, which is maintained by cholesterol-carrying lipoproteins such as low- (LDL-C) and high-density lipoprotein cholesterol (HDL-C). Both vary in composition, with LDL-C molecules, the major cholesterol distributor from the liver to peripheral tissue, containing high proportions of cholesterol (~50%), whereas HDL-C particles carry only ~20% of cholesterol and are responsible for the reverse cholesterol transport from the periphery back to the liver (Scanu & Wisdom, 1972). Dyslipidemia, an imbalance of these lipid serum levels which is determined by low plasma concentrations of HDL-C and high levels of triacylglycerides (TAG) and LDL-C, as well as hypercholesterolemia, an abnormally increased LDL-C concentration, are both clinical risk factors for ACVD (Barquera et al., 2015).

LCHF diets improve blood glucose regulation in individuals with obesity (Kirkpatrick et al., 2019; Westman et al., 2007), but can increase serum concentrations of LDL-C in others (Bueno et al., 2013; Burén et al., 2021; O'Neill & Raggi, 2020). In addition, LCHF diets have become popular among normal-weight individuals, who may also experience increased

LDL-C responses (O'Neill et al., 2020). A recent study even suggested a LCHF diet-induced LDL-C "lean mass hyper-responder" phenotype in a subset of metabolically healthy individuals (Norwitz et al., 2022). Nonetheless, LCHF ketogenic diets appear to improve lipid cardiovascular disease risk markers (Paoli et al., 2013), even with increased saturated fat intake (Feldman et al., 2022), which is known to elevate serum LDL-C concentrations (Luukkonen et al., 2018). Notably, interindividual variations in the LDL-C response to saturated fats (Griffin et al., 2021) and cholesterol-reducing medications, including statins (Descamps et al., 2015; Qamar et al., 2019) have also been reported. But as current studies are lacking extensive metabolite analyses to identify the factors accountable for the contradictory health outcomes of LCHF diets, individual responses remain unpredictable, preventing the development of guidelines for personalized dietary recommendations.

Large amounts of cholesterol enter the small intestine every day from exogenous dietary sources, including mostly animal products (~0.3–0.6 g/day), and from endogenous biosynthesis in the liver and secretion with bile acids (~0.7–0.9 g/day) (Juste & Gérard, 2021). Consequently, clinical treatments to prevent hypercholesterolemia aim to limit dietary cholesterol intake, improve the composition of dietary fats, inhibit cholesterol biosynthesis, and/or block intestinal uptake. Statins, pharmacological inhibitors of the 3-hydroxy-3-methyl-glutaryl-coenzyme A (HmG-CoA)-dependent cholesterol-generating mevalonate pathway in the liver, represent one of the most successful, widely used and best-selling drug classes worldwide (Redinbo, 2020). Lowering dietary cholesterol intake alone typically has shown limited and inconsistent effects (Gotto, 1991). Cholesterol production is tightly controlled via feedback mechanisms that adapt HmG-CoA activity to dietary cholesterol intake and cellular requirements (Luo et al., 2019) and involve insulin, which can increase or decrease HmG-CoA expression in response to higher or lower-carbohydrate diets (Kirkpatrick et al., 2019). Moreover, the clinically successful combination of statin treatment with ezetimibe, a small molecule inhibitor of cholesterol uptake from the intestine (Bays et al., 2008), demonstrates the importance of intestinal cholesterol availability. The human gut microbiota can metabolize statins (Zimmermann et al., 2019) and may contribute to adverse side effects of the medication that are observed in some individuals (Tuteja & Ferguson, 2019). Recently, heterogeneous on-target effects and adverse responses to statin therapy were associated with taxonomic gut microbiota compositions (Wilmanski et al., 2022). Similarly, the gut microbiota has been suggested to influence intestinal cholesterol availability (Kriaa et al., 2019), which could explain inconsistent clinical success with dietary intervention due to interindividual variations in

microbiota composition (Asnicar et al., 2021; Lloyd-Price et al., 2017). Further, the intestinal microbiota might be leveraged to define new diagnostic and therapeutic targets for patient stratification and optimization of pharmacological and non-pharmacological cholesterol reduction treatments.

In the intestine, cholesterol is microbially reduced to coprostanol, which is non-absorbable, stable under anoxic conditions, excreted in feces (Bull et al., 2002), undetectable in human newborns (Midtvedt AC, 1993) and germ-free rats (Kellogg & Wostmann, 1969), and reduced in antibiotically treated animals and humans (Midtvedt et al., 1990; Wainfan et al., 1952). The rate of cholesterol-to-coprostanol conversion varies between individuals, but shows a stable bimodal distribution in different human populations of cholesterol high and low-converters as classified based on the ratio of fecal coprostanol/cholesterol concentrations (Salyers et al., 1977; Veiga et al., 2005; Wilkins & Hackman, 1974). Cholesterol reduction has been attributed to a broad and diverse range of microbial taxa (Juste et al., 2021), based on *in vitro* experiments with bacterial isolates and metagenomic sequence analyses (Kenny et al., 2020). The influence of microbial cholesterol conversion on circulating cholesterol levels has been inconsistently described. An inverse relationship between serum concentrations of cholesterol and the fecal coprostanol/cholesterol ratio was reported in hospitalized patients (Sekimoto et al., 1983). Moreover, serum cholesterol levels were reduced in humans with a predicted lower capacity for microbial cholesterol reduction, based on fecal metagenomic sequence analysis (Kenny et al., 2020). Similarly, feeding rabbits with the cholesterol-reducing bacterium *Eubacterium coprostanoligenes* had a hypocholesterolemic effect (L. Li et al., 1995). However, the same bacterium did not affect plasma cholesterol levels in laying hens (L. Li et al., 1996) or germ-free mice (L. Li et al., 1998), despite evidence for at least transient intestinal colonization and cholesterol-to-coprostanol conversion. Thus, the specific details on the microbial origins of cholesterol conversion in humans, the stability and modifiability of the cholesterol converter status and its association with circulating cholesterol levels and other serum lipid markers involved in ACVD remain mostly uncharacterized.

5. Genetic modulation of the intestinal microbiota

Inflammatory bowel disease (IBD), a chronic inflammation of the gastrointestinal tract, which comprises the two forms ulcerative colitis (UC) and Crohn's disease (CD), is a prevalent gastrointestinal condition linked to microbial dysbiosis in the gut (Xavier & Podolsky, 2007). Whereas mucosal inflammation of UC is limited to the colon, transmural inflammation of CD can emerge along the entire digestive tract (Baumgart & Sandborn,

2007). Both forms manifest primarily with diarrhea, abdominal pain and rectal bleeding but can also cause extraintestinal symptoms affecting the skin, joints or bones (Larsen et al., 2010). Even though the etiology of IBD has not been fully resolved, an interplay between the intestinal microbiota, environmental factors and an individual's genetic susceptibility is thought to trigger the dysregulation of the gastrointestinal immune response (Bach, 2002; Keubler et al., 2015; Xavier et al., 2007). Although, a large meta-analysis of genome-wide association studies (GWAS) identified 163 genetic loci to be involved in IBD susceptibility, they only explained little variation in disease risk. This rather points towards interactions between the mucosal immune system and the commensal microbiota to additionally contribute to IBD pathogenesis (Jostins et al., 2012). Particularly, the imbalance between protective and potentially harmful microbes is discussed to trigger gut inflammation (Sartor, 2001, 2008). Unfortunately descriptions of IBD-associated microbial communities are inconsistent between human studies, which limits treatment approaches to target the immune system only, e.g. by immunosuppressants (Frank et al., 2007; Gophna et al., 2006; Hansen et al., 2012; Kellermayer et al., 2012).

To study microbiota alterations in the context of severe intestinal inflammation, the *interleukin-10* knockout (*IL-10*$^{-/-}$) mouse model is widely used as it mimics histological and inflammatory aspects of human IBD (Keubler et al., 2015; Maharshak et al., 2013). IL-10 is broadly expressed by various cells of the adaptive immune system, such as T helper 1 (TH1), TH17 and regulatory T cells (T_{reg}), as well as the innate immune response, e.g. dendritic (DC) or natural killer cells (NC). It is released from these cells upon activation for example by bacterial membrane molecules like lipopolysaccharides (LPS) and lipoproteins (Verma et al., 2016). The expression of IL-10 is tightly regulated by positive and negative feedback loops of IL-10 itself and the pro-inflammatory cytokine interferon gamma (IFN-γ). As a key regulator of gut homeostasis its main function is to control and dampen inflammatory responses and limit host damage (Saraiva & O'Garra, 2010). Deficiency of IL-10 in mice leads to the development of a mainly TH1/TH17-mediated spontaneous colitis after weaning due to a lack of its anti-inflammatory and immunosuppressive properties, which are mediated via the IL-10 receptor. Upon receptor binding, Janus kinase/signal transducer and activator of transcription (JAK/STAT) and phosphatidylinositol 3-kinase/protein kinase B (PI3K/Akt) cascades are activated, which lead to inhibition of pro-inflammatory cytokine production in macrophages, e.g. tumor necrosis factor α (TNFα), interleukin 12 (IL-12), and IFN-γ released by the same IL-10 producing cells (Feng et al., 2011; Riley et al., 1999).

Nevertheless, IL-10 seems to exhibit its protective properties in dependence of intestinal colonization. Several commensal bacteria, such as *Bacteroides fragilis* and *Clostridium* species have been shown to induce IL-10 production and differentiation of progenitor into IL-10 producing cells, stimulating immune tolerance towards the intestinal microbiome (Atarashi et al., 2011; Round & Mazmanian, 2010). Furthermore, *IL-10*$^{-/-}$ germ-free mice have been shown to not develop colitis unless they were colonized with a specific-pathogen free (SPF) microbiota (Sellon et al., 1998) and antibiotic treatment of neonatal *IL-10*$^{-/-}$ pups prevented intestinal inflammation up to twelve weeks of age (Madsen et al., 2000), which stresses the importance of microbial colonization for the development and maintenance of intestinal homeostasis. However, whether microbiota changes play a causative role in IBD, e.g. by infections with specific pathogens or alterations favoring intestinal inflammation, or if the microbial community composition changes subsequent to IBD onset remains unclear today.

D. Scope and aim of work

The overall aim of this thesis is to determine the impact of dietary and genetic factors on gastrointestinal microbiota compositions, as well as their metabolic and inflammatory consequences. To accomplish this, microbial communities obtained from (1) a human dietary intervention study, with the diet serving as an exemplary extrinsic, modifiable microbiota-relevant factor and (2) a genetic mouse model for intestinal inflammation, with the genetic deficiency serving as an exemplary intrinsic, non-modifiable microbiota-relevant factor, are taxonomically characterized by 16S rRNA gene amplicon sequencing and analyzed in the context of their metabolic and inflammatory implications for the host.

Low-carbohydrate high-fat diet as an extrinsic, modifiable microbiota-relevant factor, which alters microbial sterol conversion in humans

In the first part the relationship between intestinal cholesterol-to-coprostanol conversion and the gut microbiota is examined, as well as its link to serum lipid profiles, in the context of LCHF diet intervention. Metabolic profiling of fecal and serum metabolites involved in lipid and cholesterol metabolism and taxonomic profiling of the fecal microbiota of two geographically separate human cohorts reveal conserved and distinct distributions of cholesterol high and low-converter types among people with and without obesity, which are mainly attributed to the relative abundance of the bacterial species *Eubacterium coprostanoligenes*. Microbial sterol conversion, as well as the abundance of *E. coprostanoligenes* are both induced by LCHF dietary intervention in low-converters independent of the metabolic background. A subset of lean cholesterol high-converters,

which was characterized by adverse serum lipid profiles even before the LCHF diet, responded to the intervention with increased LDL-C concentrations, identifying the cholesterol high-converter type as a potential predictive biomarker for an increased LDL-C response and increased ACVD risk to LCHF in metabolically healthy lean individuals.

Genetic interleukin-10 deficiency as an intrinsic, non-modifiable microbiota-relevant factor, which promotes intestinal inflammation in a murine IBD model

To identify colitis-associated microbial markers, which characterize the fecal microbiota before and during colitis onset and in remission, female four-week-old *IL-10* knockout mice, alone or co-cohoused with matched wild type animals, were longitudinally investigated over a course of 20 weeks. Qualitative, taxonomic and quantitative microbiota profiles, as determined by sequencing and real-time polymerase chain reaction quantification of the 16S rRNA gene amplicon, as well as alterations of histopathological scores and inflammatory markers, reveal colitis susceptibility and microbiota compositions to be dependent on the mouse litter. Microbiomes of mice developing spontaneous colitis are further characterized by the presence of the genus *Akkermansia* even shortly before intestinal inflammation manifests with symptoms, suggesting that *Akkermansia* is a potential subclinical early colitis marker, which could help to predict colitis onset and improve the management of IBD therapy.

In summary, the analyses from this thesis focus on microbial community characterizations based on their metabolic and inflammatory consequences, which could be used for microbiome-dependent phenotyping and stratification of individuals to improve predicted health outcomes and personalized treatment approaches.

II. Methods

A. Study designs

1. Clinical trials for dietary interventions

For the determination of external influences on the gut microbiota, e.g. dietary modulation, two clinical trials were collectively analyzed.

The ketogenic diet (KETO) study is a single-arm before-and-after dietary intervention consisting of an isocaloric, *ad libitum*, ketosis-inducing diet with a total daily energy intake of at least 75 % fats, 12-20 % proteins and 5-10 % carbohydrates for 6 weeks, more specifically 42 days (Figure 1, page 30). Setting the focus on physical performance as primary outcome by Urbain et al. (Urbain et al., 2017), 28 healthy, lean adults (20 females, 8 males) were included with a body mass index (BMI) between 19-30 kg/m^2. Participants received counselling by a dietitian before intervention and were given a list of foods with very low carbohydrate content, which they should incorporate into their ketogenic diet according to personal preferences. Dietary compliance was assured by 7-day food questionnaires, which were collected before and during the last week of intervention, and daily self-tested measurements of urinary ketone bodies (Ketostix, Bayer Vital GmbH, Leverkusen, Germany). If necessary according to daily ketogenesis measurements, participants received dietary counselling by the dietitian via phone or in person to ensure continuous ketosis. Subjects were also advised to weigh all food items separately with a portable scale (KS 22, Beuer GmbH, Ulm, Germany) or if not possible, estimate amounts or to take photographs. Energy, macro- and micronutrient intakes were estimated based on the semi-quantitative 7-day dietary records with the Prodi 6.4 basis nutritional database software (Nutri-Science GmbH, Germany). The study was conducted at the University Medical Center Freiburg, Germany, approved by the Ethics Commission of the Albert-Ludwig University of Freiburg (494/14) and registered at the German Clinical Trials (DRKS00009605).

The CARBFUNC study is a 2-year randomized controlled trial involving 145 subjects (72 females, 73 males) with obesity (BMI > 30 kg/m^2 and/or waist circumference >102 cm for men and >88 cm for women, Figure 1, page 30). The study participants were assigned to three dietary interventions: a high-fat/low-carbohydrate diet (LCHF) including ≥ 75 energy percent (E %) from fat (with 30 E % from saturated fats) and ≤ 10 E % from carbohydrates, and two low-fat/high-carbohydrate diets differing in carbohydrate quality, containing 45 E % from carbohydrates and 38 E% from fat respectively

(https://clinicaltrials.gov/ct2/show/NCT03401970) (Horn et al., 2022). Herein, focus is set on the high-fat/low-carbohydrate diet, as comparison to the KETO study. All diets were designed with a normocaloric total energy amount of 2,000 kcal/day for women and 2,500 kcal/day for men and 17 % of total energy intake being derived from protein. CARBFUNC study participants were instructed to follow nutritional recommendations, to consume at least 500 grams of fruits and vegetables per day and generally rely on high-quality food sources. The subjects were considered healthy, as exclusion criteria were serious and/or cardiovascular disease, surgical or antibiotic treatment during the last two months before the study, chronic inflammatory bowel disease, cholecystectomy, hypo-/hyperthyroidism, menopause, pregnancy or breast feeding, food allergies or intolerances that prevent dietary adherence, medication affecting metabolism, use of statins and/or diabetes medication, smoking and alcohol consumption of more than two units per day during the study. The study was conducted at the University of Bergen, in collaboration with the Haukeland University Hospital in Bergen, Norway. The study was approved by the Regional Committee for Medical and Health Research Ethics (REC West Norway (2017/621/REK vest) and registered at ClinicalTrials.gov (NCT03401970).

All experiments adhered to the regulations of the KETO and CARBFUNC study review boards. All study procedures were performed in compliance with all relevant ethical regulations and each participant signed informed consent prior to participation.

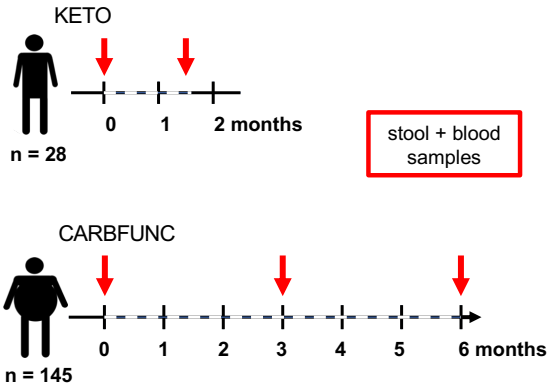


Figure 1. Observation periods of KETO and CARBFUNC subjects.
Lean KETO study participants followed an ad libitum ketosis-inducing high-fat (>75 E %) diet, whereas CARBFUNC study participants with obesity were randomly assigned to iso- and normocaloric high-fat (>=75 E %) or a low-fat (38 E %) diets. Fecal and serum samples were collected before intervention and after 6 weeks (KETO study) and three and six months (CARBFUNC study) and used for microbiota analysis and lipid profiling.

2. Murine model for genetic modulation

To study genetic impacts on the gut microbial community in the context of inflammation especially inflammatory bowel disease (IBD) progression, a murine *Interleukin-10* (IL-10) knockout (-/-) model was used. 18 female, 4-week-old specific-pathogen-free (SPF) BALB/c IL-10 deficient mice, from three distinct litters were co-housed with 4 female SPF swiss wild-type (WT) mice of the same age. The mice were allocated to six cages, with four cages co-housing WT and knockout mice and two cages with only the IL-10 deficient animals (Table 1, page 31). Within the first week after cage allocation, fecal samples were collected four times and considered as “baseline” samples to track litter and allocation effects, as well as normal short-term temporary changes within the microbial community. Mice were fed a standard chow diet (C 1000, Altromin, Lage, Germany) and no treatment besides regular fecal sampling and animal handling was performed. After the first week, fecal samples of each mouse were collected weekly and the health status, as defined by ethical requirements of the German Tierschutz-Versuchstierverordnung, daily until signs of potential colitis were noticed. These included behavior, appearance, weight, stool consistency and absolute termination criteria like rectal prolapse, bleeding from mouth or anus, paralysis, convulsions and persistent diarrhea. These parameters were scored (Table S1) and mice euthanized by gas if a score of three in one parameter or the sum of all scores greater than six was reached.

Table 1. Allocation and sampling overview of the *in vivo* murine experiment.
18 IL-10 deficient and 4 swiss wild-type mice were cohoused and sample subsets chosen for quantitative (**X**, n=542) and qualitative (**X**, n=120) microbiome analysis based on the symptomatic phenotype (according to stool consistency, weight progression and health scores), as shown for days and weeks of observation. The first mouse with colitis-related symptoms was observed after 10 weeks, initiating the symptomatic phase with more frequent sampling until final termination of the experiment after 20 weeks.

Weeks of observation			0	1			4	7	10	11				12		13	14					15		16				17						18						19						20
Days of observation			0	4	6	7	25	48	67	74	75	76	80	83	87	90	95	98	99	100	101	104	108	111	112	113	115	116	117	118	119	120	122	123	124	125	126	127	129	130	131	132	133	134	136	137
Weeks of age			4	4.5	5	5	7.5	11	13.5	14.5			15.5	16	16.5	17	17.5	18		18.5		19	19.5	20			20.5			21			21.5			22			22.5			23			23.5	
Mouse No.	Litter No.	Cage No.	Baseline				Healthy phase			Symptomatic phase																																				
56	1	1	X	X	X	X	X	X	X	X			X		X		X	X	X	X	X		X		X	X	X	X	X	X			X			X			X			X		X	X
81	2	1	X	X	X	X	X	X	X	X			X		X		X	X																											
87	2	1	X	X	X	X	X	X	X	X			X		X		X	X	X	X	X		X		X	X	X	X	X	X			X			X			X			X		X	X
103	4	1	X	X	X	X	X	X	X	X			X		X		X	X	X	X	X		X		X	X	X																		
57	1	2	X	X	X	X	X	X	X	X	X	X	X		X		X				X		X		X	X	X			X	X	X	X	X	X	X			X	X	X	X	X	X	X
82	2	2	X	X	X	X	X	X	X	X	X																																			
98	4	2	X	X	X	X	X	X	X	X	X	X	X		X		X				X		X		X		X			X	X	X	X	X	X	X			X	X	X	X	X			
104	4	2	X	X	X	X	X	X	X	X	X	X	X		X		X				X		X		X	X	X			X	X	X														
58	1	3	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X	X	X	X			X		X	X
83	2	3	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X	X	X	X			X		X	X
90	3	3	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X	X	X	X			X		X	X
99	4	3	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X	X	X	X			X		X	X
59	1	4	X	X	X	X	X	X	X	X			X		X		X				X	X	X	X			X			X	X	X	X			X			X			X		X	X
84	2	4	X	X	X	X	X	X	X	X			X		X		X				X	X	X																							
91	3	4	X	X	X	X	X	X	X	X			X		X		X				X	X	X	X			X			X	X	X	X			X			X			X		X	X
100	4	4	X	X	X	X	X	X	X	X			X		X		X				X	X	X	X			X			X	X	X	X												
85	2	5	X	X	X	X	X	X	X	X			X		X		X				X		X				X	X	X	X	X	X	X			X			X			X		X	X
92	3	5	X	X	X	X	X	X	X	X			X		X		X				X		X				X	X	X	X	X	X	X			X			X			X		X	X
101	4	5	X	X	X	X	X	X	X	X			X		X		X				X		X				X	X	X	X															
86	2	6	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X			X			X			
93	3	6	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X			X	X	X	X		X	X
102	4	6	X	X	X	X	X	X	X	X			X		X		X				X		X				X			X			X			X			X	X	X	X		X	X

Genotype

- Swiss Wild Type
- Balb/c IL-10 -/-

Sampling and condition colouring

- Regular sampling during 1st week (baseline)
- Regular sampling
- Co-housing samples of symptomatic mice
- Symptomatic mice (weight or consistency score >2)
- Euthanized mice
- Samples collected but not selected for analysis subset
- Samples not available/collected

To track the progression of colitis development and potentially spot the onset of inflammation, as soon as changes in stool consistency or other phenotypic symptoms were noticed, fecal samples were collected on two to three consecutive days for all cage mates of the conspicuous mouse. Stool consistency was defined according to a modified and simplified Bristol stool scale (Table 2, page 33) and samples grouped as "*healthy*", "*symptomatic*" and "*post colitis*" accordingly. Post colitis was characterized as healthy fecal samples derived from a previously symptomatic mouse. The first signs of colitis were observed after ~10 weeks, thereafter regular sampling took place twice a week unless no phenotypic symptoms were observed (Table 1, page 31). In addition to the three visual sample categories, a fourth group was defined as "*preC*", with referring to all healthy fecal samples of a mouse which were collected within four days prior to the first symptomatic observations. With this category, potential changes in the gut microbiome community that might be related to colitis symptoms and development should be captured. For analyses, a subset of samples representing disease progression was chosen resulting in 542 samples (Table 1, page 31).

The *in vivo* murine trial was carried out in cooperation with the group Molecular Allergology of the Department of Nutritional Medicine and Prevention of the University of Hohenheim, Germany. All animal related tasks, such as breeding, housing, monitoring and euthanizing were performed at the University of Hohenheim's Central Facility for Biological and Medical Research with Animal Husbandry.

Table 2. Modified Bristol Stool Scale for stool consistency classification of murine fecal samples.
Stool consistency scores were grouped by condition as “healthy”, “pre/post symptoms” or “symptomatic”. Example graphs modified based on Adobe Stock #434183203 (standard license)

Score	Example	Description	Condition
1		Smooth, normal consistency, soft enough to be divided by plastic spatula	Healthy/pre and post symptoms
2		Soft blobs with clear-cut edges	Symptomatic
3		Fluffy pieces with ragged edges, a mushy stool	Symptomatic
4		Watery, no solid pieces, entirely liquid	Symptomatic

B. Sample collection

For the KETO study, fecal samples were self-collected by the participants in RNAlater (Thermo Fisher Scientific, Waltham, USA) on two days before (PRE: days -2 and 0) and at the end of the dietary intervention (POST: days 40 and 42), stored in a fridge until drop-off and frozen at -80°C at the study facility. Venous blood was drawn at visits on days 0 and 42 after overnight fasting and immediately sent to the Institute for Clinical Chemistry and Laboratory Medicine of the University Medical Center Freiburg, Germany for analyses. (Urbain et al., 2017)

For the CARBFUNC study, fecal samples were self-collected before and at three and six months of the dietary intervention, stored in the freezer and dropped off at the study facility for further storage at -80°C. Blood samples were drawn at the same study visits after overnight fasting.

Fecal samples of both clinical trials were shipped on dry ice to the Microbiome and Applied Bioinformatics facility at the University of Hohenheim, Germany and stored at -80°C until processing.

For the *in vivo* mice study, mice were either placed alone in a separate cage until defecation or held in the gloved hand and the belly rubbed until defecation. Each sample was stored in a sterile tube respectively. At trial termination (~20 Weeks after the beginning) all remaining mice were euthanized and gastrointestinal tissue samples derived. Ileum, the proximal and distal colon were separated and flushed with sterile phosphate-buffered saline (PBS) to collect fecal content. Fecal samples throughout the observation, as well as at the

termination days, were collected, stored on ice and immediately brought to the Microbiome and Applied Bioinformatics facility at the University of Hohenheim, Germany, where they were frozen at -80°C until further processing. Tissue samples were shock frozen in liquid nitrogen and brought to the group Molecular Allergology of the Department of Nutritional Medicine and Prevention of the University of Hohenheim, Germany, for histopathological examinations and determination of relative cytokine gene expression levels.

C. Metabolite analyses of human blood and fecal samples

Serum concentrations of total cholesterol, triacylglycerides (TAG), low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C) were determined at the Institute for Clinical Chemistry and Laboratory Medicine of the University Medical Center Freiburg, Germany (KETO study) and at the Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway (CARBFUNC) according to standardized procedures. To track ketosis in CARBFUNC study participants, β-Hydroxybutyric acid (βHB) was measured in fasting plasma samples by gas-chromatography tandem mass spectrometry (GC-MS) (Horn et al., 2022). Fecal fatty acid, short-chain fatty acid (KETO study only) and fecal sterol/stanol concentrations (KETO and CARBFUNC study) were determined at the Department of Clinical Chemistry and Laboratory Medicine of the University Hospital Regensburg, Germany. Fecal samples were prepared as previously described by Schött el al. (Schött et al., 2018). In brief, up to 2 g of raw feces were homogenized with 2 x 2.5 mL 70%-isopropanol, using a gentleMACS™ Dissociator (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) and the dry weight (DW) determined for a 1 mL aliquot by overnight drying. Fecal homogenates were diluted to 2.0 mg DW/mL and stored at −80 °C. Short-chain fatty acids were determined by Liquid chromatography-mass spectrometry (LC-MS/MS) upon derivatization to 3-nitrophenylhydrazones (3NPH) (Liebisch et al., 2019) and other fecal fatty acids by GC-MS (Ecker et al., 2012). Fecal sterols and stanols were measured by LC-high resolution MS (LC-MS/HRMS) (Schött et al., 2018) for the KETO cohort and by triple quadrupole Gas Chromatography (GC)-MS/MS (Kunz & Matysik, 2019) for the CARBFUNC cohort. Both methods were validated and run with the same calibration solution resulting in comparable outcomes. For the CARBFUNC cohort, 17 samples did not meet the required concentration of 2.0 mg DW/mL, so they were excluded from the following analyses.

D. Histopathological examination and cytokine expression of murine colonic tissue

Histopathological examination and determination of cytokine expression was performed by the group of Molecular Allergology of the Department of Nutritional Medicine and Prevention of the University of Hohenheim, Germany. Gastrointestinal biopsies were taken of all mice after euthanizing and immediately shock-frozen in liquid nitrogen for RNA isolation or fixed in formaldehyde for embedding.

Intestinal inflammation was determined by Real-Time reverse transcription polymerase chain reaction (RT-PCR) as described by Bilotta et al. (Bilotta et al., 2021), to determine the expression of chemokine C ligand (CCL) 2 and CCL3, tumor necrosis factor (TNF), growth-regulated protein beta (Gro-beta, CXCL2), and transforming growth factor beta (TGFβ). The total RNA of proximal and distal colon biopsies was extracted and prepared with peqGOLD TriFast™ (VWR International GmbH, Erlangen, Germany) according to Hagenlocher et al. (Hagenlocher et al., 2016). The Real-Time RT reactions contained 1µL of template cDNA, 0.125µL each of sense and anti-sense primer (20 pmol stock solution, Table 3, page 36), 4µL H_2O, and 5µL SsoFast^{TM} EVAGreen Supermix (Bio-Rad Laboratories, Feldkirchen, Germany). Cytokine expression was measured using the CFX Connect Real-Time PCR System (CFX 2.1 software, Bio-Rad Laboratories, Feldkirchen, Germany) and expression profiles determined relative ($2^{-\Delta\Delta Ct}$) to the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

Table 3. Primer sequences and length used for relative cytokine expression.
Primer sequences are displayed in 5' to 3' direction.

Primer	Sense	Anti-sense	Length [bp]
CCL2	CTT CTG TGC CTG CTG CTC AT	CGG AGT TTG GGT TTG CTT GTC	273
CCL3	CTC TGC ATC ACT TGC TGC TGA CAC	CAC TCA GCT CCA GGT CGC TGA C	212
TNF	CAG ATA GAT GGG CTC ATA CCA GGG	GCC CTC TGG CCC AGG CAG TCA G	377
Gro-beta	AGT GAA CTG CGC TGT CAA TG	CTT CAG GGT CAA GGC AAA CT	154
TGFβ	GAG CTC TTC CAG ATA CTT CG	GTT GGA CTC TCT CCT CAA CA	150
GAPDH	TGG TCT CCT CTG ACT TCA AC	CCT GTT GCT GTA GCC AAA TT	128

For the histopathological examination, proximal and distal colonic tissues were fixed in 4% PBS-buffered formalin solution, embedded in paraffin and 5 µm sections each stained with hematoxylin&eosin according to Hagenlocher et al. (Hagenlocher et al., 2016) to determine tissue damage and the degree of cellular infiltration, with both scores increasing to elevated tissue alterations (Table 4, page 37). For each sample multiple cuts (proximal: 3 - 17, distal: 4 - 23) were prepared and examined at 100x or 200x magnification, resulting in an average score across all sections per biopsy.

Table 4. Scoring of histopathological examination for tissue damage and cell infiltration of colonic tissue samples.

Score	Tissue damage	Cell infiltration
0	undamaged mucosa	low number of inflammatory cells in the lamina propria
1	single lymphoepithelial damages	increased number of inflammatory cells in the lamina propria
2	surface damages of mucosa	accumulation of inflammatory cells and infiltration in submucosa
3	extensive mucosal damage and damage of deeper structures of the intestinal wall	transmural dispersion of infiltration of inflammatory cells

E. Metagenomic DNA extraction from fecal samples

Fecal samples of the KETO study were processed using a previously described combination of enzymatic digestion and mechanical disruption by bead beating (Von Rosenvinge et al., 2013). Briefly, 300 µL of the fecal slurry were centrifuged and the pellet dissolved in 800 µL enzyme mix A (5 µL of Lysozyme 10 mg/mL, 13 µL Mutanolysin 11.7 U/µL, 3.2 µL Lysostaphin 1 mg/mL, 778.8 µL 1x PBS) and transferred to a MP lysing matrix B tube (0.1 mm silica spheres, MP Biomedicals, Santa Ana, USA). The enzymatic digestion was initiated by incubation at 37°C for 30 minutes. A second enzymatic step was performed by adding 62 µL of enzyme mix B (10 µL Proteinkinase K 20 mg/mL, 50 µL SDS 10%, 2 µL RNase A 10 mg/mL) and incubation at 55°C for 45 minutes. Mechanical lysis was performed by bead beating at 6m s -f for 40 seconds (FastPrep-24, MP Biomedicals, Santa Ana, USA).

For the CARBFUNC study, 100-150 mg of the fecal samples were mechanically lysed for 40 seconds at 6m/s in MP lysing matrix B tube (0.1 mm silica spheres, MP Biomedicals, Santa Ana, USA, and FastPrep-24 5G, MP Biomedicals, Eschwege, Germany) containing 700 µL lysis buffer (Zymo Research, Irvine, USA). Metagenomic DNA was isolated from lysates using the ZR Fecal DNA Miniprep Kit (Zymo Research, Irvine, USA) according to the manufacturer's recommendation. The DNA was eluted in 100 µL RNase-free water and stored at -20°C.

For fecal murine samples, the same protocol for mechanical DNA isolation as for the CARBFUNC samples was used with slight modifications. The weight of fecal samples was determined via precision scale (ABJ120-4NM, Kern, Balingen, Germany) and the entire

fecal material per sample used for extraction. The fecal pellets were added to MP lysing matrix B tubes (0.1 mm silica spheres, MP Biomedicals, Santa Ana, USA) containing 700 µL lysis buffer (Zymo Research, Irvine, USA) and mechanically lysed 6 times for 45 seconds at 6m/s, with a 3-minute incubation on ice each (FastPrep-24 5G, MP Biomedicals, Eschwege, Germany). Thereafter, the ZR Fecal DNA Miniprep Kit (Zymo Research, Irvine, USA) according to the manufacturer's recommendation was used to extract the metagenomic DNA, which was eluted twice in 50 µL DNA Elution Buffer (Zymo Research, Irvine, USA) and stored at -20°C.

F. 16S rRNA gene amplification and sequencing

For the KETO cohort, the hypervariable V4 region of the 16S rRNA gene was amplified via polymerase chain reaction (PCR) using Golay-barcoded primers 515F and 806R (Caporaso et al., 2011), which were modified by adding 0-7 bp-long internal spacers as previously described (Fadrosh et al., 2014). For lists of primers, barcodes and spacer sequences used see Table S2. The PCR reaction was comprised of 10 µl 2x Phusion Master Mix (Thermo Fisher Scientific, Waltham, USA), 2.5 µL of each primer (final concentration 0.4 µM), 0.6 µL dimethyl sulfoxide (DMSO), and 4.4 µL template DNA and was carried out at 98°C for 2 min, with 30 cycles at 98°C for 10 s, 52°C for 15 s, and 72°C for 10 s and final extension at 72°C for 5 min. Equimolar amounts of all PCR products were extracted with the SequalPrep normalization plate kit 96 (Thermo Fisher Scientific, Waltham, USA), pooled and concentrated with the DNA Clean and Concentrator 5 kit (Zymo Research, Irvine, USA). Sequencing libraries were prepared using the TruSeq Nano DNA LT Library Prep kit (Illumina, San Diego, USA) according to the manufacturer's recommendations.

For the CARBFUNC cohort, as well as for the *IL-10*$^{-/-}$ mice study, 16S rRNA gene fragment amplification and barcoding were performed with the Quick 16S NGS Library Prep Kit (Zymo Research, Irvine, USA). Samples were diluted to an average DNA concentration of 20 ng/µL for amplification of the V3-V4 region of the 16S rRNA gene. After barcode addition samples were normalized to 30 ng per sample and pooled. As for both studies, samples numbers exceeded the maximum barcode limit, samples were sequenced on two consecutive sequencing runs, and therefore technical replicates and internal microbial standard communities were added. The final sequencing libraries were prepared according to the MiSeq System Denature and Dilute Libraries Guide (Illumina, San Diego, USA).

All libraries of the three studies were sequenced on the Illumina MiSeq instrument (MiSeq Reagent Kit v3, 600 cycles, Illumina, San Diego, USA) at the University of Hohenheim, Stuttgart, Germany.

G. Sequence processing

For the KETO study, raw sequence reads were trimmed with cutadapt v1.10 (Kechin et al., 2017) and barcodes extracted, paired-end reads merged and demultiplexed with QIIME v1.9.1 (Caporaso et al., 2010). Subsequent preprocessing steps were performed with QIIME2 v2019.7 (Bolyen et al., 2019), including open-reference operational taxonomic unit (OTU) picking based on 97% sequence similarity and classification of representative sequences with the q2-feature-classifier (Bokulich et al., 2018) against the SILVA database v132 (Quast et al., 2013). After chimera checking, 59% of sequences were retained. Singletons and sequencing artifacts were filtered if they contained less than 0.005% of total counts.

For the CARBFUNC study, raw sequences were processed with QIIME2 v2019.7 (Bolyen et al., 2019), including the DADA2 plugin (Callahan et al., 2016) for denoising, adapter trimming and chimera checking. A total of 93.8% (1st run) and 93.7% (2nd run) of all reads were retained after chimera checking, which were combined for subsequent analyses. Features of both runs were clustered based on 99% similarity with open-reference picking and representative sequences classified with the q2-feature-classifier (Bokulich et al., 2018) and mapped against the SILVA database v132 (Quast et al., 2013). Sequencing artifacts and singletons were filtered if they contained less than 0.0005% of total counts.

For the combined analysis of sequence data from both cohorts, all samples were rarefied to 1,650 sequences. Summaries of the resulting sequence data provided in Table S3. For the human cohorts additionally, taxonomic classifications, as well as a detailed list of all QIIME options, UNIX and R scripts used for sequence processing and data analysis is listed in Table S4.

As the *IL-10*$^{-/-}$ mice samples, were prepared similarly to the CARBFUNC cohort, preprocessing of raw sequence reads was also executed with QIIME2 v2019.7 (Bolyen et al., 2019) and the DADA2 plugin (Callahan et al., 2016), resulting in 77.09% reads being retained for the 1st and 74.36% for the 2nd library. After quality trimming of each sequencing run separately, both data sets were merged and processed simultaneously. Amplicon sequence variants (ASVs) were mapped against the SILVA database v132 (Quast et al., 2013) and sequencing artifacts filtered with less than 50 sequence counts (~ 0.0003% of

total counts). For diversity analyses, the sequence depth was rarefied to 10,000 sequences per sample.

H. Quantitative 16S rRNA gene expression analysis

For the *in vivo* study, the total bacterial load based on 16S rRNA gene copy numbers was determined for a subset of 120 fecal, murine samples (Table 1, page 31) via quantitative PCR (qPCR) using the commercial Femto Bacterial DNA Quantification Kit (Zymo Research, Irvine, USA) according to the manufacturer's recommendations. Samples were diluted 1:100 to obtain concentrations of 1-5 ng/µL as quantified by UV/Vis spectrophotometric determination (Nanodrop, Thermo Fisher Scientific, Waltham, USA) and both, samples and standards of known *Escherichia coli* strain JM109 genomic DNA concentrations, carried out in duplicates. Based on a standard curve, efficiencies between 87% to 92% were determined per amplification and mean DNA concentrations per sample duplicates calculated as equivalent of the *E. coli* genomic DNA. The number of 16S rRNA gene copies was quantified in reference to known genome length (5.5 Mbp) and number of 16S rRNA copies (7) per *E. coli* strain JM109 genome.

$$Nr.of\ genomes = \frac{DNA\ amount\ [ng] * Avogadro\ constant}{genome\ length\ [bp] * av.molecular\ weight\ [\frac{g}{mol} per\ bp]}$$

The 16S rRNA gene copy number (GCN) was normalized to one milligram of stool, as calculated based on the total weight of fecal matter used for mechanical lysis and the percentage of sample-lysis buffer solution subsequently used for DNA isolation.

$$\frac{GCN}{mg}stool = \frac{GCN\ in\ 50\mu L\ eluted\ DNA}{Total\ stool\ weight\ [mg] * (\%\ of\ sample\ suspension\ used * 100)}$$

I. Data Analyses

Data visualization and statistics were executed using R (v3.6.1) and the packages *vegan, biomformat, phyloseq, moments, nortest, lmerTest, emmeans, sjPlot, caret, logistf, pROC, randomForest and ComplexHeatmap*.

3. Microbial community assessment

a) Abundance

Relative abundances per taxon within a sample were calculated based on the sum of all taxon counts divided by the count of each taxon within each sample.

Absolute abundances per taxon within a sample were calculated based on the total 16S rRNA gene copy number determined for each sample multiplied by the relative abundance per taxon within each sample.

b) Diversity

The richness and variety of microbial communities is determined by alpha and beta diversity, with alpha diversity representing the diversity within a sample and beta diversity between samples. For both, various indices focusing on different characteristics of microbial communities can be used, e.g. presence/absence, or taking phylogeny and abundances into account.

Measurements of alpha diversity describe how many different microbial members are inherited in one community by considering only richness (e.g. the number of observed OTUs or ASVs) or incorporating their abundance, also referred to as evenness (e.g. Shannon- and Simpson-Index), all indicating the higher the number the greater the microbial diversity. (Kim et al., 2017; Wagner et al., 2018; Willis, 2019)

Beta Diversity determines dissimilarities between microbial communities, either based on presence/absence (e.g. Jaccard index) or abundance (e.g. Bray-Curtis Dissimilarity, Euclidean and Manhattan distance) or combining both respectively with phylogeny (e.g. weighted and unweighted UNIFRAC distances). (Knight et al., 2018; C. Lozupone & Knight, 2005; Magurran, 2021; Su, 2021) These comparisons are made for each sample pair within the dataset and result in multidimensional matrices, which can be simplified and visualized in Principal Coordinates Analysis (PCoA) plots, representing the dimensions explaining variances within the microbial community composition the best, and dissimilarities between the communities of samples pairs statistically being assessed by Analysis of similarities (ANOSIM) (Somerfield et al., 2021).

4. Comparisons of repeated measurements, technical replicates and sequence batches

Fecal samples of the KETO cohort were collected on two consecutive days, both before and after the intervention, to assess short-term intra-individual microbiota changes. Bray-Curtis dissimilarities within individuals were significantly lower between consecutive days than between time points ($p_{2\ days\ vs\ PRE|POST}$= 0.00028, $p_{2\ days\ vs\ POST|FOLLOW\text{-}UP}$= 0.02, $p_{2\ days\ vs\ PRE|FOLLOW\text{-}UP}$= 0.01, Wilcoxon rank-sum test [WRST] ^{FDR}, Figure S1). Therefore, to minimize the influence of short-term fluctuations, measurements from consecutive days were merged by using mean read counts.

Samples from the CARBFUNC study were sequenced in two separate sequencing runs. To control for batch effects, technical replicates were visually inspected based on PCoA plots of different β-diversity metrics, including weighted/unweighted UniFrac distances, Jaccard index, Euclidean and Manhattan distance (Figure S1). Bray-Curtis dissimilarities were significantly lower between technical replicates from different sequencing runs than between samples collected from the same individual at different time points ($p \leq$ 2e-16, WRST), and overall microbiota composition was not different between sequencing runs (R=0.009, p=0.29, ANOSIM), as opposed to between individuals (R= 0.98, p= 0.001, ANOSIM, Figure S1). For all subsequent analyses, the means of the technical replicates were used.

For fecal and serum lipid profiles from the CARBFUNC study, mean values were calculated from samples collected at three and six months of the dietary interventions.

Fecal samples of the murine *IL-10* knockout trial were also sequenced on two sequencing runs being controlled for by technical replicates, including different mice and mice of different time points. The overall microbial composition, as assessed based on the Bray-Curtis dissimilarity, was significantly and totally explained by the individual (R= 1, p= 0.001, ANOSIM) and not by sequencing batch (R= -0.061, p= 0.798, ANOSIM, Figure S1). Therefore, the means of technical replicates were calculated and used for subsequent analyses.

5. Statistical Analysis

Normal distribution was evaluated using Anderson-Darling and Shapiro-Wilks tests. Non-normally distributed parameters were analyzed by non-parametric tests, such as pairwise Wilcoxon rank-sum test (WRST) for group comparisons, Spearman's rank correlation (SRC) test for correlation analyses, and the chi-squared test for categorical comparisons. Corrections for false discovery rates (FDR) were performed with the Benjamini-Hochberg (BH) procedure. Unless indicated otherwise, boxplots show medians and corresponding 95% confidence intervals (CI) and significance thresholds with $p/q > 0.05$ ns, $p/q \leq 0.05$ *, $p/q \leq 0.01$ **, $p/q \leq 0.001$ ***. N values for all statistical tests and significance thresholds are listed in figure legends or are referred to in Supplemental Tables.

For differential abundance analyses and abundance correlations, pseudocounts were added to relative (0.0001 relative abundance) and absolute abundances (taxon count of 1) for each taxon if not indicated otherwise. This addition is a common method in microbiome studies which enables to include samples without reads assigned during sequencing as

classifications of sequences, simply not being present in a sample or if only too few reads to validly detect them were contained, cannot be made (Costea et al., 2014; S. Weiss et al., 2017).

a) Generalized Linear Mixed Effect Models

Generalized Linear Mixed Models (GLMMs) are extended linear models that, in addition to fixed effects, also contain random effects for linear prediction (Dean et al., 2007). The addition of random effects helps to control predictions when the response variable is clustered, e.g. longitudinal data with repeated measurements of the same individuals, or potential confounding factors such as litter or cage assignments in *in vivo* studies. Variations within the response variable, which are explained by the random effects, are quantified and predictions for the explanatory variables, also referred to as fixed effects in GLMMs, calculated accordingly (Bolker et al., 2009). Further, various data types like categorical, binary, continuous or count data, as well as nonnormally distributed data can be included by the use of link functions and exponential family distributions, e.g. normal, Poisson or binomial distribution, which is most commonly the case for microbiome data.

For all GLMMs, fits were assessed using diagnostic plots, and for model and EMMs comparisons, significance was determined after BH-based false discovery rate correction with the following thresholds for q-values: q>0.1 ns, q≤ 0.1 *, q≤ 0.05 **, q≤ 0.01 *** and q≤ 0.001 **** as not indicated otherwise. Odds ratios (OR) with 95% CI, as well as marginal and conditional R^2 were calculated for all significant linear models (Table S5 for human cohorts and Table S6 and S7 for the murine cohort).

(1) Human cohorts

For the clinical trials, associations between lipid and microbiota compositions before and after intervention were determined with GLMMs. First, taxon associations with fecal sterol conversion and blood lipids were assessed across both cohorts before dietary intervention. Only taxa with a relative abundance of >0.1% of all reads from either all high-converter samples combined or all low-converter samples combined, were considered, a pseudocount of 1 added for zero values, and the resulting relative abundances centered-log transformed. To control for cohort-specific biases, the KETO or CARBFUNC cohort grouping was included as a random effect in the GLMMs and also evaluated as a fixed effect in multiple linear fixed effects models (Tables S5). Additionally, sex was included as fixed effect to control for gender biases and only results considered, which were not associated with these confounders.

To further determine non-linear associations between the sterol converter type and the taxonomic microbiota composition, a random forest (RF) classifier was trained to predict the class of cholesterol-to-coprostanol converter type (high/low) using clr-transformed relative taxa abundances. The model performance was evaluated based on different numbers of trees (250, 500, 1000 and 2000), as well as different numbers of variables tried at each split (mtry) with values ranging from 1 to 20. The final RF classifier was run with 501 trees, 12 variables tried at each split, and a default node-size. Leave-one-out cross-validation (LOOCV) was used to evaluate the model accuracy based on recall and precision and to identify the most important features based on the mean decrease in accuracy.

Alterations in the relative abundance of microbial taxa due to the LCHF diets across both cohorts were determined in another GLMM, which was controlled for cohort and sex-specific effects and intra-individual variation as a random effect to control for repeated measurements. Relative abundances of taxa were centered-log transformed after pseudocount addition similar as described above. Microbial taxa with a relative abundance above 0.5% and which are detected in at least 4 samples, with a fold change < -0.25 or > 0.25 as determined based on estimated marginal means (EMMs) before and after the intervention, and significance of $q \leq 0.1$ based on FDR corrected Tukey test, for which no cohort or gender-bias was detected, were considered (Table S5).

Changes in the relative abundance of sterol conversion associated microbial taxa, as well as in fecal sterols and serum lipids, in response to both, the dietary intervention and the sterol conversion type, were evaluated in GLMMs with dietary intervention and the pre-existing sterol conversion type as interactive fixed effects. Similar to previous models, cohort and sex-associations were controlled for as fixed effects as well as repeated measurements as random effects per individuum. Relative abundances of taxa were centered-log transformed after pseudocount addition similar as described above (Table S5). For serum lipids and fecal sterols, different models (Gaussian and Gamma distribution) and link functions (identity, log, inverse) were compared based on Akaike information criterion (AIC) and diagnostic plots, and the best fit was determined for a log-transformed Gaussian distribution. To achieve normal distribution and improve skewness and kurtosis, pseudocounts of 0.1 instead of zero values were added for serum and fecal metabolites and concentrations log-transformed (Table S5). Differences between subgroups were further assessed by calculation of EMMs and post hoc comparisons by Tukey test. As serum lipid levels were associated with cohorts in these models, the same interaction

GLMM was also applied for each cohort separately with the exclusion of cohort as fixed effect (Table S5).

(2) Murine cohort

To identify relative and absolute abundance alterations due to colitis, shortly (up to four days prior) before and after symptomatic observations, a GLMM with the health state (healthy, pre/post colitis and colitis), litter and cage numbers as fixed effects, as well as repeated measurements per subject as random effect was carried out. Only samples after the first week of cage allocation were used to minimize allocation effects and only taxa with a relative abundance > 0.5% in one sample and that were contained in at least 3 samples were included. As described above, for relative abundances sequence counts per genus were centered-log transformed after pseudocount addition (Table S6). For absolute abundances, due to the high order of magnitude for count data and the high number of zero counts, binomial and Poisson models resulted in bad model fits. To increase the model fit, also absolute abundances were centered-log transformed after a pseudocount addition resulting in continuous data, which was then applied to the same GLMM as for the relative abundance (Table S7).

The extent to which associated factors influenced if mice had to be withdrawn from the observation due to colitis-related symptoms before 20 weeks was assessed by Firth logistic regression, which is based on the penalized likelihood estimation method (Firth, 1993). Compared to standard maximum likelihood logistic regression, a penalty term is added to the model which converges towards zero as the sample size increases to infinite observations. Thereby it accounts for issues of small sample sizes, separation and bias of prediction estimators (Firth, 1993), which applied for the murine data set with a sample size of 22 mice and quasi-complete separation of the data due to mice only being early withdrawn in litter 2 and 4. First, all previously associated factors that affect colitis development were modelled separately as single estimators and compared to the baseline line model, which is solely based on the intercept, via Likelihood ratio test. Single estimators included: litter, genotype, co-housing, differences in microbial diversity (Shannon Index and Number of observed ASVs) within and between communities, and the presence of *Akkermansia* at the day of cage allocation as well as in the first sample associated with colitis-related symptoms and up to four days prior to their observation. The presence/absence of *Akkermansia* was determined based on a relative abundance >0 (present) or equal to 0 (absent) in the respective sample revealing that the observations up to four days prior to symptoms and on the first day of symptoms resulted in identical

categorization and were thereby included only once thereafter being representative for both. Dissimilarities between communities were calculated as the mean Bray-Curtis dissimilarity between each IL-10 deficient and every wild type mouse at the day of cage allocation. The only variables that were significantly associated in the single estimator models were litter, the mean Bray-Curtis dissimilarity to wild type mice and the presence of *Akkermansia* shortly before/at the first day of symptomatic observation (Table S8). Building the final model with the three significantly associated variables resulted in a large number of probabilities being close to zero and one whenever the mean Bray-Curtis dissimilarity was added indicating a very high, almost perfect prediction, which could either indicate a statistical issue or a true biological result, both caused by interdependencies between predictors. To account for both possibilities, two final models were built, the first including litter and *Akkermansia* presence in the first sample with symptomatic observations only, and the second with all three associated predictors (Table S9). The degree of *in-sample* prediction of the logistic regression was assess by plotting the true positive rate (sensitivity) against the false positive rate (specificity) in a receiver operating characteristic (ROC) curve and calculating the area under the curve (AUC). To further validate the predictive *out-of-sample* power, meaning to forecast observations that were not part of the trained model, a stratified 10-fold cross validation splitting the dataset in a training (80%) and testing (20%) set in the same proportion as samples appeared in the population was performed. The final cross validated model was picked based on the highest accuracy measured by AIC (Tables S10).

III. Results

A. Characterization of the microbiota and sterol conversion in response to low-carbohydrate high-fat dietary intervention in two human cohorts

Interindividual variations in serum lipid responses to low-carbohydrate high-fat (LCHF) diets have been increasingly reported and the microbiota is being considered as a potential cause (Burén et al., 2021; Descamps et al., 2015; Griffin et al., 2021; Norwitz et al., 2022; O'Neill et al., 2020; Qamar et al., 2019). Cholesterol, a major key metabolite in host lipid metabolism and cardiovascular risk factor when increased in blood, is converted by the intestinal microbiota into non-absorbable coprostanol. To investigate the relationship between intestinal cholesterol-to-coprostanol conversion and the gut microbiota, as well as their link to serum lipid profiles in the context of LCHF diets, fecal and blood samples of two geographically separate human cohorts with different metabolic backgrounds were examined after six weeks or three and six months of intervention, respectively. The results of these analyses are currently under revision at *Cell Press Community*.

1. Equal distributions of cholesterol high and low-converter types among humans with and without obesity

Fecal sterol and stanol concentrations were determined by liquid chromatography-high resolution mass spectrometry (LC-MS/HRMS) in samples from 28 healthy, normal-weight participants of a German ketogenic (KETO) diet study (Urbain et al., 2017). Before dietary intervention, the microbial conversion products coprostanol (18.71 ± 13.44 nmol/mg dry weight [DW], data are shown as mean ± standard deviation [sd]) and stigmastanol (5β-sitostanol) (10.39 ± 6.71 nmol/mg DW) were the most abundant fecal stanols, followed by their animal and plant-derived sterol precursors, fecal cholesterol (15.76 ± 17.53 nmol/mg DW), β-sitosterol (4.62 ± 3.99 nmol/mg DW) and campesterol (1.43 ± 1.51 nmol/mg DW) (Figure 2a, page 49). All sterol and stanol concentrations exhibited substantial inter-individual variation and, as would be expected from a direct metabolic dependency, coprostanol and cholesterol levels were negatively correlated (Figure 2b, page 49, $R = -0.52$, $q = 0.005$, Spearman's rank correlation [SRC] corrected for false discovery rate [FDR] with the Benjamini-Hochberg procedure [SRC^{FDR}]), similarly to stigmastanol and β-sitosterol levels (Figure 2c, page 49, $R = -0.71$, $q < 1e\text{-}5$), as well as 5β-campestanol and campesterol levels (Figure 2d, page 49, $R = -0.37$, $q = 0.073$). The KETO study participants showed a bimodal distribution into high and low cholesterol-to-coprostanol converter types, as previously described for healthy individuals (Veiga et al., 2005; Wilkins et al., 1974), i.e., a greater fraction of high (coprostanol/cholesterol > 2, 61 % of the KETO study participants)

compared to low (coprostanol/cholesterol < 0.5, 14 % of the KETO study participants) cholesterol converters (Figure 2b, page 49).

To determine cholesterol high and low-converter types among individuals with different metabolic health backgrounds, fecal sterol and stanol concentrations were also characterized in 145 individuals with obesity from the CARBFUNC study, a Norwegian 2-year randomized controlled dietary intervention trial (Horn et al., 2022). The main characteristics of both study cohorts (Table S11) revealed the Norwegian cohort to be slightly older (37 ± 11.73 vs 43 ± 8.32 years). Individuals with obesity from the CARBFUNC study exhibited increased concentrations of fecal coprostanol (Table S11, $p < 1e\text{-}3$, Wilcoxon rank-sum test [WRST]), but otherwise similar sterol or stanol levels ($p > 0.05$) and compositional profiles (Figure 2e, page 49), including a negative correlation of coprostanol to cholesterol ($R = -0.47$, $q < 1e\text{-}9$) and a bimodal distribution of high (61 % of CARBFUNC study participants) and low (21 % of CARBFUNC study participants) cholesterol converter types based on coprostanol/cholesterol ratios (Figure 2f, page 49). Stratification of individuals into a larger cholesterol high-converter type fraction and a smaller low-converter type fraction based on the fecal coprostanol/cholesterol ratio therefore appears to be a conserved feature of the human fecal microbiome in individuals with and without obesity.

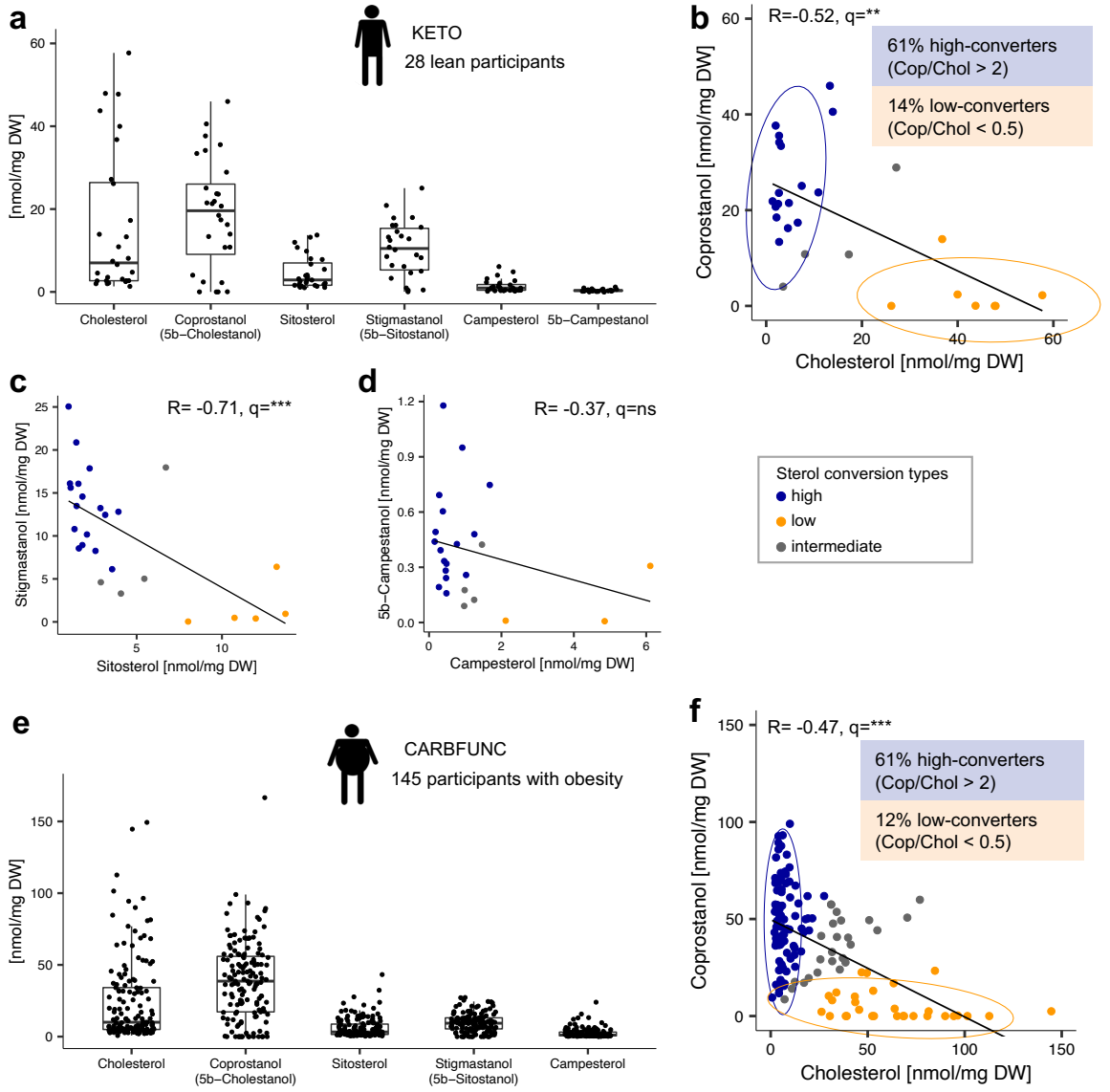


Figure 2. Equal distributions of cholesterol high and low-converter types among humans with and without obesity.

(a) Sterol and stanol concentrations as determined by LC-MS/HRMS in 28 fecal samples from the KETO study participants before dietary intervention.

(b) Negative correlation (Spearman's rank) between fecal coprostanol and cholesterol concentrations and bimodal distribution of cholesterol high (n = 17), intermediate (n = 4) and low-converter (n = 7) types, as classified based on the fecal coprostanol/cholesterol ratio.

(c) and (d) Comparable negative correlations (Spearman's rank) between the fecal concentrations of the phytosterols sitosterol and campesterol and the corresponding stanol conversion products stigmastanol (n = 26), and 5β-campestanol (n = 23).

(e) Similar fecal sterol and stanol concentration profiles in individuals with obesity from the CARBFUNC study before dietary intervention (n = 145 samples), compared to lean KETO study participants (n = 89/26/30 for cholesterol high/intermediate/low-converters).

(f) Negative correlation (Spearman's rank) of fecal coprostanol and cholesterol concentrations in CARBFUNC study participants and bimodal distribution into high and low-converter types.

Spearman's rank correlation, Benjamini-Hochberg (BH) corrected: $q > 0.05$ ns, $q \leq 0.01$ **, $q \leq 0.001$ ***

2. Distinct microbiota associations with fecal cholesterol and coprostanol

To better understand the intra-intestinal relationship of cholesterol-to-coprostanol conversion with the gut microbiota, fecal taxonomic microbiota and metabolite profiles were compared between cholesterol high and low-converters from the KETO and CARBFUNC studies. 16S rRNA sequencing resulted in 10,578,956 (KETO) and 14,910,136 (CARBFUNC) raw single reads respectively for the two cohorts. After processing and quality evaluation a total 5,518,846 (KETO) and 9,395,506 (CARBFUNC) single reads assigned to 384 (KETO) and 518 (CARBFUNC) amplicon sequence variants (ASVs) were obtained (Table S3). No difference in fecal microbiota α-diversity (Figure 3a, page 52, Shannon index, $p = 0.17$, WRST) or β-diversity (Figure 3b, page 52, Bray-Curtis dissimilarity, $R = 0.084$, $p = 0.15$, ANOSIM) was detected between lean cholesterol high and low-converters from the KETO study. However, low-converters with obesity from the CARBFUNC study had a reduced fecal microbiota α-diversity (Figure 3a, page 52, Shannon index, $p < 1e\text{-}9$, WRST) and distinct β-diversity (Figure 3b, page 52, Bray-Curtis dissimilarity, $R = 0.24$, $p = 0.001$, ANOSIM) relative to high-converters, suggesting that the microbiota relationship to cholesterol conversion was affected by obesity or other CARBFUNC cohort-specific factors.

Next, a generalized linear mixed effect model (GLMM) was used to identify shared linear associations between specific members of the fecal microbiota and fecal cholesterol and coprostanol concentrations across both the KETO and CARBFUNC studies combined (Table S5). Whereas intestinal coprostanol should be entirely produced from microbial cholesterol reduction, fecal cholesterol should originate from both endogenous and exogenous sources and therefore be more intricately regulated (Juste et al., 2021). Fecal coprostanol and cholesterol concentrations were therefore queried both independently and in combination for associations with the centered log-ratio (clr)-transformed relative abundance of specific bacterial taxa, while controlling in the corresponding GLMM for cohort and gender-specific effects (Figure 3c, page 52, Figure S2, Tables S5). Significant associations from the GLMMs ($q \leq 0.1$) were independently assessed by Spearman's rank correlation analysis (Figure 3d, e, page 52, Figure S2, SCR^{FDR}). Only two bacterial taxa, i.e. *Eubacterium coprostanoligenes* group and *Ruminococcaceae* UCG 014, showed consistent associations with cholesterol-to-coprostanol conversion across all comparisons and analyses, i.e. positive and negative correlations to fecal coprostanol and cholesterol levels, respectively (Figure 3c, d, page 54), and a positive association with the coprostanol/cholesterol ratio (Figure S2). In addition, *Lachnoclostridium* was positively and

Clostridiales XIII AD3011 negatively correlated to fecal cholesterol concentrations (Figure 3c, e, page 52, SCR^{FDR}). Detection of *Clostridiales* XIII AD3011 was limited to the CARBFUNC cohort (Figure 3e, page 52), indicating an obesity or geography-specific association of this genus with fecal cholesterol levels.

To also test for non-linear microbiota associations with the cholesterol high and low-converter types, a random forest classifier was trained on clr-transformed microbiota compositions. This classifier performed well at identifying cholesterol high-converters (84.82% precision, 94.06% recall), but lacked sensitivity for the detection of low-converters (76.92% precision, 54.05% recall). Leave-one-out cross-validation (LOOCV) identified *E. coprostanoligenes* group as the most important microbiota feature for this classification, but also several taxa, such as *Christensenellaceae.R.7* group, which the GLMM associated with fecal cholesterol and flagged with singularity fit warnings, indicating a potential overfit of the linear model (Fig. 3f, g, page 52). In summary, linear and non-linear models consistently linked *E. coprostanoligenes* to intestinal cholesterol-to-coprostanol conversion across both cohorts, but other bacterial taxa that mostly showed associations with fecal cholesterol levels may also be involved in the process.

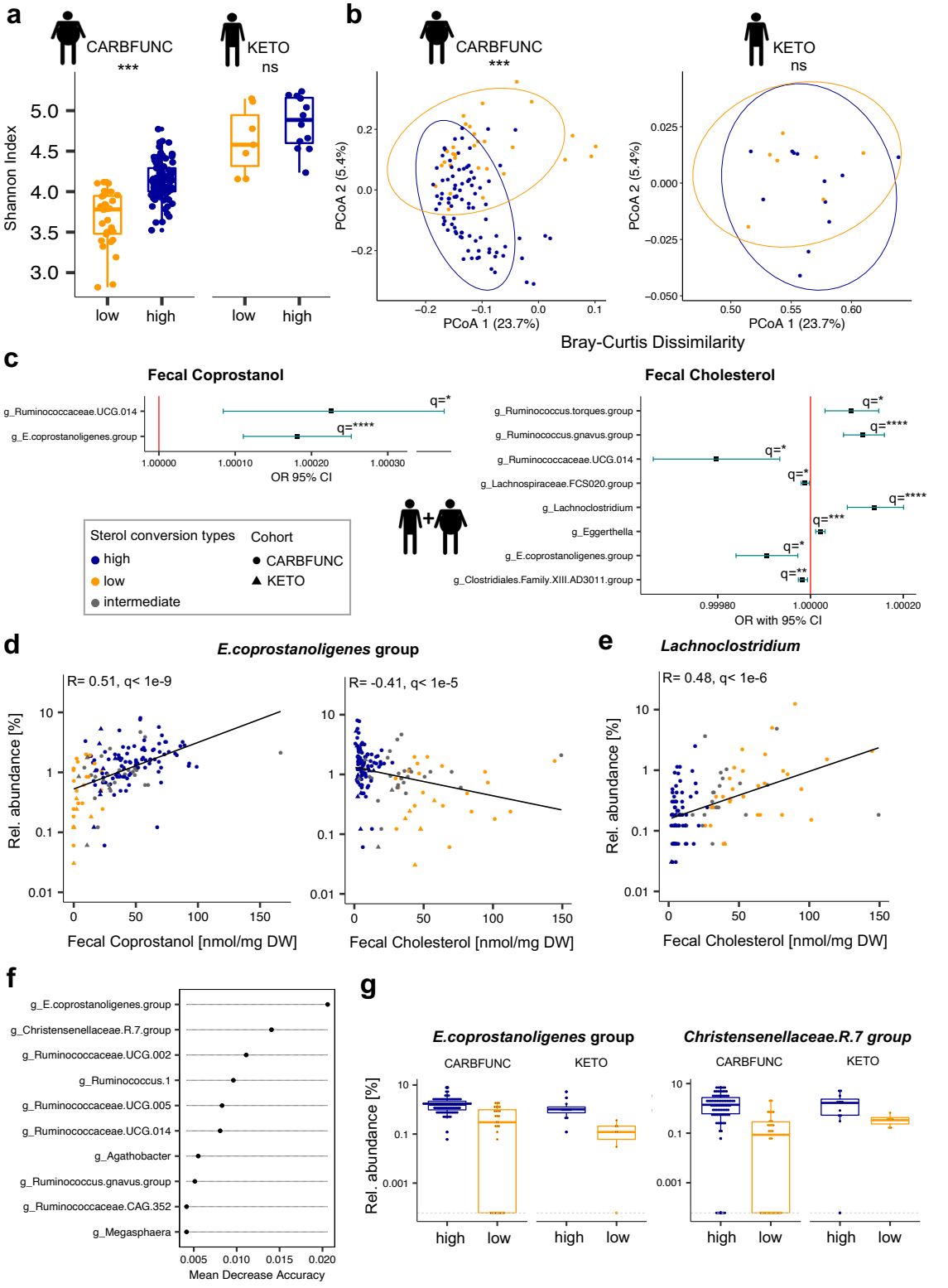


Figure 3. Distinct microbiota associations with fecal cholesterol and coprostanol. (a) and (b) Reduced diversity (Wilcoxon rank-sum) and altered composition (ANOSIM) of taxonomic microbiota profiles of cholesterol high (n = 89) compared to low-converters (n = 30) with obesity from

the CARBFUNC study, but no difference between lean cholesterol high (n = 7) and low-converters (n = 12) from the KETO study. $p > 0.05$ ns, $p < 0.001$ ***
(c) Positive and negative associations of bacterial taxa with fecal coprostanol and cholesterol concentrations, as identified by a generalized linear mixed effect model (GLMM) for the combined dataset of KETO (n = 23) and CARBFUNC (n = 145) study participants. For the GLMM input, zero values were replaced with a pseudocount and cohort and gender added as random and fixed effects (see Methods for details).
(d) and (e) Across both cohorts combined, *Eubacterium coprostanoligenes* group (n = 158) was positively correlated with fecal coprostanol and negatively correlated with fecal cholesterol concentrations, whereas *Lachnoclostridium* (n = 125), was positively correlated with fecal cholesterol concentrations.
(f) *E. coprostanoligenes* group relative abundance was the most informative microbiota feature for predicting the cholesterol converter type with a random forest model, based on leave-on-out cross-validation
(g) Increased relative abundance of *E. coprostanoligenes* group and *Christensenellaceae.R.7* group in cholesterol high-converters. Dashed lines indicate pseudocount values (0.0001 % relative abundance) of samples with zero taxon counts. Benjamini-Hochberg (BH) corrected: $q > 0.1$ ns, $q \leq 0.1$ *, $q \leq 0.05$ **, $q \leq 0.01$ ***, $q \leq 0.001$ ****

Next, fecal cholesterol and coprostanol levels were compared to short-chain fatty acid (SCFAs) concentrations in stool samples. Straight-chain SCFAs are mainly produced by microbial fermentation of non-digestible dietary fiber in the colon, whereas branched-chain SCFAs predominantly result from microbial protein fermentation (Wolter et al., 2021). Fecal cholesterol but not coprostanol was positively correlated across both studies to the straight-chain SCFAs acetate ($R = 0.38$, $q < 1e\text{-}4$), propionate ($R = 0.52$, $q < 1e\text{-}9$) and butyrate ($R = 0.33$, $q < 1e\text{-}3$, Figure 4a, page 54, SRC^{FDR}). In contrast, fecal coprostanol but not cholesterol, showed a positive correlation to the branched-chain SCFA isobutyrate (Figure 4b, page 54, $R = 0.25$, $q = 0.005$, SRC^{FDR}). An association of the cholesterol converter type with fecal SCFA levels was only identified in study participants with obesity (CARBFUNC), including decreased fecal propionate ($p < 1e\text{-}4$, WRST) and increased isobutyrate ($p = 0.023$, WRST) concentrations in high-converters. Thus, our findings are in agreement with previous reports of increased straight-chain SCFA secretion in cholesterol low-converters (Matysik et al., 2021). They indicate distinct intra-intestinal associations of cholesterol and coprostanol with specific microbial taxa and metabolites, which may be influenced by obesity or other cohort-specific parameters.

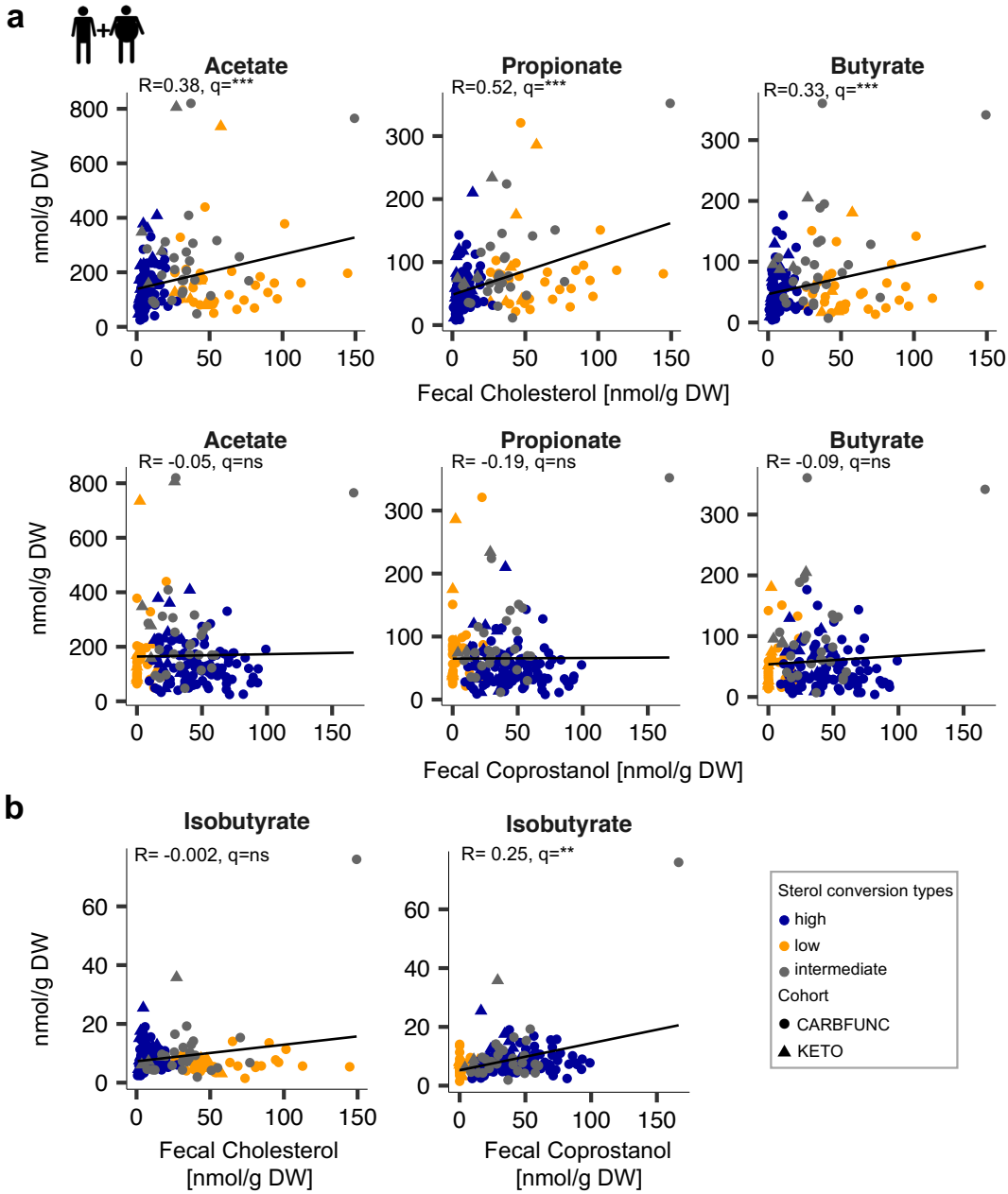


Figure 4. Distinct associations of fecal cholesterol and coprostanol concentrations with straight and branched SCFAs.
(a, b) Positive correlation of fecal cholesterol with the concentrations of the straight-chain SCFAs acetate, propionate and butyrate (a) and of fecal coprostanol with the branched-chain SCFA isobutyrate (b) in fecal samples from lean KETO study participants (n=28) and in individuals with obesity from the CARBFUNC study (n=145) before the dietary intervention. Spearman's rank correlation, BH-corrected: $q > 0.05$ ns, $q \leq 0.05$ *, $q \leq 0.01$ **, $q \leq 0.001$ ***

3. Circulating blood lipids in cholesterol high and low-converters

To determine if intestinal cholesterol-to-coprostanol conversion had a discernible impact on circulating cholesterol levels, fecal cholesterol and coprostanol concentrations with the serum levels of total cholesterol and other lipids were compared. Lean (KETO cohort) cholesterol high and low-converters showed comparable total cholesterol and low-density lipoprotein cholesterol (LDL-C) concentrations before the intervention ($p > 0.05$), but the lean cholesterol high-converters had increased triacylglycerides (TAG, $p = 0.04$) and decreased high-density lipoprotein cholesterol (HDL-C, $p = 0.04$) levels (Figure 5a, page 56, WRST). No difference between cholesterol high and low-converters with obesity (CARBFUNC cohort) was detected in serum total cholesterol, TAG, HDL-C and LDL-C levels (Figure 5b, page 56, $p > 0.05$, WRST). However, CARBFUNC study participants had increased serum TAG ($p < 1e\text{-}7$), LDL-C ($p = 0.03$) and reduced serum HDL-C ($p < 1e\text{-}13$) levels compared to KETO study participants and showed an increased TAG/HDL-C ratio ($p < 1e\text{-}11$), a marker for insulin resistance (Kim-Dorner et al., 2010), consistent with generally adverse health profiles (Table S11, WRST). In line with this, blood glucose and insulin levels were increased in CARBFUNC study participants (Table S11), but showed neither significant differences between cholesterol high and low-converters from either study (KETO: blood glucose $p = 0.092$, insulin $p = 0.77$; CARBFUNC: blood glucose $p = 0.66$, insulin $p = 0.14$, WRST), nor correlations with fecal coprostanol or cholesterol ($q > 0.05$, SCR^{FDR}, Figure S2). However, compared to cholesterol low-converters, high-converters from the CARBFUNC study had elevated serum β-hydroxybutyric acid (BHB) levels (low: 39.2 ± 33.8 vs high: 61.6 ± 59.2, $p = 0.013$, mean $\pm$ sd, WRST), indicative of increased ketosis in the cholesterol high-converters with obesity.

Neither fecal coprostanol nor cholesterol levels were significantly correlated with serum total cholesterol, TAG, LDL-C, or HDL-C in either the CARBFUNC or KETO cohorts, or individuals from both studies combined (Figure S3, $q > 0.05$, SRC^{FDR}). Similarly, no significant associations between fecal microbial taxa and serum lipid levels were identified by the GLMM after false discovery rate correction (Table S5, $q > 0.1$). The findings therefore provide no indication for a reducing effect of intestinal cholesterol-to-coprostanol conversion on circulating total cholesterol levels.

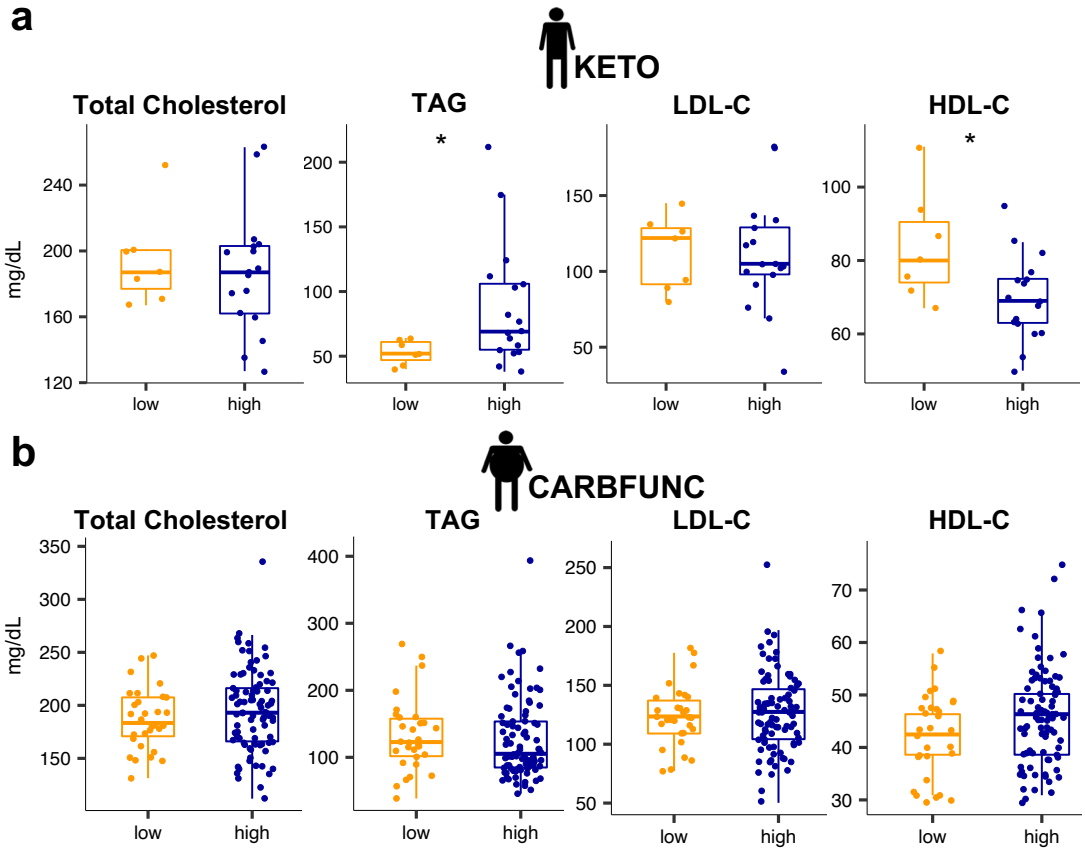


Figure 5. Circulating blood lipids in cholesterol high and low-converters.
(a) Comparable total cholesterol and low-density lipoprotein cholesterol (LDL-C), but increased serum triacylglycerides (TAG) and decreased high-density lipoprotein cholesterol (HDL-C) levels in lean cholesterol low (n = 7) compared to high-converters (n = 17) from the KETO study. (b) No significant difference in blood lipid levels between cholesterol high (n = 89) and low-converters (n = 30) with obesity from the CARBFUNC study. Wilcoxon rank-sum test, $p > 0.05$ ns, $p \leq 0.05$ *

4. Diet impact on cholesterol-to-coprostanol conversion

To identify associations between dietary habits and cholesterol-to-coprostanol conversion, lean KETO study participants were compared based on available food questionnaire data. Caloric intake from fats, fiber, carbohydrates, protein or cholesterol was comparable in lean cholesterol high and low-converters (Table 5, page 58, $p > 0.05$, WRST) and both converter types exhibited similar fecal fatty acid profiles, in terms of chain length and saturation level (Figure S4, $p > 0.05$, WRST). The ratio of animal to plant-derived fat intake was estimated for both cohorts, based on the ratio of fecal coprostanol to stigmastanol, which originate from microbial conversion of cholesterol and the phytosterol β-sitosterol, respectively (Juste et al., 2021). While no difference was detected between lean high and low-converters (Figure 6a, page 61, KETO cohort, $p > 0.05$, WRST), the fecal coprostanol/stigmastanol ratio was increased in cholesterol high-converters with obesity (Figure 6a, page 61, CARBFUNC cohort: $p < 1e\text{-}5$, WRST), indicating that, compared to low-converters, high-converters with obesity obtained a larger fraction of their fat intake from animal sources.

Table 5. Semi-quantitative food questionnaire-based dietary habits of lean KETO study participants.
Data were collected over seven consecutive days one to two weeks before (PRE) and during the last week of dietary intervention (LCHF) and used to calculate average daily intakes (mean ± standard deviations [sd]) and determine significant differences between before and on LCHF time points (paired Wilcoxon signed-rank test) and between high/low-converter types (unpaired Wilcoxon rank-sum test); n(low)=6, n(high)=17, Abbreviations: SFA = saturated fats, MUFA = mono-unsaturated fats, PUFA = poly-unsaturated fats, LCHF = low-carbohydrate high-fat

	PRE			LCHF			PRE vs LCHF	
Dietary intake	Low-converter	High-converter	Low vs High	Low-converter	High-converter	Low vs High	Low-converter	High-converter
	mean ± sd	mean ± sd	p_{PRE}	mean ± sd	mean ± sd	p_{LCHF}	$p_{PREvs\ LCHF}$	$p_{PREvs\ LCHF}$
Total kcal	2380.32 ± 394.11	2223.66 ± 454.11	0.43	2586.78 ± 835.78	2066.96 ± 425.61	0.14	0.43	0.24
%E								
Protein	13.56 ± 1.08	14.59 ± 2.21	0.35	19.83 ± 3.70	18.96 ± 3.55	0.81	0.03	<1e-3
Carbohydrate	42.16 ± 7.22	41.47 ± 6.15	0.92	5.81 ± 1.82	8.54 ± 3.09	0.06	0.03	<1e-4
Fat	40.13 ± 8.04	37.95 ± 5.30	0.66	72.15 ± 5.28	69.75 ± 5.85	0.56	0.03	<1e-4
SFA	17.17 ± 4.23	15.69 ± 2.40	0.52	26.71 ± 3.69	26.90 ± 3.96	0.92	0.03	<1e-4
MUFA	12.86 ± 3.54	12.74 ± 2.40	0.92	25.73 ± 3.35	26.20 ± 5.19	1.0	0.03	<1e-4
PUFA	5.09 ± .57	5.51 ± 1.62	0.61	9.54 ± 1.45	9.74 ± 2.99	0.81	0.03	<1e-4
Cholesterol [mg/d]	325.4 ± 165.9	304.5 ± 108.9	0.92	569.1 ± 141.5	429.8 ± 100.5	0.02	0.03	0.001
Fiber [g/d]	24.6 ± 6.2	27.5 ± 14.1	0.97	26.5 ± 11.6	20.8 ± 7.1	0.29	0.84	0.02

Both the KETO and CARBFUNC studies involved interventions with low-carbohydrate high-fat (LCHF) diets, based on ≥ 75 energy percent [E%] fat and ≤ 10 E% carbohydrate intake. Lean KETO study participants followed a 6-week *ad libitum* LCHF diet, which resulted in increased urinary and blood ketone bodies, as well as other hormonal and metabolic changes indicative of ketosis, as previously described in detail (Urbain et al., 2017; Urbain & Bertz, 2016). CARBFUNC study participants with obesity were restricted to a normocaloric LCHF diet (males: 2,500 kcal, females: 2,000 kcal, Horn et al., 2022), which was accompanied by at least transient ketosis at three months of the intervention based on increased serum BHB levels (PRE: 62.68 µM ± 68.67 vs. 3 months: 264.01 µM ± 251.74, p = 0.00031; PRE vs. 6 months: 163.26 ± 274.60, p = 0.26; mean ± sd, paired Wilcoxon signed-rank test). Fecal and serum samples were collected after six weeks (KETO study) and three and six months (CARBFUNC study) and used for microbiota analysis and lipid profiling. As both studies included individuals with different metabolic health backgrounds and involved variable time spans, first, if comparable taxonomic compositional microbiota alterations could be detected in both cohorts after LCHF diet intervention was tested. A strong and consistent shift in microbiota compositions was detected across both cohorts (Figure 6b, page 61, Table S5, GLMM), including changes in the relative abundance of bacterial genera, such as *Bifidobacterium* (mean reduction: -4.23 % ± 2.48), previously reported to be altered by ketogenic diet (Ang et al., 2020). The microbiota analysis therefore demonstrates reproducible, temporally stable LCHF diet-induced compositional microbiota changes in individuals with and without obesity.

To test whether intestinal cholesterol-to-coprostanol conversion could be dietarily modulated, we compared fecal cholesterol and coprostanol levels in KETO and CARBFUNC study participants in response to the LCHF diets (Figure 6c, page 61). The GLMMs identified cohort-specific effects for both fecal cholesterol and coprostanol concentrations but not their ratios (Figure 6d, page 61, Table S5). In cholesterol low-converters from both cohorts, LCHF diet increased cholesterol-to-coprostanol conversion, as illustrated by decreased fecal cholesterol (q < 1e-7) and increased fecal coprostanol (q < 1e-3) levels, as well as increased coprostanol/cholesterol ratios (Figure 6d, page 61, Table S5, q < 1e-9, Tukey's Test^{FDR} comparisons of GLMM based estimated marginal means [EMM]). The increased cholesterol conversion in low-converters on the LCHF diets was accompanied by a higher fecal relative abundance of *E. coprostanoligenes* (Figure 6e, page 61, Table S5, q = 0.003, Tukey's Test^{FDR}). Cholesterol high-converters responded to the LCHF diets with reduced cholesterol conversion, at least based on increased fecal cholesterol (q = 0.08) concentrations and a decreased coprostanol/cholesterol ratio (q = 0.03), although fecal coprostanol levels were not altered (q

> 0.1) (Figure 6f, page 61, Table S5, Tukey's Test^{FDR}). Neither differences in serum BHB were detected between cholesterol high and low-converters from both studies (p = 0.067, paired Wilcoxon signed-rank test), nor a positive association between increased ketosis and cholesterol-to-coprostanol conversion (Δ BHB vs. Δ coprostanol, R = 0.17, p = 0.39, SRC).

Thus, LCHF diets consistently increased cholesterol-to-coprostanol conversion in low-converters from the KETO and CARBFUNC cohorts, despite different metabolic health backgrounds and underlying fecal cholesterol and coprostanol concentrations.

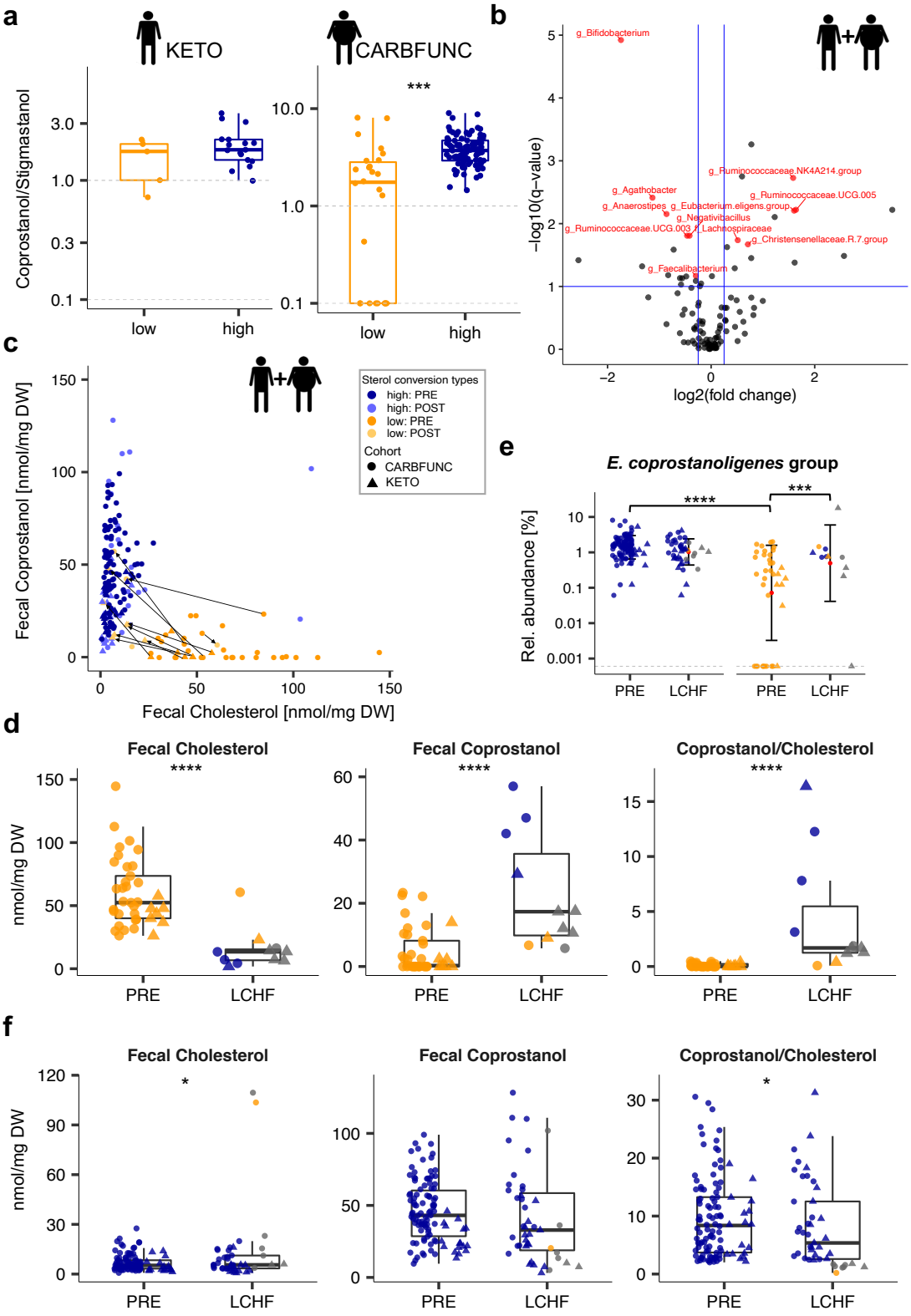


Figure 6. Diet impact on cholesterol-to-coprostanol conversion.
(a) Cholesterol high-converters with obesity (CARBFUNC study, n = 89) but not without obesity (KETO study, n = 17) exhibited an increased fecal coprostanol/stigmastanol ratio compared to low-converters

($n_{CARBFUNC}$ = 26, n_{KETO} = 5), suggesting a higher proportion of animal vs. plant-derived dietary fat intake ($p < 0.001$ ***, WRST). Zero values were replaced with a pseudocount (1 nmol / mg dry weight [DW]).
(b) LCHF diets induced consistent taxonomic microbiota alterations in the KETO and CARBFUNC cohorts, based on a combined GLMM analysis (n_{PRE} = 173, n_{LCHF} = 62). Bacterial taxa with significant changes in relative abundance ($q < 0.1$, BH-corrected Tukey's test, see horizontal blue line) and a positive or negative fold-change of > 0.25 in estimated marginal means (EMM) are marked with red dots and labels, unless they were detected by the GLMM as cohort and/or sex-associated (black dots).
(c) Distribution of cholesterol high and low-converters among all KETO and CARBFUNC study participants before (n = 143) and after (n = 54) LCHF dietary intervention (gray lines connecting pre and post-intervention samples).
(d) Increased cholesterol-to-coprostanol conversion in low-converters from both cohorts on the LCHF diets (n_{PRE} = 37, nLCHF = 11), as evidenced by reduced fecal cholesterol and increased fecal coprostanol levels and increased coprostanol/cholesterol ratios.
(e) The increased cholesterol-to-coprostanol conversion in low-converters on LCHF diets was accompanied by an increased fecal relative abundance of *E. coprostanoligenes* group ($n_{PRE/Low}$ = 37, $n_{LCHF/Low}$= 11), resulting in similar relative abundances in high and low-converters on the LCHF diets ($n_{PRE/High}$= 106, $n_{LCHF/High}$ = 43).
(f) Decreased cholesterol-to-coprostanol conversion in high-converters from both cohorts after LCHF diet intervention (n_{PRE} = 106, n_{LCHF} = 43), at least based on increased fecal cholesterol levels and an increased coprostanol/cholesterol ratio. Individuals were classified as cholesterol high/low-converters based on pre-intervention time points, with symbol colors indicating the classification during the LCHF diet. Significance determined by GLMM and post-hoc Tukey's test (Benjamini-Hochberg-corrected): $q > 0.1$ ns, $q \leq 0.1$ *, $q \leq 0.05$ **, $q \leq 0.01$ ***, $q \leq 0.001$ ****

5. Cholesterol converter type-specific dietary impact on serum lipids

To determine if the cholesterol converter type affected serum lipid responses to LCHF diets, total cholesterol, TAG, HDL-C and LDL-C concentrations were compared in high and low-converters with and without obesity. The LCHF diet-induced increase in cholesterol conversion in low-converters from both cohorts was not accompanied by altered blood lipid concentrations (Figure 7a, page 63, $q > 0.1$, Tukey's $Test^{FDR}$). However, both cholesterol high and low-converters responded to the LCHF diets with a reduction in serum TAG levels (Figure 7b, page 63, $q_{low} = 0.081$, $q_{high} < 1e\text{-}3$). This converter type-independent effect was apparent even when controlling for cohort-specific differences in serum lipids ($q \leq 0.1$, Table S5).

Cholesterol high-converters from the lean KETO cohort responded to the LCHF diet with increased serum LDL-C levels (Figure 7c, page 63, Table S5, $q = 0.015$, Tukey's $Test^{FDR}$). This effect was not explained by different intakes (E%) of total fat or saturated, mono-unsaturated, or poly-unsaturated fatty acids between lean cholesterol high and low-converters (Table 5, page 58, $p > 0.05$, WRST). Lean cholesterol high-converters from the KETO study even consumed less cholesterol during the LCHF diet than low-converters (Table 5, page 58, $p = 0.016$, WRST). Neither cholesterol high nor low-converters with obesity from the CARBFUNC study exhibited increased serum LDL-C concentrations on the LCHF diet (Table S5, $q > 0.05$), despite increased saturated fatty acid consumption (30 %E) during the

intervention (Horn et al., 2022), which has previously been suggested to increase LDL-C levels (Siri-Tarino et al., 2010).

In summary, the LCHF diet-induced increase in intestinal cholesterol conversion had no discernible effect on circulating serum lipids but, in lean individuals, the cholesterol high-converter type may be prognostic for increased LDL-C levels in response to LCHF diets.

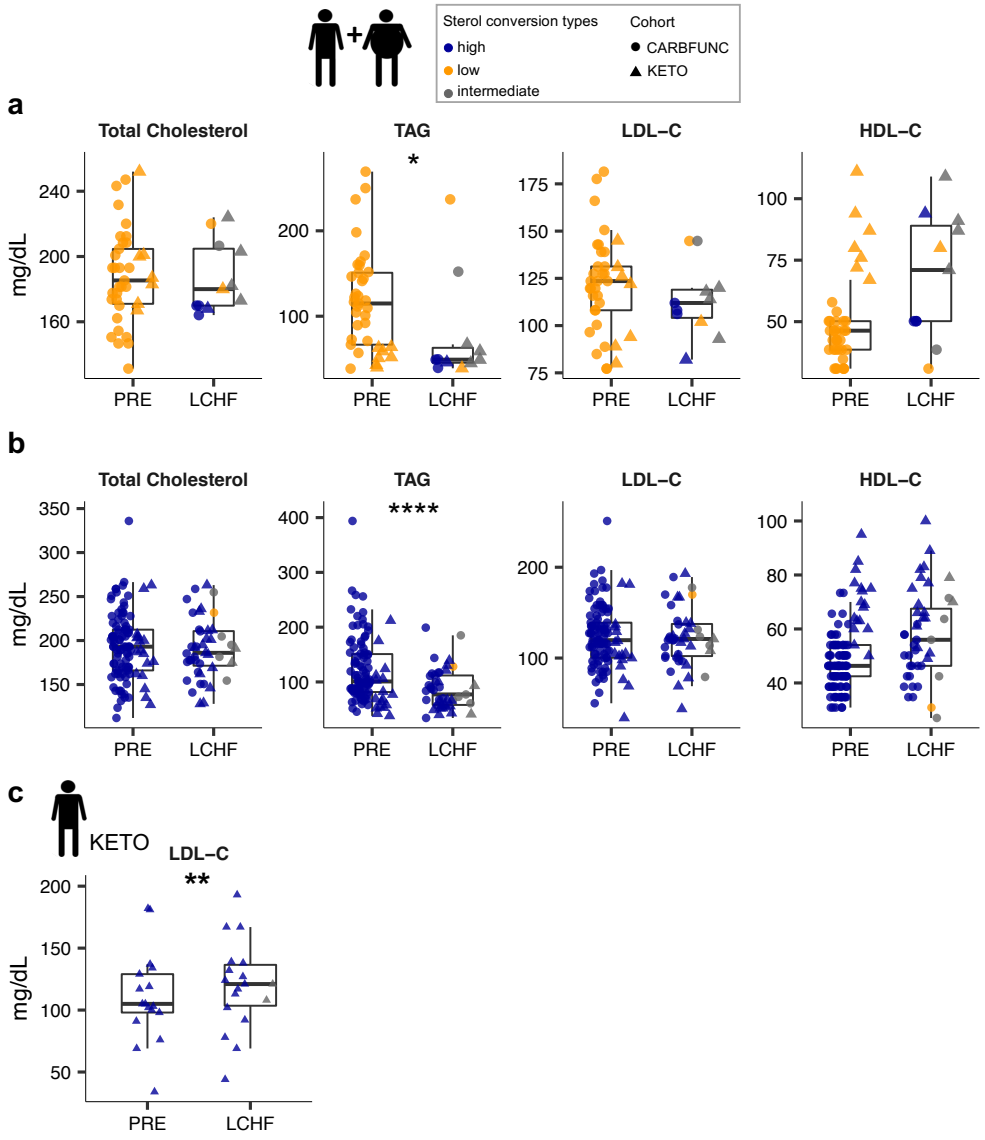


Figure 7. Cholesterol converter type-specific dietary impact on serum lipids.
(a) and (b) Decreased serum TAG levels in both cholesterol low-converters (a, n_{PRE} = 37, n_{LCHF} = 11) and high-converters (b, n_{PRE} = 106, n_{LCHF} = 43), based on estimated marginal means (EMMs), as determined by the GLMM for the combined KETO and CARBFUNC cohorts.
(c) Increased serum LDL-C levels in lean cholesterol high-converters from the KETO study on the LCHF diet (n_{PRE} = 17, n_{LCHF} = 18). Individuals were classified as cholesterol high/low-converters based on pre-intervention time points, with symbol colors indicating the classification during the LCHF diet. Significance determined by GLMM and post-hoc Tukey's test (Benjamini-Hochberg-corrected): $q > 0.1$ ns, $q \leq 0.1$ *, $q \leq 0.05$ **, $q \leq 0.01$ ***, $q \leq 0.001$ ****

B. Characterization of the microbial community during intestinal inflammation in IL-10 deficient mice co-housed with wild type animals

To investigate how intrinsic, non-modifiable factors shape the gut microbiota and how these interactions are linked to disease, more specifically colitis, a genetic knockout (KO) mouse model of mice deficient for the production of IL-10 (*IL-10*$^{-/-}$ mice) was used. *IL-10*$^{-/-}$ mice develop spontaneous colitis, which appears to be dependent on the gut microbiota (Sellon et al., 1998). BALB/c *IL-10*$^{-/-}$ mice from different litters were mixed and co-housed, including sex- and age matched Swiss wild-type (WT) mice, for twenty weeks and fecal samples were collected to examine microbiota alterations throughout the development of spontaneous colitis.

1. Phenotypic characterization of the *IL-10*$^{-/-}$ mouse model

In the *IL-10* knockout model, the well-being of mice determined by phenotypic observations is the only indicator of colitis development throughout an experiment. Intestinal inflammation though can only be verified by histopathological examination of colonic tissue after mice are sacrificed, which leads to their withdrawal from the observation. Therefore, if phenotypic observations reliably mirror inflammatory processes indicating colitis was evaluated first.

Throughout the observation, 50% of IL-10 deficient mice (9 of 18) developed colitis-related symptoms based on phenotypic observations and had to be withdrawn from the experiment early. These observations were based on a scoring system which includes the general status of mice, such as behavior, appearance, weight, and absolute indications of colitis, e.g. rectal prolapse and convulsions (Table S1). Although some wild type mice were classified as symptomatic throughout the observation, none of them showed severe signs of colitis leading to withdrawal (Table 1, page 31). Weight reductions between 10% to 20% and slightly increased stool consistency (score 2) were the only scores attributed to wild type mice, which were recovered again within a few days.

The first mouse with colitis-related symptoms was observed after eleven weeks of observation, which equaled 14 weeks of age. Up until then, the body weight increased for all mice irrespective of the genotype (mean weight gain of 10.7 g ± 1.48 g for knockout and 11.30 g ± 3.29 g for wild-type mice compared to the day of cage allocation, mean ± standard deviation [sd]). Thereafter, the body weight reached a plateau with on average variations of 0.43 g ± 0.41 g for knockout and 0.68 g ± 0.76 g for wild-type mice compared to the preceded measurement (Figure 8a, page 65). With an age of twelve weeks mice are usually considered mature adult animals (S. J. Jackson et al., 2017), which is in line with the observed stable body weight at around 14 weeks.

Wild-type Swiss mice were in general heavier than *IL-10* knockout BALB/c mice with an adult average weight of 33.04 g ± 3.04 g compared to 24.63 g ± 1.48 g for KO mice (Figure 8a, b, Table S12, WT litter 1 vs each KO litter: q≤ 1e-4, Wilcoxon rank-sum test with false discovery rate [FDR] corrected p-value according to the Benjamnini-Hochberg procedure [WRST^{FDR}]). One WT mouse (# 57) deviated from the others, with a higher body weight after 14 weeks of age. This mouse was on average 5.87 g heavier than the other WT mice (Figure S5, 36.92 g ± 1.36 g body weight of mouse #57 vs 31.05 g ± 1.01 g of all other WT mice, mean ± sd). Besides genotype differences, also body weight variations between the *IL-10*$^{-/-}$ litters were observed. Litter three mice, which was the only KO litter in which no mice had to be sacrificed due to severe colitis-related symptoms, had the highest body weight compared to the other KO litters even when considering only healthy mice (Figure 8b, Table S12, Litter 3 vs 2: 23.38 g ± 3.56 g vs 21.46 g ± 3.76 g, q≤ 1e-4; Litter 3 vs 4: 23.38 g ± 3.56 g vs 22.07 g ± 4.04 g, q= 0.0013 , mean ± sd, WRST^{FDR}).

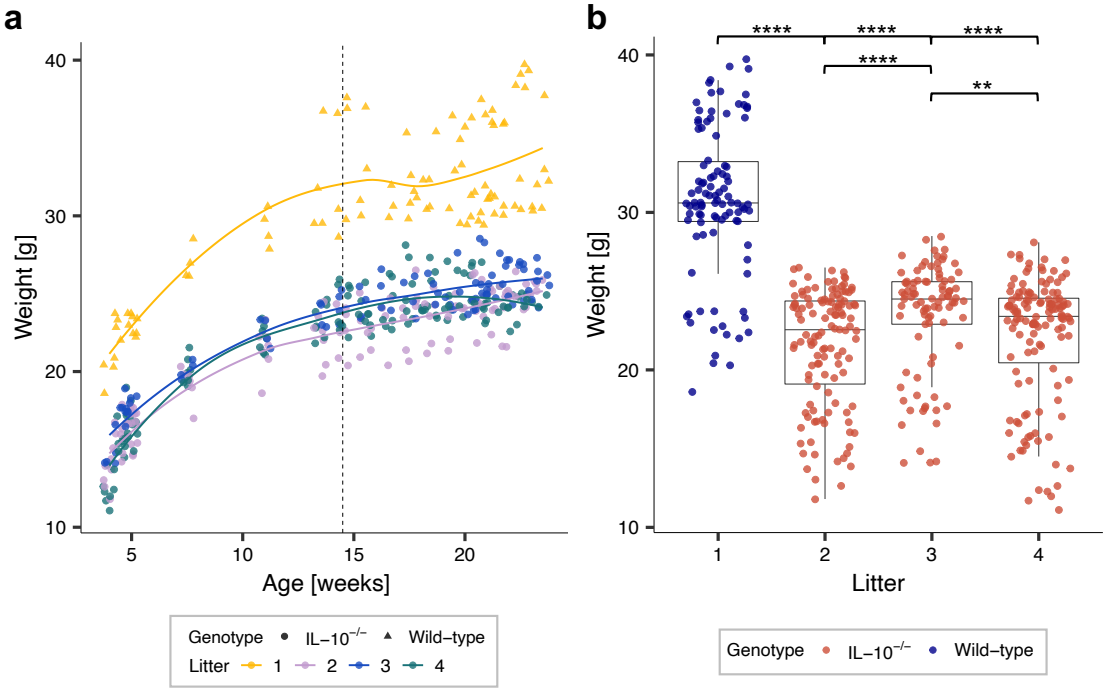


Figure 8. Body weight progression of healthy samples collected throughout twenty weeks of observation.
Absolute body weight of only healthy (scores <2) mice (a) colored and grouped by litter number using locally weighted regression and (b) per genotype for each litter. At the beginning of the observation, mice were four weeks old.
(a) The vertical dashed line indicates the first observation of colitis-related symptoms.
(b) Pairwise Wilcoxon rank-sum test with FDR corrected q-values only shown if q≤ 0.05, with q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; exact p-values for all comparisons are listed in Table S12; n(wild-type)= 97, n(knockout)= 341. Abbreviations: WRST = Wilcoxon rank-sum test, FDR = false discovery rate

To estimate the significance of observational phenotypes as indicators for colitis-related inflammation, they were correlated with histopathological assessments and cytokine expression, which were determined for every mouse after euthanasia or the end of the experiment. The sum of all phenotypic scores was significantly correlated with increased tissue damage (proximal: $\rho = 0.75$, $q = 0.00015$; distal: $\rho = 0.83$, $q \leq$ 1e-4) and cell infiltration (proximal: $\rho = 0.67$, $q = 0.0013$; distal: $\rho = 0.84$, $q \leq$ 1e-4, Spearman's rank correlation $[SRC]^{FDR}$) in the proximal and distal colon (Figure 9, page 67). The same accounted for an increased expression of inflammatory cytokines Tumor necrosis factor (TNF) (proximal: $\rho = 0.44$, $q = 0.043$; distal: $\rho = 0.77$, $q = 0.00015$), Growth-regulated protein beta (Gro-beta) (proximal: $\rho = 0.75$, $q = 0.00015$; distal: $\rho = 0.85$, $q \leq$ 1e-4), Chemokine C ligand (CCL) 2 (proximal: $\rho = 0.59$ $q = 0.0054$) and CCL3 (proximal: $\rho = 0.62$, $q = 0.0034$; distal: $\rho = 0.77$, $q = 0.00015$), as well as decreased expression of anti-inflammatory Transforming growth factor beta (TGF) (Figure 9, page 67, proximal: $\rho = -0.50$, $q = 0.022$; distal: $\rho = -0.76$, $q = 0.00015$, SRC^{FDR}). This validated the possibility of tracking inflammatory processes during colitis by phenotypic characterization based on regular scoring of the mice's well-being.

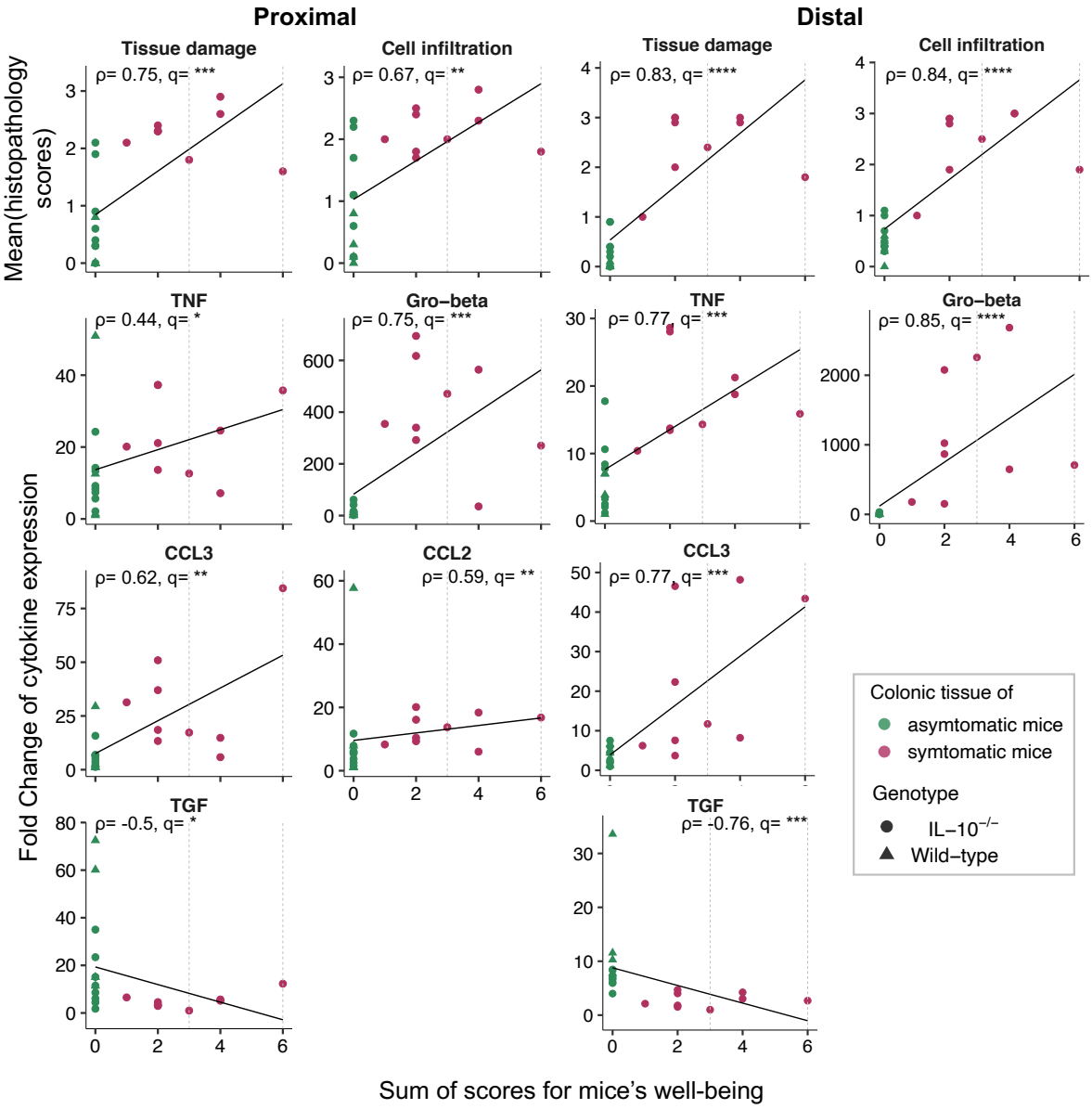


Figure 9. Intestinal inflammation was correlated with phenotypic characterizations based on scored well-being parameters in proximal and colonic tissue.
The well-being of mice was determined based on scores evaluating behavior, weight, appearance and absolute criteria. The sum of these phenotypic scores was correlated with histopathological examination and relative cytokine expression to GAPDH of proximal and distal colonic tissue and colored based on if the mice showed colitis-related symptoms (symptomatic) or not (asymptomatic) throughout the observation. Spearman's rank correlation with FDR corrected q-values, q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; n(wild-type)= 4, n(knockout)= 18. Abbreviations: CCL = Chemokine C ligand, GAPDH = Glyceraldehyde-3-phosphate dehydrogenase, Gro-beta = Growth-regulated protein beta, TNF = Tumor necrosis factor, TGF = Transforming growth factor beta

To assess the consistency of fecal samples, which has been shown to correlate with taxonomic microbiota compositions in humans (Vandeputte et al., 2016) and as diarrhea is an absolute indicator of colitis in the scoring system, a stool consistency score was introduced based on

the human Bristol Stool scale (Table 2, page 33). Additionally, to better characterize microbial alterations in the early stage of intestinal inflammation, fecal sample were collected on two to three consecutive days of an entire cage if a mouse showed unusual phenotypic symptoms as determined by any score being ≥ 2. Stool consistency was positively correlated with tissue damage (proximal: $\rho = 0.67$, $q = 0.0089$; distal: $\rho = 0.65$ $q = 0.0089$) and cellular infiltration (proximal: $\rho = 0.53$, $q = 0.030$; distal: $\rho = 0.62$, $q = 0.013$, SRC^{FDR}) in both colon biopsies (Figure 10, page 69). Alterations in cytokine expression profiles were not significantly correlated with stool consistency except for inflammatory Gro-beta (proximal: $\rho= 0.55$, $q= 0.030$; distal: $\rho= 0.52$ $q= 0.030$) and anti-inflammatory TGF (Figure 10, page 69, proximal: $\rho = -0.46$, $q = 0.049$; distal: $\rho = -0.51$, $q = 0.030$, SRC^{FDR}). For distal colonic tissue only, TNF was positively correlated with stool consistency (Figure 10, page 69, proximal: $\rho = 0.27$, $q = 0.24$; distal: $\rho = 0.53$, $q = 0.012$, SRC^{FDR}). However, in many mice regular, normal stool consistency (score: 1) was associated with tissue alterations and cytokine expression levels indicative of inflammation, indicating that stool consistency may not be as sensitive as a marker for inflammatory alterations as other phenotypic observations scoring the well-being of mice.

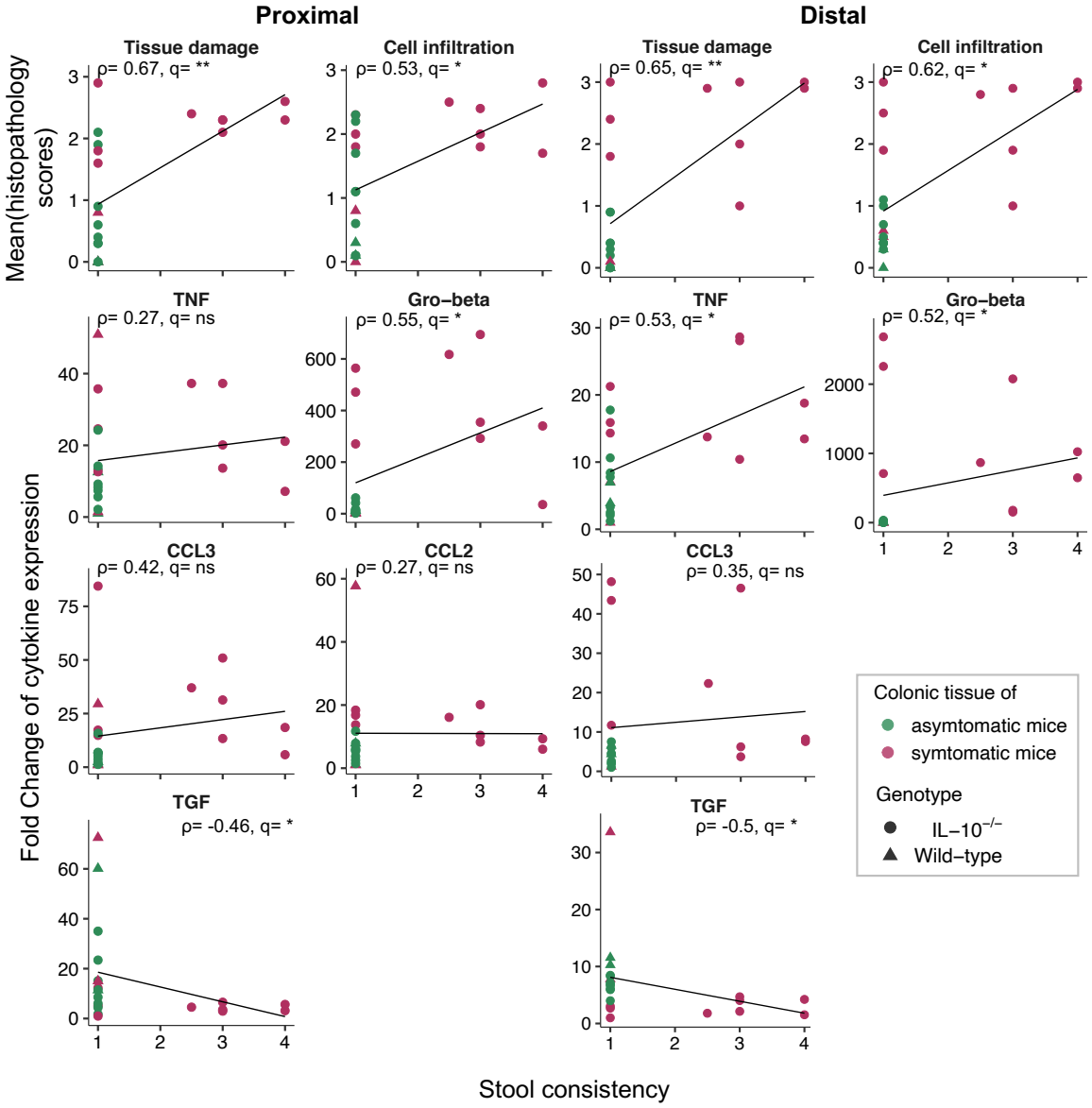


Figure 10. Stool consistency was only correlated with tissue inflammation and few cytokine levels like Gro-beta and TGF irrespective of colonic location.
Stool consistency was modified based on the human Bristol Stool scale and correlated with histopathological scores and relative cytokine expression to GAPDH of proximal and distal colonic tissue and colored based on if the mice showed colitis-related symptoms (symptomatic) or not (asymptomatic) throughout the observation. Spearman's rank correlation with FDR corrected q-values, q> 0.05 ns/not significant, q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; n(wild-type)= 4, n(knockout)= 18. Abbreviations: CCL = Chemokine C ligand, GAPDH = Glyceraldehyde-3-phosphate dehydrogenase, Gro-beta = Growth-regulated protein beta, TNF = Tumor necrosis factor, TGF = Transforming growth factor beta

2. Gut microbial characterization of the *IL-10*$^{-/-}$ mouse model

In the still largely unknow etiology of IBD, the microbiota is thought to be a key factor, either changing subsequent to inflammatory alterations in the intestinal environment or as trigger for abnormal immune responses (Keubler et al., 2015). Therefore, to characterize the role of the microbial composition and search for associations with intestinal inflammation during colitis, 16S rRNA gene amplicons were sequenced in a subset of over 500 fecal samples collected throughout the observation, resulting in 11,681,184 and 10,420,814 raw single reads respectively for two consecutive sequencing batches. After processing, quality evaluation and merging of the two batches, a total 16,731,777 single reads resulting in 900 amplicon sequence variants (ASVs) were obtained (Table S3).

a) Development and litter influences of a healthy murine gut microbial community

Since the reported descriptions of colitis-associated microbial communities vary (Dey et al., 2013; Franzosa et al., 2019; Gevers et al., 2014; Knoch et al., 2010; Schaubeck et al., 2015; Schwab et al., 2014), the determination of the healthy, normal state of the microbiota in each experimental set up is advantageous. Therefore, possible influences resulting from the experimental design, e.g. co-housing or litter number, on the microbiota of only healthy samples (with any score well-being and/or consistency < 2) were examined.

In line with the body weight progression, the dissimilarity between microbial communities based on the Bray-Curtis dissimilarity increased up to 14 weeks of age compared to the day of cage allocation for each individual, irrespective of the litter (Figure 11a, page 71). After that, again similar to the body weight, microbial changes reached a plateau at around the same time the first colitis-related symptoms were observed. Plotting the differences between mice by multidimensional scaling revealed differences between genotypes and within KO litters. The knockout litter 2 and 4 had the highest compositional dissimilarities compared to the remaining knockout litter 3 and the wild type mice (Figure 11b, c page 71; b: Genotype: $R = 0.14$, $p = 0.001$, ANOSIM; PCoA1: $q \leq 1e\text{-}4$, PCoA2: $q \leq 1e\text{-}4$, WRST^{FDR}; c: Litter: $R = 0.25$, $p = 0.001$, ANOSIM; Litter 2: PCoA1: $q \leq 1e\text{-}4$, PCoA2: $q \leq 1e\text{-}4$; Litter 4: PCoA1: $q = 0.0004$, PCoA2: $q = 0.002$, WRST^{FDR}, for exact p-values and all comparisons see Table S12). This is in line with the observed lower intra-individual dissimilarities for WT and litter 3 knockout mice compared to their microbial compositions on the day of cage allocation, as for litter 2 and 4 IL-10 deficient animals (Figure 11a, page 71). These were also the only KO litters that developed severe signs of colitis during the observation period leading to early withdrawal, which raises the

question if litter differences between IL-10 deficient mice were already present on the day of cage allocation, or if they developed with time.

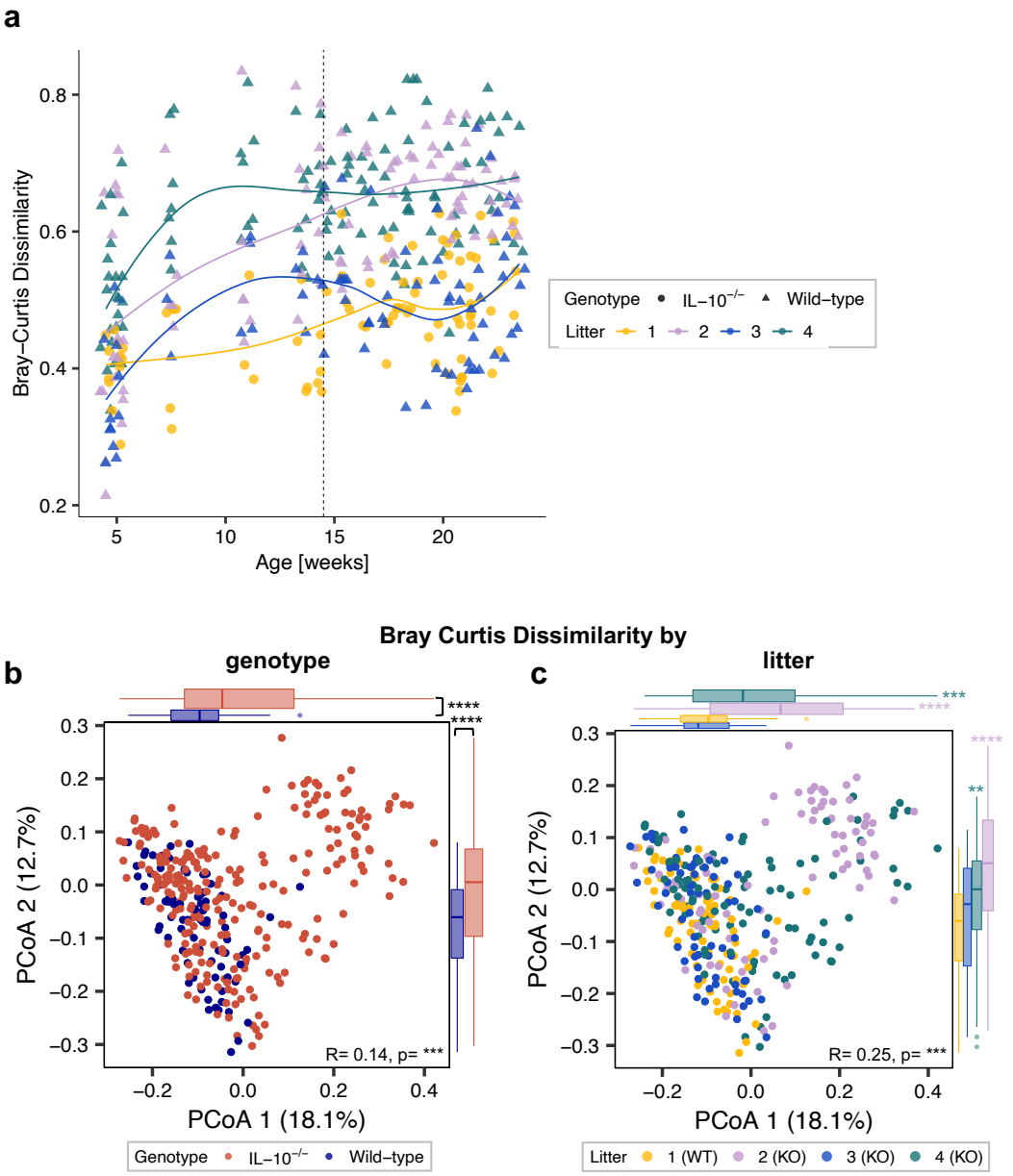


Figure 11. Dissimilarities in microbial community compositions between and within healthy mice were dependent on the litter rather than the genotype.
(a) Dissimilarities of microbial compositions within individual mice compared to the day of cage allocation are increasing until twelve to 14 weeks of age.
(b, c) Dissimilarities between mice were stronger clustered based on the litter than the genotype in general. Significances above boxplots in (c) refer to comparisons against litter 1 for each group. Clustering of healthy (scores <2) microbial community dissimilarities were based on the Bray-Curtis Dissimilarity and globally tested via ANOSIM and along each PCoA axis via pairwise Wilcoxon rank-sum test with FDR corrected q-values only shown if p/q≤ 0.05, p/q > 0.05 not significant, p/q≤ 0.01 **, p/q≤ 0.0001 ****; exact p-values for all comparisons are listed in Table S12; n(litter 1 / 2 / 3 / 4)= 79 / 84 / 66 / 101

To characterize the development of microbial diversity within the first week of cage allocation, measurements of alpha diversity considering microbial richness (e.g. the number of observed ASVs) and additionally including evenness (e.g. Shannon Index) within a sample were compared between genotypes and litters (Figure 12a, b, page 73, Figure S7). Wild-type mice had a significantly higher alpha diversity then IL-10 deficient mice irrespective of the measurement (Number of observed ASVs: p = 0.0002; Shannon Index: p = 0.0004, WRST). Further they were characterized by a trend of a lower bacterial density as determined by the number of 16S rRNA gene copies per mg stool (Figure S6, p = 0.12, WRST), which significance is likely limited by the lower sample number subset used for the quantitative examination. Discriminating litters of IL-10 deficient mice, revealed a significantly lower microbial diversity of litter 2 mice compared to all other litters (Figure 12a, b, page 73, Number of observed ASVs: $q_{\text{litter 2 vs 1}} \leq$ 1e-4, $q_{\text{litter 2 vs 3}} \leq$ 1e-4, $q_{\text{litter 2 vs 4}} = 0.0004$; Shannon Index: $q_{\text{litter 2 vs 1}} \leq$ 1e-4, $q_{\text{litter 2 vs 3}} = 0.0032$, $q_{\text{litter 2 vs 4}} = 0.0014$, $WRST^{FDR}$). *IL-10*$^{-/-}$ mice of the second knockout litter with early withdrawal (litter 4) also had a lower number of observed ASVs per sample compared to WT mice and the healthy KO litter 3 ($q_{4 \text{ vs } 1} = 0.0044$; $q_{4 \text{ vs } 3} = 0.056$, $WRST^{FDR}$), and a trend of the highest 16S rRNA gene copy number per mg feces (Figure 12a – c, page 73, $q_{4 \text{ vs } 1} = 0.17$, $q_{4 \text{ vs } 2} = 0.17$; $q_{4 \text{ vs } 3} = 0.93$, $WRST^{FDR}$). Contrarily, mice of the healthy KO litter 3 had a similar alpha-diversity as wild-type mice, despite their different genetic backgrounds and origins (Figure 12a, b, page 73, all measurements for litter 3 vs 1 $q > 0.05$, $WRST^{FDR}$, for exact p-values see Table S12). All in all, within the first week of cage allocation the bacterial diversities and densities varied substantially between mouse litters, even within genetically identical mice, suggesting that the microbiota had already been fundamentally shaped before cage allocation.

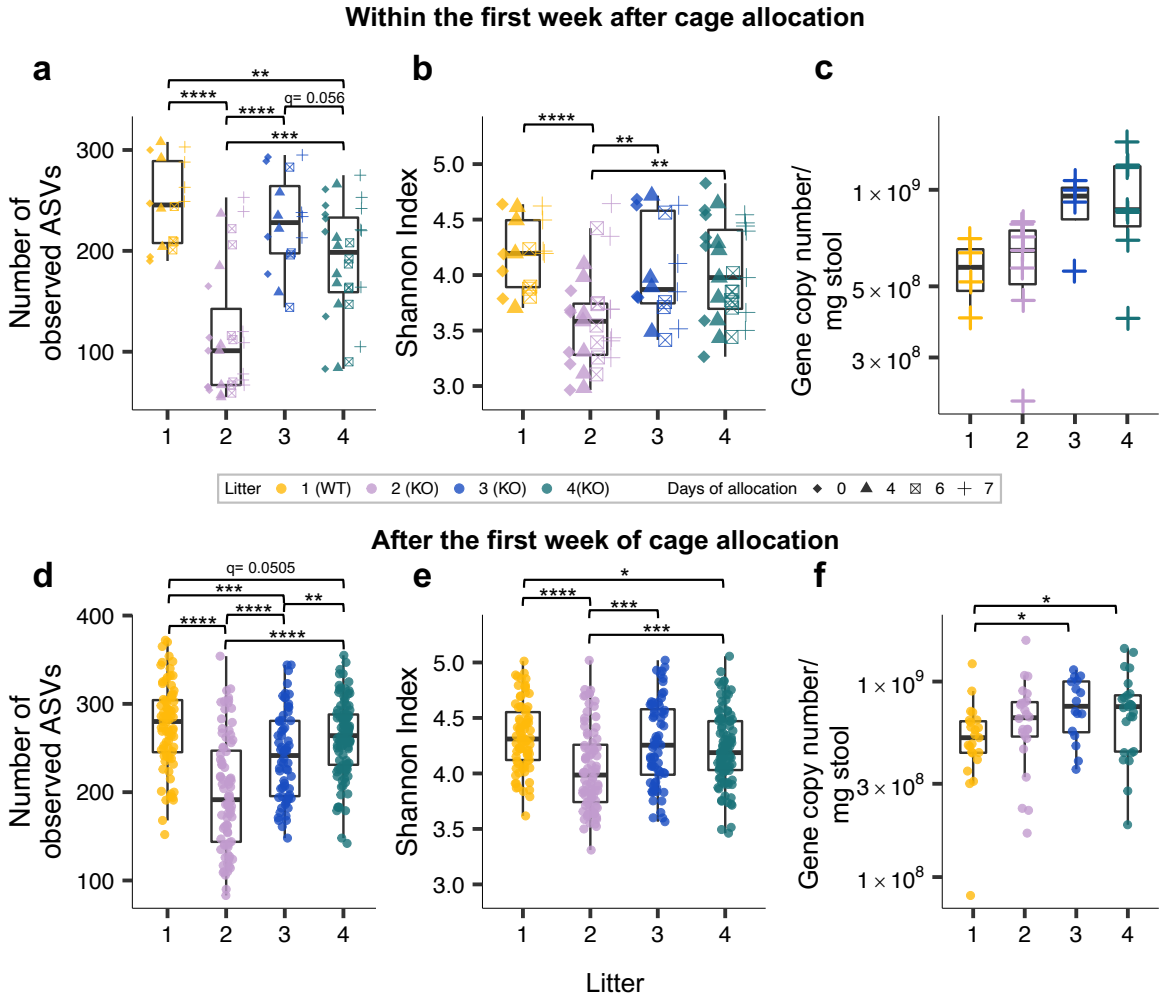


Figure 12. The microbial diversity and overall bacterial density depended on the litter and genotype in the first week of cage allocation (a-c) and under healthy conditions thereafter (d-f).
The number of observed ASVs, the Shannon Index and the total 16S rRNA gene copy number normalized to 1 mg of feces were compared between litters within the first week of cage allocation (a – c) and only under healthy (scores < 2) conditions thereafter (d – f). c and f y-axes are log10-scaled. Pairwise Wilcoxon rank-sum test with FDR corrected q-values only shown if q≤ 0.05, q > 0.05 not significant, q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; exact p-values for all comparisons are listed in Supplementary Table S12; a, b: n(litter 1 / 2 / 3 / 4)= 16 / 27 / 16 / 28; c: n(litter 1 / 2 / 3 / 4)= 4 / 7 / 4 / 7; d, e: n(litter 1 / 2 / 3 / 4)= 79 / 84 / 66 / 101; f: n(litter 1 / 2 / 3 / 4)= 23 / 24 / 16 / 25

To determine if these litter differences persisted throughout the observation or if they were compensated for by co-housing, the microbial diversity was compared between all healthy samples (scores < 2) after the first week of cage allocation until the end of the observation. Even after several weeks of co-habitation litter-dependent differences in community richness and evenness were observed. WT mice were still characterized by higher alpha diversity irrespective of the measurement (Figure S6, Number of observed ASVs: p ≤ 1e-4; Shannon Index: p = 0.0004, WRST) and a significantly lower bacterial density (Figure S6, p =0.005, WRST) compared to IL-10 deficient mice in general. Within *IL-10*$^{-/-}$ litters, the number of ASVs

and the Shannon Index continued to be significantly reduced for litter 2 mice compared to the others (Figure 12d, e, page 73, Number of observed ASVs: $q_{\text{litter 2 vs 1}} \leq$ 1e-4, $q_{\text{litter 2 vs 3}} \leq$ 1e-4, $_{\text{litter 2 vs 4}} \leq$ 1e-4; Shannon Index: $q_{\text{litter 2 vs 1}} \leq$ 1e-4, $q_{\text{litter 2 vs 3}} = 0.0006$, $q_{\text{litter 2 vs 4}} = 0.0004$, $WRST^{FDR}$). Although mice from all litters increased the number of observed ASVs throughout the observation, the increase in litter 2 and litter 4 mice was disproportionally higher, differentiating healthy litters from the ones that developed colitis again (Δ mean observed ASVs $number_{\text{after first week - within first week}}$: litter 1= 28.01, litter 2= 81.52, litter 3= 12.84, litter 4= 71.73). The higher bacterial density in knockout mice from litter 3 and 4 compared to the wild-type litter persisted throughout the entire co-habitation ($q_{\text{litter 1 vs 3}} = 0.042$, $q_{\text{litter 1 vs 4}} = 0.042$, $WRST^{FDR}$), which was only indicated by a trend for litter 2 (Figure 12f, page 73, $q_{\text{litter 1 vs 2}} = 0.12$, $WRST^{FDR}$). Taken together, the microbial diversity, as well as the overall community composition of healthy mice were characterized by the mouse litter before co-housing and even throughout it. Exact p-values as well as descriptive statistics of all comparisons are listed in Table S12.

To determine if co-housing led mice to lose or acquire bacteria, the mean relative abundance of genera per litter at each timepoint was calculated and only taxa considered being present with a mean litter abundance > 0.01 %. The taxa were then grouped by the number of litters in which they were considered being present on the day of cage allocation (day 0), e.g. taxa that were abundant in only one, two, three or all four litters before co-housing. Based on this, nine taxa were considered being present in only one litter, 14 in two and 25 in three litters at the beginning of the experiment (Table 6, page 75).

Table 6. List of the mean relative abundance as classified on the day of cage allocation and transferability of taxa between litters throughout the observation.
The mean relative abundance per litter of 110 taxa in total was calculated at each observation and grouped depending on the number of litters in which they were abundant at the day of cage allocation. Only taxa with a litter mean abundance > 0.01% were considered being present. Severe colitis = one score ≥ 3 or the sum of all scores ≥ 6 leading to early withdrawal of mice

Abundant on the day of cage allocation		In one litter only	In two litters	In three litters
Total number of taxa		9	14	25
Genotypes	Only in wild-type litter (# 1)	5	-	-
	Only in knockout litters (# 2, 3, 4)	4	5	2
	No specific pattern based on genotype	-	9	23
Phenotypic characterization	Only in litter 2	1	-	-
	In litters except litter 2	8	14	18
	Only in litters that did not develop severe colitis (# 1, 3)	7	6	-
	Only in litters that developed severe colitis (# 2, 4)	2	-	-
	No specific pattern based on phenotype	-	8	25

Wild-type mice from litter 1 had the most unique taxa at the day of cage allocation. Five taxa, the genera *Parabacteroides* and *Candidatus Soleaferrea* and three genera belonging to the families *Christensenellaceae*, *Defluviitaleaceae* and *Ruminococcaceae*, were only present in mice of litter 1 before co-housing. All of them were thereafter transferred to the knockout litters (Figure 13, page 76). Four taxa were found to be present only in one knockout litter: the genus *Mucispirillum* in litter 2, *Acetatifactor* and *Bifidobacterium* in litter 3 and *Anaerovorax* in litter 4 (Figure 14b, page 77, Figure S7). All of them, except *Bifidobacterium*, were again transferred to the remaining litters. *Bifidobacterium* on the other hand did not seem to colonize the gastrointestinal tract of wild-type at all throughout the observation (Figure S7).

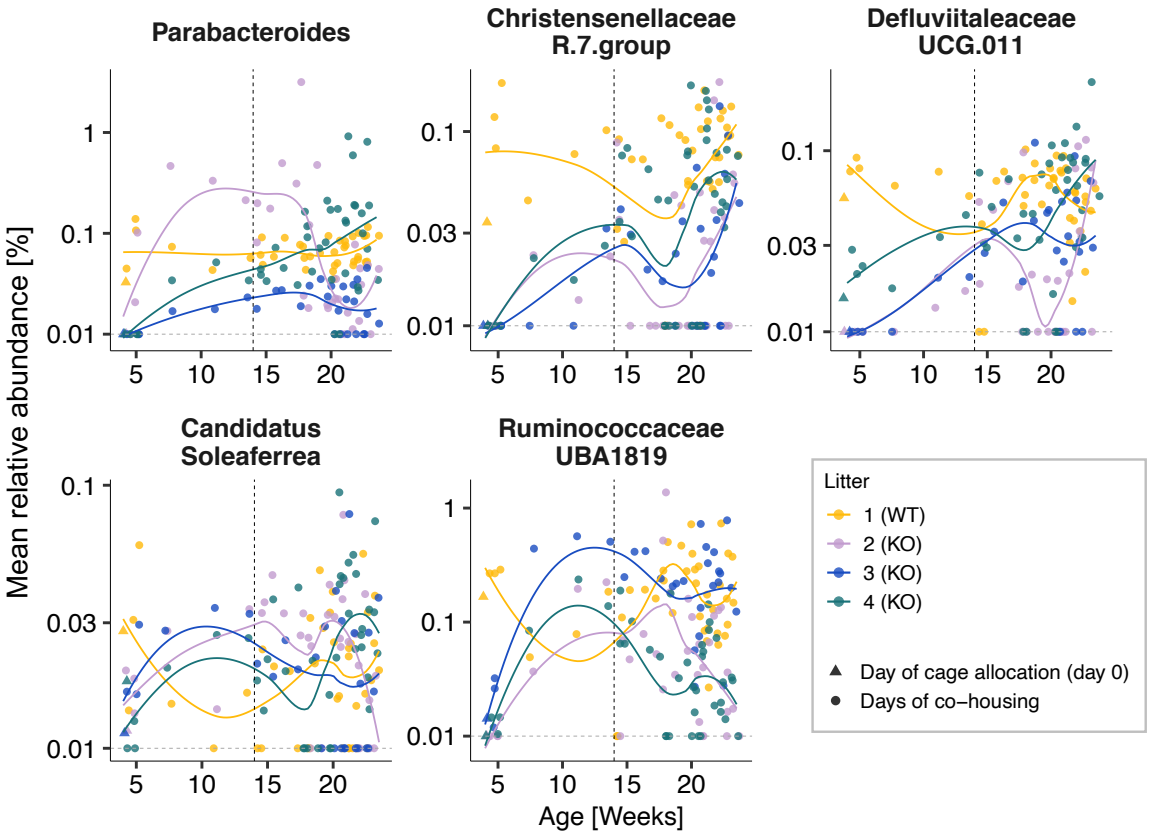


Figure 13. Five taxa were substantially more abundant in wild-type mice at the day of cage allocation.
The mean relative abundance per litter was calculated at each observation and taxa considered present with a mean relative abundance > 0.01 %. Values were log10-transformed after addition of a pseudocount of 0.0001 mean relative abundance. Only healthy samples (scores < 2) colored by litter throughout the observation are depicted. Vertical dashed line indicates the first observation of colitis-related symptoms and horizontal dashed line the pseudocount limit. Locally weighted regression was used to smooth abundance and y-axes were log10-scaled. n(litter 1 / 2 / 3 / 4) = 41 / 35 / 30 / 40

Mice from litter 2 acquired the largest number of new taxa by co-housing. At the day of cage allocation, litter 2 mice did not contain 18 taxa that were abundant in all other litters, irrespective of the genetic background (Figure 14a, page 77). These taxa were diverse belonging to families like *Ruminococcaceae* (seven genera), *Eggerthellaceae* (one genus), *Peptococcaceae* (one genus), *Lachnospiraceae* (five genera), *Erysipelotrichaceae* (one genus), *Saccharimonadaceae* (one genus), *Clostridiales Family XIII* (one genus) and the order *Mollicutes RF39* (one genus).

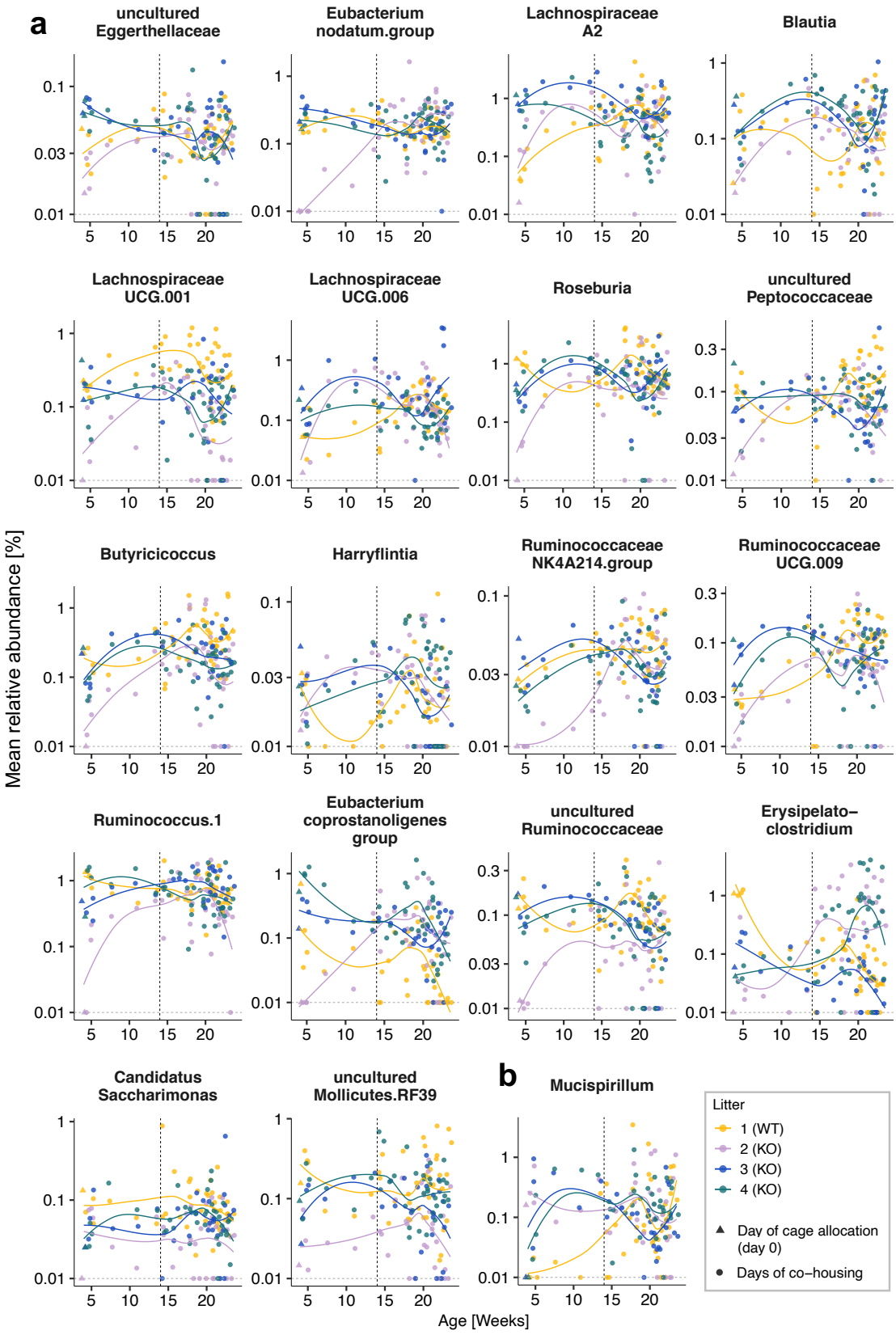


Figure 14. 18 taxa were substantially less abundant in litter 2 compared to all other litters at the day of cage allocation.

The mean relative abundance per litter was calculated at each observation and taxa considered present with a mean relative abundance > 0.01 %. Values were log10-transformed after addition of a pseudocount of 0.0001 mean relative abundance. Only healthy samples (scores < 2) colored by litter throughout the observation are depicted. Vertical dashed line indicates the first observation of colitis-related symptoms and horizontal dashed line the pseudocount limit. Locally weighted regression was used to smooth abundance and y-axes were log10-scaled. n(litter 1 / 2 / 3 / 4) = 41 / 35 / 30 / 40

Six taxa were only present in litter 1 and 3, the two litters in which no mice had be withdrawn due to colitis. These taxa belonged to the families *Ruminococcaceae* (two genera), *Lachnospiraceae* (one genera), *Erysipelotrichaceae* (one genus), and *Clostridiales Family XIII* (two genera) (Figure 15). Again, by co-housing all taxa were transferred to litter 2 and 4 mice and increased in relative abundance thereafter.

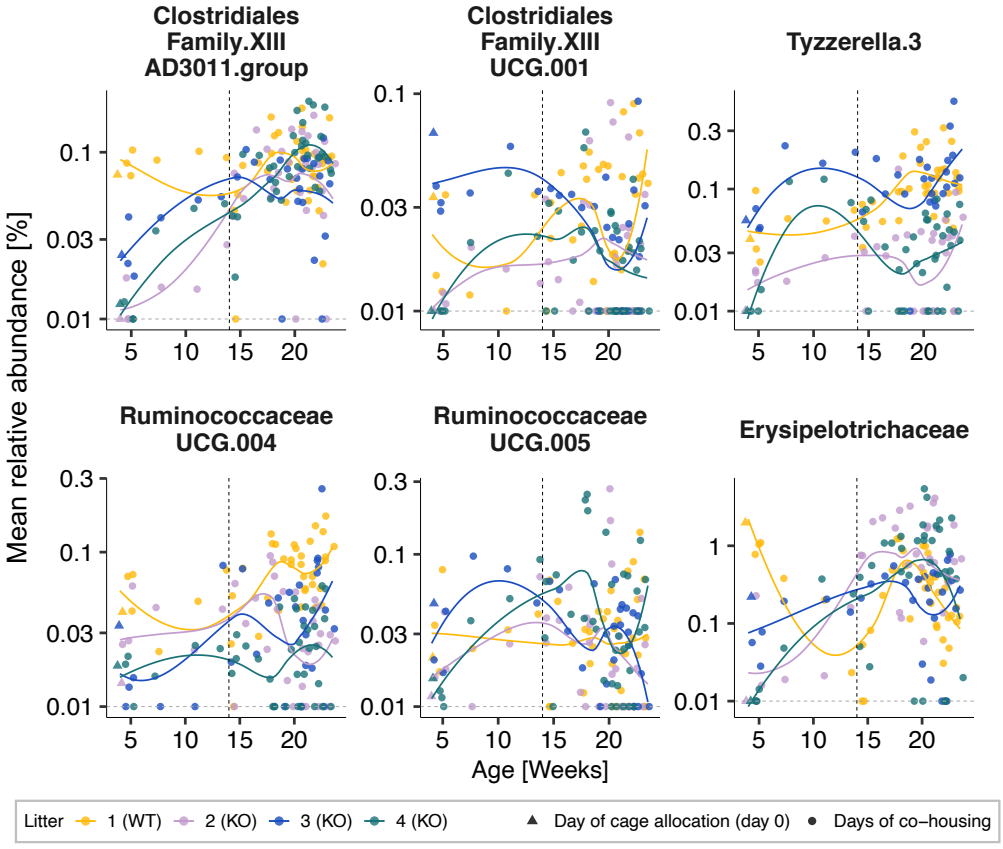


Figure 15. At the day of cage allocation, six taxa were more abundant in the two litters that did not develop severe colitis (#1 and 3).
The mean relative abundance per litter was calculated at each observation and taxa considered present with a mean relative abundance > 0.01 %. Values were log10-transformed after addition of a pseudocount of 0.0001 mean relative abundance. Only healthy samples (scores < 2) colored by litter throughout the observation are depicted. Vertical dashed line indicates the first observation of colitis-related symptoms and horizontal dashed line the pseudocount limit. Locally weighted regression was used to smooth abundance and y-axes were log10-scaled. n(litter 1 / 2 / 3 / 4) = 41/ 35 / 30 / 40

The co-housing of wild-type and knockout mice did not lead to gastrointestinal inflammation in wild-type mice, as revealed by histopathology scores as well as cytokine expression profiles at the end of the observation (Figure 9, page 67). Furthermore, they even transferred the most taxa to other litters than vice versa. Within knockout mice, the microbial community was differentially affected by co-housing dependent on the litter. Mice from the only KO litter that did not show signs of severe colitis (litter 3, based on one score ≥ 3 or the sum of all scores ≥ 6), were more similar based on microbial diversity and community composition to the wild-type than to other knockout mice (Figure 11, page 71, Figure 12, page 73, Table S12). The IL-10 deficient litter that was affected the heaviest by co-housing was litter 2, which was characterized by a microbial community changing the most throughout the observation and lower alpha-diversity in general. Taken together, under healthy conditions the microbial community development and its dynamics were affected by genotype but even stronger by litter. Although co-housing was revealed to facilitate microbial exchange, it did not completely abolish litter-dependent microbiota compositions.

b) Microbial alterations during early colitis development and progression

The development of colitis in the *IL-10*$^{-/-}$ mouse model is accompanied by intestinal inflammation (Kühn et al., 1993), which might further be intertwined with alterations of the gut microbiome. Whether the microbiota changes as a consequence of colitis-related intestinal inflammation, or if colitis is preceded by altered microbial communities, remains elusive. Comparing the bacterial density based on the absolute number of 16S rRNA genes per mg stool, revealed a decreased gene copy number in symptomatic mice (scores ≥ 2, referring to as "*colitis*" samples), which was not observed in *preC* samples, i.e. samples of phenotypically healthy mice collected up to four days before the first observation of colitis symptoms (Figure 16a, b, page 80). Any healthy samples collected of a mouse thereafter were classified as "*post colitis*", which applied to three mice throughout the trial only (# 57, # 58, # 98, Table 1, page 31). The decrease in fecal bacterial density during colitis was not accompanied by alterations in microbial diversity irrespective of the observational time point (Figure S8, Shannon Index and Number of observed ASVs, p/q > 0.05). Therefore, the decrease in bacterial density could have resulted from increased fecal water contents, as the number of 16S rRNA gene copies per mg stool was negatively correlated with stool consistency (Figure 16c, page 80, ρ = -0.47, q = 4.03e-08, SRC^{FDR}).

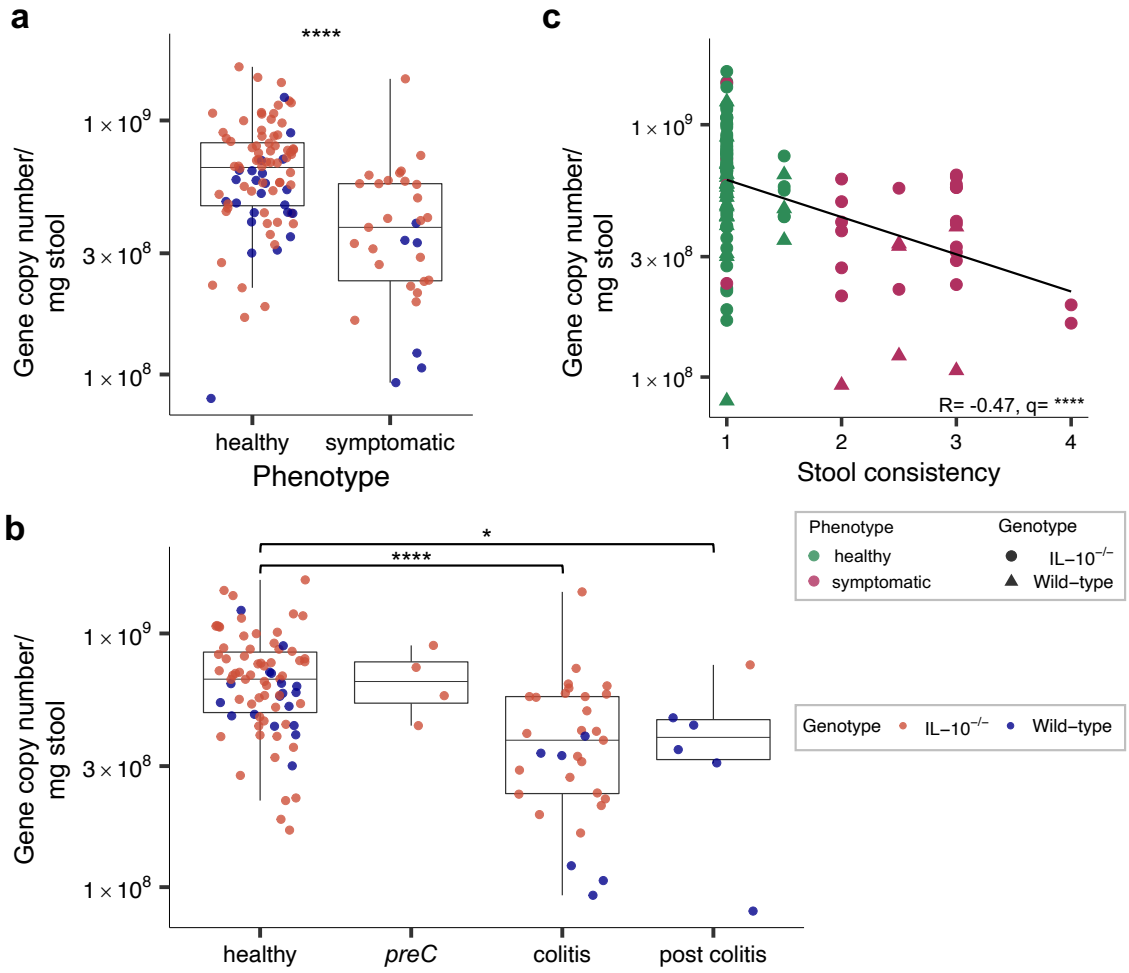


Figure 16. During colitis, the excreted fecal water content was increased and accompanied by decreased bacterial density.

(a, b) The bacterial density was based on the total number of 16S rRNA gene copies per mg stool and compared between symptomatic "*colitis*" (scores ≥ 2) and asymptomatic "*healthy*" (scores ≤ 1) mouse samples.

In (b), the healthy samples were further categorized based on the time of collection, with "*preC"* referring to healthy samples collected up to 4 days prior to the first observation of symptoms and "*post colitis*" to every sample collected after disappearance of colitis symptoms; pairwise WRST with FDR corrected q-values only shown if q≤ 0.05, exact p-values for all comparisons are listed in Supplementary Table S12;

(c) Stool consistency, colored based on if mice showed colitis-related symptoms (symptomatic) or not (healthy), was modified based on the human Bristol Stool scale and correlated with the bacterial density; Y-axes are log10-scaled; Spearman's rank correlation with FDR corrected q-values; q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; n(healthy / preC / colitis / post colitis /)= 78 / 4 / 32 / 7

While the microbial alpha-diversity was not affected by colitis, differences in community compositions between healthy and symptomatic mice were increased based on the Bray-Curtis dissimilarity (Figure 17, Table S12, R = 0.41, p = 0.001, ANOSIM; PCoA1 axis: p = 3.9e-9, WRST).

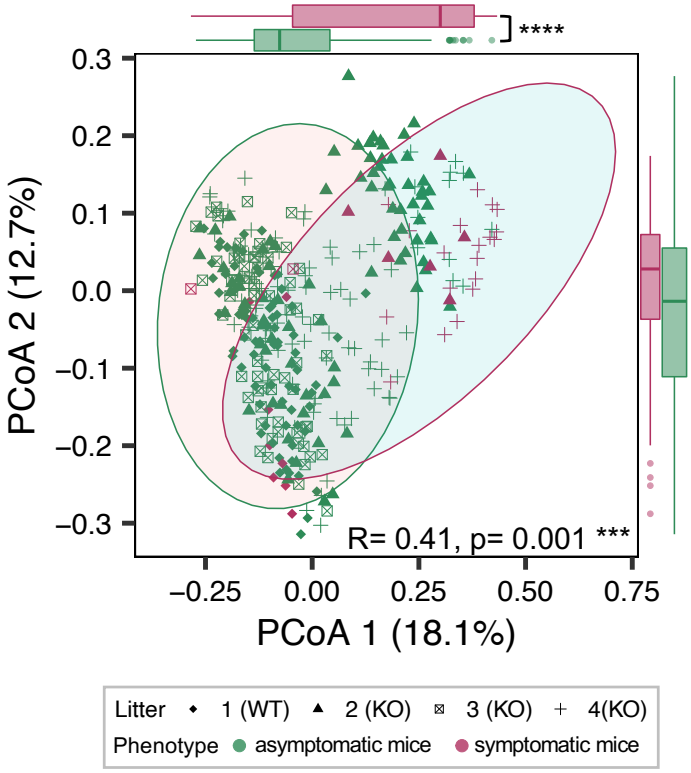


Figure 17. Altered microbiota compositions in mice with colitis.
Dissimilarities between microbial communities of mice increased in the state of colitis (n= 33) compared to asymptomatic mice (n=330), predominately along PCoA axis 1. Microbial community dissimilarities were based on the Bray-Curtis Dissimilarity and globally tested via ANOSIM and along each PCoA axis via pairwise WRST with p-values only shown if $p \leq 0.05$, $p > 0.05$ not significant, $p \leq 0.001$ ***, $p \leq 0.0001$ ****

To identify which taxon abundances were altered during colitis, a generalized linear mixed effect model (GLMM) was used, which allows to incorporate and analyse multiple variables likely to affect microbial abundances at the same time, alongside controlling for repeated measurements within individuals. In order to examine taxon abundance alterations, the state of colitis (scores ≥ 2) and the time of sample collection, i.e. the classification of the healthy samples into *preC* and *post-colitis* were included as independent variables into the model.

Four taxa were associated with the state of colitis based on both relative and absolute abundance: the two families *Erysipelotrichaceae* and *Muribaculaceae*, as well as one uncultured *Bacteroidales* genus, which also belongs to the family *Muribaculaceae*, and the genus *Lactobacillus*. All taxa were characterized by a decrease in abundance during

inflammation (*Muribaculaceae*: $q_{relative} \leq$ 1e-4, $q_{absolute}$ = 0.027; uncultured *Bacteroidales*: $q_{relative} \leq$ 1e-4, $q_{absolute}$ = 0.0003; *Lactobacillus*: $q_{relative} \leq$ 1e-4, $q_{absolute}$ = 0.041, $GLMM^{FDR}$), except for *Erysipelotrichaceae*, which increased ($q_{relative} \leq$ 1e-4, $q_{absolute}$ = 0.0082, $GLMM^{FDR}$, for log odds ratios and statistics of the GLMM see Tables S6 and S7) (Figure 18a, b, page 83, Figures S9, S10). For the relative abundance only, the same alterations for the decreasing taxa were also observed in healthy samples of mice collected after they recovered from colitis-related symptoms, referred to as *post colitis* (Figure 18a, page 83, Figure S9, *Muribaculaceae*: q = 0.0044, uncultured *Bacteroidales*: q = 0.016, *Lactobacillus*: q = 0.0051, $GLMM^{FDR}$). Interestingly, a significant alteration for the relative abundances of the uncultured *Bacteroidales* genus (q = 0.035, $GLMM^{FDR}$) and the family *Erysipelotrichaceae* (Figure 18a, page 83, Figure S9, q≤ 1e-4, $GLMM^{FDR}$), as well as a trend of the same altered absolute abundances, was found in *preC* samples (Figure 18b, page 83, Figure S10).

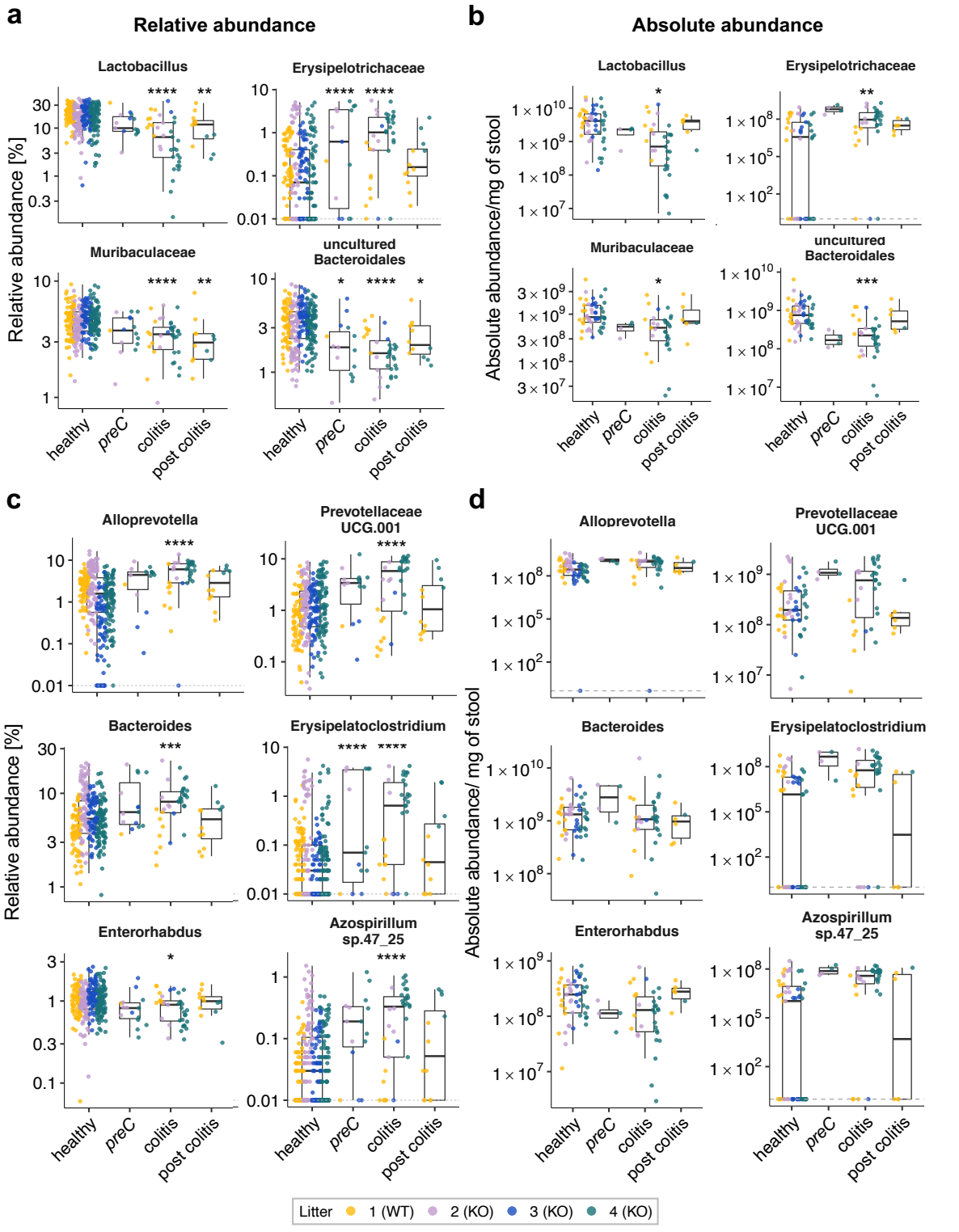


Figure 18. Colitis-specific microbiota signatures based on relative and absolute taxon abundances.

A GLMM was used to model alterations of relative (a, c) and absolute (b, d) taxon abundances of microbial community members on genus level based on the phenotype (*healthy*: well-being and consistency score < 2, *colitis*: scores ≥ 2), the time of sample collection (*preC*: healthy samples up to 4 days collected prior to symptoms, *post colitis*: healthy samples thereafter), as well as the litter and cage number for all samples collected after the first week of cage allocation. Abundances were centered and

scaled after pseudocount addition (a: 0.0001 relative abundance, b: 1 taxon count) and colored by litter. Repeated measurements within individuals were controlled for as random effects. BH-adjusted p-values only shown if $q \leq 0.05$ with $q \leq 0.05$ *, $q \leq 0.01$ **, $q \leq 0.001$ ***, $q \leq 0.0001$ ****. Significances reported refer to comparisons against the healthy state only. The horizontal dashed line indicates the threshold of pseudocount addition if no taxon counts were contained in a sample for each taxon respectively. Y-axes are log10-scaled. (a, c) n(healthy/ preC / colitis / post colitis) = 351 / 16 / 38 / 12, (b, d) n(healthy/ preC / colitis / post colitis) = 55 / 4 / 32 / 6

Additionally, the relative abundance of six genera was significantly associated with colitis, including increases of *Alloprevotella* ($q_{relative} \leq$ 1e-4), *Prevotellaceae UCG 001* ($q_{relative} \leq$ 1e-4), *Bacteroides* ($q_{relative}$ = 0.0002), *Erysipelatoclostridium* ($q_{relative} \leq$ 1e-4), and *Azospirillum sp. 47_25* ($q_{relative} \leq$ 1e-4) and a decrease of *Enterorhabdus* ($q_{relative}$ = 0.031, all $q_{absolute} > 0.05$, $GLMM^{FDR}$) (Figure 18c, d, page 83, Figures S9, S10). For three of these genera, trends of absolute abundance alterations were observed, based on FDR uncorrected $p \leq 0.05$ and large log of the odds ratios (logOR): increased *Azospirillum sp. 47_25* (p = 0.0395, logOR = 7.11, 1.03 – 13.19 95% Wald confidence interval [CI]), *Erysipelatoclostridium* (p = 0.059, logOR = 3.72, -0.09 – 7.53 95% CI) and decreased *Enterorhabdus* (p = 0.042, logOR = -0.66, -1.29 – -0.04 95% CI) (for all taxa $q_{absolute} > 0.05$, $GLMM^{FDR}$, Table S7, Figure S10). The same changes were observed as non-significant trends in *preC* samples of both datasets, indicating that microbial alterations might precede colitis onset (Figure 18c, d, page 83, Figures S9, S10). Detailed statistics of the GLMM outputs and ORs are listed in Tables S6 and S7.

c) Cage and litter additionally attribute to microbiome alterations in the context of colitis

As litter and the co-housing have been shown to determine microbial variations among asymptomatic mice (Figure 12, page 73 and Table 6, page 79), litter as well as cage were included into the GLMM as independent variables, too.

The genus *Parabacteroides* was the only taxon significantly associated with the two cages without co-housing of wild type mice (cage 5 and 6) and colitis development (Figure 19a, b, page 86, $colitis_{relative\ abundance}$: logOR = -1.0021, -1.0015 – 1.0018 95% CI, q≤ 1e-4; cage $5_{absolute\ abundance}$: logOR = -15.30, -18.73 – -11.87 95% CI, q ≤ 1e-4; cage $6_{\ absolute\ abundance}$: logOR = -16.42, -19.82 – -13.03 95% CI, q≤ 1e-4, $GLMM^{FDR}$). *Parabacteroides* increased in relative but not absolute abundance during colitis, and based on the absolute abundance was only detected in cages with Swiss WT mice (Figure 19c - f, page 86, Tables S6 and S7). In fact, *Parabacteroides* has been identified before as being only present in the WT litter at the day of cage allocation (Figure 13, page 76), suggesting a cohousing-induced transmission to *IL-10*$^{-/-}$ mice. In summary, the example of *Parabacteroides* illustrates cohousing as an important confounding factor in murine studies that in practice can make it difficult to distinguish between microbiota associations with co-housing and colitis when only relative abundance data is considered.

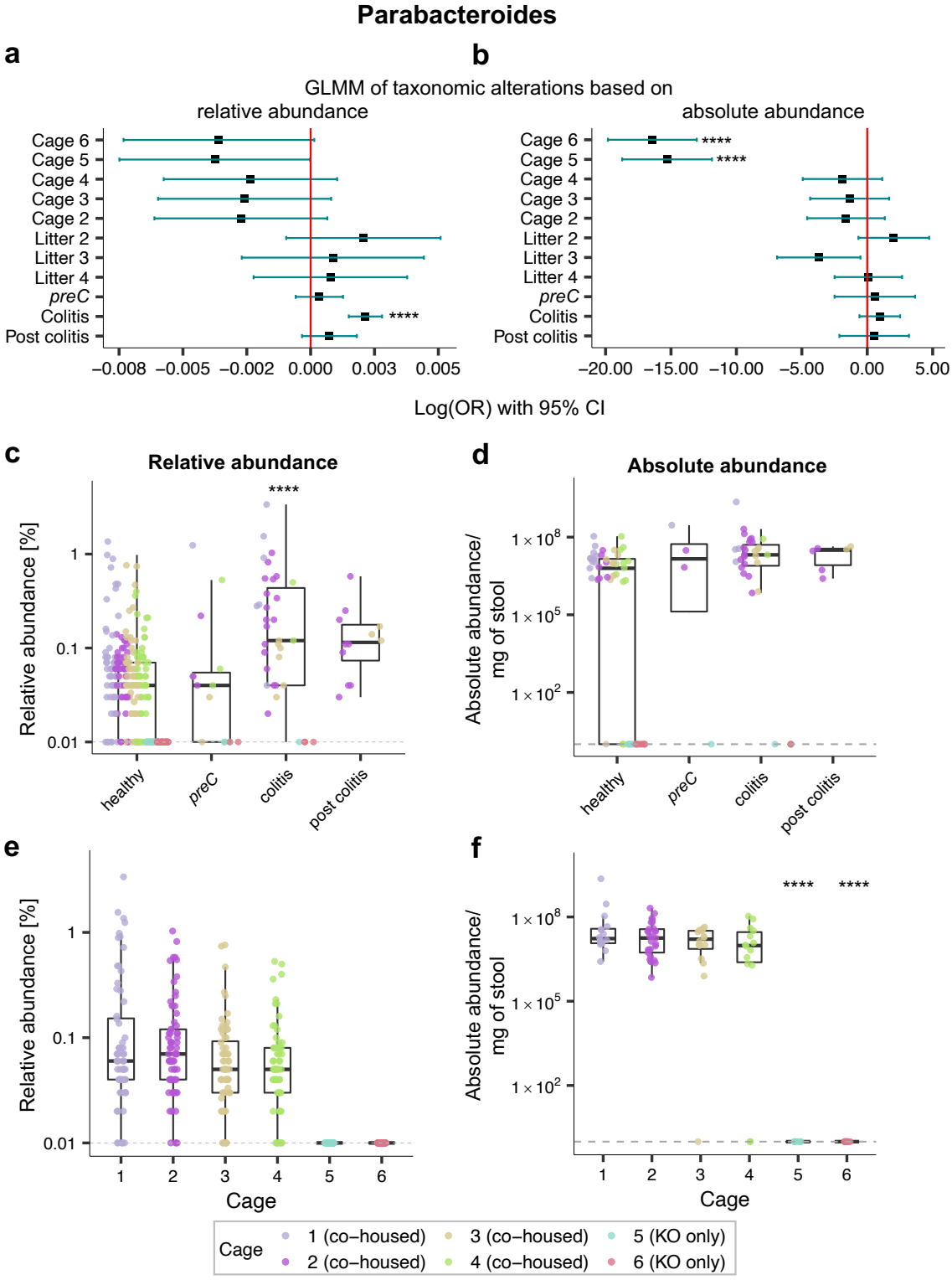


Figure 19. Association of altered *Parabacteroides* abundances with co-housing and colitis.

Variations of relative (a, c, e) and absolute (b, d, f) abundances of the genus *Parabacteroides* were modelled by GLMM based on the phenotype (healthy: well-being and consistency score ≤ 2, colitis: scores ≥ 2), the time of sample collection (*preC*: healthy samples up to 4 days collected prior to symptoms, post colitis: healthy samples thereafter), as well as the litter and cage number for all samples collected after the first week of cage allocation. Abundances were centered and scaled after

pseudocount addition (0.0001 relative abundance, 1 taxon count) and colored by the cage (c - f). Repeated measurements within individuals were controlled for as random effects. BH-adjusted p-values only shown if $q \leq 0.05$ with $q \leq 0.05$ *, $q \leq 0.01$ **, $q \leq 0.001$ ***, $q \leq 0.0001$ ****. Significances reported refer to comparisons against the healthy state (a - d), litter 1 (a, b) and cage 1 (a, b, e, f) only. The horizontal dashed line indicates the threshold of pseudocount addition if no taxon counts were contained in a sample for each taxon respectively and y-axes of (c - f) are log10-scaled. (a, c, e) n(healthy / preC / colitis / post colitis) = 351 / 16 / 38 / 12, (b, d, f) n(healthy/ preC / colitis / post colitis) = 55 / 4 / 32 / 6

Litter was linked to compositional microbiota variations in 14 genera based on the absolute abundance. Six of the 14 associated genera were additionally linked to colitis and most interestingly, all of them were associated with litter 2: uncultured *Bacteroidales* ($q_{\text{litter 1 vs 2}}$ = 0.0024), *Muribaculaceae* ($q_{\text{litter 1 vs 2}}$ = 0.00017), *Bacteroides* ($q_{\text{litter 1 vs 2}} \leq$ 1e-4), *Enterorhabdus* ($q_{\text{litter 1 vs 2}}$ = 0.034), *Prevotellaceae UCG 001* ($q_{\text{litter 1 vs 2}}$ = 0.0091) and *Akkermansia* ($q_{\text{litter 1 vs 2}}$ = 0.0091, $GLMM^{FDR}$) (Figure S10, Table S7).

Of these taxa, the genus *Akkermansia* was the only taxon which was simultaneously associated with colitis (logOR with 95% CI: 5.36, 1.87 – 8.85, q= 0.047) and the two litters with severe intestinal inflammation: litter 2 (logOR with 95% CI: 15.33, 9.3 – 21.37, q= 0.009) and 4 (logOR with 95% CI: 11.47, 5.72 – 17.21, q= 0.048) (Figure 20a, b, page 88, $GLMM^{FDR}$). The link between colitis development, litter and *Akkermansia* became even more apparent when plotting relative and absolute abundances dependent on the mice's colitis status (Figure 20c, d, page 88) and over time (Figure 20e, f, page 88). A high relative and absolute abundance of *Akkermansia*, was only detected in mice from litter 2 and 4, whereas it was mostly absent from the healthy knockout litter 3 and wild type mice despite co-habitation (Figure 20e, f, page 88). Furthermore, *Akkermansia* was the only taxon of the colitis and litter-associated bacteria, that was not present in *preC* samples of all 22 mice, suggesting that the presence (relative abundance > 0) rather than the abundance of *Akkermansia* could be a predictor of subclinical early colitis (Figure 20e, f, page 88).

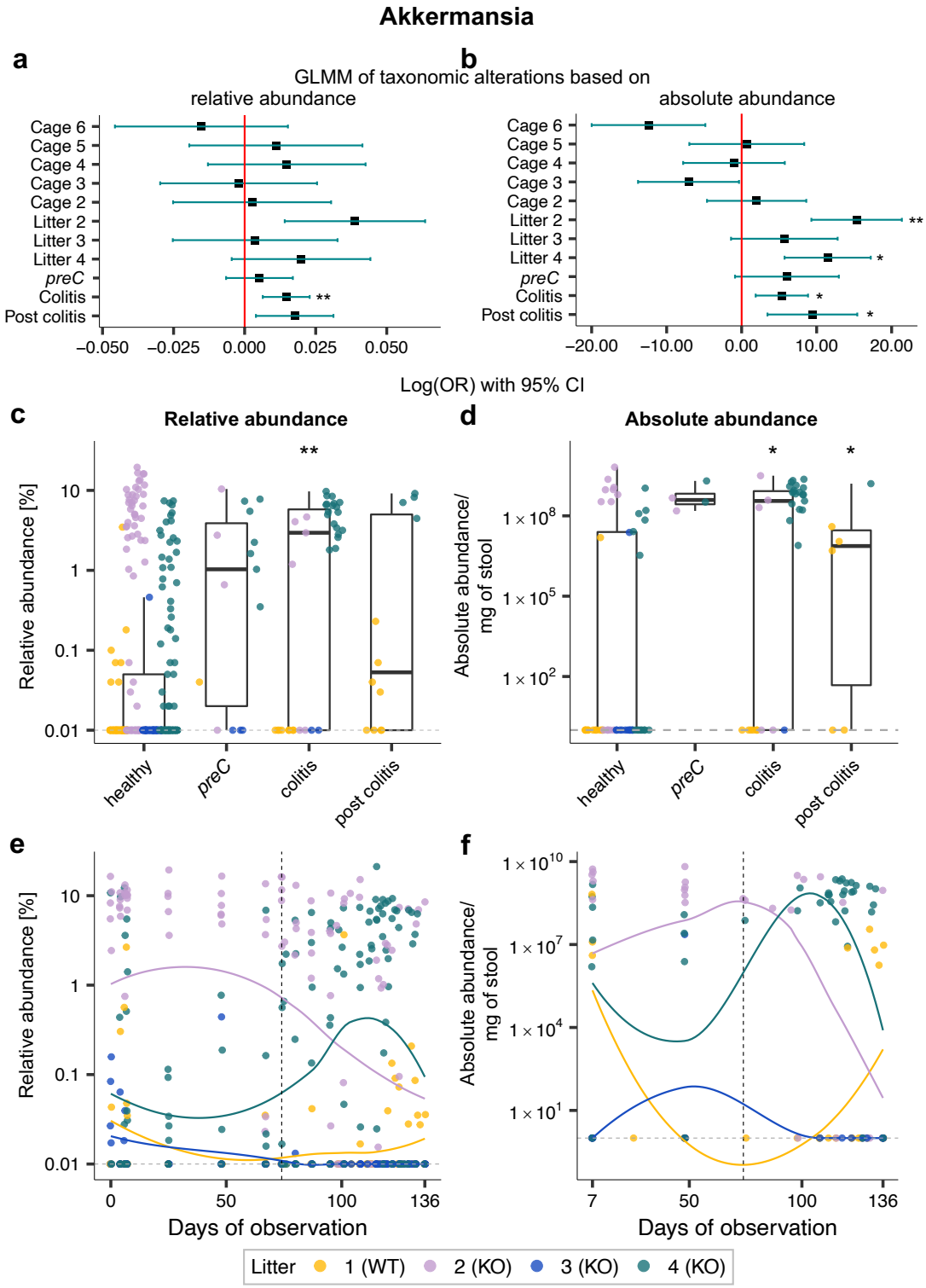


Figure 20. The genus *Akkermansia* was associated with an increase during colitis relatively and absolutely and was additionally linked to specific IL-10 deficient litter.
Variations of relative (a, c, e) and absolute (b, d, f) abundances of the genus *Akkermansia* were modelled by GLMM based on the phenotype (healthy: well-being and consistency score ≤ 2, colitis: scores ≥ 2), the time of sample collection (*preC*: healthy samples up to four days collected prior to symptoms, post colitis: healthy samples thereafter), as well as the litter and cage number for all samples collected after the first week of cage allocation. Abundances were centered and scaled after

pseudocount addition (0.0001 relative abundance, 1 taxon count) and colored by cage (c - f). Repeated measurements within individuals were controlled for as random effects. BH-adjusted p-values only shown if q≤ 0.05 with q≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****. Significances reported refer to comparisons against the healthy state (a - d), litter 1 (a, b) and cage 1 (a, b, e, f) only. The horizontal dashed line indicates the threshold of pseudocount addition if no taxon counts were contained in a sample for each taxon respectively and y-axes of (c - f) are log10-scaled. (a, c) n(healthy / preC / colitis / post colitis) = 351 / 16 / 38 / 12, (b, d) n(healthy/ preC / colitis / post colitis) = 55 / 4 / 32 / 6, (e) n(litter 1 / 2 / 3 / 4)= 105 / 128 / 92 / 150, (f) n(litter 1 / 2 / 3 / 4)= 29 / 30 / 17 / 44

d) The presence of *Akkermansia* is a good predictor for early subclinical colitis in IL-10 deficient mice

To further study the potential of predicting which mice would develop colitis and identify the factors attributing to it, firth penalized logistic regression (pLR) was used to model if mice had to be withdrawn early from the observation due to severe colitis. In these models, the outcome is a binary variable (early withdrawal: yes/no), which in the first step is modelled for each predictor separately and its prediction is then compared to the baseline model. When this single-predictor model significantly improves the prediction of the baseline model, the predictor is considered relevant and included into the final multi-predictor model. As predictors, all variables associated in this thesis either with colitis far before clinical manifestation ("very early colitis susceptibility", e.g. determined by litter, genotype, the presence of *Akkermansia* at the day of cage allocation and differences in α- and β-diversity of healthy microbial communities), immediately before clinical manifestation ("subclinical colitis onset", e.g. the presence of *Akkermansia* in *preC* samples), or the experimental set up (e.g. co-housing) were used.

The only variables that significantly improved the prediction compared to the baseline model were the presence of *Akkermansia* in *preC* samples (p= 0.00069), litter (p= 0.046) and the mean Bray-Curtis dissimilarity between the *IL-10*$^{-/-}$ and WT mice at the day of cage allocation (Table S8, p= 0.047, Likelihood ratio test). Combining them into a single pLR model revealed two findings: (1) the potential of the mean Bray-Curtis dissimilarity as very early colitis susceptibility marker is masked by interdependencies between the predictors and (2) the presence of *Akkermansia* in *preC* samples is the only significant predictor for early mice withdrawal.

Although the mean Bray-Curtis dissimilarity improved the prediction, interdependencies between predictor and the binary outcome were detected, as indicated by large regression coefficients and CIs, which are indicators of almost perfect prediction (Figure S11, Table S9, mean Bray-Curtis dissimilarity: p = 0.099, logOR = 51.63, -180.37 – 3.50 95% CI, pLR). This almost perfect prediction is determined when one or more predictors separate the binary outcome almost perfectly, also known as quasi-complete separation (D. A. Walker & Smith, 2020). In the present data, the early mice withdrawal is largely explained by the mean Bray-

Curtis dissimilarity between *IL-10*$^{-/-}$ and WT mice at the day of cage allocation. Additionally, both can be stratified by litter, resulting in early withdrawal of the mice from litter 2 and 4, which are also largely characterized by a high mean Bray-Curtis dissimilarity compared to litter 1 and 3 mice. This indicates the possibility of microbial markers that could determine colitis susceptibility very early on, e.g. dependent on the litter, even prior to subclinical mucosal inflammations, but which cannot be clearly identified with the present data.

Nevertheless, including the mean Bray-Curtis dissimilarity into the final pLR model (model 2) or not (model 1) resulted in the same finding, which revealed the presence of *Akkermansia* in *preC* samples to be the only predictor of early mice withdrawal (Figure 21a, page 91, model 1: p = 0.023, logOR = 2.47, 0.32 – 5.21 95% CI; Figure S11, model 2: p = 0.018, logOR = 5.91, 0.498 – 18.06 95% CI, pLR, Table S9). This result was further supported by assessing the performance of the models based on samples which were used for modelling (*in-sample* prediction) and based on samples the model was not trained on (*out-of*-sample prediction). *In-sample* prediction was determined by receiver operating characteristics (ROC) curves and an area under the curve (AUC) of 0.944 for model 1 (model 2: AUC = 1) (Figure 21b, page 91, Figure S11). *Out-of-sample* prediction was evaluated by a stratified 10-fold cross validation (pLR^{CV}) with the data split iteratively into training (80%) and test (20%) sets, which contained samples in the original proportion. The best model was determined based on the Akaike information criterion (AIC) with an 85% accuracy and a lambda of 0.127 for model 1 (model 2: accuracy = 95%, lambda= 0.1004), resulting in similar significant model predictions for the presence of *Akkermansia* in *preC* samples (model 1: p = 0.017, logOR = 0.986, 0.176 – 1.796 95% CI; model 2: p = 0.018, logOR = 2.312, 0.393 – 4.230) and a trend for litter 4 (model 1: p = 0.06, logOR = 0.723, -0.031 – 1.478 95% CI; model 2: p = 0.078, logOR = 0.947, -0.106 – 1.999 95% CI pLR^{CV}, Table S10). This suggests that the presence or absence of *Akkermansia* shortly prior to colitis manifestation rather than its abundance predicts severe colitis development, pointing towards *Akkermansia* being a potential subclinical biomarker of early colitis onset in *IL-10*$^{-/-}$ mice.

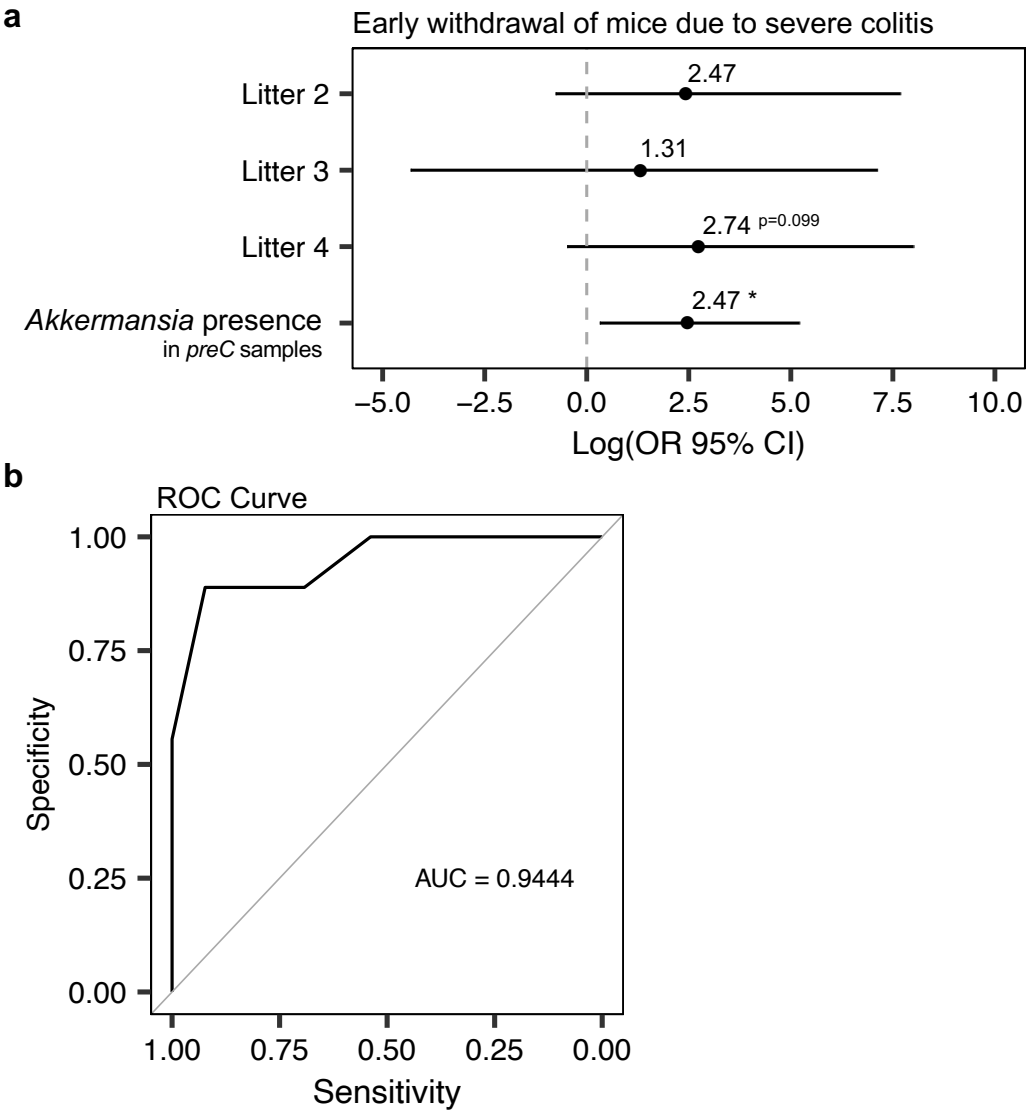


Figure 21. The presence of *Akkermansia* shortly before the first clinical manifestation of colitis is a good predictor for severe colitis development.
(a) Predictors of early mice withdrawal due to severe colitis were estimated based on Firth penalized logistic regression for litter number and the presence of *Akkermansia* up to four days prior to the first clinical manifestation (*preC*).
(b) The predictive power of the model was evaluated by receiver operating characteristics (ROC) curve indicating a good *in-sample* prediction with a high area under the curve (AUC). p> 0.05 not significant, p≤ 0.05 *, q≤ 0.01 **, q≤ 0.001 ***, q≤ 0.0001 ****; n= 22

IV. Discussion

Intestinal microbiome perturbations have been linked to multiple and diverse health conditions, including metabolic, gastrointestinal, cardiovascular and inflammatory diseases (Belizário et al., 2018; Fujimura et al., 2015; Halfvarson et al., 2017; Ohigashi et al., 2013; Turnbaugh et al., 2006a). Although there is consensus about which factors can modulate the microbiota, developing general guidelines and treatments targeting the microbial community is challenging due to huge interindividual (Falony et al., 2016; Schnorr et al., 2014) and even intraindividual variations over time (Johnson et al., 2019; C. A. Lozupone et al., 2012; Poyet et al., 2019). Therefore, the microbiotas potential as target for personalized instead of general treatment approaches is of interest. This however requires detailed characterizations of microbiota modulations and microbiome-based patient stratification, which is investigated in the context of two different microbiome-modulation factors and diseases in this thesis: (1) metabolic consequences of dietary microbiome-modulation by low-carbohydrate high-fat diet (LCHF) with relevance for atherosclerotic cardiovascular disease (ACVD), and genetic predisposition to inflammatory bowel diseases (IBD) in an *IL-10* knockout mouse model.

A. The microbiota determines cholesterol conversion and is predictive for the converter type-dependent LDL-C response to low-carbohydrate high-fat diets

The excess in blood cholesterol is a major risk factor for ACVD (Barquera et al., 2015). Besides pharmacological and dietary therapies, limitations of intestinal cholesterol absorption by regulating the microbial conversion of both endogenous and exogenous cholesterol to non-absorbable coprostanol in the intestine is conceptually appealing as a potential alternative (Kriaa et al., 2019). However, the experimental evidence from previous studies for the feasibility of this approach is limited. Several studies suggest a causal involvement of the microbiota in blood cholesterol regulation, albeit without directly implicating microbial cholesterol-to-coprostanol conversion: Germ-free rats (Danielsson, H., Gustafsson, 1959) and antibiotically treated *ApoE*$^{-/-}$ mice, which spontaneously develop hypercholesterolemia and are a widely used model to study atherosclerosis (Lo Sasso et al., 2016), exhibited elevated serum cholesterol levels (Le Roy et al., 2019), and fecal transplantation from humans with excess serum cholesterol to microbiota-depleted *ApoE*$^{-/-}$ mice could replicate these phenotypes (Le Roy et al., 2019). Kenny et al. recently identified the bacterial cholesterol dehydrogenase gene *ismA*, classified humans as encoders and

non-encoders based on the presence of *ismA*-carrying bacterial species in their fecal metagenomes, and found decreased serum total cholesterol levels in *ismA*-encoders (Kenny et al., 2020). These findings are consistent with the described reducing potential of the intestinal microbiota on cholesterol availability in the gut, although fecal cholesterol and coprostanol levels were not determined.

In the two geographically separated and metabolically distinct human cohorts, i.e. with and without obesity, in response to LCHF interventions analyzed herein, no significant association between fecal cholesterol or coprostanol concentrations and circulating cholesterol levels in the blood was observed, suggesting independent regulatory mechanisms for human cholesterol metabolism and microbial cholesterol conversion. Conceptually, this could indicate that microbial cholesterol conversion is spatially separated from and takes place after human cholesterol secretion and absorption. Microbial cholesterol conversion activities have not been mapped within the gastrointestinal tract, but may be restricted to the large intestine, where slower transit times, a lack of simple carbon sources and lower antimicrobial concentrations favor the growth of fermentative polysaccharide-degrading anaerobes, such as members of the taxonomic class *Clostridia* (Donaldson et al., 2015), which includes *E. coprostanoligenes*, a known cholesterol-to-coprostanol converter (Freier et al., 1994). In this scenario, rather than affecting cholesterol availability in the small intestine, microbial cholesterol-to-coprostanol activity could be dependent on cholesterol concentrations in the large intestine. It may thereby reflect cholesterol remainders after small intestinal passage and serve as a potential indicator of the net outcome of cholesterol intake and absorption.

The rate of cholesterol-to-coprostanol conversion varies between individuals, following a bimodal distribution of high and low-converters based on the ratio of fecal coprostanol/cholesterol concentrations (Salyers et al., 1977; Veiga et al., 2005; Wilkins et al., 1974). Previous studies reported similar fractions of high and low cholesterol-to-coprostanol converters in human populations from different European countries (Benno et al., 2009; Norin, 2008), the converter type being stable in individuals over at least several days (Kunz et al., 2019), and a decrease in the proportion of low-converters among male elder individuals (Benno et al., 2009; Norin, 2008). However, the longer-term stability of intestinal cholesterol conversion, its dependence on diet, and its relevance for human circulating cholesterol levels and metabolic health have not been conclusively determined yet (Juste et al., 2021). Herein, conserved fractions of cholesterol high and low-converter types were found in two independent, German and Norwegian, cohorts of individuals with

(CARBFUNC) and without (KETO) obesity, despite adverse metabolic health parameters in individuals with obesity (increased triacylglycerides [TAG] and low-density lipoprotein cholesterol [LDL-C] and decreased high-density lipoprotein cholesterol [HDL-C] levels). Moreover, LCHF diets led to consistent increases in cholesterol-to-coprostanol conversion in low-converters from both cohorts, despite differences in the ketogenic potential of the LCHF diets. The findings therefore suggest that cholesterol converter types represent an obesity-independent organizational feature of the human fecal microbiome that is conserved across different European regions and is amenable to modulation by LCHF dietary intervention.

Among lean study participants only, the high cholesterol-to-coprostanol converter type was associated with increased serum TAG and decreased HDL-C levels compared to the low-converter type. Furthermore, lean cholesterol high-converters responded to the ketogenic LCHF diet with a greater increase in LDL-C. In support of these distinctive LDL-C responses in lean individuals, a recent study also reported that lean people with low TAG and high HDL-C experienced a marked increase in LDL-C when adopting a LCHF diet, unlike people with obesity who showed a blunted increase (Norwitz et al., 2022). The data herein suggest that cholesterol high-conversion may contribute to diet-induced serum lipid responses or at least serve as a prognostic biomarker to predict such inter-individually variable, personalized responses. Whether the increase in LDL-C translates into an elevated cardiovascular disease risk or possibly reflects a homeostatic adaptation to redistribute cholesterol molecules between specific cells and tissues in response to increased dietary fatty acid intake, as proposed in the homeoviscous adaptation to dietary lipids model (Zinöcker et al., 2021), remains to be determined.

Cholesterol conversion has previously been demonstrated *in vitro* for a taxonomically diverse set of bacterial isolates, including from the genera *Bacteroides* (Gérard et al., 2007), *Bifidobacterium* (Zanotti et al., 2015), *Eubacterium* (Freier et al., 1994), and *Lactobacillus* (Lye et al., 2010). Kenny et al. metagenomically mapped homologs of the bacterial cholesterol dehydrogenase gene *ismA* to *E. coprostanoligenes* and different species within the phylogenetic *Clostridium* cluster IV, but found no evidence for their presence in available genome sequences from the bacterial genera listed above (Kenny et al., 2020). The findings herein do not suggest a broadly distributed microbial capacity for cholesterol-to-coprostanol conversion, as only *E. coprostanoligenes*, the well-described cholesterol-converting bacterial species originally isolated from a hog sewage lagoon (Dewei et al., 1996; Freier et al., 1994; L. Li et al., 1998), and to a lesser extent *Ruminococcaceae UCG 014*, showed

consistent positive correlations to fecal coprostanol levels, both in individuals with and without obesity. Moreover, the data argue against the previously suggested, *ismA* detection-based, categorical classification of individuals as either cholesterol converters or non-converters (Kenny et al., 2020), as *E. coprostanoligenes* was detected in the vast majority of fecal samples (96%) and individuals (97.6%) from both cohorts. The cholesterol converter phenotype therefore appears to be dependent on the relative abundance rather than the presence or absence of *E. coprostanoligenes* or other species with similar genetic potential for cholesterol-to-coprostanol conversion.

In contrast to coprostanol, which is only generated by microbial cholesterol conversion, the fecal concentration of cholesterol is dependent on a complex interplay of dietary intake, endogenous production, secretion, absorption, and microbial conversion (Juste et al., 2021). Accordingly, a positive correlation of the genus *Lachnoclostridium* with fecal cholesterol concentrations was identified, but no association with coprostanol. This genus has previously been linked to adverse phenotypes and clinical conditions, including visceral fat (Nogal et al., 2021), increased serum total cholesterol and LDL-C (Y. Wang et al., 2020), trimethylamine-N-oxide (TMAO) biosynthesis (Jameson et al., 2016), colorectal cancer (T. Li et al., 2020), and in mice, azoxymethane/dextran sulfate sodium (AOM/DSS)-induced colitis (C.-S.-E. Wang et al., 2018), and high-fat diet (Jo et al., 2021). Nevertheless, functional studies will be needed to determine if the genus *Lachnoclostridium* has a direct detrimental effect on human health and if *E. coprostanoligenes* and *Lachnoclostridium* could serve as biomarkers for distinct but complementary components of the human and microbial cholesterol metabolism.

The human dietary intervention study has several limitations: Despite the inclusion of two distinct cohorts, the overall patient numbers and especially the fraction of cholesterol low-converters, which represents only a minor proportion of both cohorts, are very small, providing limited statistical support for comparative analyses. Different metabolic health backgrounds and responses of the study participants with and without obesity to the distinct dietary interventions, i.e. a ketosis-inducing LCHF diet of six weeks (KETO study) or a LCHF diet with limited ketosis induction for six months (CARBFUNC study), introduce additional confounding factors. However, the combination of these two heterogeneous cohorts was also instrumental in the identification of what appear to be conserved structural microbiome features, i.e. the stratification of microbiomes into a larger cholesterol high-converter fraction and a smaller low-converter fraction that can be increased with LCHF dietary modulation. Moreover, no significant differences in fecal sterol and stanol concentrations

and serum lipid levels were detected between the three and six months' time points after LCHF diet intervention in the CARBFUNC study ($p > 0.05$, Wilcoxon ranked-sum test), attesting to the temporal robustness of the reported findings. Finally, a shotgun metagenomics approach for fecal microbiota analysis could potentially identify associations between specific orthologous gene groups or metabolic pathways and microbial cholesterol conversion. However, since the mechanistic and genetic basis of cholesterol-to-coprostanol conversion, which based on the correlation analysis appears to be mostly restricted to *Eubacterium coprostanoligenes*, remains largely unknown, the added value of a metagenomic sequence analysis may still be limited at this point.

In summary, the findings are in agreement with a model that explains individual intestinal cholesterol-to-coprostanol converter types as a result of long-term dietary or other habits. They do not indicate a hypocholesterolemic effect of intestinal microbial cholesterol conversion but point to an adverse response of increased LDL-C levels to ketogenic diet in lean cholesterol high-converters. Finally, the data suggest potential relevance of the cholesterol converter type as a personalized microbiome marker for metabolic health and response to dietary intervention.

B. The presence of *Akkermansia* is a subclinical marker for early colitis in IL-10 deficient mice and colitis susceptibility is associated with litter-specific microbiota compositions

The knockout of the anti-inflammatory cytokine *IL-10* in mice leads to spontaneous colitis development and resembles histopathological alterations similar to IBD in humans, particularly of Crohn's disease (CD), which are characterized by inflammatory cell infiltration into the lamina propria and submucosa, and transmural, discontinuous inflammatory lesions (Bleich et al., 2004; Kühn et al., 1993). Furthermore, they are accompanied by altered expression profiles of increased proinflammatory cytokines like Tumor necrosis factor alpha (TNF), Chemokine C ligand (CCL) 2 and CCL3, and a decrease of anti-inflammatory Transforming growth factor (TGF) (Berg et al., 1996; Hagenlocher et al., 2016). As these pathological changes can only be evaluated after withdrawing mice from an observation, characterizations of colitis based on the phenotypic scoring of the mice's well-being have been proposed: disproportional weight loss (> 10%), changes in behavior and appearance, and absolute criteria indicating a severe state of disease like rectal prolapse or bleeding (Hagenlocher et al., 2016). Herein, the phenotypic scoring parameters mirrored the state of inflammation during colitis well based on histopathological examinations and cytokine expression, which validates the classification of mice into asymptomatic/healthy (any score based on the mice's well-being < 2) and symptomatic (any score based on the mice's well-being ≥ 2) mice throughout the observation.

IBD diagnosis and differentiation of IBD subtypes in humans is also limited to gastrointestinal endoscopy and histopathological evaluation of mucosal tissue (Maaser et al., 2019), which are conducted commonly only after symptomatic manifestations. But recent studies report inflammatory processes in the mucosa to precede symptoms (Rodríguez-Lago et al., 2020; Torres et al., 2021), which imposes the question of how colitis should be defined and measured to include pre-clinical periods. As routinely gastrointestinal endoscopies are cost-intensive and inconvenient for the patient, the identification of non-invasive, subclinical biomarkers is of main interest.

Stool consistency, as measured by the Bristol Stool Scale (Lewis & Heaton, 1997), is proposed to be one of these potential biomarkers since pathological alterations of IBD are often accompanied by changes in stool consistency, mostly diarrhea (Mowat et al., 2011). It is also associated with microbiome markers like species richness and the growth of specific bacterial community members, most likely due to altered gastrointestinal transit times (Vandeputte et al., 2016). Although a positive correlation of increasing stool

consistency and histopathological scores was observed in the presented work, associations with the cytokine expression profile were inconsistent. Stool consistency was further negatively correlated with the fecal bacterial density as estimated based on the 16S rRNA gene copy number per mg stool, which was significantly reduced in symptomatic mice. This suggests stool consistency as an important but independent parameter of microbial ecology, which is not as sensitive as indicator of colitis in mice as the established scores. It rather changes in later stages of inflammation, subsequent to early events, as it is associated with signs of severe inflammation that lasted over a longer period, e.g. tissue alterations, rather than cytokine expressions, which are released from cells in short-terms upon activation (J.-M. Zhang & An, 2007).

Colonization of germ-free IL-10 deficient mice with conventional and specific pathogen free (SPF) microbiota was shown to lead to intestinal inflammation within one week after treatment (Sellon et al., 1998). One perception in the multihit model etiology of IBD is that an abnormal immune response against commensals of the intestinal microbiota triggers colitis in genetically susceptible individuals rather than the presence of particular pathogens (Keubler et al., 2015). But whether alterations of the commensal community provoke the immune response or if the microbiota changes subsequently to it, is still an open question. Despite inconsistent descriptions of IBD-associated bacterial communities in humans, they appear to generally be characterized by reduced microbial diversity (α-diversity) (Franzosa et al., 2019; Sellon et al., 1998). In the present study, no reduction in neither microbial richness nor eveness due to intestinal inflammation was found. Nevertheless, colitis-related communities were associated with higher dissimilarity of the microbiota composition (β-diversity) compared to healthy animals, as well as decreased overall bacterial microbiota density as estimated based on the 16S rRNA gene copy number per mg stool. This suggests compositional and quantitative microbiota alterations to happen during colitis, without a concomitant loss of α-diversity in *IL-10*$^{-/-}$ mice. One study previously reported reduced fecal microbial cell densities in patients with CD compared to a healthy cohort based on the enumeration of microbial cells via flow cytometry (Vandeputte et al., 2017), which is in line with the herein reported lower bacterial density of 16S rRNA gene copies per mg stool in symptomatic mice measured by quantitative polymerase chain reaction (qPCR). In the present work, the bacterial density was also negatively correlated with stool consistency which could result from shorter transit time during intestinal inflammation and therefore higher water and mucus content and lower numbers of bacteria excreted per mg feces (Lewis et al., 1997; Vandeputte et al., 2016).

Although the knockout of the *IL-10* gene sufficiently induces spontaneous colitis in mice, colitis susceptibility varies between *IL-10*$^{-/-}$ murine strains depending on their genetic background. Inbreed IL-10 deficient C57BL/6J mice are relatively resistant to spontaneous intestinal inflammation, whereas others are highly susceptible and show first signs of inflammation shortly after weaning, e.g. 129/SvEv mice (Berg et al., 1996; Bristol et al., 2000; Mähler & Leiter, 2002), which has been attributed to multiple variations in quantitative trait loci (Beckwith et al., 2005; Farmer et al., 2001). The IL-10 deficient mice used in the present study have a BALB/c genetic background and are characterized by intermediate colitis susceptibility when comparing the percentage of sick mice and colonic disease scores to IL-10 deficient C57BL/6J and 129/SvEv mice (Berg et al., 1996). Berg et al. found 100% of BALB/c *IL-10* knockout mice to be affected by colitis after twelve weeks, similar to the 129/SvEv strain, but with lower severity scores and intestinal inflammation (Berg et al., 1996). Compared to this study, mice herein were less susceptible to colitis, with only one of 18 mice (5.5 %) being withdrawn from the observation at three months of age. By the end of the experiment, after 20 weeks, 50% (9 of 18) of IL-10 deficient mice showed severe signs of intestinal inflammation. These variations relative to the literature could be attributable to environmental factors such as the animal facility. Even mice with the same genetic background, as well as their subsequent F1 generations have been reported to show consistently reduced histopathology scores in one facility over another, arguing for influences of the mice's surrounding and habitation setting (Mähler et al., 2002). Also, the experimental set up could explain the delayed spontaneous intestinal inflammation, as IL-10 deficient mice were co-housed with wild type mice, leading to transmission of fecal bacteria between cohabitated mice due to their coprophagic trait, which could potentially affect colitis susceptibility.

For various measurements independent and dependent on the microbiota, a stratification based on the *IL-10*$^{-/-}$ litter was observed. Although all IL-10 deficient mice reached a stable body weight at 14 weeks of age, the average mature weight differed between litters, revealing litter 3 to be the heaviest and therefore more similar to the wild type mice. This litter was also the only knockout litter that did not develop severe intestinal inflammation, which classifies litter 2 and 4 as colitis-sensitive litters and litter 3 as resistant litter.
Multiple conditions of microbial ecology were linked to the same classification. Differences in microbial composition based on the Bray-Curtis dissimilarity between *IL-10*$^{-/-}$ and WT mice on the day of cage allocation clustered litters with and without severe colitis development and improved the prediction of early mice withdrawal. This points towards

potential microbial colitis susceptibility markers very early on associated with litter, but which cannot be identified further with the present data. The same cluster into colitis-resistant and colitis-sensitive litters was obtained when including only healthy samples collected after the first week of cage allocation, arguing for litter-dependent microbial signatures that cannot be completely abolished by co-housing. This is in line with recent findings of Singh et al., who associated the microbial variation in a co-housing set up between wild-type and colitis-prone-mdr1a deficient mice significantly with 98% accuracy to maternal trait over 80% by cage via random forest modelling. Also, the genotype was not significantly associated at all in the model (Singh et al., 2021). Consistent differences in microbial richness and evenness within the first week after cage allocation and thereafter identified higher bacterial diversity in the colitis-resistant relative to the colitogenic litters, which is in line with other studies that distinguished more and less IBD-sensitive microbial communities based on α-diversity measurements (Knoch et al., 2010; Maharshak et al., 2013; Wohlgemuth et al., 2011).

The influence of the microbiota on colitis susceptibility was further supported by the fact, that litter 2, the litter that was most prone to early intestinal inflammation, was generally characterized by low α-diversity and distinct microbial compositions before cohousing compared to the others litters. At the day of cage allocation, 18 taxa were abundant in all litters except for litter 2, revealing it to be the most influenced by co-housing, as it acquired all of these taxa later on. Contrarily, the *IL-10*$^{-/-}$ litter 3 which did not develop severe colitis, was the most similar to the wild type mice based on the microbiota composition.

Differences between IL-10 deficient litters could not be explained by familial relationships as all litters contained different parental animals. Furthermore, none of the mothers were directly related, which has been reported to influence microbial communities even over more than one generation (Hufeldt et al., 2010; Ley et al., 2005; Singh et al., 2021; Stappenbeck & Virgin, 2016). Litter 2 and 4 mice were only indirectly related as they shared the same grandfather on the father's site. Bodden et al. recently reported the phenotype but only minorly the microbiota to be altered in mice when their fathers were fed a Western style diet (Bodden et al., 2022). Although paternal inheritance has been shown to influence the next generation, e.g. by epigenetic modification of the germ cells (A. M. Edwards & Cameron, 2014), to my knowledge no reports of microbial alterations spanning two generations and specifically in IL-10 deficient mice have been published yet. Therefore, and as no further information about the paternal grandparents is available, the potential influence of the father's epigenetic modification on colitis susceptibility cannot be completely ruled.

Taken together, these findings revealed colitis susceptibilities to differ between genetically similar mice and to potentially dependent on the microbiota composition they adapted early on in life within their litter. These litter-specific traits could contribute to heterogeneous descriptions of microbial communities between studies and identify litter as an important confounding factor *in vivo* studies. This is concerning and should be critically considered regarding mouse studies that typically aim to control for such effects with a short co-housing adaptation phase.

Differences in the early developmental window in the life of mice might also attribute to litter-specific colitis susceptibilities. The body weight, as well as dissimilarities between microbial communities increased continuously and simultaneously within the first twelve to 14 weeks of age and reached a stable state thereafter irrespective of the mice's' genotype. Interestingly, spontaneous colitis development was also first observed after 14 weeks of age. The definition of mature, adult mice differs between studies, ranging from six to 20 weeks of age. Most often sexual maturation around six weeks of age is used as a hallmark (Drickamer, 1981), which does not necessarily represent fully developed animals as many systems take longer to mature, e.g. the maturation of T-cell responses can take up to eight weeks of age (Holladay & Smialowicz, 2000) and T- and B-lymphocyte production has been shown to increase up to 26 weeks after birth (Kincade, 1981). Similarly in humans, the development of a resilient microbial community takes place especially in early stages of life within the first two to three years, reaching an equilibrium under healthy conditions in adults (J. E. König et al., 2011; Yatsunenko et al., 2012). In the present work, a direct impact on the microbiome of dietary transition from nursing to solid food can be neglected as all mice had access to solid food before weaning, which usually happens at three weeks of age, and mice commonly start to nibble on solid food two weeks after birth (B. König, 1994; B. König & Markl, 1987; Krackow & Hoeck, 1989). Taken together, this suggesting a co-dependency between the adulthood of mice and the maturation of the immune system as well as the microbiota for colitis development, being in line with the multihit model etiology hypothesis for IBD (Keubler et al., 2015).

When general community changes due to colitis were assessed using generalized linear mixed effect models (GLMM), ten bacterial taxa, which are all commensals of the murine gut (Gordon & Dubos, 1970; Yang et al., 2019), were significantly associated with the state of colitis based on their relative abundance: the genera *Lactobacillus*, *Alloprevotella, Bacteroides, Erysipelatoclostridium, Enterorhabdus* and *Azospirillum sp 47_45*, as well as

a genus of the families *Prevotellaceae* and *Erysipelotrichaceae* each, and two genera of the family *Muribaculaceae.*

Associations between colitis and six of these taxa (*Alloprevotella, Bacteroides, Erysipelatoclostridium, Enterorhabdus, Prevotellaceae UCG 001* and *Azospirillum sp 47_45*) could not be validated by absolute abundances determined by qPCR of the universal 16S rRNA gene. Only trends similar to the changes in the compositional data were observed. This could either mean that (1) these taxon alterations are not characteristic for colitis-specific microbiota compositions at all, or that (2) quantitative measurements are limited in statistical support. The first reason could be explained by characteristics of compositional data. As they are based on the fractions of each samples' composition, they often overestimate relative abundance alterations. Changes in abundance of one taxon alter the relative abundances of all other community members as well although their actual numbers might not have changed (Gloor et al., 2016). The second reason could be attributed to differences in sample sizes. For the compositional approach over 500 samples were analyzed whereas only a subset of 120 samples was used for quantifications of the 16S rRNA gene, suggesting also lower sample sizes to diminish statistical significances. Although all of these six taxa have been described in the context of intestinal inflammation before and more or less consistent between studies, their potential involvement in colitis development in *IL-10*$^{-/-}$ mice cannot be clarified based on the present data.

Contrarily, evidence for colitis-specific abundance alterations of the remaining four bacteria was supported by the quantitative measurements, associating a decrease of *Lactobacillus* and two genera belonging to the family *Muribaculaceae*, as well as an increase of the family *Erysipelotrichaceae* with colitis.

Members of the *Lactobacillus* genus have been immensely studied and used as probiotic for treatment of IBD and gastrointestinal infections (Macfarlane GT, 2002; Madsen, 2001) due to reported immune modulating properties, e.g. regulating cytokine production in dendritic and intestinal mucosa cells (Christensen et al., 2002; Peña et al., 2004; Perdigó et al., 2002). Furthermore, supplementation with a strain of *L. reuteri* attenuated inflammation in neonatal IL-10 deficient 129 Sv/Ev mice. These mice are characterized by high colitis susceptibility and are simultaneously reported to contain decreased relative abundances of *Lactobacillus* two weeks after birth, which the authors interpreted as a potential microbiota indicator of increased colitis susceptibility. (Madsen et al., 1999)

Members of the family *Muribaculaceae* have been reported as potential mucus degraders due to their capability to hydrolyze mucus-derived carbohydrates (Lagkouvardos et al., 2019). They also have been suggested as ecological gatekeepers for healthy microbial gut

communities with resilience to *Clostridioides difficile* infection by limiting nutrient availability and preventing colonization (Pereira et al., 2020). Furthermore, culture-independent studies linked the abundance of *Muribaculaceae* genera to short-chain fatty acid (SCFA) production in mice promoting gut barrier functions by increasing availability of energy substrates for intestinal epithelial cells (Ormerod et al., 2016; Smith et al., 2019). Therefore, decreased levels of these beneficial commensals in the state of colitis could reduce maintenance of gut homeostasis in the host leading to a loss of stability and function favoring inflammation. The family *Erysipelotrichaceae*, which was increased during colitis in IL-10 deficient mice, has been found to be enriched during acute inflammation in DSS-treated wild type C57BL/6 mice (Schwab et al., 2014). Immunogenic properties of this family have been linked to Human Immunodeficiency Virus (HIV) and IBD, as one *Erysipelotrichaceae* species is identified to be highly coated with intestinal immunoglobulin A (IgA) relative to other gut microbes (Palm et al., 2014). The transfer of highly IgA coated bacterial consortia containing *Erysipelotrichaceae* spp. to germ-free mice further triggered more severe DSS-induced colitis compared to IgA-negative bacterial communities as shown by the authors. Schaubeck et al. reported *Erysipelotrichaceae* enrichment in mice with TNF-driven CD-like transmural inflammation (Schaubeck et al., 2015), supporting the findings of a positive correlation between Erysipelotrichi relative abundance and TNF levels in chronic HIV patients receiving suppressive antiretroviral therapy compared to healthy controls (Dinh et al., 2015). Nevertheless, results of human and murine studies have also been contrary especially in the IBD context, as a decrease in *Erysipelotrichaceae* was found in patients with recurring CD (Dey et al., 2013) and new-onset CD (Gevers et al., 2014), arguing for inter-host variation potentially resulting from differences in microbial gut communities between mice and humans and/or sensing of bacterial ligands in innate immune responses (Nguyen et al., 2015; Zschaler, J., Schlorke, D., and Arnhold, 2014). Furthermore, the current lack of well-characterized members of the *Erysipelotrichaceae* family limits transferability from *in vivo* studies to humans, as differences between species and even strains within a taxonomic family could results in varying metabolic or immunogenic properties (Kaakoush, 2015; Lloyd-Price et al., 2017; Podlesny et al., 2022; Schirmer et al., 2019).

As colitis-specific microbiota alterations, such as of *Erysipelotrichaceae* and one genus of *Muribaculaceae*, were also observed before the onset of colitis phenotypes, microbiota alterations specific to colitis in IL-10 deficient mice could be indicative of hidden inflammations even before clinical manifestation of colitis.

Litter and co-housing were additionally identified by the GLMM to explain variation within the taxonomic composition. Numbers of 16S rRNA gene copies per mg stool attributed to *Parabacteroides* revealed its absence in cages without wild-type mice. Also, comparisons on the day of cage allocation identified only wild-type mice to contain *Parabacteroides*, suggesting that alterations of *Parabacteroides* abundance might rather be related to transmission between genotypes via co-habitation than to colitis development. This is further supported by comparisons between knockout mice that developed colitis and those that did not, showing that also healthy mice acquired *Parabacteroides* throughout the observation and that *IL-10* knockout mice absent of this genus still developed intestinal inflammation. *Parabacteroides*, a gut commensal found in healthy wild-type mice (J. Wang et al., 2019) and humans (Falony et al., 2016), has been identified in the context of intra-abdominal infections first, but specific species such as *P. goldsteinii* have also been associated with the improvement of insulin resistance, obesity and chronic obstructive pulmonary disease (Chang et al., 2015; Wu et al., 2019). The *P. goldsteinii* strain MTS01 has also been recently proposed as a probiotic to mitigate *Helicobacter pylori*-induced infections (Lai et al., 2022). Further, Cuffaro et al. isolated *P. distasonis* from human neonatal and adult intestines, which showed the potential to restore gut barrier functions, reinstate the epithelial barrier, exert anti-inflammatory potential *in vitro*, and to protect mice from chemically induced colitis by Trinitrobenzene sulfonic acid (Cuffaro et al., 2020). They determined the anti-inflammatory potential to be mediated by IL-10 induction, which could be a reason why no beneficial effect on the IL-10 deficient mice was observed in the present study. Taken together, positive associations of *Parabacteroides* relative abundance and colitis are more likely to be compositional artifacts than to contribute actively to colitis development. This genus is not a commensal in the present *IL-10* knockout mouse model and was only acquired by them via the co-housing with wild-type mice, which did not further seem to affect the colitis development in *IL-10*$^{-/-}$ mice after acquisition.

Mice that showed colitis-related symptoms were characterized by either very high (>1% relative abundance, >10^7 attributed gene copies per mg stool) or no abundance of the genus *Akkermansia.* This clustering was also litter-specific revealing high abundances in mice with severe inflammation, which belonged only to litter 2 and 4, and absence of *Akkermansia* in mice of the two litters with only mild and temporary colitis-related symptoms (litter 1 and 3). Assessing *Akkermansia* abundance over time and including asymptomatic samples, further revealed only samples from litter 2 and 4 mice to contain high relative and absolute abundances despite their co-housing with the colitis-resistant *IL-10*$^{-/-}$ litter 3. Additionally, only the presence of *Akkermansia* shortly before (up to four days) the first symptomatic

observation, was determined to be a good predictor of early mice withdrawal due to severe colitis. These observations point towards *Akkermansia* being a potential subclinical marker of early colitis rather than triggering its development actively.

Akkermansia, a gut commensal represented by only one known species up to date *A. muciniphila*, is controversially discussed in the literature (Cirstea et al., 2018; de Vos, 2017; Derrien et al., 2017). Due to its ability to degrade mucin, it has been characterized as resident of the mucus layer, where it is thought to exhibit beneficial effects, e.g. via the production of propionate and acetate (Derrien et al., 2010). Animal and human studies have associated reduced relative abundance of *Akkermansia* with metabolic conditions like obesity (H. Zhang et al., 2009), dyslipidemia (Brahe et al., 2015) and type 2 diabetes (X. Zhang et al., 2013). But high *A. muciniphila* fecal relative abundances have also been reported in neurological diseases including Parkinson's disease (Heintz-Buschart et al., 2018; Hill-Burns et al., 2017), multiple sclerosis (Berer et al., 2017; Cekanaviciute et al., 2017) and Alzheimer's disease (Vogt et al., 2017). Not only between but also within conditions the same controversial role is discussed, e.g. in IBD. Most studies report reduced levels of *A. muciniphila* during colitis (Lopez-Siles et al., 2018; Papa et al., 2012; Png et al., 2010; T. Zhang et al., 2021) contradicting findings of others (Danilova et al., 2019; Ganesh et al., 2013). These contrary findings are likely to be attributed to strain variability, the pathogenic context and *in vivo* models used as explained in the following.

For example, reported anti-inflammatory properties of *A. muciniphila* are identified to only be exhibited by certain strains, such as human but not murine isolates (Guo et al., 2017; Zhai et al., 2019). Contrarily, administration of *Akkermansia* to SPF IL-10 deficient mice was enough to induce spontaneous colitis development (Seregin et al., 2017). The authors demonstrated this induction to be mediated by lipopolysaccharide (LPS), a component of the outer cell wall of gram-negative bacteria such as *Akkermansia*, which is sensed by Toll-like receptor (TLR) 4 and MyD88 signaling during spontaneous colitis.

Furthermore, *Akkermansia* might only exhibit detrimental effects depending on the pathogenic context as shown in co-occurrence with the pathogen *Salmonella typhimurium* in mice colonized with a human mimicking microbiota (Ganesh et al., 2013). Under this condition, *Akkermansia* is thought to exhibit harmful effects passively, e.g. by reducing the mucus barrier and elevating accessibility of pathogens to the intestinal mucosa, which facilitates their translocation (Desai et al., 2016; Seregin et al., 2017; A. W. Walker, Sanderson, et al., 2011; T. Zhang et al., 2021). Also, only mild colitis was found in *A. muciniphila* monocolonized germ-free IL-10 deficient mice compared to severe inflammation in SPF mice gavaged with *A. muciniphila* (Seregin et al., 2017), which

supports the context-dependent colitogenicity and potentially synergistic effect of *Akkermansia* and the microbial community.

As *Akkermansia* is a resident of the mucosal layer, a high fecal abundance might indicate a higher excretion of mucus, which is characteristic for IBD patients and a sign of intestinal inflammation (Hendrickson et al., 2002), stressing the role of high *A. municiphila* relative abundance in feces as an early indicator of colitis development. Taking everything into consideration, the importance of *Akkermansia* entering the microbial community in susceptible hosts and its subsequent growth and high excretion is highlighted to be a good indicator of subclinical early colitis in the *IL-10* knockout mouse model.

In summary, the microbial community characterization of IL-10 deficient mice throughout their maturation and the development of spontaneous colitis revealed that (1) intestinal inflammation is characterized by global microbial alterations, (2) susceptibility to spontaneous colitis is influenced by the mouse litter and (3) not attenuated by co-housing with wild-type mice. The increased litter-dependent susceptibility was further associated with the presence of *Akkermansia* in mature mice, which was revealed to be a good predictor for severe colitis development shortly before symptoms manifest.

C. Individual microbial traits are important to consider for clinical outcomes

The herein examined taxonomic microbiota characterizations during LCHF dietary interventions, as example of an extrinsic, modifiable microbiota-relevant factor and genetic deficiency of IL-10 in mice exemplary of an intrinsic, non-modifiable microbiota-relevant factor both identified individual traits as being important for intervention outcomes. Personalized medicine approaches, in which treatments and medical decisions are tailored based on an individual's genetic profile or other biomarkers, are emerging in the last two decades since the human genome sequence has been steadily decoded (Yamamoto et al., 2022). In cancer and rare diseases therapy, precision medicine approaches are most advanced and hold the promise to improve treatment outcomes, reduce adverse events and decrease healthcare costs (Kashyap et al., 2017). With improving resolution and lower costs of sequencing techniques, the microbiota has become of increasing interest for personalized medicine approaches. As it not only contributes to inter-individual variability in health as well as in disease, it is also a promising modifiable target for therapies, e.g. by dietary modulation. For example, glycemic responses to different bread types can be predicted solely based on the microbiome (Korem et al., 2017; Zeevi et al., 2015), which highlights the ability of microbiome-based stratification to improve host physiology in personalized dietary interventions. Similarly, the LCHF-diet induced increased serum LDL-C concentration reported herein, with potentially associated elevated cardiovascular risks, was limited only to the subgroup of lean cholesterol high-converters. Phenotyping and stratification of individuals before any treatment, e.g. into sterol conversion types, could potentially predict treatment responses and prevent adverse side effects. For example, by excluding lean high-converters from LCHF diet recommendations due to an elevation of serum LDL-C levels.

Whereas personalized approaches are gaining importance in human studies, they have been less often considered in *in vivo* studies due to the identical genetic background of animals. But even among genetically identical mice, microbial variations depending on the mouse litter were found in the IL-10 deficient mouse model. Furthermore, these litter-depending microbiota communities improved the prediction of high and low colitis susceptibility, which could potentially affect the outcome of studies and conclusions drawn when not considered. The alleged advantage of genetic similarity could in fact turn out to be a pitfall when litters with low and high susceptibility are not evenly distributed among cages or intervention arms in a trial.

Taking litter effects into account revealed *Akkermansia* to be a potential biomarker for impending severe intestinal inflammation. Solely the presence (relative abundance > 0) of

Akkermansia up to four days before colitis manifested predicted the disease in mice, pointing towards *Akkermansia* as a promising microbial biomarker to detect subclinical early colitis. Identifying this state of early inflammation before symptoms manifest is of main concern in IBD diagnosis and management, and could be used as an indicator of treatment success (Dragoni et al., 2021). Although patients are already treated for IBD pharmacologically, e.g. by anti-inflammatory and immunosuppressing drugs, subclinical mucosal inflammations can persist, which destroy the intestinal epithelium and increase the risk of colorectal cancer (Choi et al., 2017). Up to date, these mucosal inflammations can only be diagnosed by gastrointestinal endoscopy, which is cost-intensive, very unpleasant and inconvenient for IBD patients, and decreases their compliance to regular re-assessment (Dragoni et al., 2021). The identification of this state of early inflammation by cost and time-effective, non-invasive, *Akkermansia*-specific PCR from fecal samples could improve the quality of life of patients and the clinical management of IBD.

Limitations of the studies are in line with common pitfalls in microbiome research, e.g. the use of compositional data and the detection problem of sparse microbiota members, which inflate sequence count data with large numbers of zeros.

In the human intervention, results are based on compositional data only as too little of the samples were left for additional quantitative measurements. This limits drawing conclusions about the totality of the microbial composition and taxon-specific alterations independent of overall community changes (Gloor et al., 2016), as discussed earlier.

Large numbers of zeros in sequence counts are an important and common feature of microbiome data (Kaul et al., 2017). They either result from not all taxa being present in all samples or from taxa being present below the detection limit of current sequencing techniques. Statistically and conceptually this is difficult to handle, as many analyses depend on logarithmic transformation, which results in undefined values for zero counts and therefore exclusion from the data set. Conceptually, it can be of biological relevance if certain bacteria are completely absent from a sample or only less abundant. To overcome these issues, pseudocounts, small numbers usually ≤ 1, are added to count data (Kaul et al., 2017). As the addition can influence analysis, e.g. by skewing the distribution of data, they need to be applied with caution and interpreted in the context of the data (Costea et al., 2014). For example, the cage effect and transmission by co-housing from wild type to *IL-10* knockout mice of *Parabacteroides* could not have been detected if samples with zero counts, indicating absence, were excluded in the GLMM without the addition of pseudocounts. On the other hand, pseudocount addition did not influence correlations between fecal sterols and relative taxon abundances in the human cohorts, which were

based on alterations of the abundance, stressing context-dependent interpretation of pseudocount added data and their relevance to detect community alterations based on presence/absence.

The herein reported results characterize taxonomic microbial alterations, but not functional properties, and are limited in resolution to the genus or species level as they are based on 16S rRNA gene sequencing (Fricker et al., 2019). To gain further insights into functional profiles of the microbiome and strain variability of taxa of interest, such as *E. coprostanoligenes* and *A. muciniphila,* metagenomic whole-genome shotgun sequencing could be applied. This would further enable tracking bacterial transmission between individuals, e.g. in co-housed mice, or of changes within individuals over time, e.g. before and during dietary intervention, and identification of metabolic differences between strains of the same species (Vicedomini et al., 2021; Yan et al., 2020). The latter could be of special interest to further uncover *Akkermansia*'s role in colitis development, especially as strain-specific functional properties of its strains have been described before (Guo et al., 2017; Zhai et al., 2019). It still remains inconclusive if all litter 2 and 4 mice contained the same *Akkermansia* strain at the beginning or if different strains persisted within different litters over time. Additionally, it could help to elucidate if the *Akkermansia* strain increasing in feces during colitis, possessed inflammatory properties itself or if its enrichment is rather a marker for higher mucus excretion due to intestinal inflammation.

To further improve causation between the presence of certain bacteria and the observed phenotype, individual bacterial strains could be isolated and cultured. Their cultures could then be used for *in vitro* and *in vivo* studies to induce or prevent the associated clinical response, e.g. colitis, changes in sterol conversion ratio or altered blood lipid levels. These experiments could be used to evaluate intervention strategies with the goal to manipulate the microbiota in the future.

Taken together, fecal microbial community characterization in response to LCHF-diets in humans and throughout the development of intestinal inflammation in a colitis mouse model, highlight the necessity for personalized, more individual-dependent research even in *in vivo* models to account for interindividual differences and improve treatment responses.

V. Conclusion

In the presented work, the taxonomic characterizations of gut microbiomes under the impact of two major community shaping factors, i.e. dietary modulation and genetic deficiency, revealed classifications of individuals based on their microbiota composition as an important trait to predict clinical outcomes. In two independent studies, comprising human as well as mouse cohorts, alterations of the gut microbiota were linked to specific metabolic and inflammatory consequences, which are associated with phenotypes of increased health risk.

First, in the context of dietary interventions as an exemplary extrinsic, modifiable microbiota-relevant factor, microbial and metabolic profiling of two human cohorts with and without obesity stratified individuals independently of their metabolic background into high or low-cholesterol conversion types, which was largely dependent on the relative abundance of the bacterial species *Eubacterium coprostanoligenes*. Importantly, lean high-converters responded to low-carbohydrate high-fat dietary (LCHF) intervention with increased low-density lipoprotein cholesterol (LDL-C) serum concentrations, revealing the cholesterol high-converter type as a potential predictive biomarker for an increased atherosclerotic cardiovascular disease risk in lean adults in response to LCHF diets. This suggested identification of a potential high risk subpopulation for which LCHF diets could be harmful, is of relevance as LCHF diets are popular even among metabolically healthy, lean individuals.

The analysis of a genetic deficiency serving as an exemplary intrinsic, non-modifiable microbiota-relevant factor in an *Interleukin-10* knockout mouse model for inflammatory bowel disease (IBD), identified potential microbial markers for very early colitis susceptibility and subclinical colitis onset. Whereas colitis susceptibility was dependent on global microbiome characteristics acquired early on in life by each mouse litter, colitis onset shortly before clinical symptoms manifested was strongly predicted by the presence of *Akkermansia*, which suggests it to be a subclinical colitis marker. Together with other global microbiome characteristics which were linked to IBD such as changes in stool consistency, *Akkermansia* presence could help to identify subclinical colitis onset and improve IBD management, which is a current burden in IBD therapy.

In summary, these analyses reveal microbiome-dependent phenotyping and classification of individuals as an important trait to improve the prediction of clinical outcomes and guide personalized treatment approaches.

References

Ang, Q. Y., Alexander, M., Newman, J. C., Tian, Y., Cai, J., Upadhyay, V., Turnbaugh, J. A., Verdin, E., Hall, K. D., Leibel, R. L., Ravussin, E., Rosenbaum, M., Patterson, A. D., & Turnbaugh, P. J. (2020). Ketogenic Diets Alter the Gut Microbiome Resulting in Decreased Intestinal Th17 Cells. *Cell*, *181*(6), 1263-1275.e16. https://doi.org/10.1016/J.CELL.2020.04.027

Asnicar, F., Chan, A. T., Huttenhower, C., Spector, T. D., Segata, N., Berry, S. E., Valdes, A. M., Nguyen, L. H., Piccinno, G., Drew, D. A., Leeming, E., Gibson, R., Le Roy, C., Al Khatib, H., Francis, L., Mazidi, M., Mompeo, O., Valles-Colomer, M., Tett, A., … Ordovás, J. M. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nature Medicine*, *12*(11), 13. https://doi.org/10.1038/s41591-020-01183-8

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, I. I., Umesaki, Y., Itoh, K., & Honda, K. (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. *Science*, *331*(6015), 337–341. https://doi.org/10.1126/SCIENCE.1198469

Bach, J.-F. (2002). The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *Https://Doi.Org/10.1056/NEJMra020100*, *347*(12), 911–920. https://doi.org/10.1056/NEJMRA020100

Barquera, S., Pedroza-Tob, A., Medina, C., Ia Hern Andez-Barrera, L., Bibbins-Domingo, K., Lozano, R., & Moran, A. E. (2015). Global Overview of the Epidemiology of Atherosclerotic Cardiovascular Disease. *Arch Med Res*, *46*(5), 328–338. https://doi.org/10.1016/j.arcmed.2015.06.006

Baumgart, D. C., & Sandborn, W. J. (2007). Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet*, *369*(9573), 1641–1657. https://doi.org/10.1016/S0140-6736(07)60751-X

Bays, H. E., Neff, D., Tomassini, J. E., & Tershakovec, A. M. (2008). Ezetimibe: cholesterol lowering and beyond. *Expert Review of Cardiovascular Therapy*, *6*(4), 447–470. https://doi.org/10.1586/14779072.6.4.447

Bear, T., Dalziel, J., Coad, J., Roy, N., Butts, C., & Gopal, P. (2021). The Microbiome-Gut-Brain Axis and Resilience to Developing Anxiety or Depression under Stress. *Microorganisms*, *9*(4). https://doi.org/10.3390/MICROORGANISMS9040723

Beckwith, J., Cong, Y., Sundberg, J. P., Elson, C. O., & Leiter, E. H. (2005). Cdcs1, a Major Colitogenic Locus in Mice, Regulates Innate and Adaptive Immune Response to Enteric Bacterial Antigens. *Gastroenterology*, *129*(5), 1473–1484. https://doi.org/10.1053/J.GASTRO.2005.07.057

Belizário, J. E., & Faintuch, J. (2018). Microbiome and Gut Dysbiosis. *Experientia Supplementum (2012)*, *109*, 459–476. https://doi.org/10.1007/978-3-319-74932-7_13/COVER

Benno, P., Midtvedt, K., Alam, M., Collinder, E., Norin, E., & Midtvedt, T. (2009). *Examination of intestinal conversion of cholesterol to coprostanol in 633 healthy subjects reveals an age-and sex-dependent pattern*. https://doi.org/10.1080/08910600500519854

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., & Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(44), 18933–18938. https://doi.org/10.1073/PNAS.1007028107/SUPPL_FILE/PNAS.201007028SI.PDF

Berer, K., Gerdes, L. A., Cekanaviciute, E., Jia, X., Xiao, L., Xia, Z., Liu, C., Klotz, L., Stauffer, U., Baranzini, S. E., Kümpfel, T., Hohlfeld, R., Krishnamoorthy, G., & Wekerle, H. (2017). Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10719–10724. https://doi.org/10.1073/PNAS.1711233114/SUPPL_FILE/PNAS.201711233SI.PDF

Berg, D. J., Davidson, N., Kühn, R., Müller, W., Menon, S., Holland, G., Thompson-Snipes, L., Leach, M. W., & Rennick, D. (1996). Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4+ Th1-like responses. *Journal of Clinical Investigation*, *98*(4), 1010–1020. https://doi.org/10.1172/JCI118861

Bevins, C. L., & Salzman, N. H. (2011). Paneth cells, antimicrobial peptides and maintenance of

intestinal homeostasis. *Nature Reviews Microbiology 2011 9:5*, *9*(5), 356–368. https://doi.org/10.1038/nrmicro2546

Bilotta, S., Bhargavi Paruchuru, L., Feilhauer, K., Köninger, J., & Lorentz, A. (2021). Resveratrol Is a Natural Inhibitor of Human Intestinal Mast Cell Activation and Phosphorylation of Mitochondrial ERK1/2 and STAT3. *International Journal of Molecular Sciences Article*. https://doi.org/10.3390/ijms22147640

Bjarnason, I., Peters, T. J., & Veall, N. (1983). A persistent defect in intestinal permeability in coeliac disease demonstrated by a 51Cr-labelled EDTA absorption test. *The Lancet*, *321*(8320), 323–325. https://doi.org/10.1016/S0140-6736(83)91628-8

Blaser, M. J., & Dominguez-Bello, M. G. (2016). The Human Microbiome before Birth. *Cell Host and Microbe*, *20*(5), 558–560. https://doi.org/10.1016/j.chom.2016.10.014

Bleich, A., Mä, M., Most, C., Leiter, E. H., Liebler-Tenorio, E., Elson, C. O., Hedrich, H. J., Schlegelberger, B., & Sundberg, J. P. (2004). Refined histopathologic scoring system improves power to detect colitis QTL in mice. *Mammalian Genome*, *15*, 865–871. https://doi.org/10.1007/s00335-004-2392-2

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., Spector, T. D., Keinan, A., Ley, R. E., Gevers, D., & Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, *16*(1). https://doi.org/10.1186/S13059-015-0759-1

Bodden, C., Pang, T. Y., Feng, Y., Mridha, F., Kong, G., Li, S., Watt, M. J., Reichelt, A. C., & Hannan, A. J. (2022). Intergenerational effects of a paternal Western diet during adolescence on offspring gut microbiota, stress reactivity, and social behavior. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *36*(1). https://doi.org/10.1096/FJ.202100920RR

Bokulich, N. A., Dillon, M. R., Bolyen, E., Kaehler, B. D., Huttley, G. A., & Caporaso, J. G. (2018). *q2-sample-classifier: machine-learning tools for microbiome classification and regression*. *3*(30), 934. https://doi.org/10.21105/joss.00934

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/J.TREE.2008.10.008

Bolyen, E., Ram Rideout, J., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Titus Brown, C., Callahan, B. J., Caraballo-Rodramp, M., Chase, J., … Gregory Caporaso, J. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vich Vila, A., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S. P., Zhernakova, D. V, Jankipersadsing, S. A., Jaeger, M., Oosting, M., Cenit, M. C., Masclee, A. A. M., Swertz, M. A., Li, Y., Kumar, V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature GeNetics VOLUME*, *48*. https://doi.org/10.1038/ng.3663

Brahe, L. K., Le Chatelier, E., Prifti, E., Pons, N., Kennedy, S., Hansen, T., Pedersen, O., Astrup, A., Ehrlich, S. D., & Larsen, L. H. (2015). Specific gut microbiota features and metabolic markers in postmenopausal women with obesity. *Nutrition & Diabetes*, *5*(6). https://doi.org/10.1038/NUTD.2015.9

Bristol, I. J., Farmer, M. A., Cong, Y., Zheng, X. X., Strom, T. B., Elson, C. O., Sundberg, J. P., & Leiter, E. H. (2000). Heritable Susceptibility for Colitis in Mice Induced by IL-10 Deficiency. *Inflammatory Bowel Diseases*, *6*(4), 290–302. https://doi.org/10.1097/00054725-200011000-00006

Bueno, N. B., De Melo, I. S. V., De Oliveira, S. L., & Da Rocha Ataide, T. (2013). Very-low-carbohydrate ketogenic diet v. low-fat diet for long-term weight loss: A meta-analysis of Randomised controlled trials. *British Journal of Nutrition*, *110*(7), 1178–1187. https://doi.org/10.1017/S0007114513000548

Bull, I. D., Lockheart, M. J., Elhmmali, M. M., Roberts, D. J., & Evershed, R. P. (2002). The origin of faeces by means of biomarker detection. *Environment International*, *27*(8), 647–654. https://doi.org/10.1016/S0160-4120(01)00124-6

Burén, J., Ericsson, M., Damasceno, N. R. T., & Sjödin, A. (2021). A Ketogenic Low-Carbohydrate High-Fat Diet Increases LDL Cholesterol in Healthy, Young, Normal-Weight Women: A

Randomized Controlled Feeding Trial. *Nutrients*, *13*(3), 1–12. https://doi.org/10.3390/NU13030814

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods 2016 13:7*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camilleri, M., Madsen, K., Spiller, R., Van Meerveld, B. G., & Verne, G. N. (2012). Intestinal barrier function in health and gastrointestinal disease. *Neurogastroenterology and Motility*, *24*(6), 503–512. https://doi.org/10.1111/J.1365-2982.2012.01921.X

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pẽa, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods 2010 7:5*, *7*(5), 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4516–4522. https://doi.org/10.1073/PNAS.1000080107/SUPPL_FILE/PNAS.201000080SI.PDF

Carabotti, M., Scirocco, A., Antonietta Maselli, M., & Severi, C. (2015). *The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems*. *28*, 203–209. www.annalsgastro.gr

Caugant, D. A., Levin, B. R., & Selander, R. K. (1984). Distribution of multilocus genotypes of Escherichia coli within and between host families. *Epidemiology & Infection*, *92*(3), 377–384. https://doi.org/10.1017/S0022172400064597

Cekanaviciute, E., Yoo, B. B., Runia, T. F., Debelius, J. W., Singh, S., Nelson, C. A., Kanner, R., Bencosme, Y., Lee, Y. K., Hauser, S. L., Crabtree-Hartman, E., Sand, I. K., Gacias, M., Zhu, Y., Casaccia, P., Cree, B. A. C., Knight, R., Mazmanian, S. K., & Baranzini, S. E. (2017). Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10713–10718. https://doi.org/10.1073/PNAS.1711235114/SUPPL_FILE/PNAS.1711235114.SAPP.PDF

Chang, C.-J., Lin, C.-S., Lu, C.-C., Martel, J., Ko, Y.-F., Ojcius, D. M., Tseng, S.-F., Wu, T.-R., Chen, Y.-Y. M., Young, J. D., & Lai, H.-C. (2015). Ganoderma lucidum reduces obesity in mice by modulating the composition of the gut microbiota. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms8489

Cho, I., & Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. *Nature Reviews Genetics 2012 13:4*, *13*(4), 260–270. https://doi.org/10.1038/nrg3182

Choi, C. H. R., Bakir, I. Al, Hart, A. L., & Graham, T. A. (2017). Clonal evolution of colorectal cancer in IBD. *Nature Reviews Gastroenterology & Hepatology 2017 14:4*, *14*(4), 218–229. https://doi.org/10.1038/nrgastro.2017.1

Christensen, H. R., Frøkiaer, H., & Pestka, J. J. (2002). Surface Markers in Murine Dendritic Cells Expression of Cytokines and Maturation Lactobacilli Differentially Modulate. *J Immunol References*, *168*, 171–178. https://doi.org/10.4049/jimmunol.168.1.171

Cirstea, M., Radisavljevic, N., & Finlay, B. B. (2018). Good Bug, Bad Bug: Breaking through Microbial Stereotypes. *Cell Host and Microbe*, *23*(1), 10–13. https://doi.org/10.1016/J.CHOM.2017.12.008

Costea, P. I., Zeller, G., Sunagawa, S., & Bork, P. (2014). *A fair comparison*. https://doi.org/10.1038/nmeth.2897

Cuffaro, B., Assohoun, A. L. W., Boutillier, D., Súkeníková, L., Desramaut, J., Boudebbouze, S., Salomé-Desnoulez, S., Hrdý, J., Waligora-Dupriet, A. J., Maguin, E., & Grangette, C. (2020). In Vitro Characterization of Gut Microbiota-Derived Commensal Strains: Selection of Parabacteroides distasonis Strains Alleviating TNBS-Induced Colitis in Mice. *Cells*, *9*(9). https://doi.org/10.3390/CELLS9092104/CELLS_09_02104_PDF.PDF

Danielsson, H., Gustafsson, B. (1959). On serum-cholesterol levels and neutral fecal sterols in germ-free rats. *Arch. Biochem. Biophys.*, *83*, 482–485. https://doi.org/10.1016/0003-9861(59)90056-6

Danilova, N. A., Александровна, Д. Н., Abdulkhakov, S. R., Рустамович, А. С., Grigoryeva, T. V, Владимировна, Г. Т., Markelova, M. I., Ивановна, М. М., Vasilyev, I. Y., Юрьевич, В. И.,

Boulygina, E. A., Александровна, Б. Е., Ardatskaya, M. D., Дмитриевна, А. М., Pavlenko, A. V, Владимирович, П. А., Tyakht, A. V, Викторович, Т. А., Odintsova, A. K., … Аббасович, А. Р. (2019). Markers of dysbiosis in patients with ulcerative colitis and Crohn's disease. *Terapevticheskii Arkhiv*, *91*(4), 13–20. https://doi.org/10.26442/00403660.2019.04.000211

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V, Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*, 559. https://doi.org/10.1038/nature12820

de Vos, W. M. (2017). Microbe Profile: Akkermansia muciniphila: a conserved intestinal symbiont that acts as the gatekeeper of our mucosa MICROBE PROFILE. *Microbiology*, *163*, 646–648. https://doi.org/10.1099/mic.0.000444

Dean, C. B., Nielsen, J. D., & Nielsen, J. D. (2007). Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal*, *13*, 497–512. https://doi.org/10.1007/s10985-007-9065-x

Derrien, M., Belzer, C., & de Vos, W. M. (2017). Akkermansia muciniphila and its role in regulating host functions. *Microbial Pathogenesis*, *106*, 171–181. https://doi.org/10.1016/J.MICPATH.2016.02.005

Derrien, M., van Passel, M. W. J., van de Bovenkamp, J. H. B., Schipper, R. G., de Vos, W. M., & Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes*, *1*(4), 254–268. https://doi.org/10.4161/GMIC.1.4.12778/SUPPL_FILE/KGMI_A_10912778_SM0001.ZIP

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, *167*(5), 1339-1353.e21. https://doi.org/10.1016/J.CELL.2016.10.043

Descamps, O., Tomassini, J. E., Lin, J., Polis, A. B., Shah, A., Brudi, P., Hanson, M. E., & Tershakovec, A. M. (2015). Variability of the LDL-C lowering response to ezetimibe and ezetimibe + statin therapy in hypercholesterolemic patients. *Atherosclerosis*, *240*(2), 482–489. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2015.03.004

Dewei, R., Ling, L., Schwabacher, A. W., Young, J. W., & Beitz, D. C. (1996). Mechanism of cholesterol reduction to coprostanol by Eubacterium coprostanoligenes ATCC 51222. *Steroids*, *61*(1), 33–40. https://doi.org/10.1016/0039-128X(95)00173-N

Dey, N., Soergel, D. A. W., Repo, S., & Brenner, S. E. (2013). Association of gut microbiota with post-operative clinical course in Crohn's disease. *BMC Gastroenterology*, *13*(1), 1–11. https://doi.org/10.1186/1471-230X-13-131/FIGURES/3

Dinh, D. M., Volpe, G. E., Duffalo, C., Bhalchandra, S., Tai, A. K., Kane, A. V, Wanke, C. A., & Ward, H. D. (2015). Intestinal Microbiota, Microbial Translocation, and Systemic Inflammation in Chronic HIV Infection. *J Infect Dis .*, *211*(1), 19–27. https://doi.org/10.1093/infdis/jiu409

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(26), 11971–11975. https://doi.org/10.1073/PNAS.1002601107

Dominguez-Bello, M. G., Godoy-Vitorino, F., Knight, R., & Blaser, M. J. (2019). Role of the microbiome in human development. *Gut*, *68*(6), 1108–1114. https://doi.org/10.1136/GUTJNL-2018-317503

Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2015). Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology 2015 14:1*, *14*(1), 20–32. https://doi.org/10.1038/nrmicro3552

Doyle, A., McGarry, M. P., Lee, N. A., & Lee, J. J. (2012). The construction of transgenic and gene knockout/knockin mouse models of human disease. *Transgenic Research*, *21*(2), 327–349. https://doi.org/10.1007/S11248-011-9537-3/FIGURES/7

Dragoni, G., Innocenti, T., & Galli, A. (2021). Biomarkers of Inflammation in Inflammatory Bowel Disease: How Long before Abandoning Single-Marker Approaches? *Digestive Diseases*, *39*(3), 190–203. https://doi.org/10.1159/000511641

Drickamer, L. C. (1981). Selection for age of sexual maturation in mice and the consequences for population regulation. *Behavioral and Neural Biology*, *31*(1), 82–89. https://doi.org/10.1016/S0163-1047(81)91114-6

Ebino, K. Y., Amao, H., Suwa, T., Kuwabara, Y., Saito, T. R., & Takahashi, K. W. (1987). Coprophagy in the Germfree Mouse. *Exp. Anim*, *36*(1), 33–37.

Ecker, J., Scherer, M., Schmitz, G., & Liebisch, G. (2012). A rapid GC–MS method for quantification of positional and geometric isomers of fatty acid methyl esters. *Journal of Chromatography B*, *897*, 98–104. https://doi.org/10.1016/J.JCHROMB.2012.04.015

Edwards, A. M., & Cameron, E. Z. (2014). Forgotten fathers: Paternal influences on mammalian sex allocation. *Trends in Ecology and Evolution*, *29*(3), 158–164. https://doi.org/10.1016/J.TREE.2013.12.003

Edwards, C. J. (2008). Commensal Gut Bacteria and the Etiopathogenesis of Rheumatoid Arthritis. *The Journal of Rheumatology*, *35*(8).

Estaki, M., Pither, J., Baumeister, P., Little, J. P., Gill, S. K., Ghosh, S., Ahmadi-Vand, Z., Marsden, K. R., & Gibson, D. L. (2016). Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome*, *4*(1), 1–13. https://doi.org/10.1186/S40168-016-0189-7/FIGURES/7

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*(1), 1–7. https://doi.org/10.1186/2049-2618-2-6/FIGURES/3

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., Clemente, J. C., Knight, R., Heath, A. C., Leibel, R. L., Rosenbaum, M., & Gordon, J. I. (2013). The long-term stability of the human gut microbiota. *Science*, *341*(6141). https://doi.org/10.1126/SCIENCE.1237439/SUPPL_FILE/FAITH.SM.REVISED-2.PDF

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M. J., Valles-Colomer, M., Vandeputte, D., Tito, R. Y., Chaffron, S., Rymenans, L., Verspecht, C., Sutter, L. De, Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., … Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, *352*(6285), 560–564. https://doi.org/10.1126/SCIENCE.AAD3503/SUPPL_FILE/TABLE_S9.XLSX

Farmer, M. A., Sundberg, J. P., Bristol, I. J., Churchill, G. A., Li, R., Elson, C. O., & Leiter, E. H. (2001). A major quantitative trait locus on chromosome 3 controls colitis severity in IL-10-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 13820–13825. https://doi.org/10.1073/PNAS.241258698

Fasano, A., & Shea-Donohue, T. (2005). Mechanisms of Disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nature Clinical Practice Gastroenterology & Hepatology 2005 2:9*, *2*(9), 416–422. https://doi.org/10.1038/ncpgasthep0259

Feldman, D., Huggins, S., & Norwitz, N. G. (2022). Short-term hyper-caloric high-fat feeding on a ketogenic diet can lower low-density lipoprotein cholesterol: the cholesterol drop experiment. *Current Opinion in Endocrinology, Diabetes, and Obesity*, *29*(5), 434–439. https://doi.org/10.1097/MED.0000000000000762

Feng, T., Qin, H., Wang, L., Benveniste, E. N., Elson, C. O., & Cong, Y. (2011). Th17 Cells Induce Colitis and Promote Th1 Cell Responses through IL-17 Induction of Innate IL-12 and IL-23 Production. *The Journal of Immunology*, *186*, 6313–6318. https://doi.org/10.4049/jimmunol.1001454

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, *80*(1), 27. https://doi.org/10.2307/2336755

Fluhr, L., Mor, U., Kolodziejczyk, A. A., Dori-Bachash, M., Leshem, A., Itav, S., Cohen, Y., Suez, J., Zmora, N., Moresi, C., Molina, S., Ayalon, N., Valdés-Mas, R., Hornstein, S., Karbi, H., Kviatcovsky, D., Livne, A., Bukimer, A., Eliyahu-Miller, S., … Elinav, E. (2021). Gut microbiota modulates weight gain in mice after discontinued smoke exposure. *Nature*, *600*, 713. https://doi.org/10.1038/s41586-021-04194-8

Frank, D. N., St. Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(34), 13780. https://doi.org/10.1073/PNAS.0706625104

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., Vatanen, T., Hall, A. B., Mallick, H., McIver, L. J., Sauk, J. S., Wilson, R. G., Stevens, B. W., Scott, J. M., Pierce, K., Deik, A. A., Bullock, K., Imhann, F., Porter, J. A., … Xavier, R. J. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*, *4*(2), 293. https://doi.org/10.1038/S41564-018-0306-4

Freier, T. A., Beitz, D. C., Li, L., & Hartman, P. A. (1994). Characterization of Eubacterium coprostanoligenes sp. nov., a cholesterol-reducing anaerobe. *International Journal of Systematic Bacteriology*, *44*(1), 137–142. https://doi.org/10.1099/00207713-44-1-137

Fricker, A. M., Podlesny, D., & Fricke, W. F. (2019). What is new and relevant for sequencing-based microbiome research? A mini-review. *Journal of Advanced Research*, *19*, 105–112. https://doi.org/10.1016/J.JARE.2019.03.006

Fujimura, K. E., & Lynch, S. V. (2015). Microbiota in allergy and asthma and the emerging relationship with the gut microbiome. *Cell Host and Microbe*, *17*(5), 592–602. https://doi.org/10.1016/j.chom.2015.04.007

Fujiwara, S. (2018). Humanized mice: A brief overview on their diverse applications in biomedical research. *Journal of Cellular Physiology*, *233*(4), 2889–2901. https://doi.org/10.1002/JCP.26022

Furness, J. B., Kunze, W. A. A., & Clerc, N. (1999). Nutrient tasting and signaling mechanisms in the gut II. The intestine as a sensory organ: Neural, endocrine, and immune responses. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *277*(5 40-5). https://doi.org/10.1152/AJPGI.1999.277.5.G922/ASSET/IMAGES/LARGE/AGIJ31104002X.JPEG

Ganesh, B. P., Klopfleisch, R., Loh, G., & Blaut, M. (2013). Commensal Akkermansia muciniphila Exacerbates Gut Inflammation in Salmonella Typhimurium-Infected Gnotobiotic Mice. *PLOS ONE*, *8*(9), e74963. https://doi.org/10.1371/JOURNAL.PONE.0074963

Garber, A. J., Menzel, P. H., Boden, G., & Owen, O. E. (1974). Hepatic Ketogenesis and Gluconeogenesis in Humans. *The Journal of Clinical Investigation*, *54*(4), 981–989. https://doi.org/10.1172/JCI107839

Gensollen, T., Iyer, S. S., Kasper, D. L., & Blumberg, R. S. (2016). How colonization by microbiota in early life shapes the immune system. *Science*, *352*(6285), 539–544. https://doi.org/10.1126/SCIENCE.AAD9378

Gérard, P., Lepercq, P., Leclerc, M., Gavini, F., Raibaud, P., & Juste, C. (2007). Bacteroides sp. strain D8, the first cholesterol-reducing bacterium isolated from human feces. *Applied and Environmental Microbiology*, *73*(18), 5742–5749. https://doi.org/10.1128/AEM.02806-06

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C., González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., … Xavier, R. J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host & Microbe*, *15*(3), 382–392. https://doi.org/10.1016/J.CHOM.2014.02.005

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine 2018 24:4*, *24*(4), 392–400. https://doi.org/10.1038/nm.4517

Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., & Egozcue, J. J. (2016). It's all relative: analyzing microbiome data as compositions. *Annals of Epidemiology*, *26*(5), 322–329. https://doi.org/10.1016/J.ANNEPIDEM.2016.03.003

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., & Ley, R. E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe*, *19*(5), 731–743. https://doi.org/10.1016/J.CHOM.2016.04.017

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G., & Ley, R. E. (2014). Human genetics shape the gut microbiome. *Cell*, *159*(4), 789–799. https://doi.org/10.1016/J.CELL.2014.09.053/ATTACHMENT/9E974D9A-E7D0-4CBF-A4FF-77001365A165/MMC2.XLSX

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F., & Veldhuyzen Van Zanten, S. J. O. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *Journal of Clinical Microbiology*, *44*(11), 4136–4141. https://doi.org/10.1128/JCM.01004-06

Gordon, J. H., & Dubos, R. (1970). The anaerobic bacterial flora of the mouse cecum. *Journal of Experimental Medicine*, *132*(2), 251–260. https://doi.org/10.1084/JEM.132.2.251

Gotto, A. M. (1991). Cholesterol Intake and Serum Cholesterol Level. *New England Journal of Medicine*, *324*, 912–913.

Griffin, B. A., Mensink, R. P., & Lovegrove, J. A. (2021). Does variation in serum LDL-cholesterol

response to dietary fatty acids help explain the controversy over fat quality and cardiovascular disease risk? *Atherosclerosis*, *328*, 108–113. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2021.03.024

Guo, X., Li, S., Zhang, J., Wu, F., Li, X., Wu, D., Zhang, M., Ou, Z., Jie, Z., Yan, Q., Li, P., Yi, J., & Peng, Y. (2017). Genome sequencing of 39 Akkermansia muciniphila isolates reveals its population structure, genomic and functional diverisity, and global distribution in mammalian gut microbiotas. *BMC Genomics*, *18*(1), 1–12. https://doi.org/10.1186/S12864-017-4195-3/FIGURES/6

Gutzeit, C., Magri, G., & Cerutti, A. (2014). Intestinal IgA production and its role in host-microbe interaction. *Immunological Reviews*, *260*(1), 76–85. https://doi.org/10.1111/IMR.12189

Hagenlocher, Y., Hösel, A., Bischoff, S. C., & Lorentz, A. (2016). Cinnamon extract reduces symptoms, inflammatory mediators and mast cell markers in murine IL-10-/- colitis. *Journal of Nutritional Biochemistry*, *30*, 85–92. https://doi.org/10.1016/J.JNUTBIO.2015.11.015

Halfvarson, J., Brislawn, C. J., Lamendella, R., Vázquez-Baeza, Y., Walters, W. A., Bramer, L. M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., McClure, E. E., Dunklebarger, M. F., Knight, R., & Jansson, J. K. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology*, *2*(February), 1–7. https://doi.org/10.1038/nmicrobiol.2017.4

Hall, A. B., Tolonen, A. C., & Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nature Reviews Genetics 2017 18:11*, *18*(11), 690–699. https://doi.org/10.1038/nrg.2017.63

Hansen, R., Russell, R. K., Reiff, C., Louis, P., McIntosh, F., Berry, S. H., MGBRhopadhya, I., Bisset, W. M., Barclay, A. R., Bishop, J., Flynn, D. M., McGrogan, P., Loganathan, S., Mahdi, G., Flint, H. J., El-Omar, E. M., & Hold, G. L. (2012). Microbiota of de-novo pediatric IBD: Increased faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *American Journal of Gastroenterology*, *107*(12), 1913–1922. https://doi.org/10.1038/AJG.2012.335

Hapfelmeier, S., Lawson, M. A. E., Slack, E., Kirundi, J. K., Stoel, M., Heikenwalder, M., Cahenzli, J., Velykoredko, Y., Balmer, M. L., Endt, K., Geuking, M. B., Curtiss, R., McCoy, K. D., & Macpherson, A. J. (2010). Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science*, *328*(5986), 1705–1709. https://doi.org/10.1126/SCIENCE.1188454/SUPPL_FILE/HAPFELMEIER.SOM.PDF

Heintz-Buschart, A., Pandey, U., Wicke, T., Sixel-Döring, F., Janzen, A., Sittig-Wiegand, E., Trenkwalder, C., Oertel, W. H., Mollenhauer, B., & Wilmes, P. (2018). The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Movement Disorders*, *33*(1), 88–98. https://doi.org/10.1002/MDS.27105

Hendrickson, B. A., Gokhale, R., & Cho, J. H. (2002). Clinical aspects and pathophysiology of inflammatory bowel disease. *Clinical Microbiology Reviews*, *15*(1), 79–94. https://doi.org/10.1128/CMR.15.1.79-94.2002

Hill-Burns, E. M., Debelius, J. W., Morton, J. T., Wissemann, W. T., Lewis, M. R., Wallen, Z. D., Peddada, S. D., Factor, S. A., Molho, E., Zabetian, C. P., Knight, R., & Payami, H. (2017). Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Movement Disorders*, *32*(5), 739–749. https://doi.org/10.1002/MDS.26942

Holladay, S. D., & Smialowicz, R. J. (2000). Development of the murine and human immune system: Differential effects of immunotoxicants depend on time of exposure. *Environmental Health Perspectives*, *108*(SUPPL. 3), 463–473. https://doi.org/10.1289/EHP.00108S3463

Hollander, D., Vadheim, M., Brettholz, E., Petersen, G., Delahunty, T., & Rotter, J. (1986). Increased intestinal permeability in patients with Crohn's disease and their relatives. A possible etiologic factor. *Annals of Internal Medicine*, *105*(6), 883. https://doi.org/10.7326/0003-4819-105-6-883

Horn, C., Laupsa-Borge, J., Andersen, A. I. O., Fasmer, K. E., Holmefjord, M.-A., Revheim, I., Johannessen, K. K., Næsheim, N. T., Storås, I., Leikanger, T., Amundsen, K., Skjerve, K. L., Dyer, L., Spjelkavik, C., Haldorsen, I., Lindseth, I., Dierkes, J., Mellgren, G., & Dankel, S. N. (2022). Diets differing in carbohydrate cellularity and amount similarly reduced visceral fat in people with obesity - a randomized controlled trial (CARBFUNC). *Clinical Nutrition*, *0*(0). https://doi.org/10.1016/J.CLNU.2022.08.028

Hufeldt, M. R., Nielsen, D. S., Vogensen, F. K., Midtvedt, T., & Hansen, A. K. (2010). Family relationship of female breeders reduce the systematic inter-individual variation in the gut microbiota of inbred laboratory mice. *Laboratory Animals*, *44*, 283–289.

https://doi.org/10.1258/la.2010.010058

Imhann, F., Vich Vila, A., Bonder, M. J., Fu, J., Gevers, Di., Visschedijk, M. C., Spekhorst, L. M., Alberts, R., Franke, L., Van Dullemen, H. M., Ter Steege, R. W. F., Huttenhower, C., DIjkstra, G., Xavier, R. J., Festen, E. A. M., Wijmenga, C., Zhernakova, A., & Weersma, R. K. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*, *67*(1), 108–119. https://doi.org/10.1136/GUTJNL-2016-312135

Jackson, M. A., Verdi, S., Maxan, M. E., Shin, C. M., Zierer, J., Bowyer, R. C. E., Martin, T., Williams, F. M. K., Menni, C., Bell, J. T., Spector, T. D., & Steves, C. J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-05184-7

Jackson, S. J., Andrews, N., Ball, D., Bellantuono, I., Gray, J., Hachoumi, L., Holmes, A., Latcham, J., Petrie, A., Potter, P., Rice, A., Ritchie, A., Stewart, M., Strepka, C., Yeoman, M., & Chapman, K. (2017). Does age matter? The impact of rodent age on study outcomes. *Laboratory Animals*, *5*(2), 160–169. https://doi.org/10.1177/0023677216653984

Jameson, E., Doxey, A. C., Airs, R., Purdy, K. J., Murrell, J. C., & Chen, Y. (2016). Metagenomic data-mining reveals contrasting microbial populations responsible for trimethylamine formation in human gut and marine ecosystems. *Microbial Genomics*, *2*(9), e000080. https://doi.org/10.1099/MGEN.0.000080

Janeway, C. A., & Medzhitov, R. (2003). Innate Immune Recognition. *Https://Doi.Org/10.1146/Annurev.Immunol.20.083001.084359*, *20*, 197–216. https://doi.org/10.1146/ANNUREV.IMMUNOL.20.083001.084359

Jo, J. K., Seo, S. H., Park, S. E., Kim, H. W., Kim, E. J., Kim, J. S., Pyo, J. Y., Cho, K. M., Kwon, S. J., Park, D. H., & Son, H. S. (2021). Gut Microbiome and Metabolome Profiles Associated with High-Fat Diet in Mice. *Metabolites*, *11*(8). https://doi.org/10.3390/METABO11080482

Johansson, M. E. V., & Hansson, G. C. (2016). Immunological aspects of intestinal mucus and mucins. *Nature Reviews Immunology 2016 16:10*, *16*(10), 639–649. https://doi.org/10.1038/nri.2016.88

Johansson, M. E. V., Sjövall, H., & Hansson, G. C. (2013). The gastrointestinal mucus system in health and disease. *Nature Reviews Gastroenterology & Hepatology 2013 10:6*, *10*(6), 352–361. https://doi.org/10.1038/nrgastro.2013.35

Johnson, A. J., Vangay, P., Al-Ghalith, G. A., Hillmann, B. M., Ward, T. L., Shields-Cutler, R. R., Kim, A. D., Shmagel, A. K., Syed, A. N., Walter, J., Menon, R., Koecher, K., & Knights, D. (2019). Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host and Microbe*, *25*(6), 789-802.e5. https://doi.org/10.1016/J.CHOM.2019.05.005

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, P. L., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Barrett, J. C., & Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, *491*, 119. https://doi.org/10.1038/nature11582

Juste, C., & Gérard, P. (2021). Cholesterol-to-coprostanol conversion by the gut microbiota: What we know, suspect, and ignore. *Microorganisms*, *9*(9). https://doi.org/10.3390/microorganisms9091881

Kaakoush, N. O. (2015). Insights into the Role of Erysipelotrichaceae in the Human Host. *Frontiers in Cellular and Infection Microbiology | Www.Frontiersin.Org*, *5*, 84. https://doi.org/10.3389/fcimb.2015.00084

Kang, D. W., Adams, J. B., Coleman, D. M., Pollard, E. L., Maldonado, J., McDonough-Means, S., Caporaso, J. G., & Krajmalnik-Brown, R. (2019). Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-42183-0

Kashyap, P. C., Chia, N., Nelson, H., Segal, E., & Elinav, E. (2017). Microbiome at the Frontier of Personalized Medicine. *Mayo Clin Proc*, *92*(12), 1855–1864. https://doi.org/10.1016/j.mayocp.2017.10.004

Kaul, A., Mandal, S., Davidov, O., & Peddada, S. D. (2017). Analysis of Microbiome Data in the Presence of Excess Zeros. *Front. Microbiol*, *8*, 2114. https://doi.org/10.3389/fmicb.2017.02114

Kechin, A., Boyarskikh, U., Kel, A., & Filipenko, M. (2017). cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol.*, *24*(11), 1138–1143. https://doi.org/10.1089/CMB.2017.0096

Keita, Å. V., & Söderholm, J. D. (2010). The intestinal barrier and its regulation by neuroimmune factors. *Neurogastroenterology & Motility*, *22*(7), 718–733. https://doi.org/10.1111/J.1365-2982.2010.01498.X

Kellermayer, R., Mir, S. A. V., Nagy-Szakal, D., Cox, S. B., Dowd, S. E., Kaplan, J. L., Sun, Y., Reddy, S., Bronsky, J., & Winter, H. S. (2012). Microbiota separation and C-reactive protein elevation in treatment-naïve pediatric granulomatous crohn disease. *Journal of Pediatric Gastroenterology and Nutrition*, *55*(3), 243–250. https://doi.org/10.1097/MPG.0B013E3182617C16

Kellogg, T. F., & Wostmann, B. S. (1969). Fecal neutral steroids and bile acids from germfree rats. *Journal Lipid Research*, *10*, 495–503. https://doi.org/10.1016/S0022-2275(20)43040-8

Kenny, D. J., Plichta, D. R., Shungin, D., Koppel, N., Hall, A. B., Fu, B., Vasan, R. S., Shaw, S. Y., Vlamakis, H., Balskus, E. P., & Xavier, R. J. (2020). Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. *Cell Host & Microbe*, *28*(2), 245-257.e6. https://doi.org/10.1016/J.CHOM.2020.05.013

Keubler, L. M., Buettner, M., Häger, C., & Bleich, A. (2015). A multihit model: Colitis lessons from the interleukin-10-deficient mouse. *Inflammatory Bowel Diseases*, *21*(8), 1967–1975. https://doi.org/10.1097/MIB.0000000000000468

Khaleghi, S., Ju, J. M., Lamba, A., & Murray, J. A. (2016). The potential utility of tight junction regulation in celiac disease: focus on larazotide acetate. *Therapeutic Advances in Gastroenterology*, *9*(1), 37–49.

Kim-Dorner, S. J., Deuster, P. A., Zeno, S. A., Remaley, A. T., & Poth, M. (2010). Should triglycerides and the triglycerides to high-density lipoprotein cholesterol ratio be used as surrogates for insulin resistance? *Metabolism - Clinical and Experimental*, *59*(2), 299–304. https://doi.org/10.1016/J.METABOL.2009.07.027

Kim, B. R., Shin, J., Guevarra, R. B., Lee, J. H., Kim, D. W., Seol, K. H., Lee, J. H., Kim, H. B., & Isaacson, R. E. (2017). Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J. Microbiol. Biotechnol.*, *27*(12), 2089–2093. https://doi.org/10.4014/JMB.1709.09027

Kincade, P. W. (1981). Formation of B Lymphocytes in Fetal and Adult Life. *Advances in Immunology*, *31*(C), 177–245. https://doi.org/10.1016/S0065-2776(08)60921-9

Kirkpatrick, C. F., Bolick, J. P., Kris-Etherton, P. M., Sikand, G., Aspry, K. E., Soffer, D. E., Willard, K. E., & Maki, K. C. (2019). Review of current evidence and clinical recommendations on the effects of low-carbohydrate and very-low-carbohydrate (including ketogenic) diets for the management of body weight and other cardiometabolic risk factors: A scientific statement from the National Lipid Association Nutrition and Lifestyle Task Force. *Journal of Clinical Lipidology*, *13*(5), 689-711.e1. https://doi.org/10.1016/J.JACL.2019.08.003

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L. I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410–422. https://doi.org/10.1038/S41579-018-0029-9

Knoch, B., Nones, K., Barnett, M. P. G., McNabb, W. C., & Roy, N. C. (2010). Diversity of caecal bacteria is altered in interleukin-10 gene-deficient mice before and after colitis onset and when fed polyunsaturated fatty acids. *Microbiology*, *156*(11), 3306–3316. https://doi.org/10.1099/MIC.0.041723-0/CITE/REFWORKS

Koh, A., De Vadder, F., Kovatcheva-Datchary, P., & Bäckhed, F. (2016). From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell*, *165*(6), 1332–1345. https://doi.org/10.1016/J.CELL.2016.05.041

König, B. (1994). Behavioral Ecology and Sociobiology Components of lifetime reproductive success in communally and solitarily nursing house mice-a laboratory study. *Behav Ecol Sociobiol*, *34*, 275–283.

König, B., & Markl, H. (1987). Maternal care in house mice - I. The weaning strategy as a means for parental manipulation of offspring quality. *Behavioral Ecology and Sociobiology*, *20*(1), 1–9. https://doi.org/10.1007/BF00292161

König, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl*(Suppl 1), 4578–4585. https://doi.org/10.1073/PNAS.1000081107

Korem, T., Zeevi, D., Zmora, N., Weissbrod, O., Bar, N., Lotan-Pompan, M., Avnit-Sagi, T., Kosower, N., Malka, G., Rein, M., Suez, J., Goldberg, B. Z., Weinberger, A., Levy, A. A., Elinav, E., & Segal, E. (2017). Bread Affects Clinical Parameters and Induces Gut Microbiome-Associated Personal Glycemic Responses. *Cell Metabolism*, *25*(6), 1243-1253.e5. https://doi.org/10.1016/J.CMET.2017.05.002

Krackow, S., & Hoeck, H. N. (1989). Sex ratio manipulation, maternal investment and behaviour during concurrent pregnancy and lactation in house mice. *Animal Behaviour*, *37*(PART 2), 177–186. https://doi.org/10.1016/0003-3472(89)90108-5

Kriaa, A., Bourgin, M., Potiron, A., Mkaouar, H., Jablaoui, A., Gérard, P., Maguin, E., & Rhimi, M. (2019). Microbial impact on cholesterol and bile acid metabolism: current status and future prospects. *Journal of Lipid Research*, *60*(2), 323. https://doi.org/10.1194/JLR.R088989

Kühn, R., Löhler, J., Rennick, D., Rajewsky, K., & Müller, W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. *Cell*, *75*(2), 263–274. https://doi.org/10.1016/0092-8674(93)80068-P

Kunz, S., & Matysik, S. (2019). A comprehensive method to determine sterol species in human faeces by GC-triple quadrupole MS. *The Journal of Steroid Biochemistry and Molecular Biology*, *190*, 99–103. https://doi.org/10.1016/J.JSBMB.2019.03.014

Lagkouvardos, I., Lesker, T. R., Hitch, T. C. A., Gálvez, E. J. C., Smit, N., Neuhaus, K., Wang, J., Baines, J. F., Abt, B., Stecher, B., Overmann, J., Strowig, T., & Clavel, T. (2019). Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome*, *7*(1), 1–15. https://doi.org/10.1186/S40168-019-0637-2/FIGURES/4

Lai, C.-H., Lin, T.-L., Huang, M.-Z., Li, S.-W., Wu, H.-Y., Chiu, Y.-F., Yang, C.-Y., Chiu, C.-H., & Lai, H.-C. (2022). Gut Commensal Parabacteroides goldsteinii MTS01 Alters Gut Microbiota Composition and Reduces Cholesterol to Mitigate Helicobacter pylori-Induced Pathogenesis. *Frontiers in Immunology*, *0*, 3297. https://doi.org/10.3389/FIMMU.2022.916848

Larsen, S., Bendtzen, K., & Nielsen, O. H. (2010). Extraintestinal manifestations of inflammatory bowel disease: Epidemiology, diagnosis, and management. *Https://Doi.Org/10.3109/07853890903559724*, *42*(2), 97–114. https://doi.org/10.3109/07853890903559724

Latorre, R., Sternini, C., De Giorgio, R., & Greenwood-Van Meerveld, B. (2016). Enteroendocrine cells: a review of their role in brain–gut communication. *Neurogastroenterology & Motility*, *28*(5), 620–630. https://doi.org/10.1111/NMO.12754

Le Roy, T., Lécuyer, E., Chassaing, B., Rhimi, M., Lhomme, M., Boudebbouze, S., Ichou, F., Haro Barceló, J., Huby, T., Guerin, M., Giral, P., Maguin, E., Kapel, N., Gérard, P., Clément, K., & Lesnik, P. (2019). The intestinal microbiota regulates host cholesterol homeostasis. *BMC Biology*, *17*(1). https://doi.org/10.1186/S12915-019-0715-8

Leamy, L. J., Kelly, S. A., Nietfeldt, J., Legge, R. M., Ma, F., Hua, K., Sinha, R., Peterson, D. A., Walter, J., Benson, A. K., & Pomp, D. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biology*, *15*(12), 552. https://doi.org/10.1186/S13059-014-0552-6/TABLES/3

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D., & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology*, *24*(2), 160–168. https://doi.org/10.1016/J.COPBIO.2012.08.005

Lee, S. H., Yun, Y., Kim, S. J., Lee, E. J., Chang, Y., Ryu, S., Shin, H., Kim, H. L., Kim, H. N., & Lee, J. H. (2018). Association between Cigarette Smoking Status and Composition of Gut Microbiota: Population-Based Cross-Sectional Study. *Journal of Clinical Medicine*, *7*(9), 282. https://doi.org/10.3390/JCM7090282

Lewis, S. J., & Heaton, K. W. (1997). Stool form scale as a useful guide to intestinal transit time. *Scandinavian Journal of Gastroenterology*, *32*(9), 920–924. https://doi.org/10.3109/00365529709011203

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 11070–11075. https://doi.org/10.1073/PNAS.0504978102

Li, L., Batt, S. M., Wannemuehler, M., A Dispirito, D., & Beitz, C. (1998). Effect of feeding of a cholesterol-reducing bacterium, Eubacterium coprostanoligenes, to germ-free mice. *Lab Anim Sci*, *48*(3), 253–255.

Li, L., Baumann, C. A., Meling, D. D., Sell, J. L., & Beitz, D. C. (1996). Effect of orally administered Eubacterium coprostanoligenes ATCC 51222 on plasma cholesterol concentration in laying hens. *Poultry Science*, *75*(6), 743–745. https://doi.org/10.3382/PS.0750743

Li, L., Buhman, K. K., Hartman, P. A., & Beitz, D. C. (1995). Hypocholesterolemic effect of Eubacterium coprostanoligenes ATCC 51222 in rabbits. *Letters in Applied Microbiology*, *20*(3), 137–140. https://doi.org/10.1111/j.1472-765X.1995.tb00410.x

Li, T., Nakatsu, G., Chen, Y. X., Yau, T. O., Chu, E., Wong, S., Szeto, C. H., Ng, S. C., Chan, F. K. L., Fang, J. Y., Sung, J. J. Y., & Yu, J. (2020). A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. *Gut*, *69*(7), 1248–1257. https://doi.org/10.1136/GUTJNL-2019-318532

Liebisch, G., Ecker, J., Roth, S., Schweizer, S., Öttl, V., Schött, H. F., Yoon, H., Haller, D., Holler, E., Burkhardt, R., & Matysik, S. (2019). Quantification of Fecal Short Chain Fatty Acids by Liquid Chromatography Tandem Mass Spectrometry-Investigation of Pre-Analytic Stability. *Biomolecules*, *9*(4). https://doi.org/10.3390/BIOM9040121

Lim, M. Y., Yoon, H. S., Rho, M., Sung, J., Song, Y. M., Lee, K., & Ko, G. (2016). Analysis of the association between host genetics, smoking, and sputum microbiota in healthy humans. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP23745

Lim, M. Y., You, H. J., Yoon, H. S., Kwon, B., Lee, J. Y., Lee, S., Song, Y. M., Lee, K., Sung, J., & Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut*, *66*(6), 1031–1038. https://doi.org/10.1136/GUTJNL-2015-311326

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature 2017 550:7674*, *550*(7674), 61–66. https://doi.org/10.1038/nature23889

Lo Sasso, G., Schlage, W. K., Boué, S., Veljkovic, E., Peitsch, M. C., & Hoeng, J. (2016). The Apoe −/− mouse model: a suitable model to study cardiovascular and respiratory diseases in the context of cigarette smoke exposure and harm reduction. *J Transl Med*, *14*, 146. https://doi.org/10.1186/s12967-016-0901-1

Lopez-Siles, M., Enrich-Capó, N., Aldeguer, X., Sabat-Mir, M., Duncan, S. H., Garcia-Gil, L. J., & Martinez-Medina, M. (2018). Alterations in the Abundance and Co-occurrence of Akkermansia muciniphila and Faecalibacterium prausnitzii in the Colonic Mucosa of Inflammatory Bowel Disease Subjects. *Frontiers in Cellular and Infection Microbiology*, *8*(SEP), 281. https://doi.org/10.3389/FCIMB.2018.00281/BIBTEX

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230. https://doi.org/10.1038/NATURE11550

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005/ASSET/CD76613D-18C6-418B-996A-AB8D3D6CA216/ASSETS/GRAPHIC/ZAM0120562270003.JPEG

Luo, J., Yang, H., & Song, B. L. (2019). Mechanisms and regulation of cholesterol homeostasis. *Nature Reviews Molecular Cell Biology 2019 21:4*, *21*(4), 225–245. https://doi.org/10.1038/s41580-019-0190-7

Luukkonen, P. K., Sädevirta, S., Zhou, Y., Kayser, B., Ali, A., Ahonen, L., Lallukka, S., Pelloux, V., Gaggini, M., Jian, C., Hakkarainen, A., Lundbom, N., Gylling, H., Salonen, A., Oresic, M., Hyötyläinen, T., Orho-Melander, M., Rissanen, A., Gastaldelli, A., … Yki-Järvinen, H. (2018). Saturated Fat Is More Metabolically Harmful for the Human Liver Than Unsaturated Fat or Simple Sugars. *Diabetes Care*, *41*(8), 1732–1739. https://doi.org/10.2337/DC18-0071

Lye, H. S., Rusul, G., & Liong, M. T. (2010). Removal of cholesterol by lactobacilli via incorporation and conversion to coprostanol. *Journal of Dairy Science*, *93*(4), 1383–1392. https://doi.org/10.3168/JDS.2009-2574

Maaser, C., Sturm, A., Vavricka, S. R., Kucharzik, T., Fiorino, G., Annese, V., Calabrese, E., Baumgart, D. C., Bettenworth, D., Nunes, P. B., Burisch, J., Castiglione, F., Eliakim, R., Ellul, P., González-Lama, Y., Gordon, H., Halligan, S., Katsanos, K., Kopylov, U., … Stoker, J. (2019). ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *Journal of Crohn's and Colitis*, 144–164. https://doi.org/10.1093/ecco-jcc/jjy113

Macfarlane GT, C. J. (2002). Probiotics, infection and immunity. *Curr Opin Infect Dis.*, *15*(5), 501–

506. https://doi.org/10.1097/00001432-200210000-00008
Madsen, K. L. (2001). The use of probiotics in gastrointestinal disease. *Canadian Journal of Gastroenterology*, *15*(12), 817–822. https://doi.org/10.1155/2001/690741
Madsen, K. L., Doyle, J. S., Jewell, L. D., Tavernini, M. M., & Fedorak, R. N. (1999). Lactobacillus species prevents colitis in interleukin 10 gene–deficient mice. *Gastroenterology*, *116*(5), 1107–1114. https://doi.org/10.1016/S0016-5085(99)70013-2
Madsen, K. L., Doyle, J. S., Tavernini, M. M., Jewell, L. D., Rennie, R. P., & Fedorak, R. N. (2000). Antibiotic therapy attenuates colitis in interleukin 10 gene–deficient mice. *Gastroenterology*, *118*(6), 1094–1105. https://doi.org/10.1016/S0016-5085(00)70362-3
Magurran, A. E. (2021). Measuring biological diversity. *Current Biology*, *31*(19), R1174–R1177. https://doi.org/10.1016/J.CUB.2021.07.049
Maharshak, N., Packey, C. D., Ellermann, M., Manick, S., Siddle, J. P., Huh, E. Y., Plevy, S., Sartor, R. B., & Carroll, I. M. (2013). Altered enteric microbiota ecology in interleukin 10-deficient mice during development and progression of intestinal inflammation. *Gut Microbes*, *4*(August 2013), 316–324.
Mähler, M., & Leiter, E. H. (2002). Genetic and Environmental Context Determines the Course of Colitis Developing in IL–10-Deficient Mice. *Inflammatory Bowel Diseases*, *8*(5), 347–355. https://doi.org/10.1097/00054725-200209000-00006
Mason, K. L., Huffnagle, G. B., Noverr, M. C., & Kao, J. Y. (2008). Overview of gut immunology. *Advances in Experimental Medicine and Biology*, *635*, 1–14. https://doi.org/10.1007/978-0-387-09550-9_1/COVER
Matysik, S., Krautbauer, S., Liebisch, G., Schött, H. F., Kjølbæk, L., Astrup, A., Blachier, F., Beaumont, M., Nieuwdorp, M., Hartstra, A., Rampelli, S., Pagotto, U., & Iozzo, P. (2021). Short-chain fatty acids and bile acids in human faeces are associated with the intestinal cholesterol conversion status. *British Journal of Pharmacology*, *178*(16), 3342–3353. https://doi.org/10.1111/BPH.15440
Matzinger, P. (2002). The Danger Model: A Renewed Sense of Self. *Science*, *296*(5566), 301–305. https://doi.org/10.1126/SCIENCE.1071059
McKnite, A. M., Elisa Perez-Munoz, M., Lu, L., Williams, E. G., Brewer, S., Elisa, M., & M, J. W. (2012). Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0039191
Midtvedt AC, M. T. (1993). Conversion of cholesterol to coprostanol by the intestinal microflora during the first two years of human life. *J Pediatr Gastroenterol Nutr*, *17*(2), 161–168.
Midtvedt, T. (1974). Microbial bile acid transformation. *Am. J. Clin. Nutr.*, *24*, 1341–1347.
Midtvedt, T., Lingaas, E., Carlstedt-Duke, B., Hoverstad, T., Midtvedt, A. C., Saxerholt, H., Steinbakk, M., & Norin, K. E. (1990). Intestinal microbial conversion of cholesterol to coprostanol in man. *APMIS*, *98*(7–12), 839–844. https://doi.org/10.1111/J.1699-0463.1990.TB05004.X
Mowat, C., Cole, A., Windsor, A., Ahmad, T., Arnott, I., Driscoll, R., Mitton, S., Orchard, T., Rutter, M., & Younge, L. (2011). Guidelines for the management of inflammatory bowel disease in adults. *Gut*, *60*(5), 571–607. https://doi.org/10.1136/gut.2010.224154
Mueller, N. T., Whyatt, R., Hoepner, L., Oberfield, S., Dominguez-Bello, M. G., Widen, E. M., Hassoun, A., Perera, F., & Rundle, A. (2014). Prenatal exposure to antibiotics, cesarean section and risk of childhood obesity. *International Journal of Obesity 2015 39:4*, *39*(4), 665–670. https://doi.org/10.1038/ijo.2014.180
Nguyen, T. L. A., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease Models & Mechanisms*, *8*(1), 1–16. https://doi.org/10.1242/DMM.017400
Nogal, A., Louca, P., Zhang, X., Wells, P. M., Steves, C. J., Spector, T. D., Falchi, M., Valdes, A. M., & Menni, C. (2021). Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat. *Frontiers in Microbiology*, *12*, 1943. https://doi.org/10.3389/FMICB.2021.711359/BIBTEX
Norin, E. (2008). Intestinal cholesterol conversion in adults and elderly from four different European countries. *Annals of Nutrition & Metabolism*, *52 Suppl 1*(SUPPL. 1), 12–14. https://doi.org/10.1159/000115341
Norwitz, N. G., Feldman, D., Soto-Mota, A., Kalayjian, T., & Ludwig, D. S. (2022). Mechanisms Elevated LDL Cholesterol with a Carbohydrate-Restricted Diet: Evidence for a “Lean Mass Hyper-Responder” Phenotype. *Curr Dev Nutr .*, *6*(1). https://academic.oup.com/cdn/.
O’Neill, B., & Raggi, P. (2020). The ketogenic diet: Pros and cons. *Atherosclerosis*, *292*, 119–126.

https://doi.org/10.1016/J.ATHEROSCLEROSIS.2019.11.021

Ohigashi, S., Sudo, K., Kobayashi, D., Takahashi, O., Takahashi, T., Asahara, T., Nomoto, K., & Onodera, H. (2013). Changes of the Intestinal Microbiota, Short Chain Fatty Acids, and Fecal pH in Patients with Colorectal Cancer. *Dig Dis Sci*, *58*, 1717–1726. https://doi.org/10.1007/s10620-012-2526-4

Ormerod, K. L., Wood, D. L. A., Lachner, N., Gellatly, S. L., Daly, J. N., Parsons, J. D., Dal'molin, C. G. O., Palfreyman, R. W., Nielsen, L. K., Cooper, M. A., Morrison, M., Hansbro, P. M., & Hugenholtz, P. (2016). Genomic characterization of the uncultured Bacteroidales family S24-7 inhabiting the guts of homeothermic animals. *Microbiome*, *36*. https://doi.org/10.1186/s40168-016-0181-2

Palm, N. W., De Zoete, M. R., Cullen, T. W., Barry, N. A., Stefanowski, J., Hao, L., Degnan, P. H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J. H., Goodman, A. L., & Flavell, R. A. (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*, *158*(5), 1000–1010. https://doi.org/10.1016/J.CELL.2014.08.006

Paoli, A., Rubini, A., Volek, J. S., & Grimaldi, K. A. (2013). Beyond weight loss: A review of the therapeutic uses of very-low-carbohydrate (ketogenic) diets. *European Journal of Clinical Nutrition*, *67*(8), 789–796. https://doi.org/10.1038/ejcn.2013.116

Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S. P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., Schauer, D. B., Ward, D. V., Korzenik, J. R., Xavier, R. J., Bousvaros, A., & Alm, E. J. (2012). Non-Invasive Mapping of the Gastrointestinal Microbiota Identifies Children with Inflammatory Bowel Disease. *PLOS ONE*, *7*(6), e39242. https://doi.org/10.1371/JOURNAL.PONE.0039242

Peña, J. A., Li, S. Y., Wilson, P. H., Thibodeau, S. A., Szary, A. J., & Versalovic, J. (2004). Genotypic and Phenotypic Studies of Murine Intestinal Lactobacilli: Species Differences in Mice with and without Colitis. *Applied and Environmental Microbiology*, *70*(1), 558–568. https://doi.org/10.1128/AEM.70.1.558-568.2004

Perdigó, G., Galdeano, M., Valdez, J. C., & Medici, M. (2002). Interaction of lactic acid bacteria with the gut immune system. *European Journal of Clinical Nutrition*, *56*, 21–26. https://doi.org/10.1038/sj.ejcn.1601658

Pereira, F. C., Wasmund, K., Cobankovic, I., Jehmlich, N., Herbold, C. W., Soo Lee, K., Sziranyi, B., Vesely, C., Decker, T., Stocker, R., Warth, B., von Bergen, M., Wagner, M., & Berry, D. (2020). Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. *Nature Communications*, *11*(1), 5104. https://doi.org/10.1038/s41467-020-18928-1

Png, C. W., Lindén, S. K., Gilshenan, K. S., Zoetendal, E. G., McSweeney, C. S., Sly, L. I., McGuckin, M. A., & Florin, T. H. J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *American Journal of Gastroenterology*, *105*(11), 2420–2428. https://doi.org/10.1038/AJG.2010.281

Podlesny, D., Arze, C., Dörner, E., Verma, S., Dutta, S., Walter, J., & Fricke, W. F. (2022). Metagenomic strain detection with SameStr: identification of a persisting core gut microbiota transferable by fecal transplantation. *Microbiome*, *10*(1), 1–15. https://doi.org/10.1186/S40168-022-01251-W/FIGURES/5

Podlesny, D., & Fricke, W. F. (2021). Strain inheritance and neonatal gut microbiota development: A meta-analysis. *International Journal of Medical Microbiology*, *311*(3). https://doi.org/10.1016/J.IJMM.2021.151483

Poyet, M., Groussin, M., Gibbons, S. M., Avila-Pacheco, J., Jiang, X., Kearney, S. M., Perrotta, A. R., Berdy, B., Zhao, S., Lieberman, T. D., Swanson, P. K., Smith, M., Roesemann, S., Alexander, J. E., Rich, S. A., Livny, J., Vlamakis, H., Clish, C., Bullock, K., … Alm, E. J. (2019). A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nature Medicine 2019 25:9*, *25*(9), 1442–1452. https://doi.org/10.1038/s41591-019-0559-3

Qamar, A., Giugliano, R. P., Keech, A. C., Kuder, J. F., Murphy, S. A., Kurtz, C. E., Wasserman, S. M., Sever, P. S., Pedersen, T. R., & Sabatine, M. S. (2019). Interindividual Variation in Low-Density Lipoprotein Cholesterol Level Reduction With Evolocumab: An Analysis of FOURIER Trial Data. *JAMA Cardiology*, *4*(1), 59–63. https://doi.org/10.1001/JAMACARDIO.2018.4178

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*. https://doi.org/10.1093/NAR/GKS1219

Redinbo, M. R. (2020). The Microbiome Revolution Turns to Cholesterol. *Cell Host and Microbe*, *28*, 154–156. https://doi.org/10.1016/j.chom.2020.07.011

Riley, J. K., Takeda, K., Akira, S., & Schreiber, R. D. (1999). Interleukin-10 Receptor Signaling through the JAK-STAT Pathway. *Journal of Biological Chemistry*, *274*(23), 16513–16521. https://doi.org/10.1074/jbc.274.23.16513

Rodríguez-Lago, I., Ramírez, C., Merino, O., Azagra, I., Maiz, A., Zapata, E., Higuera, R., Montalvo, I., Fernández-Calderón, M., Arreba, P., Carrascosa, J., Iriarte, A., Muñoz-Navas, M., Cabriada, J. L., & Barreiro-de Acosta, M. (2020). Early microscopic findings in preclinical inflammatory bowel disease. *Digestive and Liver Disease*, *52*(12), 1467–1472. https://doi.org/10.1016/j.dld.2020.05.052

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature 2018 555:7695*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Round, J. L., & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12204–12209. https://doi.org/10.1073/PNAS.0909122107

Salyers, A. A., Sperry, J. F., Wilkins, T. D., Walker, A. R., & Richardson, N. . (1977). Neutral steroid concentrations in the faeces of North American White and South African Black populations at different risks for cancer of the colon. *S. Afr. Med. J*, *51*, 823–827.

Saraiva, M., & O'Garra, A. (2010). The regulation of IL-10 production by immune cells. *Nature Reviews Immunology 2010 10:3*, *10*(3), 170–181. https://doi.org/10.1038/nri2711

Sartor, R. B. (2001). Induction of mucosal immune responses by bacteria and bacterial components. *Current Opinion in Gastroenterology*, *17*(6), 555–561. https://doi.org/10.1097/00001574-200111000-00013

Sartor, R. B. (2008). Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology*, *134*(2), 577–594. https://doi.org/10.1053/J.GASTRO.2007.11.059

Scanu, A. M., & Wisdom, C. (1972). Serum Lipoproteins structure and function. *Annu. Rev. Biochem.*, *41*, 703–730.

Scharrer, E. (1999). Control of food intake by fatty acid oxidation and ketogenesis. *Nutrition*, *15*(9), 704–714. https://doi.org/10.1016/S0899-9007(99)00125-2

Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S. B., Jehmlich, N., Basic, M., Dupont, A., Hornef, M., Von Bergen, M., Bleich, A., & Haller, D. (2015). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut*, *65*(2), 225–237. https://doi.org/10.1136/gutjnl-2015-309333

Scheiman, J., Luber, J. M., Chavkin, T. A., MacDonald, T., Tung, A., Pham, L. D., Wibowo, M. C., Wurth, R. C., Punthambaker, S., Tierney, B. T., Yang, Z., Hattab, M. W., Avila-Pacheco, J., Clish, C. B., Lessard, S., Church, G. M., & Kostic, A. D. (2019). Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism. *Nature Medicine 2019 25:7*, *25*(7), 1104–1109. https://doi.org/10.1038/s41591-019-0485-4

Scheinberg, P. (1965). Observations on cerebral carbohydrate metabolism an man. *Annals of Internal Medicine*, *62*, 367–371. https://doi.org/10.7326/0003-4819-62-2-367

Schirmer, M., Garner, A., Vlamakis, H., & Xavier, R. J. (2019). Microbial genes and pathways in inflammatory bowel disease. *Nature Reviews. Microbiology*, *17*(8), 497. https://doi.org/10.1038/S41579-019-0213-6

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M., & Bork, P. (2012). Genomic variation landscape of the human gut microbiome. *Nature 2012 493:7430*, *493*(7430), 45–50. https://doi.org/10.1038/nature11711

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., De Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A. G., & Crittenden, A. N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, *5*, 3654. https://doi.org/10.1038/ncomms4654

Schött, H. F., Krautbauer, S., Höring, M., Liebisch, G., & Matysik, S. (2018). A Validated, Fast Method for Quantification of Sterols and Gut Microbiome Derived 5α/β-Stanols in Human Feces by Isotope Dilution LC-High-Resolution MS. *Analytical Chemistry*, *90*(14), 8487–8494.

https://doi.org/10.1021/ACS.ANALCHEM.8B01278
Schoultz, I., & Keita, Å. V. (2020). The Intestinal Barrier and Current Techniques for the Assessment of Gut Permeability. *Cells*, *9*(8), 1909. https://doi.org/10.3390/cells9081909
Schroeder, B. O., Birchenough, G. M. H., Ståhlman, M., Arike, L., Johansson, M. E. V., Hansson, G. C., & Bäckhed, F. (2018). Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration. *Cell Host & Microbe*, *23*(1), 27-40.e7. https://doi.org/10.1016/J.CHOM.2017.11.004
Schwab, C., Berry, D., Rauch, I., Rennisch, I., Ramesmayer, J., Hainzl, E., Heider, S., Decker, T., Kenner, L., Mü Ller, M., Strobl, B., Wagner, M., Schleper, C., Loy, A., & Urich, T. (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *The ISME Journal*, *8*, 1101–1114. https://doi.org/10.1038/ismej.2013.223
Sekimoto, H., Shimada, O., Makanishi, M., Nakano, T., & Katayama, O. (1983). Interrelationship between serum and fecal sterols. *Japanese Journal of Medicine*, *22*(1), 14–20. https://doi.org/10.2169/INTERNALMEDICINE1962.22.14
Sellon, R. K., Tonkonogy, S., Schultz, M., Dieleman, L. A., Grenther, W., Balish, E., Rennick, D. M., & Sartor, R. B. (1998). Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection and Immunity*, *66*(11), 5224–5231. https://doi.org/10.1128/IAI.66.11.5224-5231.1998/ASSET/4D668248-FEBB-4CBA-B1B6-82A04342A8CF/ASSETS/GRAPHIC/II1180410004.JPEG
Sender, R., Fuchs, S., & Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*, *164*(3), 337–340. https://doi.org/10.1016/J.CELL.2016.01.013
Sengupta, S., Peterson, T. R., Laplante, M., Oh, S., & Sabatini, D. M. (2010). mTORC1 controls fasting-induced ketogenesis and its modulation by ageing. *Nature 2010 468:7327*, *468*(7327), 1100–1104. https://doi.org/10.1038/nature09584
Seregin, S. S., Golovchenko, N., Schaf, B., Chen, J., Pudlo, N. A., Mitchell, J., Baxter, N. T., Zhao, L., Schloss, P. D., Martens, E. C., Eaton, K. A., & Chen, G. Y. (2017). NLRP6 protects IL10−/− mice from colitis by limiting colonization of Akkermansia muciniphila. *Cell Reports*, *19*(4), 733. https://doi.org/10.1016/J.CELREP.2017.03.080
Singh, G., Brass, A., Cruickshank, S. M., & Knight, C. G. (2021). Cage and maternal effects on the bacterial communities of the murine gut. *Scientific Reports*, *11*, 9841. https://doi.org/10.1038/s41598-021-89185-5
Siri-Tarino, P. W., Sun, Q., Hu, F. B., & Krauss, R. M. (2010). Saturated fat, carbohydrate, and cardiovascular disease. *The American Journal of Clinical Nutrition*, *91*(3), 502–509. https://doi.org/10.3945/AJCN.2008.26285
Smith, B. J., Miller, R. A., Ericsson, A. C., Harrison, D. C., Strong, R., & Schmidt, T. M. (2019). Changes in the gut microbiome and fermentation products concurrent with enhanced longevity in acarbose-treated mice. *BMC Microbiology*, *19*(1), 1–16. https://doi.org/10.1186/S12866-019-1494-7/FIGURES/5
Somerfield, P. J., Clarke, K. R., & Gorley, R. N. (2021). A generalised analysis of similarities (ANOSIM) statistic for designs with ordered factors. *Austral Ecology*, *46*(6), 901–910. https://doi.org/10.1111/AEC.13043
Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., Gregory Caporaso, J., Knights, D., Clemente, J. C., Nakielny, S., Gordon, J. I., Fierer, N., & Knight, R. (2013). Cohabiting family members share microbiota with one another and with their dogs. *ELife*, *2*(2). https://doi.org/10.7554/ELIFE.00458
Sonnenburg, E. D., & Sonnenburg, J. L. (2014). Starving our microbial self: The deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metabolism*, *20*(5), 779–786. https://doi.org/10.1016/J.CMET.2014.07.003
Stappenbeck, T. S., & Virgin, H. W. (2016). Accounting for reciprocal host–microbiome interactions in experimental science. *Nature 2016 534:7606*, *534*(7606), 191–199. https://doi.org/10.1038/nature18285
Stewart, J. A., Chadwick, V. S., & Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *Journal of Medical Microbiology*, *54*(12), 1239–1242. https://doi.org/10.1099/JMM.0.46189-0/CITE/REFWORKS
Strandwitz, P., Kim, K. H., Terekhova, D., Liu, J. K., Sharma, A., Levering, J., McDonald, D., Dietrich, D., Ramadhar, T. R., Lekbua, A., Mroue, N., Liston, C., Stewart, E. J., Dubin, M. J., Zengler,

K., Knight, R., Gilbert, J. A., Clardy, J., & Lewis, K. (2018). GABA-modulating bacteria of the human gut microbiota. *Nature Microbiology 2018 4:3*, *4*(3), 396–403. https://doi.org/10.1038/s41564-018-0307-3

Su, X. (2021). Elucidating the Beta-Diversity of the Microbiome: from Global Alignment to Local Alignment. *MSystems*, *6*(4). https://doi.org/10.1128/MSYSTEMS.00363-21/ASSET/B7EB79F6-8035-4759-86A8-0CB86E6EE7FA/ASSETS/IMAGES/MEDIUM/MSYSTEMS.00363-21-F001.GIF

Tlaskalová-Hogenová, H., Tpánková, R., Kozáková, H., Hudcovic, T., Vannucci, L., Tuková, L., Rossmann, P., Hrní, T., Kverka, M., Zákostelská, Z., Klimeová, K., Pibylová, J., Bártová, J., Sanchez, D., Fundová, P., Borovská, D., Rtková, D., Zídek, Z., Schwarzer, M., … Funda, D. P. (2011). The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cellular & Molecular Immunology 2011 8:2*, *8*(2), 110–120. https://doi.org/10.1038/cmi.2010.67

Torow, N., & Hornef, M. W. (2017). The Neonatal Window of Opportunity: Setting the Stage for Life-Long Host-Microbial Interaction and Immune Homeostasis. *The Journal of Immunology*, *198*(2), 557–563. https://doi.org/10.4049/JIMMUNOL.1601253

Torres, J., Committee, S. W. S., Halfvarson, J., Committee, S. W. S., Rodríguez-Lago, I., Committee, S. W. S., Hedin, C. R. H., Committee, S. W. S., Jess, T., Committee, S. W. S., Dubinsky, M., Committee, S. W. S., Croitoru, K., Committee, S. W. S., Colombel, J.-F., & Committee, S. W. S. (2021). Results of the Seventh Scientific Workshop of ECCO: Precision Medicine in IBD—Prediction and Prevention of Inflammatory Bowel Disease. *Journal of Crohn's and Colitis*, *15*(9), 1443–1454. https://doi.org/10.1093/ECCO-JCC/JJAB048

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2008). A core gut microbiome in obese and lean twins. *Nature 2008 457:7228*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006a). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*, 1027–1031. https://doi.org/10.1038/nature05414

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006b). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Tuteja, S., & Ferguson, J. F. (2019). Gut Microbiome and Response to Cardiovascular Drugs. *Circulation. Genomic and Precision Medicine*, *12*(9), 421–429. https://doi.org/10.1161/CIRCGEN.119.002314

Urbain, P., & Bertz, H. (2016). Monitoring for compliance with a ketogenic diet: what is the best time of day to test for urinary ketosis? *Nutr Metabol (Lond)*, *4*(13), 77. https://doi.org/10.1186/s12986-016-0136-4

Urbain, P., Strom, L., Morawski, L., Wehrle, A., Deibert, P., & Bertz, H. (2017). Impact of a 6-week non-energy-restricted ketogenic diet on physical fitness, body composition and biochemical parameters in healthy adults. *Nutrition & Metabolism*, *14*(1). https://doi.org/10.1186/S12986-017-0175-5

Ussar, S., Griffin, N. W., Bezy, O., Fujisaka, S., Vienberg, S., Softic, S., Deng, L., Bry, L., Gordon, J. I., & Kahn, C. R. (2015). Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome. *Cell Metabolism*, *22*(3), 516–530. https://doi.org/10.1016/J.CMET.2015.07.007

van de Merwe, J. P., Stegeman, J. H., & Hazenberg, M. P. (1983). The resident faecal flora is determined by genetic characteristics of the host. Implications for Crohn's disease? *Antonie van Leeuwenhoek*, *49*(2), 119–124. https://doi.org/10.1007/BF00393669

van Passel, M. W. J., Kant, R., Zoetendal, E. G., Plugge, C. M., Derrien, M., Malfatti, S. A., Chain, P. S. G., Woyke, T., Palva, A., de Vos, W. M., & Smidt, H. (2011). The Genome of Akkermansia muciniphila, a Dedicated Intestinal Mucin Degrader, and Its Use in Exploring Intestinal Metagenomes. *PLOS ONE*, *6*(3), e16876. https://doi.org/10.1371/JOURNAL.PONE.0016876

Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R. Y., Joossens, M., & Raes, J. (2016). Stool

consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*, *65*(1), 57–62. https://doi.org/10.1136/gutjnl-2015-309618

Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R. Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., & Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature 2017 551:7681*, *551*(7681), 507–511. https://doi.org/10.1038/nature24460

Veiga, P., Juste, C., Lepercq, P., Saunier, K., Béguet, F., & Gérard, P. (2005). Correlation between faecal microbial community structure and cholesterol-to-coprostanol conversion in the human gut. *FEMS Microbiology Letters*, *242*(1), 81–86. https://doi.org/10.1016/J.FEMSLE.2004.10.042

Verma, R., Balakrishnan, L., Sharma, K., Khan, A. A., Advani, J., Gowda, H., Srikanth, &, Tripathy, P., Suar, M., Pandey, A., Gandotra, S., Keshava Prasad, T. S., & Shankar, S. (2016). A network map of Interleukin-10 signaling pathway. *Journal of Cell Communication and Signaling*, *10*, 61–67. https://doi.org/10.1007/s12079-015-0302-x

Vicedomini, R., Quince, C., Darling, A. E., & Chikhi, R. (2021). Strainberry: automated strain separation in low-complexity metagenomes using long reads. *Nature Communications 2021 12:1*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-24515-9

Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., Carlsson, C. M., Asthana, S., Zetterberg, H., Blennow, K., Bendlin, B. B., & Rey, F. E. (2017). Gut microbiome alterations in Alzheimer's disease. *Scientific Reports 2017 7:1*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-13601-y

Von Rosenvinge, E. C., Song, Y., White, J. R., Maddox, C., Blanchard, T., & Fricke, W. F. (2013). Immune status, antibiotic medication and pH are associated with changes in the stomach fluid microbiota. *The ISME Journal*, *7*(7), 1354–1366. https://doi.org/10.1038/ISMEJ.2013.33

Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., & Belkaid, Y. (2020). Host variables confound gut microbiota studies of human disease. *Nature*, *587*. https://doi.org/10.1038/s41586-020-2881-9

Wagner, B. D., Grunwald, G. K., Zerbe, G. O., Mikulich-Gilbertson, S. K., Robertson, C. E., Zemanick, E. T., & Harris, J. K. (2018). On the use of diversity measures in longitudinal sequencing studies of microbial communities. *Frontiers in Microbiology*, *9*(MAY), 1037. https://doi.org/10.3389/FMICB.2018.01037/BIBTEX

Wainfan, E., Henkin, G., Rice, L. I., & Marx, W. (1952). Effects of antibacterial drugs on the total cholesterol balance of cholesterol-fed mice. *Archives of Biochemistry and Biophysics*, *38*(1), 187–193. https://doi.org/10.1016/0003-9861(52)90023-4

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., & Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ISMEJ.2010.118

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G., & Petrovska, L. (2011). *High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease*. https://doi.org/10.1186/1471-2180-11-7

Walker, D. A., & Smith, T. J. (2020). Logistic Regression Under Sparse Data Conditions. *Journal of Modern Applied Statistical Methods*, *18*(2), 25. https://doi.org/10.22237/jmasm/1604190660

Wang, C.-S.-E., Li, W.-B., Wang, H.-Y., Ma, Y.-M., Zhao, X.-H., Hong Yang, J.-M. Q., & Li, J.-N. (2018). VSL#3 can prevent ulcerative colitis-associated carcinogenesis in mice. *World Journal of Gastroenterology*, *24*(37), 4254–4262.

Wang, J., Lang, T., Shen, J., Dai, J., Tian, L., & Wang, X. (2019). Core gut bacteria analysis of healthy mice. *Frontiers in Microbiology*, *10*(APR). https://doi.org/10.3389/FMICB.2019.00887

Wang, Y., Ye, X., Ding, D., & Lu, Y. (2020). Characteristics of the intestinal flora in patients with peripheral neuropathy associated with type 2 diabetes. *The Journal of International Medical Research*, *48*(9). https://doi.org/10.1177/0300060520936806

Weiss, G. A., & Thierry Hennet, ·. (2017). Mechanisms and consequences of intestinal dysbiosis. *Cell. Mol. Life Sci*, *74*, 2959–2977. https://doi.org/10.1007/s00018-017-2509-x

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1),

1–18. https://doi.org/10.1186/S40168-017-0237-Y/FIGURES/8

Westman, E. C., Feinman, R. D., Mavropoulos, J. C., Vernon, M. C., Volek, J. S., Wortman, J. A., Yancy, W. S., & Phinney, S. D. (2007). Low-carbohydrate nutrition and metabolism. *The American Journal of Clinical Nutrition*, *86*(2), 276–284. https://doi.org/10.1093/AJCN/86.2.276

Wilkins, T. D., & Hackman, A. S. (1974). Two Patterns of Neutral Steroid Conversion in the Feces of Normal North Americans. *Cancer Research*, *34*, 2250–2254.

Willis, A. D. (2019). Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*, *10*, 2407. https://doi.org/10.3389/FMICB.2019.02407/BIBTEX

Wilmanski, T., Kornilov, S. A., Diener, C., Conomos, M. P., Lovejoy, J. C., Sebastiani, P., Orwoll, E. S., Hood, L., Price, N. D., Rappaport, N., Magis, A. T., & Gibbons, S. M. (2022). Heterogeneity in statin responses explained by variation in the human gut microbiome. *Med (New York, N.Y.)*, *3*(6), 388-405.e6. https://doi.org/10.1016/J.MEDJ.2022.04.007

Wohlgemuth, S., Keller, S., Kertscher, R., Stadion, M., Haller, D., Kisling, S., Jahreis, G., Blaut, M., & Loh, G. (2011). Intestinal steroid profiles and microbiota composition in colitic mice. *Www.Landesbioscience.Com Gut Microbes 159 Gut Microbes*, *2*(3), 159–166. https://doi.org/10.4161/gmic.2.3.16104

Wolter, M., Grant, E. T., Boudaud, M., Steimle, A., Pereira, G. V., Martens, E. C., & Desai, M. S. (2021). Leveraging diet to engineer the gut microbiome. *Nature Reviews Gastroenterology & Hepatology 2021 18:12*, *18*(12), 885–902. https://doi.org/10.1038/s41575-021-00512-7

Wu, T.-R., Lin, S., Chang, J., Lin, T.-L., Martel, J., Ko, Y.-F., Ojcius, D. M., Lu, C.-C., Young, J. D., & Lai, H.-C. (2019). Gut commensal Parabacteroides goldsteinii plays a predominant role in the anti-obesity effects of polysaccharides isolated from Hirsutella sinensis. *Gut*, *68*(2), 248–262. https://doi.org/10.1136/gutjnl-2017-315458

Xavier, R. J., & Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, *448*, 427–434. https://doi.org/10.1038/nature06005

Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., Ward, K. J., Jackson, M. A., Xia, Y., Chen, X., Chen, B., Xia, H., Xu, C., Li, F., Xu, X., Al-Aama, J. Y., Yang, H., Wang, J., Kristiansen, K., … Jia, H. (2016). Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. *Cell Systems*, *3*(6), 572-584.e3. https://doi.org/10.1016/J.CELS.2016.10.004

Yamamoto, Y., Kanayama, N., Nakayama, Y., & Matsushima, N. (2022). Personalized Medicine Review Current Status, Issues and Future Prospects of Personalized Medicine for Each Disease. *Journal of Personalized Medicine*, *12*, 444. https://doi.org/10.3390/jpm12030444

Yan, Y., Nguyen, L. H., Franzosa, E. A., & Huttenhower, C. (2020). Strain-level epidemiology of microbial communities and the human microbiome. *Genome Medicine 2020 12:1*, *12*(1), 1–16. https://doi.org/10.1186/S13073-020-00765-Y

Yang, J., Park, J., Park, S., Baek, I., & Chun, J. (2019). Introducing Murine Microbiome Database (MMDB): A Curated Database with Taxonomic Profiling of the Healthy Mouse Gastrointestinal Microbiome. *Microorganisms*, *7*(11), 480. https://doi.org/10.3390/microorganisms7110480

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Gloria Dominguez-Bello, M., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*, 222–267. https://doi.org/10.1038/nature11053

Zanotti, I., Turroni, F., Piemontese, A., Mancabelli, L., Milani, C., Viappiani, A., Prevedini, G., Sanchez, B., Margolles, A., Elviri, L., Franco, B., van Sinderen, D., & Ventura, M. (2015). Evidence for cholesterol-lowering activity by Bifidobacterium bifidum PRL2010 through gut microbiota modulation. *Applied Microbiology and Biotechnology*, *99*(16), 6813–6829. https://doi.org/10.1007/S00253-015-6564-7/FIGURES/8

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., … Segal, E. (2015). Personalized Nutrition by Prediction of Glycemic Responses. *Cell*, *163*(5), 1079–1094. https://doi.org/10.1016/J.CELL.2015.11.001

Zhai, R., Xue, X., Zhang, L., Yang, X., Zhao, L., & Zhang, C. (2019). Strain-specific anti-inflammatory properties of two Akkermansia muciniphila strains on chronic colitis in mice. *Frontiers in Cellular and Infection Microbiology*, *9*(JUL), 239. https://doi.org/10.3389/FCIMB.2019.00239/BIBTEX

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell,

M. D., Wing, R., Rittmann, B. E., & Krajmalnik-Brown, R. (2009). Human gut microbiota in obesity and after gastric bypass. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2365. https://doi.org/10.1073/PNAS.0812600106

Zhang, J.-M., & An, J. (2007). Cytokines, Inflammation and Pain. *Int Anesthesiol Clin*, *45*(2), 27–37. https://doi.org/10.1097/AIA.0b013e318034194e

Zhang, T., Ji, X., Lu, G., & Zhang, F. (2021). The potential of Akkermansia muciniphila in inflammatory bowel disease. *Applied Microbiology and Biotechnology*, *105*, 5785–5794. https://doi.org/10.1007/s00253-021-11453-1/Published

Zhang, X., Shen, D., Fang, Z., Jie, Z., Qiu, X., Zhang, C., Chen, Y., & Ji, L. (2013). Human Gut Microbiota Changes Reveal the Progression of Glucose Intolerance. *PLOS ONE*, *8*(8), e71108. https://doi.org/10.1371/JOURNAL.PONE.0071108

Zhang, X., Yang, S., Chen, J., & Su, Z. (2019). Unraveling the Regulation of Hepatic Gluconeogenesis. *Frontiers in Endocrinology*, *9*(JAN). https://doi.org/10.3389/FENDO.2018.00802

Zhao, X., Zhang, Z., Hu, B., Huang, W., Yuan, C., & Zou, L. (2018). Response of gut microbiota to metabolite changes induced by endurance exercise. *Frontiers in Microbiology*, *9*(APR), 765. https://doi.org/10.3389/FMICB.2018.00765/BIBTEX

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A. V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S. A., Joossens, M., Cenit, M. C., Deelen, P., Swertz, M. A., Weersma, R. K., … Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science (New York, N.Y.)*, *352*(6285), 565–569. https://doi.org/10.1126/SCIENCE.AAD3369

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, *570*, 464–467. https://doi.org/10.1038/s41586-019-1291-3

Zinöcker, M. K., Svendsen, K., & Dankel, S. N. (2021). The homeoviscous adaptation to dietary lipids (HADL) model explains controversies over saturated fat, cholesterol, and cardiovascular disease risk. *The American Journal of Clinical Nutrition*, *113*(2), 277–289. https://doi.org/10.1093/AJCN/NQAA322

Zoetendal, E. G., Akkermans, A. D. L., Akkermans-van Vliet, W. M., De Visser, J. A. G. M., & De Vos, W. M. (2001). The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease*, *13*(3), 129–134. https://doi.org/10.1080/089106001750462669

Zschaler, J., Schlorke, D., and Arnhold, J. (2014). Differences in innate immuneresponse between man and mouse. *Crit. Rev. Immunol.*, *34*, 433–454. https://doi.org/10.1615/critrevimmunol.2014011600

Appendix

Supplementary Data

Supplementary tables and figures are deposited on Zenodo, an open-access repository developed under the European OpenAIRE program and operated by CERN or are currently under revision at *Cell Press Community* (Bubeck et al., *Microbial cholesterol conversion and converter type-dependent LDL-C response to low-carbohydrate high-fat diet*). The supplementary data can be accessed under doi:10.5281/zenodo.7513747 (Bubeck, 2023).

Table 7. List of digital supplementary tables deposited on Zenodo.

Label	Table description
Table S1.	Scoring parameter to assess the well-being of mice.
Table S2.	Fecal microbiota genus-level taxa counts of the two LCHF human cohort.
Table S3.	Summary of sequencing runs for all three cohorts.
Table S4.	QIIME options, R and UNIX scripts and commands used for sequence processing and data analysis of the two LCHF human cohort.
Table S5.	Associations between taxonomic microbiota compositions, serum lipids and cholesterol-to-coprostanol conversion as determined by generalized linear mixed effects models.
Table S6.	Statistics of the generalized linear mixed effect model estimating relative taxon abundances of the *IL-10* knockout mouse cohort.
Table S7.	Statistics of the generalized linear mixed effect model estimating absolute taxon abundances of the *IL-10* knockout mouse cohort.
Table S8.	Statistics of single predictor models during model building of the *IL-10* knockout mouse model.
Table S9.	Statistics of the final logistic regression models for the prediction of early mouse withdrawal.
Table S10.	Cross validation of final logistic regression models for the prediction of early mouse withdrawal.

Continued on the next page

Table 7 continued

Table S11.	Overview of KETO and CARBFUNC study cohort parameters.
Table S12.	Statistics for main figure group comparisons of the *IL-10* knockout mice trial.
Table S13.	Statistics for supplementary figure group comparisons.

Table 8. List of digital supplementary figures deposited on Zenodo.

Label	Figure description
Figure S1.	Evaluation of intra-individual short-term microbial alterations of the human KETO cohort (a) and batch effect evaluation of the human CARBFUNC (b - e) and the murine *IL-10* knockout/wild-type co-housing cohort (f, g)
Figure S2.	Associations of fecal taxonomic microbiota composition with the fecal coprostanol/cholesterol ratio.
Figure S3.	No associations between fecal cholesterol and coprostanol concentrations and serum lipid levels in both cohorts.
Figure S4.	Similar fecal fatty acid profiles in cholesterol high and low-converters from the KETO cohort before the LCHF diet intervention.
Figure S5.	Mouse 57 had a higher body weight than other wild type mice after maturation.
Figure S6.	The microbial diversity and overall bacterial density depended on genotype in the first week of cage allocation and under healthy conditions thereafter.
Figure S7.	Three taxa were abundant in only litter 3 or 4 mice at the day of cage allocation.
Figure S8.	Colitis did not affect alpha diversity within samples.
Figure S9.	Relative abundance alterations of various bacteria were associated with the phenotype rather than litter and cage.
Figure S10.	Absolute abundance alterations of various bacteria were associated with the phenotype rather than litter and cage.
Figure S11.	The mean Bray-Curtis Dissimilarity between each IL-10 deficient and all wild type mice is an almost perfect predictor of early mice withdrawal.

Additional publications

In addition to the work presented in this thesis, the following research articles have been published with me as an author. These publications are independent of the work presented in this thesis.

Appendix 1: Bacterial microbiota diversity and composition in red and white wines correlate with plant-derived DNA contributions and botrytis infection

Alena M. Bubeck, Lena Preiss, Anna Jung, Elisabeth Dörner, Daniel Podlesny, Marija Kulis, Cynthia Maddox, Cesar Arze, Christian Zörb, Nikolaus Merkt & W. Florian Fricke (2020). Bacterial microbiota diversity and composition in red and white wines correlate with plant-derived DNA contributions and botrytis infection. *Scientific Reports* 10, 13828 (2020). https://doi.org/10.1038/s41598-020-70535-8

Personal contribution:

I re-processed the existing sequencing data and applied new bioinformatical methods. Additionally, I performed quantification of the 16S rRNA gene amplicon via qPCR and conducted all subsequent statistical analyses and graphical visualizations leading to this first author publication. I also wrote and revised the manuscript.

Credit author contributions: Study Design: C.Z., N.M., W.F.F.; Data acquisition: A.M.B., A.J., C.M., E.D., L.P., M.K.; Data analysis: A.M.B., C.A., D.P.; Writing of manuscript: A.M.B., W.F.F.

SCIENTIFIC REPORTS
nature research

Check for updates

OPEN

Bacterial microbiota diversity and composition in red and white wines correlate with plant-derived DNA contributions and botrytis infection

Alena M. Bubeck[1], Lena Preiss[1], Anna Jung[1], Elisabeth Dörner[1], Daniel Podlesny[1], Marija Kulis[1], Cynthia Maddox[1,3,4], Cesar Arze[1,3,5], Christian Zörb[2], Nikolaus Merkt[2] & W. Florian Fricke[1,3]✉

Wine is a globally produced, marketed and consumed alcoholic beverage, which is valued for its aromatic and qualitative complexity and variation. These properties are partially attributable to the bacterial involvement in the fermentation process. However, the organizational principles and dynamic changes of the bacterial wine microbiota remain poorly understood, especially in the context of red and white wine variations and environmental stress factors. Here, we determined relative and absolute bacterial microbiota compositions from six distinct cultivars during the first week of fermentation by quantitative and qualitative 16S rRNA gene amplification and amplicon sequencing. All wines harboured complex and variable bacterial communities, with *Tatumella* as the most abundant genus across all batches, but red wines were characterized by higher bacterial diversity and increased relative and absolute abundance of lactic and acetic acid bacteria (LAB/AAB) and bacterial taxa of predicted environmental origin. Microbial diversity was positively correlated with plant-derived DNA concentrations in the wine and *Botrytis cinerea* infection before harvest. Our findings suggest that exogenous factors, such as procedural differences between red and white wine production and environmental stress on grape integrity, can increase bacterial diversity and specific bacterial taxa in wine, with potential consequences for wine quality and aroma.

Wine is a popular alcoholic beverage, which is cherished for its versatile aroma and complexity worldwide. Although it is globally produced and marketed, regional wine varieties include prominent, often historic and legally protected, geographic pedigrees and appellations. While specific wine "terroirs" or phenotypic characteristics have been associated with quantifiable molecular markers, such as chemical and metabolite profiles[1,2] and sensory attributes[3–5], many of the underlying mechanisms for the development of colour, aroma and flavour variations remain poorly understood. The most important intrinsic and extrinsic factors that have been identified include grape-specific differences in secondary microbial metabolite diversity and composition; soil, weather, and climate; geological conditions and environmental stress factors; viticulture and the winemaking process itself[6–8].

As wine colour, aroma and flavour are substantially affected by microbial fermentation of the grape must, the taxonomic composition and functional repertoire of the wine microbiota, as well as its dependence on environmental influences, are of great interest[9]. Besides eukaryotic yeasts as the drivers of alcoholic fermentation, bacteria are known to contribute to malolactic acid fermentation and other metabolic processes[10,11]. The diversity, composition and biogeography of the fungal and bacterial microbiota of wine has been illustrated by

[1]Department of Microbiome Research and Applied Bioinformatics, Institute of Nutritional Sciences, University of Hohenheim, Stuttgart, Germany. [2]Department of Plant Quality and Viticulture, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany. [3]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. [4]Present address: Personal Genome Diagnostics, Baltimore, MD, USA. [5]Present address: Ring Therapeutics, Cambridge, MA, USA. ✉email: w.florian.fricke@uni-hohenheim.de

	Sampling schedule (Days/hours)												*B. cinerea* (%)[a]	Total acid (g/L)	
Day	**0**	**1**			**2**			**3**	**4**	**5**	**6**	**7**	**0**	**0**	**7**
Hour	**0[b]**	**2**	**8**	**15**	**23**	**31**	**39**	**50[b]**	**73**	**97**	**121**	**145[b]**	**0**	**0**	**145**
Red wines															
Regent/REG	x	x	x	x	x	x	x	x	x	x	x	x	5	7.5	7.3
Schwarzriesling/SCH	x	x	x	x	x	x	x	x	x	x	x	x	30	8.8	7.6
White wines															
Merzling/MER	x	x	x	x	x	x	x	x	x	x	x	x	5	7.7	6.7
Seyval/SEY	x	x	x	x	x	x	x	x	x	x	x	x	30	7.3	6.5
Helios/HEL	x	x	x	x	x	x	x	x	x	x	x	x	40	9.6	8.1
Bacchus 1/BAC1	x	x	x	x	x	x	x	x	x	x	x	x	15	6.7	6.1
Bacchus 2/BAC2	x	x	x	x	x	x	x	x	x	x	x	x	15	6.4	5.9

Table 1. Sample and metadata overview. The first sample of each wine was collected immediately before inoculation with the yeast starter culture. BAC2 grapes were treated with combi-protec with insecticide SpinTor (Belchim Crop Protection). [a]Infectious burden as percentage of affected grapes. [b]Samples selected for qPCR.

several recent cultivation-independent, high-throughput amplicon sequencing studies[11–17]. Differences in wine microbiota profiles have been associated with the grapevine, including distinct cultivars[14,15,18], plant organs[17,19], and vintages[14]. Environmental sources of microbes in wine have been identified as plant leaves and roots[17,19], soil[15,17,20], and the winery setting[21]. Not surprisingly, the wine microbiota is affected by the fermentation type, i.e. spontaneous or natural fermentation as opposed to inoculated fermentation with yeast starter cultures[22,23], and fungal infections, including sour rot and Botrytis bunch rot[23,24]. Interestingly, bacterial microbiota differences between red and white wines have received limited recognition, despite distinct production processes and well-characterized metabolic and aromatic profile differences. For white wine production, red or white grapes are crushed and pressed and only the clarified juice is used for fermentation, whereas a mixture of juice with skin and seeds from crushed red grapes is used for red wine fermentation[25]. As a result, red wines are characterized by increased concentrations of secondary plant metabolites, such as flavonoids, including anthocyanins, flavonols and flavanols; non-flavonoids, including hydroxybenzoic and hydroxycinnamic acids and stilbenes; and other phenolic compounds, many of which have been associated with health-promoting benefits[26,27].

In order to improve our understanding of the composition, organization and temporal dynamics of the red and white wine bacterial microbiota, we determined relative and absolute microbiota compositions from six distinct cultivars during the first week of fermentation by 16S rRNA gene amplification and amplicon sequencing. All wines were found to harbour complex bacterial communities with substantial variations between red and white wines, distinct cultivars and even separate fermentations from the same cultivar. Increases or decreases in the relative abundance of specific bacterial taxa during the fermentation were associated with changes in total bacterial concentrations and the observed differences between time points, wine types and cultivars were most strongly correlated with microbial diversity. Variations in diversity could be attributed to plant-derived DNA contributions and the influence of environmental stress factors, such as *Botrytis cinerea* fungal burdens of grapes before harvest. Our findings point to exogenous factors contributing to bacterial microbiota diversity in wine with both potentially desirable and undesirable consequences for wine quality and aroma.

Results

Wine sampling and microbiota analysis overview. Metagenomic DNA was isolated from seven batches of fermenting wines, including two red wines (Regent/REG, Schwarzriesling/SCH) and four white wines (Helios/HEL, Merzling/MER, Seyval/SEY, Bacchus/BAC) (Table 1). Two independent BAC fermenting batches were included, BAC1 from regular grapevines as opposed to BAC2, which had been treated with a commercial combination of insect attractant and insecticide (Combi-protec, Belchim; BAC2). All grapes were grown on the same vineyard, harvested within three weeks, processed and fermented in close proximity at the same winery at the viticulture unit of the University of Hohenheim, Stuttgart, Germany. Available metadata for all wines included metabolic differences between batches at the beginning and/or end of the fermentation period of 14 days (pH, total acid, alcohol and sugar content), changes during fermentation (°Brix/must weight) and rates of infection of grapes with the fungal plant pathogen *Botrytis cinerea* at the time of harvest (Supplementary Table S1). Longitudinal samples were collected 3–4 times daily for the first three days and once per day for each remaining day during the first week of fermentation. (Table 1). The first sample was collected within two hours after grape pressing and before inoculation with commercial yeast starter cultures (Supplementary Table S1). Bacterial taxonomic microbiota compositions were determined by 16S rRNA gene amplicon sequencing, resulting in 5.38 million taxonomically assigned sequences from a total of 84 samples. Of these, ~ 58% were classified as plant-derived reads, i.e. chloroplast and mitochondrial sequences, which showed grape vine (*Vitis vinifera*) and rootstock (*Vitis riparia*) as the closest matches in public databases (Supplementary Table S2). For most analyses, unless indicated otherwise, plant-derived reads were removed, resulting in a dataset of 2.21 million sequences. After rarefaction to 3,500 reads per sample, 520 distinct bacterial species equivalent or operational taxonomic units (OTUs) were identified. Quantitative 16S rRNA gene amplifications were carried out at three

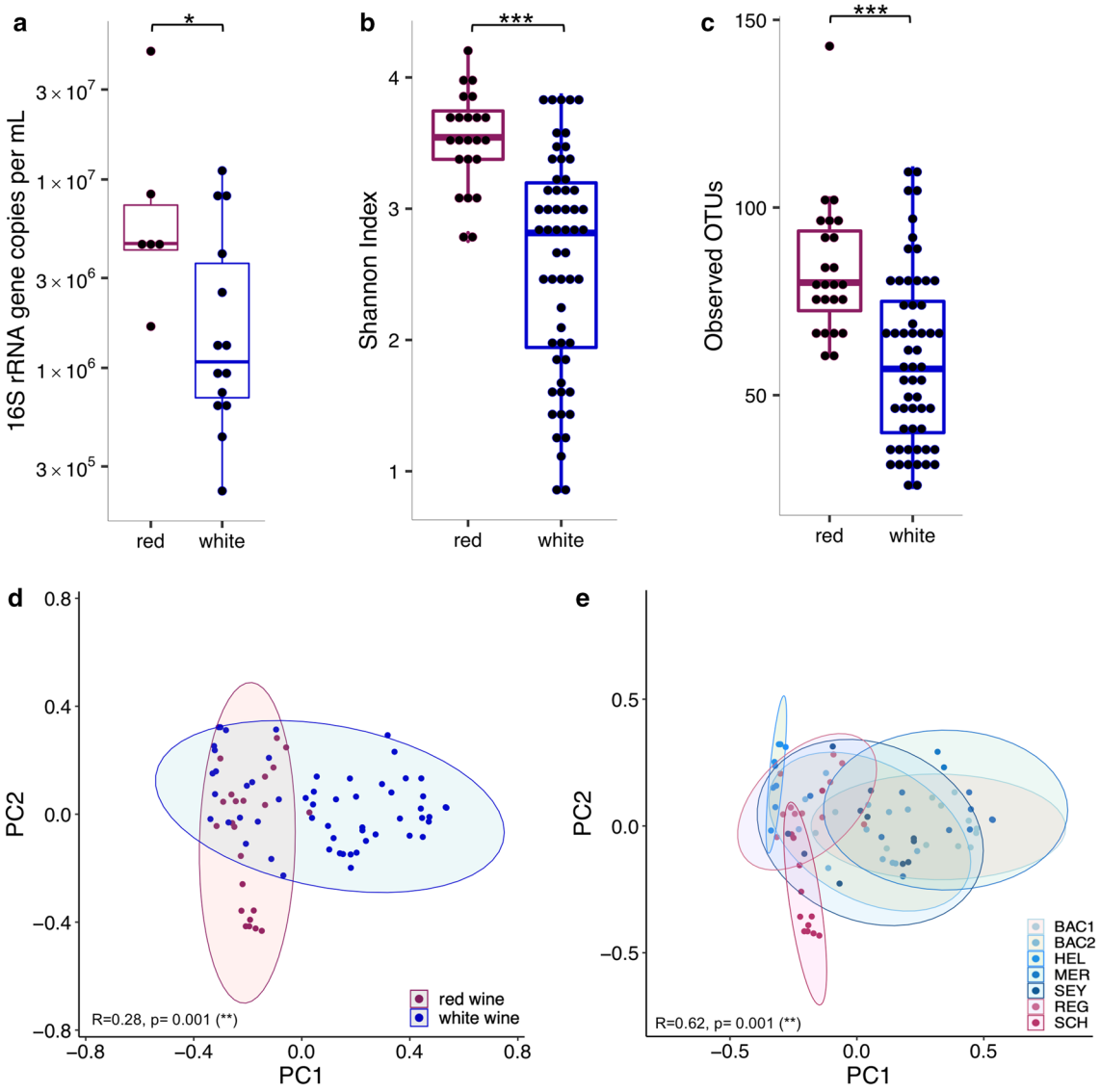


Figure 1. Red and white wine bacterial microbiota comparison. Red and white wine samples across all time points were compared based on 16S rRNA gene copy numbers per mL (**a**), and bacterial microbiota diversity, based on Shannon Index (**b**) and observed OTUs (**c**), as well as taxonomic composition, based on Bray–Curtis dissimilarity (**d**). Taxonomic distances were also compared for all red and white wines separately (**e**). Significance was calculated based on Wilcoxon rank-sum test, corrected with the Benjamini–Hochberg procedure (**a**–**c**) and ANOSIM with 999 permutations (**d**, e) with *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Sample numbers: $n_{red\ wine} = 23$, $n_{white\ wine} = 56$, $n_{BAC1} = 12$, $n_{BAC2} = 12$, $n_{HEL} = 9$, $n_{MER} = 12$, $n_{SEY} = 11$, $n_{SCH} = 11$ and $n_{REG} = 11$.

time points, on days 0, 3 and 7, and used to determine absolute bacterial abundances or bacterial loads, i.e. 16S rRNA gene copies per millilitre of wine. Summary statistics of sequencing output, taxonomic compositions and quantitative analyses are shown in Supplementary Tables S3 and S4.

Differences in bacterial microbiota concentration, diversity and composition between red and white wines. Compared to white wines, red wines harboured increased bacterial loads of larger taxonomic diversity, as determined based on higher concentrations of bacterial 16S rRNA genes (Fig. 1a), increased alpha-diversity (Fig. 1b) and more observed OTUs (Fig. 1c). Taxonomic bacterial microbiota compositions showed minor but significant differences between red and white wines (Fig. 1d), as well as between separate red and white wine batches (Fig. 1e).

The genus *Tatumella* from the phylum *Proteobacteria* was the most abundant bacterial taxon across all samples, both in terms of relative (26 ± 3%) and absolute abundance ($10^6 \pm 5*10^5$ 16S rRNA gene copies/mL). The latter was calculated as the taxon-specific fraction of the total bacterial 16S rRNA gene copy number per sample. While the relative abundance of *Tatumella* was higher in white compared to red wines (Fig. 2a), there was no difference in absolute abundance between both wine types (Fig. 2b), suggesting higher loads of additional other bacteria in red wines as the source of reduced relative abundances of *Tatumella*, which demonstrates the utility

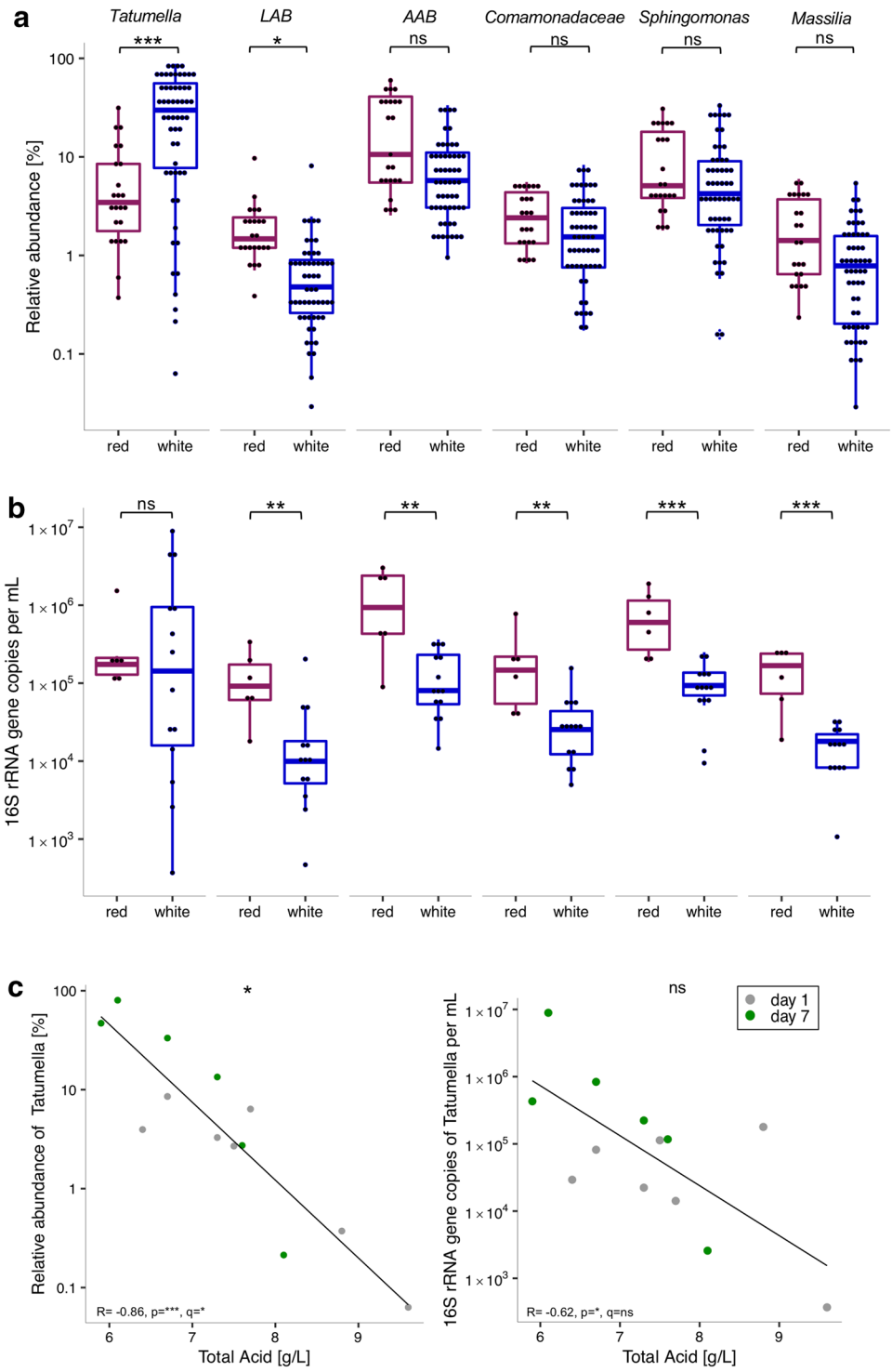


Figure 2. Bacterial groups with differential relative (**a**) and absolute (**b**) abundance in red and white wines and association with total acid content in wine (**c**). The relative abundances, based on 16S rRNA gene amplicon sequencing (**a**) and absolute abundances, based on 16S rRNA gene copy number concentrations (**b**) in red and white wines were compared for the proteobacterial genus *Tatumella*, lactic acid bacteria (LAB), acetic acid bacteria (AAB) and the proteobacterial family *Comamonadaceae* and genera *Sphingomonas* and *Massilia*. Relative and absolute abundance of *Tatumella* and total acid content (g/L) were negatively correlated, based on available metadata (**c**). Significance was calculated based on ALDEx analysis with 128 DMCs (**a**, **b**) and Spearman's rank correlation test (**c**), all corrected with the Benjamini–Hochberg procedure, with ns = not significant, *p/q < 0.05, **p/q < 0.01, ***p/q < 0.001. Effect sizes for ALDEx analyses in (**a**): *Tatumella* = 0.87, LAB = − 0.47, AAB = − 0.14, *Comamonadaceae* = 0.21 *Sphingomonas* = 0.21, *Massilia* = − 0.11.

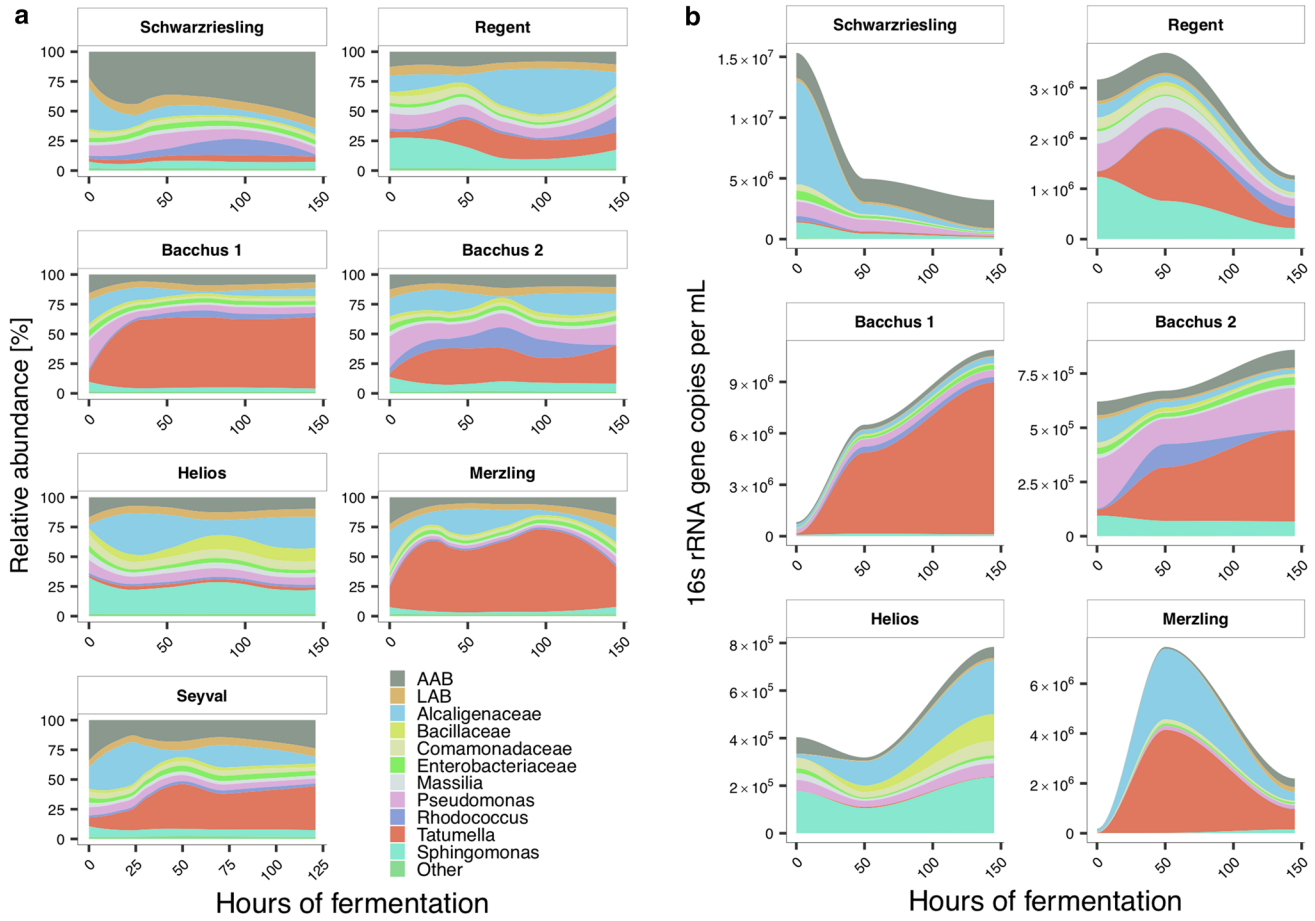


Figure 3. Relative (**a**) and absolute (**b**) bacterial microbiota composition changes during fermentation. For the visualization an abundance threshold was set of either ≥ 1% in all samples or ≥ 5% in at least one sample. The eleven most abundant assigned taxa on the genus level are shown in the figure. Locally weighted regression was used to smooth relative abundances over time.

of comparing both metrics for microbiota analysis. *Tatumella* relative and absolute abundance were negatively correlated with total acid content (Fig. 2c), which was the only significant association between individual bacterial taxa and the available metabolic metadata after correcting for multiple testing.

The groups of acetic acid bacteria (AAB) and lactic acid bacteria (LAB) are often used to refer to bacteria that can negatively and positively affect the wine fermentation process, aroma and quality[10,11]. For our analysis, we combined all OTUs from those genera typically assigned to AAB (i.e. *Acetobacter*, *Gluconacetobacter*, and *Gluconobacter*) and LAB (*Lactobacillus*, *Oenococcus*, and *Lactococcus*) and found LAB to be overrepresented in red compared to white wines, both in terms of relative (Fig. 2a) and absolute abundance (Fig. 2b). Increased bacterial loads in red wine were also found for AAB and the family *Comamonadaceae*, the genera *Sphingomonas* and *Massilia* (abbreviated from here on as CSM, Fig. 2b).

Changes of the bacterial wine microbiota during fermentation. While fermentation was generally associated with alterations in relative and absolute bacterial microbiota compositions in both red and white wines, changes varied between individual wine batches (Fig. 3a,b). Differences in taxonomic microbiota compositions were smaller between longitudinally collected samples from the same wine batch than between samples from distinct wine types or batches (Fig. 4a; Wilcoxon rank-sum test on Bray–Curtis dissimilarity). There was a trend towards a positive correlation of bacterial loads in white wines with the duration of the fermentation (Fig. 4b; $p < 0.05$, $q = ns$) and while both the relative and absolute abundance of *Tatumella* increased in all white wines during fermentation, differences were not significant (Supplementary Fig. S1a). Increased absolute abundances of the family *Bacillaceae* and the genus *Oenococcus* by the end of the fermentation period were the only significant changes observed across all red and white wines (Supplementary Fig. S1a), suggesting that pre-fermentation factors have a stronger impact on the bacterial wine microbiota than fermentation itself.

Association of differences in microbial diversity with plant-derived read fractions and environmental stress factors. Bacterial microbiota diversity differed between red and white wines (Fig. 1b) and considerably varied among white wines (Shannon diversity: 0.8–3.9), indicating that diversity was influenced by

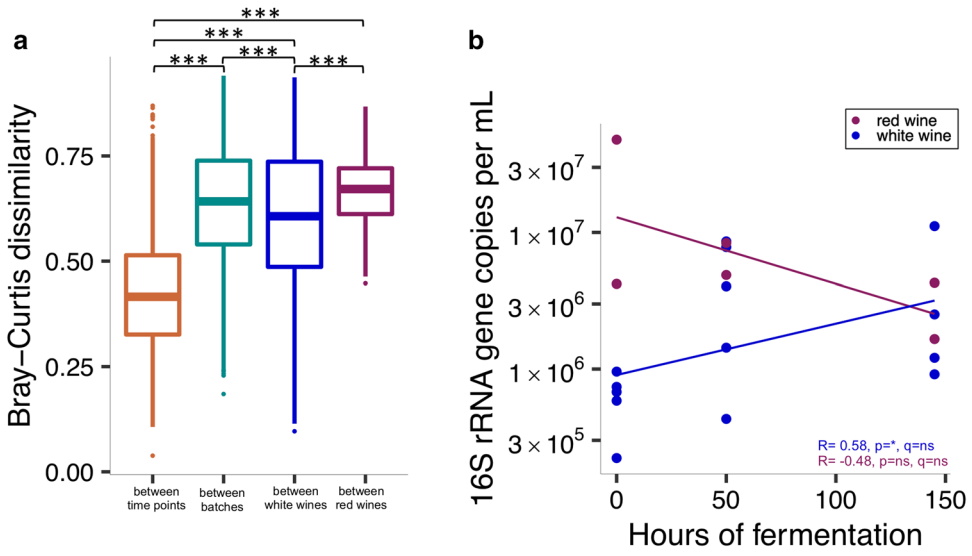


Figure 4. Bacterial microbiota changes during fermentation. Taxonomic microbiota compositions were compared between samples, based on Bray–Curtis dissimilarity, showing that samples from the same wine were more similar to each other and that red wine samples were more heterogeneous than white wine samples (**a**). Bacterial abundances of red and white wines showed opposing trends during fermentation, based on 16S rRNA gene copy numbers (**b**). Significance was calculated based on Wilcoxon rank-sum test (**a**) and Spearman's rank correlation test (**b**), all corrected with the Benjamini–Hochberg procedure, with ns = not significant, *p/q < 0.05, ***p/q < 0.001.

additional factors, independently of the wine type. Indeed, across all wine samples, alpha diversity was negatively correlated with the genus *Tatumella* and positively correlated with AAB, LAB, and CSM, including relative and absolute abundances (Fig. 5a,b).

We hypothesized that differences between red and white wines that were involved in the processing of grapes, i.e. an additional mechanical destemming step for red but not white wine grapes, and in the fermentation process, i.e. a prolonged exposure of red wines to grape skins during fermentation, could increase microbial contributions from skin-attached environmental bacteria and result in higher bacterial microbiota diversity. We therefore compared plant-derived 16S rRNA amplicon read fractions from mitochondria and chloroplasts across all wine batches and time points (Supplementary Fig. S1b), as a potential marker for contact with grape skins and other plant tissues. As expected, we found higher relative and absolute contributions of plant-derived DNA in red compared to white wines (Fig. 6a). While there was no correlation between plant-derived read fractions and fermentation time (p = ns), the relative abundance of plant-derived reads varied substantially among white wines (Fig. 6a) and showed a positive association with bacterial microbiota diversity across red and white wine samples combined (Supplementary Fig. S2a). In addition, plant-derived read fractions were negatively correlated with the relative but not absolute abundance of *Tatumella* and positively correlated with the relative but not absolute abundances of AAB, LAB, and CSM (Fig. 6b and Supplementary Fig. S2c). The fact that both microbial diversity and plant-derived read fractions showed comparable associations with the same bacterial taxa could suggest that the mechanism that led to increased plant-derived reads was also responsible for increased microbial diversity. However, as correlations were stronger for microbial diversity, additional factors besides those that increased plant-derived reads would need to have contributed to microbial diversity in our wine.

To account for putative environmental influences on the bacterial wine microbiota, we compared wines based on infectious burdens of *B. cinerea* (Table 1). The rate of *B. cinerea* infection was positively correlated with microbial diversity in white wines but there was no significant correlation in red wines (Supplemental Fig. S2b). Again, as with plant-derived read fractions, the relative abundance of *Tatumella* was negatively and with CSM positively correlated with botrytis infection in white wines (Supplemental Fig. S3). Despite similar associations, botrytis infection rates and plant-derived read fractions were not significantly correlated with each other (data not shown), suggesting that they reflect independent mechanisms for increased bacterial microbiota diversity, reduced relative abundance of *Tatumella* and increased relative abundance of CSM in wine.

Of note, BAC2 grapevines, which had been experimentally treated with a combination of insect attractant and insecticide, showed a higher infestation with the insect fruit pest *Drosophila suzukii* than untreated BAC1 grapevines. We therefore compared wines from treated BAC2 and untreated BAC1 grapevines and found BAC2 samples to be characterized by increased plant-derived read fractions, as well as similar bacterial microbiota alterations as had been associated with botrytis infection, i.e. reduced relative abundance of *Tatumella* and increased relative abundances of *Sphingomonas* and *Massilia* (Supplemental Fig. S4).

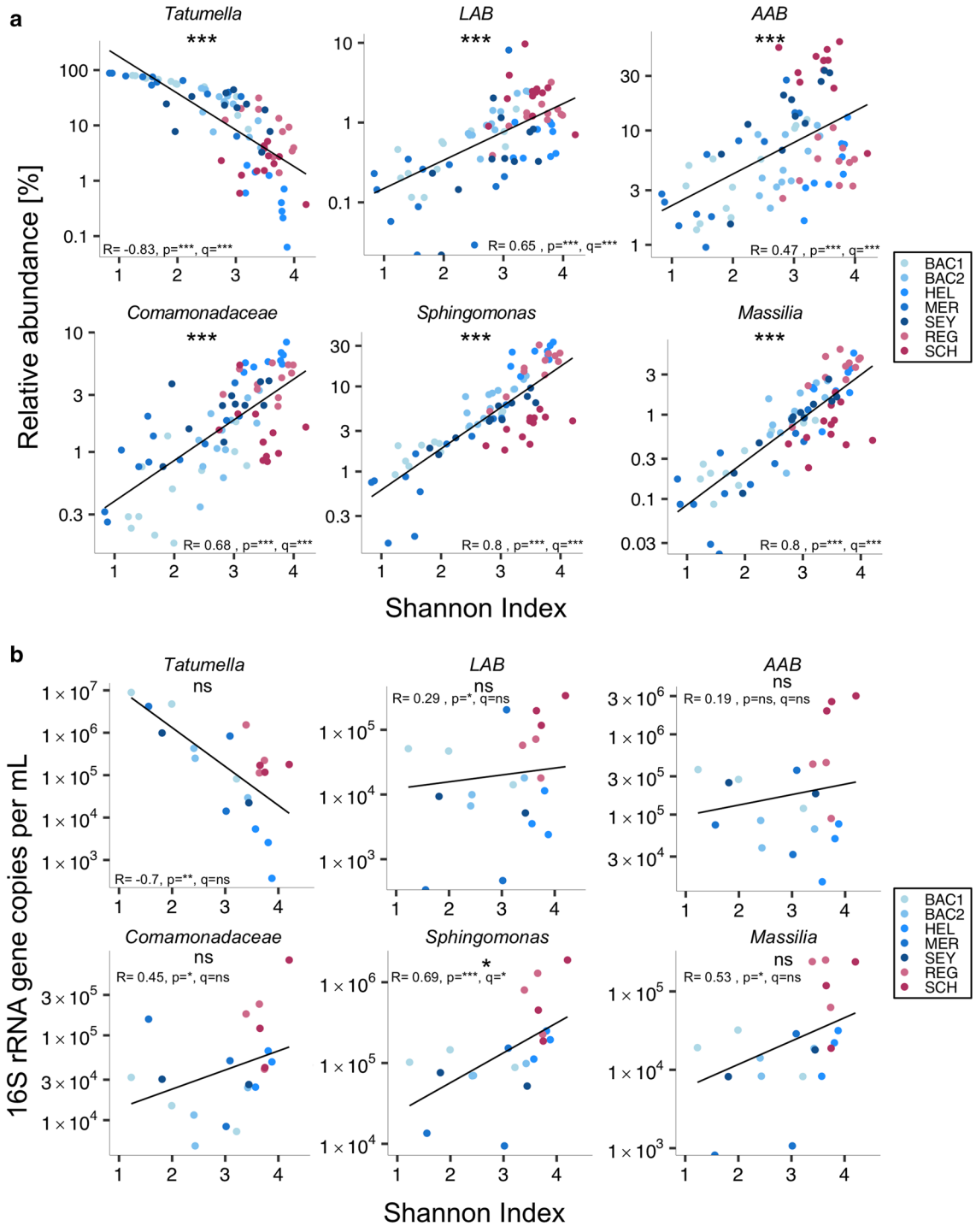


Figure 5. Relative and absolute abundances of *Tatumella*, acetic acid (AAB) and lactic acid bacteria (LAB) and other taxa were correlated with microbial diversity. Significance was calculated based on Spearman's rank correlation test (**a**, **b**) corrected with the Benjamini–Hochberg procedure, with ns = not significant, *p/q < 0.05, ***p/q < 0.001.

Discussion

Our study supports previous 16S rRNA gene amplicon sequencing-based reports of a complex and shifting bacterial microbiota in wine[11–15,17], which we have expanded with complementary quantitative 16S rRNA gene abundance analysis. We show that red wine harbours a more diverse bacterial microbiota with a higher density of bacterial 16S rRNA genes per millilitre than white wine. The chromosomal copy number of the 16S rRNA gene is known to vary substantially between taxa[28], but assuming on average four 16S rRNA genes per bacterial genome, we would expect red wines from our sample set to contain ~ 10^7 bacterial genomes per mL or about

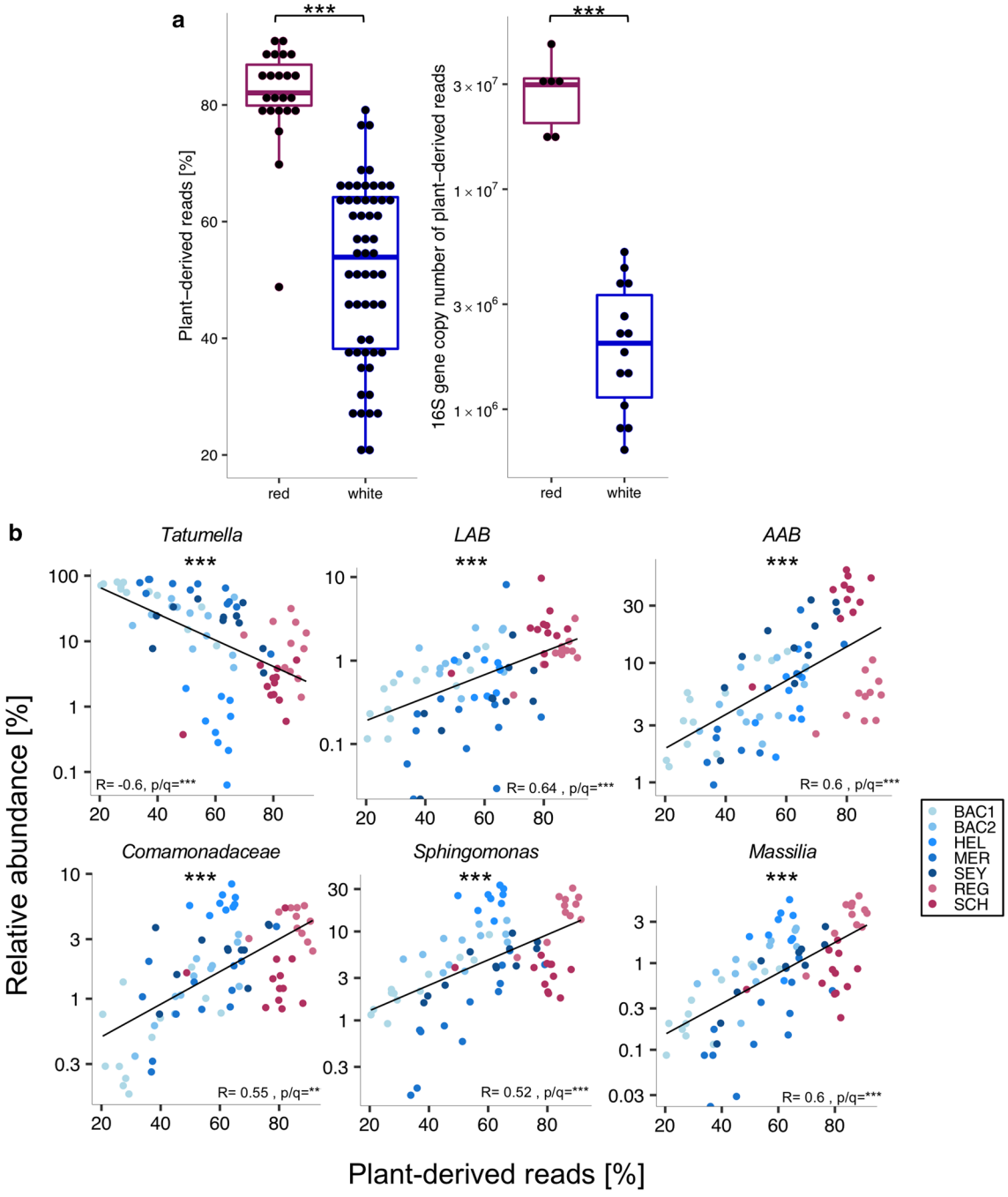


Figure 6. Plant-derived read fractions were correlated with the relative abundance of *Tatumella*, AAB, LAB and other taxa in wine. Significance was calculated based on Wilcoxon rank-sum test (**a**) and Spearman's rank correlation test (**b**), all corrected with the Benjamini–Hochberg procedure, with ns = not significant, ***p/q < 0.001.

tenfold more genomes than white wines. While not significant after correcting for multiple tests ($q > 0.05$), there was a clear trend towards increasing 16S rRNA gene copy numbers during white wine fermentation ($p < 0.05$), which resulted in comparable densities in red and white wines by the end of the one-week observation period. However, in our longitudinal analysis, fermentation time was associated with only moderate bacterial microbiota changes, as few bacterial taxa increased or decreased during fermentation and compositional differences between time points were small compared to those seen between wine types and batches. Calculating the fraction of taxon-specific 16S rRNA genes based on relative abundance values from the amplicon sequence analysis, we show that increases in overall 16S rRNA gene copy number density are mostly due to increased abundance of *Tatumella*, suggestive of active proliferation of this bacterial genus. Species from the genus *Tatumella* have been associated with fruits, including pineapple and mandarin orange, but also soil samples and even human patient

specimens[28,29]. Whereas relative microbiota profiling alone would suggest reduced contributions of *Tatumella* to the red wine microbiota, the combination with quantitative microbiota profiling demonstrated comparable 16S rRNA gene densities for this genus, suggesting instead higher overall bacterial densities and additional contributions of other bacteria to the red wine microbiota. We identified three bacterial groups with higher absolute abundance in red compared to white wines, which are likely main contributors to the increased bacterial loads of red wines: acetic acid bacteria (AAB, from the genera *Acetobacter*, *Gluconacetobacter*, and *Gluconobacter*), lactic acid bacteria (from the genera *Lactobacillus*, *Oenococcus*, and *Lactococcus*) and a third group bacteria (CSM, including the family *Comamonadaceae* and the genera *Sphingomonas* and *Massilia*). The Gram-positive AAB are frequently found in wine but undesirable for wine production due their ability to efficiently convert ethanol to acetate, which is associated with spoilage, alters the wine aroma, and reduces its commercial value[30]. Grape skins have been suggested as a major source for AAB in wine, consistent with findings of reduced AAB abundance on berries washed by rain[31]. The LAB are physiologically more diverse and include *Oenococcus oeni*, which is typically responsible for the favourable malolactic fermentation[32], but also *Lactobacillus* and *Pediococcus* species that can cause additional spontaneous fermentations with undesirable aromatic consequences[10,33]. The generally lower concentrations of LAB in wine, consistent in our samples, have been attributed to the mostly anaerobic lifestyles, suggesting competitive advantages for yeasts and AABs under the aerobic conditions of the grape[10]. *Sphingomonas* and *Massilia* species have been identified in rhizosphere and soil microbiomes[34,35] and correspondingly, are frequently isolated from soil and water samples[36,37]. Thus, at least for AAB and bacteria from the CSM group an exogenous origin in wine is likely, e.g. through direct intake of grape skin-attached bacteria or through indirect environmental sampling of soil or water-associated bacteria that come into contact with grape skins. Martins et al. described the presence of a culturable, epiphytic bacterial microbiota on external grape berry and leaf surfaces, which overlapped with the CSM group from our study[38], which could explain the increased relative and absolute contributions of these bacteria to the red wine microbiota.

As bacterial diversity and abundance of AAB, LAB and CSM in wine differed not only between red and white wines, but also among different grapevine cultivars and wine batches, we searched for additional factors and mechanisms that could explain variable exogenous contributions to the bacterial microbiota in wine. While we initially removed plant-derived 16S rRNA gene amplicon read fractions from chloroplasts and mitochondria as contaminants from our analysis, we later hypothesized that plant-derived read fractions could also represent biomarkers for wine exposure to plant-associated bacteria, particularly from grape skins. In line with this assumption, we found increased plant-derived read fractions in red wines and robust positive correlations with microbial diversity and the abundance of bacteria from putative exogenous sources, i.e. AAB, LAB and CSM, across all red and white wine samples. As the microbial ecology of grapes, including microbial burdens and species diversity is largely affected by the grape health status[10], we also searched for factors associated with differences in plant-derived DNA contributions among white wines. We assumed that, by increasing interactions of grape juice with exogenous, skin-attached bacteria, interference with berry integrity before harvest could induce similar effects in white wines as extended grape skin contact during fermentation in red wines. Our collection of wines was exposed to two environmental stress factors with potential disruptive effects on grape skin integrity before harvest: (I) grapevines were differentially affected by infections with the fungal pathogen *Botrytis cinerea* and (II) as a consequence of the experimental application of an insect attractant/insecticide combination, Bacchus grapevines on adjacent sections of the vineyard were differentially exposed to infestations with the spotted fruit fly, *Drosophila suzukii*. *B. cinerea*, the causative agent of botrytis bunch rot in viticulture colonizes different plant organs and can penetrate the protective cuticle covering the grape epidermis[39–41]. Fermentations of botrytis-infected grapes have been shown to be enriched for bacteria and fungi and to contain increased microbial diversity and acetic acid bacteria concentrations[23], in line with the higher bacterial microbiota diversity and abundance of AABs in those white wines from our collection that had a higher rate of botrytis infection. Similarly, Bokulich et al. showed that botrytized wine, i.e. fermentations from *B. cinerea*-infected but then dried, partially raisined grapes affected by "noble rot", were characterized by high bacterial diversity[13]. *D. suzukii* could induce comparable bacterial microbiota effects, as these insects puncture the grape skin for oviposition[42] and *D. suzukii* exposure has previously been associated with increased bacterial loads, most importantly of AABs[43], in accordance with our findings from the direct comparison of the two Bacchus batches.

Our detailed qualitative and quantitative wine microbiota analysis supports the presence of a diverse bacterial microbiota in wine, which appears to be shaped by both endogenous and exogenous factors. On the one hand, comparable absolute abundance of the genus *Tatumella* in red and white wines, with a trend towards increasing density in white wines during fermentation, suggests a putative endogenous, grape juice-derived source of these bacteria in all wines, largely unaffected by external factors such as wine processing procedures and pathogen burdens before grape harvest. Robust positive associations of AAB, LAB and CSM with markers of plant tissue contributions and pre-harvest pathogen burdens, on the other hand, suggest a putative exogenous, grape skin-derived source of these bacteria in all wines, resulting in increased loads in red wines and wines affected by *B. cinerae* or *D. suzukii*. Additional studies will be needed to further our understanding of endogenous and exogenous contributions to the wine microbiota, including larger sample sets that span a wider variety of wine types, cultivars, and environmental conditions.

The specific metabolic contributions of those bacteria classified here as of putative exogenous origin to the extended secondary plant metabolite spectrum in red wines remain largely unknown. However, it is conceivable that exogenous factors, ascribed to plant tissue contributions, botrytis and fruit fly infestation in our analysis, could be leveraged to deliberately increase, modify, or expand the aromatic quality of wine. In fact, botrytized white wines would represent an example, as well as the production of "orange", or skin-contact, white wines that are characterized by increased phenolic concentrations of antioxidant potential[44–47].

Methods

Grapes, metadata and sample collection. All grapevines, including the two red wine cultivars Regent (REG) and Schwarzriesling (SCH) and the four white wine cultivars Merzling (MER), Seyval (SEY), Helios (HEL), and Bacchus (BAC) were grown on the vineyards of the University of Hohenheim, Stuttgart, Germany in 2015. Bacchus grapevines were divided into two separate groups: plants from the first group did not receive special treatment (BAC1), whereas plants from the second (BAC2) underwent treatment with a combination of insect attractant and insecticide (Combi-protec with insecticide SpinTor, Belchim Crop Protection, Belgium), which was associated with higher infestation of spotted wing drosophila, *Drosophila suzukii*, in the BAC2 grapevine. All grapes were harvested within three weeks (September/October 2015). Within two hours after collection grapes were processed for red and white wine production, with grapes intended for red wine production undergoing a mechanical destemming before crushing. The first samples were collected before and after the grape must was inoculated with a commercial *Saccharomyces cerevisiae* yeast starter culture (NT2000/NT50/NT112, Oenobrands SAS, France). Additional samples were collected three times per day for the first three days and daily for the remaining days during the first week of fermentation. All samples were immediately stored at − 80 °C until further processing.

Oenological parameters. Oenological parameters were measured as previously described[48]. Total soluble solids (°Brix) of wine were determined using a refractometer (Opton, Zeiss, Germany) and total acid and pH by titration (TiroLine easy, Schott, Germany). High performance liquid chromatography (Merck-Hitachi, Germany) was carried out to determine wine alcohol contents (flow rate: 0.5 mL/min; detection at 210 nm), using a Rezex ROA-Organic Acid H + (8%), LC Column 300 × 7.8 mm, Ea (Phenomenex, Germany) in combination with a SecurityGuard Cartridge, Carbo-H 4 × 3.0 mm pre-column (Phenomenex, Germany).

DNA extraction. Metagenomic DNA of all samples was isolated using a previously described method[49] from our group, which combines both enzymatic digestion and mechanical disruption by bead beating. In brief, aliquots of 500 µL per sample were thawed on ice and centrifuged for 15 min at 4,000×*g*. The pellet was washed in 1 mL PBS, centrifuged for 5 min at 8,000×*g*, resuspended in 700 µL PBS and transferred to a Lysis B Matrix tube (MP Biochemicals, France) for bead beating. Enzymatic cell lysis (lysozyme, mutanolysin, lysostaphin, proteinase K and RNase) was initiated as described in the method above. The resulting cell lysate was processed with the ZR Fecal DNA mini-prep kit (Zymo Research, USA) according to the manufacturer's recommendation and eluted in ultra-pure water. As controls, blank DNA extractions and extractions from the yeast starter cultures were included and further processed along with the wine samples. The DNA was stored at -20 °C until further processing.

16S rRNA gene amplification and sequencing. Hypervariable region V4 of the 16S rRNA gene was amplified from metagenomic DNA via PCR using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, USA) and Golay-barcoded primers 515F and 806R adapted from Caporaso et al.[50] and additionally modified by adding internal spacers of 0 to 7 bp, adapted from Fadrosh et al.[51]. Primer, barcode and spacer sequences are listed in Supplementary Table S5. The PCR reaction contained 10 µL of 2 × Phusion Master Mix (Thermo Fisher Scientific, USA), 0.8 µL of each primer (final concentration 0.4 µM), 0.6 µL dimethyl sulfoxide (DMSO), and 7.8µL template DNA. PCR amplifications were carried out as described previously[52], with an initialization step at 98 °C for 2 min, followed by 30 cycles of 98 °C for 10 s, 52 °C for 15 s and 72 °C for 15 s, and a final extension at 72 °C for 5 min. The SequalPrep normalization plate kit 96 (Thermo Fisher Scientific, USA) was used to select equimolar PCR product amounts, which were subsequently pooled and concentrated with the DNA Clean and Concentrator 5 kit (Zymo Research, USA). Sequencing libraries were generated with the TruSeq Nano DNA LT Library Prep kit (Illumina, USA) for sequencing on the Illumina MiSeq platform (MiSeq Reagent Kit v3, 600 cycles, Illumina, USA) at the University of Hohenheim, following the manufacturer's recommendations.

Quantitative 16S rRNA gene amplification. Metagenomic DNA was diluted to a concentration of approximately 1 ng/µL, of which 2 µL were used as template for quantitative PCR amplification of the universal bacterial 16S rRNA gene with the Femto bacterial DNA quantification kit (Zymo Research, USA) according to the manufacturer's recommendations. All reactions were carried out in duplicates. Genomic DNA from *E. coli* strain JM109 (Zymo Research, USA) was used as an internal standard to estimate bacterial 16S rRNA gene copy numbers. The PCR was run on a CFX96 Touch real-time detection system (Bio-Rad, USA). Samples were considered non-amplified if the quantification cycle (Cq) value was greater than 39. Samples with quantification cycle (Cq) values greater than 39 were discarded; average Cq values of duplicates were calculated for each sample and used to determine 16S rRNA gene copy numbers. Bacterial 16S rRNA gene copy numbers were calculated for each sample as concentrations per 1 mL of wine. Taxon-specific absolute abundances were determined by multiplying relative abundance values from the 16S rRNA gene amplicon sequence analysis with total 16S rRNA gene copy numbers (see Supplementary Table S3).

Microbiota analysis and statistical methods. Pre-processing of raw sequence data was performed with QIIME v1.9.1[53,54], cutadapt v1.10[55] and custom Python scripts, including trimming of spacer and primer sequences, merging of raw paired end reads using bbmerge v9.02[56], barcode extraction and demultiplexing of samples. The processed sequences were imported into QIIME2 v2018.2. OTUs were generated using open-reference OTU picking with a similarity threshold of 97% and classified with the q2-feature-classifier[57] against the Greengenes database v13_8 (greengenes.lbl.gov). Chimera checking was performed with vsearch[58], as part of

the QIIME2 pipeline. OTU's taxonomically assigned to chloroplast and mitochondria were searched against the NCBI nucleotide database by BLAST[59], which identified grapevine, *Vitis vinifera*, and its grafted rootstock, *Vitis riparia*, as closest matching sequences. These sequence reads accounted for 58% of the sequences and were removed from the analysis, unless noted otherwise. For diversity analyses, all samples were rarefied to 3,500 reads per sample. Detailed information about sequenced OTU's and taxonomic assignment per sample is listed in Supplementary Tables S6–S13. Unless indicated differently, the final dataset consisted of n = 56 white and n = 23 red wine samples for sequenced data. For qPCR data, the final set consisted of n = 14 white and n = 6 red wine samples. Statistical testing and data visualization were carried out in R (www.R-project.org/), using the packages *vegan*, *biomformat*, *nortest* and *ALDEx2*[60]. All parameters were tested for normal distribution with Anderson–Darling and Shapiro–Wilk tests. Features were filtered to only include OTUs with a relative abundance of ≥ 1% in all samples or of ≥ 5% in at least one sample. Not normally distributed data were analysed using non-parametric tests. For group comparisons, pairwise Wilcoxon rank sum test was used, for correlation analyses Spearman's rank correlation tests and for differential abundance analyses ANOVA-Like Differential Expression (ALDEx) Analysis[60] with 128 Dirichlet Monte-Carlo Instances (DMC). All tests were corrected with the Benjamini–Hochberg procedure, based on the number of features (n = 79). Comparisons of dissimilarities between communities were done by analysis of similarities (ANOSIM)[61], with the number of permutations set to 999, of the Bray–Curtis dissimilarity. Significance levels were determined as $p > 0.05$ (ns), $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). Unless indicated, mean values are presented with standard error of means (SEM). Detailed information about all bioinformatic scripts and commands used for the analysis is provided in Supplementary Table S13.

Data availability

Pre-processed, trimmed and merged paired-end read contigs have been deposited in the European Nucleotide Archive under primary accession number PRJEB37054 (secondary accession number ERP120343).

Received: 27 April 2020; Accepted: 27 July 2020
Published online: 14 August 2020

References

1. López-Rituerto, E. *et al.* Investigations of La Rioja terroir for wine production using 1H NMR metabolomics. *J. Agric. Food Chem.* **60**, 3452–3461 (2012).
2. Son, H.-S. *et al.* Metabolomic studies on geographical grapes and their wines using 1H NMR analysis coupled with multivariate statistics. *J. Agric. Food Chem.* **57**, 1481–1490 (2009).
3. Fischer, U., Roth, D. & Christmann, M. The impact of geographic origin, vintage and wine estate on sensory properties of Vitis vinifera cv Riesling wines. *Food Qual. Preference* **10**, 281–288 (1999).
4. Robinson, A. L. *et al.* Influence of geographic origin on the sensory characteristics and wine composition of *Vitis vinifera* cv. Cabernet Sauvignon Wines from Australia. *Am. J. Enol. Viticult.* **63**, 467–476 (2012).
5. Kontkanen, D., Reynolds, A. G., Cliff, M. A. & King, M. Canadian terroir: sensory characterization of Bordeaux-style red wine varieties in the Niagara Peninsula. *Food Res. Int.* **38**, 417–425 (2005).
6. González-Barreiro, C., Rial-Otero, R., Cancho-Grande, B. & Simal-Gándara, J. Wine aroma compounds in grapes: a critical review. *Crit. Rev. Food Sci. Nutr.* **55**, 202–218 (2015).
7. Styger, G., Prior, B. & Bauer, F. F. Wine flavor and aroma. *J. Ind. Microbiol. Biotechnol.* **38**, 1145–1159 (2011).
8. Robinson, A. L. *et al.* Origins of grape and wine aroma. Part 1. Chemical components and viticultural impacts. *Am. J. Enol. Viticult.* **65**, 1–24 (2014).
9. *Wine Microbiology*. (Springer US, 2007).
10. Barata, A., Malfeito-Ferreira, M. & Loureiro, V. The microbial ecology of wine grape berries. *Int. J. Food Microbiol.* **153**, 243–259 (2012).
11. Piao, H. *et al.* Insights into the bacterial community and its temporal succession during the fermentation of wine grapes. *Front. Microbiol.* **6**, 809 (2015).
12. Bokulich, N. A. *et al.* Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. *MBio* **7**, e00631 (2016).
13. Bokulich, N. A., Joseph, C. M. L., Allen, G., Benson, A. K. & Mills, D. A. Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PLoS ONE* **7**, e36357 (2012).
14. Bokulich, N. A., Thorngate, J. H., Richardson, P. M. & Mills, D. A. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proc. Natl. Acad. Sci. USA* **111**, E139–E148 (2014).
15. Mezzasalma, V. *et al.* Geographical and cultivar features differentiate grape microbiota in Northern Italy and Spain Vineyards. *Front. Microbiol.* **9**, 946 (2018).
16. Morgan, S. C. *et al.* The effect of sulfur dioxide addition at crush on the fungal and bacterial communities and the sensory attributes of Pinot gris wines. *Int. J. Food Microbiol.* **290**, 1–14 (2019).
17. Zarraonaindia, I. *et al.* The soil microbiome influences grapevine-associated microbiota. *MBio* **6**, 025767 (2015).
18. Portillo, M. C., Franquès, J., Araque, I., Reguant, C. & Bordons, A. Bacterial diversity of Grenache and Carignan grape surface from different vineyards at Priorat wine region (Catalonia, Spain). *Int. J. Food Microbiol.* **219**, 56–63 (2016).
19. Marasco, R., Rolli, E., Fusi, M., Michoud, G. & Daffonchio, D. Grapevine rootstocks shape underground bacterial microbiome and networking but not potential functionality. *Microbiome* **6**, 3 (2018).
20. Berg, G. & Smalla, K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol.* **68**, 1–13 (2009).
21. Bokulich, N. A., Ohta, M., Richardson, P. M. & Mills, D. A. Monitoring seasonal changes in winery-resident microbiota. *PLoS ONE* **8**, e66437 (2013).
22. Stefanini, I. *et al.* Dynamic changes in microbiota and mycobiota during spontaneous 'Vino Santo Trentino' fermentation. *Microb. Biotechnol.* **9**, 195–208 (2016).
23. Lleixà, J., Kioroglou, D., Mas, A. & Portillo, M. D. C. Microbiome dynamics during spontaneous fermentations of sound grapes in comparison with sour rot and Botrytis infected grapes. *Int. J. Food Microbiol.* **281**, 36–46 (2018).
24. Nisiotou, A. A., Rantsiou, K., Iliopoulos, V., Cocolin, L. & Nychas, G.-J.E. Bacterial species associated with sound and Botrytis-infected grapes from a Greek vineyard. *Int. J. Food Microbiol.* **145**, 432–436 (2011).
25. Amerine, M. A. & Berg, H. W. *The Technology of Wine Making* (Avi Pub Co, West Port, 1980).
26. German, J. B. & Walzem, R. L. The health benefits of wine. *Annu. Rev. Nutr.* **20**, 561–593 (2000).

27. Jezek, M., Zörb, C., Merkt, N. & Geilfus, C.-M. Anthocyanin management in fruits by fertilization. *J. Agric. Food Chem.* **66**, 753–764 (2018).
28. Sun, D.-L., Jiang, X., Wu, Q. L. & Zhou, N.-Y. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* **79**, 5962–5969 (2013).
29. Brady, C. L. *et al.* Transfer of Pantoea citrea, Pantoea punctata and *Pantoea terrea* to the genus *Tatumella emend* as *Tatumella citrea* comb.. nov. *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov. *Int. J. Syst. Evol. Microbiol.* **60**, 484–494 (2010).
30. Bartowsky, E. J. & Henschke, P. A. Acetic acid bacteria spoilage of bottled red wine: a review. *Int. J. Food Microbiol.* **125**, 60–70 (2008).
31. González, A., Hierro, N., Poblet, M., Mas, A. & Guillamón, J. M. Application of molecular methods to demonstrate species and strain evolution of acetic acid bacteria population during wine production. *Int. J. Food Microbiol.* **102**, 295–304 (2005).
32. Grandvalet, C. Oenococcus oeni: queen of the cellar, nightmare of geneticists. *Microbiology* **163**, 297–299 (2017).
33. Pan, W., Jussier, D., Terrade, N., Yada, R. Y. & de Orduña, R. M. Kinetics of sugars, organic acids and acetaldehyde during simultaneous yeast-bacterial fermentations of white wine at different pH values. *Food Res. Int.* **44**, 660–666 (2011).
34. Schlatter, D. C., Yin, C., Hulbert, S. & Paulitz, T. C. Core Rhizosphere microbiomes of dryland wheat are influenced by location and land use history. *Appl. Environ. Microbiol.* **1**, 4. https://doi.org/10.1128/AEM.02135-19 (2020).
35. Wei, Y.-J. *et al.* High-throughput sequencing of microbial community diversity in soil, grapes, leaves, grape juice and wine of grapevine from China. *PLoS ONE* **13**, e0193097 (2018).
36. Gallego, V., Sánchez-Porro, C., García, M. T. & Ventosa, A. *Massilia aurea* sp. nov., isolated from drinking water. *Int. J. Syst. Evol. Microbiol.* **56**, 2449–2453 (2006).
37. Takeuchi, M., Hamana, K. & Hiraishi, A. Proposal of the genus Sphingomonas sensu stricto and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.* **51**, 1405–1417 (2001).
38. Martins, G. *et al.* Characterization of epiphytic bacterial communities from grapes, leaves, bark and soil of grapevine plants grown, and their relations. *PLoS ONE* **8**, e73013 (2013).
39. Coertze, S., Holz, G. & Sadie, A. Germination and establishment of infection on grape berries by single airborne conidia of *Botrytis cinerea*. *Plant Dis.* **85**, 668–677 (2001).
40. Kolattukudy, P. E. Enzymatic penetration of the plant cuticle by fungal pathogens. *Annu. Rev. Phytopathol.* **23**, 223–250 (1985).
41. Williamson, B., Tudzynski, B., Tudzynski, P. & van Kan, J. A. L. Botrytis cinerea: the cause of grey mould disease. *Mol. Plant Pathol.* **8**, 561–580 (2007).
42. Ioriatti, C. *et al.* Drosophila suzukii (Diptera: Drosophilidae) and its potential impact to wine grapes during harvest in two cool climate wine grape production regions. *J. Econ. Entomol.* **108**, 1148–1155 (2015).
43. Ioriatti, C. *et al.* Drosophila suzukii (Diptera: Drosophilidae) contributes to the development of sour rot in grape. *J. Econ. Entomol.* **111**, 283–292 (2018).
44. Carbone, K. & Fiordiponti, L. Colour evaluation, bioactive compound content, phenolic acid profiles and in vitro biological activity of Passerina del Frusinate white wines: influence of pre-fermentative skin contact times. *Molecules* **21**, 960 (2016).
45. Bene, Z. & Kállay, M. Polyphenol contents of skin-contact fermented white wines. *Acta Aliment.* **48**, 515–524 (2019).
46. Salemnia, S., Garcia-Torres, R., Herman, D. & Fajardo-Lira, C. Red, White, And...Orange? A New Look into an Old Wine (P20-007-19). *Curr. Dev. Nutr.* **3** (2019).
47. Darias-Martín JJ, Rodríguez OM, Díaz ES, Lamuela-Raventós RM. Effect of skin contact on the antioxidant phenolics in white wine. *Food Chem.* **71**, 483–487 (2000).
48. Lang, C. P., Merkt, N., Klaiber, I., Pfannstiel, J. & Zörb, C. Different forms of nitrogen application affect metabolite patterns in grapevine leaves and the sensory of wine. *Plant Physiol. Biochem.* **143**, 308–319 (2019).
49. von Rosenvinge, E. C. *et al.* Immune status, antibiotic medication and pH are associated with changes in the stomach fluid microbiota. *ISME J.* **7**, 1354–1366 (2013).
50. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **108**, 4516–4522 (2011).
51. Fadrosh, D. W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).
52. Wurm, P. *et al.* Qualitative and quantitative DNA- and RNA-based analysis of the bacterial stomach microbiota in humans, mice, and gerbils. *mSystems.* https://doi.org/10.1128/mSystems.00262-18 (2018).
53. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
54. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet. J.* **17**, 10–12 (2011).
56. Bushnell, B., Rood, J. & Singer, E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
57. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
58. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
60. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS ONE* **8**, e67019 (2013).
61. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Austral Ecol.* **18**, 117–143 (1993).

Acknowledgements

We thank Johanna Beeck and Frederic Fürstenberger for producing the wines at the University of Hohenheim Viticulture Unit.

Author contributions

Study Design: C.Z., N.M., W.F.F.; data acquisition: A.M.B., A.J., C.M., E.D., L.P., M.K.; data analysis: A.M.B., C.A., D.P.; writing of manuscript: A.M.B., W.F.F. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-70535-8.

Correspondence and requests for materials should be addressed to W.F.F.

Appendix 2: What is new and relevant for sequencing-based microbiome research? A mini-review

Alena M. Fricker*, Daniel Podlesny* & W. Florian Fricke (2019). What is new and relevant for sequencing-based microbiome research? A mini-review. *Journal of Advanced Research*, Vol 19, https://doi.org/10.1016/j.jare.2019.03.006

* A.M. Fricker and D. Podlesny contributed equally to this work.

Personal contribution:

I contributed to this review by writing several sections and the revision of the manuscript.

Credit author contributions: Writing of manuscript: A.M.F., D.P. and W.F.F.

Journal of Advanced Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Review

What is new and relevant for sequencing-based microbiome research? A mini-review

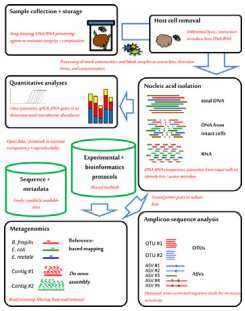
Alena M. Fricker [a,1], Daniel Podlesny [a,1], W. Florian Fricke [a,b,*]

[a] Dept. of Microbiome Research and Applied Bioinformatics, Institute for Nutritional Sciences, University of Hohenheim, Stuttgart, Germany
[b] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

HIGHLIGHTS

- Sample storage and nucleic acid isolation influence microbiota compositions.
- Error-corrected amplicon sequence variants (ASVs) improve 16S rRNA analysis.
- Contamination and host cells confound and complicate microbiota analysis.
- Quantitative and active microbiota analyses can complement existing methods.
- Open data and protocol sharing increases transparency and reproducibility.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 December 2018
Revised 20 March 2019
Accepted 20 March 2019
Available online xxxx

Keywords:
Microbiome
Contamination
Amplicon sequence variant
Quantitative profiling
Active microbiota
Open data

ABSTRACT

Microbiome research has transformed the scientific landscape, as reflected by the exponential increase in microbiome-related publications from many different disciplines. Host-associated microbial communities play a role for almost all aspects of human, animal and plant biology and health. Consequently, there are tremendous expectations for the development of new clinical, agricultural and biotechnological applications of microbiome research. However, the field continues to be largely shaped by descriptive studies, the mechanistic understanding of microbiome functions for their hosts remains fragmentary, and direct applications of microbiome research are lacking. The aim of this review is therefore to provide a general introduction to the technical opportunities and challenges of microbiome research, as well as to make experimental and bioinformatic recommendations, i.e. (i) to avoid, reduce and assess the confounding effects of sample storage, nucleic acid isolation and microbial contamination; (ii) to minimize non-microbial contributions in host-associated microbiome samples; (iii) to sharpen the focus on physiologically relevant microbiome features by distinguishing signals from metabolically active and inactive or dead microbes and by adopting quantitative methods; and (iv) to enforce open data and protocol policies in order increase the transparency, reproducibility and credibility of the field.

Introduction

Most microbiome projects today apply large-scale parallel sequencing to taxonomically and functionally characterize previously described and not-yet-cultivated, uncharacterized

Peer review under responsibility of Cairo University.
* Corresponding author.
E-mail address: w.florian.fricke@uni-hohenheim.de (W.F. Fricke).
[1] A.M. Fricker and D. Podlesny contributed equally to this work.

https://doi.org/10.1016/j.jare.2019.03.006
2090-1232/

microorganisms. The widespread application of high-throughput genomic approaches has been afforded by next-generation sequencing platforms that are easy to install and maintain. In addition, widely established experimental and bioinformatic protocols exist for sample processing, nucleic acid isolation, sequence target amplification, library preparation, sequence data processing and statistical analysis. Other high-throughput methods for system-wide microbiome analyses, such as metaproteomics or metabolomics/metabonomics [1], are less well established and widely used but are often successfully combined with genomics for systems-level approaches to simultaneously study different aspects of the microbiome. Cultivation-based isolation and characterization of individual microorganisms from microbiome samples can further complement nucleic acid sequencing-based and other 'omic approaches [2]. In the following, the microbiota will be referred to as the 'assemblage of microorganisms present in a defined environment' and the microbiome as the 'entire habitat, including the microorganisms . . . , their genomes . . . , and the surrounding environmental conditions' [1]. As sequencing-based microbiome analysis continues to be the most popular technique across the field, this review focuses on the discussion of experimental and bioinformatic aspects of this approach to highlight current problems and pitfalls as well as future chances and possibilities (Fig. 1).

Genomics and bioinformatics techniques of microbiome analysis

Sequencing-based characterizations of entire microbial communities, as well as their individual components and functions in unprecedented detail, is largely afforded by two main techniques: amplicon sequencing and metagenomics. The first method generates taxonomic compositional microbiota profiles at relatively moderate costs that allow even small research groups to run

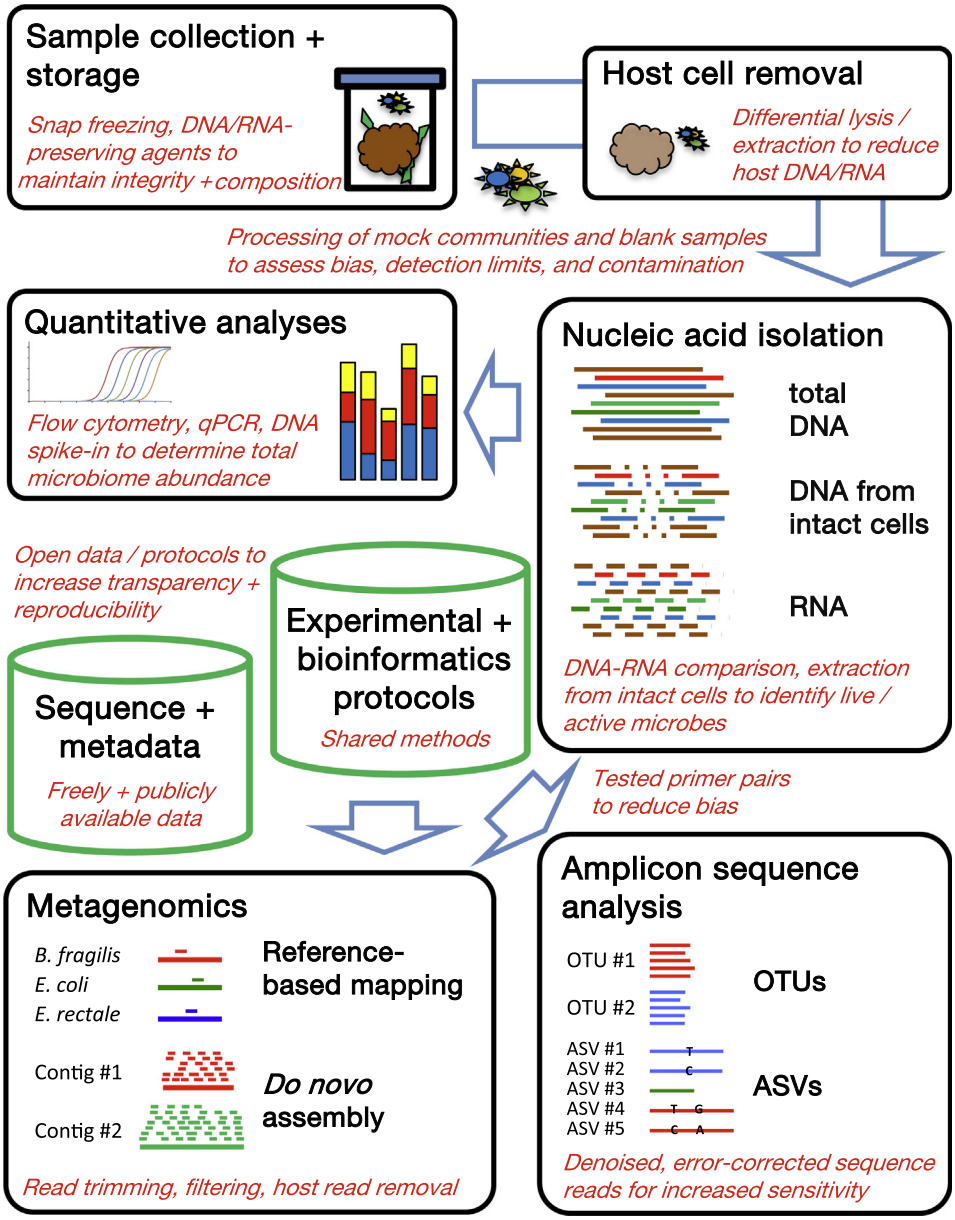


Fig. 1. Overview of recommendations for improved sequence-based microbiome analysis. Important technical components of typical laboratory and bioinformatic microbiome analysis projects (black boxes) and the bioinformatic resources that are generated in these projects (green columns) are shown, together with specific recommendations to expand and improve existing protocols (in red). Abbreviations: qPCR, quantitative real-time PCR; OTUs, operational taxonomic units; ASVs, amplicon sequence variants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

large-scale bacterial microbiota analysis projects. The latter method generally affords a more comprehensive, but also more costly, taxonomic and functional analysis of the entire viral, bacterial and eukaryotic microbiota [3]. Both approaches have been scaled up to include thousands of samples in a single study. Best practice recommendations for microbiome analysis, including laboratory and bioinformatic procedures are available, for example, from the U.S. Microbiome Quality Control [4] project.

Taxonomic microbiome profiling by amplicon sequencing

Amplicon sequencing methods rely on the selective binding of universal primer pairs to highly conserved regions within the genomes of specific microbiome members of interest and the sequencing of the resulting PCR products, which encompass taxon-specific hypervariable regions [5]. The most commonly used target amplicon for microbiome analysis is the bacterial 16S rRNA gene, but universal primer pairs have also been described for archaeal and eukaryotic small subunit ribosomal RNA genes, internal transcribed spacers (ITS) of the fungal and other ribosomal RNA operons and other conserved genomic loci [6]. Within the bacterial 16S rRNA gene numerous primer combinations have been proposed to amplify different hypervariable regions and to generate PCR products of variable lengths suitable for different sequencing platforms (e.g., Pacific Biosciences vs. Illumina) [5]. However, even "universal" primers can preferentially bind specific bacterial taxa, leading to compositional study biases that vary between microbiome types (e.g. gut vs. vaginal microbiome) and should be considered in the project planning phase [7,8].

Sequence variations in 16S and 18S rRNA genes, ITS regions and other metagenomic loci contain phylogenetic information that can be used to infer the taxonomic relationships of their microbial hosts. However, natural genetic variations are not easily distinguishable from sequencing errors, which even on the relatively accurate Illumina sequencing platform affects ~0.1% of all sequenced nucleotides [9]. Given the scale of current microbiome studies, bioinformatic protocols therefore have to account for millions of wrong base calls per project.

For amplicon sequencing-based microbiota analysis, sequences are traditionally clustered into *operational taxonomic units* (OTUs) based on arbitrarily defined thresholds of sequence similarity. For example, 16S rRNA gene fragments of >97% sequence identity are clustered into separate OTUs that reflect the phylogenetic boundaries of distinct bacterial species. Sequence clustering can be guided by bacterial reference genomes, yet common methods often also include *de novo* clustering to identify previously unknown species [10]. OTU picking assigns similar, but slightly different sequences to the same taxon, assuming a shared biological origin. Clustering therefore diminishes the impact of technical variation on the analysis results, but at the expense of reduced sensitivity in detecting biological variation. Fungal microbiota analysis by ITS amplicon sequencing follows similar principles as bacterial 16S rRNA analysis but sequence clustering and classification are complicated by inconsistent amplicon lengths and varying sequence similarities between fungal species [11]. The UNITE project represents an effort to generate a resource to represent the growing, known diversity of ITS sequence data [12], similar to the well-established SILVA database for pro- and eukaryotic small and large subunit rRNA genes [13].

To differentiate between biological and technical sequence variations, reference-free statistical denoising methods such as Deblur or Dada2 [14,15] have recently been implemented in QIIME2, a popular open-source software package for 16S rRNA analysis [16]. These tools generate error profiles of amplicon sequence datasets, which are then used to resolve sequencing errors and achieve single-nucleotide resolution for each amplicon sequence. Compared to OTU-based approaches, analysis of the resulting *amplicon sequence variants* (ASVs) provides improved sensitivity and specificity and reduces the problem of inflated microbiota datasets due to falsely identified distinct OTUs originating from mis-clustered sequences [17]. In addition, OTU clustering results are bound by the specific sequence data from which they were inferred and are therefore non-reproducible with modified or expanded datasets. The latest denoising algorithms overcome this limitation by recovering independent biological sequences as ASVs, fostering the reproducibility and comparability of amplicon-based microbiome analysis [18].

Taxonomic and functional profiling of the entire microbiome by metagenomics

Metagenomics uses the whole-genome shotgun approach to fragment and sequence the entire DNA of a microbiome sample instead of 16S rRNA gene fragments or other target amplicons alone. Correspondingly, the generated reads can originate from phages, viruses, bacteria, archaea, fungi and other eukaryotes and include plasmids and other extra-chromosomal elements as well as host, chloroplast and mitochondrial DNA. Compared to 16S rRNA analysis, this method needs significantly more data to obtain the sequencing depth that is required to identify and characterize rare microbiota members, often reaching several terabases per study and increasing costs and bioinformatic demands. However, as metagenomics potentially allows for functional microbiota characterization and, in theory, affords taxonomic resolution down to the level of individual microbial strains, it has become increasingly popular in microbiome research [19].

Quality control measures for metagenomic shotgun sequencing with new tools, such as KneadData, combine quality-based metagenomic read trimming and filtering with the bioinformatic detection and removal of human, plant and other eukaryotic host DNA (http://huttenhower.sph.harvard.edu/kneaddata). Metagenomic sequence data are typically analysed either by *de novo* assembly or by comparing reads individually to reference databases in a mapping-based process [20]. The *de novo* assembly of microbial genomes can help identify and comprehensively characterize previously unknown members of the microbiota [21]. However, because assembly requires substantial sequencing depth, assembly-based methods are typically restricted to the genomic reconstruction of highly abundant microbiome members. Marker gene-based sequence mapping with tools such as MetaPhlAn2 can be used for taxonomic profiling of entire microbial communities, including rare microbiome members [22].

Microbiome sample handling and processing

Maintaining microbiome integrity during sample collection and storage

Among many other factors, the accuracy of sequencing-based microbiota analysis depends on how well the original structure of the microbial community can be preserved between the time of sample collection and processing. Distinct members of the human, plant and environmental microbiota respond differently to extended periods of sample storage by dying or by suspending, retaining or increasing metabolic activity. Problematic artefacts for taxonomic or functional microbiota analysis can also arise from unintended disruption of the sample environment due to freeze–thaw cycles; exposure to oxygen, UV light, or osmotic stress; storage buffer components, etc. As a consequence, storage

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

conditions can affect microbiome analysis and lead to biased results [23].

Snap freezing of microbiome samples in liquid nitrogen and their long-term storage at −80 °C are generally considered as the gold standard for sample preservation [24]. However, commercial nucleic acid-preserving reagents and sampling kits that are used to maintain sample integrity in studies involving the collection of environmental samples or self-collected human specimens outside of the laboratory environment have generally been reviewed favourably [23]. Studies have suggested that temperature shifts alone have minor effects on taxonomic compositions and inter-individual differences in human gut microbiota analyses [24]. Chu et al. (2017) found the living bacterial microbiota of faecal samples to be most strongly affected by oxygen exposure, rather than by other factors, even repeated freeze–thaw cycles [25]. The same accounts for fungal microbiome samples, which are commonly stored with nucleic acid-preserving agents [26]. As mycorrhizal soil fungi colonize plant root tissues, the disruption of root connections after sampling can reduce mycorrhizal mycelial abundance and subsequently, induce the growth of mycelium-dependent other fungal opportunists, highlighting a specific protential problem for plant-associated fungal microbiota analysis [27].

Avoiding selective enrichment and depletion of microbes during nucleic acid isolation

Obtaining personalized gut microbiome analysis results from consumer microbiome testing services, journalist Tina Saey was surprised to receive substantially different results, particularly with respect to the relative abundance of the two dominant bacterial gut phyla *Firmicutes* and *Bacteroidetes* [28]. While numerous confounding factors might account for these observed variations, differences between nucleic acid isolation protocols have been known to introduce biases in taxonomic microbiota analysis. Even widely used commercial kits for DNA and RNA isolation differ in their efficiency in lysing specific microbes, including Gram-positive and Gram-negative bacteria, such as *Firmicutes* and *Bacteroidetes*, respectively [29,30]. Host-associated and environmental microbiome samples typically contain heterogeneous mixtures of viral, archaeal and eukaryotic microorganisms, including live and dead, active and inactive, vegetative and sporulated cells; cellular debris; free nucleic acids and other macromolecules. Microbial lysis protocols differ in their capacity to break open these different types of microbial components for nucleic acid isolation. Humic acids, melanin, polysaccharides, polyphenols and other sample components can interfere with DNA and RNA isolation and downstream applications, such as nucleic acid amplification or concentration determination [31].

Most microbiome analysis protocols include combinations of physical and enzymatic disruptions of microbial cells for nucleic acid isolation [4], which can be amended based on project-specific requirements, e.g., by adding specific polysaccharide-degrading enzymes such as lyticase for fungal microbiome analysis projects [32]. However, protocol variations lead to study-specific biases, which is one reason for the scarcity of meta-analyses of microbiome data [33–35]; these meta-analyses have had trouble with, for example, the identification of universal, disease-specific biomarkers across separate human microbiome studies. Depending on the microbiome sample type and specific microbial taxa of interest, testing and evaluating different nucleic acid extraction protocols on mock communities of diverse, defined microbial composition should be part of the early project planning phase. But project-specific technical biases are difficult to completely avoid, and consistency of the applied methods within specific microbiome studies might be most useful and practical.

Reducing, assessing and characterizing microbiome contamination

The interpretation of microbiome data can be complicated by contamination from sources other than the original sample [36]. The high sensitivity of sequencing-based microbiome analysis, particularly 16S rRNA gene amplicon sequencing, in detecting previously unknown, rare, and often non-cultivable microbiome members can also be problematic when contamination leads to false positive results. Laboratory consumables, reagents and even DNA extraction kits contain trace amounts of microbial DNA, and to some extent, sample collection, handling and processing always lead to low-level contamination [37,38]. Salter et al. (2014) ran microbiome analyses on serial dilutions of the same clonal culture of *Salmonella bongori* and identified a diverse microbiome that included both environmental and host-associated bacteria from the human skin and gut [37]. Importantly, the relative abundance of bacterial signals from contamination was positively correlated with the dilution factor of the original culture, demonstrating that the microbiome signal from contamination becomes more significant with decreasing amounts of sample starting material. Thus, contamination is less relevant for the analysis of faecal or soil samples of high microbial density than for host-associated human or plant microbiome studies of low microbial biomass, such as skin and vaginal swabs, tissue biopsies, urine, and the phyllosphere [39,40].

A prominent example of a controversially discussed microbiome finding concerns the placenta [41]. While several prominent publications reported on the presence of a unique placental microbiome in clinically asymptomatic women [42,43], these reports have been challenged as contradicting the paradigm of a tightly immune-controlled sterile womb and the practice of surgically removing sterile mouse pups from pregnant mice to generate germ-free mice [41]. Lauder et al. (2016) compared human placenta samples with vaginal swabs and experimental controls, including sterile and 'air swabs', and found the bacterial density and taxonomic composition of the healthy placental samples to be indistinguishable from those of microbiome-negative controls [44].

A three-tiered approach has been proposed to address the contamination problem [36]: First, good laboratory practice measures can reduce the chance of contamination when handling and preparing microbiome samples. This includes using purified, DNA-free reagents and kits, whenever possible, as well as spatially separating sample processing and DNA isolation, PCR setup and subsequent steps in the lab. Besides bacterial cells and genomic DNA from environmental sources, amplified PCR products can pose an important laboratory source of contamination for 16S rRNA analysis [37]. Second, the extent of contamination should be assessed by including technical replicates and internal controls in every step of the sample preparation protocol. Negative, microbiome-free, extraction controls and positive controls of microbial mock communities in defined concentrations can be used to determine the upper and lower limits of detection. Third, contamination controls should be sequenced and analysed together with the biological samples to characterize the influence of contamination on analysis results. For example, similarities between microbiome profiles of biological samples and negative controls can be quantified to compare the effect sizes of biological findings against contamination signals. However, the general exclusion of putative contamination signals from the analysis, by removing taxa from negative controls, can also distort microbiome analysis results and should be avoided. As contamination often originates from the laboratory environment, it can be directly influenced by related projects and include microbial signals that are similar to those from the original samples [37].

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

Reducing the impact of host DNA

Non-microbial DNA from human, animal or plant hosts is another major concern for sequencing-based microbiome analysis. Inadequate removal of host DNA can significantly increase the cost of host-associated microbiome projects or even make them practically impossible if the sequencing effort to obtain sufficient coverage of the microbial metagenome becomes prohibitively large. Healthy human faeces typically contain <10% human DNA, but up to 90% of sequence reads from low-microbial biomass samples such as saliva, nasopharyngeal, skin and vaginal swabs can be assigned to the human host [40]. While bacterial concentrations in urine increase during bladder infection, the concomitant increase in host DNA from epithelial cell damage can complicate microbiome analyses [Fricke, unpublished data]. As chloroplast and mitochondrial genomes from eukaryotic cells also carry 16S rRNA genes, host DNA can be problematic for 16S rRNA analysis, especially for food or plant microbiome projects [45]. Finally, host sequence removal may be mandatory before newly generated sequence data can be released in public databases to secure the privacy and confidentiality of human study participants, as required by most journals and funding agencies prior to publication, or to protect proprietary information from genetically modified or patented crops.

The relative level of host DNA can be reduced experimentally, either by removing host cells before DNA extraction or by selectively enriching microbial DNA after DNA extraction, or host DNA can be deleted bioinformatically by identifying and removing host reads from resulting sequence data, as described above. To remove host cells before DNA extraction, differential lysis can be used to selectively release and degrade host DNA before microbial (bacterial and fungal) DNA is isolated since mammalian cells are less robust than most microbial cells [46]. Density gradient centrifugation has also been used to separate host tissue from bacterial cells in plant samples [47]. However, microbiome samples, such as human faeces, also contain free microbial DNA from dead bacteria or cells that were disrupted during sample collection or storage, and certain microbes may be more susceptible to eukaryotic lysis regimens than others. Therefore, differential lysis protocols can reduce total yields of isolate nucleic acids [48] and bias subsequent compositional microbiota analyses towards specific taxa such as hard-to-lyse gram-positive bacteria [49]. Commercial solutions have become available to detect and remove vertebrate DNA by binding methylated CpG sequence motifs, which are abundant in eukaryotic but rare in microbial genomic DNA [50]. The latter method has been used to enrich bacterial and protist DNA for subsequent analysis of human and fish samples [50]. As an alternative approach to reduce the number of host-derived, non-bacterial PCR products, Lundberg et al. (2013) developed synthetic oligomers that bind as peptide nucleic acid (PNA) PCR clamps specifically to plant chloroplast and mitochondrial 16S rRNA gene sequences and block them from amplification [45]. In a similar approach, Agler et al. (2016) used specific nested primers, or "blocking oligos", inside the 16S rRNA gene of unwanted plant organelle DNA, to avoid amplification of the full-length PCR product for subsequent analyses [51].

New perspectives: Quantitative analysis and identification of active microbes

Adopting methods for quantitative microbiome profiling

Without accounting for potential differences in absolute microbial abundance between samples, the vast majority of microbiome projects today aim to characterize microbial communities based on compositional data [52]. These studies typically determine fractions of an unknown total number of microbial species, 16S rRNA gene copies, and other taxon-specific genes or functional gene categories. Unfortunately, compositional data tend to be misinterpreted as suggesting absolute shifts, reductions or increases in specific microbial taxa, gene functions or other microbiome parameters. Changes in absolute abundance of microbiome features can be biologically and clinically relevant, e.g. in small intestinal bacterial overgrowth (SIBO) [53], but tend to be ignored in standard microbiota projects. Vandeputte et al. (2017) found the bacterial load of human faeces to vary between healthy people and in individuals over time and bacterial density correlated with faecal enterotype [54]. Moreover, the authors demonstrated that quantitative microbiota profiling can change clinical perspectives. In this case, compared to previous reports based on relative faecal microbiota profiling, different bacterial taxa could be identified as specific biomarkers for inflammatory bowel diseases [54].

Different experimental approaches have been proposed to gather quantitative microbiome information, including cell counting by flow cytometry [54], quantitative or real-time PCR of the universal bacterial 16S rRNA gene [55] and normalization of bacterial relative abundances based on defined cell numbers that are spiked into the samples before nucleic acid isolation [56]. While the first approach is technically more demanding, commercial kits have become available to easily integrate quantitative analyses into microbiome project workflows. Importantly, sequencing depth, i.e., the number of reads assigned to each sample after 16S rRNA gene amplicon sequencing, should not be used to infer quantitative information, as inconsistent read counts between samples are typically technical artefacts that do not reflect quantitative differences [54]. However, the sequencing depth per sample does affect the alpha- and beta-diversity parameters of the microbiota and should be controlled, e.g., by bioinformatically rarefying read counts to equal numbers prior to statistical analysis [57].

Differentiating between total and active microbes

Sequencing-based microbiome studies typically rely on DNA as sole evidence for the existence of a microbiota in a sample. However, DNA from dying cells or spores or cell-free DNA in a sample may be evidence for microbial contact, but it does not necessarily indicate microbial life and an active microbiota in the sample. For example, the existence of a blood microbiome remains controversial, despite PCR-based evidence for bacterial 16S rRNA genes in blood DNA extracts from non-septic individuals, as attempts to culture bacteria from the same samples have mostly been unsuccessful [58]. While bacterial adaptation to the harsh conditions of the stomach has been demonstrated, metabolically active microbes in the stomach are difficult to distinguish from ingested, inactive microbes from other, adjacent body sites or food using DNA-based microbiota surveys alone [59]. To address this problem, a number of experimental and bioinformatic approaches have recently been proposed to identify metabolically active microbes reflective of a thriving microbiota.

Propidium monoazide (PMA) intercalates into double-stranded DNA, preventing it from being amplified by PCR and has been used by Chu et al. (2017) to remove free DNA from dead microbes prior to 16S rRNA gene amplicon sequencing [25]. Several groups have shown that 16S rRNA-based taxonomic microbiota compositions differ between RNA and DNA fractions isolated from the same sample [59]. This has been used to differentiate between transcriptionally active bacteria, which are identified on the basis of RNA evidence, and all other bacteria, which are identified on the basis of DNA evidence. Moreover, if DNA- and RNA-based analyses are combined with quantitative microbiota profiling, the ratio of 16S rRNA transcript-to-gene copies can be used to quantify

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

transcriptional activity and stratify bacterial taxa [59]. However, recent studies on soil bacteria also found 16S rRNA transcripts to remain stable for extended periods of time [60] and 16S rRNA gene and transcript compositions to be indistinguishable [61], suggesting that RNA-based methods to measure metabolic activity do not work equally well for all microbiome types. Importantly, experimental protocols need to support the simultaneous isolation of DNA and RNA from the same sample and extracted RNA should be carefully controlled for contamination with trace amounts of DNA, in order to avoid selectively enriching specific microbial taxa with separate lysis protocols or erroneously interpreting DNA-based signals as indicators of transcriptional activity, respectively [59].

An interesting approach to bioinformatically infer microbial growth rates from metagenome sequence data has been proposed by Korem et al. (2015) [62]. The authors demonstrated a positive correlation between bacterial growth and replication activity *in vitro* that is reflected by relatively increased concentrations of DNA from genomic regions around the origin compared to that from the terminus of replication. By mapping metagenomic sequence reads to bacterial reference genomes, a 'peak-to-trough' coverage ratio was calculated by comparing the origin and terminus DNA concentrations for each individual genome. This ratio was then used to stratify gut bacteria according to replicational activity and to statistically associate specific active bacteria with diseases such as inflammatory bowel disease and type II diabetes [62].

Release of published microbiome data and protocols

Microbiome research benefits from the availability of research data and protocols, and efforts should be made to establish and maintain open data and protocol policies across the entire field of microbiome research [63]. Progress in human microbiome research is increasingly driven by large, multi-centre studies based on the processing, sequencing and analysis of thousands of samples, often using custom laboratory and bioinformatic protocols to generate a statistical basis to detect subtle microbiome phenotypes. As a consequence, newly generated raw data and metadata, tools and protocols represent a substantial scientific resource to the broader research community that allows others to reproduce and expand published findings, recombine datasets for meta-analyses and develop new analytical approaches. For this reason, raw sequence and other omics data, associated sample metadata, and experimental and bioinformatic protocols for sample processing and analysis from published studies need to be made fully, freely and easily accessible. Accurate, detailed and complete bioinformatics analysis protocols should all scripts and precise commands that are needed to allow for full reproduction of raw data processing, data analysis and the generation of published figures. Although most funding agencies and journals in theory have set formal policies for data availability, access can be complicated in practice due to incomplete or inconsistent datasets, missing metadata information, and simple technical difficulties. Authors can be reluctant to comply with formal requirements that journals and funding agencies are struggling to enforce. Universal mandatory data and protocol release before manuscript submission would facilitate and improve peer review and allow journals to check for data availability as part of the submission process.

Conclusions and future perspectives

Microbiome research continues to excite both the scientific community and the public at large. However, the field has also been blamed for overselling findings and not producing reliable, applicable results [64]. While the mechanistic understanding of microbiota functions may yet remain too fragmentary to allow for the immediate development of diagnostic and therapeutic applications, there is little doubt about the general importance of human, animal and plant microbiomes for their hosts. To foster successful microbiome research in the future, it will be important for researchers, authors, reviewers, journals and funding agencies to (i) push the field towards the more widespread application of carefully controlled protocols for sample storage, nucleic acid isolation, contamination, amplification, sequencing and bioinformatic analysis; (ii) develop, optimize and standardize appropriate, improved analysis protocols; (iii) adopt and combine new experimental techniques, such as DNA- and RNA-based, relative and quantitative microbiota profiling; and (iv) increase the transparency and outreach of microbiome research by releasing data, metadata and protocols from published studies (Fig. 1).

Acknowledgements

Daniel Podlesny is supported by funds from the German Research Foundation (DFG) under project number 316130265.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with Ethics Requirement

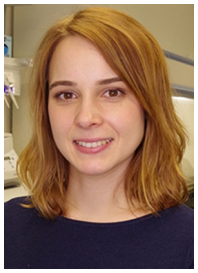
This article does not contain any studies with human or animal subjects.

References

[1] Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome 2015;3:31.
[2] Clavel T, Gomes-Neto JC, Lagkouvardos I, Ramer-Tait AE. Deciphering interactions between the gut microbiota and the immune system via microbial cultivation and minimal microbiomes. Immunol Rev 2017;279(1):8–22.
[3] Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. Nat Rev Microbiol 2015;13(6):360–72.
[4] Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 2017;35(11):1077–86.
[5] D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 2016;17:55.
[6] Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI, et al. Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. PLoS One 2013;8(2). e47879.
[7] Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, et al. Impact of preservation method and 16S rRNA hypervariable region on GutMicrobiota profiling. mSystems 2019;4(1).
[8] Graspeuntner S, Loeper N, Kunzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. Sci Rep 2018;8(1):9678.
[9] Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 2015;43(6). e37.
[10] Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ 2014;2. e545.
[11] Halwachs B, Madhusudhan N, Krause R, Nilsson RH, Moissl-Eichinger C, Hogenauer C, et al. Critical issues in mycobiota analysis. Front Microbiol 2017;8:180.
[12] Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res 2019;47(D1):D259–64.
[13] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41(Database issue). D590-6.

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

[14] Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2017;2(2).
[15] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13(7):581–3.
[16] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7(5):335–6.
[17] Kopylova E, Navas-Molina JA, Mercier C, Xu ZZ, Mahe F, He Y, et al. Open-source sequence clustering methods improve the state of the art. mSystems 2016;1(1).
[18] Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 2017;11 (12):2639–43.
[19] Garud NR, Good BH, Hallatschek O, Pollard KS. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol 2019;17(1). e3000102.
[20] Nayfach S, Pollard KS. Toward accurate and quantitative comparative metagenomics. Cell 2016;166(5):1103–16.
[21] Brown CT, Sharon I, Thomas BC, Castelle CJ, Morowitz MJ, Banfield JF. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. Microbiome 2013;1(1):30.
[22] Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 2015;12(10):902–3.
[23] Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep 2015;5:16350.
[24] Bundgaard-Nielsen C, Hagstrom S, Sorensen S. Interpersonal variations in gut microbiota profiles supersedes the effects of differing fecal storage conditions. Sci Rep 2018;8(1):17367.
[25] Chu ND, Smith MB, Perrotta AR, Kassam Z, Alm EJ. Profiling living bacteria informs preparation of fecal microbiota transplantations. PLoS One 2017;12 (1). e0170922.
[26] Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, et al. Identification of eukaryotic open reading frames in metagenomic cDNA libraries made from environmental samples. Appl Environ Microbiol 2006;72(1):135–43.
[27] Lindahl BD, de Boer W, Finlay RD. Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. ISME J 2010;4(7):872–81.
[28] Saey TH. The bizarre side of science [Internet]. Gory Details: Science News. 2014. Available from: https://www.sciencenews.org/blog/gory-details/here%E2%80%99s-poop-getting-your-gut-microbiome-analyzed.
[29] Dineen SM, Rt Aranda, Anders DL, Robertson JM. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. J Appl Microbiol 2010;109(6):1886–96.
[30] Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017;35(11):1069–76.
[31] Tebbe CC, Vahjen W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl Environ Microbiol 1993;59(8):2657–65.
[32] Leaw SN, Chang HC, Sun HF, Barton R, Bouchara JP, Chang TC. Identification of medically important yeast species by sequence analysis of the internal transcribed spacer regions. J Clin Microbiol 2006;44(3):693–9.
[33] Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 2017;8(1):1784.
[34] Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. MBio 2016;7(4).
[35] Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vazquez-Baeza Y, et al. Meta-analyses of studies of the human microbiota. Genome Res 2013;23(10):1704–14.
[36] Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: Issues and recommendations. Trends Microbiol 2018.
[37] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014;12:87.
[38] Hiergeist A, Reischl U. Priority program intestinal microbiota consortium/ quality assessment p, Gessner A. Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. Int J Med Microbiol 2016;306(5):334–42.
[39] Turner TR, James EK, Poole PS. The plant microbiome. Genome Biol 2013;14 (6):209.
[40] Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome 2018;6(1):42.
[41] Perez-Munoz ME, Arrieta MC, Ramer-Tait AE, Walter J. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. Microbiome 2017;5(1):48.
[42] Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci Transl Med 2014;6(237):237ra65.
[43] Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. Sci Rep 2016;6:23129.
[44] Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, et al. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome 2016;4(1):29.
[45] Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. Practical innovations for high-throughput amplicon sequencing. Nat Methods 2013;10 (10):999–1002.
[46] Hunter SJ, Easton S, Booth V, Henderson B, Wade WG, Ward JM. Selective removal of human DNA from metagenomic DNA samples extracted from dental plaque. J Basic Microbiol 2011;51(4):442–6.
[47] Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, et al. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc Natl Acad Sci U S A 2009;106(38):16428–33.
[48] Ferretti P, Farina S, Cristofolini M, Girolomoni G, Tett A, Segata N. Experimental metagenomics and ribosomal profiling of the human skin microbiome. Exp Dermatol 2017;26(3):211–9.
[49] Horz HP, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. J Microbiol Methods 2008;72 (1):98–102.
[50] Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. PLoS One 2013;8(10). e76096.
[51] Agler MT, Mari A, Dombrowski N, Haquard S, Kemen EM. New insights in host-associated microbial diversity with broad and accurate taxonomic resolution. bioRxiv; 2016.
[52] Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol 2016;26(5):322–9.
[53] Quigley EM. Small intestinal bacterial overgrowth: what it is and what it is not. Curr Opin Gastroenterol 2014;30(2):141–6.
[54] Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature 2017;551(7681):507–11.
[55] Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology 2002;148(Pt 1):257–66.
[56] Stammler F, Glasner J, Hiergeist A, Holler E, Weber D, Oefner PJ, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome 2016;4(1):28.
[57] Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 2017;5(1):27.
[58] Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory diseases. FEMS Microbiol Rev 2015;39(4):567–91.
[59] Wurm P, Dorner E, Kremer C, Spranger J, Maddox C, Halwachs B, et al. Qualitative and quantitative DNA- and RNA-based analysis of the bacterial stomach microbiota in humans, mice, and gerbils. mSystems 2018;3(6).
[60] Papp K, Mau RL, Hayer M, Koch BJ, Hungate BA, Schwartz E. Quantitative stable isotope probing with H2(18)O reveals that most bacterial taxa in soil synthesize new ribosomal RNA. ISME J 2018;12(12):3043–5.
[61] Papp K, Hungate BA, Schwartz E. Microbial rRNA synthesis and growth compared through quantitative stable isotope probing with H2(18)O. Appl Environ Microbiol 2018;84(8).
[62] Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science 2015;349(6252):1101–6.
[63] Langille MGI, Ravel J, Fricke WF. "Available upon request": not good enough for microbiome data! Microbiome 2018;6(1):8.
[64] Hanage WP. Microbiology: Microbiome science needs a healthy dose of scepticism. Nature 2014;512(7514):247–8.



Alena M. Fricker is a PhD student in the group of Dr. Fricke, in the Department of Microbiome Research and Applied Bioinformatics at the University of Hohenheim, Stuttgart, Germany. She completed her M.S. in Molecular Nutritional Sciences at the University of Hohenheim and is now pursuing a Ph.D. project in human microbiome research focusing on culture- and sequencing-based methods to study the influence of dietary interventions on the gut microbiota.

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

Daniel Podlesny also completed a M.S. in Molecular Nutritional Sciences at the University of Hohenheim and is now a PhD student in Dr. Fricke's group at the University of Hohenheim, Stuttgart, Germany. His work involves the development of new bioinformatic tools for strain-level metagenomic sequence analysis to study the microbiome dynamics in response to fecal transplantation.

W. Florian Fricke is Professor for Microbiome Research and Applied Bioinformatics at the University of Hohenheim in Stuttgart, Germany and Adjunct Assistant Professor at the Institute for Genome Sciences, University of Maryland School of Medicine in Baltimore, Maryland, USA. He received his PhD at the University of Goettingen, Germany, and was a post-doctoral fellow at The Institute for Genomic Research (TIGR), Rockville, Maryland. His group uses microbial genomics and applied bioinformatics to study the dynamics of the human gastrointestinal microbiota in the context of health and disease. His research focuses on the stomach microbiota, fecal microbiota transplantation and the development of functional biomarkers for host-microbe interactions.

Please cite this article as: A. M. Fricker, D. Podlesny and W. F. Fricke, What is new and relevant for sequencing-based microbiome research? A mini-review, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2019.03.006

Appendix 3: High intake of orange juice and cola differently affects metabolic risk in healthy subjects

Franziska Büsing, Franziska A. Haegele, Alessa Nas, Laura-Verena Doebert, **Alena Fricker**, Elisabeth Doerner, Daniel Podlesny, Julian Aschoff, Tobias Poehnl, Ralf Schweiggert, W. Florian Fricke, Reinhold Carle, Anja Bosy-Westphal (2018). High intake of orange juice and cola differently affects metabolic risk in healthy subjects. *Clinical Nutrition*, Vol 38, https://doi.org/10.1016/j.clnu.2018.02.028

Abstract:

Background: Higher consumption of sugar-containing beverages has been associated with an elevated risk of type 2 diabetes and gout. Whether this equally applies to cola with an unhealthy image and orange juice (OJ) having a healthy image remains unknown.

Methods: In order to investigate whether OJ and cola differently affect metabolic risk 26 healthy adults (24.7 ± 3.2 y; BMI 23.2 ± 3.3 kg/m2) participated in a 2x2-wk intervention and consumed either OJ or caffeine-free cola (20% Ereq as sugar from beverages) in-between 3 meals/d at ad libitum energy intake. Glycemic control, uric acid metabolism and gut microbiota were assessed as outcome parameters. Results: Fecal microbiota, body weight, basal and OGTT-derived insulin sensitivity remained unchanged in both intervention periods. Levels of uric acid were normal at baseline and did not change with 2-wk cola consumption (-0.03 ± 0.67 mg/dL; $p > 0.05$), whereas they decreased with OJ intervention (-0.43 ± 0.56 mg/dL; $p < 0.01$) due to increased uric acid excretion (þ130.2 ± 130.0 mg/d; $p < 0.001$). Compared to OJ, consumption of cola led to a higher daylong glycemia (DiAUC: 36.9 ± 83.2; $p < 0.05$), an increase in glucose variability (DMAGE-Index: 0.29 ± 0.44; $p < 0.05$), and a lower 24 h-insulin secretion (DC-peptide excretion: -31.76 ± 38.61 mg/d; $p < 0.001$), which may be explained by a decrease in serum potassium levels (-0.11 ± 0.24 mmol/L; $p < 0.05$).

Conclusion: Despite its sugar content, regular consumption of large amounts of OJ do not increase the risk of gout but may even contribute to lower uric acid levels. The etiology of impaired insulin secretion with cola consumption needs to be further investigated.

Personal contribution:

In this publication, I performed the 16S rRNA gene sequencing analysis, read processing following the QIIME pipeline and statistical analyses of the first cohort, which was examined between February and March 2016. I also reviewed and provided feedback on the manuscript.

Credit author contributions: Writing of the manuscript: ABW, FB, LVD, WFF; data acquisition: FB, FH, LVD, AF, ED, DP, AN, TP, RS; data analysis: FB, FH, LVD, AF; discussion of data and proof-reading of the manuscript: ABW, FB, FH, LVD, AF, ED, DP, WFF, JA, RS, RC; study design: ABW.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor Prof. Dr. W. Florian Fricke for giving me the opportunity to pursue my doctorate in his group and for his supervision over the past years. Your support, guidance, and the fruitful discussions we have had made this an inspiring experience for me. You have encouraged me to keep a flexible mind and to look at things from a different angle. Thank you for believing in me and appreciating my work.

I also want to thank my thesis committee members Prof. Dr. Thomas Kufer and Jun.-Prof. Dr. Amélia Camarinha Silva for taking the time as my mentors and to evaluate my work.

From the bottom of my heart, I would like to say a big thank you to all the group members of the Department of Microbiome and Applied Bioinformatics with whom I have worked with. Thanks to Daniel, Lisa, and Moni who welcomed and supported me from the very beginning. Daniel, thank you, especially for your patience and advice on bioinformatics and statistics. Working with you has been inspiring and your calm and pleasant manner has always been a rock in stormy seas. A special thanks also to Hannah R. and Nina, who made leaving this group even harder for me. I cannot imagine two better people to continue my work and hand over my precious projects to, although I wish we have met earlier. I am also immensely happy for all the friendships that I have been able to build during my time in this group. Christina, Hannah T., and Coco, I enjoyed working with you to the fullest and I am looking forward see you all succeed professionally and personally, which I am entirely sure you all will. My thanks also goes to Emily, Claudia, Sergio, and all of the other students, for times full of laughter and joy and the delightful cake and coffee breaks.

On a more personal note, I am deeply grateful to my family for their support and the freedom they give me to always pursue whatever feels best for me. Knowing that there is a safe place to return to whenever I'm in need has encouraged me to step out of my comfort zone multiple times.

I also thank my friends for all the encouragement and distraction they provided when I needed them the most.

Lastly, special thanks to my husband Roman, for your unconditional love, tireless support, and understanding. You've always been by my side no matter how stressed or enthusiastic I was. You've provided a shoulder to lean on and have been a steady part of my life, which I can always depend on.

Declaration in lieu of an oath on independent work

Declaration in lieu of an oath on independent work

according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences, Natural Sciences, and Business, Economics and Social Sciences

1. The dissertation submitted on the topic

> **Characterization of Dietary and Genetic Influences on the Gastrointestinal Microbiota**

is work done independently by me.

2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.

3. I did not use the assistance of a commercial doctoral placement or advising agency.

4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Stuttgart, 13.01.2023 _____

Place, Date Signature

Curriculum Vitae

PERSONAL INFORMATION

Name	**Alena Marie Bubeck** née Fricker
Place of birth	Stuttgart, Germany
Nationality	German

WORK EXPERIENCE

July 2018 - present	Research Scientist, University of Hohenheim, Germany Microbiome research in the Dept. for Microbiome and Applied Bioinformatics.
October 2017 – May 2018	Student Intern, F . Hoffmann-La Roche AG, Basel, Switzerland Pre-clinical R&D in Safety Pharmacology, Drug Disposition and Safety department within Pharma R&D Development
February 2014 – April 2014	Student Intern, Chemisches und Veterinäruntersuchungsamt Stuttgart, Fellbach, Germany Food safety inspections in the Dept. of Food Chemistry and Microbiology

EDUCATION

July 2018 - present	Doctoral Degree, University of Hohenheim, Germany Doctoral studies focused on the human intestinal microbiota and supervised by Prof. W. F. Fricke.
October 2012 – June 2018	Master of Science, University of Hohenheim, Germany Program: Molecular Nutritional Science. Thesis with a focus on dietary interventions on the intestinal microbiota: "Consumption of 20% of total energy requirement in the form of Coca Cola increases variance in taxonomic fecal microbiota diversity and composition compared to orange juice".
October 2012 – July 2011	Bachelor of Science, University of Hohenheim, Germany Program: Nutritional Science Thesis with a focus of inflammatory responses to wheat enzymes: “Wheat amylase trypsin inhibitors activate human intestinal mast cells”

SCIENTIFIC PRESENTATIONS

June 2019	12th Seeon Conference: Microbiota, Probiotics and Host. Seeon, Germany High/Low Sterol Conversion Is Associated with Different Microbiota and Metabolite Responses to Ketogenic Diet.
October 2015	7th European Mast Cell and Basophil Research Network (EMBRN) Congress, Marseille, France Wheat amylase trypsin inhibitors activate human intestinal mast cells.

PUBLICATIONS & PRE-PRINTS (FIRST-AUTHOR)

April 2022	Microbial cholesterol conversion and converter type-dependent LDL-C response to low-carbohydrate high-fat diet.
submitted	A.M. Bubeck, P. Urbain, C. Horn, A.S. Jung, L. Ferrari, H.K. Ruple, D. Podlesny, S. Zorn, J. Laupsa-Borge, C. Jensen, I. Lindseth, G. Arslan Lied, J. Dierkes, G. Mellgren, H. Bertz, S. Matysik, S. Krautbauer, G. Liebisch, H.F. Schoett, S.N. Dankel, W.F. Fricke. Under revision at *Cell Press Community*
August 2020	Bacterial microbiota diversity and composition in red and white wines correlate with plant-derived DNA contributions and botrytis infection. A.M. Bubeck, L. Preiss, A. Jung, E. Dörner, D. Podlesny, M. Kulis, C. Maddox, C. Arze, C. Zörb, N. Merkt and W.F. Fricke. *Scientific Reports*. doi: 10.1038/s41598-020-70535-8
March 2019	What is new and relevant for sequencing-based microbiome research? A mini- review. A.M. Fricker*, D. Podlesny*, W.F. Fricke. Journal of Advanced Research. *contributed equally. doi: 10.1016/j.jare.2019.03.006

Stuttgart, 13.01.2023

_____

Alena M. Bubeck